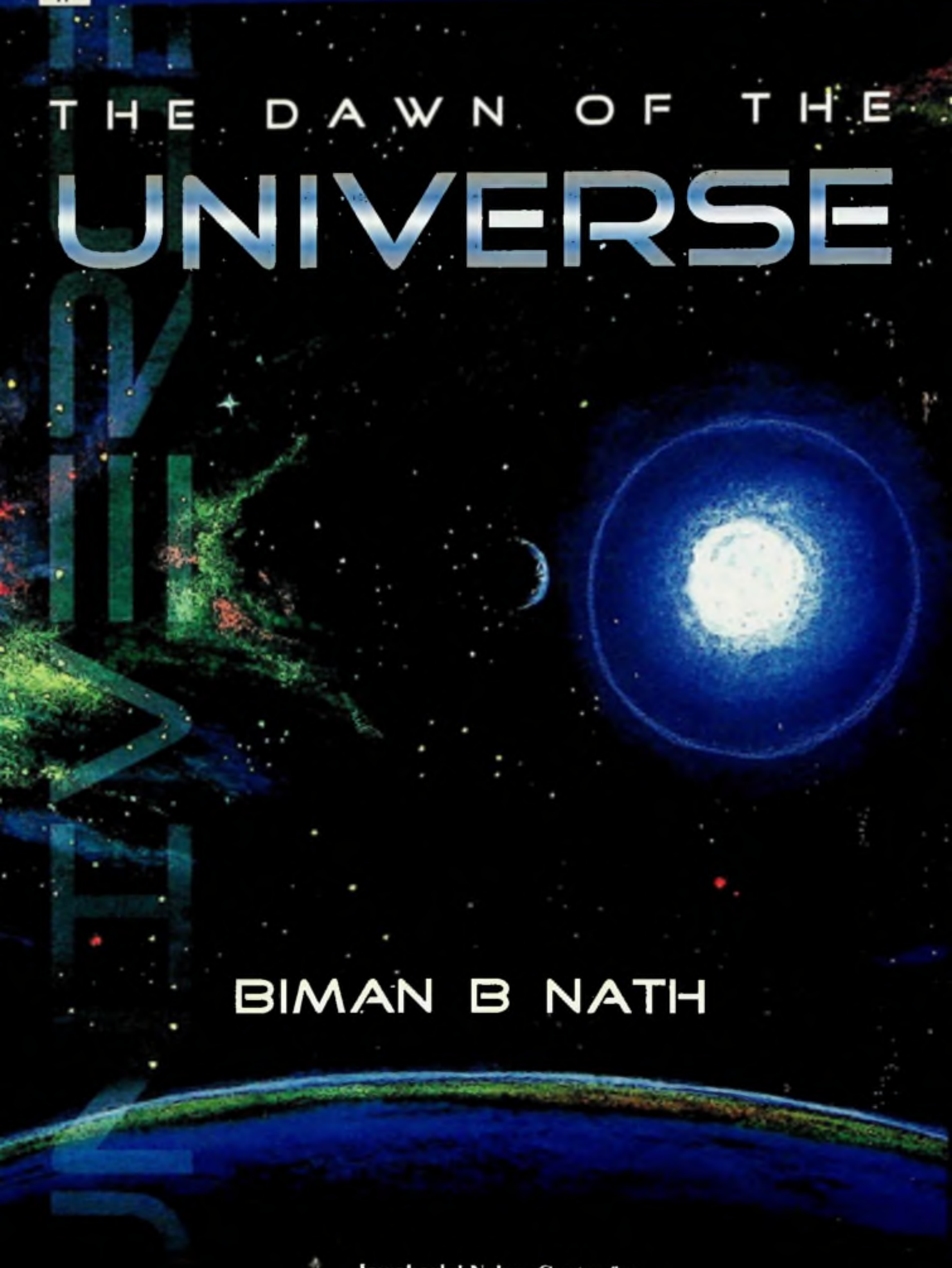
Universities Press

EDUCATIONAL MONOGRAPHS



THE DAWN OF THE UNIVERSE

BIMAN B NATH

THE DAWN OF THE UNIVERSE

THE DAWN OF THE UNIVERSE

BIMAN B NATH

Raman Research Institute
Bangalore

Jawaharlal Nehru Centre for
Advanced Scientific Research

Universities Press

Universities Press (India) Private Limited

Registered Office
3-5-819 Hyderguda, Hyderabad 500 029 (A.P.), India
e-mail: hyd2_upilco@sancharnet.in

Distributed by
Orient Longman Private Limited

Registered Office
3-6-752 Himayatnagar, Hyderabad 500 029 (A.P.), India

Other Offices
Bangalore / Bhopal / Bhubaneshwar / Chennai
Ernakulam / Guwahati / Hyderabad / Jaipur / Kolkata
Lucknow / Mumbai / New Delhi / Patna

First published in Bengali 2002
English translation from the Bengali by Biman B. Nath
First published in English 2005

ISBN 81 7371 520 3

Set in Times 10 on 12 by
OSDATA, Hyderabad 500 029

Printed in India at
Graphica Printers, Hyderabad.

Published by
Universities Press (India) Private Limited
3-5-819, Himayatnagar, Hyderabad 500 029

For my brother

Foreword

The Jawaharlal Nehru Centre for Advanced Scientific Research established by the Government of India in 1989 as part of the centenary celebrations of Pandit Jawaharlal Nehru, has completed fifteen years of its existence. The Centre is an autonomous institution devoted to advanced scientific research. It promotes programmes in chosen frontier areas of science and engineering and supports workshops and symposia in these areas. It also has programmes to encourage young talent. The Centre has now been recognised as a Deemed University by the University Grants Commission.

In addition to the above activities, the Centre has a programme of publishing high quality Educational Monographs written by leading scientists and engineers in the country addressed to students at the graduate and postgraduate levels, and the general research community. These are short accounts introducing the reader to interesting areas in science and engineering in an easy-to-read manner so that further study in greater depth and detail are facilitated.

This monograph is one of the series being brought out as part of the publication activities of the Centre. The Centre pays due attention to the choice of authors and subjects and style of presentation, to make these monographs attractive, interesting and useful to students as well as teachers. It is our hope that these publications will be received well both within and outside India.

M.R.S. Rao
President

Contents

Preface

This book tells the story of the emergence of the first luminous objects in the Universe. It is not intended for specialists in cosmology but for those who are interested in physics and cosmology and who may wish to learn more about the recent interesting developments in the field. Although it is written in a non-technical manner, without using equations as much as possible, a few mathematical derivations are sketched in the *Notes* for the curious reader.

The rise of cosmology as a science has almost been the proverbial rags-to-riches story. Few topics in science have seen a rise as spectacular. Even a few decades ago, it was often the subject of jokes, and its practitioners the object of derision and even pity. Most looked upon it as 'a science, but not Science' (*eine Wissenschaft, aber nicht Wissenschaft*), as Immanuel Kant had derided the study of chemistry in his time. Today, cosmology is an essential part of physics, and provides the context in which many new discoveries in astrophysics and particle physics can be understood, just as the concept of evolution provides a context for most biological studies.

It is hard to overemphasise the role of technological advances in the present revolution in cosmology. Recent developments in observational techniques have enabled cosmology to evolve from the data-starved enterprise that it was at the beginning of the last century to a data-enriched science. This aspect of modern cosmology has been highlighted in this book along with the presentation of theoretical ideas. In this regard, I have drawn the attention of the readers to the observational facilities which are being used in India and which will contribute more in the near future.

To begin with, in the first chapter, the reader is given an idea of the different structures present in the Universe, by embarking on an imaginary journey outward from the Earth. Basic ideas of the theory of relativity are sketched in the second chapter, as they are essential for understanding the concepts of cosmology. The third chapter discusses various aspects of the expansion of the Universe, and the next chapter briefly discusses the major events in the history of the Universe. These few chapters provide the background material that one requires to understand the formation and evolution of galaxies.

The fifth chapter describes the basic ideas of structure formation by gravitational instability. It also describes the way observations are being used to test these theoretical ideas. The next chapter discusses the formation of stars in galaxies and the emergence of the first stars and galaxies in the Universe. It also discusses various effects of these first-generation luminous objects, including some effects on the diffuse gas in the Universe which are being used as clues to learn more about this interesting epoch. The last chapter discusses a few aspects of the evolution of galaxies since then and a few other observational tools which have become important in modern cosmology.

I wrote a book on the same topic in Bengali a few years ago, and the present book grew out of the suggestions from a number of friends and colleagues to translate it for a wider readership. I have taken this as an opportunity to enlarge and update the content of the book.

I am indebted to Professor N. Mukunda for his encouragement and also for carefully reading the manuscript. Bikram Phookun was the first person to have read the whole manuscript and his comments have been extremely valuable. I thank both of them for wading through the morass of my English prose, and alerting me to glaring errors in it. I fear my revisions may not satisfy them completely, but I hope they will find the text much improved. I owe Palash Baran Pal a debt of gratitude for encouraging me to write the original Bengali book and editing it enthusiastically. I should hasten to add that none of them are responsible for any errors that still remain in book. I shall be most grateful to readers for pointing out any factual mistake.

I thank Shashi Bhushan Pandey, Richard Perley, David Mailn, National Astronomical Observatory of Japan, Space Telescope Science Institute, and many others for making available some of the images used in the book and giving permission to reproduce them.

I also thank Ms. Jebah S. David and others in the editorial staff of Universities Press for their patient editing and Madhu Reddy, Director, Universities Press for all the help during the publication process. I am also grateful to Professor M K Chandrashekaran (JNCASR) for his encouragement, and also to Jawaharlal Nehru Centre for Advanced Scientific Research for including this book in its Educational Monograph series.

It is to my elder brother, whose love for science was most inspiring during my childhood, that I dedicate this book.

Finally, I must thank Sourak for allowing me to spend more time on the book than with him. And special thanks to my wife, Shampa, for sharing the burden and joys of writing.

Biman Nath
Bangalore 2004

1

Our Place in the Universe

... He burned his house down for the fire insurance
And spent the proceeds on a telescope
To satisfy a lifelong curiosity
About our place among the infinities.

Robert Frost, *The Star-splitter*

Cosmology is the science of the Universe as a whole. Unlike astrophysicists who study stars and galaxies and other heavenly bodies, cosmologists wonder about the whole of the Universe – its origin, evolution and its possible future. By the term 'Universe' here one means the entity to which everything belongs. It is by no means easy for us to appreciate its vastness. The large distances that we are used to in our daily life correspond at the most to the size of the Earth. The circumference of the Earth is around 40000 kilometres. To comprehend the size of the Universe, we will need to deal with distances many times this tiny length scale. As a matter of fact, by the time the reader finishes reading this sentence (say, a minute), our Galaxy would have travelled a distance of 35000 km towards the nearest galaxy cluster. We have to get used to such vast numbers if we are to comprehend the size, or the age, of the Universe. We have to learn not only about the Solar System outside of the Earth, but also about the Milky Way that contains it, and even about the worlds outside it.

It will be a bewildering experience for a person to see the map of the whole world, if he has never travelled outside his town, or thought about the world outside. It will not be easy for him to place himself on this map. If instead, we first show him the map of the town he lives in, and then the map of the district that the town belongs to, and then that of the state and the country, and then the position of the country in the whole world, it will be easier for him to appreciate his position in the world. We would have to venture on a similar imaginary journey, if we wish to place ourselves in the Universe – to first get acquainted with the world in our vicinity and then look at the Universe from progressively bigger perspectives.

1.1 Solar System

We begin our journey outward with a tour of our nearby space. It has been known since ancient days that there are five luminous objects, apart from the Sun and the Moon, which wander around in the sky. The stars appear to be more or less fixed in the sky. These 'wanderers' are what we call 'planets' today—Mercury, Venus, Mars, Jupiter and Saturn. Since the invention of the telescope, three others have been added to this list, which are too faint to be seen with the naked eye—Uranus, Neptune and Pluto (Fig. 1.1).

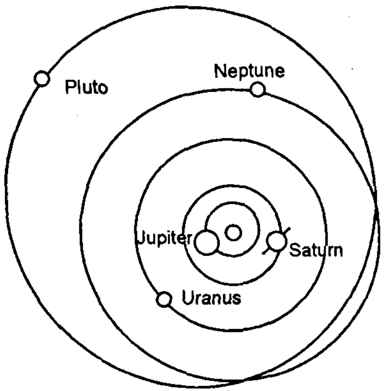


Fig. 1.1 *The orbits of the four outermost planets in our Solar System.*

The Greeks were the first to try to understand the rather complicated paths that these planets chalk out in the sky. Initially it was thought that they all circle around the Earth. It is common to begin with a simple assumption in trying to understand a phenomenon. It is not only the prerogative of scientists; it happens all the time in our daily lives. Suppose we find one of our fellow students absent someday. We would immediately think of some reasons behind it: perhaps he had taken ill, or there was an emergency at home. This is what we call a 'hypothesis' with which we try to explain our observation (of finding him absent). We may later call him at home to find out the actual reason, and that would be a test of our hypothesis – if it was true after all.

It was natural for the Greeks to start with the assumption that the planets had circular orbits, since this is the simplest assumption that one could make. In this case, the planet would always be at the same distance from the central object. Any other type of orbit will have to be more complicated than this. It is natural to begin with the simplest hypothesis, and make them progressively more complicated, if the simplest one fails the tests. For example, in the earlier example, if we find out from the phone call (the test of the hypothesis) that our friend was neither ill nor was there an emergency at home, we would then think of some complicated hypothesis: perhaps he had suddenly decided on a day trip somewhere. We would then find out if this new hypothesis were true.

The ancient Greeks had a similar situation with the planets. They found that the planets move about in the sky in a far more complicated pattern than could be predicted from simple circular orbits around the Earth. In fact, some planets, such as Mars, have the habit of turning around completely after moving in a seemingly straight path. After reversing its path, Mars then does another somersault after a few months and continues in the original direction. The Greeks were compelled to think of more complicated orbits to explain such bizarre loops in the sky. They thought of smaller circles superposed on the circular orbit, but that made the orbits look quite odd.

Finally Kepler showed that the easiest way to understand the movements of the planets was to let them move in elliptical orbits around the Sun. The initial hesitation to displace Earth from the centre of the Solar System was brushed aside, once Galileo's observations brought forth evidence for the new look of the Solar System. He observed that Venus showed waxing and waning phases like the Moon. One is then forced to keep the Sun at the centre to explain this. Such an observation could never have been explained by the system of the Greeks. There was also a reluctance to think of non-circular orbits, but Newton showed mathematically that ellipses are more natural than circles in the case of a force like gravitation[1].

We then have nine such planets moving around the Sun in our Solar System. Apart from the planets, there are also smaller objects, which sometimes show up as meteors or comets when they collide with or come near the Earth. Consider the size of this system of planets. We are about 1500,00,000 km away from the Sun. (Throughout the book, we will write large numbers like this as 1.5×10^8, which means 1.5 multiplied by 100 million. This way of writing it immediately gives a rough idea of the largeness of the number by looking at the power of ten.) Pluto is about 40 times more distant (from the Sun) than this. As far as time span is concerned, Pluto takes about 250 Earth years to go around the Sun once.

We will have to go far beyond the Solar System if we want to find the neighbouring stars. Let us first find out what is known about stars before we encounter the world of stars and our Galaxy.

1.2 Stars

Most stars are essentially giant balls of hot gas. A large fraction of this gas is hydrogen, the most abundant element in the Universe, and a smaller fraction, helium. The other heavier elements are present in much smaller quantities. These spheres of gas are bound by their own gravity. In fact, the gravitational force due to their own mass is so large that there is a tremendous pressure inside the sphere of gas. This high pressure is what makes the gas very hot. We encounter this relation between the pressure and temperature of a gas when we use a bicycle pump; as we put pressure on it, the air inside becomes hot. The temperature of the gas in the centre of our Sun is very high, about 15 million degrees Celsius[2].

The central temperature of a star depends on its mass. If the star is very massive then the inward pressure due to gravity is also large, and consequently the central temperature is high. The temperature of the outer layers is somewhat lower than the central temperature but is still indicative of the mass of the star. For example, in the Sun, while the central temperature is 15 million degrees, the outermost layers are about 6000 degrees hot. A star more massive than the Sun will have higher central and surface temperature than this.

One can estimate the mass of a star from the colour of its light, since the colour depends on the temperature of its outer layers. We know that as the temperature of an object rises, it first appears red, then yellow and blue at very high temperatures. The air near the wick of a candle is very hot and it appears blue there. The air a little outside is slightly cooler than this and is orange or yellow in this part. By the same reasoning, massive stars appear to be blue and less massive stars look redder. For example, the star Aldebaran in the constellation of Taurus is redder and less massive than Vega in the Lyra constellation[3].

The high central temperatures inside stars makes it possible to ignite thermonuclear reactions, as in a hydrogen bomb. Let us look at this process in detail. When the temperature becomes higher than a few thousand degrees, hydrogen atoms become 'ionised', which means that they get robbed of their electrons. The hydrogen atom is the simplest of all atoms; it has a proton at its centre and an electron moving around it (Fig. 1.2). The central proton holds the electron by electrostatic attraction, since they have opposite electrical charges. At high temperatures, atoms move around vigorously and they often collide. These collisions become more frequent as the temperature rises, and protons can lose their hold on the electrons as a result of these collisions[4].

Higher temperatures than this can have far more serious consequences. Usually the protons avoid each other, since they have similar electrical charge (which repel one another). In other words, there is a barrier to their coming too close to one another. At very high temperatures, though, an interesting thing happens following the laws of quantum mechanics. When the protons happen to come close to one another, there is a possibility that they may 'tunnel' through this barrier and merge with one another. Two protons can then produce a heavier particle, a deuteron, with a proton and a neutron. This deuteron can then react with two other protons in succession and finally produce a helium nucleus, with two protons and two neutrons (Fig. 1.2).

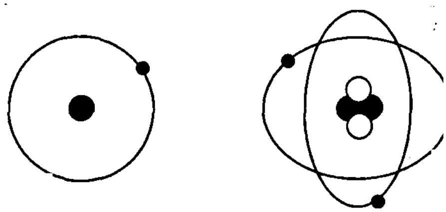


Fig. 1.2 *The structures of hydrogen and helium atoms are shown schematically.*

The interesting part of this set of reactions is that the mass of the resulting helium nucleus is slightly less than the total mass of four individual protons (Fig. 1.3)[5]. This is a common occurrence in nature. It is as if one was paying tax to Nature for living together by shedding some mass. We also know that mass is equivalent to energy ($E = mc^2$) as Einstein had shown. When four protons, therefore, combine to form a helium nucleus, the deficit mass then appears as energy. This is the main idea behind the hydrogen bomb, and such reactions take place inside stars all the time. This energy is radiated in the form of light which makes the stars shine.

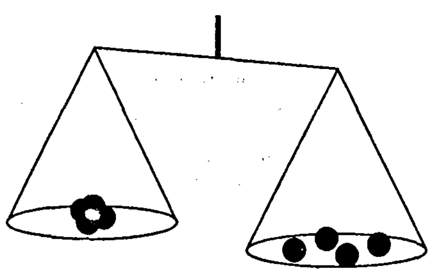


Fig. 1.3 *The combined mass of four protons (on the right) is larger than the mass of a helium nucleus produced by thermonuclear fusion (on the left). The mass discrepancy is approximately 7 parts in a thousand, and is released as radiation energy.*

The production of this thermonuclear energy in the centres of stars helps them to counteract the inward pressure due to gravity. A balance is struck between these two opposite forces, one trying to contract and the other trying to expand the star. This balance keeps the star in a steady state. Most stars in the sky shine in this manner.

This state of steady shining can however continue only for a limited duration. The reservoir of energy available for thermonuclear burning is naturally finite. Our Sun, for example, has been shining in this manner for the last 4.5×10^9 years. It has lost much of the initial amount of hydrogen in the centre in this process, having converted it to helium. Its supply of hydrogen fuel will naturally come to an end at some point of time, although the Sun still has enough to shine for about 5×10^9 years more.

When the fuel for thermonuclear reactions at the centre of stars is used up, the balance between the outward and inward pressures is undone. There is no more production of thermonuclear energy and there is nothing to stop the inward pressure due to gravity. The centre of a star then contracts. This increases the density of gas in the star. The ultimate fate then depends on the total mass of the star.

When matter becomes very dense, the laws of quantum mechanics become important in determining its behaviour. Particles like electrons avoid being too close to one another in the quantum world. At high densities, when electrons are forced to be closely packed, they therefore repel one another, giving rise to an outward pressure[6,7]. This pressure can hold out against the inward pressure of gravity only if the star is not too massive compared to our Sun[8]. Stars in this state can therefore stay in a steady state even without any thermonuclear reactions. The star cools slowly by radiation from the hot outer layer. These are called 'white dwarf' stars; they are called white because the outer layers are hot enough to radiate light that is more on the blue side than the red, and they are dwarfs because they are tiny compared to normal stars.

Our Sun, for example, will shrink to a size comparable to that of the Earth when it becomes a white dwarf at the end of its thermonuclear burning stage (its present diameter being about hundred times that of the Earth). Being tiny in size means that its density will be very high. The average density of gas in the Sun is close to that of water. White dwarfs are much denser than this; a spoonful of white dwarf matter will weigh a few hundred kilograms.

It is possible for this unusual pressure from electrons to balance the gravitational pressure only for low mass stars. The gravitational force of stars much more massive than the Sun is impossible to be won over in this way. Stars, in this case, keep on contracting and becoming denser. At some stage, the density becomes so large that protons and electrons are forced to combine to form neutrons (which are slightly more massive than protons but have no electrical charge). The whole star then becomes an ocean of neutrons. Neutrons and electrons are similar in one respect. Neutrons, like electrons, avoid one another when packed at high density. They then provide an outward pressure, just as electrons do for white dwarfs. These stars whose gravitational pressure is kept in balance by pressure from neutrons are called 'neutron stars'. They are even smaller in size than white dwarfs, with a typical neutron star being the size of a city (with a diameter of around ten kilometres). They are also very highly dense; a spoonful of neutron star matter will weigh as much as the Himalayas.

Neutrons stars can stay peacefully in this manner for a long time, as white dwarfs do. But not all stars are so lucky. Objects that have masses more than 3–4 times that of the Sun are doomed to a tragic end. No known physical process can then stop the influence of gravity in these objects and they shrink to become what scientists call a 'black hole'. The gravitational force in these objects is so large than nothing, not even light, which has the largest speed in the universe, can escape. Without any radiation coming out of these objects, they naturally appear black to us. We only get to see the signs of their presence, when a star or some other object comes too close for comfort, and are sucked up by the monstrous gravity of the black hole.

To summarise what we have learned about stars so far, most stars shine as a result of thermonuclear reactions in them. At the end of this phase, they reach their final stages, of a white dwarf, or neutron star or a black hole, depending on their mass.

The way to the final stage also depends on the mass of the star. For stars like the Sun, the centre will contract at the end of the thermonuclear burning stage, and the outer layer will expand. The outer surface will then look red and the Sun will appear to have a gigantic size. At this stage, our Sun will be larger than the orbits of Mercury and Venus, and the temperature on Earth will be too high for life to survive.

There are many such geriatric stars in the sky. Aldebaran, Betelguese and Antares are all 'red giants' like these (they do appear red to the naked eye). In some cases, the star cannot hold on to its outer layer any longer because of the diminished force of gravity at the periphery after its expansion, and the gaseous matter there spreads out in space. We see this gas as a ring, with the central star having contracted to the white dwarf stage. The famous 'Ring nebula' in the constellation of Lyra (near Vega) is a fine example of this (Fig. 1.4).

The tug of war between gravity and energy from thermonuclear reactions is somewhat protracted for very massive stars. The reactions do not stop with the production of helium from hydrogen. It then begins to burn helium to form heavier elements from it. Massive stars have higher temperatures in their cores and nuclear reactions can then make more and more complicated elements. First the helium nuclei combine to form carbon. When the helium fuel gets over, the core starts to burn carbon to form nitrogen. Other heavier elements like oxygen, sodium and so on are then created one after another.

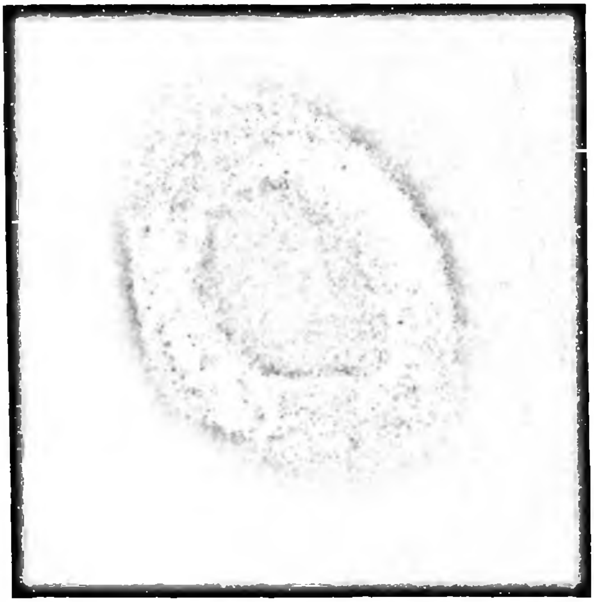


Fig. 1.4 *Ring nebula in the Lyra constellation. The diffuse gas in the form of a shell has been flung from the outer layers of a red giant star, whose core now shines as a white dwarf star in the centre. (Picture courtesy: STScI and NASA)*

A time comes, however, when it becomes impossible to extract energy from such nuclear reactions. This happens when iron is formed at the core of stars. It turns out that if iron is used in a nuclear reaction to form a heavier element than this, then the mass of the resulting nucleus is *not* smaller than the total mass of the burnt elements. In other words, one needs to add energy in this case and energy cannot be extracted from burning iron. When the stellar core is made of iron, the thermonuclear reactions come to an end.

This sudden pause means that gravity again wins the war between inward and outward pressure. We have seen that (the cores of) massive stars in this case contract to form either neutron stars or black holes, depending on the mass (of the core). This sudden withdrawal of the outward pressure wreaks havoc in the rest of the star, which collapses on to the frigid core, and the ensuing stampede sends a shockwave through the outer layer of the star, spitting it out into space in a gigantic explosion. All the elements created for eons inside the star are expelled into space. The heat of the explosion also induces more complicated nuclear reactions (which had stopped at iron), and creates elements like gold, silver, lead, mercury and others. Gas enriched with all these material is gurgled out into space. Such violent stellar funerals are called 'supernovae' (Fig. 1.5).

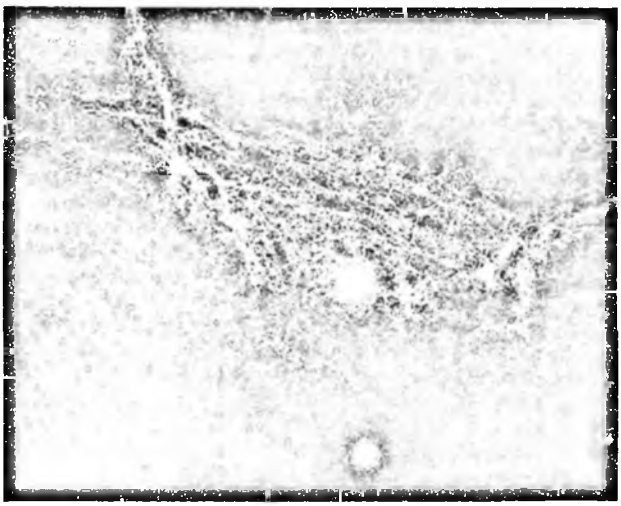


Fig. 1.5 *The remnants of a supernova explosion in the constellation of Vela. The explosion occurred 11000 years ago and the shell of ionised gas will take a few million years to slow down and cool (@ Anglo-Australian Observatory/ David Malin Images).*

This is not just science fiction. Astronomers routinely observe such explosions in the sky, which are so powerful that for a while they can be among the brightest objects in the sky. Their peak brightness can equal that of billions of normal stars. This luminosity makes them detectable from large distances. There are also historical records of such supernovae in the past. Chinese astronomers were surprised to see a new bright object in the sky in 1054 AD, in the constellation of Cancer[9]. They recorded it as a 'guest star' and their records say that it was bright for a few days before dimming away slowly. In 1987, astronomers caught a nearby star red-handed when it exploded. Their careful observation showed that nuclear reactions during the explosion created elements heavier than iron, like cobalt and lead, just as theoreticians had predicted earlier.

Such explosions throughout the age of the universe have seeded it with important elements that are essential for life. The carbon in our cells, the calcium in our bones, the oxygen in the air – everything has been cooked in the stellar furnaces, and have been spread out with the help of supernovae. Newer worlds like our Solar System have been created from the ashes of such explosions.

Let us briefly recapitulate what we have learned about the 'lives' and 'deaths' of stars. If the star is almost as massive as our Sun, then it will expand after a few billion years of running thermonuclear reactions and will turn into a giant red star. Its core will then become a white dwarf and the outer layer will slowly disperse into space. If the star is very massive then it will create heavier elements than helium through nuclear reactions. It will then destroy itself in an explosion and spread these elements in space, while the core will turn into either a neutron star or a black hole.

What are the time scales over which these processes take place? We would do well to learn about the ages of stars. The age of a star depends on its energy resource and the way it spends its energy, that is, the way it radiates. The energy resource is simply proportional to the total mass (from the equivalence of mass and energy), whereas the brightness depends on the mass in a complicated way. Massive stars spend their energy in a wild manner. Stars that are ten times more massive than the Sun are ten thousand times brighter than the Sun. Their lives are then a thousand times shorter than that of the Sun. (If someone has ten times as much money as I have and spends it ten thousand times faster than I do, he would become a pauper a thousand times faster than I would.) We have learned that the life span of the Sun is about ten billion years. Stars ten times more massive than it, therefore, end their lives within a hundred million years in a supernova explosion.

It is important to learn what the astronomers know about the birth of stars if we wish to discuss the emergence of the first stars in the universe. Since stars have a finite life span, it must be that stars are being formed at the present epoch in our Galaxy. The Milky Way would have otherwise plunged into darkness after a few billion years—the lifespan of an average star—of its formation. The actual sites of star formation have long been sought after. Astronomers have recently found what we can call nurseries of stars, where new stars are being born out of gaseous clouds. The details of the process of star formation are still uncertain, but here is a brief summary of what has been found so far.

When we look at the starry sky at night, we might think that space is empty except for a number of stars. We now know that the space between the stars is not a vacuum. It is full of gas and dust, distributed at places in a diffuse manner, and elsewhere bundled up in giant, dense clouds. Even in the diffuse phase, the density of gas in the interstellar space is much smaller than that of air in our atmosphere. If the gas in our atmosphere could be diluted by some means by a factor of 10^{18} (a billion billion times), then it would become as light as the interstellar gas. This gas is often rendered hot after it has been shocked by the debris from a nearby supernova explosion. At many places, however, the gas is very cold, and has coagulated into a giant cloud. The density of gas is larger here than elsewhere. This dense gas is too cold to radiate photons of visible wavelengths; on the contrary it absorbs light from nearby stars.

In the night sky, one can easily find dark 'holes' in the otherwise bright distribution of stars. Stars are rare in these patches of the sky, and it appears as if some 'cloud' has obstructed our view of the background stars in these regions. These are the signatures of interstellar clouds. Many of these clouds are very large and dense, and are called molecular clouds. This is because most of the gas in these clouds is very cold and in the form of molecules. According to astronomers, stars are born in these giant molecular clouds. They have found evidence for this by observing the clouds with light rays that are not absorbed by the gas there. Infrared rays are the best tools for this type of observation. Astronomers have recently seen the innards of giant molecular clouds by observing the infrared rays radiated from them, and there are evidences of protostars shrouded by gas. At present, there is a lot of research going on to understand the important steps in the birth of stars in such a situation.

1.3 Milky Way

Consider now the world outside our Solar System and the neighbouring stars. Before embarking on such a journey, we should change our units of distances from kilometres to something else. The distances we have encountered so far are tiny compared to the average distances between stars. We certainly do not use inches and centimetres to measure distances between cities on Earth, whereas that might still be useful to discuss lengths of sentences printed in a book, for example. Astronomers commonly use the unit of 'parsec' while discussing distances in interstellar space. One parsec is approximately three light years, which means that light takes approximately three years to cross a parsec. In standard units, one parsec is approximately thirty trillion kilometres[10].

We have already encountered our nearest star which is the Sun. If we could view our Sun from a distance of a few parsecs, then we would notice a few other neighbouring stars. We would find that the typical distance between the stars is around a parsec. We would see Sirius and Altair on two sides of our Sun, and would find that it is immersed in a tenuous, hot gas that appears to blow in from a nearby association of stars (as shown in the top right corner of Fig. 1.6).

If we view the solar neighbourhood from a larger distance, say a thousand parsecs (a kilo parsec), we will then notice a few molecular clouds in the vicinity. These clouds are as big as a few hundred parsecs across. We will notice that stars seem to stay in groups. There are many such 'associations' in our neighbourhood region. For example, there is a large molecular cloud and a stellar association behind the star Sirius (in the constellation of Orion in our sky). This association can even be seen with a pair of good binoculars, towards the sword, south of the belt of Orion.

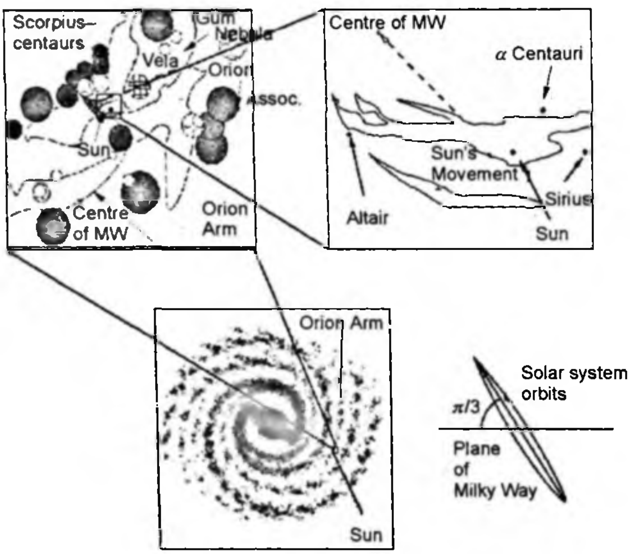


Fig. 1.6 *A schematic diagram showing the neighbourhood of the Sun and its place in the Milky Way. The upper right panel shows the region close to the Sun, with stars like Sirius and Altair seen along with hot gas engulfing the Sun. The size of this region is about twenty parsecs (Altair is 16.8 parsecs from us). The direction of the Sun's movement and the direction of the centre of the Milky Way are indicated by dotted arrows.*

This is embedded in a bigger region, shown on the top left, with a few molecular clouds (shown as dark circles) and stellar associations (shown as white circles filled with smaller circles). The Orion stellar association (labeled 'Orion Assoc') is seen at the right and the Scorpius–Centaurus association is at the top left corner. The remnants of the Vela supernova in the central region are embedded within a gaseous nebula called the Gum nebula (top right). The extent of this whole region is around 1000 light years.

This region is in turn contained in our Galaxy, the Milky Way, shown at the bottom. The sequence of pictures can also be viewed in the reverse manner, starting with the Orion arm of our Milky Way, and then zooming on this region to the top left and then again focussing on the solar neighbourhood in the top right corner.

Finally the plane of the Solar System is juxtaposed against that of the Milky Way. The plane of the orbits of most planets of the Solar System makes an angle of approximately 60 degrees with the plane of the Milky Way, as shown in the bottom right corner.

(We should remember that although Sirius and the stars in the Orion constellation appear close to one another in the sky, they are at different distances in space. They appear to be close in the sky because they are in similar directions when viewed from Earth.)

Furthermore, we will find hot gas from a nearby supernova explosion, whose remnants can be found in the constellation of Vela. We will also be able to locate the stellar association from where the hot gas engulfing our Sun has been blowing. This group of massive stars is in the constellation of Scorpio. Massive stars not only radiate their energy copiously, but also throw away chunks of hot gas which spread out in space.

We can now place our Solar System in a bigger perspective in the world of stars. Let us continue to zoom out to an even bigger perspective; we will find that our neighbourhood is an average community of stars in the backwaters of a massive conglomeration of stars, called the Milky Way.

Our Galaxy, the Milky Way, contains approximately one hundred billion stars. From a distance it would appear as a thin disc, with a bulge in the central region. From this bulge, a number of spiral 'arms' seem to spring out, like an octopus. Our sun is contained in one of these 'arms', and is around 8 kiloparsecs (approximately 24000 light years) away from the centre of this massive system of stars. We are almost near the rim of the disc of the Milky Way, tucked away in a suburb. The most interesting thing in this region seems to be the Orion molecular cloud and the stellar association near it, which is why the 'arm' to which we belong is called the 'Orion arm' of the Milky Way. The Sun revolves around the centre of the Milky Way with an approximate speed of 220 kilometres per second, which means that it takes around 200 million years for the Sun to complete a revolution.

We can see the Milky Way as a bright ribbon spanning the summer night sky on Earth. This is because the Milky Way has a shape of the disc and we are looking at it from inside the disc. This disc is very thin, with a thickness that is much smaller than its radius (the ratio being smaller than one-tenth). Viewed from the top it looks like a circle, and from the edge it looks like a thick line. We see the nearest stars in all directions as we live deep inside the disc. When we look at the farthest stars though, it is like viewing the disc from its edge, and we find them inside a thick line or a ribbon.

It is very fortunate for life on Earth that we are so far from the hustle and bustle of the centre of the Milky Way. It is not only that the density of stars is large there, with frequent interactions between them, but astronomers have also recently discovered a massive black hole in the centre. As a result of the violent interactions of gas and stars with this black hole, energy is radiated in a myriad wavelengths, some of them being harmful for the survival of any life form.

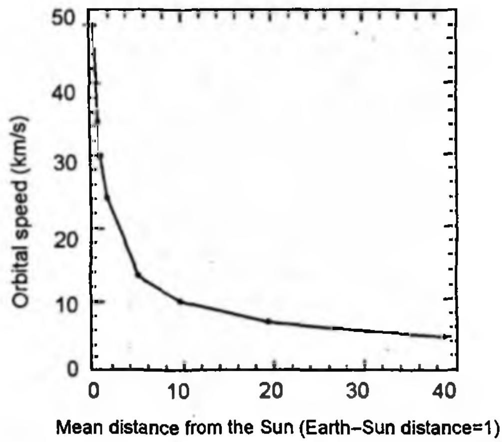


Fig. 1.7 *Orbital speeds of planets in the Solar System are plotted against the mean distance from the Sun, in the unit of the mean Earth–Sun distance.*

Before continuing with our journey outward, let us briefly find out about the mass of our Galaxy. It is not easy to determine the mass though. As a matter of fact, there are a few puzzling observations that have kept scientists worried for more than three decades now. An important aspect of our Universe has its seeds in this observation.

Consider a stone tied to a string whose other end is tied to a finger, and suppose the stone is being swung around the finger. The faster one tries to move the stone, the larger must be the force applied on the string. If the length of the string is increased somehow, a smaller force must be applied to get the stone rotated with the same speed. In other words, one can 'measure' the force applied on the string, by the speed of the stone and the length of the string. One can determine the magnitude of the force knowing these two values. Now, consider a similar situation with the stone being replaced by a planet, the finger by our Sun and the force on the string by the force of gravitation. We would then come to a similar conclusion that the speed and distance of the planet would give a measure of the gravitational force between the Sun and the planet.

We know from Newton's law of gravitation that this force decreases with the inverse square of the distance. If the distance is doubled, the force between two objects is reduced to a quarter. If a planet is moved to a greater distance from the Sun (increasing the length of the string in the earlier example) then the force on it will decrease and so will its speed around the Sun. This is true for all planets in. our Solar System, for example. Pluto moves along its orbit with a speed (about 5 kilometre a second) that is much less than that of Earth (about 30 kilometres per second; see Fig. 1.7).

Suppose we were to observe that the speed does not decrease with the distance of the planet. One way to explain the situation would be to discard our notion of gravitational force. Scientists however have a lot of faith in Newton's law of gravitation. Although Einstein modified it somewhat in the last century, in this situation both would have come to the same conclusion.

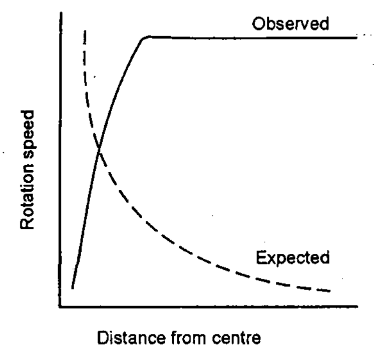


Fig. 1.8 *The observed relation between the rotational speed of stars and gas vs. the distance from the centre of Milky Way is shown by the dotted curve. The expected curve, assuming mass to be concentrated in the centre, is shown by the solid curve.*

There can then be only one reasonable explanation. Gravitational force not only depends on the distance, but also naturally on the gravitating mass. The only way to increase the force keeping the same distance is to increase the gravitating mass. We would have to distribute this extra mass in such a way as to compensate for the decrement of gravitational force with distance. If the speed of the planets around the Sun did not decrease with distance, it would then mean that there was a large amount of mass hidden somewhere in the Solar System; it must be invisible, otherwise we would have detected it anyway and would have included it in our calculation to begin with.

It turns out that astronomers have found a case of such additional but invisible mass in the case of stars revolving around the centre of the Milky Way. They have discovered that the speed of stars around the centre of the Milky Way does not decrease with the distance from the centre. As a matter of fact, the speed remains more or less a constant as one travels farther from the centre (Fig. 1.8). Leaving apart the central region, most stars (and gas) in our Milky Way revolve around the centre with a speed of approximately 220 km per second. If we use Newton's law of gravitation, then there is only one explanation for this observation. It must be that the mass of the Milky Way is not concentrated in its centre, but is distributed far and wide throughout the Galaxy. On the contrary, it is found that stars are actually concentrated towards the centre and their numbers decrease drastically as one goes away from the centre. This means that the additional mass needed to explain the observation of rotation speeds *cannot* be due to stars, but something else. It must be a material that does not shine well, but only gravitates[11].

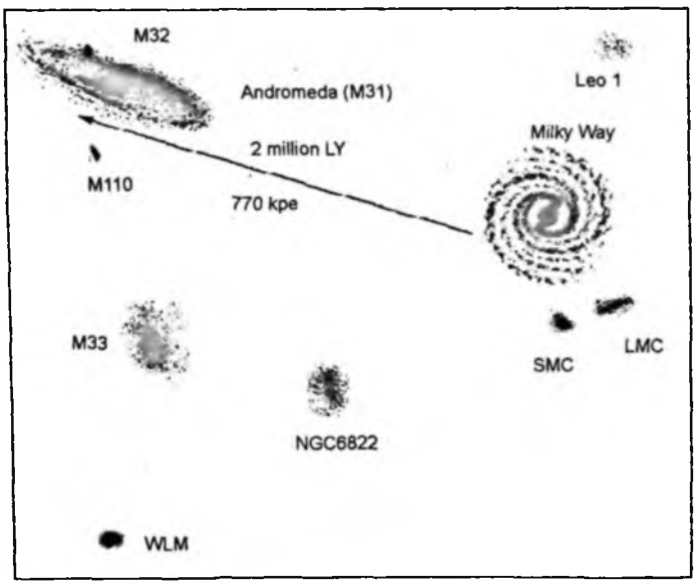


Fig. 1.9 *The Local Group of galaxies to which our Milky Way belongs. The immediate neighbours of Milky Way, the small and large Magellanic Clouds are shown (labelled SMC and LMC) on the right, and a group of galaxies near Andromeda galaxy is shown on the top left. A few other prominent members of the Local group are also shown as a collage of their images (sizes of individual galaxies are not to scale).*

This 'dark matter', as scientists call it, is still not very well understood. Its fundamental constituents and their nature are a topic of intense research at present. It is however now known that dark matter exists not only in our Galaxy, but in all the galaxies that astronomers have observed so far. It is not only bundled up in galaxies, dark matter also seems to be distributed throughout the universe[12]. We will discuss this in more detail in the later chapters. Let us summarise here what scientists know about the total mass and the size of our Galaxy.

Astronomers believe that although the visible matter of our Milky Way is contained in a thin disc, the dark matter is spread out much beyond it, perhaps in a large spherical region and containing the relatively small disc of stars and gas. It is possible that the size of the region containing the dark matter of our Galaxy is as large as a hundred kiloparsec. The distance of our Sun from the centre of the Galaxy, for comparison, is only about 8 kiloparsec. The total mass, including the visible and invisible matter, of our Galaxy is equivalent to about 150 billion Suns. (The mass of our Sun is 2×10^{30} kg, which is called a 'solar mass'. In this unit, the total mass of our Galaxy is 150 billion solar mass, i.e., 1.5×10^{11} $M_{\odot}$.)

1.4 Beyond the Milky Way

If we could somehow view our Milky Way from outside, what else will we notice around it? Firstly we will find two small galaxies very close, at a few hundred kiloparsecs (almost a billion light years) from the Milky Way. They have irregular shapes, which are not spherical or elliptical, nor disc shaped like our Galaxy. We can see them with the naked eye in the sky on Earth; they appear as two fuzzy 'clouds' in the sky from the Southern hemisphere. Magellan the explorer was the first to have recorded sighting them, and they are called the 'small Magellanic cloud' and 'large Magellanic cloud' in his honour. They are tiny compared to the Milky Way in their mass and they orbit our Galaxy being bound by its gravity.

A little beyond these two satellite galaxies of ours, we will find another spiral galaxy like ours, Andromeda, which is approximately 770 kiloparsecs from the Milky Way. Andromeda can also be seen with the naked eye on a clear night on Earth as a hazy fuzz. In many respects, Andromeda is a galaxy similar to ours, with a mass slightly larger than that of the Milky Way. We will then find that there are about two dozen small galaxies which move around due to the combined gravitational force of the Milky Way and Andromeda. The size of this assembly is approximately a million parsec (a megaparsec), and astronomers call this assembly of galaxies, 'the Local Group'(Fig. 1.9). Most galaxies in this group, apart from the two irregular shaped Magellanic clouds and the spiral galaxies like the Milky Way and Andromeda, have elliptical shapes and are small in size. The spiral and elliptical are the most common shapes of galaxies in the Universe.

In our discussion on our Milky Way, we found that there are associations of stars. Galaxies too tend to crowd into groups and associations and our Local Group of galaxies is an example of such an assembly. There are many such groups of galaxies in our Universe. Then there are regions with a much larger assembly of galaxies, a cluster of galaxies, with hundreds or even thousands of member galaxies. These clusters are often as large as a few megaparsecs. As a matter of fact, our Local Group of galaxies is situated close to such a cluster of galaxies, which can be seen from the Earth in the constellation of Virgo, and which is called the 'Virgo galaxy cluster' (Fig. 1.10). It is approximately 16 megaparsecs from us, and is pulling our Local Group of galaxies towards it with its gravitational force. The galaxies in our group are hurtling through space towards it with a mind-boggling speed of approximately 600 kilometres per second. This is the speed referred to at the beginning of this chapter.

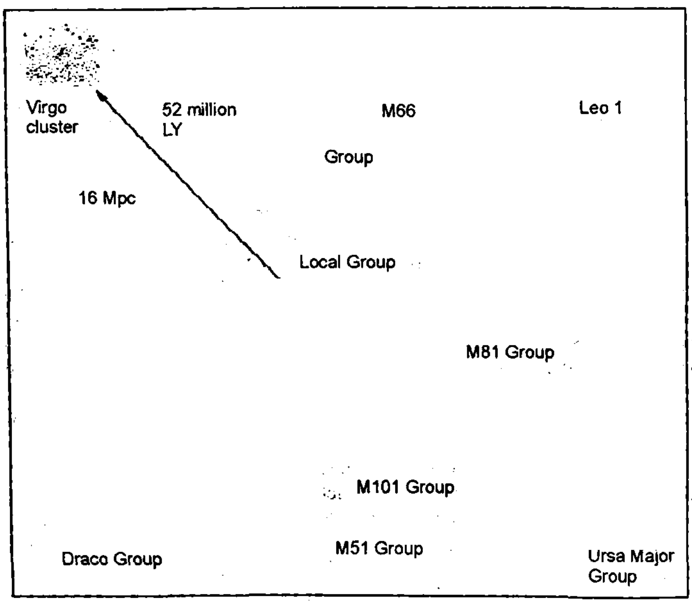


Fig. 1.10 *The environment of the Local Group of galaxies, with a few adjacent groups of galaxies and the Virgo cluster of galaxies (top left corner) is shown as a collage of images.*

One would be tempted to guess at this point that perhaps even clusters of galaxies combine to form even larger associations, probably superclusters of galaxies. A question then naturally arises whether or not this tendency of forming bigger and bigger groups continues without any end, or whether there is some limit to this tendency.

It is not easy to answer such questions. A map of a much bigger region of the universe is needed, which will give accurate positions of thousands of galaxies, and whose distribution in space will possibly give clues to the clustering tendencies of galaxies. This is by no means an easy task. It is difficult to estimate the distances of individual galaxies, especially when they are far away and appear dim in our sky. Technological advances, however, have helped astronomers enormously in this regard. They can now obtain the distance of several galaxies at the same time with the help of fibre optics technology. It now takes a few years instead of decades to draw up a map of the nearby Universe.

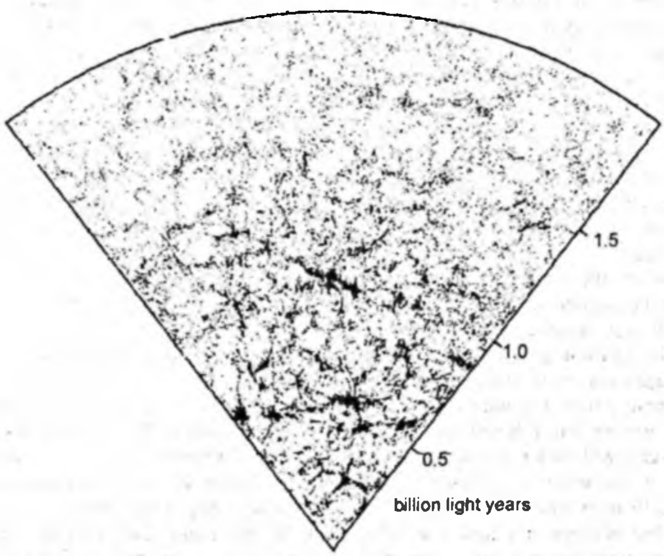


Fig. 1.11 *A part of the '2dF survey' map of the nearby Universe. Our Milky Way is the bottommost point in this map and every other point represents a galaxy. This map contains around 100000 galaxies (Picture courtesy: The 2dF Galaxy Redshift Survey Team).*

Astronomers have recently compiled the distances of approximately 220000 nearby galaxies, in a survey called the '2dF galaxy survey'. The individual. patches of the sky that were imaged had angular sizes of a couple of degrees, from which it derives its name. The central point in this map (Fig. 1.11) represents our Milky Way and every other point represents a galaxy with its distance from us indicated[13].

The 2dF map of galaxies clearly shows that galaxies are not spread out in the universe in an even manner. They form groups of various sizes; groups of galaxies combine to form clusters, which then forms superclusters of galaxies. It however appears from a careful study of the map that the tendency of clustering does not continue unabated, at least not with the same vigour.

One can quantify this clustering tendency with the help of the inhomogeneity in the density of galaxies, by which we mean the number of galaxies in a certain volume of the universe. If the galaxies were spread out evenly, the density of galaxies will then be a constant everywhere. Any deviation from constancy in this density will signify the clustering tendency. One can also attempt to measure this density averaged over regions of different sizes, which will signify the tendency of clustering for assemblies of different sizes. From a careful statistical study of the map of the nearby Universe, it seems that the density of galaxies averaged over rather small regions is a varying quantity, which means that there is a large non-uniformity in the distribution of galaxies when averaged over small regions. This non-uniformity, or inhomogeneity, however, decreases as the density is averaged over larger and larger regions. The inhomogeneity in the density of galaxies become less than a per cent when the distribution is averaged over length scales in excess of 100 megaparsecs. In other words, the universe looks more and more homogeneous, evenly spread, as one zooms out to view it from bigger and bigger perspectives[14].

The farthest galaxy in the 2dF map is approximately a few thousand megaparsecs away from us (10^9 parsec). If the estimates made by astronomers from the map of this region is confirmed, it will mean that the Universe beyond this will not be markedly different. As we have found, the Universe will look more or less the same if seen from very big perspectives. This is true even for the largest scales encountered within the periphery of the 2dF map and is expected to be true for scales larger than this.

Alert readers may find something amiss at this point. The map of the nearby universe may not really represent the map at the present instant. This is because of the finite speed of light. It takes about 8.5 minutes for light to reach us from the Sun, meaning that the Sun that we see at a given instant of time is actually the picture of the Sun as it was 8.5 minutes ago. When we see the light from the new born stars in the molecular clouds of Orion, it has taken more than a millennium to reach us. We would have to wait for approximately 2 million years to learn about any new phenomenon in the Andromeda galaxy. We may wonder if the Andromeda galaxy at the present instant is at the same position in space, or if it actually looks different. The questions become more pertinent when we consider the farthest galaxies in the 2dF map, for example. Light from these galaxies would have taken a few billion years to reach us. Has our discussion then been too naive? Is it fair to compare one part of the Universe with another which is separated not only in space but also in time? How would the evolution of the Universe affect this comparison? It is as if history and geography were intertwined when it comes to matters of the whole Universe.

We will first have to learn what is known about the history of the Universe to get a proper perspective to appreciate this puzzling question. Before we do so, we will need to briefly familiarise ourselves with the notions of relativity, since it is one of the pillars of modern cosmology.

2

The Relativistic Universe

No speed of wind or water rushing by
But you have speed far greater. You can climb
Back up a stream of radiance to the sky,
And back through history up the stream of time.
And you were given this swiftness, not for haste
Nor chiefly that you may go where you will,
But in the rush of everything to waste,
That you may have the power of standing still–

Robert Frost, *The Master speed*

One of the most important discoveries of the twentieth century is that of the expansion of the Universe. It is not only an example of how technological advancement has enriched the topic of astrophysics, but also of the enormous progress in theoretical science. Mankind had never considered, and never had to consider until then, a changing Universe. It had never occurred to him that the Universe could be both infinite and bound. These revolutionary thoughts became possible after the momentous discovery of the theory of relativity by Einstein. We will therefore have to understand the basic principles of relativity to appreciate the discoveries that have made cosmology a science.

It was in a paper written in 1905 that Einstein first formulated his theory of relativity. It was still not general enough to be applied to all cases, but was valid only for special situations, and is called the 'special theory of relativity'. It is applicable to frames of reference moving with constant velocity, which means when there is no acceleration. When Einstein discovered the 'general theory of relativity' almost a decade later, it was complete enough to include the effects of frames with acceleration. It also gave a novel way to think of gravitation, which made it possible to appreciate better the discoveries of modern astrophysics.

2.1 Special Theory of Relativity

The principle of relativity advocates that all inertial frames of reference are equivalent in describing Nature. The laws of physics which describe natural phenomena need not be changed when the observer changes its state of uniform motion. This basic principle of relativity is not new and dates back to Galileo who had noticed that physics according to different observers need not be different.

A simple example will perhaps explain this better. Consider the situation inside a cabin in a ship, with all its windows closed so that one cannot look out. Suppose the ship is moving with a constant speed, in the same direction, so that there is no acceleration. Let us also imagine that it is a completely calm sea and that there are no waves at all. In this situation, if we consider the motion of a ball thrown vertically upward inside the cabin, we can ask what will happen to the ball as it comes down. It is possible that the ball may come down to the same position with respect to the sea outside, which will mean that it will seem to fall backward to an observer in a closed cabin. It is also possible that it will come down to the same position relative to this observer. We can also think of a similar experiment in a bus or a train, but we should ensure that there is no acceleration, or, no wavering in the motion and no jolts.

Common sense says that the ball will fall back to the same position with respect to the observer in the cabin. The situation for this observer is equivalent to throwing a ball upward from a static ground since in the closed cabin moving with a constant speed, one cannot tell if one is moving at all. Since the motion of the ball also seems equivalent to the case of a static observer throwing a ball upward, it means that laws of physics are the same for all frames moving with constant speeds. Nature can be described by all frames of reference moving with constant speeds with respect to each other by the same laws. All such frames of reference are equivalent in this respect.

This much was known since Galileo. Einstein added a simple postulate to this principle of relativity. This revolutionary postulate has to do with the speed of light. We need to digress a bit here in order to appreciate the need for bringing in light and its speed. Physicists had faced a difficult problem at the end of the nineteenth century. As we have seen earlier, the observed motion of an object depends on the state of motion of the observer. If we observe a moving train from another moving train with the same speed, it will appear to be at rest with respect to us. We are used to this in our daily life. Physicists however had difficulties understanding certain experiments with the speed of light with this simple law of addition of speeds. They had to conclude that somehow the speed of light could not be added to or subtracted from the motion of the observer. This was strange indeed. In the game of cricket, the fast bowlers often have a long run up to increase the speed of the ball. The speed they acquire during the run up adds to the speed with which they would have bowled from rest. If instead of the ball they had to hurl a light beam, then it would not be advantageous to have a long run up, because the speed of light cannot be increased by hurling it from a moving body. In the imaginary game of light-cricket, everyone would bowl with the same speed, the speed of light.

This situation does not at all match our common sense expectation. Physicists therefore built some strange models by which light would *appear* to have the same speed, although the common expectation of addition and subtraction of velocities could be still applied to it. The clash between the theory of light, that is the theory of electromagnetism, and the laws of mechanics still sounded discordant to many. The fact that light did not obey the laws of mechanics meant that there was a disparity between these two important parts of physics. Einstein made his bold postulate in order to bring these two parts under the same umbrella.

In his paper in 1905, he put forth the suggestion that it is not just in appearance, but the speed of light is a universal constant. This hypothesis immediately solves all the problems faced by experimentalists. But if we accept this postulate, many of the common ideas of speed will have to be changed. Since speed means the amount of space covered in a given duration of time, it will also mean a revision of our ideas of space and time. We usually think of time and space as things that are the same for everyone. The distance measured between two points, or the duration of time between two events, are the same no matter how and by whom the measurements are done. We will soon find that these common ideas will have to be modified if we were to say that light has the same speed no matter who measures it.

Consider a simple clock made of two parallel mirrors and a photon (a particle of light) being reflected between them. Let us first find out what an observer sitting with the clock will find. According to him, the photon will successively reflect from the top and the bottom mirror, and thereby keep time like the swinging of a pendulum clock (Fig. 2.1, top part). If one knows the speed of light and the distance between the two mirrors, one can easily use this 'light-clock' to keep time by counting the number of times the photon has bounced back and forth between the mirrors. Now, suppose we put this clock on a moving train and observe it from outside the train. According to this second observer, light will appear to travel at an angle and reflect off the two mirrors (Fig. 2.1, lower part).

We however know from simple geometry that the diagonal path is longer than the vertical path between the mirrors. There are then two possibilities for the second observer outside the train. He will either think that light has a greater speed than when the train was static (by adding the train speed to that of light), or, he will think along the lines of Einstein and say that light had the same speed. In the first case, he will think that 'although the diagonal path is longer than the earlier path, light now has greater speed to compensate for it, and so the clock will keep the same time'. If Einstein were correct, then he will have to argue, 'since the diagonal path is longer and the speed of light is the same, it would take longer for the photon to reflect off from the mirrors, and the clock would therefore appear to run slow.'

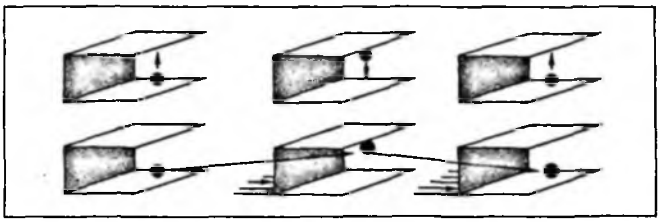


Fig. 2.1 *An imaginary clock to explain the relativity of time. The upper set of snapsh ts represent how the clock would appear to the first observer at rest, and the lower set, to the second observer.*

We then have to conclude that time will run slower in a moving clock than in a static one. The clock on the train will appear to run in slow-motion to the observer standing outside. It might be argued that we have a puzzling situation here because we have used a very unconventional clock in our example. It seems difficult to accept that something as constant as time could vary. Could time really wait for anyone?

To pursue this question, let us provide our observers with two ordinary wrist watches. According to the principle of relativity, the wrist watch will have to keep the same time as the 'light-clock' even when the train starts moving. Otherwise, from the difference in the behaviour of the two clocks before and after the train begins to move, the observer will be able to find out the speed of the train. In the principle of relativity, one should not be able to determine one's own state of motion by any physical experiment, since the laws of physics are the same for everyone (without acceleration)[1].

In other words, the wrist watch will always keep the same time as the light-clock. We therefore have no other alternative but to accept that time indeed flows with different rates for observers in different states of motion. However strange it might seem at first, we will find that everything moved slow in the moving train, including the heartbeat of the first observer sitting in the train. The moving train has indeed slowed down the flow of time.

It is not just time; Einstein proved that even our common perception of length will have to be changed. We could imagine a 'light-ruler' along the same line as our 'light-clock', to measure lengths with the help of the speed of light. By the same arguments as we have discussed earlier, we will find that things in the moving train will appear to be shorter to an observer standing outside the train.

The difference is however very very small for our daily life situations. The slowing down of time and shortening of rulers will be very insignificant for the speeds we are used to. According to the laws of special relativity, if the train has a speed of 60 km per hour, then the difference in the rate of the clocks inside and outside the train will be as small as 0.000002 per cent. It is not only impossible to perceive such differences, but also difficult to measure them, even by sensitive instruments. This is why we are used to thinking of space and time as being the same in different frames. The differences however become important at very high speeds.

One could cite the example of the decaying muons here. Muons are particles which decay (like radioactive materials) after a certain time and give rise to electron and neutrinos. The speed of muons created in laboratories is not very large and their decay time is about two-millionth of a second. In a loose way, one could think of muons as being suicidal and committing suicide after this 'life time' is over. But physicists observing muons moving at very high speed find something very different from this suicidal tendency. When muons move with a speed that is 99.5 per cent of the speed of light (approximately a billion km per hour), one finds that their decay time, or their 'life-time', is increased by a factor of ten. One can explain this phenomenon by the special theory of relativity, arguing that the clocks of moving muons run slower, and so by the time they commit suicide, a much longer time has elapsed in the clocks of the physicists observing them. From our point of view we therefore think that their decay time has increased by a large amount. This is an experimental proof of the slowing down of time.

2.2 Spacetime

According to Einstein, it is our common sense that is to blame if these new ideas of space and time sound preposterous to us. Since Newton's times, we have been led to think of space and time as being the stage on which the material phenomena of the world occur. They were supposed to only act as the background and not participate in or interfere with the phenomena taking place in them. According to Einstein, this background of space and time cannot be the same for everyone. Space and time therefore cannot be a passive background for events, but are involved in an active manner in the events. There is no escape from bringing them to the forefront of the arena of physical events.

What Einstein achieved was not only to change our notions of space and time, but also to bind them into a single entity in an unprecedented manner. He showed that time and space are parts of a much more fundamental entity called 'spacetime'. It is our habit of separating them into time and space that causes all the troubles with time dilation or length contraction at high speed.

Let us consider the example of a car that has a constant speed of, say, 60 km per hour. Its speed never changes although its direction of motion may change. Let us imagine this car speeding on a wide road along the north–south direction. Let us also suppose that the length of the road is 60 km and the end is marked by a brick wall. If the car moves exactly in the north–south direction, it will hit the wall exactly after an hour of starting. Let us now suppose that somehow the driver is feeling drowsy and his sense of direction is not very accurate anymore. He may then drive the car in a slightly slanted manner, and drive a bit in the east–west direction and not exactly in the north–south direction. In this case, he will take longer than an hour to hit the brick wall.

Suppose an observer tries to measure the speed of the car by recording the exact time of the crash of the car with the wall. If the car hits the wall exactly an hour after starting then he deduces a speed of 60 km per hour, and if it hits later, he deduces a slower speed. The speed measured by this observer will not be the real speed of the car, but the speed in the north–south direction. The real speed of the car is always a constant, but when the car moves at an angle to the north–south direction, a part of its actual speed goes into having some speed in the east–west direction and the speed in the north–south direction becomes less (Fig. 2.2).

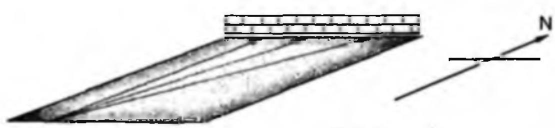


Fig. 2.2 *The speed of a car estimated by its arrival time at the brick wall. This speed will actually be its speed in the north–south (NS) direction. If a car moves exactly towards the north, then its actual speed and the NS speed will be equal. If the car however strays somewhat in the east–west direction, it will arrive a bit later at the wall, making its NS speed somewhat lesser than the actual speed with which the car moved on the ground.*

In other words, when the car moves exactly in the north–south direction, all of its speed of 60 km per hour is directed in the north–south direction. If by mistake it moves at an angle, then a part of this speed is used in going in the east–west direction. Since the actual speed is a constant, the resulting speed in the north–south direction is bound to be less or in other words, since the combined value of speeds in the two directions is a constant, speed in one direction must decrease to compensate for the increase in the other direction.

According to Einstein, our speed through time and space is somewhat like the speed in the north–south and east–west direction. Usually we think of speed as some amount of space being covered during an interval of time. We can also think of a sort of speed through time, which will essentially be the rate of flow of time. Suppose this is equivalent to the speed in the north–south direction in our example, and the usual speed (through space) is equivalent to the speed in the east–west direction. The constant speed of the car will then be equivalent to the constant speed of light.

Imagine that our combined speed through space and time, or spacetime, is constant, like the actual speed of the car in the example. When we stay put at some position in space, we only move through time, and the speed in space is zero. This is equivalent to the car moving exactly in the north–south direction, with zero speed in the east–west direction. If we now move in space and acquire a speed through space, it will be like moving at an angle and acquiring an east–west speed at the expense of the speed in the north–south direction. In other words, if we acquire a speed through space, our speed in time is bound to slow down. The higher our speed in space, the more prominent is the slowing down of time.

We can ask if this slowing down of time can ever make it sit still. As a matter of fact, it will, for objects moving with the speed of light. Time stands still for photons. It has the 'master speed' in the language of Frost (as quoted at the beginning of the chapter).

We can then say that we all move with the speed of light through spacetime. Photons use all of their speed in going through space and so their speed through time is reduced to zero. And when we stay fixed at a position in space, we are using all of this speed in moving through time. All other phenomena take place somewhere between these two extreme cases, where the speed through time and space is divided up into two parts. When one increases, the other decreases to compensate for it[2].

This way of looking at spacetime makes the relativity of time and space appear less strange. The simplicity of this approach also makes the combined picture of spacetime very elegant.

2.3 General Theory of Relativity

We have found that it is cumbersome to think of time and space separately, as was previously done. It makes more sense to think of them combined into a single entity. We have however kept our discussion limited to the cases of objects moving with constant velocity, without any acceleration. One can then ask how this combined background of spacetime changes as we include acceleration and begin to change the speed.

Einstein added one more postulate here. Although simple in content, the implications of this postulate have also been revolutionary, like his first postulate of constancy of light speed. It will be easy to explain it with the help of an example from daily life. Consider going up in an elevator. When it begins to move up we suddenly feel an attractive force downward. For a moment, we feel heavier than normal. Similarly, when an elevator begins its journey downward, for a few moments we feel lighter as if the gravitational attraction of the Earth had decreased.

A point to note is that this does not happen when the elevator goes up or down at a constant speed; we feel the change in the gravitational force only at the time of change of this speed, when the elevator begins to move up or down.

It will then appear that the change in speed is related to gravitation in some way. Einstein therefore postulated that acceleration and gravitation have the same effect on objects. Imagine an observer going up in an elevator in space and suppose the elevator gains speed steadily. If all windows are then closed, this observer will feel some 'weight', almost as if the elevator were standing on the ground of a massive planet.

If this equivalance between gravity and acceleration is accepted, then we do not need to limit the principle of relativity—the principle that laws of nature are independent of the reference frame—to the cases of constant speed. We can extend this principle to the cases including acceleration. We would then say that the experiences of these observers (moving with changing speed) could also be explained by the usual laws of physics—one only needs to include the effects of an equivalent gravitational field.

There is an obvious problem with this algorithm though. If we were to calculate the effect of this equivalent gravitational field according to Newton's law, we would need to determine the distances between objects. In Newton's law, the gravitational force depends on this distance. But we have found that distance, or the measurement of space, depends on the observer. Whose measurement of distance should then be used in this calculation? We cannot extend the principle of relativity to cases with acceleration if we cannot answer this question. Einstein therefore realised that it is impossible to do this extension in patches, using Newtonian ideas here and new approaches there. One will have to completely change the way we look at gravity. And without waiting for anyone else, Einstein started on this difficult road by himself and was able to build a theory which would put the principle of relativity on a pinnacle.

Imagine putting our elevator inside a rocket and performing a simple experiment in it. We can shine a beam of light from its wall on one side to the opposite side. We know that light travels in a straight line. Photons do not like to waste time by going in a zigzag path. When the rocket stays on the ground, the light beam will then naturally travel straight and fall on a spot on the opposite wall at the same height. If we think a bit, we will find that this will be the case even when the rocket goes up with a constant speed (since the experiences of a static observer and an observer with constant speed must match with one another).

Let us now imagine launching this rocket into space (the diagram on the right of Fig. 2.3) and gradually increase its speed. In this case, by the time the light beam reaches the other wall, the rocket will have gained some speed and travelled an extra bit of distance than in the case of constant or no speed at all. The beam of light will not be able to keep up with this extra forward motion, will lag behind and create a spot on the opposite wall at a lesser height than earlier. In other words, the path of the light will be curved.

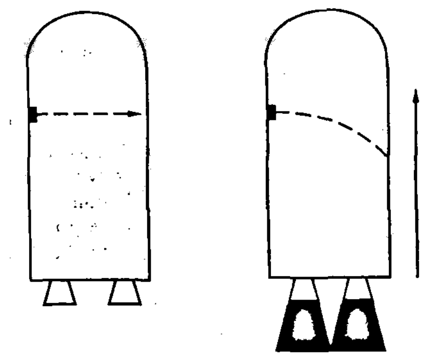


Fig. 2.3 *Light will travel in a curved path inside an accelerated elevator.*

We however know that light cannot travel in a curved line. Light always takes the path with the shortest time of travel and therefore it travels straight because a straight line between two points is the shortest distance between them. Why should light then take a curved path if it is not the shortest distance?

Einstein took a bold step here and proclaimed that the space inside an accelerated elevator is 'bent' in such a way that the curved path becomes the shortest path between two points.

It is not difficult to think of a situation though in which a curved path is the shortest path. Imagine flying in an aeroplane from Delhi to Washington, D.C. If the pilot chooses the shortest path then the plane will travel close to Greenland. It becomes easy to understand if we draw this path on a globe. But if we draw the same path on a plain map, the same path would look far from being a straight line. It is then because of the curved face of the Earth that a curved line is the shortest path between two points (Fig. 2.4).

We should remember that it is the same path but described on background maps of two different styles and they appear different. We must use the proper background to interpret its relevance. It is almost like using a word in two different languages which can make its meaning completely different. 'Gift' in English means a present whereas it means poison in German. A severe misunderstanding can ensue if we do not know in which language it is being used. Similarly we may have to face such puzzles as 'how can a curved path be the shortest path?' if we were not aware of the background map.

Fig. 2.4 *The shortest path on the curved face of the Earth will appear to be curved when described on a map drawn on plain paper, as shown by the appearance of the same path – the shortest path between New Delhi and Washington D.C. – on the surface of the round Earth (top) and on plain paper (bottom).*

Going back to the space inside an accelerated rocket, we can then say that the space inside the rocket is bent, which will make a curved light path the shortest distance between the two walls. The walls or the rocket are not bent, it is the space itself that is bent. It may be difficult to appreciate a bent region of space for three-dimensional beings like us. We can bend a line by taking its two ends and moving them in a direction perpendicular to the original straight line. We can also bend a plain piece of paper by moving some of its parts in the perpendicular direction. If we were now to bend the three-dimensional space that we live in, in which direction do we move it around?

Since we cannot perceive of any other direction that is perpendicular to the three dimensions of length, breadth and height, it is impossible for us to visualise curved space. One way to think about it is to consider the laws of geometry in a curved space and compare them with that in flat space.

In the usual Euclidean geometry we know that the sum of three angles in a triangle is 180 degrees. If a triangle is drawn on a curved space like the surface of the Earth, then the sum of the angles will be larger than 180 degrees. It is also possible to think of a curved space on which the sum will be less than 180 degrees. The space will in this case have the shape of a saddle (Fig. 2.5).

Fig. 2.5 *Three possible geometries of spacetime: from left to right—curved like the surface of a ball, curved like a saddle and with zero curvature. In this picture, we have suppressed one dimension of space and shown the bending of space with two dimensions of length and width. We should remember that space also has another dimension of height. We have to imagine that this three-dimensional space is bent in some intangible 'direction'. In that case too there will be the above-mentioned three possibilities for the sum of three angles of a triangle.*

Our traditional (that is, Euclidean) ideas of parallel lines also get changed in curved space. One important axiom in Euclidean geometry has it that two parallel lines never intersect even when extended to infinity on a flat plain. If we however draw two parallel lines on a ball they will tend toward each other when extended. Two parallel lines on a saddle will move away from one another.

We can then express the result of our thought-experiment inside the rocket as being that the geometry of space inside a rocket changes when it is accelerated. Of course we should not only talk about the geometry of space, but that of spacetime. We have also learned earlier that the effect of acceleration is equivalent to that of gravity. Einstein therefore arrived at the conclusion that *gravity changes the geometry of spacetime.*

2.4 Gravity and Geometry

Let us try to understand with the help of an example what this actually means. Consider two travellers beginning their journey northward from the equator on Earth. Suppose the distance between them at the outset (while they are on the equator) is 200 km. Let us also suppose that according to their beliefs they are standing on a flat surface and not a curved one. In other words they are firm believers of a flat Earth. If they now both go northward with the help of a compass, and move at the same speed, then they would discover after a while that the distance between them has decreased from the initial 200 km (Fig. 2.6).

They would then wonder and puzzle over the fact that their intermediate distance has decreased although they had not deviated from their northward path. For the sake of dramatising our example, suppose one of the travellers to be Newton himself. He would then say that this could be easily explained if there was an attractive force between them.

Fig. 2.6 *Two travellers going north from the equator on Earth would find their intermediate distance decrease with time.*

If we now imagine Einstein to be travelling on a spaceship outside the Earth and watching his predecessor on Earth, he would naturally realise that the root of the puzzle faced by the travellers is their denial of a curved surface. They are bound to come together although they walk on exactly parallel lines, since they are actually on a curved surface. As a matter of fact, both would bump into one another at the north pole at some point of time. If one denies all this and incorrectly believes in a flat Earth, one would have to invent forces of attraction to explain the observations.

According to Einstein, therefore, one should abandon the explanation of gravitation as a force of attraction and think of it instead as something which changes the geometry of spacetime. Why are the planets like Earth moving around the Sun? Newton would have said that this is because of the gravitational attraction between the Sun and the planets. Einstein would rather say that the gravity of the Sun changes the spacetime in its vicinity so that a straight line gets bent and the orbits of planets are nothing but bent straight lines. Although we cannot visualise the bending of three-dimensional space around us, we can compare it with a bent piece of cloth or paper.

If the Sun did not exist then the spacetime around the Earth would have been like a plain piece of cloth that is kept taut. If we imagine the Earth to be a small ball on this cloth it would then move in a straight line if given a push in some direction. If we now put a massive ball on this cloth, it would distort the cloth as the gravity of the Sun distorts the spacetime around it. The ball representing the Earth would not move in a straight line if given a push; it would take a curved path (Fig. 2.7).

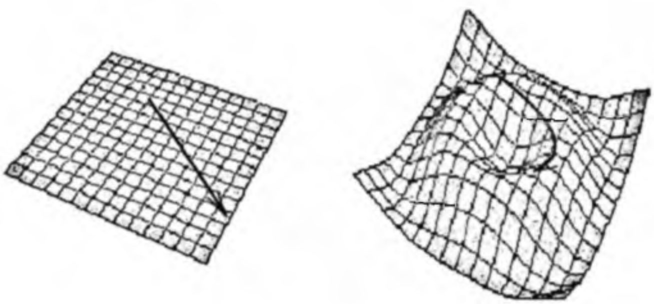


Fig. 2.7 *An object will move on a straight line (without being subjected to any force) on a flat surface (left), but it will move in a curved path if the background surface is curved (right).*

This analogy is an incomplete one since we have likened the four-dimensional spacetime to a two-dimensional piece of cloth. It does help us understand the basic ideas of the general theory of relativity, though in a limited way. If the mass of the heavy ball is increased then the distortion of the piece of cloth would be more severe. Almost similarly, the degree of distortion of spacetime around an object depends on its mass. Roughly speaking, mass (or energy) changes the geometry of spacetime, and the geometry determines the trajectory of an object. This is the essence of Einstein's equation and the general theory of relativity. Space and time have not only been combined here, they have also been given a geometry and gravity has acquired a new and elegant interpretation.

Simplicity and elegance are however not proofs of a scientific theory. One has to prove it correct by doing experiments and testing its predictions. The curvature of spacetime around objects like the Earth or even the Sun is very very small, like the minute time differences between observers inside and outside of a moving train. Nevertheless, careful experiments have proved Einstein's theory right. One such experiment in 1919 detected the bending of light near the Sun as predicted by Einstein's theory.

Observations during the solar eclipse of that year showed that light rays coming from distant stars get bent while passing close to the Sun (Fig. 2.8). It is almost like the bending of light rays through a lens. Naturally this observation had to be done only during a solar eclipse as it is impossible to observe stars near the Sun otherwise. Astronomers found that the bending of light rays, closely obeyed the predictions of Einstein's theory[3].

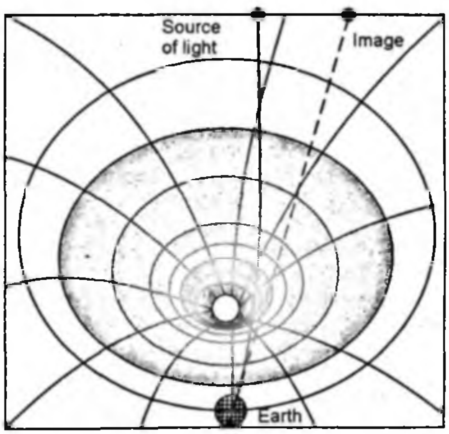


Fig. 2.8 *The bending of a ray of light from a distant star while passing by the Sun makes the star appear at a different place than it would have been without any bending. The spacetime is represented as a two-dimensional grid which is distorted near a massive object like the Sun. The ray of light coming from a source of light (top of the diagram) is shown to move along a curved line as a result of the distortion in the spacetime in the vicinity of the Sun. When this ray is observed at Earth, it would appear to have come along the dashed line instead, which is the extrapolation of the last portion of the real path, and the object would appear at a different position in the sky on Earth. Normally it is impossible to observe the stars near the Sun except at the time of a solar eclipse. Astronomers first observed an apparent shift of the position of stars during the solar eclipse of 1919. They recorded the position of the stars during the eclipse and compared them with those observed six months earlier, when the Sun was at a different part of the sky. This comparison proved Einstein's prediction to be correct.*

This bending of light is because of the curved nature of spacetime in the vicinity of the Sun. According to Einstein's theory, it is also possible that spacetime can be so severely distorted that light rays cannot come out of a region. This is where a black hole appears. When an object comes too close to a region of such severely distorted spacetime, it gets hurled into the black hole. In other words, a black hole sucks into it any matter that comes too close[4].

It is not only the path of the light that gets affected by gravity, the energy of a particle of light (photon) also decreases when in the vicinity of a massive object. The amount of this decrement depends on the trajectory of the photon and the mass of the object. There have been many experiments to test this prediction and all have proved Einstein's theory to be correct. It is not only in the laboratories that this theory has been tested, it has now also begun to affect our daily life in many ways. In these days where aeroplanes try to land during bad weather using computers and not under the manual manoeuvring of a pilot, even a small mistake in the calculation of distance and time could prove fatal. The softwares used by these computers now take into account the minute effects predicted by the theory of relativity[5]. We will find later that the effects of a curved spacetime on light will be very useful in our discussion of observational cosmology.

2.5 Relativistic Cosmology

Since mass affects the geometry and shape of spacetime, one can then ask how the total mass of the Universe affect its spacetime. In other words, one could ask how the spacetime of the whole Universe looks. Here too one finds that there are three possibilities: (1) perhaps the spacetime is curved like the surface of a ball—where two parallel lines slowly converge and where the sum of the angles in a triangle is larger than 180 degrees; (2) or, perhaps it is curved like a saddle—with parallel lines in it diverging out and sum of the angles in a triangle being less than 180 degrees; (3) or, possibly it is just flat, where the normal geometrical laws of Euclid are valid.

We will discuss later which of these three possibilities is corroborated by modern cosmological research. One important aspect of the spacetime of the Universe is whether or not it is closed. If the geometry resembles that of a surface of a ball, then the spacetime will be closed onto itself. It will not have any edge though, just as the Earth has no edge. The Universe will then be finite but without any boundary. If the geometry is like that of a saddle then the Universe will be infinite and without any boundary. We should note that this is the first time a finite but edge-less Universe has been contemplated by mankind.

Scientists were naturally fascinated by these new ideas about the Universe in the beginning of the twentieth century. At the same time a theoretical calculation surprised most of them, including Einstein. When the theory of general relativity was applied to the case of the Universe it was realised that whatever be the geometry of its spacetime, it could never be static. Spacetime of the Universe must either inflate or deflate. This restless nature of spacetime appeared bizarre to scientists. Even Einstein who was the epitome of the iconoclastic scientist of that era, could not accept it initially. In his own calculations, he tried to halt the expansion or contraction of the Universe by some other means.

It is around this time that a few astronomers discovered a very strange behaviour of galaxies in the Universe. This is the beginning of modern cosmology.

3

The Expanding Universe

> ...The Universe may or may not be very immense,
> As a matter of fact there are times when I am apt
> To feel it close in tight against my sense
> Like a caul in which I was born and still am wrapped.
>
> Robert Frost, *Skeptic*

In the beginning of the twentieth century, an American named Edwin Hubble who had studied law at Oxford grew disenchanted with it and decided all of a sudden to try his luck in astronomy. He would soon make one of the most important discoveries of the century and proclaim the Universe to be expanding. At around the same time, a metereologist in Russia sent Einstein some of his calculations on the dynamics of spacetime. Einstein at first hesitated but later accepted the results of these calculations by Alexander Friedmann. A Belgian priest called Georges Lemaître who was trying to apply Einstein's ideas to the spacetime of the Universe also arrived at similar conclusions. He had gone a step further than Friedmann though and also thought about the possible ways of testing these theoretical ideas. These ideas came to the notice of the British astronomer, Arthur Eddington who wrote an essay on it.

It was by reading this essay that Edwin Hubble found the similarity between his discovery and the contemporary theoretical ideas of the Universe, and realised how momentous his discovery was.

Friedmann (and Lemaître) discovered that in the general theory of relativity, the Universe must either expand or contract. It cannot remain static. By the expansion of space, one means the increase in the distance between material objects in the Universe. If a physicist observes the Universe he will find the rest of the Universe moving away from him since the intermediate space is expanding. In a contracting Universe, one will find the opposite, to be true. Lemaître noticed that this expansion (or contraction) of space will change the nature of light coming from distant galaxies. To understand this we must first understand the nature of light itself.

3.1 Cosmological Redshift

Light is essentially an electromagnetic disturbance that is moving. Just like the waves created on the surface of water by throwing a stone into it, an electrical (or magnetic) disturbance somewhere gives rise to electromagnetic waves or light. These waves can have various lengths just like water waves of different lengths. The nature of light depends on its wavelength. If the wavelength of light is about ten-millionth of a centimetre (10^{-7} cm), we call it an X-ray. If the wavelength is increased to 0.00004 (4×10^{-5}) cm, it appears as blue light to our eyes. When the wavelength is gradually increased it appears as light of different colours in the visible spectrum. For red light, the wavelength is around 0.00007 cm. One then gets infrared rays (0.0001 to 0.01 cm), microwave radiation (1 mm to 10 cm) and radio waves in general (one to hundred metres). Although we call them by different names, they are actually the same sort of waves with different wavelengths.

When space expands or contracts, the measures of length–width–height change, and the wavelength of light passing through this space also changes. To simplify our discussion, let us take the example of a one-dimensional Universe. Suppose that we all are confined to a very narrow (with almost zero width) rubber band. This width and heightless band *is* our Universe and there is nothing outside of it. Suppose there are (again one-dimensional) bugs inhabiting this Universe; they are uniformly spaced and are marching in unison with a uniform speed (Fig. 3.1). Consider the case in which this band is being gradually stretched along its length. These uniformly spaced bugs can then represent the crests (or, equivalently, the troughs) of a light wave moving through an expanding Universe. As the rubber band stretches, as in an expanding Universe, one finds that the distance between the bugs, or equivalently, the wavelength of light, increases with time.

As space expands, the wavelength of light changes and makes blue light appear red. In an expanding Universe, therefore, we will find distant light to be redder than usual. A galaxy in an expanding Universe will appear redder than in the case of a static Universe. This shift of wavelength is called the 'cosmological redshift'. The more distant the source of light, the more of (expanding) space it has to traverse, the more redshifted it will appear to us. It will therefore appear redder to us.

It is difficult to detect this redshift in white light though, since all colours equally contribute to it, and if blue light is shifted to red, ultraviolet light will get shifted to blue and fill the gap. The spectrum of completely white light therefore will not show any change. Although, if some colour dominates the spectrum of light from a source, it will be possible to detect the redshift in it. This is what happens in the case of light from distant galaxies.

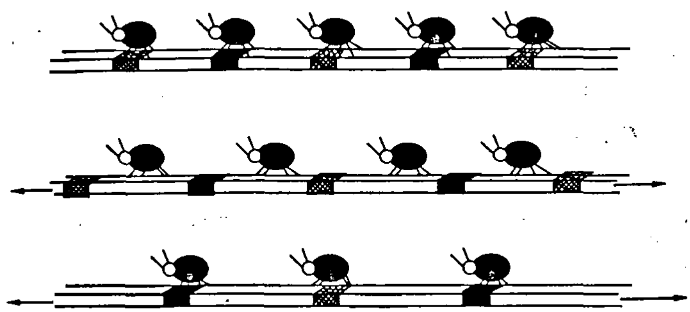


Fig. 3.1 *A simple one-dimensional Universe is being represented here by a thin rubber band. Suppose this rubber band gradually stretches, just as the Universe expands. A line of bugs marching uniformly on this stretching band is analogous to a light wave travelling through an expanding Universe. The bugs are uniformly spaced and they move at the same speed relative to the rubber band. The rubber band is uniformly marked and initially the bugs line up with the markers. A sequence of three snapshots during which the central bug moves one marker ahead is shown in the picture. By this time all the other bugs have also moved one marker ahead and so they remain equidistant from one another. The distance between them however increases in direct proportion to the stretching of the rubber band. If each bug represents the crest or the trough of a light wave, this sequence shows that the wavelength of light will increase with time in an expanding Universe.*

This is exactly what was found by Edwin Hubble. He found that most of the galaxies observed by him showed signs of redshifted light. Apart from some nearby galaxies (like Andromeda), all galaxies seemed to show that space was expanding. This immediately suggested that the theoretical calculations of Einstein (and Friedmann–Lemaître) were not simply figments of the imagination. Hubble announced his discovery of the expanding Universe in 1929.

The expanding space between us and the distant source of light makes it appear to recede from us. All galaxies in an expanding Universe will therefore appear to move away from us. The expansion of space is therefore made tangible by the objects moving with it. Surely the associated speed of recession of galaxies is related to the redshift suffered by the light coming from them. It turns out that for very small redshifts, or equivalently, for very nearby galaxies, the redshift is proportional to the speed. This approximate proportionality holds in the region of the Universe probed by Hubble.

Hubble recorded this recessional speed of galaxies and their respective distances from the Earth. He then discovered a simple relation between the speed and the distance which bolstered the theoretical speculations of that time. He found that the speed is proportional to the distance. Distant galaxies moved away from us with a greater speed than the nearby ones, with a simple proportionality between speed and distance[1]. Let us try to understand what this really means.

3.2 Universe without a Centre

We have learned in the first chapter that the galaxies in the Universe appear to be more or less homogeneously distributed when averaged over a large length scale, or, equivalently, when seen from a very large perspective. Einstein did not have access to such data of positions of our nearby galaxies. He did make a guess though, that the matter in the Universe is more or less homogeneously distributed. He also assumed that the Universe looked the same in all directions from a given position— it was both homogeneous and isotropic. There is no privileged position and no centre in this sort of Universe. He (and Friedmann) did his calculations for the spacetime of the Universe based on these assumptions, which led to the predictions of an expanding (or a contracting) Universe. If we think a bit we will find that Hubble's discovery matches these theoretical ideas closely.

Let us again take the example of a one-dimensional Universe, of a narrow rubber band. Let us suppose that we are positioned on this band equidistant from each other. In the case of a stretching rubber band, i.e., in an expanding Universe, we will find our neighbours slowly receding from us. Suppose this speed of recession is proportional to the distance. Distant neighbours are therefore moving away from us with a greater speed than the nearby ones. The apparent speeds of our neighbours are shown in Fig. 3.2 with arrows of corresponding lengths and directions. Suppose our nearest neighbour on the right has a speed of v. The nearest neighbour on the left will then have a speed that is opposite in direction to this, that is, a speed of $(-v)$. The next nearest neighbours will have speeds of $2v$ and $(-2v)$, since their distances from us are twice that of the nearest neighbours. Consider now the observations made by one of our neighbours. We will try to observe from the vantage point of our nearest neighbour on the left. In this case, our neighbour will naturally think himself to be static and that we are moving away from him, with a speed of : $0+v=v$. The speed of our nearest neighbour on the right will appear to him as : $v+v=2v$. On the other side, his nearest neighbour will seem to move with a reduced speed of $(-2v)-(-v)=-v$.

If one compares the speeds and distances of objects as perceived by us and our neighbours we will then find that the proportionality between speed and distance holds in both cases (Fig. 3.2). Our neighbour will also discover that speed is proportional to the distance. In other words, everyone in this rubber band Universe will think that everyone else is moving away from him in the same way (keeping the proportionality between speed and distance). The expansion of the rubber band Universe will thus appear to be the same for everyone.

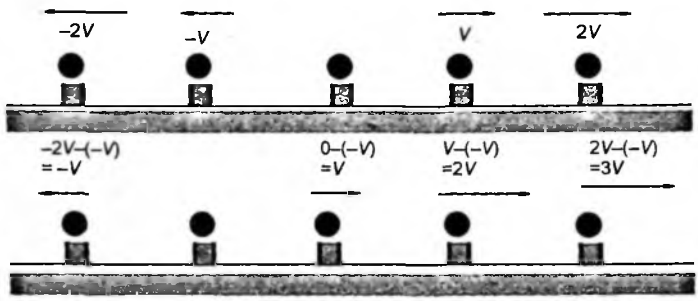


Fig. 3.2 *A simple one-dimensional Universe of a rubber band is being shown here. The rubber band is gradually stretching like our expanding Universe. In this case (the top panel) the apparent speed of our neighbours as perceived by us (the observer in the middle) are shown with arrows. The length of the arrows correspond to the magnitude of speed. Suppose this speed is proportional to the distance. The bottom panel shows the speeds as perceived by our neighbour.*

If we now think of our three-dimensional Universe instead of this simplified rubber band Universe we will again find the same result. If the recession speed of galaxies is proportional to their distances to us, then an observer on any other galaxy will also find the same relation. He will also observe that galaxies are moving away *from him*, with a speed that is proportional to the distance *to him*. This means that galaxies observed by Hubble are not simply moving away from our Milky Way, but all galaxies are moving away from all other galaxies—and in the same manner. The expansion of the Universe is therefore universal and without any centre. This is exactly what was speculated by Einstein and his colleagues. We then find that the law of expansion of the Universe as discovered by Hubble tallies with the theoretical speculations[2].

One could ask here if this expansion of space is also shared by the space inside our Milky Way. Is our galaxy expanding? What about the Solar System? Is our Earth expanding its horizons too?

The answer is no. Although the Universe is expanding in a global sense, there can be exceptions to it locally, for example, in regions where there is a strong local gravitational field. This is what happens inside our Milky Way (or any other galaxy) or in its vicinity. In these regions, the local gravitational field is large enough to make the spacetime there not participate in the global expansion. For the region containing our Galaxy, the gravitational field is not only large inside the Milky Way, it is also large in its vicinity. It is large enough to make our neighbouring galaxy Andromeda come closer to the Milky Way instead of the two galaxies moving away from one another. As one observes more distant galaxies, the local gravitational pull of the local galaxies begin to fade away and one then finds them moving away from us.

One might also wonder if the Universe is expanding *into* something. The answer is that it is not expanding into anything, since everything belongs to the Universe and there is nothing outside of it. All of space is expanding.

There have been many attempts in the last century to accurately measure the speed of expansion of the Universe since Hubble announced his discovery in 1929. It has been very difficult to measure this speed since it is not easy to determine the distances of galaxies. There has been a lot of progress though in the last decade and astronomers now think that the speed of expansion is such that a galaxy situated a million parsecs from us will recede with a speed of approximately 70 km per second. According to Hubble's law, a galaxy at twice this distance will have a recession speed that is also twice this speed. This value of the expansion speed—of 70 km per second per megaparsec— is called the 'Hubble constant'. It is a constant since it is shared by all matter in the Universe, at any given time (although its value may change with time).

3.3 Past and Future of Expansion

It can be asked if the speed of expansion of the Universe stays constant, or if it changes with time. One can also ask if there is any way to determine the history of the expansion, if it has remained the same forever and if it will change in the future. The answer to this question lies in the calculations done by Friedmann. According to this calculation, the rate of expansion of the Universe depends on the density of the Universe. By density, one means the total matter and energy density, including both dark and visible matter, and any other kind of energy that might be available in the Universe.

One could give a simple analogy to illustrate this point. Consider throwing a stone up in the air on Earth. In most cases, the stone will stop at some height and return to Earth. The height reached by the stone will naturally depend on the initial speed with which it is thrown. More speed will ensure greater heights reached by the stone. If this speed is gradually increased, we will find that the stone at some point will escape from the gravitational field of the Earth. According to Newton's law, this situation arises when the speed is approximately 11 km per second or more. This is the threshold speed for rockets when they are launched into space from the surface of Earth.

This escape speed naturally depends on the gravitational field of the Earth, which depends on its total mass and size. If one does this experiment on the Moon or Mars, one will find that the escape speed there is smaller than that on Earth, since the gravitational forces on the Moon and Mars are less than that on Earth. It would then be possible to throw a stone out of the gravity of Moon or Mars with a lesser speed. We thus find that the fate of a stone (for a given initial speed) depends on the mass and size of the object on whose surface the experiment is being done. We can also say that it depends on the density of this object (since density is determined by mass and size).

In an almost similar way, the fate of the Universe depends on its density. The speed of its expansion will decrease with time, just as the speed of the stone thrown up. If the density is larger than a critical value, the speed of its expansion will reverse sign at some point. The Universe will then begin to contract. This is analogous to the return of the stone after reaching a maximum height. If the density is less than this critical value, gravity will never be able to put the expansion on hold, analogous to the escape of a stone. The speed of expansion in this case will remain non-zero even at infinite extent and the expansion in this case will continue forever. If the density of the Universe is exactly equal to the critical value (analogous to the stone having exactly the escape speed), the speed of expansion will slowly decrease but the rate of this decrement will be so slow that the Universe will never be able to stop and contract. This value of the density is called the 'critical density'. According to the calculations of Friedmann, this critical density for our Universe (with the value of the Hubble constant as mentioned earlier) is equivalent to approximately five hydrogen atoms inside a volume of one cubic metre. This may sound negligibly small, but we should remember that the density in question is not the density of matter on Earth, or in the Solar System or even in the Milky Way. The density of matter in these regions is much larger than the global density of the Universe. We should for this purpose take an average over a distance scale at which the distribution of matter in the Universe looks more or less homogeneous[3].

Cosmologists describe the ratio of the density of the Universe to this critical density by the Greek letter Ω. When the density is equal to the critical density, the value of Ω is unity.

According to Friedmann's calculations, the geometry of spacetime also depends on the mass density of the Universe. When the density is larger than the critical density, spacetime has a global curvature like that of a surface of a ball. In this case the Universe is said to be closed and the expansion will cease at some point making the Universe contract. When the density is smaller than the critical density, it has a curvature like that of a saddle. The Universe in this case is said to be open and there is no stopping of the expansion in this case. When the density is equal to the critical density, the geometry is that of an Euclidean plane (Fig. 3.3).

The connection between the dynamics of the Universe and the geometry of spacetime is somewhat more complicated than this in reality. There exists an extra parameter in Einstein's equation for gravity that can complicate matters. Einstein called it the 'cosmological constant'. He thought that there was a case, made possible by his equation, which would endow some energy even to empty space. This energy would be able to make empty space expand or contract, even without the presence of any mass. This would correspond to some inherent energy of emptiness or the vacuum. Einstein himself was not very comfortable with this idea and once even said that thinking about the cosmological constant was one of the biggest blunders in his life. Physicists have however not abandoned it completely and as a matter of fact very recent observations have suggested the presence of this bizarre energy.

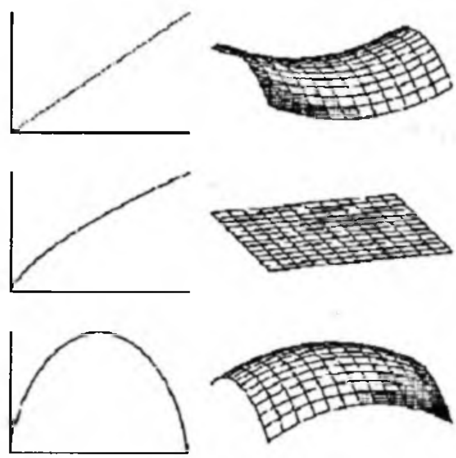


Fig. 3.3 *Three possibilities of the rate of expansion of the Universe are being shown here. The variation of the speed of the expansion with time is shown on the left, along with the corresponding geometry of spacetime on the right. In the topmost panel, it is shown that the expansion continues unabated in an open Universe. The density of the Universe is less than the critical density. When the density is equal to the critical density, the geometry becomes Euclidean and although space expands forever, the speed of expansion decreases and becomes zero at infinity (middle panel). The bottom panel shows that in the case of density being larger than the critical density, the expansion of the Universe stops after some time and then the Universe starts to contract.*

It seems that this energy has bestowed on the spacetime of our Universe a repulsive force which is making it expand at an accelerated manner, unlike the expectation of a slowed down expansion as discussed earlier.

Cosmologists are studying this 'dark energy' in detail at present. Many of them are of the opinion that this dark energy is not a constant, but one whose magnitude changes with time. It is possible that this dark energy is the manifestation of a yet unknown particle. Some call this hypothetical particle (or field, in the language of quantum mechanics) the 'quintessence' taking the cue from the five elements of Aristotle. Aristotle had thought that the basic constituents of everything in the Universe were five elements, four of which are familiar to us—earth, water, fire and wind. It was thought that the heavenly bodies were made of the fifth essence, 'quintessence'[4].

If the presence of dark energy is confirmed by future observations, the expansion of the Universe will never cease no matter what the geometry of spacetime is. The repulsive nature of this dark energy will keep up the expansion irrespective of the gravitational energy of its mass density. In fact observations indicate that the expansion of the Universe is accelerating at present.

3.4 Olbers' Paradox

Let us discuss an old puzzle before going into the details of these recent observations. The puzzle is simply stated: why is the night sky dark? It may sound trivial and many will think that the answer is that the night sky is dark because the Sun is then not visible in the sky.

If we however think a bit deeper we will find that this puzzle is not trivial after all. Heinrich Olbers, a German astronomer in the nineteenth century, first thought about this problem. Suppose we are in the middle of a forest with many trees. There are trees whichever direction we may choose to look at. If we stand inside a small forest, we may be able to see the region outside the forest in the directions where there are no trees. If the forest is big enough, it may not be possible to find such a direction, because there will be many more trees in this case. Our line of sight will be bound to cross some tree or the other. In the case of an infinite forest, we could say with certainty that all lines of sight will hit a tree somewhere (Fig. 3.4).

Consider now the case of an observer in an infinite Universe. In this case, our line of sight is bound to be cut short by some luminous object or another (star or a galaxy), irrespective of the direction we may choose (Fig. 3.5). In other words, it would not matter if the Sun were there in the sky or not, all direction paths in the sky would point toward a source of light. The whole sky would then be bright and would be bright no matter if it is day or night. This is the basic point of Olbers' paradox[5].

There were many attempts to resolve this paradox initially. The answer to the puzzle has however come with the discoveries made in the twentieth century. There are mainly two factors that make the night sky dark. We will discuss the secondary reason first.

We have learned that the wavelength of light increases with the expansion of the Universe. We have also learned that this causes the redshift of distant sources of light. The wavelength, or the frequency, of light is intimately related to the inherent energy of light. This energy increases with the frequency of light, or equivalently, decreases with wavelength. In other words, the energy of an X-ray photon is much larger than that of a blue photon, which in turn is larger than that of a red photon, and so on.

Fig. 3.4 *Any line of sight in an infinite forest will cross a tree at some distance or the other. This will make all lines of sight blocked by a tree trunk at some distance.*

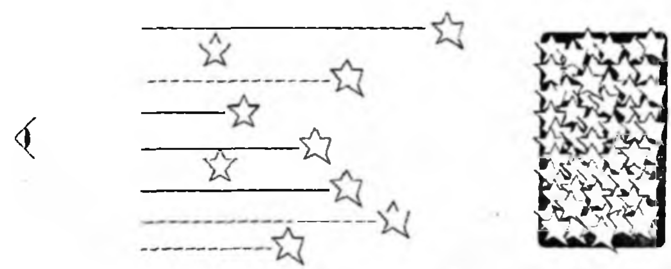


Fig. 3.5 *In a similar manner, any line of sight for an observer in an infinite Universe will cross a luminous object somewhere and even the night will be bright.*

As the Universe expands, light therefore not only suffers a shift in its colour, but is also degraded in energy. The expansion of the Universe therefore necessarily reduces the energy of light coming from a distant source. The redshift of light therefore would reduce the brightness of the night sky.

This however does not completely explain the darkness of the night sky. Although it does reduce the amount of energy that our eyes would receive in an infinite Universe, it does not explain the level of darkness we are used to in the night sky. The primary reason for this has now been understood based on the discoveries of the twentieth century.

We saw in the first chapter that the light from a distant source tells us about it as it was in the past; the more distant the source, the more in the past that it corresponds to. When we observe the Sun, it corresponds to the Sun as it was eight and a half minutes ago. When we observe the Andromeda galaxy, we find it as it was about two million years ago. As our line of sight advances to further distances, it will also go back further in the past. Let us now think again of the infinite Universe in Olbers' puzzle. The main argument for a bright night sky was that a line of sight in an infinite Universe is bound to intercept a luminous source somewhere in space. We should however remember that these lines of sight would also take us back in time. And we should not forget that stars and galaxies have a finite lifetime. They do not exist, and have not been there, for ever. As we then take our lines of sight back into the past, we would come to a point when there were no galaxies or stars, or any luminous sources. This is of course the main point of discussion in this book, the birth of luminous objects and the dawn of the Universe. Even without going into the details of when and how the first luminous objects appeared in the Universe, we can however say that the main argument in Olbers' paradox is not convincing. It is not possible to extend the lines of sight without any limits. And it is not possible to guarantee that all lines of sight would intercept a luminous source somewhere in space, or time, since there must have been a time when there were no luminous objects in the Universe.

We can also add another argument along the same line. If we imagine the state of the Universe in the past, keeping in mind the expansion of space, we will find that galaxies were closer to each other in the past than they are at present. We can imagine the 'film' of the evolution of the Universe running backward. The Universe will then progressively look denser as one goes back more and more into the past. There will certainly be a point of time when everything will be on top of each other. At this point, the extent of the Universe, will not be more than a point. This moment would then correspond to the birth of the Universe since when it has been expanding. It is impossible for us to imagine the state of the Universe before this moment[6]. This moment is commonly known as the Big Bang[7].

We therefore find that although the Universe may be infinite, as the lines of sight are advanced in space, there necessarily comes a point in time beyond which they cannot be extended. There is a limit to the extension of the line of sight. The main argument behind the Olbers's puzzle is therefore not justified and this is the main reason why the night sky is dark[8].

Cosmologists call this scenario for the evolution of the Universe the Big Bang model. One might ask here where in our Universe did this 'big bang', or big explosion, actually occur. We should remember though, that the whole Universe was at a point at this moment, with infinite density. The Big Bang therefore occurred throughout the whole Universe.

3.5 The Fate of the Universe

Let us now discuss the recent observations of astronomers which shed light on the history and future of the Universe. Astronomers use the concept of redshift to discuss the history of the Universe. We have learned that the wavelength of light is stretched by the expansion of space. If space expands by a factor of two, the wavelength is doubled. In the language of astronomers, the 'redshift' of light will be 1. When the wavelength of light increases by a factor of three, the corresponding redshift will be 2[9].

Cosmologists refer to the present epoch as being at zero redshift. If we then discuss a particular epoch in the history of the Universe when it was half its present size, then the wavelengths of light were also half their values now. The Universe has since then doubled in size. Cosmologists refer to this epoch as being at redshift epoch of 1. When the Universe was a third of its present size, it was at a redshift epoch, or simply redshift, of 2.

As an analogy to understand this better, let us imagine that we have all decided to mark the different epochs of our childhood by our corresponding heights at those epochs relative to our height now. We would then refer to some particular epoch of someone's life as being at a height-shift of 1, or something similar. This would mean that the height of this person at that particular epoch was half of his height at present. The exact timing of the epoch, or the age of the person at this epoch, would however depend on the growth rate of this person as a child, which would depend on a myriad of factors like nutrition, exercise and so on. Let us imagine that we do not have access to such information and we also do not know our exact age at present. Suppose we have decided to remember every interesting epoch of our childhood by the measure of 'height-shift'. For example, we may remember our first day at school not by our age then but by the fact that our height-shift was approximately 1, or that our height was approximately half our height now.

Cosmologists are also faced with a similar situation. The rate of expansion of the Universe depends on its mass (and energy) density, and on the existence of the cosmological constant, as well as on the value of the Hubble constant. We still do not know the exact values of most of these parameters, and for some of them we are not even confident of their existence (e.g., dark energy). It therefore makes sense not to talk of the ages of the epochs that cosmologists are interested in, but discuss them in terms of their redshifts. Cosmologists therefore mark the different phases of the history of the Universe with their corresponding redshifts.

One often comes across inaccurate statements regarding this in popular articles in newspapers and magazines. Writers of these articles often announce some new cosmological observations with colourful statements to the effect that some new object has been discovered 'so many billion light years away, at the edge of the Universe'. It is erroneous to talk of distances in this regard and one has to be careful about their usage. The distances between galaxies and objects are always changing. The distance to this new object when the light was emitted from it (and with which the discovery has been made) was much smaller than the distance at present. Which distance is being referred to in this statement? It is also very difficult to estimate distances to objects and there are often lively debates among astronomers about their favourite methods of measuring distances. This uncertainty is the reason behind the fact that the Hubble constant could not be measured with good accuracy until very recently. It is also erroneous to talk in terms of time without stating what assumptions have been made about the constituents of the Universe (and the value of the Hubble constant) to arrive at the estimate. It is of course difficult to explain these subtleties in the limited space available for an article in a newspaper or a magazine. There also remains the temptation of making the article colourful with sensational phrases.

What the astronomers actually gather firsthand from the light received from a distant source is its redshift which is easily determined from the position of the signature lines in the spectrum of the source. The distance of a source and the age of the Universe at the time of emission of light from this source are however not easily estimated, but depend on the assumptions of several parameters. It is therefore sensible to discuss the observations in terms of redshift instead of distance or time.

One can here give an idea of the relation between the age of the Universe at different epochs and the corresponding redshifts. Consider the simple case of a Universe with flat spacetime, and with the density equal to the critical density (that is, with $\Omega = 1$). Let us also assume that the present day value of the Hubble constant is 70 km per second for a distance of one million parsec, as mentioned earlier. In this case, the epoch of redshift 1 would correspond to a time that is about 6 billion years before the present epoch. Redshift of 2 was 7.45 billion years and redshift of 3 was 8.18 billion years ago. If we continue to go back in the past, we would reach the Big Bang at some point, when the redshift would be infinity. In this case this would happen about 9.3 billion years ago. This would then be the age of the Universe. We should however remember that this estimate is only valid for the case when the Universe has critical density and there is no cosmological constant[10].

3.6 Accelerated Expansion of the Universe

Let us now come back to the discussion of the recent discovery of an accelerated Universe. Consider a very distant source of light. An observer first determines the redshift of the source from its spectrum. If we can somehow determine its actual brightness then by comparing this with its observed brightness at Earth we will be able to determine how its brightness has decreased with distance. Astronomers have been studying the rate of expansion of the Universe from these two parameters.

One could try to explain this with the help of an analogy. We know that the buying power of money decreases with time because of inflation. What we can now buy with a given amount of money is much less than what our forefathers could with the same money. Consider the case of an imaginary country where some natural calamity has destroyed all economic and trade records. There now remains no copy of any record of its economic history. Fortunately, however, the coins used at various times in the past have survived the calamity. Suppose some historians now decide to discover the economic history of the country on the basis of these surviving coins. They are especially interested in the history of inflation of the country, and how the buying power of money has decreased or increased with time and with what rate, and so on.

It turns out that the past governments of the country had some foresight and had printed some important information on the coins whenever new coins were distributed in the country. They had not only marked the date of distribution of the coins, but also the amount of, say, rice (of some specified variety), that could be bought with a unit of currency. For example, suppose it was printed on the coins '1 Rupee : 1990 : 10 gram'. A coin from an earlier year could perhaps be marked as '1 Rupee : 1989 : 20 gram', and so on. (The rate of inflation in this example is highly exaggerated to illustrate the point better.) Encouraged by this discovery, the historians set out to unearth all the coins that they could and decipher the markings on them.

In a short while they could collect a handful of such old coins and they set out to analyse the data collected from them. They then decide to record on the one hand the year when the coin was first distributed in the market, and on the other hand record the relative amount of rice that could be bought at that time compared to the present. Suppose one could at present buy only 1 gram of rice with 1 rupee. In other words, one could only buy ten percent of what one could in 1990. The historians decide to record this information in the following way —(1990; 10%). The information for the year 1980 could, for example, be recorded as (1980; 5%), and so on. This is how they decide to record the relative prices for different years. When they plot these data points the graph would look like the one in Fig. 3.6. The years would be marked on the horizontal axis and the relative prices would be marked on the other axis. The figure shows different possible histories of inflation.

Suppose the rate of inflation in the country was steady for years. This would mean that the buying power had steadily decreased over time, or equivalently, the relative price was small in the past. Historians in this case would find that their collected data would amount to the thick solid line in the figure. If however the rate of inflation had actually fallen over the years, the historians would then find the data following the dashed line.

In this case the relative price would still decrease as one goes back in the past but at a smaller rate compared to the earlier case of steady inflation. If the rate of inflation had actually increased over time, then one would find that the dotted bottommost line best respresents the data.

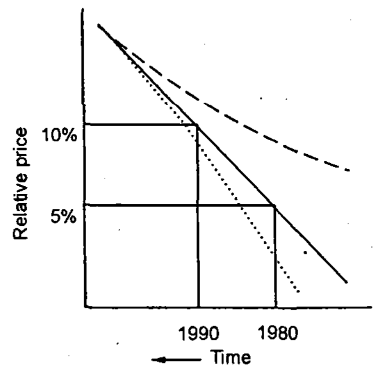


Fig. 3.6 *The relative price plotted against the year of appearance of a coin in the imaginary country.*

We then find that it is possible to determine the history of inflation if one knows the date and the relative buying power of money. In an almost similar manner, astronomers have attempted to determine the history of expansion of the Universe. In their case the date marked on the coin would correspond to the redshift of a distance source, and the amount of rice bought with it would correspond to its actual brightness. As the historians compared the original buying power with that at present, the astronomers would compare the actual brightness of the object with its observed brightness.

Fortunately for the historians in our example, a properly calibrated value of the buying power at different times was recorded. This calibration was done by using a fixed amount of rice of a standard kind.

It is however difficult for astronomers to find objects whose actual brightnesses are well calibrated. A distant luminous object could be a normal or a peculiar galaxy, or even the manifestation of some bizarre phenomenon. Recently astronomers have used a particular kind of supernova in this regard. According to them this type of supernova always reaches the same peak brightness. It is like having a bulb of a fixed wattage[11].

Astronomers have been able to find interesting clues to the history of expansion of the Universe using the relative brightnesses of these supernovae and their redshifts (Fig. 3.7). According to them the rate of expansion has been increasing in the recent past. Comparing the data with Friedmann's equations, they find that the total mass density of the Universe is about 30% of the critical density and that there is some dark energy, perhaps in the form of a cosmological constant.

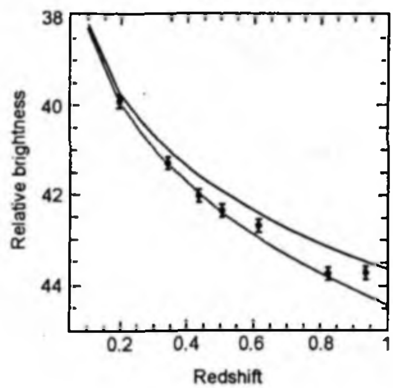


Fig. 3.7 *The relative brightnesses of supernovae are plotted against their redshifts. (The numbers on the y-axis are the difference between the apparent and absolute magnitude; magnitude being a parameter that increases with decreasing brightness. Larger differences in magnitude, i.e., toward the lower side of the plot, signify relatively dimmer supernovae.) If the data points followed the solid curve on the top, it would have meant a decelerating Universe. Since the data points follow the curve in the bottom, an accelerating Universe is suggested.*

We will learn in later chapters that the geometry of the spacetime of our Universe is Euclidean according to new observations. We had already learned that the spacetime is Euclidean in a Universe with the critical density, although we were cautioned that the presence of dark energy and cosmological constant changes all these conclusions. Combining the information of a flat spacetime with the above-mentioned information from supernovae data, it seems that the dark energy constitutes about 70% of the critical (mass–energy) density. The deficit in the mass density of the Universe is compensated by the dark energy. This is the energy of empty space that Einstein was reluctant to include in his equation. It now seems that this dark energy is responsible for the accelerated growth of the Universe. These observations are however only in the preliminary stages and there is a lot of research going on at present to try to understand the data and get more data. If however the trend is confirmed by future observations, the fate of the Universe will be to expand in this manner for ever[12].

As mentioned earlier, the relation between the age of the Universe and the redshift also changes in the presence of dark energy. If the relative importance of gravitating mass and dark energy is as mentioned earlier, then the present day age of the Universe would be around 13.7 billion years.

It is important to check whether or not this estimate of the age of the Universe is at variance with other data. It is however too difficult to estimate the age of the Universe from any direct observation. Astronomers have used three methods to estimate the age. Firstly, we know the age of the Earth or the Solar System from the observations of radioactive materials. One can date a sample of radioactive material from its decay rate. The age of the Earth has been determined in this manner to be about 4.5 billion years, and that of the Solar System is about 4.6 billion years. It is difficult to do radioactive dating for stars and galaxies. Recently however a few astronomers have discovered the presence of uranium and thorium in a very old star in our Milky Way, and have estimated its age from the relative abundance of these two elements to be between 11.7 and 18.7 billion years. The age of another star has been determined to be between 9.5 and 15.5 billion years from the decay rate of the uranium in it.

Secondly, astronomers can estimate the age of old stars from their knowledge of the evolution of stars. Such observations indicate that the oldest stars in our Milky Way are about 10.2 to 12.8 billion years old.Thirdly, astronomers also try to estimate the age of white dwarf stars from their spectra using their knowledge of how they should cool in time after their formation. They have therefore tried to discover the oldest white dwarfs in the Milky Way and their observations indicate these old stars to be of an age of around 10 billion years.

Naturally all these estimates of ages of stars must be less than the age of the whole Universe. There still remains a lot of uncertainty in the estimates of these stellar ages. It seems however that there is no discord of these estimates with the age of the Universe estimated from the supernova observations.

Finally, let us recapitulate what we have learned about the expansion of the Universe. Einstein's equations first showed that the Universe must be restless and it must either expand or contract. Edwin Hubble then discovered with the help of redshifts of distant galaxies that space is indeed expanding. This expansion is universal and does not have a centre. The rate of expansion depends on the mass (energy) density of the Universe. If the density is larger than a critical density then the Universe will contract after some point of time. Recently however astronomers have found from their observations that there is a large amount of energy associated with empty space, and that this dark energy is responsible for an accelerated growth of the Universe although the mass density of the Universe falls short of the critical density.

4

The Early Universe

> ...They claim they recollect the morn
> When unto Earth her first was born...
>
> Robert Frost, *A Wishing Well*

It took a few decades after the discovery of the expansion of the Universe by Hubble for physicists to begin thinking earnestly about the history of the Universe. One reason for the hesitation was that there was no scope for testing their ideas and theories with concrete observations. And science cannot progress without the ability to test the hypotheses made by scientists. It was therefore only a few maverick physicists who set out to use physics to learn about the history of the Universe. Notable among them was George Gamow.

Gamow was the first physicist to use the principles of physics to determine the state of the Universe after the Big Bang and its evolution. With the observations made by Hubble and the calculations of Friedmann at his disposal, he wanted to draw the history of the early Universe.

4.1 The First Three Minutes of the Universe

We have established earlier that galaxies in the past were closer to each other than they are at present. We have to think in terms of the primordial gas if we are to think of the times before the galaxies were born. The particles in this gas would come closer to each other as one goes back in the past. This would mean that the gas would progressively become denser as we go back further and further into the past. Its density has been decreasing in time because of the expansion.

There is another characteristic of the gas, other than its density, that changes with time. It is its temperature. When a balloon filled with hot gas is allowed to expand on its own, the gas inside it gradually cools. The total energy of the gas has not changed in this case; the temperature has decreased because of the increased volume. The gas in the Universe also cools with time because of its expansion. Conversely, the gas was both dense and hot in the past.

Gamow studied the thermonuclear reactions in the centre of the Sun before he began thinking about the early Universe. As a matter of fact, the mechanism obeying the laws of quantum mechanics by which protons fuse together to build helium nuclei was discovered by Gamow himself. Naturally when he began thinking about the hot and dense early Universe, he wondered if there were any thermonuclear reactions in the Universe then. He guessed that elements must have been forged in the heat of the nascent Universe, just like in the centres of stars.

We now know that his guess was correct. At least partially so, in the sense that at least some elements must have been created in the primordial Universe. If it were not so then there would be a puzzling gap in the constituents of the Universe. We learned in the first chapter that elements like helium and others are created in the centres of stars. These heavy elements are then dispersed in space by supernova explosions at the end of the lives of stars and new stars are born from this 'enriched' gas (astronomers prefer to think of this in terms of enrichment and not of pollution). These new generation stars create more heavy elements in their centres. The abundance of heavy elements thus steadily increases, being created, by every new generation of stars.

When astronomers observe very old stars they find that they are correspondingly deficient in heavy elements. There was however one set of observations that did not fit in this scenario. They had noticed that stars with very low abundance of heavy elements always had a minimum amount of helium in them. In other words, the abundance of helium was far from being low even in stars with very low abundance of other heavier elements. It was as if there was a minimum amount of helium in stars, however old and anemic in other elements they might be. This minimum amount was approximately 24% of the total mass of the star (Fig. 4.1).

Imagine a census of a nation's citizens in which statistics of their wealth is also being collected. Suppose it is required to find the distribution of wealth among its citizens. Suppose now that there is found to be a minimum amount of wealth that one can have, and that no matter how poor one may seem from other accounts, everyone has a minimum amount of possessions. This would has point towards the possibility that someone has distributed this minimum amount to everybody, perhaps the government.

Astronomers concluded from their observations of a minimum amount of helium in stars that this much of helium must have come from an earlier phase in the history of the Universe during which time it was distributed uniformly. They also noticed that the confirmation of this idea would provide strong support to the ideas of Gamow on the primordial synthesis of elements, especially helium. This early synthesised helium would be distributed uniformly in the Universe, accounting for the minimum amount of helium found in all stars.

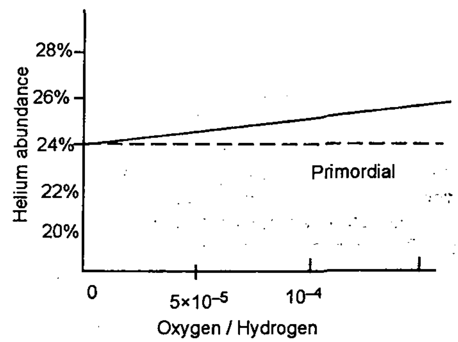


Fig. 4.1 *Observations of stars otherwise extremely poor in heavy elements such as oxygen show a minimum abundance of helium much larger than zero. This plot shows the observed abundance of helium (as a percentage of the total mass) of stars against the relative abundance of oxygen with respect to hydrogen. It is clear that there is a substantial amount of helium in stars which have extremely small amounts of oxygen, which points towards a primordial abundance of heilum (shown by the dashed horizontal line).*

Gamow and his students carried out detailed calculations of primordial nucleosynthesis in the 1940s. It was somewhat complicated compared to the calculations of nucleosynthesis in the centres of stars. Although protons in this case will also fuse together to form deuterons and then helium nuclei, the expansion of the Universe makes the process complicated to analyse. On the one hand, the particles combine to form heavier nuclei and on the other hand, the expansion of the Universe tends to take them apart. It is not just that they are moving apart, but one also has to remember that all nuclear reactions have a threshold temperature below which they cannot take place. The temperature of the gas however keeps on cooling due to the expansion. The rates of reactions therefore slow down to become negligible at some point.

In the early Universe, the reactions which would have required a larger threshold temperature took place early and they then slowly froze out, in the sense that the state of the reactants become frozen in the absence of reactions. The reactions which required lower temperatures then dominated the state of affairs. One can then sketch the history of the Universe in the following manner.

When the Universe was younger than a hundred-thousandth of a second, particles as massive as protons did not exist. The Universe consisted of particles that make up heavier particles like protons and neutrons, and are called quarks (Fig. 4.2). Usually the quarks bind together to form protons and neutrons and other particles at the temperatures we are more used to. It is only at very very high temperatures that it is possible for them to move about freely and form a sea of quarks. When the temperature of the Universe decreased somewhat, the quarks came together to form particles like protons and neutrons. There were some reactions taking place at that time which could change protons to neutrons and vice versa. When the Universe become a few seconds old, those reactions too became irrelevant. The protons and neutrons could not change into one another any longer and became 'frozen' in their respective states. They then began forming heavier nuclei like deuteron and helium in earnest. This went on for about three minutes[1]. The density and temperature of the Universe became so less after this epoch that it was not possible to sustain any nuclear reactions anymore. Although a small amount of a few nuclei heavier than helium—like lithium and berylium—was forged during this process, very heavy elements like carbon and so on did not get an opportunity to be synthesised in this phase. The era of primordial nucleosynthesis came to an end before their turn could come up.

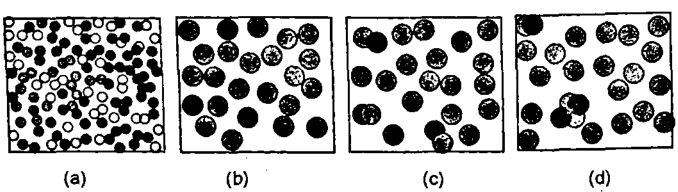


Fig. 4.2 *The sequence of the important epochs in the early Universe is shown here. In the beginning, when the Universe was younger than a hundred-thousandth of a second, it was the realm of quarks (a). As the temperature decreased after a while, quarks came together to form protons (grey) and neutrons (black) (b). Neutrons then combined with protons to form deuterons (c) and then helium nuclei (d).*

Physicists calculated that soon after the end of the metamorphosis between protons and neutrons, their ratio was such that for every 14 protons there were 2 neutrons[2]. In other words, out of every 16 particles there were 2 neutrons. All these neutrons got incorporated in some or other helium nuclei. A helium nuclei however has 2 protons and 2 neutrons. The 2 neutrons mentioned earlier will then combine with 2 protons to form a helium nucleus, and (14 – 2 =)12 protons will remain as leftovers. These protons will ultimately form hydrogen atoms. We then find that 4 out of 16 particles were converted to helium. In other words, approximately 25% of all particles formed helium nuclei. Since the masses of protons and neutrons are approximately of the same order one can say that the primordial nucleosynthesis gave rise to helium particles that amounted to 25% of the total mass of the Universe.

This fraction of 25% does sound familiar from our earlier discussion of the minimum fraction of helium in stars. The estimate of helium created in the very early Universe from nucelosynthesis does match well with the observations of astronomers. This above estimate depended only on the calculations of Friedmann (which are based on the general theory of relativity) and the observations by Hubble. It is so straightforward that there is hardly any room for a large change in the estimate. This match was therefore a big success for cosmologists and provided a major support for the Big Bang model. It was clear that a hot early Universe was not just a figment of the imagination and it was a big step toward making cosmology a scientific enterprise[3].

4.2 Microwave Background Radiation

We have so far neglected an important constituent of the Universe in our discussion. Photons are particles which contribute a large amount of energy to the total energy content of the Universe. One should also take into account the role of radiation in the history of the Universe. Radiation plays a major role in reactions in which particles metamorphise into one another or change their states. On the one hand, the energy of radiation will give rise to new particles and on the other hand, particles will annihilate antiparticles to form pure radiation.

Gamow realised that radiation during this era of the Universe would have a distinct property which it acquires on account of its close relationship with matter. Since the Universe was hot and dense, a state of thermal equilibrium was possible between matter and radiation. Particles would have absorbed some radiation and then emitted some to keep the balance. This balance or the state of equilibrium endows the radiation with a tell-tale sign of the strong interaction between matter and radiation.

If we want to visualise a situation where matter absorbs all radiation falling on it, and emits some amount of radiation, we will need to ensure that there is no reflection of light. The process of reflection of light means that not all the photons interacting with matter are getting absorbed. It is difficult to think of any such situation in daily life. One could however find an approximate situation in the case of an oven. The chances of reflection of light from outside sources is very small deep inside an oven. Any photon found inside an oven must have been emitted by the (hot) walls, and given enough time to settle down at a certain temperature, the radiation will be in equilibrium with the matter making up the oven. Let us find out how the properties of this radiation change with the temperature of the matter in equilibrium with it.

Here we need to change our unit of temperature from Celsius or Fahrenheit to that of kelvin. Physicists use this unit of temperature as its zero corresponds to a very special temperature, when the inherent movement of atoms and molecules is zero (at least classically). It is not difficult to switch from Celsius to kelvin units—the steps in this unit are similar to that of Celsius scale only that the zero here corresponds to −273 degrees Celsius. In other words, one gets the temperature in Kelvin scale by adding 273 to the temperature in Celsius. 37 degrees Celsius will therefore be equivalent to 310 kelvin, and so on.

Going back to the case of radiation in equilibrium with matter at different temperatures, if the temperature is about 4000 kelvin (or, 3727 degrees Celsius) the radiation will look fairly red. There will of course be photons of all colours in it, but those of red colour will dominate. If the temperature of the oven is increased now to 5000 kelvin, then the intensities of all the colours will increase somewhat, but photons of yellow colour will dominate now. If the temperature is further increased, the most dominant colour will be blue. The intensity of the blue part of the spectrum of this radiation will be the largest (Fig. 4.3).

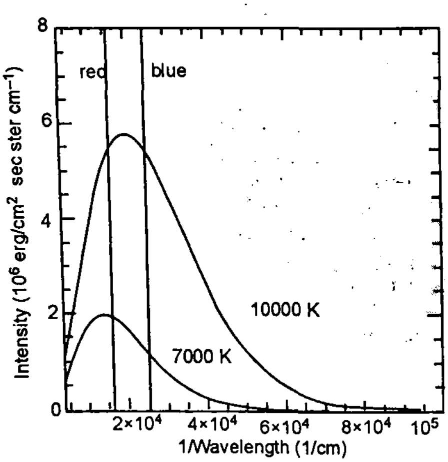


Fig. 4.3 *The diagram shows the spectrum of the radiation that is in thermal equilibrium with matter at different temperatures; that is the dependence of the intensity of radiation on the wavelength is being shown for different temperatures. The vertical axis plots the intensity (in the unit of 0.1 joule (one million erg) per square cm per second per steradian per (1/cm)) and the horizontal axis plots the inverse of the wavelength (in the unit of 1/cm). The wavelengths that are in the range of visible light are shown by their colours (blue, red). It is seen from the diagram that radiation in thermal equilibrium with matter kept at 7000 kelvin is dominated by red colour. As the temperature is increased, shorter wavelengths begin to dominate. At 10000 kelvin the radiation is much bluer than the earlier case.*

Astronomers concluded from their observations of a minimum amount of helium in stars that this much of helium must have come from an earlier phase in the history of the Universe during which time it was distributed uniformly. They also noticed that the confirmation of this idea would provide strong support to the ideas of Gamow on the primordial synthesis of elements, especially helium. This early synthesised helium would be distributed uniformly in the Universe, accounting for the minimum amount of helium found in all stars

If we now recall our earlier discussion on the relation between energy and colour of photons, we will realise that as temperature increases, the dominance of short wavelength photons, or photons with high energy becomes conspicuous. One can give similar examples from our daily life apart from that of the oven. When a piece of iron is heated it initially looks red, then orange and then yellow, as its temperature is increased. Similarly, one finds the portion of air closest to the wick of a candle to be blue and the air outside of it, where the temperature is relatively small, to be yellow or orange.

Physicists had discovered this special behaviour of radiation at the end of the nineteenth century. They found that the spectrum of the radiation that is in thermal equilibrium with matter depends only on the temperature and not on the composition of the matter. It does not matter if the oven is made of iron or gold. If the temperature is the same then the radiation will be similar. In fact one often calls it the temperature of the radiation. Since we have in mind objects which do not reflect any light to us but absorb every bit of radiation falling on them, this radiation is also called the 'black body radiation'.

The discovery of the black body spectrum, as it is usually called, is associated with the names of many important physicists. Max Planck was the first physicist who obtained the full characteristics of this radiation and correctly determined how the intensity varied with wavelength[4]. The black body spectrum is therefore often called the Planckian spectrum. A deeper understanding of the spectrum however came much later, with the work of Satyendranath Bose. It was then understood that this spectrum describes not just photons but also other particles of a similar nature. These particles are called 'bosons' in honour of Satyendranath Bose.

Let us now go back to the arguments put forward by George Gamow. According to him the radiation in the very early Universe interacted strongly with matter and was in a state of thermal equilibrium with it. It then followed that this radiation will have a Planckian spectrum. When the temperature of the Universe was very high, the radiation was then dominated by very short wavelength photons. As the Universe expanded and cooled, longer wavelength photons gradually took over.

This means that such radiation should still exist, only it would be dominated by very low energy or long wavelength photons. Whatever the energy of the dominant photons might be, a Planckian spectrum is expected of the radiation. It is an inevitable outcome of the Universe once having been very hot and dense in the past. This radiation would therefore support such a scenario.

Gamow had predicted this in a paper written in 1948 with two of his students[5]. They had shown that the temperature of the radiation is closely related to the sequence of events in the nucleosynthesis epoch. They predicted that the present day temperature of this radiation would be a few kelvin. From Planck's formula it then follows that this radiation at present will be dominated by a very low energy microwave radiation. The wavelength of such a radiation is about a few millimetres. One uses these radiations for television or communication over a large distance, and also in microwave ovens. The intensity of the cosmic microwave radiation will however be very low. One will need a very sensitive instrument to be able to detect it.

Such instruments did not exist at the time of Gamow's article, soon after the end of the second world war. Physicists soon forgot about this prediction and of trying to build instruments to look for it. Perhaps these calculations on the early Universe were perceived to be too exotic to warrant much attention. Gamow and his students also failed to inspire any astronomer to look for such a radiation.

After a span of fifteen years a few physicists began to think about this radiation again. Yakov Zel'dovich of Moscow started investigating the effects of an early hot Universe along the lines of Gamow. Two of his students, Andrei Doroshkevich and Igor Novikov wrote an article urging astronomers to look for this radiation. They argued in favour of such an experiment saying that the instrument best suited to this purpose was a radio telescope in New Jersey, USA, which was operated by AT&T Bell Laboratories. Unfortunately most physicists remained unaware of their article, which was written in Russian.

At the same time, in 1965, the AT&T Bell Laboratory was doing some experiments on communication via artificial satellites using the radio telescope at New Jersey. Arno Penzias and Robert Wilson tried to find out how one could detect very feeble signals from space with it. They however soon noticed that no matter how sensitive they tried to make the instrument, a noise persisted in the detector system that precluded the possibility of detecting very feeble signals. They tried to get rid of the noise by many means, including removing pigeon droppings from the antenna by crawling inside it. They were however unable to decrease the noise below a certain point. They also noticed that this noise had the same intensity no matter which direction in the sky the telescope was aimed at. From the intensity of the noise they estimated that if this were due to a blackbody radiation, the corresponding temperature will be around 3 kelvin. This was puzzling as they were unaware of any known source of such a radiation in the Universe.

Within a few kilometres of Penzias and Wilson's telescope, a few physicists at Princeton were rediscovering the calculations of a hot early Universe, unaware of the fact that Gamow and his students had already published their calculations on the same. They realised, independently of Doroshkevich–Novikov and Gamow and his students, that they needed to detect very faint microwave radiation from space to test these ideas. Unlike others though, they proceeded to build an instrument for this purpose and began their observations with it.

It is not uncommon to encounter such a situation in the history of science, when scientists seem to be blissfully unaware of the existence of ideas and proofs right under their noses. These situations are at best temporary though, and in this case too it turned out to be so[6]. Physicists very soon realised that what they had discovered was the background radiation that pervades the whole Universe and was a relic of the early hot Universe[7,8].

We often perceive this radiation in a curious way. When a television set stops receiving signals from a TV station due to some problem, we find the screen to be full of luminous points flickering in it. When one removes the TV cable and tries to receive the signal through an antenna, one finds the same spots. About a few per cent of these innumerable flickering points is due to the cosmic microwave background radiation(Fig. 4.4). This is because microwave radiation is used to telecast TV signals and a television set occasionally receives microwave photons from the cosmic background.

Fig. 4.4 *When a television set does not detect signals from any station due to some failure, one sees innumerable luminous points flickering in the screen, about a percent of which is due to the cosmic microwave background radiation.*

Physicists however had to wait for a few decades to fully determine the spectrum of this radiation . They could analyse the spectrum with great accuracy only in 1990 with the help of data from a satellite named COBE (Cosmic Background Explorer). This observed spectrum matched closely the Planckian formula for a blackbody spectrum, just as one would expect if matter and radiation were in a state of thermal equilibrium, or, equivalently, interacted strongly with each other, at some point in the history of the Universe. That is, if the Universe were hot and dense at some point or if the Big Bang model of the Universe had any truth in it.

The match between the spectrum observed with COBE and the theoretical Planckian spectrum is indeed striking (Fig. 4.5). We had earlier mentioned that it is difficult to find a good example of black body radiation in real life; the example of an oven was only an approximation to the theoretician's idealised situation. It now turns out that the cosmic microwave background radiation is the perfect example of a black body radiation. And this characteristic of the cosmic radiation is a big support for the Big Bang model of the Universe.

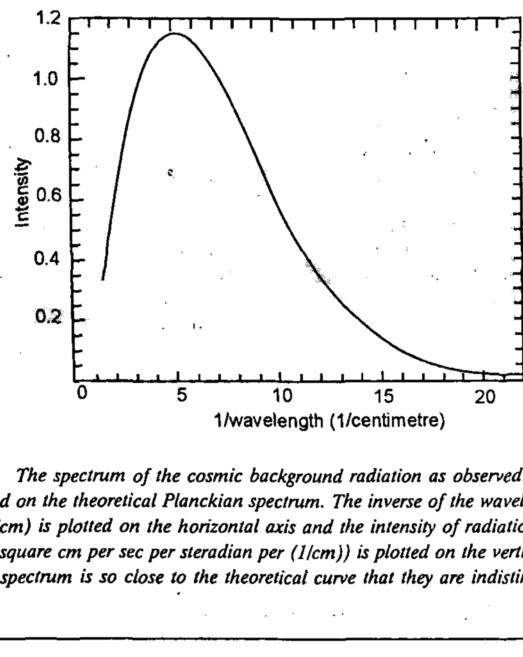


Fig. 4.5 *The spectrum of the cosmic background radiation as observed by COBE is superposed on the theoretical Planckian spectrum. The inverse of the wavelength (in the unit of 1/cm) is plotted on the horizontal axis and the intensity of radiation (in* 10^{-11} *joule per square cm per sec per steradian per (1/cm)) is plotted on the vertical axis. The observed spectrum is so close to the theoretical curve that they are indistinguishable in this plot.*

This discovery created a ripple in the study of cosmology and elevated it into an exact science. Earlier its practitioners were ridiculed as being wild speculators, and famous physicists like George Gamow were not spared. From this point on, however, it became difficult to ignore cosmology and keep it isolated outside the mainstream of physics.

4.3 A Brief History of the Universe

After having discussed the microwave background radiation, we can now go back to our earlier discussion of the early Universe. We now have all the ingredients to discuss the important epochs in the history of the Universe. We will indicate the different epochs by their redshift and approximate age of the Universe at that time. We had earlier seen how the redshift gives an idea of the relative size of the Universe at different epochs. If one has in mind the redshift epoch of 5, it means that the size of the Universe then was a sixth of its size now. We will also indicate the corresponding temperature of the Universe at different epochs.

Very early Universe

Redshift : more than a trillion
(Or, *equivalently, the size of the Universe is a trillionth of its size at present*)
Temperature : more than a trillion (10^{12}) kelvin
Approximate age : less than a hundred thousandth of a second

The Universe was so hot at this stage that matter was in the form of quarks and particles of the same kind. Protons and neutrons, and other more familiar particles are made of a few quarks. At very high temperatures, the quarks are able to move freely, and particles like protons cannot exist. Some recent experiments in particle accelerators have attempted to create this phase of matter and there is a lot of research going on about it.

There was an important event during this very early phase of the Universe. There are some indications from observations of the present day Universe which has led physicists to consider the possibility of a period of enormous growth of the Universe at this phase. This epoch of 'inflation', as it is called, has left its mark on the structure of the Universe today.

We will find out in the later chapters that observations indicate a spacetime geometry that is Euclidean and flat. It is, however, difficult to understand why it should be so with no *a priori* bias on the possible geometries. Physicists guess that although the spacetime is apparently flat inside our observable Universe, it might not be so at a much larger scale. Although the Earth is round it does seem flat on a small scale, when length scales much smaller than the horizon are considered. It is therefore possible that the geometry is different on a much larger scale of the Universe and it is only flat within the small region of the observable Universe (Fig.4.6). According to the physicists, a period of inflation of the Universe in the beginning would have ensured that any small region of the Universe would look flat enough. The actual reasons and mechanism for this hypothetical period of inflation remain inconclusive though.

Apart from the inflation, some other physical process may have taken place during this phase of the Universe about which we do not have much knowledge yet. Our knowledge of the properties of matter at such high temperature and density is very limited. Research on the fundamental particles, like quarks, will make valuable contributions to the study of cosmology in the future. We will simply remember here that when the temperature of the Universe dropped below about a trillion degrees, familiar particles like protons and neutrons appeared in the Universe and the era of nucleosynthesis gradually began.

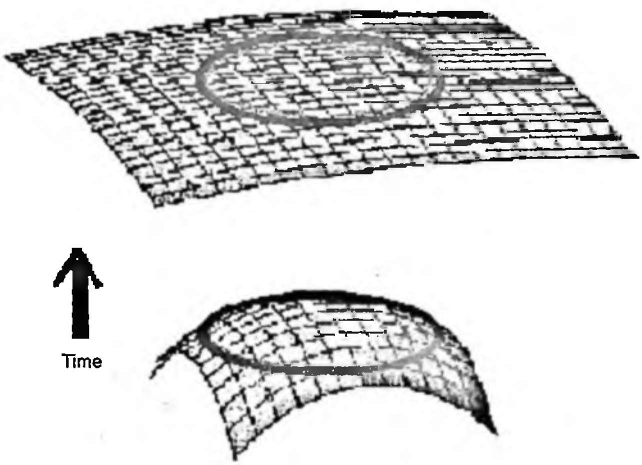


Fig. 4.6 *A small region on a surface of any global shape will seem to be flat when inflated by a large amount. Consider being on a curved surface and suppose the circle drawn on it denotes the horizon. As the curved surface is inflated the region inside the horizon will progressively become flatter.*

Primordial nucleosynthesis epoch

Redshift : About one billion
Temperature : About one billion (10^9) kelvin
Approximate age : Three minutes [1]

This is a landmark epoch in the history of the Universe. Primordial nuclear reactions have slowly come to an end at this phase. The reactions have so far converted 25% of the total mass of the Universe into helium nuclei, and a bit of deuteron, lithium and beryllium. As the Universe continues to expand and cool, these once pervasive nuclear reactions slowly fizzle out.

We learned in the first chapter that many heavy elements are created in the nuclear reactions at the centres of stars. The results of these stellar nuclear reactions are visible as the iron in our blood or the calcium in our bones. It is difficult to find any relic residue from the primordial nucleosynthesis epoch in our surroundings. The minimum abundance of helium in stars is a mere indicator of this epoch. There is however one clear residue of the primordial nucleosynthesis era and it is deuterium. The deuterium atom has a nucleus (called deuteron) with one proton and a neutron. When helium is fused in the stellar cores, deuterium nuclei are produced in the intermediate stage, but all of them are assimilated into the final helium nuclei and no deuteron survives after the process. It therefore means that the small amount of deuterium present in the Universe dates back to the primordial nucleosynthesis epoch, because no fresh deuteron has been synthesised after that. This means that the deuterium in heavy water, which is very important for the production of atomic energy these days, owes its existence to this important epoch in the Universe.

Recently astronomers have made some interesting observations about the abundance of deuterium in the Universe. We had earlier seen how the calculations for the abundance of helium from nucleosynthesis matches well with the observed values. The calculation for the abundance of deuterium is slightly more complicated and depends strongly on the density of the Universe.

We know that a deuterium nucleus is formed out of the fusion of two protons. One of the protons becomes a neutron (releasing a positron and a neutrino) in the process of fusion. This deuteron then fuses with two other protons in succession to form a helium nucleus. If the density of the matter is high then the rate of formation of these helium nuclei from deuterons is also enhanced. This is because the rate of reaction is proportional to the density of participant nuclei. As the density of the Universe decreases with time, the rate of this reaction slowly decreases. There is therefore a competition between deuterons coming close to fuse and moving apart due to the expansion of the Universe. At some point of time, the fusion reactions between deuterons become rare and negligible. The left-over deuterons will then combine with electrons to form the deuterium atoms in the Universe.

This means that the higher the density of matter, the larger the fraction of deuterium nuclei fusing to form helium and the smaller the fraction of left-over deuterium nuclei. In other words, the present day abundance of deuterium atoms would depend crucially on the density of matter. Astronomers have therefore been interested in this important but difficult observation, as it will directly measure the density of the Universe[9].

According to their observations, the density of matter is about five per cent of the critical density. Alert readers would be suspicious here of a mismatch with the density of matter reported earlier. We had learned in the third chapter that the supernovae observations indicate a density of about 30% of the critical density. Here we find that deuterium observations indicate a much smaller density of about 5% of the critical density. Whatever happened to the other 25%?

As a matter of fact one has two different kinds of matter in mind here. The density obtained from supernovae observations pertain to the *total* mass density of the Universe. It is not clear if the inventory of all types of matter includes some exotic particles which do not participate in nucleosynthesis reactions. If there are indeed such particles, then the density obtained from deuterium abundance will pertain only to the density of familiar particles like protons and neutrons, which participate in nuclear reactions as we know. The exotic particles, which nevertheless gravitate, will then be contained in the other 25% of matter. This means that our familiar kind of matter constitute only about 5% of the critical density, and particles still unknown make up about 25%. and together the total mass density of matter in the Universe is about 30% of the critical density. We are therefore driven to conclude that most of the matter in the Universe is of a kind that is still unfamiliar to us.

Going back to the history of the Universe, as the Universe expanded and cooled, matter still remains too hot for a long time to form atoms. Different nuclei need to combine with electrons to form neutral atoms. The temperature of the Universe at this stage was however too high for electrons to be bound to atoms. Electrons therefore roamed free and interacted strongly with other particles, including radiation. And the Universe remained in this phase for a few hundred thousand years.

Recombination epoch

Redshift : 1100
Temperature : 4000 kelvin
Age : Approx. 3×10^5 years

The temperature now permitted electrons to bind to different nuclei. Since most of the matter was made of hydrogen anyway, their nucleus, which means a proton, was now capable of holding on to an electron and forming a neutral atom.

This created some problem for the photons though. So far the photons had interacted strongly with free electrons. The interaction between photons and electrons bound inside atoms is of a very limited nature. Only photons of certain wavelengths are allowed in this case to interact. It turns out that the cosmic background radiation at this epoch was dominated by infrared photons and could not interact strongly with newly formed atoms.

The photons of the cosmic radiation therefore began to travel freely without interacting much with matter any longer. Cosmologists call this the 'decoupling era' *signifying the divorce between radiation and matter.*

The degree of ionisation of matter depends not only on its temperature but also its density. This dependence was derived by Meghnad Saha, and is widely used in stellar astrophysics to deduce the degree of ionisation of different elements at various temperatures in stars. Saha's ionisation law can also be used here to estimate the redshift of the decoupling era. This redshift turns out to be of the order of 1100, which means an approximate age of a few hundred thousand years.

The cosmic background radiation slowly shifted outside the realm of visible and infrared radiation after this epoch. The dominant photons of this background radiation slowly acquired the wavelengths of microwave radiation. The Universe became literally covered with darkness at this stage.

The matter in the Universe continued to cool and become tenuous with age. The three important epochs are schematically explained in Fig. 4.7.

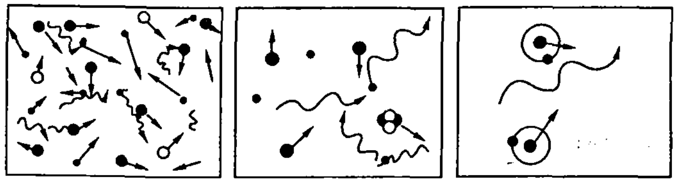


Fig. 4.7 *Three important epochs in the history of the Universe are schematically explained here. The first panel on the left shows the state of the Universe at the end of the very early phase, when protons (black filled circle), neutrons (white filled circle) and electrons (black point) and other particles interacted strongly with radiation (wavy lines). The second panel in the middle shows the state at the end of the nucleosynthesis era, when a few helium nuclei (with two protons and two neutrons) have formed. The last panel on the right shows the state after the recombination epoch, when protons and electrons have combined to form atoms but the interaction with photons has plummeted. It has also been indicated here how the wavelength of radiation and distances between particles have increased with time.*

This dark age of the Universe lasted for a few billion years. The matter in the Universe was not however completely dormant, and silently went about the task of creating structures in the Universe, the structures that later manifested itself in the form of stars, galaxies and clusters of galaxies. The warmth of the first stars slowly lifted the Universe out of its dark age. We will discuss this process in detail in the next two chapters. There are a few interesting aspects of the cosmic background radiation that will be useful later and we will discuss them here. This discussion will also emphasise how cosmology has become part and parcel of modern astrophysics.

4.4 Spots in the Sky

We have seen how the emergence of neutral atoms brings about the parting of ways between matter and radiation in the Universe. Even after this divorce though the spectrum of the radiation keeps intact its Planckian nature. This spectrum remains a souvenir of the close relation between matter and radiation in the past. Interestingly, the observed radiation also carries some information about the present day state of affairs in the Universe. For example, astronomers have discovered the motion of our Galaxy through the observations of this background radiation.

They used a phenomenon discovered in the nineteenth century for this observation. It says that the wavelength of radiation from a source suffers a change when the source is in a state of motion. When the source moves toward the observer, the wavelengths would appear to decrease. Similarly, for a receding source, of radiation, the wavelengths would appear to increase. The degree of change will depend on the relative velocity between the source and the observer, and is in fact proportional to it.

This phenomenon is common to all types of waves, from sound waves to that of light. We often encounter this effect in acoustic waves when we hear the siren of a police car speeding away from or toward us. In the first case the pitch of the siren, which indicates the wavelength, goes down and in the second case it goes up. In other words, the wavelengths of acoustic waves increase or decrease depending on the relative velocity. As a matter of fact, if instead of the source we were to move toward it, the wavelength would again appear to shorten[10] (see Fig. 4.8).

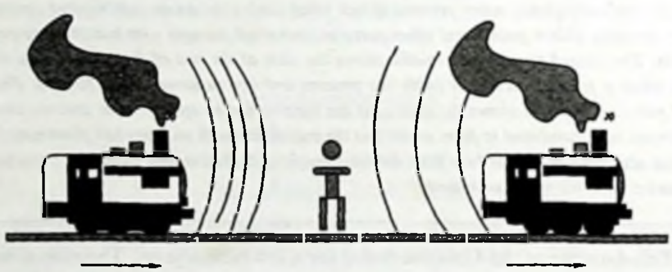


Fig. 4.8 *When the source of radiation (a train in the picture) moves toward us, the distance between the crests of the waves decreases (left panel). In other words, the wavelength will decrease. Similarly, for a receding source of radiation (right panel), the wavelength will increase.*

The same argument also holds for light waves. Photons from a source of light moving away from the observer will appear to have wavelengths larger than those from a static source. It will be the same if the source is static and the observer moves away from it.

Let us now turn our attention to the cosmic background radiation. We know that this radiation pervades the whole Universe, and that it cools with the expansion of the Universe which takes galaxies away from one another in a homogeneous and isotropic fashion. This means that if the galaxies follow the expansion of the Universe, the background radiation will be isotropic, that is, look the same in all directions. If however a galaxy deviates slightly from the general expansion of the Universe, say, as a result of local gravitational push and pull, then the radiation will not be isotropic. From the vantage point of the observer in this galaxy, the radiation will appear to have shorter wavelengths in the direction of its (deviant) motion and longer wavelengths in the opposite, as a result of the Doppler effect. The radiation will appear to be slightly bluer and hotter in the forward direction and redder and cooler in the opposite direction.

We had seen earlier that galaxies often deviate from the universal expansion because of local gravitational fields. The motion of the Milky Way and the Andromeda galaxy is an example. The centre of gravity of our Local Group, comprising these two major galaxies and a number of smaller entities, is also moving toward the neighbouring Virgo cluster of galaxies. These deviant and peculiar motions of galaxies will then cause the cosmic background radiation to have a slightly higher temperature in one direction in the sky and a slightly lower temperature in the opposite direction (Fig. 4.9).

Fig. 4.9 *The temperature of the cosmic background radiation in different parts of the sky is shown here. The temperature map is plotted in the Galactic coordinate system, in which the central position points towards the centre of the Milky Way, and the extreme point on the right (which is coincident with the extreme point on the left) points in the opposite direction, as seen from the Earth. The temperature is slightly higher in one direction and equally lower in the opposite direction (Picture courtesy : NASA). [Also see Plates]*

Astronomers have found the tell-tale signature of this effect in the cosmic microwave background radiation. Although it is more or less isotropic and has a uniform temperature all over the sky, there is a small difference of order 0.2% (two parts in a thousand). The temperature of the radiation is slightly higher by this amount in one direction and lower by the same amount in the opposite direction.

This signature is so clear that it is difficult to think of any other explanation than that of the Doppler effect due to the peculiar motion of our

Milky Way. As a matter of fact, the direction of the hotter temperature coincides with that of the Virgo cluster of galaxies and it is reasonable to assume that the gravity of this cluster is the culprit behind our peculiar motion. It is then straightforward to use the Doppler formula to estimate the speed with which our Milky Way is moving toward the Virgo cluster, and it turns out to be approximately 600 km per second[11]. This is the speed that was referred to in the first chapter.

Astronomers have now found differences in the temperature of the cosmic radiation in different parts of the sky at an even finer level. They found in 1992 with the help of observations from COBE that this radiation is anisotropic to some extent (Fig. 4.10). The level of anisotropy is much smaller than that of the Doppler effect mentioned earlier. It is of the order of ten parts in a million. Although small, cosmologists attach a lot of significance to this minute anisotropy.

Fig. 4.10 *The finer anisotropy of the microwave background radiation, as detected recently by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite (Picture courtesy: NASA/WMAP Science Team). [Also see Plates]*

According to them this is the signature, like a fossil from the past, of the process of structure formation in the Universe. Cosmologists have attempted to sketch the history of the first stars and galaxies following this line of approach, and we will discuss it in the next chapter.

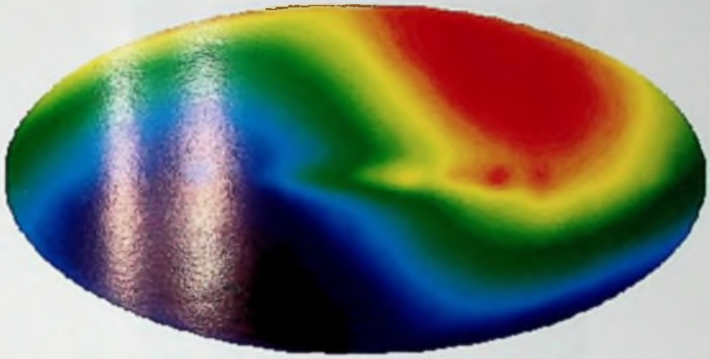


Fig. 4.9 *The temperature of the cosmic background radiation in different parts of the sky is shown here. The temperature map is plotted in the Galactic coordinate system, in which the central position points towards the centre of the Milky Way, and the extreme point on the right (which is coincident with the extreme point on the left) points in the opposite direction, as seen from the Earth. The temperature is slightly higher in one direction and equally lower in the opposite direction. (Picture courtesy : NASA)*

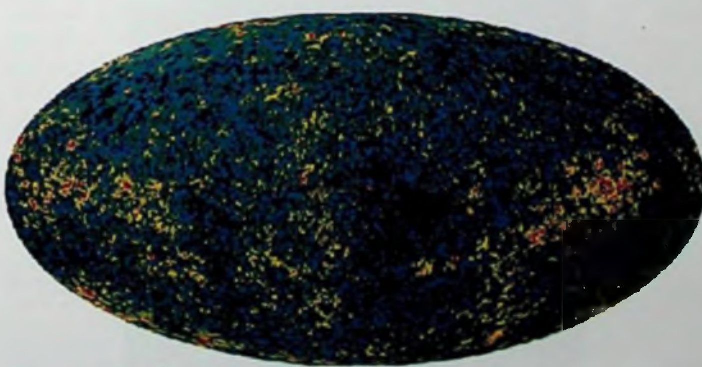


Fig. 4.10 *The finer anisotropy of the microwave background radiation, as detected recently by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite (Picture courtesy: NASA/ WMAP Science Team).*

5

Structures in the Universe

> ... The Universe is but the Thing of things,
> The things but balls all going round in rings.
> Some of them mighty huge, some mighty tiny,
> All of them radiant and mighty shiny...
>
> Robert Frost, *Accidentally on Purpose*

We might be puzzled if we compare the present day state of the Universe with what we have gathered about the early Universe. We saw that particles of matter (and radiation) were in close proximity and interacted strongly with each other in the early phase. It is then implied that matter was very homogeneously distributed in space since close interactions would tend to wash out any patchiness and wrinkles in the distribution. The present day state of the Universe is however far from being homogeneous. There are structures like galaxies (and stars inside them). It has also been pointed out in the first chapter that the distribution of galaxies is far from being uniform. Galaxies tend to cluster into groups and larger clusters of galaxies, which then form even bigger associations.

The prime question that confronts modern cosmologists is that of explaining the emergence of these structures out of a more or less homogeneous Universe. After a few decades of intense research it now seems that gravity plays an important role in this process.

5.1 Gravitational Instability

A peculiar aspect of gravitational interaction is that there is only one 'charge'—there is no duality of positive and negative mass. This makes gravity always attractive in Nature. This characteristic of gravity makes it stand apart from other forces in nature. The existence of two different electrical charges, for example, makes it possible to have either attraction

Figure 6.7 *In this picture of M82, a starburst galaxy, taken by the Subaru telescope in Hawaii, shows hot gas coming out of it. The elongated object is the galaxy itself. The temperature of this hot gas is around a million degrees. (Courtesy: NAOJ. Copyright Subaru Telescope, NAOJ. All rights reserved)*

Figure 7.4 *Many distant galaxies can be seen in this picture taken by Hubble Space Telescope. (Courtesy: NASA and STScI)*

5

Structures in the Universe

... The Universe is but the Thing of things,
The things but balls all going round in rings.
Some of them mighty huge, some mighty tiny,
All of them radiant and mighty shiny...

Robert Frost, *Accidentally on Purpose*

We might be puzzled if we compare the present day state of the Universe with what we have gathered about the early Universe. We saw that particles of matter (and radiation) were in close proximity and interacted strongly with each other in the early phase. It is then implied that matter was very homogeneously distributed in space since close interactions would tend to wash out any patchiness and wrinkles in the distribution. The present day state of the Universe is however far from being homogeneous. There are structures like galaxies (and stars inside them). It has also been pointed out in the first chapter that the distribution of galaxies is far from being uniform. Galaxies tend to cluster into groups and larger clusters of galaxies, which then form even bigger associations.

The prime question that confronts modern cosmologists is that of explaining the emergence of these structures out of a more or less homogeneous Universe. After a few decades of intense research it now seems that gravity plays an important role in this process.

5.1 Gravitational Instability

A peculiar aspect of gravitational interaction is that there is only one 'charge'—there is no duality of positive and negative mass. This makes gravity always attractive in Nature. This characteristic of gravity makes it stand apart from other forces in nature. The existence of two different electrical charges, for example, makes it possible to have either attraction or repulsion between two arbitrary charges. There is no scope for repulsion in the case of gravity (leaving aside the role of dark energy in cosmology).

This attractive nature of gravity produces an instability which has created the structures in the Universe. Consider very tiny inhomogeneities in the distribution of matter in the Universe in the past. It can then be shown that gravitational instability makes this tiny non-uniformity grow into tangible structures in the Universe. An instability is a runaway process which continues to take a system away from stability. In the case of gravitational instability, tiny clusters of matter may grow into denser and larger clusters, because of the attractive nature of gravity.

One could give an analogy from daily life to explain the nature of instabilities. Consider the schedule of buses on a certain route in a city. Suppose that the bus stops are at equal distances from each other and suppose that buses take ten minutes to travel the distances between them. Let us also suppose that the schedule of buses has been fixed after taking into account the average number of persons likely to wait at every stop and the time taken by them to get on the buses. Usually the buses run on time and so the duration between them remains the same. Suppose that one day there is a large crowd waiting for a bus at a certain stop. There is then a delay in taking all the passengers aboard at this stop. The bus then arrives at the next stop a bit later than the scheduled time of arrival. In the mean time, the second stop now has a larger than average number of people owing to this delay, and the bus is further delayed in taking these people aboard. Suppose the schedule of this bus for arrival at different stops is as follows : 10 am, 10:10 am, 10:20 am, 10:30 am and so on. On that day, the actual arrival times then become, say, 10 am, 10:11 am, 10:22 am, 10:33 am. It should be noticed that the bus is being delayed at every stop.

Let us now consider the next bus following this one. At the first stop, the first bus had waited for an extra duration and so people who had come slightly late could also get in. There would then be a smaller than average number of people waiting for the second bus. The second bus would therefore take a shorter time than usual in taking on the passengers at the first stop. It would also reach the second stop slightly early, and find a smaller number of people at this stop as well. This would make it leave even earlier. It would therefore continue to arrive earlier at all the stops. Suppose the schedule of arrivals for this bus is : 10:10 am, 10:20 am, 10:30 am and so on. The actual timings on that day would become, say : 10:09 am, 10:18 am, 10:27 am and so on.

This will essentially decrease the time interval between the buses. Instead of a fixed interval of ten minutes at every stop, the interval will shrink to 9 minutes at the second stop, to 8 at the next and so on. At some point they may perhaps arrive at some stop around the same time.

This is a common phenomenon in nature in which small differences increase without any control and is an example of instability. The attractive nature of gravity also gives rise to a type of instability. Suppose the distances between particles of matter are equal to begin with. If now there is a small deviation somewhere, i.e., there is a small difference in the distances between them in some region, this deviation will tend to grow as a result of gravity.

Particles which were slightly closer as a result of the initial deviation will tend to attract one another more. This is because the gravitational force increases with decreasing distance. This will bring them closer, which will make the force between them even larger, bringing them even closer, and so on. On the other hand, the particles which were slightly farther apart than usual will now have their interaction diminished to some extent. These two factors will ensure that small deviations grow in time (Fig. 5.1).

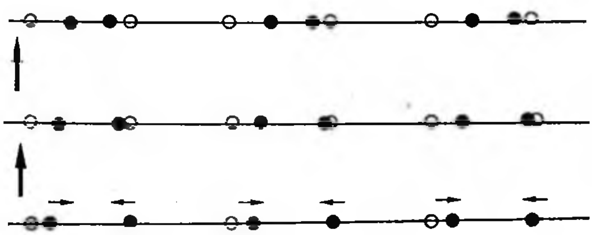


Fig. 5.1 *The black circles in this figure show the (actual) positions of particles. The white circles show the case if they had been equidistant from one another initially and had remained so. If however the particles happen to be slightly closer in some regions initially (the lowest panel), then as time grows, the deviations (difference between the idealized white circles and actual black circles) grow. At some point, the particles may even collide with one another.*

Deviations in the inter-particle distances also mean deviations from the uniform density of matter. The density is larger than average where particles have come together and is smaller than average in the regions from where the particles have moved away. In other words, gravitational instability increases the non-uniformity or inhomogeneity in the matter density (Fig. 5.2).

Research in the last few decades has shown almost conclusively that gravitational instability is at the heart of the origin of structures, and the clumpiness in the distribution of galaxies in the present Universe[1]. This conclusion stems from a detailed comparison between the predictions of theoretical ideas and careful observations. We will not be able to appreciate the present state of cosmology by only knowing the mechanism behind the process of structure formation in the Universe. Cosmology has acquired its current status because of its capability of quantitatively testing its hypotheses. It will be sensible to go deeper into the details of formation of structures and to know how cosmologists have been testing their ideas.

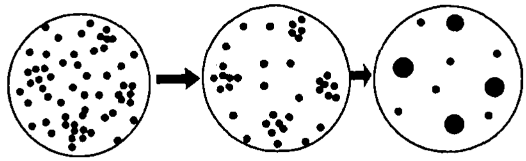


Fig. 5.2 *The three panels in this figure schematically shows the effect of gravitational instability. The first panel (on the left) shows the state of the Universe at a very early phase, with more or less uniform distribution of matter and with very small inhomogeneities. The panels on the right show how this non-uniformity increases with time, and how finally particles come close to form structures in the Universe.*

If we think in detail about the process of structure formation, we will realise that constituents of the Universe other than matter have serious effects on the evolution of structures. We will find that radiation can make the evolution of clumps of matter somewhat complicated. In addition, the presence of dark energy or a cosmological constant can further complicate the situation.

5.2 Matter and Radiation

Although matter and radiation were in close contact with each other in the early Universe, their fundamentally different natures made them evolve differently. Although photons are also particles they have some special characteristics. They are massless, for example. Apart from this, there is a very important difference which can be explained by considering two balloons filled with radiation and matter particles. We know that a balloon full of hot gas and with higher pressure than the surroundings, will slowly expand on its own until the pressure inside equalises with the pressure outside. A hypothetical balloon filled with photons will also expand and thereby lower the pressure inside. There will however be a difference between the *rate* of decrease of pressure, or, equivalently, the energy density inside. It turns out that the energy density in the balloon filled with photons will decrease *faster* than in the balloon filled only with particles[2]. In other words, radiation cools faster than matter when it is allowed to expand on its own.

Let us now go back to our discussion of the early Universe keeping in mind this essential difference between matter and radiation. Since matter cools slower than radiation as it expands, it must also heat slower than radiation when it contracts. We know that the energy density of radiation at present is much smaller than that due to matter. As we now run the film of the Universe backward in time and allow the Universe to contract, we will find that both matter and radiation energy density increase with time, but radiation energy density would increase faster. At some point in the past, therefore, it would take over and become the most dominant form of energy in the Universe.

It therefore means that the early Universe was dominated, not by matter, but by radiation. As the Universe expanded and cooled, the importance of radiation over matter diminished. At present, we only see the relic of this once-important-but-now-worn-out radiation in the form of the microwave background.

It can be shown that matter and radiation were of equal importance, as far as the relative energy density is concerned, when the Universe was around ten thousand years old. The corresponding redshift epoch is approximately ten thousand years. This means that the Universe was about ten thousand times smaller than its present day size at this important epoch. Cosmologists call this epoch the 'matter–radiation equality epoch', or simply, the 'equality epoch'. This epoch is also important from the point of view of structure formation in the Universe as we will find out shortly.

Another essential difference between radiation and matter that is relevant here lies in their speeds. Most of the matter, especially the part that remains strongly coupled to one another, begins to move slowly once the Universe cools down somewhat. Photons however always move with their constant 'master speed'. Matter particles which decouple early from the rest of the crowd move about freely enough and can still retain a very high speed when other matter particles have slowed down. If we however ignore for a moment these decoupled particles and concentrate on the particles which are still strongly interacting with each other, the difference between the speeds of matter and radiation becomes stark.

This difference in speeds had a negative effect on the formation of structures in the Universe. As long as matter and radiation are strongly coupled, as they were before the equality epoch, it would have been difficult for matter particles to cluster since the frenzy of photons tend to break away any such nascent cluster of matter. This tendency of radiation to inhibit clustering of matter will certainly depend on the relative importance of radiation over matter in the Universe. Naturally, radiation had its way before the epoch of matter radiation–equality mentioned earlier.

After this epoch of equality, radiation would find it difficult to inhibit clustering of matter, but it continued to exert considerable influence on the process. This continued until the epoch of recombination, or the decoupling epoch, when neutral atoms formed and the interaction between radiation and matter plummeted. Between these two epochs (after the equality epoch and before the decoupling epoch), clustering of matter proceeded in a hesitant manner. Although radiation did not completely wash out any cluster of matter particles, its outward pressure was felt by clumps of matter that were shrinking under their own gravity. Most of the clumps at this phase went through an oscillation due to the tug of war between gravity and the outward pressure from radiation. This phenomenon was first described in detail by Joseph Silk in the 1960s.

These ideas are not confined to the domain of theoretical cosmology any more. Astronomers have recently discovered the signatures of these oscillations of early clumps of matter. Before we discuss these observations in detail, let us find out about another important factor that can inhibit the growth of clustering of matter. According to cosmologists, the effect of the cosmological constant or dark energy is also such that it tends to inhibit the clustering of matter. We have earlier encountered the repulsive nature of dark energy in the overall accelerated growth of the Universe. It turns out that the same repulsive nature also obstructs the clustering of matter. Cosmologists therefore have to work out the relative importance of radiation, matter and dark energy in their calculation of the evolution of clumping of matter.

5.3 Power Spectrum of Clustering

In our discussion on clustering of matter so far, we have not yet thought about an important aspect of growth of clumpiness of matter. It is not enough to discuss the tendency of clustering. This tendency may also differ depending on the size of clusters.

Let us think of a situation in which there is clustering of some parameter not in space but in time. For example, we can think of the typical heartbeat pattern of a healthy person. The pattern of heartbeat is the result of a complex process involving many factors. Although we may think of the pattern as quite orderly, actual data show that the interval between heartbeats is not constant even under normal circumstances, and there are fluctuations in this interval. What is interesting is that these erratic fluctuations cover a large range of time intervals. There are fluctuations of the order of seconds, as well as deviations that operate for a few minutes, or tens of minutes (Fig. 5.3). Fluctuations of different time scales possibly arise from different physical processes. But even without going into such details, one can describe the features in the pattern by saying that there exist fluctuations in the interval between heartbeats from a fraction of a minute to tens of minutes.

Contrast this situation with the pattern of intervals of a simple pendulum. The pattern in this case is orderly and absolutely periodic. Another way of saying this is that there is a particular frequency associated with the system. In the case of the heartbeat pattern, the erratic pattern consisting of fluctuations of different time scales can be described as consisting of a *number of frequencies*, and not a single frequency. Yet another way of describing the pattern in the case of a simple pendulum is to say that the energy of the system is confined to movements with a single, particular frequency. In the heartbeat pattern, one can then say that the energy is *distributed* in various processes with different frequencies (Fig. 5.4).

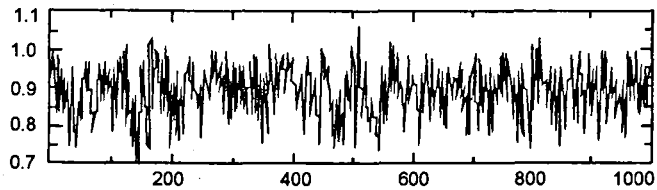


Fig. 5.3 *A typical heartbeat pattern of a healthy person. The vertical axis shows the time (in seconds) between heartbeats and the horizontal axis is labelled with the beat number. Notice that there are variations of all 'sizes', from very rapid variations to changes which occur over a longer duration.*

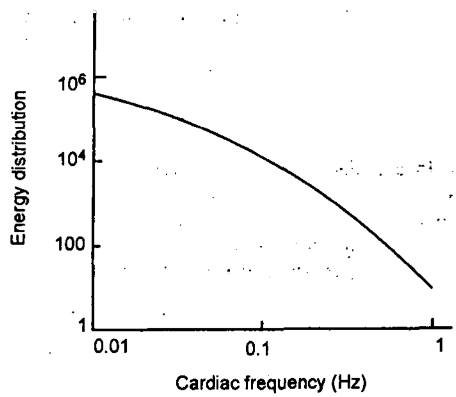


Fig. 5.4 *A sketch of the power spectrum obtained from the heartbeat pattern shown earlier. The horizontal axis covers the range of frequencies in the unit of Hertz (one cycle per second).*

It turns out that this distribution of energy in different frequencies in a system that is marked with erratic movements describes well the statistical nature of the system. The statistical nature of the fluctuations is well defined by this distribution, which is called the 'power spectrum'[3]. It essentially shows how some processes with certain frequencies can dominate other processes. For example, the observed power spectrum of the heartbeat pattern of a healthy person depends on the frequency in a rather simple way— the higher the frequency (that is, the smaller the interval between heartbeats), the smaller the energy associated with it. This roughly means that extremely rapid fluctuations are less frequent compared to variations over long durations.

Consider now fluctuations in space instead of time. This is somewhat more abstract than our example of heartbeat pattern and its fluctuations in time. Patterns can occur in space just as they can appear in time. An ordered periodic time signal is analogous to precise ordering in space. A row of houses with equal distances between them is analogous to the periodicity of a pendulum. Let us now consider matter distributed in space more or less uniformly with some small fluctuations in density here and there. One could then describe the statistical nature of this (density) fluctuation in terms of a power spectrum, although now this would mean distribution over different *spatial* frequencies. This frequency in space has the units of the inverse of unit length, just as ordinary frequency has the unit of the inverse of a second. Since this is a bit difficult to think of, it might be easier to think in terms of length itself. We will therefore discuss space in terms of a power spectrum of fluctuations over different length scales, instead of spatial frequencies.

Take the analogy of lengths of words in a sentence. Suppose a sentence consists of words which are all of the same length, say four letters each. This is highly improbable, but in any case, the power spectrum of distribution of the word lengths for this sentence will consist of a single length scale, of four letters. In a realistic sentence, however, there will be words of a variety of sizes, from single letter words like 'a' to larger words. The power spectrum will then be distributed over different word lengths. (Interestingly, the power spectrum in this case is similar to that of the heartbeat pattern, with smaller power being associated with large words, essentially saying that very long words occur rarely.)

A power spectrum of fluctuations can therefore describe well the statistical qualities of a varying quantity. Cosmologists employ the concept of power spectrum of varying density in the Universe in their study of formation and evolution of structures. (Figs. 5.5 & 5.6) If we recall our discussion in the previous section, we can say that gravitational instability increases the clustering of matter. Leaving aside the extreme cases of clustering in which particles collide with one another (in which case processes other than gravity also become important), gravity increases clustering at all scales. If there are density fluctuations of a given length scale to begin with, the inhomogeneity at this scale will increase with time. Fluctuations in all length scales will therefore increase with time. In other words, we can say that *the power spectrum of density fluctuations increases* with time as a result of gravitational instability.

We can also say that radiation and dark energy tend to inhibit the increase in the power spectrum. As we have learned, radiation exerted most of its influence before the matter–radiation equality epoch. The influence of the cosmological constant on clustering also changes with time. A detailed discussion of these dependences is however beyond the scope of this book.

One could mention another complication to give an idea of the intricacies involved in this calculation. Although we have discussed the existence of dark matter, physicists are still far from discovering its exact nature. According to most cosmologists, dark matter consists of very massive particles, but some think that it may consist of light particles like neutrinos. These particles move with speeds not much less than that of light. It turns out that the clustering of matter and its evolution in time depend crucially on the nature of dark matter. If dark matter consists of particles like neutrinos[4] moving with high speed, it will have an effect on clumping similar to that of radiation i.e., it would inhibit the formation of clumps, especially clumps which are small in size. The power spectrum of density inhomogeneities would therefore be very different in the case of massive dark matter and neutrino-like dark matter.

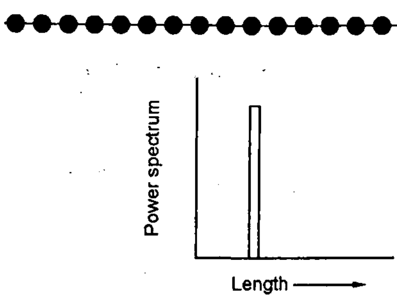


Fig. 5.5 *Clumping of matter is shown in a one-dimensional Universe. If matter clumps to form a periodic array of objects, the power spectrum of fluctuations will have contributions from a single length scale. It is analogous to the power spectrum of a single pendulum with a definite frequency.*

Cosmologists think that these are the main ingredients in the process of formation of structures in the Universe. The most important ingredient, of course, is gravitation although various characteristics of the Universe can also be influential. We have already learned about the effect of radiation and the cosmological constant. In addition, one should also consider the effects of the geometry of spacetime, the value of the Hubble constant and other parameters. For example, if the radiation energy density or dark energy density were increased somehow, the power spectrum of structures in this imaginary Universe would be different from that in our Universe. It would be difficult to cluster at small length scales in this Universe owing to the existence of the negative influence of radiation or dark energy. In the case of larger clumps though, gravity may still win over these inhibiting factors. This imaginary Universe would be dominated by large size inhomogeneities or clumps and may be deficient in small clumps.

One can turn this around and try to predict the characteristics of the

Universe (e.g., the importance of radiation or dark energy in it) from the *present state* of the structures in the Universe. It is analogous to guessing the recipe of a cooked dish from the way it tastes. Just as a veteran chef can tell the ingredients used in a dish from its taste, cosmologists try to use their knowledge of the structures in the Universe to estimate different parameters that define our Universe. There is a lot of activity in this front at present and cosmologists have been largely successful in this endeavour. Let us try to understand the essentials of this study.

5.4 Evolution of Structures in the Universe

Alert readers may be puzzled over one aspect at this point. Although we have concluded that gravitation tends to increase small inhomogeneities into large ones, we have not yet discussed the origin of the *initial* small clumpiness of matter. In the analogy of bus timings, the irregularities were set off at a particular stop by a larger than average number of passengers there. There must have been some reason behind this increase. When and how did the original clumps of matter, however small, arise in the Universe? Without the initial density fluctuations, gravity will not be able to create the structures that we find today.

Cosmologists are still not sure of the origin of these inhomogeneities although there are a few reasonable theories. They think that the inflationary period in the very early Universe had an important role to play in this aspect. It was then that very small, microscopic fluctuations, perhaps arising from quantum mechanical uncertainty, were magnified to large, macroscopic inhomogeneities. These initial inhomogenities are thought to have had a large range of length scales; i.e., there were inhomogeneities ranging from small to large sizes. Cosmologists have estimated the statistical nature of these primordial density fluctuations as if they originated in the inflationary epoch. In other words, they have tried to determine the power spectrum of such initial fluctuations.

According to cosmologists, the power spectrum in this case has a special property. It is thought that the power spectrum should be a simple decreasing function of the length scale of the inhomogeneity. In other words, there is more 'power' for small size inhomogeneities and the power spectrum decreases with increasing length scales. This basically means that the Universe should look more and more uniform if seen from progressively bigger perspectives (Fig. 5.6). A random selection of small regions may look very different from one another, but large regions of the Universe selected at random will look fairly similar to one another. Cosmologists have studied such power spectrum in the last three decades because of their special characteristics[5]. The pioneers in this research were Edward Harrison and James Peebles of USA and Yakov Zel'dovich of the then USSR who began their studies on this kind of power spectrum in the 1970s.

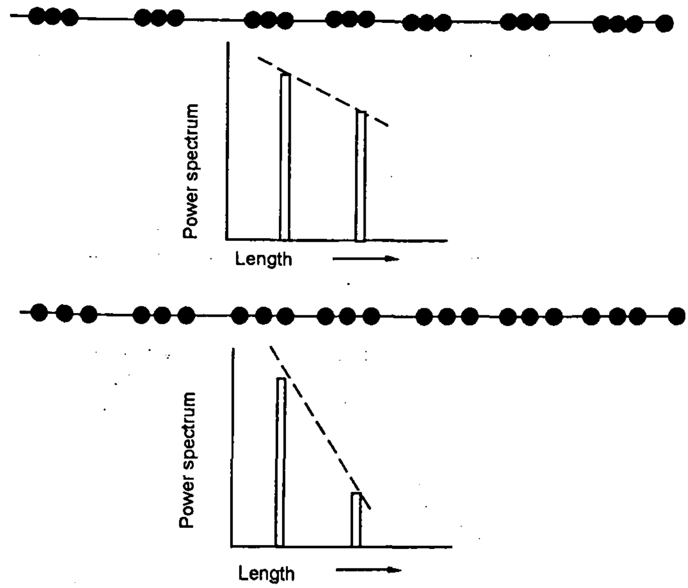


Fig. 5.6 *The power spectrum of density fluctuations of matter is explained for two different cases. Consider the distribution of matter clumped on a single length scale as in Fig. 5.5.*

Consider then the case of a power spectrum with additional contribution from a larger length scale. This will make these objects (black circles) cluster at a large length scale, with a strength depending on the 'power' at this scale. This case is shown by the picture on top. If the power at the larger length scale is decreased (picture at bottom) then the clustering at this scale will be less strong. Regions of the Universe selected at random would then look more similar to one another than the earlier case. In other words, the Universe will look more uniform from a large perspective if the power spectrum has a smaller value at a large length scale.

In reality, the power spectrum has contributions from many length scales giving rise to various hierarchy, or levels, of clustering, instead of just two levels of clustering as in this simplified example.

We have concluded earlier that gravitational instability tends to increase the power spectrum at all length scales. It turns out that the power spectrum not only increases in its value but also changes its shape in time. The main reason for this is the effect of radiation on clustering. Fluctuations on very small length scales were inhibited in the early hot Universe from growing too large, compared to fluctuations on much larger length scales. These small-scale fluctuations therefore grew timidly compared to large-scale fluctuations. The power in small fluctuations therefore did not grow as large as the power in larger fluctuations. This caused a change in the *shape* of the power spectrum.

We however know that the effect of radiation was mostly important before the epoch of equality between matter and radiation. The extent to which radiation could inhibit the clumping of matter was limited because of the limited duration within which it could operate. It could therefore exert its negative influence mostly on very small size inhomogeneities. The age of the Universe at the time of matter and radiation equality was approximately fifty thousand years[6]. How far could a photon travel during this time? It could have travelled at most a distance of around fifty thousand light years, or around fifteen kiloparsec. We must however remember that the corresponding redshift of the Universe was about ten thousand. The Universe has expanded by a factor of around ten thousand since then. A distance of ten kiloparsec then will now correspond to a length scale of around a hundred megaparsec.

What does this mean? Initially the power spectrum of clustering was simple and monotonically decreased with length[5]. Radiation then began to exert its influence on the clustering, but its negative influence was mostly effective only until the epoch of equality of matter and radiation. This means that the effect of radiation was limited to a length scale of ten kiloparsec, which today corresponds to a length scale of a hundred mega parsec. Inhomogeneities smaller than this were inhibited by radiation in the early Universe, compared to inhomogeneities larger than this. In other words, the power spectrum for these smaller length scales has not grown as much as that for large size inhomogeneities.

This has made the power spectrum of clustering look a bit more complicated. It got *bent* at a particular length scale, corresponding to the above estimate. It has been increasing in its value since the equality epoch but it has more or less retained the bent shape it acquired around that epoch. Cosmologists therefore predicted that this would be the shape of the power spectrum of clustering in the present day Universe. It would have a 'bump' at a length scale of approximately hundred megaparsec (Fig. 5.7).

It is by no means an easy task to check this prediction with the underlying power spectrum of clustering in the present Universe. One needs data for thousands of galaxies to be able to perform a statistical analysis with some confidence. It has only become possible very recently with the help of technological advances, as mentioned in the first chapter, to map the positions of many galaxies at a short time. A recent example of such a galaxy survey is the '2dF galaxy survey'.

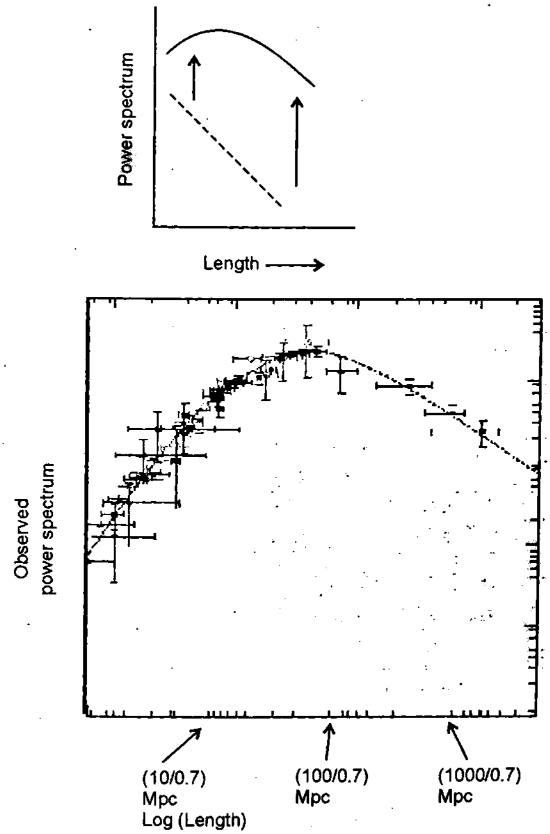


Fig. 5.7 *The evolution of the matter power spectrum is sketched here. The plot on the top shows the theoretical idea of the initial power spectrum as the dashed line. Both x (length) and y (power spectrum) axes are in logarithmic scale. The initial power spectrum is inversely proportional to length. The arrows show how the power spectrum gradually increases and changes its shape. The observed power spectrum, from various observations, is shown at the bottom. Determination of length scales in the Universe depends on the value of the Hubble constant, which has been assumed to be 70 km/s per Mpc, which is why the length scales appear as multiples of (1/0.7). The data points at large length scales are from cosmic microwave background radiation (see Section 5.5), at intermediate length scales from the 2dF galaxy survey, and a few other observations.*

It has been possible to determine the power spectrum of clustering of galaxies from such maps (Fig. 5.8). What is thrilling is that the observed power spectrum matches remarkably well with the theoretical prediction in many aspects. Firstly, the power spectrum decreases with length scale for large size inhomogeneities. A decrease in the power of clustering means that seen in a large enough perspective, the distribution of galaxies would be close to uniform. In other words, the deviations from uniformity decrease as one considers large enough regions of the Universe. This was referred to in the first chapter. Secondly, there is indeed a 'bump' in the power spectrum, at about a hundred megaparsec, approximately the length scale predicted by theoreticians.

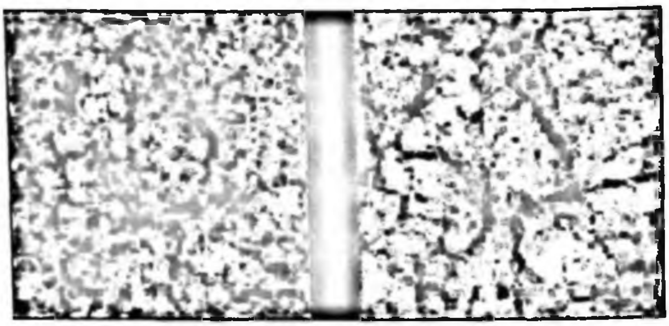


Fig. 5.8 *Cosmologists can determine the characteristics of clustering of matter in different kinds of Universes with the help of simulations in computers. In the picture, a small cubic region of an imaginary Universe is being shown, whose sides measure 500 mega parsec, or 1500 million light years. Cosmologists begin their simulation with a situation in which matter is more or less uniformly distributed, with small inhomogeneities present in it. One can then simulate the evolution of clustering of matter into the present state.*

The picture on the left shows the state of affairs in a Universe in which there is no cosmological constant. The picture on the right shows the situation in a Universe dominated by cosmological constant. Compared to the picture on the left, this Universe has less clustering of matter at small length scales. The typical size of structures in this Universe is much larger than that on the left. We should however remember that clustering of matter can produce structures of different shapes, somewhere more or less spherical but elsewhere more flattened and at some places like thin filaments. The typical size of all such structures on the right is larger than on the left. (Picture courtesy : Virgo Consortium)

This success has been very encouraging. The uncanny match between observations and theoretical expectations suggests that cosmologists now know the main ingredients for the recipe of structure formation in the Universe. What remains to be done is to determine the power spectrum of clustering to a good accuracy and to find out the parameters of the Universe from it. For example, one wishes to know if there is really a dark energy and to what extent it dominates the total energy density of the Universe. Or, the density of matter in the Universe, and so on.

Cosmologists are still busy working these things out. Preliminary results seem to show that the matter density of the Universe is about 30% of the critical density and that the cosmological constant (or dark energy) dominates the energy density of the Universe. It also appears that the main constituent of dark matter cannot be neutrino-like particles, but some yet unknown massive particles.

It is indeed heartwarming to find that these results match well with other observations, like that of the acceleration of the Universe from supernova observations. It is then tempting to think that perhaps cosmologists now understand the process of structure formation well and that the knowledge about the parameters that define our Universe is close at hand. One should however remember that there is a lot of research going on in these topics at present and the results quoted above are only preliminary.

There is another tool that has come to be handy to cosmologists in their pursuit of the knowledge of our Universe. It is the observation of inhomogeneity of the cosmic microwave background radiation.

5.5 Patches in the Background Radiation

Cosmologists asked an important question in the 1960s about the clustering of matter soon after the discovery of the background radiation. Did the clumping of matter in the Universe leave any signature on the microwave background radiation in the Universe? The basis of this suspicion lay in Einstein's theory of relativity, which predicts that photons lose energy while travelling through spacetime that is curved due to gravity. If there is clustering of matter then the spacetime in the vicinity is expected to be curved and the energy of a photon travelling through this region is expected to be affected. The energy of photons is related to their colour and the temperature that one associates with them. It is then natural to expect that the photons of background radiation may bear some signature of the early clustering of matter.

Photons in the early Universe however interacted strongly with matter. Whenever there was any loss or gain of the photon energy, frequent interactions ensured that there was no large-scale non-uniformity in the photon energy distribution. This phase of strong interactions ended though with the decoupling of matter and radiation at the recombination (or decoupling) epoch. Photons started moving freely after this and they are now detected as background radiation by our telescopes. Photons should therefore carry intact the signatures of any changes in energy that occurred just prior to the decoupling era. In other words, photons of microwave background radiation, which interacted last with matter at the decoupling era, must carry information about the clustering of matter at that era, just as a fossil carries information from a bygone era.

We learned earlier that the energy of a photon changes if the geometry of the spacetime it is moving through is non-Euclidean. Now, as mass in the Universe clusters, it deforms the spacetime geometry locally and changes the energy of the photons that move in it. If there is a strong interaction between matter and radiation, photons could compensate for the change of energy at one position by changes acquired elsewhere. At the epoch of decoupling, however, photons stopped interacting much with matter. The distribution of photon energy must have been patchy at this epoch. As the photons began to move freely after this epoch, they must have retained this patchiness in the distribution of energy. One must therefore observe how the energy of the microwave background photons is distributed in different directions of the sky to discover this effect.

We have earlier found that cosmologists characterise the microwave background radiation with the temperature associated with it—the temperature of a blackbody which would emit radiation with the same spectrum. Since changing energies of photon changes its spectrum, one can then say that the events at the decoupling era must change the temperature of the background radiation at different regions in the Universe, or at different directions in the sky. One therefore expects the microwave background radiation to have slight variations in its temperature in different directions of the sky.

Cosmologists in the 1960s estimated that the expected fractional variation of the temperature was approximately one part in hundred thousand (10^{-5})[7]. This is the minute variation of the temperature referred to at the end of the last chapter (see Fig. 4.10). The *COBE* (Cosmic Background Explorer) satellite discovered this signature of deviation in different directions, or anisotropy, in the temperature. This discovery in 1992 was another milestone in the study of structure formation in the Universe. It has confirmed that the basic theoretical ideas of cosmologists for the evolution of clustering in the Universe are correct. This signature is an inevitable effect of clustering of matter. The patchiness in the temperature distribution of the background radiation in our sky is therefore like a fossil from the early Universe, giving us clues about the Universe at the decoupling era.

The study of the anisotropy of the microwave background radiation is an intense area of research now. It is expected that there will be many discoveries in this field in the next decade or so. Let us try to understand the salient questions and how they are being approached.

If we concentrate on the map of temperature of the background radiation in the sky (Fig. 4.10), we will first notice that the temperature is not uniform. We will also notice that like the distribution of sizes of inhomogeneities of matter in the Universe, the sizes of the deviant portions are also not the same all over the sky. The analogy is of course not coincidental. The deviations in radiation temperature naturally arise because of the underlying clustering of matter in the Universe. One therefore expects that the pattern of temperature deviation in the sky will also have contributions from patches of various sizes. By the 'size' of a patch, one here means its angular extent.

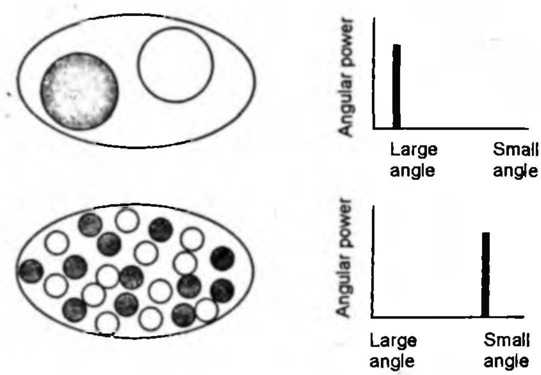


Fig. 5.9 *The basic idea behind the angular power spectrum is schematically explained here. If the temperature distribution of the background radiation is such that patches of different colour (or, equivalently, temperature) are of large angular size, then the angular power spectrum will have contributions from only large angles (upper panel). If instead the patches are all of small sizes then the power spectrum will be dominated by contributions at small angles (bottom panel). The horizontal axis in the plot for the angular power spectrum has the unit of the inverse of angle in the sky. Large angles appear towards the left and small ones toward the right of the horizontal axis.*

Since the tendency of clustering of matter at different length scales is different, it then follows that the dominance of patches of various angular sizes in the pattern of temperature variation will also be different. One could then describe the statistics of the pattern of temperature variation with an 'angular' power spectrum of sizes of spots. This power spectrum will indicate the relative dominance of spots of different angular sizes in the overall pattern. If the pattern is dominated by only variations over small angles (analogous to rapid variation in the heartbeat pattern, say), the power spectrum will be concentrated only at small angles. If in addition there are variations over large angles, then the power spectrum will also extend to large angular sizes (see Fig. 5.9).

Cosmologists therefore describe their observations of temperature variation in the background radiation with the help of an angular power spectrum.

This is then compared with expectations from theoretical considerations. What sort of power spectrum does one expect from clustering of matter?

Consider the time just prior to the decoupling era. We have earlier discussed how clumps of matter were going through oscillations because of the tug of war between gravity and effects of radiation. Photons in the vicinity of these clumps must have gone through changes in energy because of the curved spacetime around them. In addition, there would be a change of energy due to the movement of matter in the oscillating clumps, from the Doppler effect. We have discussed earlier how a moving source of radiation has its frequency, and hence energy, changed due to its motion. The oscillations of the clumps would go through a cycle of maximum and minimum speeds, just like a pendulum has zero speed at the moment of turning and maximum speed when it is vertical. The change in energy due to motion, therefore, would be minimum for clumps which are at the point of turning over (from expansion to contraction). Since clumps of different sizes would be at different phases of oscillation at the decoupling era, the shift in energy from motion would vary according to the size of the clump. Moreover, since clumps of different length scales would appear to have different angular sizes in the sky, the shift in energy or temperature would vary, in an oscillating manner, with the angular size of patches in the background radiation temperature map.

Just as in the case of inhibition of clustering by radiation before the equality epoch, the effect of radiation on the oscillation of clumps will be limited to a certain extent. The oscillations of the clumps is similar to the cycle of compression and rarefaction of air when sound waves propagate through it. This limiting size of clumps will correspond roughly to the distance travelled by acoustic waves from the Big Bang to the decoupling era. Clumps of bigger size will not be affected by radiation and they will keep on contracting slowly due to gravity.

Detailed calculations of these different sources of energy change in radiation indicate that the resulting angular power spectrum will have an oscillating pattern. Since oscillation is limited to clumps smaller than a certain size, depending on the age of the Universe at the era of decoupling and the distance travelled by acoustic waves until then, the angular power spectrum will show an oscillating pattern for angles smaller than a particular value. The power spectrum for patches larger than this angle will have a simple shape. This means that the angular power spectrum will have a bump or a peak at a particular angle and then have a series of such peaks for angles smaller than this. These are called the 'acoustic peaks' in the angular power spectrum.

Cosmologists have been actively searching for such signatures in the temperature map of the background radiation. It is difficult, however, to determine the power spectrum of pattern of temperature variation to accuracy as one needs a large amount of data to perform the statistical analysis and finding the pattern for a small region of the sky does not suffice for this purpose. It is also difficult to find the fine-scale pattern, since one needs large instruments to observe with a high resolution. In addition, one needs a sensitive instrument since the fractional variation of temperature is very small[8].

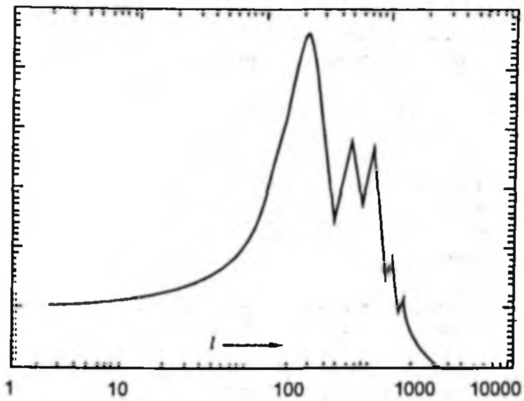


Fig. 5.10 *An example of the expected angular power spectrum of anisotropy in the background radiation. As in Fig. 5.9, large angles correspond to the left of the horizontal axis and small angles are represented toward the right. Here 'l' is a measure of the inverse of angular separation in the sky. A value of $l = 180$ corresponds to an angle of about 1 degree. It is seen in the figure that the power spectrum for very large angles (in the left) has a simple shape whereas it has a wavy pattern for smaller angles. The observations of COBE probed the coarse-grained pattern in the sky and found the power spectrum of variation over large angles, and confirmed the simple shape of the power spectrum for these angles. Recently the first few 'peaks' have also been confirmed.*

The shape of the angular power spectrum depends on various characteristics of the Universe. For example, the density of (dark and normal) matter, the density of dark energy or the cosmological constant, the value of the Hubble constant and so on. It is hoped that a detailed comparison between this theoretical plot with the observed power spectrum will enable one to determine these important parameters of our Universe.

Technological advances have now enabled cosmologists to determine the angular power spectrum to a reasonable accuracy and extend it to small enough angles to be able to compare the results with theoretical predictions (Fig. 5.10). They determined the first 'Doppler peak' in the angular power spectrum with observations from a few balloon-borne instruments in 2000. A couple of more 'peaks' have now been well determined from observations by a satellite named *WMAP* (Wilkinson Microwave Anisotropy Probe) that was launched in 2001[9]. Already a number of exciting new results have come out of the analysis of these data and it is now a very active area of research.

Incidentally, the observation of the first Doppler peak has important clues about the geometry of the spacetime of the Universe. We have seen that the angle of the first peak corresponds to the maximum size of clumps that oscillated just prior to the decoupling epoch, and we roughly estimated this limiting size from the age of the Universe at this epoch. The corresponding angle for these clumps in our sky will, however, depend on the geometry of spacetime in the Universe. The relation between angle and length naturally depend on the rules of geometry. If there is a line segment and one extends two imaginary lines from its ends to the observer, the angle subtended by them will depend on the geometry of the plane on which these lines are being drawn. A line segment on the equator of the Earth, for example, will subtend a larger angle at the pole than the angle subtended at the same distance by a line of equal length on a plane surface (Fig. 5.11).

Turning it around, one can say that if one knows the length and the angle subtended at a point, one can deduce the geometry involved. In the case of the first Doppler peak of the background radiation anisotropy, the length scale of the relevant clumps is known from theoretical considerations. If one can determine the corresponding angle subtended in our sky, by finding the angle corresponding to the first peak, one can easily deduce the geometry of the Universe. It turns out from observations that the spacetime geometry is Euclidean to a reasonable accuracy.

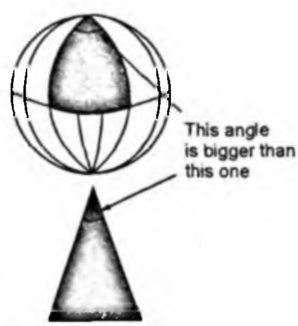


Fig. 5.11 *A line segment of a certain length at the equator on Earth subtends a larger angle at the pole than it does at an equidistant point in space drawn on a plane (shown at the bottom).*

A large number of scientists are now involved in the detailed study of the temperature variation of the background radiation. Besides the space-borne instruments, there are many ground-based telescopes that are being used for this. Scientists have even braved the extreme weather in Antarctica to observe the microwave sky from there, as the climate there is very dry and the lack of absorption due to water in the atmosphere makes it a good site of observation of microwave radiation.

There is a plan to launch another satellite to study the temperature variation pattern with even greater angular resolution than that by *WMAP*. European countries have planned to send a satellite called *PLANCK* (named in honour of Max Planck) around 2007[10]. There is already a lot of excitement in anticipation of interesting results from it. The analysis and interpretation of these future data will not be very easy. The slight temperature variation that is being sought is extremely small, and may be swamped by radiation at these frequencies from various sources in the sky. One has to also take into account the fact that the gas in our Milky Way emits radiation in these frequencies. Since they are randomly distributed in the sky, they also contribute to variation of temperature in different parts of the sky. One must then subtract this 'foreground noise' to be able to see the cosmological signal in the background. It is like straining to hear an injured soldier groaning in the midst of ear-splitting artillery noise in a battle field. One must then understand well the nature of radiation from objects in our Galaxy and in the nearby Universe, so that one can perform this subtraction well.

We have so far learned how the detailed study of galaxy distribution and background radiation anisotropy have encouraged cosmologists by the match between theory and observations. Incidentally, they have employed yet another tool in this research and the preliminary results are also encouraging on this front.

5.6 Gravitational Lens

If we recall our discussion on the general theory of relativity, a photon not only suffers a change of energy in a curved spacetime but also changes its path. We have seen how gravity acts like a lens and how the path of a photon passing close by the Sun can get deflected from a straight line path.

Just like the distorted images one sees in a hall of mirrors, this deviation from a straight line path can distort the apparent shape of the source, depending on the characteristics of the lens and its orientation with respect to the source and the observer. The shape of a background galaxy whose light passes through a region of curved spacetime, perhaps due to some agglomeration of mass in the foreground, will then appear distorted. According to general relativity, the amount of distortion depends on the mass density of the foreground object as projected in the sky. Most often the mass of the object in the foreground is dominated by dark matter, which does not shine and swamp the brightness of the background galaxy. One can then clearly see the distorted shape of the background galaxy.

If the density of the foreground mass is not very large then the distortion would be very small. Also the extent of distant galaxies which can act like background source are also small. Searching for such examples of gravitational lens therefore requires observing the sky at a very fine resolution. These examples first appeared in radio wavelengths where it is easier to have high angular resolution studies, than in optical wavelengths, owing to the turbulent atmosphere on Earth. It has now also been possible to find optical examples with the help of space-borne instruments like the Hubble Space Telescope (Fig. 5.12).

Consider now the case of clustering of matter in the Universe. This process naturally distorts the spacetime geometry depending on the amount and size of clustering. This distortion is expected to be very very small, much smaller than the case of spacetime around a fully formed galaxy or cluster of galaxies. Nevertheless one does expect that the shape of very distant galaxies would be distorted not only due to massive objects in the foreground, but also due to slight inhomogeneities of matter distributed in the foreground. The distortion in this case would certainly be very minute, but astronomers have been working to find such signatures in the sky.

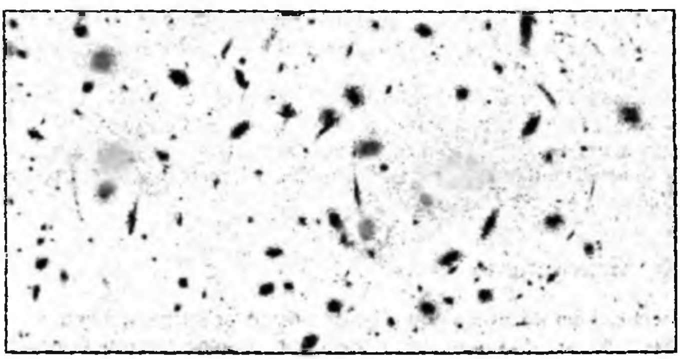


Fig. 5.12 *This is a picture of the galaxy cluster Abell 2218 taken by the Hubble Space Telescope. Notice that many galaxies are stretched to thin arcs. These galaxies are background galaxies whose light crosses the galaxy cluster in the foreground and gets distorted by lensing. (Picture courtesy : STScI and NASA)*

The first impetus to this experiment came from the study by Nicholas Kaiser of Canada in the 1990s. He developed a method with which one could analyse the data of background galaxies to deduce the clustering property of the matter in front. He not only did the theoretical calculations, but also worked out the details of how such an observation should be done, and was involved in these observations. It is indeed due to such combinations of theoretical insight and detailed observations that modern cosmology has attained its present stature. This is what makes the nature of cosmological research at present so different from what it was in the initial stages.

Various groups of astronomers are now involved in this study of 'weak lensing', so-called because of the degree-of distortion. In 2000 they found, from a preliminary study of their observations that it was indeed possible to deduce the clustering property of the Universe, although uncertainties still exist to a large extent. The power spectrum of clustering deduced from gravitational lensing also point toward a Universe dominated by a cosmological constant or by dark energy, and with a third of the total energy density being contributed by matter.

It is indeed thrilling to arrive at the same result from three independent routes. The apparent match between the deduced parameters of the Universe from different approaches, like galaxy surveys, anisotropy of background radiation and weak gravitational lensing, has encouraged cosmologists enormously.

5.7 Hierarchy of Structures

Let us recapitulate what we have learned about the formation and evolution of structures in the Universe. Very small primordial density fluctuations were most probably magnified into macroscopic inhomogeneities during the inflationary epoch in the early Universe. These inhomogeneities in the distribution of matter have grown since then due to gravitational instability. We have learned how radiation and dark energy affect the clustering of matter, and we have found out how cosmologists are trying to test these theoretical ideas with different observations.

There is an important aspect of the evolution of this clustering. With the power spectrum of clustering that is inferred from observations, it turns out that the strength of clustering at small scales is larger than at large length scales. We have found earlier that this property gave the Universe a rather uniform look when seen from a large perspective. As clustering proceeds, at some point, matter within a cluster becomes bound by its own gravity and decouples from the rest of the Universe. In other words, the cluster then forms objects familiar to us like galaxies or clusters of galaxies. The form of the power spectrum then means that the earliest objects that cross the threshold to form bound objects are very small clumps of matter since clustering at small length scales is most vigorous.

The next clumps to cross the threshold to form bound entities would be slightly larger. Put in a different way, the early small objects merge with one another and form bigger objects or galaxies. Small galaxies merge to form large galaxies, large galaxies then come together to form clusters of galaxies. This is how structures at different levels must have appeared in the Universe.

This does not however mean that all galaxies belong to some galaxy clusters, or that there is no matter left in the Universe today outside galaxies and other bound objects. The above scenario is correct only in a statistical sense. At one phase, the clustering in the Universe was dominated by forming small galaxies. Again it does not mean that no large galaxies appeared then. Just as exceptional students appear in a class of average students, rare and large density deviations at places could have created large galaxies. The next phase would have been dominated by the formation of larger galaxies. Cosmologists think that at present the clustering is dominated by the formation of supercluster-sized objects in the Universe (Fig. 5.13).

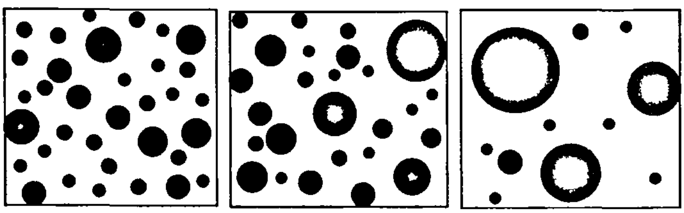


Fig. 5.13 *The evolution of clustering in the Universe is schematically shown here. Starting with the state of affairs in the past at the left, the panels on the right show the progress of clustering in time. Small objects were abundant in the early phase of clustering and as time went on, larger objects have appeared in the Universe, out of mergers of the parent small objects.*

The basic ingredient of this scenario is the continual merging of objects into bigger and more massive objects. Astronomers have by now seen numerous examples of strong interactions and the merging of galaxies (Fig. 5.14). It is a very interesting area of research in astronomy. It is not only that they have caught galaxies red-handed while merging, in some cases, they have also found that such interactions give rise to a bout of star formation in the galaxy. Observations in different wavelengths from radio to sub-millimetre to infrared to optical, are being brought together to study these merging galaxies.

Moreover, deep observations with telescopes like the Hubble Space Telescope have shown that the average size of galaxies in the past was smaller than those present now and these early galaxies have been found in a state of interaction with one another[11]. This roughly matches the expectations of cosmologists, although a detailed comparison between theory and observations is difficult. What one records in these observations is light emanating from stars in a galaxy, whereas the clumps studied by theorists are dominated by dark matter. We still do not know exactly how normal matter behaves when clumps dominated by dark matter merge and interact.

Fig. 5.14 *An example of interacting galaxies NGC 2207 and IC 2163 (Picture courtesy: STScI and NASA).*

We must then try to understand how stars form in the clumps of matter created by the clustering in the Universe. This will lead us to the study of the first stars and first galaxies in the Universe. As dark matter slowly clustered in the Universe, some clumps lit up and began to shine because of star formation in them. It is like lighting up apartments after one has created the basic structure with concrete. This is the topic of our discussion in the next chapter.

6

First Luminous Objects

... And yet he is a power of light
And could in one burst overwhelm
And dayify the darkest realm ...

Robert Frost, *Two Leading Lights*

We know that most of the matter in our Universe is invisible; the normal component, which produces stars that shine, for example, contains only a small fraction of the total mass density (around 10–15% per cent). As the clustering of matter in the Universe progressed, the main participant in this process was dark matter. This type of matter does not interact much with normal matter (or even with itself) other than through gravity, since otherwise radiation would have been produced in such interactions. In the clusters of matter that gathered slowly in the darkness of the Universe, there was a small fraction of normal matter which collected around dark matter because of gravity. This normal matter then at some point created stars and began to shine. This marked the epoch of the first luminous objects in the Universe.

We would now like to ask what causes the normal matter component to form stars, and when it begins to do so in earnest. We would also like to find out when (and where) the requirements for star formation were met for the first time in the Universe. We will try to understand these conditions in this chapter.

6.1 Conditions for Star Formation

Consider a clump of matter which has become so dense as to become bound by its own gravity. Most of the matter is naturally dark, but let us concentrate on the small fraction of normal matter in it. In a state of equilibrium, the normal matter, which we will simply refer to as 'gas' hereafter, is hot and has a pressure because of the gravitational force acting on it. It is just like the temperature and pressure inside a star. Gravity keeps the gas particles on their toes, making them run helter-skelter. This random motion of gas particles manifests itself as gas pressure which keeps the gas in a state of equilibrium against the burden of gravity.

The gas however slowly radiates away energy. A part of the thermal energy content of the gas is slowly lost through radiation. Cosmologists now believe that the necessary condition for star formation in a clump of matter crucially depends on the *rate* of this energy loss through radiation.

Consider the situation when the loss of energy is very slow. In this case, pressure inside the gaseous matter will decrease slowly and the gaseous cloud will contract slowly. Since the rate of energy loss is small, the gas cloud will remain hot and at high pressure while contracting. This is hardly fertile ground for star formation. It may sound paradoxical but one needs cold clouds to be able to form stars, like the giant molecular clouds in our galaxy that we encountered in the first chapters. Dense and cool material in these clouds can collapse under some perturbation, and the resulting heat in the very dense core can ignite thermonuclear reactions necessary for stellar birth. It is difficult for hot and tenuous gas to collapse to form stars. Such proto-galaxies will not form stars.

What is required to form stars is that the gaseous cloud should cool fast, which would make it fragment into smaller portions, which would be dense and cool and which could form stars in them. The requirement for fragmentation is that the gas pressure, which keeps its shape and size intact, should plummet. And for the gas pressure to decrease rapidly, one needs to arrange for the internal energy of the gas to be drained away rapidly. In other words, if the loss of energy through radiation is fast enough, the gas cloud will fragment into denser and cooler portions which will have the potential to turn into molecular clouds and ultimately form stars in them (Fig. 6.1).

We then find that the condition for clumps of matter to form stars is that gas should radiate away its energy fast enough. To be precise, the time scale for energy loss should be smaller than that for collapse under gravity. This concept was first discussed in detail in 1977 by a few astronomers—Joseph Silk, Martin Rees and Jeremiah Ostriker, based on earlier suggestions by Fred Hoyle. They discovered an interesting result from their calculations.

It turns out that gas radiates away energy by different mechanisms at various temperatures. At very high temperatures it is due to electrostatic interactions between protons and electrons, whereas at lower temperatures it is due to electrons jumping from a higher energy level inside atoms to a lower level (see Section 6.2 for details). The calculation takes into account the rate of energy loss due to various mechanisms which are at work inside objects of different masses and temperatures.

They found that if a clump of matter is more massive than a few thousand billion solar masses (10^{12} $M_\odot$), it cannot cool fast enough to form stars in it. The gas in these clumps cannot radiate away energy fast enough to induce fragmentation of the gaseous cloud. Clumps of matter with mass lower than this have the potential to become luminous with stars[1].

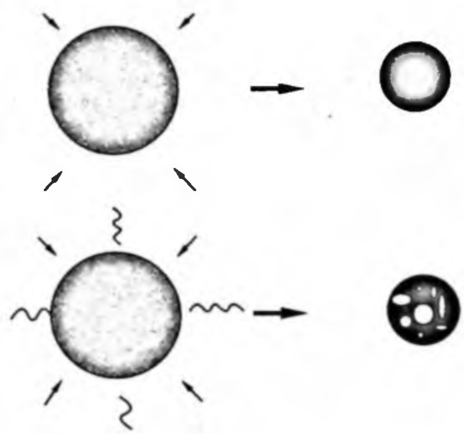


Fig. 6.1 *When gas in a clump of matter radiates energy slowly (top), the gas remains hot while contracting under gravity and does not form stars. When the gas radiates away its energy fast (bottom), it fragments while contracting and forms stars. The radiation of energy is shown here as waves emanating from the clump.*

Careful readers may remember our discussion on galactic masses in the first chapter. We found that our Milky Way had a mass much smaller than this limit of 10^{12} $M_{\odot}$. As a matter of fact most galaxies have masses which are much smaller than this. It is indeed interesting to find concurrence between the theoretical idea and the actual masses of galaxies in the Universe. The success of such a simple analysis has been very encouraging in the study of formation of galaxies.

It is reasonable to assume that there are physical reasons for the typical sizes and masses of everything in Nature. Why do stars typically have the mass of a few solar masses? If the mass were very small then the temperature at its core would not be high enough to ignite the thermonuclear reactions necessary for it to shine. If the mass were very high, then pressure due to its radiation would rip the star apart.

We now understand the basic reason behind the typical masses of galaxies. For a clump of matter much bigger than the limiting mass, the gas in it does not cool fast enough to form stars. Of course, not all galaxies have masses close to this limiting mass. There are many galaxies with masses much smaller than this. These galaxies, with masses around a billion solar masses (10^9 $M_\odot$) are called dwarf galaxies, for their small size and mass.

Cosmologists therefore think of the process of formation of galaxies in the following way—when clustering of matter in the Universe created clumps of different sizes, they contracted under their own gravity. Both kinds of mass, the dark and the normal, participated in this contraction. When the total mass was less than 10^{12} solar masses then the rate of energy loss through radiation from normal matter was very high, which created an environment suitable for star formation. The gas in this case fragmented into many dense and cool parts in which stars could form. This is how galaxies were lit up.

We have seen that in the initial phase of clustering in the Universe it was mostly the small sized clumps that were created in abundance. This was because of the power spectrum of clustering in the Universe. These small clumps then merged with one another to form bigger clumps. Clearly, it was the dwarf galaxies that appeared first in the Universe which then merged with one another to form increasingly bigger galaxies like our Milky Way or Andromeda.

A question naturally arises here. *When* did the first luminous galaxies appear in the Universe?

The background radiation in the Universe has long since moved out of the visible range of electromagnetic waves. It was soon after the epoch of formation of neutral atoms (recombination epoch), when the Universe was approximately a few hundred thousand years old, that the background radiation first moved out of the visible spectrum into the infrared region. Later its wavelength became longer than this and gradually came to be in the microwave region. The Universe therefore was literally dark for a long time. Clustering of matter progressed in silence in this darkness. The question is : when did the Universe come out of the dark ages?

6.2 The First Galaxies

This question can be asked in a slightly different way. Since we know that in the hierarchy of structures, small structures formed first and then the bigger structures appeared, we can ask if there is a minimum mass of a clump that can harbour star formation in it? And typically when did objects of *this* mass come about in the history of the Universe? A somewhat detailed analysis of the rate of energy loss of gas is required to answer this question.

We know that the pressure and temperature arising out of gravitational force inside an object depend on its mass and size. This is true in the case of stars as well as galaxy-sized clumps of mass. In general, the temperature of gas inside a clump of matter in the Universe increases with its mass. To give an idea of the relevant temperatures one could cite the example of the proto-galactic matter that gave birth to our Milky Way. The gas in this proto-galaxy was initially at a temperature of about a million kelvin.

The gas has cooled since then and in the process created generations of stars. If a clump of matter is as small as ten million solar masses then the temperature is a mere few thousand kelvin[2].

It turns out that the temperature of the gas is crucial in determining the *rate* of its energy loss through radiation. Gas at different temperatures radiates energy through different mechanisms. When the temperature exceeds ten million degrees then free electrons are responsible for most of the radiation. At these temperatures the gas is mostly ionised, with free electrons and the residual (positively charged) part of atoms, called ions, zooming around in space. When these electrons happen to come too close to any ion they decelerate because of electrostatic attraction, and lose energy through radiation[3].

When the temperature is somewhat less than this, say around ten thousand kelvin, most electrons are bound inside atoms. They then radiate energy obeying the laws of transition from one energy level to another. Atoms jiggle around in space with a smaller speed than in the earlier case of high temperature but they still manage to nudge electrons into high energy levels once in a while. The energetic electrons then jump to lower levels, emitting photons. The net result is that a fraction of the original energy in the jiggling of atoms is radiated away from the gas, and the atoms cool down in the process.

If the temperature is even lower than ten thousand degrees or so, then the speed of the atoms is not large enough to push electrons up to higher energy levels. Transitions between atomic energy levels therefore cannot help the gas get rid of its thermal energy. It turns out that energy levels in molecules have smaller steps in between and electrons can be nudged to higher levels if they belong to molecules since only a small amount of energy is needed to do so.

Since we are considering the conditions for the first stars, we cannot use any molecules other than hydrogen molecules in our discussion, as other heavy elements are created inside stars and they were synthesised much after the first stars appeared. There are two hydrogen atoms in a hydrogen molecule. Two electrons from these two atoms orbit both atoms together. According to quantum mechanics, the energy levels for the electrons in this case have finer steps than in the case of electrons inside individual atoms. These energy levels help the gas radiate away its energy. It is therefore essential to have some amount of hydrogen molecules in the gas to radiate energy fast if the temperature is too low. Since our aim is to find the minimum mass of an object that can form stars, we must try to understand the conditions for the formation of hydrogen molecules which can help form stars in small objects, with low temperature gas in them.

It is however extremely difficult to create hydrogen molecules in the first place. It is not enough to bring two hydrogen atoms close; there must be some sort of a deep union between the two. The laws of quantum mechanics make it very difficult for this to happen for hydrogen atoms, with a probability of one in a hundred thousand. The hydrogen molecules in our Milky Way are created in a different manner. There are catalysts which can increase the chances of the atoms forming a molecule. Dust grains in space act as these catalysts.

It may be suprising to many to know that there is a large amount of dust in space, in the region between stars[4]. These dust grains are in most cases like the dust on Earth, complicated compounds of silicon and carbon. These grains are created in the outer layers of cool stars (like red giants) and later dispersed in space. When a hydrogen atom collides with such a grain in space it sticks on to its surface for a while. If during this time, another atom collides with the grain and it too attaches itself on the surface, then there is a greater chance of forming a molecule between these two atoms, than there would be if there were no dust grains to act as mediators. The dust grain here increases the probability by providing a meeting place and increasing the time of interaction between the two atoms.

This is how molecular hydrogen is created in present day galaxies, where there is an abundance of heavy elements that have accumulated over time, and where there are favourable sites for the formation of dust grains. We are however concerned here with a phase in the history of the Universe when there were no stars, let alone any heavy elements which are produced by stars, and therefore no dust grains. The common process of forming molecular hydrogen on dust grain surfaces could not have taken place then.

One must therefore find out if there is any other way in which hydrogen molecules could have formed in the early Universe. It turns out that there is one such process. Although most of the electrons in the Universe combined with protons and other ions, at the recombination or decoupling epoch, to form neutral atoms, there were a few free electrons left over from this era. It is believed that these free electrons helped forge a reasonable amount of molecular hydrogen in the early Universe. First a hydrogen atom meets a free electron and forms an odd atom with two electrons in it. This extraordinary atom then combines with a normal atom forming a hydrogen molecule, setting one of the electrons free in the process (see Fig. 6.2).

The rates of all these intermediate step reactions have been measured in the laboratory. One then uses the knowledge of clustering of matter in the Universe along with these rates to determine the amount of molecular hydrogen that could have formed inside clumps of matter. This detailed calculation was done in 1997 in a collaboration of many cosmologists, building on earlier basic work on molecular hydrogen[5].

One can summarise their results in the following way. It is possible to have a tiny amount of molecular hydrogen inside objects in the early Universe from the above process. The background radiation inhibits this process though to some extent, especially at the intermediate step when the odd atom carries an extra electron in it. This electron is somewhat loosely bound and energetic photons of the background radiation can break this bond. The energy of the background radiation photons however decreases in time. At a certain point of time it loses the ability to destroy the odd atom with two electrons. From this point of time molecular hydrogen can form in earnest. The accumulation of molecular hydrogen increases the rate of energy loss through radiation in gas, as explained earlier. Some clumps of matter can become potential sites for star formation in this process. According to the calculations, when the redshift of the Universe was approximately 30, i.e., when the age of the Universe was approximately 100 million years, a few clumps of matter could have formed stars in this way. The masses of these proto-galaxies were in the range of 10^5–10^6 solar masses, that is, about a million times smaller than our Milky Way and even smaller than typical dwarf galaxies.

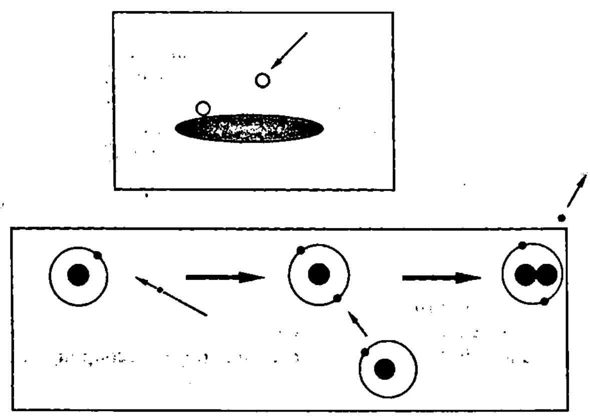


Fig. 6.2 *Dust grains help to form hydrogen molecules in modern galaxies like the Milky Way (upper panel). The elongated grey object represents a dust grain. Suppose an atom collides with it and sticks to it for a while. If another atom comes near it when it is still attached to the grain, there is a chance that the atoms form a molecule. In the early Universe though, there were no dust grains, which can only form out of heavy elements and which did not exist initially. Hydrogen molecules can still form through a lengthy process (see text) (bottom panel).*

These tiny galaxies were the homes of the first stars in the Universe. The size of the Universe then was approximately 30 times smaller than its present size. Clustering of matter in the Universe had created gravitationally bound objects of around a million solar masses. The temperature of the gas in these objects was a few thousand kelvin. A small amount of hydrogen molecules had slowly accumulated in this gas over time. The amount was small, with only a fraction of one part in a million by mass, but enough for the gas to radiate away its energy fast. The gas then fragmented to form smaller clouds in which the first stars in the Universe formed, marking the end of the dark ages of the Universe.

6.3 Flash of First Light

We have so far discussed the epoch of the first stars in the Universe. One might ask about the effect of the first starlight on matter in the Universe. How did the first stars affect the evolution of matter in the Universe? There is still an active debate on this topic because of the number of ways the first starlight could have influenced the surrounding material.

There has been a surprising result in this study. It has been found that the first starlight could have been fatal to the molecular hydrogen in the vicinity of the first stars or galaxies. Molecular hydrogen is vulnerable to ultraviolet photons which can destroy the bond between two atoms in it. Ultraviolet rays from the first stars could have therefore destroyed the molecular hydrogen in the nearby objects or would-be-galaxies. These surrounding proto-galaxies would then be incapable of cooling fast enough to form stars.

If this calculation is correct then the first starlight could have had a negative effect on the matter in the vicinity. A few precocious stars would first light up at some place in the Universe but their radiation would destroy the chance of stars appearing in the neighbourhood. The very first stars would therefore disable other clumps of matter in their neighbourhood from creating stars in them. The first stars naturally lived for a finite duration (depending on their mass). The Universe would then soon plunge into darkness again after their 'death'.

If this unfortunate scenario really occurred then the Universe would have been covered in darkness again. This second period of darkness would have continued for a while.

We have seen how gas with temperatures in excess of ten thousand degrees radiates photons without the help of molecular hydrogen. This is because at this temperature electrons inside atoms can help the gas get rid of its thermal energy. Gas with this temperature can however occur in objects with masses much higher than we have considered here, and needs objects with about a hundred million solar masses. This corresponds to the masses of dwarf galaxies. We also know that clustering in the Universe creates, on average, progressively larger objects. Thus, when the Universe became dark again due to the first stars in tiny galaxies, the darkness would have continued until the first dwarf galaxies appeared, which could form stars without the help of the precious molecular hydrogen. These dwarf galaxies then lit up the Universe again.

As already mentioned this phase in the history of the Universe is under a lot of debate among cosmologists. There still remains a lot of uncertainty. On the one hand, the exact spectrum of light from the first stars remains to be properly determined. We must remember that these stars were devoid of any heavy elements other than helium, and therefore should have a very different structure as well as emerging radiation compared to the typical stars we find today. It should also be remembered that the amount of ultraviolet rays in the total radiation from a star depends on the mass and the mix of heavy elements in it. The typical mass of the first stars is also uncertain. Some believe that the fragmentation of gas clouds leads only to massive stars, much more than the typical stars at present. We have seen how massive stars are hot, and therefore blue, implying a prominence of ultraviolet radiation from them. All these factors are crucial in determining the exact effect of the first stars on the matter in the Universe.

Another uncertainty in this aspect lies in the possibility of the appearance of black holes in the Universe and their effects. Let us first find out how black holes in the present day Universe influence the radiation emerging from a galaxy.

6.4 Black Hole and Active Galaxies

We have so far discussed only normal galaxies like our Milky Way. The radiation emanating from them arises mainly in stars. There are however many galaxies in the Universe in which the major contribution to the total brightness does not come from stars but something else. These galaxies are also extraordinarily luminous. They are typically as bright as a few hundred to a few thousand Milky Ways put together. In addition, one often finds the evidence of matter being thrown out of these objects, in the form of jets, at speeds comparable to the speed of light. The material in the jet also radiates energy after interacting with matter in the vicinity.

These bizarre objects are called active galaxies. They were first discovered in the 1960s and since then have been a major topic of research in astrophysics. The detailed study of the radiation from them—their spectrum and variation in time—has convinced astronomers that the main sources of their activity are black holes at their centres.

The black holes responsible for the fireworks in active galaxies are very massive, typically of around a billion solar masses. When gas in the vicinity of these black holes gets sucked into them, it gets very hot because of the intense gravitational field it is subjected to. Finally jets of material are flung out in space in two opposite directions with tremendous energy. This ejected material can travel a large distance before it is decelerated by interaction with matter. At times this distance can be many times larger than the size of the galaxy in which the black hole resides. The energy of the jet finally manifests itself in the form of a copious emission of photons when it hits the ambient medium and spreads out in a 'lobe' (see the example in Fig. 6.3).

Although black holes cannot be detected through direct observations, there are ample reasons to believe in their existence. Firstly, no one has come up with an alternative explanation for all observations of active galaxies without using black holes. Secondly, there now exist detailed observations for a few nearby galaxies that point toward the existence of black holes. These observations involve the movement of gas and stars around a central object that is so massive and compact that it could not be anything other than a black hole. As a matter of fact, many astronomers now believe that not only active galaxies, but even normal galaxies carry black holes in their central regions. The masses of these black holes seem to vary according to the masses of the 'host' galaxy. The evidence for the existence of a black hole in the centre of the Milky Way (toward the direction of the Sagittarius constellation in our sky) is now very compelling. It is evident from the way stars and gas move around this object. Just as one can estimate the mass of the Sun from the motion of the planets around it, the motion of stars very close to the centre of the Milky Way are being used to determine the mass of the central object. The mass of this black hole is not very large though, around a million solar masses. Also it seems to be in a dormant state, not accreting enough gas around it. The fireworks associated with it are therefore less stunning than in active galaxies.

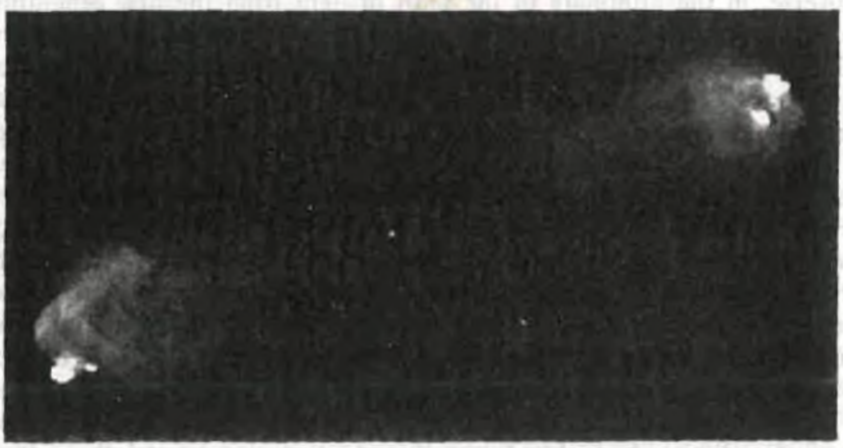


Fig. 6.3 *An example of an active galaxy. This picture is not a photograph of visible photons, but photons of radio wavelengths. This object is seen in the Cygnus constellation in our sky. Notice the two jets coming out of two opposite sides of the central active galaxy. The length of these jets is approximately 160,000 light years. It is believed that a black hole lurking in the active galaxy in the centre is responsible for the powerful jets. (Picture courtesy: Dr. Richard Perley)*

Recently a satellite was launched with a very sensitive X-ray telescope on board. It has been named *Chandra* in honour of the distinguished astrophysicist Subramanyam Chandrasekhar. Observations with the help of

Chandra have now revealed the existence of black holes in a few nearby galaxies, including Andromeda, by detecting the X-ray photons emitted by gas prior to plunging into the black hole[6].

It is still uncertain though how these black holes formed in the first place and how they grew in mass so much as to turn the host galaxy into an active galaxy. This is an interesting topic of research at present. There have been a few interesting results lately which indicate that the formation of black holes can be crucial to the formation and evolution of galaxies in general. According to these observations, the mass of a central black hole is proportional to the mass of the whole galaxy. It is not only proportional to the mass, but also related to the average speed of stars in the whole galaxy[7]. This observed relation has not yet been explained convincingly. It is not clear if the black hole forms first and its mass somehow constrains the total mass of the galaxy in which it resides, or is the other way round and somehow the mass of the galaxy limits the growth of the central black hole and therefore its mass. Perhaps research in the near future will shed more light on this.

Let us recall our discussion on the fate of molecular hydrogen in the early Universe, in the aftermath of the flash of light from the first generation stars. It turns out that the spectrum of radiation from active galaxies is very different from that of normal galaxies, in that there is a large proportion of high energy photons, e.g., X-rays and gamma rays, in them compared to that of normal galaxies. If it so happens that black holes form early enough to seed activity in galaxies around the time when the first stars appear, then the effect of these high energy photons on molecular hydrogen in the vicinity could be complicated. We may be tempted to think that these high energy photons would add to the destruction of molecular hydrogen by ultraviolet photons. It however turns out that these very high energy photons may free a number of electrons, by knocking them off their parent atoms. These free electrons may then boost the formation of molecular hydrogen by helping in the intermediate level of the process already mentioned.

The final outcome of this complicated process will depend on the spectrum of radiation and abundance of active galaxies around this epoch. Detailed calculations are still being done for these processes and the jury is still out on this question.

6.5 Intergalactic Matter

Another interesting aspect of the emergence of the first stars is their effect on matter that has not yet clustered to form any object. As matter in the Universe proceeds with clustering at various levels (with different length scales), there is a fraction of matter that does not get to form any gravitationally bound object. This fraction of residual matter decreases with time though, as the process of clustering slowly engulfs all matter. But this fraction is far from being zero at the present epoch. There is a significant amount of matter that is still at large and has so far eluded being a part of any bound object.

This 'intergalactic' gas is naturally very tenuous, since it is not strongly clustered anywhere. As a matter of fact, the present day density of this gas is ridiculously small. It is similar to the density that one would get if one distributed the content of a tiny snowflake in the volume occupied by the Earth[8].

One would have thought that this tenuous gas would be beyond the possibility of detection. Cosmologists, however, have become ambitious enough to observe this gas. In fact they have been using these observations to deduce the effect of clustering of matter on the rest of the Universe.

We have earlier seen how normal matter in the Universe cools down to form neutral atoms, by the combination of positive nuclei— of mostly hydrogen and helium—with free electrons. This happened at the recombination epoch, when the Universe was approximately three hundred thousand years old. We have also found that one way to dislodge the bound electrons from their residence inside the atom is to increase the temperature of the gas. This increases the random movements of atoms and frequent collisions can ionise the gas by ridding the atoms of their electrons.

There is yet another way to ionise matter. When high energy photons interact with atoms, its electron can absorb the photon and in the process get kicked out of the atom. This is called the 'photoelectric effect' and was first explained by none other than Einstein.

Consider now the effect of the light of the first stars in the Universe on the surrounding matter. As mentioned earlier these stars emit a large amount of ultraviolet photons. These photons have high enough energy to help electrons in hydrogen atoms get away and become free. The surrounding gas is therefore likely to become ionised, as it used to be earlier, before the epoch of recombination. Initially only the nearby gaseous matter is ionised and as time proceeds more and more gas becomes ionised due to prolonged radiation of ultraviolet photons by the stars. Gas hidden inside large bound objects is not as vulnerable to this harmful ultraviolet radiation as the intergalactic gas, since photons would lose energy in penetrating the objects and would not be effective deep inside any large clump of matter. Most of the intergalactic gas may therefore become ionised, or re-ionised (since it was ionised in the very early Universe anyway), sometime after the first stars lit up the Universe.

There are telltale signs of this re-ionisation of the Universe. The observations of the intergalactic gas at the present epoch, and in the recent past, show that it is highly ionised. It is indeed fascinating that we not only know the existence of the tenuous gas lurking between galaxies, but we also know whether or not it is neutral. How does one infer the physical state of this gas?

One can use the analogy of observing the bones of a patient using X-rays to explain this. The patient is irradiated with X-ray photons from one side which are detected on the other side as they come out of the body. The X-rays travel through most of the flesh unimpeded but get absorbed by the bones, which then show up as dark objects in the X-ray photograph. The dark absorption features in this photograph reveal the existence of the bones.

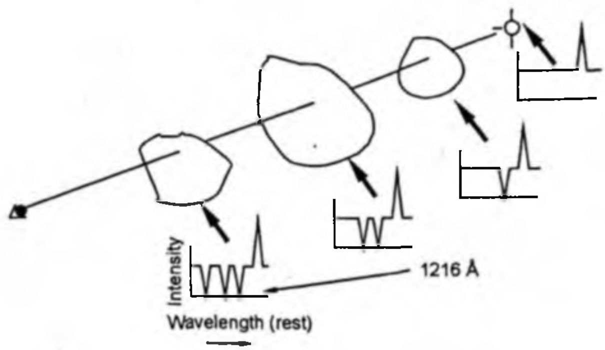


Fig. 6.4 *This schematic diagram shows how the spectrum of light of a distant active galaxy (top right corner) acquires absorption features owing to the absorption by neutral hydrogen atoms in intervening matter. The initial spectrum (top right) shows the emission line due to emission of Lyman-α photons (at rest wavelength* 1216 Å*) from the distant galaxy. Photons with wavelength shorter than this get absorbed – after their wavelengths have been stretched to become comparable to* 1216 Å *by the expansion of the Universe – by neutral hydrogen in the intervening space. Discrete intergalactic 'clouds' would imprint isolated absorption lines in the observed spectrum.*

Consider now a distant luminous object in the place of the X-ray source. Let us suppose this object is an active galaxy, since they are very bright and can be detected easily from a large distance. The radiation from this galaxy contains photons of various wavelengths. A few of these photons can get absorbed by hydrogen atoms, if there are any, in the intergalactic space. Electrons which are bound inside atoms can absorb photons only of certain wavelengths, or energies (unlike free electrons which can interact with photons of all wavelengths). This is because electrons are allowed to possess only certain amounts of energy inside an atom. Consider a hydrogen atom with its electron in the lowest energy level. It can only absorb a photon whose energy can take it to the next or a higher energy level. It will not absorb photons of any wavelengths that will not take it to any of the allowed energy levels.

The hydrogen atoms, if there are any, in the intergalactic gas would mostly have their electrons in the lowest energy state. The most relevant photon energy for them to absorb is that which will take it to the next energy level. Photons with this energy have a wavelength of approximately 0.0001216 mm, and are called 'Lyman-α' photons. The physicist Theodor Lyman had investigated photons which take an electron in hydrogen atom in the lowest energy level to various upper levels of energy. This series of photons is called the Lyman series and the first in this series (which can take the electron to the next level) has the tag of 'α'.

Photons travelling from the active galaxy that we have considered will however get stretched by the expansion of the Universe on their way toward us. Suppose a photon had a wavelength at the source less than that of a Lyman-α photon. On its way toward us, its wavelength will be stretched to larger values. At some point its wavelength will become equal to a Lyman-α photon. If the intergalactic matter here has neutral hydrogen atoms in it, there is then a chance for this stretched photon to be absorbed at this region. One can argue that this is going to happen to all photons which had wavelengths shorter than the Lyman-α photon at source. As a matter of fact, *all* photons which initially have wavelengths shorter than the Lyman-α photon will be liable to get absorbed by hydrogen atoms in the intergalactic space at *some* intervening region (Fig. 6.4). It is straightforward to determine the probability of this occurrence, and it is found that even a small amount of neutral atoms in the intergalactic gas can absorb all these photons from a background source. Just like the bones in a body absorb X-rays from a background source.

When astronomers discovered active galaxies in the 1960s, James Gunn and Bruce Peterson drew the attention of cosmologists to this possibility. They claimed that one could probe the existence of neutral atoms in the intergalactic space by examining the spectrum of radiation of these distant active galaxies. One needs to look at the part of the spectrum with photons which would have had initial wavelengths less than 0.1216 micron, that of Lyman-α photon. If there is even a small amount of hydrogen atom in the intergalactic gas, this part of the spectrum would be devoid of any photons, and would be dark there. The radiation from distant galaxies should be devoid of these particular photons[9].

Astronomers have frantically searched for such signatures in the spectra of distant active galaxies in the last four decades. They have however failed to see any clear signature of *all* photons shorter than the Lyman-α wavelength dropping out of the spectrum in any active galaxies. There are some signs of absorption of some photons at some discrete regions in the intervening space though, but no sign of any *en masse* drop out.

This can only be explained if hydrogen in the intervening space is ionised to a large degree and there are hardly any neutral atoms there. The discrete absorption features at various parts of the spectrum show the existence of neutral atoms at *some* isolated regions. It is now believed that these regions are slightly denser than average, where intergalactic matter has become clustered to some extent and is somewhat clumpy. The intergalactic gas by definition lies outside objects produced by the clustering process in the Universe, and can therefore be taken as the residual material from this process. It now appears that even this residual material has clustered here and there, although by a small amount and not enough to create any gravitational bound objects. These small variations in the intergalactic matter produce the variation of density of hydrogen atoms in space, which in turn show up as dark lines in the spectra of distant galaxies. On the average, however, the intergalactic gas is very highly ionised.

It is thought that the cause of this re-ionisation is connected to the formation of galaxies in the Universe. As structures evolved in the Universe and stars appeared with copious production of ultraviolet photons, they have re-ionised the Universe.

Fig. 6.5 *The process of re-ionisation of the Universe is schematically shown here. Dark regions indicate the presence of hydrogen in the atomic state and the white regions denote ionised hydrogen. As the first galaxies light up they ionise the hydrogen gas in their vicinity with the help of ultraviolet radiation (first picture on the left: possible redshift epoch – approximately 30. Slowly more hydrogen in the Universe gets ionised (second picture in the middle), and at some point the intergalactic hydrogen in the whole Universe gets re-ionised (last picture: possible redshift epoch – sometime between 6–20).*

The very first stars in the Universe probably initiated the process of ionisation in their vicinity. Slowly the ionised regions would have increased in size and gradually overlapped with one another to make the whole Universe ionised (see Fig. 6.5). The exact evolution of this process must have depended on the details of the star formation history of the Universe. For example, it is crucial to know if the molecular hydrogen was destroyed effectively by ultraviolet photons and if there was a lull in the formation of stars. This would have stopped the ionisation process temporarily or may even have allowed the ionised gas to revert to the neutral phase for a while.

Recently astronomers have made some interesting observations in this regard. They have been conducting a survey of objects in the Universe with the ultimate goal of finding redshifts of a million galaxies, much larger than the 2dF survey mentioned in the first chapter. This survey, called the Sloan Digital Sky Survey (SDSS), is being done with the help of a medium-sized telescope dedicated to this project. The surrey has already discovered many distant galaxies, a good fraction of them being active galaxies[10].

The SDSS survey has found several active galaxies at redshifts of around six whose spectra show the existence of neutral atoms in the diffuse intergalactic gas, just as Gunn and Peterson had predicted four decades ago. They have also found that active galaxies with slightly lower redshifts, that is at a later time, until the present epoch, do not show any absorption from neutral atoms, as mentioned earlier. It is then possible that the state of the intergalactic gas in the Universe around the redshift of six was just prior to being completely re-ionised (i.e., at a phase between the last two panels of Fig. 6.5). It then had a small amount of neutral hydrogen atoms which were soon ionised.

Very recently cosmologists have also used another tool to study the history of re-ionisation. We know that the photons of the cosmic microwave background radiation have been travelling more or less freely since the epoch of decoupling. This is because matter formed atoms at that point, which do not interact as liberally with photons as free electrons do. If however atoms get re-ionised at a later epoch, setting the electrons free once more, the photons would again interact with these electrons. This renewed scattering can leave some imprint on the photons which cosmologists hope to study in detail.

Photons have a spin like many other quantum particles (which in classical parlance is called *polarisation* of electromagnetic waves). It turns out that scattering with electrons can alter its polarisation. The photons of microwave background in different directions of the sky show variations of its polarisation, because of scattering with electrons just prior to the epoch of decoupling. This has been confirmed by observations from Antarctica and more recently by *WMAP*. The scattering with electrons later in the re-ionised gas also leaves its mark in the polarisation pattern of photons in the sky.

Recent preliminary results from the observations by *WMAP* satellite seem to indicate the history of re-ionisation of the Universe was far from simple. The amount of scattering of photons of background radiation that is indicated by data is not consistent with a simple and sudden re-ionisation at a redshift of six. It is now believed that the history of re-ionisation was a prolonged one and perhaps proceeded with fits and starts. It is difficult to say anything more than this with the present data. There is a flurry of activity in research on this topic at present.

6.6 Future Observations

The renewed interaction of photons with electrons in the intergalactic gas will also change the anisotropy pattern of the background radiation temperature. It will change some of the signatures in the fossil that is carried by the background radiation photons, as we discussed in the last chapter. On one hand, the scattering of photons by the born-again electrons will wash out some anisotropic patches of temperature distribution, and on the other hand, the movement of the re-ionised gas will produce some new patches. In short, the angular power spectrum of anisotropy of microwave background will get distorted because of re-ionisation of the Universe. The new spots embossed on the background radiation by the re-ionised gas are expected to be of small angular scales.

It will be possible to test these theoretical ideas by sensitive observations in the near future. As mentioned earlier, the European countries plan to launch a mission called *PLANCK* in 2007, which will observe the fine structure of the temperature pattern of microwave background. *PLANCK* is also being designed to detect the polarisation of these photons to a high accuracy. Observations with it will soon test many theoretical ideas about the history of re-ionisation of the Universe. Another experiment with an array of telescopes, *ALMA* (Atacama Large Millimetre Array), planned for a site in the Atacama desert of Chile, is also expected to detect the temperature variation at high resolution[10].

Another group of cosmologists is planning to use not electrons, but the endangered species of neutral hydrogen atoms towards the end of the re-ionisation process, to probe this epoch. Astronomers routinely use photons of a particular wavelength that hydrogen atoms emit to probe these atoms. This radiation has a rather long wavelength, of 21 cm, that makes it less vulnerable to being scattered by dust particles in space. It has therefore become a part and parcel of astronomical observations of hydrogen atoms in the Universe. It is often simply called the '21 cm radiation'. Most radio telescopes in the world routinely use this radiation to study hydrogen atoms in the Universe.

If we however plan to detect this '21 cm radiation' from neutral atoms during the re-ionisation process, we should remember that these photons would get redshifted by the time they reach us. If re-ionisation occurred at a redshift of, say, six, then this radiation will get redshifted by a factor of $(1 + 6) = 7$. This means that we will detect this photon with a wavelength of $7 \times 21 = 147$ cm, slightly longer than a metre.

It has now become extremely difficult to detect radiation of these wavelengths coming from space because of the increasing use of facilities like mobile phones, FM radio, and so on. Mobiles phones use radiation of similar metre-sized wavelengths for communication. The noise created by frequent use of these devices smear out the faint extraterrestrial signals at these wavelengths.

The use of these devices in India is still less than that in the Western countries, especially in places far away from large cities (although it is fast catching up with Western countries). Indian astronomers therefore planned to build a large radio telescope to study radiation at these wavelengths from space. The Giant Metre-wave Radio Telescope (GMRT), designed by Govind Swarup, has been recently built at Khodad near Pune (Fig. 6.6). Astronomers have already begun to use this telescope in earnest and have already made some interesting observations[12].

Some cosmologists believe that if one detects the redshifted 21 cm radiation from hydrogen atoms just prior to the complete re-ionisation era, the intensity of this radiation will not be the same all over the sky. Firstly we will detect this radiation only from regions with a significant amount of neutral atoms. During the last stages of the re-ionisation process, when ionised regions overlapped with one another, the distribution of neutral atoms in the Universe was rather patchy. Secondly, the temperature of these atoms was also not uniform. Martin Rees and his colleagues have recently studied this possibility in some detail. According to their calculations, the hydrogen gas in the vicinity of early active galaxies could have been at a higher temperature than far away from them. This effect would cause the intensity to vary across the sky, which could then be tested by instruments like GMRT. Such an observation would be a novel way of probing the epoch of re-ionisation of the Universe.

Fig .6.6 *A few antennas of GMRT (Giant Metre-wave Radio Telescope) situated near Pune in Western India are shown. It combines the signals detected in 30 antennas, each 45 metre in diameter, and produces a combined image of the sky in radio wavelengths. (Picture courtesy: NCRA, Pune)*

Astronomers are also planning to build a larger instrument than the GMRT in the future. The effective size of this telescope will be a square kilometre. The design of this 'Square Kilometre Array' (SKA) has not yet been decided, but cosmologists are already excited at the prospect of observations with it.

6.7 Other Effects of Early Galaxies

The effect of early galaxies on the surrounding matter is not confined to ionising the gas by photons radiated by them. Cosmologists have been studying a few other possible effects and the possibilities of observing their signatures.

We have seen that the early generation of stars appeared mostly inside dwarf or even smaller galaxies. These galaxies have very small masses which makes their gravitational fields very feeble. This may have a calamitous effect on the gas contained in them. If we recall our discussion in the first chapter on the evolution of stars, massive stars end their phase of thermonuclear reactions at their cores in spectacular explosions that rip them apart. Such an explosion not only destroys the parent star (apart from the inert core), but also has tremendous effect on the matter in the vicinity, the diffuse gas in the galaxy.

The gas surrounding a supernova explosion gets slapped by the momentum of the ejected material. Ultimately a substantial fraction of the kinetic energy of the explosion is imparted to the gas as thermal energy, essentially heating it up. This increases the random motion of the gas particles. A part of this energy is of course lost through radiation but if there are frequent explosions in the galaxy to compensate for it, the gas can remain hot. A hot gas is however liable to escape from the clutches of gravity, when the random movement of gas particles exceeds the escape velocity of the object. This of course does not cause any problem in a large galaxy, since its gravity is large enough to contain the hot gas.

It can be a problem though in two cases. Firstly if the rate of formation of stars increases for some reason, then the rate of input of energy into the gas through supernova explosions increases, and can make the gas very hot. This gas can then become vulnerable enough to escape even from a large galaxy. This is called a 'galactic wind' and a galaxy can lose a substantial fraction of its gas content by this process.

Secondly, if the galaxy has a small mass, its feeble gravitational field may not be able to contain the hot gas even if the rate of formation of stars is not abnormally high.

These are not figments of theoretician's imaginations any longer. Hot gas has been clearly seen to develop a wind in galaxies with pathological star formation rates (see Fig. 6.7) and also from nearby dwarf galaxies. Often this hot gas emits X-rays and recently launched X-ray telescopes have seen many such examples.

Fig. 6.7 *In this picture of M82, a starburst galaxy, taken by the Subaru telescope in Hawaii, shows hot gas coming out of it. The elongated object is the galaxy itself. The hot gas near the central region (seen in red) is seen to be rising out of the galaxy. The temperature of this hot gas is around a million degrees. (Courtesy : NAOJ. Copyright Subaru Telescope, NAOJ. All rights reserved) [Also see Plates]*

Consider now the case of the early generation stars inside small galaxies. It is believed that the effect of supernova explosions in these galaxies could make the gas hot enough to spill out of the galaxy in a wind. The details of this process are still being worked out at present, and if these possibilities are confirmed by detailed calculations, then the early galaxies could have influenced the Universe in many ways.

It is not only the energy input from supernova explosions that may drive gas out of a galaxy. There are also a few other ways in which this can happen. We have earlier found that galaxies often interact and collide with each other. These interactions must have been more frequent in the past since galaxies were closer to one another then. These interactions can have disastrous effects on the galaxies by stripping them of their gas in the wake of the collision. It is possible that frequent collisions in the past also could have taken some gas out of the galaxies.

This gas is of course not pristine and is enriched with heavy elements spewed out by supernova explosions. Cycles of star formation and supernova explosions inside the galaxy are expected to slowly increase the abundance of heavy elements, like carbon, silicon and so on. When this enriched gas goes out of the galaxies and mixes with the intergalactic gas, it will also enrich them to some extent.

It is intriguing that astronomers have recently discovered just such signatures of heavy elements in the tenuous intergalactic gas. They have detected elements, like carbon, silicon and others, in the same way they deduced the existence of hydrogen atoms, with the observation of spectra of bright active galaxies in the background. They have detected absorption features in these spectra which could only be due to absorption by heavy elements. The abundance of these elements in the intergalactic gas is very small, much smaller than in the Sun for example, but it still points to the interesting ways in which early galaxies could have left their marks in the Universe.

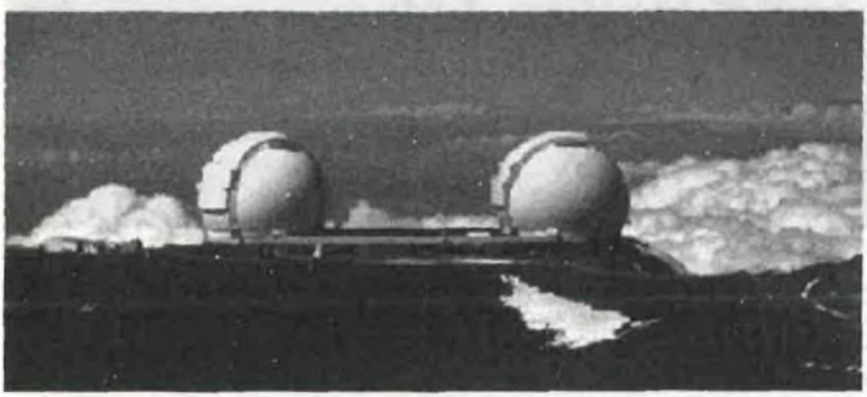


Fig. 6.8 *The largest optical telescope in the world, the Keck telescope is situated in the Hawaii islands. Each of the twin telescopes has a mirror 10 m in diameter.*

These new observations have again been possible with remarkable technological advances in the recent years. The large Keck telescope (Fig. 6.8) with a mirror as large as ten metres in diameter helped astronomers detect these faint signatures in the spectra of distant galaxies. It is thrilling enough to know that astronomers can detect the presence of the tenuous intergalactic gas. But to detect the presence of extremely small amounts of elements like carbon and silicon in this tenuous gas is a feat that would have been unimaginable even a few years ago.

European astronomers have now constructed a large telescope in Chile, called the Very Large Telescope (VLT), which consists of a few separate telescopes. Signals from these telescopes are being combined so that the effective size of the whole telescope is many times the size of individual mirrors, which are large enough any way. They have also built a replica of the Keck telescope by its side and efforts are on to combine the signals from the two. Apart from these, Japanese astronomers have built a large telescope, called the Subaru, at Hawaii islands, which has a mirror of 8 metres in diameter.

6.8 More Planned Observations

The effects we have discussed so far of the early galaxies, including ionisation and enrichment with heavy elements, are still indirect at best. A few astronomers have been planning more ambitious projects, which may bring forth more direct clues about the very early generation stars and galaxies. These experiments would be very difficult. One would need giant infrared telescopes to be able to observe galaxies at a redshift of, say, ten. Optical photons at that epoch would get redshifted to the infrared range at present epoch, and this is why one needs an infrared telescope. As a matter of fact, a group of European astronomers announced in early 2004 the discovery of a galaxy at redshift of 10 using the infrared detectors at the Very Large Telescope in Chile.

A few infrared telescopes are being planned to be launched into space in the near future. NASA has already sent the Spitzer Space Telescope in 2003, which will be able to observe galaxies at a redshift of around five. A much larger infrared telescope is being planned after that although its design has not yet been decided. It is tentatively scheduled to be launched in 2011 and has been named the 'James Webb Space Telescope'. It will most probably have a mirror of diameter 6.5 metres—more than three times the size of the present Hubble Space Telescope[13].

7

Evolution of Galaxies

... Everything was there,
Every single thing
Waiting was to bring,
Clear from hydrogen
All the way to men...

Robert Frost, *A Never Naught Song*

Galaxies have lit up the Universe since time immemorial. We have discussed in the earlier chapters the process of initial star formation in gravitationally bound clumps of matter. The characteristics of galaxies have certainly changed over time, owing to many factors such as the slow depletion of gas in producing stars, or interaction with neighbouring galaxies, and so on. We will discuss a few aspects of the evolution of galaxies in this final chapter, especially those galaxies which have received much attention in recent years because observations have made them worth discussing in detail.

7.1 Morphology and Dynamics of Galaxies

The evolution of galaxies is still not well understood. There are many aspects of it which remain uncertain. On the one hand, it is difficult to theoretically analyse a complex system like a galaxy. A typical galaxy like our Milky Way contains approximately a hundred billion stars. There is of course a large amount of dark matter, and then some amount of gas in it. It is difficult to predict from simple analysis how such a system should evolve with time, and how the image and the spectrum of a galaxy changed on account of this evolution. Some astrophysicists try to simulate the evolution with the help of computers but even then it remains a very difficult topic to study. Although one knows the trajectory of a particle well, the evolution of a system of particles, with different properties, becomes difficult to track if there are a large number of particles in a system.

On the other hand, it is only very recently that it has been possible to study galaxies from the past in good-detail. Before the advent of the Hubble Space Telescope it was almost impossible to think of observing galaxies from redshifts larger than one. This field of extragalactic astronomy has seen a spectacular growth in the last few years because of a number of large, ground-based telescopes as well as the Hubble space telescope. These difficulties in understanding the evolution in theoretical terms and testing the theories with observations have made the study of this topic very challenging.

There are a number of questions that have kept astronomers busy in this regard. For example, why are some galaxies spiral and others elliptical? One has a flattened structure like a disc whereas the other is more or less spherical, with a certain elongation along some direction. One has its stars and gas rotating around the centre—for example, our Sun rotates around the centre of the Milky Way with a speed of about 220 km per second. Others do not have any definite axis of rotation and the random motions of stars dominate over any small directed motion they may have[1].

There are other differences between spiral and elliptical galaxies. Spiral galaxies like the Milky Way contain a large amount of gas and dust particles, much more than that compared to a typical elliptical galaxy. Spiral galaxies (Fig. 7.1) seem to have a keen interest in star formation and one finds a large number of young stars in them, whereas a typical elliptical galaxy (Fig. 7.2) looks more like an old age home for stars where the phase of active star formation seems like a distant memory. We know that massive stars die young and are blue in colour, whereas stars which stay around to shine for a long time are mostly low in mass and red in colour. Spiral galaxies therefore have more blue stars than elliptical galaxies which show a preponderance of old, red stars.

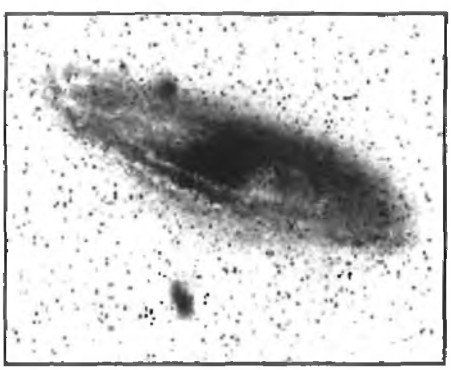


Fig. 7.1 *Andromeda is a spiral galaxy like our Milky Way. (Picture Courtesy :*Jason Ware)

One is naturally tempted to ask if this dichotomy in the morphology of galaxies is an outcome of the process of evolution of galaxies, or if this dichotomy is etched in the galaxy population from the very beginning. It is a question of whether 'Nature' is more important then 'nurture', or the other way round. We have so far seen that when the clustering process in the Universe makes objects that are gravitationally bound, cooling of the gas in it slowly creates an environment for star formation. Are there some clumps of matter that are destined to become spiral (or elliptical, for that matter) because of some characteristics acquired at birth? In this case, one wonders what these characteristics could be. Or, is it that clumps of matter are all born alike and they acquire their looks much later, due to some evolutionary process? In this case, it is possible that the environment of a galaxy is important for its evolution. It is similar to questions that parents would like to ask when their children grow up to become very different from each other: is the difference due to the inner constitution of genes, or due to differences in environments in which they grow up?

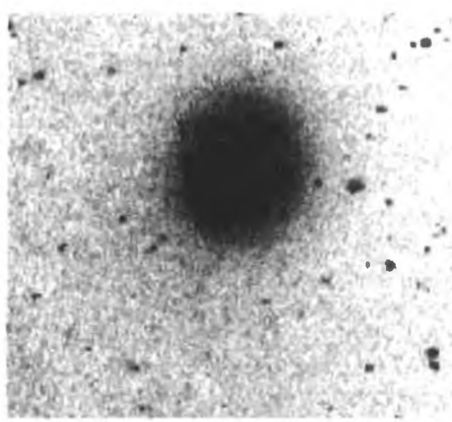


Fig. 7.2 *A neighbouring galaxy of Andromeda (M32) is an elliptical galaxy. (Picture Courtesy : Jason Ware)*

One characteristic of a clump of matter which we have not yet discussed could be important in this regard. Consider the motions of stars in a spiral and an elliptical galaxy. Physicists use the concept of momentum—the product of mass and velocity—to describe the motion of an object. It is not just the mass or the velocity that determines the effect of some force on a moving object, for example, but its momentum. This is true for the case of linear trajectory, or motion in a single direction. Similarly in the case of rotational motion one uses the concept of 'angular momentum' of an object, which is a combination of the mass, its distance from the axis of rotation and its speed. For example, the fate of a rotating top on a plane, and the way any small undulation in the plane may affect its rotation, will depend on its angular momentum[2].

The *extent* of the object is important for its angular momentum. If a given amount of mass is redistributed in such a way as to make the constituent particles more distant from the axis of rotation, keeping their speeds the same, the angular momentum will increase. The angular momentum of an object however does not change unless there is a force acting on it in such a way as to encourage or discourage its rotation. In the absence of this kind of influence on an object from outside, its angular momentum remains a constant. In our case of an object where the mass is being redistributed to make the particles more distant from the axis of rotation, the speeds of particles will have to change to keep the angular momentum a constant.

Examples of such a change in speed are very common. For example, when a diver curls up in mid-air while plunging into water, the speed of tumbling through air also increases. During the curling up process, his limbs come closer to the axis of rotation and accordingly the speed of rotation increases to keep the angular momentum a constant.

The stars in an elliptical galaxy move around the galactic centre in all possible directions and there is no common axis of rotation. The net angular momentum therefore is small since the angular momentum from constituent stars tend to cancel each other's contributions. The stars in a spiral galaxy however rotate around a certain axis, and their angular momenta add up to make a subtantial contribution to the whole galaxy. This is then an important distinction between the dynamics of spiral and elliptical galaxies, in that the spiral galaxy possesses a large total angular momentum compared to a typical elliptical galaxy.

How do galaxies acquire any angular momentum at all? We have so far discussed the formation of clumps of matter but we have not discussed any rotational motion in these clumps. In 1969, based on an earlier work by Fred Hoyle in 1949, James Peebles suggested a mechanism by which this could happen. According to this scenario, clumps of matter acquire some amount of angular momentum during the process of their formation. Gravity is again behind this phenomenon as in the case of clustering of matter.

The strength of gravity decreases with distance. If an object is large in size, the gravitational attraction due to another object on all parts of its body will not be uniform. The parts of it which are closer to the gravitating object will feel more attraction than other, more distant parts of it. This is the reason, for example, behind the tides in the oceans on Earth.

Consider two clumps of matter, one bigger and more extended than the other. Let us also suppose that both objects are somewhat elongated. It is of course natural for objects to form with different kinds of shape and an exactly spherical object formed out of clustering in the Universe is an unrealistic expectation. These objects are of course moving past one another in space. If we consider such an encounter between a small and large object from the vantage point of the large object, we will find that the part of the small object that is nearer to us would get more attracted than other parts. The gravitational attraction between the objects may slow them down somewhat. In addition to this, the greater attraction for the nearer side of the small object would also make it *turn* a bit. This would set the object rotating around itself, or, in other words, give it a net angular momentum.

The large object will also turn but by a smaller amount than the other. Cosmologists now believe that this is how clumps of matter that later became galaxies acquired angular momentum (Fig. 7.3). Once acquired, the angular momentum is unlikely to change unless there is some interaction with other galaxies. Therefore, as the clump of matter contracts under gravity, the speed of rotation would increase to keep its total angular momentum a constant.

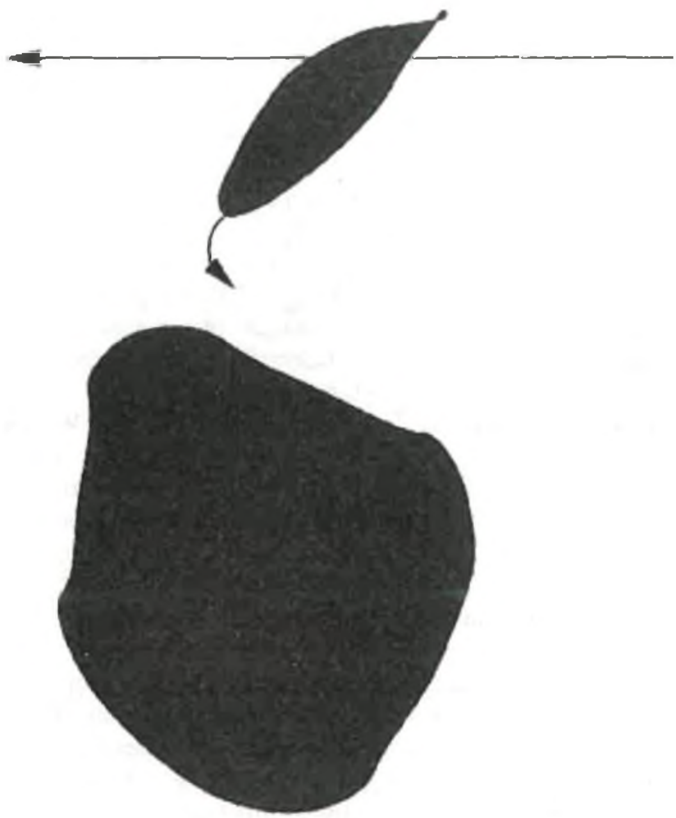


Fig. 7.3 *The tidal origin of angular momentum of galaxies is shown schematically.*

The angular momentum acquired by clumps of matter through this tidal mechanism is obviously not uniform for all clumps. Some will acquire larger and some smaller angular momenta than average and there will be a range in the magnitude of angular momentum thus acquired. It is then possible that the clumps of matter poor in angular momentum are destined to become elliptical galaxies, and clumps with better luck may later become a spiral galaxy. Detailed calculations and computer simulations seem to lend some credence to this idea, but one can hardly ignore the effects of 'nurture' from environments in the later phases of evolution of a galaxy.

There is now some support for the idea that interactions between galaxies can significantly alter their shapes and dynamics. It has been noticed that there is an abundance of elliptical galaxies in regions where galaxies crowd together to form galaxy clusters. Spiral galaxies are hard to come by in such regions, especially in the centre of the crowded region. Some believe that this is because close and frequent interactions between spirals in these crowded regions have destroyed their fragile spiral arms and discs. It has also been noticed that some elliptical galaxies bear the mark of past interactions, in the motion of their member stars, and sometimes in the type of stars in them. Collisions of galaxies can stir up the gas content of galaxies, making them vulnerable enough to form stars. Some elliptical galaxies in fact have very young stars, an exception to their largely old populations of stars, which shows the sign of recent interactions. But it may not be true that all elliptical galaxies have become so because of interactions. A number of elliptical galaxies at high redshift, i.e., in the past, look very similar to elliptical galaxies at present. This is an active topic of research now and no firm conclusions have been reached yet.

Another interesting aspect of interaction between galaxies is that by stirring up the gas it tends to feed the black hole at the centre, like stoking a furnace. Usually the central black hole is starved of its diet in the circumstances of a normal galaxy, like our Milky Way. If however a large amount of gas is pushed towards the centre, the increased rate of accretion by the black hole may trigger jets and other manifestations of an active galaxy. Collisions between galaxies may push a normal galaxy towards a more frenzied state and turn it into an active galaxy. It is also possible that the stirring up of the gas increases the rate of formation of stars to a pathological magnitude. These are called 'starburst galaxies' [?], which we encountered earlier in the context of galactic winds. It is intriguing that most such starburst galaxies do show imprints of earlier interactions.

These are active areas of research at present. The emphasis is on studying the evolution of galaxies at various wavelengths of electromagnetic radiation, and piecing together a coherent story from these clues. Different wavelengths carry signatures of various physical processes, and studies done at one wavelength often complement those at other wavelengths. One of the strengths of modern astronomy is the ability to put together evidences of different kinds from these multi-wavelength studies into a complete description. A number of telescopes operating at various wavelengths have been brought into the arena of this multi-wavelength research, without which modern astronomy, and cosmology, would have lagged far behind.

7.2 History of Star Formation in the Universe

Another question that comes to mind when we think about the evolution of galaxies since the earliest epochs is how the illumination of the Universe, at various wavelengths, has changed with time. One wonders if the level of illumination which we find today has remained the same throughout the history of galaxy evolution, or if it was higher or lower in the past.

Fig. 7.4 *Many distant galaxies can be seen in this picture taken by Hubble Space Telescope. (Courtesy : NASA and STScI) [Also see Plates].*

It is again to the credit of modern technology that significant progress has taken place lately in this field. It was very difficult, if not next to impossible, even a decade ago to think of observing galaxies at redshifts larger than or comparable to unity in some detail. This redshift corresponds roughly to an era when the age of the Universe was about a third of its present age. Higher redshifts were thought to be in the realm of science fiction. The advent of the space telescope has changed all that.

In December of 1996, the Hubble space telescope (Fig. 7.4) took a long exposure photograph of a region of the sky (near the Great Bear constellation) at a stretch for ten days. Faint objects can be captured in a photograph if the plate is exposed for a long time giving a large number of photons the opportunity to leave their signatures. Moreover this is a telescope where the image is hardly influenced by the turbulence in our atmosphere, which so often blurs precious details in images taken by ground-based telescopes. This is the advantage of having a telescope in outer space. This long exposure picture taken by astronomers has become a treasure house for them, with a fantastic amount of information hidden in it, enough to keep them busy for a while. They have recently taken a much longer exposure photograph with the Hubble space telescope, the results of which were announced in February 2004[4]. These pictures have already given them some clues about the history of the background light in the Universe.

They have detected in this picture the presence of galaxies at redshifts even as large as five. This has really been a giant leap for astronomers. It is not only that they have searched for distant (i.e., early) galaxies, but they have also pieced together the information from various epochs to build a history of star formation.

We have seen in the first chapter that heavy stars are short-lived and that blue and ultraviolet photons contribute substantially to their total budget of radiation. The stars look bluish for this reason. One could turn this argument around to say that the presence of blue stars in a galaxy would point towards a recent phase of star formation. In fact the fraction of blue photons in its spectrum would be a diagnostic of the rate at which stars are being formed there. One could therefore use the average colour of a galaxy to estimate the formation rate of stars in it.

Astronomers have analysed the colours of galaxies at various epochs detected in the Hubble Deep Field and have drawn up the history of average star formation rate in the Universe (see Fig. 7.5). According to this study, this rate has decreased over time in the recent past and has come to an all time low at the present epoch. It appears that the rate of star formation at a redshift of unity was about ten times larger than it is now. (In a flat Universe, without any cosmological constant, this epoch was when the Universe was approximately one-third of its present age but we should remember that the relation between age and redshift has many uncertain factors.)

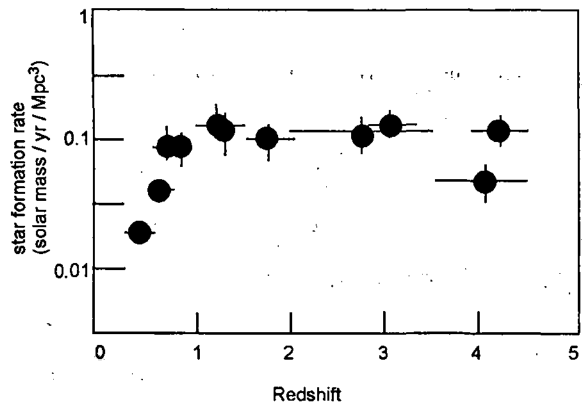


Fig. 7.5 *The plot shows the history of star formation rate in the Universe. It shows how the average rate of formation of stars has changed from the early epochs (high redshift) to the present day (redshift of zero: shown at the left end). The star formation rate (on the y-axis) is shown in the unit of solar mass per year per cubic megaparsec. The data points have large uncertainties at high redshift and the evolution of star formation rate before the redshift of three still remains somewhat inconclusive.*

The rate of star formation seems to reach a plateau beyond a redshift of 1.5 or so until a redshift of three. It is still not very clear how the plot extends beyond the redshift of three because of paucity of observations and also because of various uncertainties. Blue photons are easily scattered by dust particles, more than photons toward the red end of the spectrum (like the efficient scattering of blue light by air molecules in our sky which makes it blue). The observed amount of blue photons therefore does not give a complete picture, depending on the dust content of a galaxy. According to some astronomers the plateau continues beyond the redshift of three whereas some tend to believe that it should either decrease or increase.

Another window of wavelength that has opened up recently is that of sub-millimetre wavelengths, straddling the range between radio and infrared photons. Photons that were emitted in the infrared range at redshift of five would have been stretched to sub-millimetre waves by the present epoch. The dust grains which could have absorbed blue and visible light in distant galaxies, would heat up and re-radiate in infrared wavelengths. These photons would be available at present in the sub-millimetre range. It is however extremely difficult to detect these photons, which is why this window has opened up only recently. It is difficult to detect faint signals in sub-millimetre wavelengths because the detector itself would radiate in this range to some extent. If we recall our discussion on blackbody spectrum, we will find that to suppress this radiation the detector should be cooled to low temperatures. As a matter of fact to be able to detect signals from the distant past, one needs to cool the detectors to about 0.1 kelvin (i.e., approximately −272.9 Celsius). Recently a few sub-millimetre telescopes have been built and it has now become an exciting area of research.

It is hoped that with the help of these telescopes, and the upcoming infrared telescopes such as the *Spitzer Space Telescope* and the *James Webb Space Telescope*, astronomers will be able to extend their knowledge of star formation well beyond the redshift of three. They will then be able to complete the story of evolution of star formation from the very early generation stars to the present epoch.

The history of star formation as plotted in Fig. 7.5 is of course a gross average taken over a large region of space containing many galaxies. One wonders about the contributions in this average plot from different types of galaxies, such as ellipticals and spirals. Did the relative contributions of different types of galaxies change with time? One also wonders about the difference in star formation in galaxies which cluster and which prefer to remain isolated. Are there differences in the evolution of galaxies depending on their environment? These are some of the questions that astronomers hope to answer in the future.

7.3 Gamma Ray Bursts

Astronomers have recently stumbled upon a rather interesting tool in the study of star formation in the Universe. We have earlier discussed the use of different wavelengths from radio to infrared to visible to ultraviolet. This new tool is associated with photons of much higher energy than these—in the range of gamma rays, which are more powerful than X-rays. The technology for detecting these photons is different, and the physical processes that give rise to gamma radiation are also very different from those responsible for milder photons.

There was a puzzling observation made by American scientists in the 1960s, at the height of the cold war between USA and USSR. Scientists from both sides were worried about possible secret nuclear explosions and tests carried out by the other side. American scientists had launched a secret satellite with detectors on board to probe any such tests. Nuclear explosions give rise to a large amount of gamma rays, which are not produced in abundance by usual terrestrial phenomena. Gamma rays detected by the secret satellite would have been sure signs of any such covert nuclear experiments. To the surprise of American scientists, the satellite picked up gamma radiation very frequently, even when they were very sure that there were no nuclear tests around (from other indicators like seismic studies). After more than a decade, when the military scientists were convinced that these gamma rays had nothing to do with happenings on the Earth, and that they were cosmic in origin, they de-classified their data and it became a puzzle which astronomers were encouraged to try to explain.

A number of theories were suggested in the 1980s to explain this phenomenon. What astronomers observed was a sudden burst of gamma radiation in a certain part of the sky which would continue for seconds to minutes. It was as if there were some explosions emitting copious amount of these gamma rays in a short time. They were also frequent, with big bursts happening almost once a day. It is difficult to pinpoint the direction of arrival of gamma rays to a high accuracy, because of their tremendous energy content. (Being energetic, they penetrate too deep into any 'mirror' to focus them and makes the 'mirror' useless, unlike less energetic visible photons which can be brought to focus and form an image easily with mirrors.) It was not possible to connect the bursts to any possible celestial object because of lack of directionality. One therefore could not estimate the distances of the sources of these bursts. And without any idea of the distances, one did not know the actual brightness, and the energy content, of the phenomenon involved. A less bright star can outshine a really bright star if it happens to be at a smaller distance from Earth. One cannot determine the actual brightness of an object without any estimate of the distance to it. Astronomers had hoped that if they could pinpoint the direction from which the gamma rays arrived, there would be a possibility of studying that region of the sky in detail to find a probable culprit, perhaps a deviant star or a bizzare galaxy, or perhaps something completely different. All these hopes came to naught because of lack of any accurate knowledge of the direction.

There were some indirect evidences though that these bursts were taking place at a great distance from us. In the beginning of the 1990s, there were indications that the distribution of these bursts was more or less uniform over the sky and not confined to any particular region. Since our galaxy has the shape of a disc, if the sources of bursts were confined to our galaxy, the bursts would have appeared only in the patch of the sky where one finds the Milky Way. The distribution of the bursts would have been non-uniform and confined to a long patch in the sky. Since this was not the case, the suspicion was that the bursts originated at much larger distances, or equivalently, in the distant past[5].

The situation has become significantly better in the last few years. It has been found that although the burst in gamma radiation lasts only a few seconds, at the most a minute or so, the source then appears to radiate in a lower frequency (smaller energy) range, in X-rays and then in visible wavelengths, even down to radio wavelengths. Detecting and pinpointing a source of X-rays is a relatively easy task compared to that of detecting gamma rays. It has been possible to quickly pinpoint the direction of the source by detecting the 'afterglow' in X-rays with the help of X-ray telescopes on satellites. With this information at hand, large telescopes on Earth as well as the Hubble space telescope have been directed toward the source to find out the nature of the host galaxy of the burst, and also study the 'optical' afterglow of the burst. There is also a flurry of activity on the theoretical front to understand the physical mechanism of the explosive event and to discover the basic cause of these gamma ray bursts. It is heartwarming to note that a couple of optical telescopes in India (in Nainital and in Hanle, near Leh) and GMRT (near Pune) which was mentioned earlier have joined in the study of these sources, and observations from them are being used to constrain theoretical models.

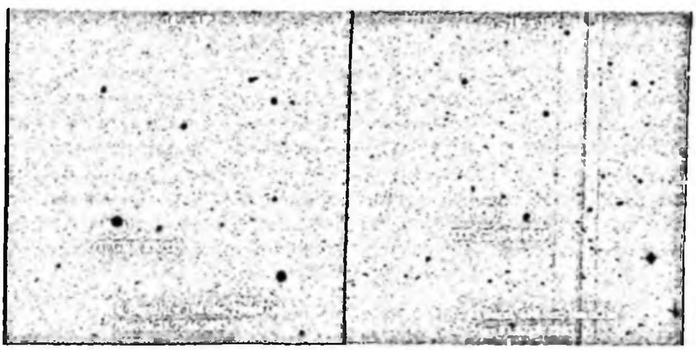


Fig. 7.6 *The dimming of the optical afterglow of the gamma ray burst of 29th March 2003 (the image on the left was taken on 29th March and on the right, on 1st May 2004). The position of burst is shown by circle. Images were taken at the Aryabhatta Research Institute of Observational Science, Nainital, India. (Courtesy : ARIES, Nainital)*

Astronomers have found clear signatures of absorption by intergalactic gas in the spectrum of these sources, just as one finds in the spectra of distant active Galaxies. These observations have confirmed that the sources typically lie much beyond the confines of our Galaxy. One can easily determine the redshift of the intergalactic gas from the absorption features in the spectrum. The source of gamma ray burst must lie *behind* or beyond this redshift. In this way a lower limit on the distance to the source can be determined. Already the redshifts of a number of such sources have been determined. The largest redshift (i.e., the most distant burst) encountered so far is approximately six.

There is still considerable uncertainty as to the cause of such explosions. A common belief for a number of years has been that it is caused by the collision and merger of two neutron stars, or two black holes. Many however believe that it is caused by a super-supernova, or a 'hypernova', which takes place when an extremely massive star ends its phase of nuclear fusion and when a black hole core appears in its centre. We have learned in the first chapter that a typical supernova produces a neutron star in its centre and rips itself apart in the process. A hypernova is believed to be an extreme version of a supernova, with added complications from the appearance of a black hole in the centre.

There is however some concurrence among astronomers in the connection of gamma ray bursts with formation of stars. Whether or not it is due to collisions of neutron stars or a hypernova of a massive star, they seem to be somehow related to stars. In this case, there should be a connection between the frequency of these bursts with the rate of formation of stars. It is then hoped that the study of gamma ray bursts, their evolution with redshift in particular, will provide important clues to the rate of formation of stars. Since the explosions are very bright, and are easily detected over a large distance, this then provides us with a handy tool to study the early Universe. Already the preliminary studies done with the bursts at high redshift seem to yield the *same* rate of star formation that was deduced from other observations, as described earlier. It is hoped that this interesting probe of the early Universe will provide more information in the near future.

7.4 Clusters of Galaxies

Galaxies are seldom found completely isolated. Most galaxies in the Universe are members of small groups of galaxies (like the Local Group of which our Milky Way is a member) or large clusters of galaxies. Structures as large as galaxy clusters have arisen in the Universe only recently. As the process of clustering in the Universe has progressed, the power of clustering for large length scales have increased, making the inhomogeneities of this length scales palpable in recent times. Equivalently, galaxies have come together to become gravitationally tied to each other to form clusters of galaxies. A very important question in the study of galactic evolution is whether or not the environment of galaxies affects its internal evolution.

Clusters of galaxies are spectacularly different from what one might naively think. They are not just a menagerie of galaxies. Apart from the member galaxies, they contain a large reservoir of diffuse gas confined by gravity. The gravity is, of course, dominated by contribution from dark matter. This gas is very hot, being mainly heated by motions of galaxies, which essentially reflect the strength of gravity inside clusters. The temperature of the gas varies between ten to hundred million degrees in different clusters, depending on the total mass and size. Curiously, this gas is not pristine, but has heavy elements besides hydrogen and helium. The abundance of these heavy elements is somewhat smaller than (about a third of) what is found in the Sun, for example.

We have seen how the tenuous intergalactic gas is also enriched with heavy elements but its level of contamination is far less than that in the gas in clusters. One is naturally tempted to ask when and how this gas was enriched and how it has managed to enrich itself to far more than the intergalactic gas at large. It is also not clear where most of this gas has come from, and whether or not much of this gas has been gurgled out of galaxies.

Astronomers have recently made a lot of progress in the study of these clusters of galaxies with a number of X-ray telescopes aboard satellites. The hot gas emits a copious amount of X-ray photons and since the clusters harbour a large amount of this gas, clusters of galaxies are very luminous in X-rays. As a matter of fact this was one of the unexpected discoveries made by the first generation X-ray satellites. (Riccardo Giacconi, who was a pioneer in the development of X-ray satellites, was awarded the Nobel Prize in physics in 2002.)

It appears that close interactions inside the crowded region of a cluster can have significant effects on galaxies, especially if there is a small galaxy involved in the encounter. This 'galaxy harassment' has been claimed to render some galaxies gas-less and in some cases completely change their identity. It is also possible that some interactions lead to the triggering of activity of inner black holes and produce active galaxies in a cluster. The effect of the jets coming out of these active galaxies on the gas content of the cluster could be interesting. It may heat up the gas in the vicinity of the active galaxy. It may also happen that the material in the jet, which is composed of very energetic particles and has low mass density compared to the gas, may produce bubbles rising through the gas, like bubbles in a soda bottle. These bubbles of energetic particles have been detected in radio wavelengths even in the outer periphery of some clusters. The dynamics and the evolution of these processes are now topics of active research.

It is not just the luminous clusters of galaxies, but also the galaxies in smaller groups which have become interesting objects of study lately, with the help of deep observations that are possible now. We would like to mention one intriguing observation in this regard. It has been noticed that there is a deficit of small satellite galaxies in these groups compared to what one expects from the ideas of clustering in the Universe. This has been found in our Local Group of galaxies as well as in a few other groups. The theory of clustering in the Universe, as described in detail in Chapter 5, can predict the number of galaxies within a given mass range in a certain volume, given the power spectrum of clustering. Using a power spectrum that is consistent with all other data (of galaxy surveys, for example), one can then predict the number of small galaxies, in a given range of mass, within a group of galaxies, and compare them with observed statistics. It has been found that there is a distinct paucity of small galaxies in groups.

It is not yet clear why there should be a large discrepancy, given the level of agreement of the power spectrum with other data. It is possible that there is something fundamentally wrong with our knowledge of clustering of matter. Some astronomers have thought of a less radical explanation.

The statistics of small galaxies would depend crucially on the ability of observing a clump of matter. The power spectrum of clustering predicts the number of clumps of matter in a given mass range. There is of course a difference between a clump of matter and a luminous galaxy. If a clump of matter is somehow devoid of stars, it will not be visible in the sky and will be left out of the statistics gathered by astronomers. This will happen if some galaxies, especially small galaxies, could somehow be made infertile and somehow inhibit star formation in them.

One way to stop star formation in a clump of matter is to hinder the infall of gas inside it. We have seen that initially it is the dark matter that gathers and forms clumps, because most of the matter in the Universe is dark. Normal gas then follows suit and gets trapped by the gravity of the clump of dark matter. If the temperature of the gas is high, however, the random motion of gas particles could be larger than the escape velocity of very small clumps, and it will be difficult to trap this hot gas inside the small clump. If the intergalactic gas is somehow heated up, small clumps of matter would never be able to get their share of gas trapped in them, and they could be devoid of stars. This would then reconcile the discrepancy between the theoretical prediction of small galaxies and their observed statistics. One would say that the small clumps of matter are there as predicted by theory, it is only that they cannot form stars and are therefore not visible. What could however heat up the intergalactic gas?

It is now believed that the process of re-ionisation of the Universe also increased the temperature of the intergalactic gas, in addition to ionising it and enriching it with heavy elements. When ultraviolet photons from the first galaxies knock off electrons from atoms and ionise them, the ejected electrons can be energetic enough to heat the gas. As a matter of fact, the temperature is expected to be as large as ten thousand degrees soon after the Universe re-ionised. This temperature is enough to make the intergalactic gas avoid getting into small sized clumps of matter and keep them dark and invisible in the Universe.

It is interesting that a clear signature of the re-ionisation of the Universe which was caused by its early generation of stars is present in the Universe today. Future studies will certainly bring out not only the details of the early epochs of galaxies in the Universe, but also shed light on how the Universe has come to be the way it is today.

8

Notes

... But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep.

Robert Frost, *Stopping by Woods on a Snowy Evening*

Chapter 1 : Our Place in the Universe

1. Traditionally, calculus is used to explain the importance of elliptical orbits in the case of gravitation. The reader can however find a splendid explanation using geometry in *Feynman's Lost Lecture: The Motion of Planets Around the Sun,* W. W. Norton & Co., 1996.

2. One can estimate the core temperature of a star from basic principles of physics in the following way. If ionised hydrogen gas has density of ρ then the number of particles per unit volume is $2\rho/m_p$, m_p being the mass of a proton, and the factor of 2 takes into account an electron and a proton for a total mass of proton, neglecting the electron mass. The pressure that it exerts is then $\sim 2\rho k_B T/m_p$, where $k_B = 1.38\times10^{-23}$ joule per kelvin is a constant. (In these *Notes* we will frequently estimate the approximate values of quantities and will use the symbol $\sim$ for 'almost equal to' rather than the '=' sign for exact equality.) This pressure balances the inward pressure due to gravity. Suppose the total mass of the star is M and its radius is R, then the pressure at the centre due to gravity is approximately (gravitational force) $\div$ (volume) $\sim GM^2/R^4$. Here G is the gravitational constant. Equating these two pressure, one gets $T \sim GM m_p/k_B R$. Using the solar radius of $R = 7 \times 10^8$ m and solar mass of $M = 2 \times 10^{30}$ kg, one obtains the core temperature of $T \sim 1.5 \times 10^7$ kelvin.

   In reality this calculation is much more complicated, where one has to equate the difference in the pressure between every layer of the star to the difference in the gravitational force. Consider concentric spherical shells in the star. At a radius r, the pressure difference between two infinitesimally separated shells (with separation Δr), can be written as $\Delta p(r)$. The force outward will then be (pressure) $\times$ (area)$= -\Delta p(r) 4\pi r^2$. The weight of the shell is given by (total mass) $\times$ (g), or $(4\pi r^2 \Delta r \rho) \times (GM(r)/r^2)$, where $M(r)$ is the mass inside the radius r. Equating this two forces, one gets for the condition of equilibrium,

$$\frac{dp(r)}{dr} = -\frac{GM(r)\rho(r)}{r^2}. \tag{8.1}$$

In making the rough estimate above, we have approximated $\frac{dp}{dr}$ to (P/R), where P is the pressure at the centre and R is the radius of the star. It is interesting that the estimate made above is a very good approximation to the real temperature at the centre of the Sun.

3. To be precise, the spectrum of the light from a star depends on its mass. Meghnad Saha was the first to show how such an analysis could be done. The degree of ionisation of a gas naturally depends on its temperature and density. Meghnad Saha theoretically calculated the relation between the degree of ionisation and the temperature of a gas, and one can apply this relation to the degree of ionisation of various gases in the outer layers of stars as deduced from its spectrum, and find out the temperature of the star. This temperature is related to the mass of the star.
4. A gas is said to be 'ionised' in this state of isolated electrons and leftover positive ions. This state of matter is also called 'plasma'.
5. In this reaction, two more particles appear along with deuteron: one is a neutrino and the other a positron. A positron has the same mass as an electron but with an opposite charge. Even when the masses of these positron and neutrino are added, the final mass of a helium nuclei is smaller than that of four individual protons.
6. There are other kinds of particles, called 'bosons', which do not repel one another in this way. Photons, the particles of light, are particles of this kind.
7. This outward pressure becomes important when the mean separation between particles like electrons becomes comparable to the de Broglie wavelength, $\lambda \sim h/p$, where h is the Planck's constant ($= 6.7 \times 10^{-34}$ joule s) and p is the momentum of the particle. According to the 'exclusion principle' in quantum mechanics, the particles are inhibited from sharing the same quantum state, or equivalently, having separations much smaller than the de Broglie wavelength. These particles are called Fermions. Matter in this state is said to be 'degenerate'.

8. To be precise, it is possible for stars with mass less than 1.44 times that of the Sun. This was theoretically discovered by Subramanyam Chandrasckhar. Legend has it that he arrived at this conclusion while on the ship from Madras to Cambridge, UK, at the age of nineteen. He was awarded the Nobel Prize in 1983 for this important discovery.
9. This important event is not recorded in any medieval Indian text. It is possible that the political climate of the country then was not conducive to scientific research.
10. The word 'parsec' was coined from 'parallax' and 'second'. The 'second' here refers to a measure of an angle in the sky: an arcsecond is one-sixtieth of an arcminute, which is in turn one-sixtieth of a degree. If a star is situated at a distance of a parsec from the Earth, then the radius of the orbit of Earth around the Sun would subtend an angle of an arcsecond at this star. It then follows from geometry that a parsec is equal to 3.08×10^{16} metre.
11. It is not difficult to appreciate the implication of this observation even without detailed calculations. The observations of astronomers say that the brightness of the Milky Way drops drastically away from its centre. In other words, the 'visible' mass is concentrated towards the centre. Consider now a star (of mass m) at a distance r from the centre, and assume that the total mass of the Milky Way inside this radius is $M(r)$. According to Newton's law, the gravitational force on this star is $GM(r)m/r^2$. If the rotational speed of the star is v, the centrifugal force due to its rotation around the centre is mv^2/r. In a steady state situation, these two forces would balance each other, and we will have,

$$\frac{mv^2}{r} = \frac{GM(r)m}{r^2} \tag{8.2}$$

which readily gives us $v \propto \sqrt{M(r)/r}$. In the case of the Solar System, the total mass is concentrated in the Sun, and one has $M(r) = M$, independent of r, and so $v \propto \sqrt{1/r}$, and so the speed decreases with distance. Since astronomers find v to be almost a constant far away from the centre, it would mean that $M(r) \propto r$, meaning that the total mass inside radius r should increase proportionally with the radius. Since the brightness decreases rapidly with distance from the centre of the Milky Way, the total brightness inside a given radius r grows very slowly with radius. Comparing these two trends, one concludes that there is a large amount of 'invisible' mass in the Milky Way, especially in the outer regions.

12. There are a few possible constituents of dark matter. It is possible that this matter like ordinary matter but is packed in an object that cannot

radiate energy. Planets like Jupiter, or which are slightly more massive than Jupiter, cannot radiate light whereas they contain non-negligible amount of matter. It is also possible that the dark matter consists of an ensemble of black holes. The other possibility is that the dark matter is made of neutrinos, which do not interact with matter or with themselves very much, and therefore remain 'invisible'. Finally it is possible that the dark matter is made of some exotic particle whose existence we are still not aware of, but whose properties can be estimated from the ideas of particle physics. Many physicists are currently involved in the attempts to detect such particles with the help of clever experiments.

13. Readers can find more information on this survey at the internet website `http://www.aao.gov.au/2df`

14. A minority of cosmologists believe that the tendency of clustering continues *ad infinitum*. According to them the large-scale structure of the Universe is like a fractal. Without going to the mathematical definition of a fractal, one can give the example of a cauliflower. A small part of a cauliflower is the miniature version of the whole thing. The tendency of branching out or producing clusters is the same for a small segment and a large part of a fractal.

Chapter 2 : The Relativistic Universe

1. For those readers who like to think mathematically, one can explain the relativity of time in the following way: if the distance between the mirrors is h, the speed of the train is v and if t is the time elapsed, then from Pythagorus's theorem, the length of the oblique path of light is $\sqrt{h^2 + v^2(t/2)^2}$, since during half the total time elapsed t, the train travels a distance of $v(t/2)$ (see Fig. 8.1). The total distance of the oblique path is then $2\sqrt{v^2(t/2)^2 + h^2}$. If the speed of light c is a constant, then light will take

$$t = 2\sqrt{\frac{v^2(t/2)^2 + h^2}{c}} \tag{8.3}$$

to traverse this distance. The solution of this equation is given by

$$t = \frac{2h}{\sqrt{(c^2 - v^2)}} \tag{8.4}$$

Since this will be the time elapsed as measured by the second observer, let us write this as t_2. We therefore have, $t_2 = 2h/\sqrt{(c^2 - v^2)}$. The time elapsed as measured by the first observer will simply be $t_1 = 2h/c$. We can combine these two expressions as $t_1 = \frac{t_2}{\sqrt{1-v^2/c^2}}$. This equation clearly says that t_1 is always larger than t_2. In other words, the time measured by the observer sitting inside the train will run slower than that of the observer standing outside.

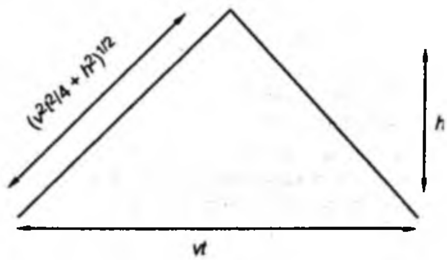


Fig. 8.1 *The geometry of the light path in the example of a light-clock with vertical separation h, travelling on a train moving with speed v is schematically shown.*

2. For readers with some knowledge of special relativity and four-dimensional spacetime, one can show this mathematically in the following way. One can define the speed in the four-dimensional spacetime (three dimensions of space and one dimension of time) as $u = \frac{dx'}{d\tau}$, where x' has four components $[x' = (ct, x, y, z)]$ and τ is the 'proper time', which is defined as $d\tau^2 = dt^2 - \frac{1}{c^2}[dx^2 + dy^2 + dz^2]$. The magnitude of this speed in spacetime is given by $|u| = \sqrt{\frac{dx'^2}{d\tau^2}} = \sqrt{\frac{c^2dt^2 - dx^2 - dy^2 - dz^2}{dt^2 - \frac{1}{c^2}(dx^2+dy^2+dz^2)}}$ = c=constant. One can also write this equation as, $c^2(\frac{dt}{d\tau})^2 - [(\frac{dx^2+dy^2+dz^2}{d\tau^2})] = c^2$; or, in a slightly different form, as, $c^2(\frac{d\tau}{dt})^2 + [(\frac{dx^2+dy^2+dz^2}{dt^2})] = c^2$ If we identify the first part as the speed in time and the second part as the speed in space, then their sum is a constant, and if one increases, the other is bound to decrease.

3. According to Einstein's theory, a light ray grazing the surface of an object of radius R and mass M will get bent by an angle $4GM/(Rc^2)$, where G is the usual gravitational constant and c is the speed of light. Using the radius and mass of the Sun in this expression, one gets an angle that is 1/3000 of a degree. Physicists have used this aspect of gravity—its ability to bend light rays—in various ways. We have learnt in the first chapter that a part of dark matter could be massive planets like Jupiter or compact objects like white dwarf and neutron stars. If such a compact object happens to cross the line of sight of

a distant star, the starlight would momentarily appear to be brighter as a result of the lens action of gravity (of the compact and dark object). Astronomers have recently found such signatures of lensing of distant stars by dark compact objects in our Galaxy.

4. In Newtonian physics, one can estimate the threshold criterion for forming a black hole by requiring the escape velocity from an object of mass M and size R to be c, the speed of light. This gives one the condition, $c^2/2 = GM/R$, or $R = 2GM/c^2$. Actual calculations in general relativity shows that any object within this radius cannot communicate with the world outside. This is called the 'Schwarzschild radius', honouring Karl Schwarszchild who first derived it.

   The Schwarzschild radius of the Sun, for example (of mass 2×10^{30} kg) is ≈ 3 km, which is very small compared to its actual size, ~ 700 million metre. The Sun would have been a black hole had its size been smaller than or comparable to its Schwarzschild radius.

5. For more of such experiments which test the theory of relativity, the reader can look up the book *Was Einstein Right?: Putting General Relativity to the Test* by Clifford Will (HarperCollins, 1993).

Chapter 3 : The Expanding Universe

1. Huble's law can be written as $v = Hr$, where r is the distance to a galaxy and v its speed of recession. 'H' is a constant of proportionality which is called the Hubble constant. Current estimate of H is approximately 70 km/s per megaparsec.

2. Curious readers can check easily that any other relation between the speed and distance, would not be consistent with the assumption of a homogeneous and isotropic Universe. If Hubble had discovered that the speed is proportional to the square of the distance, say, then one can prove in the case of our one-dimensional rubber band Universe that the expansion would not be uniform. The expansion would then appear to proceed at different rates from different vantage points in the Universe.

3. One can try to estimate this critical density even without going into the details of the general theory of relativity. Consider the state of expansion of a region of space around us. Suppose this region is spherical with a radius R. The volume of this region of space is then $(4\pi/3)R^3$. If the mean density of the Universe is ρ, the total mass contained in this region would be (density $\times$ volume), or, $M = (4\pi/3)R^3\rho$. Consider now the case of a galaxy situated on the edge of this region. Since the distance from us is R, its speed of recession from us would be $v = H_0 \times R$ (where H_0 is the present day

value of the Hubble constant). Let us now compute the total energy of this galaxy. A part of this energy would be due to its speed; its magnitude is $(1/2)mv^2$, or $(1/2)m(H_0R)^2$. Here m is the mass of the galaxy. The rest of the energy would be due to gravitational force : $(-GMm/R)$. The total energy is then

$$E = \left(\frac{1}{2}\right) m(H_0R)^2 - Gm\left(\frac{4\pi}{3}\right) R^2\rho \tag{8.5}$$

This is essentially the Friedmann equation for the Universe . (For a Universe dominated by radiation energy, one uses the corresponding radiation energy density ϵ instead of ρ in this equation by putting $\rho = \epsilon/c^2$.)

If the total energy is positive then the galaxy will recede from us for ever. In the case of negative energy, the galaxy will stop its expansion and return towards us at some point. We can compare this situation with the expansion of the Universe. The case of zero energy would then correspond to the case of density being equal to the critical density.

When does the total energy become zero ($E = 0$) in this case? This happens when

$$Gm\frac{4\pi}{3}R^2\rho = \left(\frac{1}{2}\right) m(H_0R)^2 \tag{8.6}$$

In other words, when

$$\rho = \frac{3H_0^2}{8\pi G} \tag{8.7}$$

This is the value of the critical density. Using the value of the present day Hubble constant, we obtain a density of $\sim 10^{-26}$ kg per cubic metre, which approximately means five hydrogen atoms in a volume of a cubic metre.

We can also estimate the age of the Universe with the above equation. Writing the speed in the language of calculus as $\frac{dr}{dt}$ we can rewrite the equation as:

$$\frac{dr}{dt} = \sqrt{\frac{2GM}{r}} \tag{8.8}$$

If we say that the distance between galaxies is $r = 0$ at the time of the Big Bang, that is, at $t = 0$, then the solution of this equation is given by

$$\frac{2}{3}r^{3/2} = \sqrt{(2GM)}\,t \tag{8.9}$$

In other words, the mean distance between galaxies in the Universe will increase as : $r \propto t^{2/3}$. If we now assume that the present age of the Universe is t_0 and that the present day value of the Hubble constant is H_0, then we can write:

$$\sqrt{\left(\frac{2GM}{r}\right)} = \frac{dr}{dt} = v = H_0\,r \tag{8.10}$$

whose solution is

$$\sqrt{2GM} = H_0 r^{3/2} \tag{8.11}$$

Using the above equation again, we can write,

$$t_0 = \left(\frac{2}{3}\right)\left(\frac{1}{H_0}\right) \tag{8.12}$$

This is the age of the Universe if the density equals the critical density.

For a radiation dominated Universe, the energy density decreases faster (see *Notes* for the next chapter), as $\rho \propto r^{-4}$. In this case, following the same method as above, one can show that, $r \propto R^{1/2}$.

4. An exotic characteristic of this 'particle' is that it has negative pressure. Usually we only enounter material with positive pressure. A more or less simplified account of research on this can be found in *Quintessence* by Laurence Krauss (Basic Books, 2000).

5. Readers may think that the brightness of a distant source would diminish and make the night sky dark. Although the brightness of individual sources does decrease with distance, this would not have any effect on the brightness of the night sky. Let us imagine a spherical region of the Universe around us and suppose that this sphere is divided into concentric shells. Let us estimate the total brightness at the centre (at Earth) due to the stars distributed in these shells. Suppose a particular shell has radius R, then the distance to a star in this shell would be $\approx R$. We know that the brightness of light decreases in proportion to $1/R^2$. Stars on a shell with larger R would seem fainter. The surface area of the shell would also however increase with R. This area is proportional to R^2. We then have the following situation: the brightness of individual stars decrease with $1/R^2$ but the number of stars at distance R increase as the surface area of a shell at this distance, and hence, as R^2. The brightness at Earth due to any shell is thus a constant, and independent of its

distance from Earth. If the whole Universe is now imagined to be made of such shells, one can easily reach the conclusion that the night sky should be infinitely bright.

6. Some bold cosmologists do however contemplate upon the birth of our Universe in the context of an ensemble of many universes (multiverse). Curious readers are referred to *Before the Beginning : Our Universe and Others* by Martin Rees (Helix Books, 1998).

7. Cosmologists have also thought about other possibilities for the history of the Universe. Notable among them is the 'steadystate theory' which was the focus of intense research at one point. Hermann Bondi, Fred Hoyle and Thomas Gold postulated that the Universe retains a steady state even as it expands, by creating matter to keep the mass density a constant in time. The Universe, in this hypothesis, has been and will remain for ever. This Universe does not have any beginning. Recent observations however show that the Universe has been evolving in the past. A few cosmologists have recently come up with the suggestion of a 'quasi-steady state theory', although there has not yet been any supporting observation for it.

8. *Darkness At Night: A Riddle of the Universe* by Edward Harrison (Harvard University Press, 1989) has a marvellous account of the Olbers's paradox.

9. If the wavelength of light at the moment of emission is λ_e and the observed wavelength is λ_o, then the redshift z is defined as $z = (\lambda_o/\lambda_e) - 1$. If spacetime expands by a factor of two to double the wavelength of light, that is, if $(\lambda_o/\lambda_e) = 2$, then this would correspond to a redshift of $2 - 1 = 1$.

10. Equation (8.11) shows how the Hubble constant varies with time. Since the lengthscale r is proportional to $1/(1+z)$, it follows that

$$H(z) = H_0(1+z)^{3/2} \tag{8.13}$$

when Ω=1, and there is no cosmological constant or dark energy. Integrating equation (8.10) to a certain redshift (or, t), one then gets an age of the Universe at that redshift, as,

$$t = \frac{2}{3H_0(1+z)^{1.5}} \tag{8.14}$$

where z is the redshift epoch. One recovers the present day age of the Universe by putting $z = 0$ here, since the present day corresponds to $z = 0$. The age will be larger if the value of Ω is smaller. If there is dark energy or a cosmological constant, then the relation between age and redshift becomes complicated.

11. These are called 'Type 1a' supernovae and they are a bit different from the supernovae described as dying explosions of stars in the first chapter of this book. Type 1a supernovae occur when a white dwarf star accretes enough gas from its companion to cross the mass limit for a white dwarf star, the Chandrasekhar mass limit. At this point, a runaway thermonuclear explosion occurs which rips the star apart. Since the Chandrasekhar mass limit depends on the fundamental constants of nature, it is believed that this type of supernovae reaches a peak brightness that is also a constant.

12. Readers can get updated information on these observations on the Internet at this address: `http://www-supernova.lbl.gov`

Chapter 4: The Early Universe

1. For a radiation dominated Universe, the Friedmann equation for a critical density Universe is given by (see *Notes* for the previous chapter),

$$\left(\frac{dr}{dt}\right)^2 = \frac{8\pi G\rho r^2}{3} \qquad (8.15)$$

Since the radiation energy density decreases as $\rho \propto r^{-4} = Ar^{-4}$, with A being a constant, one finds, $(r\frac{dr}{dt})^2 = 8\pi GA/3$, or,

$$rdr = \sqrt{\left(\frac{8\pi GA}{3}\right)}dt \qquad (8.16)$$

which upon integration (with the initial Big Bang condition that $r = 0$ at $t = 0$) gives,

$$\frac{r^2}{2} = \sqrt{\left(\frac{8\pi GA}{3}\right)}t \qquad (8.17)$$

so that, $A = 3r^4/(32\pi Gt^2)$. This means that $\rho = 3/(32\pi Gt^2)$. Since the radiation energy density (equivalent to ρc^2) is given by $4\sigma T^4/c$ (see *Note* No. 4 below), it follows that,

$$T = 1.5 \times 10^{10} t^{-1/2} \qquad (8.18)$$

Using now a temperature of $\sim 10^9$ K, the age of the Universe is found to be ≈ 100 s.

2. As long as protons and neutrons are kept in equilibrium by electrons, positrons, neutrinos and antineutrinos, the relative numbers of protons and neutrons is given by the Boltzmann distribution of particles in equilibrium. According to this distribution, in thermal equilibrium at a certain temperature T, the relative number of particles with energy difference of ΔE is given by $\exp(-\Delta E/kT)$, where k is the Boltzmann constant ($= 1.38 \times 10^{-23}$ joule per kelvin). The relative number of protons and neutrons would therefore be

$$\frac{n_n}{n_p} = \exp\left(\frac{-(m_n - m_p)c^2}{kT}\right) \tag{8.19}$$

where the difference in the masses of neutrons and protons is $m_n - m_p \approx 0.0014 m_p$. Using $m_p \approx 1.67 \times 10^{-27}$ kg, one gets a ratio of $n_p/n_n \sim 5$ at $T \sim 10^{10}$ K, the temperature of the Universe before neutrons and protons fell out of equilibrium. After they fall out of equilibrium, this relative number of neutrons and protons would remain 'frozen', irrespective of the temperature, since the reactions that convert neutrons to protons and vice versa become negligible.

This ratio, however, changes slightly with time, since free neutrons, decay in time (to protons, electrons and antineutrino) on timescales of about 900 seconds. (Bound neutrons do not decay.) We must find the ratio of neutrons and protons by the time dueterium nuclei are formed, when the temperature has dropped to 10^9 K, or when the age of the Universe is about 100 s (see previous note). In about 100 s, a fraction of neutrons would decay into protons, increasing the relative number of protons to neutrons somewhat. Using the half-life time of decaying neutrons, can show that this relative number becomes ~ 7 in 100 s, when the temperature of the Universe is $\sim 10^9$ K. This means that there was a neutron for every 7 protons in the Universe.

3. The book *The First Three Minutes : A Modern View of the Origin of the Universe* by Steven Weinberg (Basic Books, 2nd Edition, 1993) is a classic among books on cosmology written for general readers. The Nobel prize winning physicist has explained the basic principles of modern cosmology in a clear and easy language. A must-read for everyone interested in cosmology.

4. The Planckian spectrum of radiation of a black body at temperature T is given by (with dimensions of energy per unit area per unit time per unit solid angle per unit frequency),

$$B_\nu(T) = \frac{2h\nu^3/c^2}{\exp(h\nu/kT) - 1} \tag{8.20}$$

where h is Planck's constant ($= 6.7 \times 10^{-34}$ joule s) and ν is the frequency of radiation. The energy density of this radiation can be calculated by integrating this spectrum over frequencies, dividing by c and multiplying by a factor of 4π representing the total solid angle. This gives the energy density as,

$$u(T) = \frac{4\pi}{c}\int_0^\infty B_\nu(T)d\nu = \frac{8\pi^5 k^4}{15c^3h^3}T^4 = aT^4 \tag{8.21}$$

where $a = 7.5\times10^{-16}$ joule per cubic metre per kelvin^4 is a constant.

One aspect of this radiation that we will use later is the relation between the temperature and the volume within which it is confined. The corresponding pressure due to radiation is one-third of the energy density. Since the energy density is $\epsilon = aT^4$, we have $p = aT^4/3$. If the size of the confining 'box' is R, the volume is R^3 and the total energy inside the box can be written as ϵR^3. From the first law of thermodynamics, one has for an adiabatic (without any input or output of energy) expansion of the box, $dE = -pdV$, which implies,

$$\begin{aligned} &d(\epsilon R^3) = -\epsilon R^2 dR \\ &\Rightarrow 4T^3dTR^3 + 3T^4R^2dR = -T^4R^2dR \end{aligned} \tag{8.22}$$

which leads to,

$$\frac{dT}{T} = -\frac{dR}{R} \tag{8.23}$$

This implies that the radiation temperature decreases with the expansion of the Universe as $T \propto R^{-1}$. Since length scales in the Universe increases as $1/(1+z)$ (z being the redshift), it follows that $T \propto (1+z)$.

5. One can explain the crux of Gamow's argument without going into much detail in the following manner. Firstly, there are four important parameters that govern the result of nuclear reactions: the abundance and speed of particles, the duration available and the extent to which the particles interact. The threshold temperature for nuclear reactions to produce a deuteron is $\sim 10^9$K. This is because the binding energy of the deuteron nuclei is $E_b \sim 2.2$ MeV and the corresponding temperature is of the order E_b/k_B, where $k_B = 1.38\times10^{-23}$ joule K^{-1} is the Boltzmann constant mentioned earlier. The average thermal speed of particles (like proton) at this temperature is $v \sim 5000$ km per second. One can use the Friedmann equations to determine the age of the Universe to be $t \sim 100$ seconds when it had a temperature of 10^9 K (see the first *Note* of this chapter). The last factor is expressed by the area of cross-section which measures the extent with which the participant nuclei interact. One has from the physics of nuclear

reactions that the relevant cross-section is $\sigma \sim 10^{-33}$ square metre. Suppose the density of particles is n per unit volume. The number of nuclear reactions to have taken place within a duration t is of the order $n v t \sigma$. If one now requires primordial nuclear reactions to nave produced a significant amount of deuteron nuclei, one could express the requirement mathematically as $n v t \sigma \sim 1$. This condition readily gives the required density as $n \sim 10^{24}$ per cubic metre, and for particles like proton, the mass density is $\sim 10^{-3}$ kg per cubic metre.

Gamow then compared this density with the present-day density of the Universe. He used observations made by Hubble which indicated a mass density of $\sim 10^{-27}$ kg per cubic metre. This means that the volume of the Universe has increased by a factor of $\sim 10^{24}$ since the era of nucleosynthesis. In other words, the length scales have increased by a factor of $(10^{24})^{1/3} = 10^{8}$. The temperature of the radiation however decreases inversely proportional to the length scales (see previous *Note*), and so must have dropped to a temperature of $(10^{9}/10^{8}) \sim 10$ K. This was the essence of Gamow's argument. We should remember here that we have glossed over many important details. It is still interesting that this simple argument gives the right order of magnitude of the temperature of the background radiation.

George Gamow was not only a legendary physicist; he also was a great communicator of science to the general public. He was also known for his sense of humour. When he and his student, Herman Alpher, wrote the above-mentioned estimate in a paper, Gamow thought that if there were an author with a name sounding like 'beta', between his name and that of his student, it would sound like the first three letters of the Greek alphabet. He then thought of Hans Bethe who was an authority on nuclear reactions in stars and who agreed to Gamow's idea when he approached him. This famous article was therefore published in the trio's name (H. Alpher, H. Bethe, G. Gamow (1948), *Physical Review*, vol. 73, p.803).

6. One can find a lot of similarity between this story and that described in Aurther Koestler's *The Sleepwalkers* on the discovery of elliptical orbits by Johannes Kepler. There were a lot of indications that the shapes of the orbits calculated by him were elliptical but he kept on missing the conclusion for various reasons. According to Koestler, it was as if Kepler were sleepwalking through the clear evidences for elliptical orbits.

7. Penzias and Wilson were awarded the Nobel prize in 1978 for their momentous discovery.

8. Recent observations by COBE find the temperature of this radiation to be 2.728 ± 0.002.

9. To be precise, one should say that the abundance of deuterium depends on the ratio between the energy densities of matter and radiation. The smaller the value of this ratio, or equivalently, the larger the relative importance of radiation over matter in the Universe, the larger is the amount of residual deuterium. Astronomers estimate this ratio by determining the abundance of deuterium. They then readily determine the density of (ordinary) matter, since the radiation energy density is known from the observations of the cosmic background radiation. And this turns out to be about 5% of the critical density of the Universe.

10. This relation between relative speed and the change in wavelength of radiation was discovered by Christian Doppler. If the speed of the signal (sound or light) is c, and the relative speed (along the direction connecting the source and the observer) is v, the frequency of the signal ν is shifted to

$$\frac{\nu'}{\nu} = \frac{1}{1 + v/c} \tag{8.24}$$

It then follows that for speeds much less than that of light, a photon of wavelength (at rest) λ_0 will appear as a photon of wavlength λ, when the relative speed is v, where,

$$\frac{\lambda - \lambda_0}{\lambda_0} \approx \frac{v}{c} \tag{8.25}$$

For speeds comparable to that of light, one needs to take into account relativistic corrections and the Doppler shift is reduced in the case of redshift (recession) and increased in the case of blueshift (approach). In the case of redshift, contraction of length from special relativity decreases the wavelength at the same time the Doppler effect is increasing it, reducing the effect of the Doppler shift. In the case of blueshift, relativistic contraction reduces the wavelength further, increasing the effect of the Doppler shift.

11. It is possible to show that black body radiation of temperature T will retain its Planckian spectrum to an observer moving with speed v, but its temperature will change by a fractional amount,

$$\frac{\Delta T}{T} \approx \frac{v}{c} \tag{8.26}$$

If the fractional change in temperature is approximately 0.2%, then the speed of the observer relative to the frame of reference of the radiation (that is, the peculiar speed) will be 0.2% of the speed of light. This means a speed of ~ 600 km per second.

Chapter 5 : Structures in the Universe

1. In the last few decades many models which involved processes other than gravitational instability, such as the effect of radiation on matter have been thought of. For example, one model showed that in the presence of a uniform radiation field, two clumps of matter would mutually shadow each other, making the space between them devoid of radiation pressure. The radiation pressure outside would then induce these clumps to come closer, making the clustering in the Universe grow with time. Another model hypothesised that large explosions from an early generation of galaxies created bubbles in the Universe and formed later generation galaxies on the surfaces of these bubbles, giving rise to the bubble-like structures seen in galaxy surveys.

2. The reason behind this difference is that on the one hand, the energy density of matter is ρc^2 (ρ being the mass density and c the speed of light), and this density is inversely proportional to volume ($\rho c^2 \propto \frac{1}{R^3}$), since $\rho \propto \frac{1}{R^3}$. On the other hand, the radiation energy density is proportional to T^4 (see equation 8.21), where T is the temperature of the blackbody. This temperature depends on the extent as $T \propto \frac{1}{R}$. This means that $T^4 \propto (\frac{1}{R^3})^{4/3}$. This is why radiation energy density decreases faster than matter energy density if the volume is increased.

3. Consider a certain distribution of matter in space, with density $\rho(\mathbf{r})$ varying with position $\mathbf{r}$ (which is a vector). Suppose the mean density (averaged over a very large region) is denoted by $\bar{\rho}$. The fractional 'overdensity' at a certain point $\mathbf{r}$ can then be defined as $\delta(\mathbf{r}) = (\rho(\mathbf{r}) - \bar{\rho})/\bar{\rho}$. Let us consider this function $\delta(\mathbf{r})$.

   One can decompose this function which varies with position into a linear combination of plane waves $\sin kr$, where k is the wave number ($= 2\pi/\lambda$ for waves with wavelength λ). This is called a Fourier decomposition. The mean square of the amplitude of these waves for a certain wave number k is called the power spectrum, $P(k)$. Mathematically, the Fourier transformation of δ, and its inverse, are given by (for a certain volume V),

$$\delta = \frac{V}{(2\pi)^3} \int \delta_k e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k \tag{8.28}$$

The power spectrum is defined as $P(k) \equiv |\delta_k|^2$.

The interested reader can consult specialised books to find out more about this. For example, *Cosmological Physics* by John Peacock (Cambridge University Press, 1999), Principles of Physical Cosmology by P. J. E. Peebles (Princeton University Press, 1993), *Structure Formation in the Universe* by T. Padmanabhan (Cambridge University Press, 1993).

4. Such candidates for dark matter are called 'hot dark matter'. These particles moved with relativistic speed when they decoupled from the rest of the matter. Candidate particles which decoupled with speeds far less than that of light are termed 'cold dark matter'.

5. Mathematically, the Harrison–Zel'dovich power spectrum is $P(k) \propto k$. It can be shown that the fluctuations in gravitational potential energy in this case is independent of the length scale. In the absence of any characteristic length scale of density fluctuations, this is thought to be the simplest type of fluctuation that is possible, and has been used as the initial power spectrum of density fluctuation. It can also be shown that density fluctuations averaged over a length scale L, in this case, will be proportional to $\delta \propto L^{-2}$, which means that the Universe looks uniform when averaged at very large length scales.

6. One can roughly estimate the redshift of the equality epoch and the corresponding age of the Universe in the following way. For a blackbody at temperature T, the radiation energy density is given as $7.5 \times 10^{-16} \times T^4$ joule per cubic metre (see equation (8.21)). Since the present day temperature of the background radiation is approximately 2.73 kelvin, the corresponding energy density is approximately 4.2×10^{-14} joule. Naturally the radiation energy density in the past was higher. We have seen that the temperature of the radiation has decreased with the expansion of the Universe. This means that at the redshift epoch of z, when the size of the Universe was $\frac{1}{(1+z)}$ of its present size, the temperature of the radiation was $1+z$ times that at present. That is, the radiation energy density at this epoch was $(1+z)^4$ times that at present. In other words, if the redshift corresponding to the equality epoch is denoted as z_{eq}, then the radiation energy density then was $4.2 \times 10^{-14}(1+z_{eq})^4$ joule per cubic metre.

   We should compare this with matter energy density. Suppose the matter density in the Universe is a fraction Ω_m of the critical density (Ω_m is a number less than or equal to unity). We have earlier found that the critical density is approximately 10^{-26} kg per cubic metre (if the present day value of the Hubble constant is 70 km per second per megaparsec). The present day matter density is given by $\Omega_m \times 10^{-26}$

kg per cubic metre. Suppose also that the matter at present does not have much kinetic energy from its motion, which seems to be true from observations. Using the relation between mass and energy, that the rest energy corresponding to mass m is mc^2, one can then say that the matter energy density per cubic metre is approximately $10^{-9}\Omega_m$ joule. This clearly shows that matter contributes much more to the energy density of the Universe than radiation. Now, the matter density in the past was larger than at present. For example, at the redshift epoch of z, when the size of the Universe was $\frac{1}{(1+z)}$ of its present size, the volume was $\frac{1}{(1+z)^3}$ of its present value. This means that the matter energy density was $(1+z)^3$ times that today. At the equality epoch, therefore, the matter energy density was $10^{-9}\Omega_m(1+z_{eq})^3$ joule per cubic metre.

Equating these two energy densities—of matter and radiation—one finds the redshift of the equality epoch to be $z_{eq} \sim 25000\,\Omega_m$ (for the assumed value of the Hubble constant of 70 km per second per megaparsec). Using the precise numbers, one has $z_{eq} \sim 12000\,\Omega_m$. If the value of Ω_m is 0.3, then z_{eq} is approximately 4000. We have here neglected the fact that neutrino-like particles would also contribute to the radiation energy density.

If we now use the relation between age and temperature (equation 8.18) in a radiation-dominated Universe, we can find the age of the Universe at the equality epoch. Since the temperature at the equality epoch will be $T \sim 2.73(1+z_{eq})$, one can write the age as

$$t_{eq} = \frac{(1.5 \times 10^{10})^2}{(4000 \times 2.73)^2}\text{s} \approx 5 \times 10^4\,\text{yr} \tag{8.27}$$

8. Although a detailed calculation is very complicated, one can very crudely estimate the amount of anisotropy. One knows from general relativity that the change of energy of a photon depends on the strength of the gravitational field. This means that the variation of temperature would depend on the variation of the gravitational field. Now, one measure of the variation of this field is the average speed of matter arising from it. Galaxies have peculiar speeds, in addition to the speed of expansion from one another, because of gravitational attraction. We have seen how our Milky Way is moving toward the Virgo cluster of galaxies because of the inhomogeneity of the gravitational field in that direction over other directions. We had also found that this speed is approximately 0.2% of the speed of light, or two parts in a thousand. According to relativity, the variation of the gravitational field responsible for this speed will change the energy of photons by a fraction $(0.002)^2$, or 4 parts in a million. This is a crude estimate of the temperature variation in the background radiation due to clustering of matter.

8. The following website of Max Tegmark has a compilation of results from all experiments on the anisotropy of background radiation: `http://www.hep.upenn.edu/~ max/cmb/experiments.html`
9. For details of the *WMAP* satellite and results from it, see :`http://map.gsfc.nasa.gov`
10. For more on this mission see : `http://sci.esa.int/planck`
11. A particular observation done with the help of the Hubble Space Telescope is being referred to here. See below for more discussion on the Hubble Deep Field observations.

Chapter 6 : First Luminous Objects

1. To understand this limiting mass, let us consider a spherical object of radius R and mass density ρ. Its volume is therefore $(4\pi/3)R^3$ and the total mass $M = (4\pi/3)R^3\,\rho$. Suppose this object is contracting under its gravity, and suppose it takes time t_g to shrink to the centre. Since the speed of contraction at its surface is $v = (GM/R)^{1/2}$, this time for contraction is roughly $t_g \sim R/v \sim (G\rho)^{-1/2}$.

   We should compare this time taken scale with the time taken to lose energy through radiation. Suppose we denote this radiation time scale by t_c. If the temperature of the gas T is less than approximately 3×10^5 kelvin, then one can determine this time scale as $t_c \sim 5\times10^{-16}T^{3/2}\rho^{-1}$ second (where the mass density of the gas is ρ). The temperature of the gas will however depend on the gravitational field inside the object: $T \sim \frac{GMm_p}{k_B R}$, where m_p is the mass of a proton and k_B is a constant whose value is 1.38×10^{-23} joule per degree kelvin (see the next *Note* below).

   If one now demands that $t_c < t_g$, then one easily obtains that the mass should be less than a limiting value of $M < 5\times10^{11}\,M_\odot$. We have here assumed that the total mass of the object is normal and there is no dark matter. If it is assumed that only ten per cent of the total matter is contributed by gas, then the above limiting mass changes to $M < 1.5\times10^{12}\,M_\odot$.

2. One can estimate this temperature using the 'virial theorem'. According to this theorem, the time-averaged kinetic energy (KE) and the gravitational potential energy (PE) of a system of particles in equilibrium are related as (−2 KE) = PE. We also know that the temperature of an ensemble of particles in (thermal) equilibrium is related to the average speed of random motion of particles. If we consider particles in an ionised gas, most of the mass comes from protons than electrons, and the kinetic energy of protons will be

relevant here. Using these two arguments, we can estimate the temperature of an ensemble of particles that is bound by its gravity and is in equilibrium to be $T \sim GMm_p/Rk_B$, where M is the mass and R is the radius of the spherical object. Here m_p is the mass of a proton and k_B is a constant whose value is 1.38×10^{-23} joule per degree kelvin. Using a mass of $M = 10^7$ solar mass, i.e., $M = 2 \times 10^{37}$ kg, and $R = 1$ thousand parsec (the typical size of a galaxy with this mass), one estimates the temperature to be few thousand kelvin.

Careful readers may have noticed that this relation between temperature and mass is similar to the estimate for temperature at the cores of stars as described in the first chapter. The similarity is not a coincidence.

3. The scientific name of this radiation is Bremsstrahlung. When the electrostatic attraction between an electron and a relatively slow moving ion decelerates, or puts a break on the motion of the electron, it radiates away some energy. '*Brems*' means 'break' and '*strahlung*' means radiation in German, from which this term has been coined.

4. Study of dust particles in space constitutes an important part of astrophysics. The chemical and physical structure of dust grains and their effect on the matter in space are main topics of research. For example, the effect of dust grains on the temperature of the surrounding gas is often important. Diffuse gas in our Milky Way, for example, is hot near star forming regions, and often dust grains are the reasons behind its high temperature. Ultraviolet photons from heavy stars collide with these dust grains and eject electrons with high energy (by photoelectric effect). These ejected electrons heat up the surrounding gas.

5. The details can be found in the paper by Max Tegmark and his colleagues which appeared in the *Astrophysical Journal* (1997), vol. 474, p.1.

6. One can find more information on the results obtained by *CHANDRA* at the following website: `http://chandra.nasa.gov`

7. According to new observations by John Magorrian and his colleagues, the mass of the central black hole is proportional to v^n, where v is the average speed of the stars in the galaxy and n lies between 4 and 5.

8. Current estimates of the amount of matter in the intergalactic medium are approximately 5% of the critical density of the Universe.

9. Although this is commonly known as the 'Gunn–Peterson' test, there were a few other astrophysicists who independently discussed this at

around the same time, including Peter Scheuer of Cambridge, UK, and Igor Sklovsky of the then USSR.

10. This survey aims to determine the redshift of a million galaxies in one-quarter of the sky. In the process, it will have determined the positions and brightnesses of more than 100 million objects in the sky. For more information on this survey, look at : `http://www.sdss.org`

11. For more information on this experiment, look at : `http://www.alma.nrao.edu`

12. For more information on this instrument, look at : `http://www.ncra.tifr.res.in`

13. For more information on these two planned missions, look at : Spitzer Space Telescope `http://www.spitzer.caltech.edu`, James Webb Space Telescope `http://www.jwst.nasa.gov`

Chapter 7 : Evolution of Galaxies

1. There are also galaxies in the Universe which are neither spiral nor elliptical. For example, the small and large Magellanic clouds near our Milky Way mentioned in the first chapter are neither spiral nor elliptical in shape. These galaxies without any regular shape are called irregular galaxies. There also exist spiral galaxies which have a rod-like structure near their centre, and are called 'barred spiral galaxies'. It is believed that these 'bars' arise from interactions between galaxies.

2. If a particle of mass m rotates at a distance of r around an axis with speed v, the magnitude of its angular momentum is $L = mrv$. Therefore, it is not only the speed and mass, but also the extent of the body which is important for angular momentum. If for some reason the extent of a body is shrunk toward the axis of rotation, i.e., if r is decreased, then v will increase to keep the total angular momentum a constant.

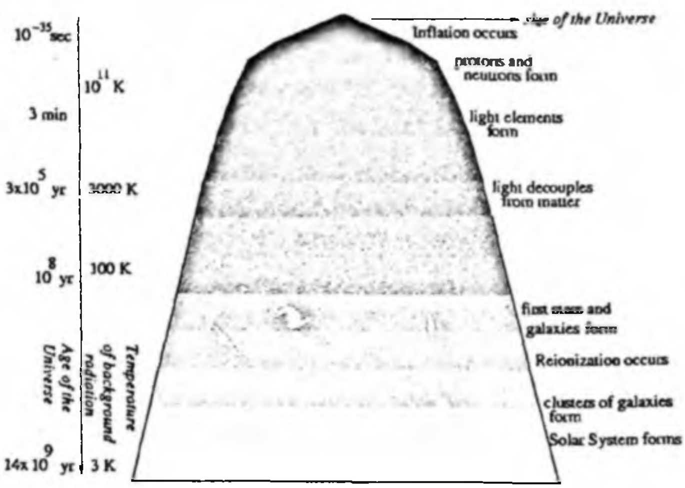
3. Astronomers define the star formation rate in a galaxy by the mass that is being used to form stars in a given amount of time. The typical formation rate of stars in a normal galaxy like the Milky Way is of the order of a few solar masses per year. Observations of starburst galaxies often suggest star formation rates of the order of a hundred solar mass per year.

4. This observation and the resulting photograph was announced to the astronomical community in a novel way. Usually when a new discovery is made, one needs to wait for a scientific paper to appear in a journal to be able to find the details of the observation

and the data. This certainly takes a long time, and usually other scientists, who are not involved in the observation, cannot make further studies for a while until details are available. The scientists associated with the Hubble project took a revolutionary decision in this case. They made the complete data of this observation available to public through internet within a span of fifteen days. The effect of this decision was also spectacular. There was a flurry of research papers from around the world within months of this announcement. For more information on the Hubble Deep Field, see `http://www.stsci.edu/ftp/science/hdf/hdf.html`

5. One may argue that uniformity of distribution can also mean that the sources are *very* close to us, perhaps on the edge of the solar system. This possibility has been examined in detail and proved wrong.

The Time-line of the Universe



Index

www.universitiespress.com

Since time immemorial man has asked the question, "When was the Universe first lit up?" Modern cosmology is now poised to answer this question. The purpose of the book is to explain in simple language the recent discoveries and future trends in cosmology. Although written mainly for the general reader, it also has detailed mathematical notes for the more curious reader.

It tells the story of the emergence of the first luminous objects in the Universe. Beginning with an imaginary journey across the Universe which gives the reader an idea about our place in the Universe and the enormous scales involved, the author provides a basic sketch of the theory of relativity, the expansion of the Universe and major events in its history. The book then goes on to describe the formation and evolution of the Universe.

Biman B Nath received his PhD in astronomy from the University of Maryland, USA, in 1992, and is currently at the Raman Research Institute in Bangalore. Apart from his professional research, he is interested in the popularisation of science, and his articles have appeared in many national newspapers.

Cover illustration: Biman B Nath
Cover design: OSDATA, Hyderabad

Rs 125.00

The Educational Monographs published by Universities Press in collaboration with JNCASR address the needs of students and the teaching and research community

ISBN 81 7371 520 3
9788173715204

Universities Press

Biman B Nath: *The Dawn of the Universe*