The Wondrous Universe

Creation without Creator?

Bearbeitet von
Gerhard Börner

1. Auflage 2011. Taschenbuch. VII, 184 S. Paperback
ISBN 978 3 642 20103 5
Format (B x L): 15,5 x 23,5 cm
Gewicht: 299 g

2. The World at Large: From the Big Bang to Black Holes

First of all we want to look in detail at some insights of modern cosmology and physics, since we want to lay a solid foundation of facts for the scientific view of the world and not just tell a kind of fairy tale.

2.1 Immediate Experiences: A Play of Thoughts

Let us travel in our imagination away from our home on the Earth, even away from the Earth into outer space. As we move away farther and farther, we find that the familiar outlines of houses and streets become more and more vague. From a height of 10 km we see a colorful map, and from 100 km away the circular edge of the terrestrial sphere comes into view. Oceans, continents, and many clouds dominate the picture. From a distance of 100,000 km we see the Earth floating like a blue sphere in the black sky of outer space (see Fig. 2.1).

Our imagined journey then takes us past the Moon. We reach the neighboring planet Mars, and move on past Jupiter and Saturn.

Looking back we see the Sun as a fiery ball surrounded by its planets orbiting in a plane (Fig. 2.2).

The Sun is a star, i.e., it shines from its own power, whereas the planets just reflect the sunlight. Like all stars it is a gigantic fusion reactor producing energy by the fusion of atomic nuclei in its interior, and radiating light and heat away from its hot surface. From a distance of 10,000 billion kilometers (10^{13} km), our solar

G. Börner, *The Wondrous Universe*, Astronomers' Universe,
DOI 10.1007/978-3-642-20104-2_2,

Fig. 2.1 The Earth, as seen by the crews of the Apollo flights, floats like a splendid, colorful sphere in space (*courtesy of NASA, Apollo 17*)

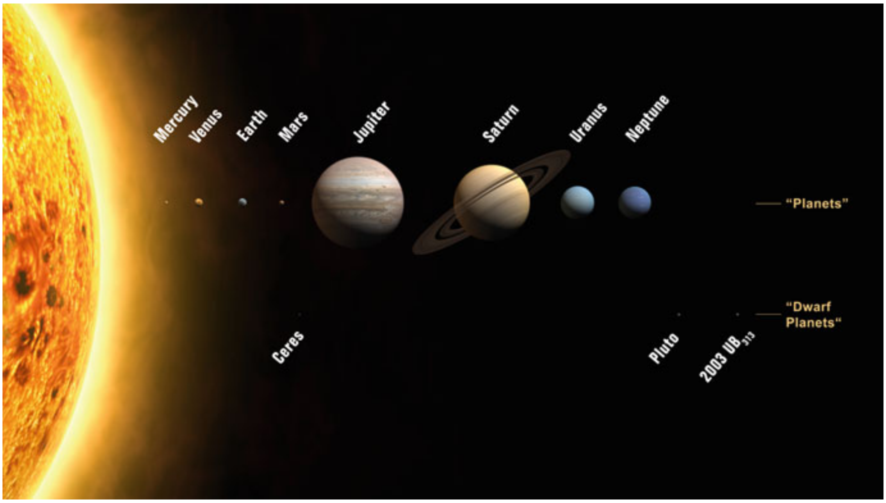

Fig. 2.2 This picture shows the Sun, and from left to right the planets. Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Pluto in approximately the correct ratio of their sizes (*courtesy of IAU*)

system appears lonesome and somehow lost in the gigantic largeness of space. The distances have grown so big now that to state them in terms of kilometers would lead to large and impractical numbers. Therefore we choose a new scale, the light-travel time:

We measure distances by the time it takes for light to traverse them. The velocity of light is about 300,000 km s^{-1}. From the Moon to the Earth (384,000 km) the light needs a little more than a second (1.28 s), from the Sun to the Earth about 8 min. Thus we say that the Earth–Sun distance is 8 light-minutes.

We have traveled already 10,000 billion kilometers, i.e., one light-year. Now we approach the nearest neighboring star, and when we continue our flight, we meet stars like our Sun again and again, seemingly without end.

Nothing can move faster than light, but in our imagination we can, of course, exceed that speed limit. After 100,000 light-years we have apparently passed the assembly of stars, because we leave a stellar system behind which resembles a flat disk with a central bulge. It contains about 100 billion stars together with gas and diffusely spread out solid matter called "dust" by astronomers. This is our Milky Way, the "Galaxy."

Extragalactic space seems to be empty, but in the distance we can discern a big stellar system next to us – Andromeda, a galaxy at a distance of two million light-years. It appears similar to our galaxy with spiral-shaped extensions. In our faster-than-light travel we meet such galaxies again and again. They seem to fill all of space. But, when we come to a standstill, we find that all those galaxies rush away from each other in rapid flight, like the fragments of a huge explosion. In addition we observe radiation signals from very large distances, signals from a gradually burning out fireball. We cannot see any further, the origin of the cosmic flight remains hidden to us.

This cosmic system in rapid expansion presents an amazing sight, much more complex than the peaceful, uniform distribution of stars in our neighborhood suggests. How can we obtain an intelligible picture of these conditions in space and time?

Deep in thought we return to the Earth, to the starting point of our voyage, and look at the table at which we sit. A solid piece of furniture, no doubt, supporting us reliably. But the solid, brightly polished surface is only seemingly so. As soon as we begin in our imagination a journey to smaller and smaller separations, the polished surface at first turns out to be a rough landscape of valleys and steeply rising peaks. When we penetrate to dimensions of a hundred millionth of a centimeter, we are surrounded by the electron shells of the atoms, orbiting and oscillating electric charges which are arranged in various regular patterns. The electrons dance around a nucleus which carries practically the whole mass of the atom concentrated in a tiny volume which extends only to one part in a hundred thousand of the scale of the electron shell. (These big numerical factors are written by the physicists in powers of ten: one hundred thousand is 10^5, and one hundred thousandth, or one part in one hundred thousand is 10^{-5}.)

Nothing is there between shell and nucleus – just empty space. The table shows itself as not a very solid object. It is a porous, almost empty thing, but it appears solid to us, because from the point of view of an atomic nucleus we are also a porous structure. We also consist of electrons and atomic nuclei, just like the table. The tiny nuclei of the atoms again are composed of protons and neutrons, the building blocks of matter. At this level all inanimate and living things in the world are equal: an assembly of protons, neutrons, and electrons.

Different forms and shapes are created from these identical small building blocks by arranging them in various ways according to the laws of physics.

To be sure, protons and neutrons are not elementary particles yet. If we imagine a look inside those nuclear particles, we find empty space, and point-like particles, the quarks. Both neutron and proton contain three quarks. The electrons, like the quarks, are indivisible point-like particles. Inside the nuclear particles we cannot look at the world as easily as in our normal human environment, where there are tables, houses, cats, human beings, and much more. The elementary particles do not keep such a well-defined identity, they become much more vague, merge in some sense with the forces acting upon them, and can no longer be discerned as tiny objects in space. It is difficult to write about impressions on the way to even smaller dimensions, because we do not have adequate conceptions for that in the classical world. We can still regard the quarks as a form of material objects, but if we try to probe more deeply, we find that the material properties fade away. The seemingly point-like particles are actually concentrated packets of energy produced by the vibrations of a diminutive string. To be sure, we are also gradually losing any orientation in space and time, as we venture into this regime of dimensions less than 10^{-33} cm. It seems that even space and time perish in the ups and downs of string vibrations.

Deeply impressed by this vision we return to our reliable environment of solid objects.

Does this view of the world on its largest and smallest dimensions truly describe reality? First of all it is nothing but pictures and mathematical constructions invented by us to help us grasp the world around us. Of course, these images are coined by our senses, our reason and mind, i.e., by our brain which has been shaped during a long biological evolution, and is thus dependent on the world too. All scientific insights, and our daily experiences as well, are filtered by these conditions of our sensory equipment. Nevertheless, it looks as if a reality existed independently of ourselves, as if we could succeed in unraveling its properties step by step, even demonstrate and explain its counter-intuitive aspects.

At this point let us leave these preliminary remarks, and turn in detail to some of the things we have seen on our excursion.

2.2 Cosmology

The big-bang model of the universe can be comprehended in illustrative pictures, but the description of the path from astronomical observations and theoretical considerations to that model requires a discussion of many astronomical and physical details. We have to inspect the wealth of detailed results which can be combined to yield the present-day view of the cosmos.

If we want to understand how well the standard model is established, we have to consider stellar evolution, the spectral analysis of the light received from distant galaxies, the properties of the cosmic microwave radiation, and some fundamental features of Albert Einstein's theory of general relativity.

2.2.1 The Darkness of the Night Sky

There are a few easy cosmological observations which neither require expensive telescopes nor satellites. The cheapest entry to cosmology is right above you, when you stand in front of your doorstep at night and look at the stars in the sky. Why is the sky between the stars dark?

If the stars were distributed uniformly in space, were shining forever without change, then there would be no gap between the stars. In every direction you would see a star – some close, some far away. The night sky would be everywhere as bright as the surface of a star. The night sky is dark, however, and therefore this assumption about the stars cannot be correct.

Johannes Kepler in 1610 had already noticed that the darkness of the night sky contradicted some older ideas about the structure of the world, especially the view of Giordano Bruno, who held that the cosmos was infinite and unchanging. Later on Kepler's arguments were repeated several times, for instance in 1823 by the physician and astronomer Heinrich Wilhelm Olbers of Bremen. They are named after him "Olbers' Paradox," although they are not paradox, and were not invented by Olbers. Interestingly enough, even at the beginning of the twentieth century most astronomers believed in a static world. This was the motivation for Albert Einstein to look for a uniform, static cosmological model as a solution of his theory.

Today we know that the world is not static, that all stars came into being a finite time ago, and that they will all perish. Therefore there is only a minimal chance to find a star in any direction, and the night sky appears dark.

Besides the light from the stars we also receive radiation from the hot plasma of the early universe which surrounds us in huge distances like a giant hollow sphere. The cosmic expansion stretches all wavelengths, and shifts this radiation out of the visible range into the microwave region of the electromagnetic spectrum. Clearly, it deserves its name "cosmic microwave background (CMB)," and it surely does not disturb the darkness of the night sky. Thus this everyday, or rather every-night, commonplace fact of the darkness of the night sky tells us that the world is expanding, and that the stars have arisen a finite time ago.

2.2.2 The Life-Cycle of the Stars

In dark nights away from city lights we can see the bright band of the Milky Way stretching across the sky – billions of stars which

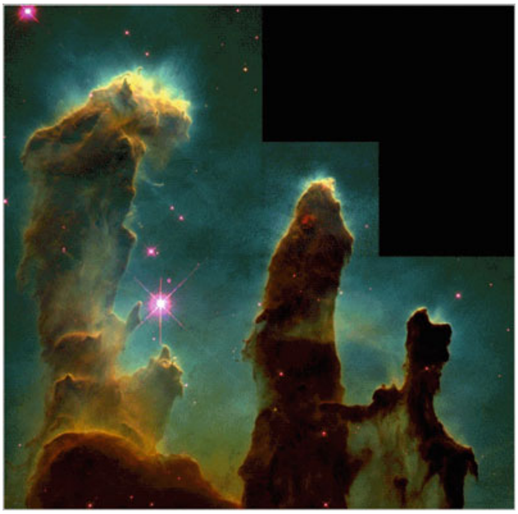

Fig. 2.3 Obtained with the Hubble Space Telescope (HST) this picture of a region in the constellation "Eagle" (Aquila) shows structures where new stars are forming. Columns of cold hydrogen gas and dust are illuminated by the UV radiation of the newly formed stars. A color composition is chosen such that an optically appealing impression results (*courtesy of the Space Telescope Science Institute (STScI)*)

send out their energy like the Sun. In reality we can, just using our eyes – the "unarmed" eye as militant astronomers love to put it – see only about a thousand of the nearest stars.

Our Sun is a very typical star. It was born in the condensation of an interstellar gas cloud. The initial clump of gas contracted more and more under the action of its own gravity, until finally the center became hot and dense enough to start the fusion of hydrogen into helium (Fig. 2.3).

In this reaction four atomic nuclei of hydrogen, i.e., four positively charged nuclear particles (four protons), merge to form one atomic nucleus of helium which consists of two protons and two neutrons. The mass of the four individual nuclear particles adds up to a mass larger than that of the helium nucleus by 0.7%. According to Einstein's famous equation ($E = mc^2$: energy is equal to mass times velocity of light squared) this mass deficit is translated into energy by the fusion reaction. The fusion energy per gram of matter is about a million times larger than the energy set free in a chemical reaction like a combustion of fuel or an explosion. Only the explosion of an hydrogen bomb displays the dramatic example of fusion energy suddenly set free.

In the interior of the Sun the hydrogen fusion runs as a slow, controlled process. It has made the Sun shine for 4.5 billion years. In about another 5 billion years the hydrogen at the Sun's core will have been used up. Nuclear fusion in the center will stop when about 12% of the hydrogen supply are consumed. Then hydrogen starts to burn in a spherical shell around the central region. The Sun tries to establish a new equilibrium and blows up its outer layers enormously – it turns into a "red giant." When the Sun reaches this stage, its outer layers will extend to the Earth's orbit. The Sun, like any other similar star, will exist only for a relatively short time, only about 500 million years, in the red giant stage. After that, helium burning starts in the center followed by further short-lived fusion reactions. Eventually the outer shell of about one quarter of the mass is expelled. The remaining core shrinks to a very dense object of about the Earth's dimension – to a "white dwarf." As a white dwarf the Sun will shine with a bluish light scarcely brighter than the full Moon on the burnt-out Earth. This evolutionary history can be accurately predicted, because it simply follows from the laws of physics which govern the nuclear reactions inside the Sun. But there is no reason to panic – the Sun will exist as a quietly shining star for a substantial amount of time. Mankind has just started its evolution as an intelligent life-form. If they do not perish prematurely, our offspring will advance, during the billions of years of their future, far beyond Earth to distant solar systems. Even if now the Earth was the only planet carrying life, there would be sufficient time for life to spread out over all of the Milky Way, and even to other galaxies.

The evolutionary path of other stars can deviate considerably from that of the Sun.

Stars with larger mass produce more energy in their interior, are more luminous, and remain for a much shorter time span in the phase of hydrogen burning. A star which is about 10 times more massive than the Sun enters the red giant stage already after 10 million years. A smaller star with about 10% of the Sun's mass uses its nuclear fuel very economically, shines quite faintly (only at about one thousandth of the solar luminosity), and exists for about 10,000 billion years.

The final stage of massive stars is quite dramatic: Since the core of a massive star contains too much mass for a stable white dwarf, it will collapse further driven by the relentless pull of gravity, and end up at a much smaller radius, and a much higher density. A neutron star (an extreme object resembling a gigantic, solar mass atomic nucleus with a radius of 10 km) or a black hole may be the final state. In a certain mass range the core may even be disrupted completely. In any case a huge eruption is triggered, the so-called supernova explosion, which hurls the outer parts of the star into the surrounding space. A supernova shines for some time with extreme brightness, often surpassing the whole host galaxy in luminosity. The remnants in many cases radiate actively as pulsars or X-ray sources (Fig. 2.4).

Our Milky Way is populated by all these different types of stars: Blue, very bright, massive, and short-lived stars which are formed again and again from gas and dust, many stars like our Sun, red stars of small mass which are long-lived and faint, luminous red giants, white dwarfs, pulsars, and X-ray sources, all are contained in this huge stellar system.

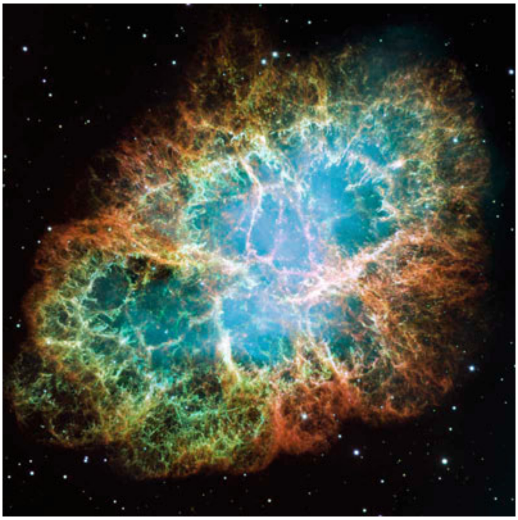

Fig. 2.4 The Crab nebula shown on this HST image is the remnant of a supernova observed in 1054 AD by Chinese astronomers. A "pulsar" at the center of this nebula emits periodic radio signals with a period of 33 ms. This pulsar is a neutron star which rotates 30 times per second around its axis (*courtesy of the STScI*)

Every 100 years or so a supernova explodes in a typical galaxy. The last time such an event could be seen in our Milky Way was in 1604 (Kepler's supernova). In the year 1987 a spectacular supernova could be observed in the Large Magellanic Cloud, a small stellar system close to the Milky Way at a distance of only about 180,000 light-years. For astronomers supernovae are naturally of great interest, but they are important for the whole of mankind too:

The heavy elements, present also in the human body, like carbon, oxygen, silicium, iron, etc., have all been formed in the interiors of massive stars, and have been distributed in space during the explosion of these stars. In that sense we are children of the supernovae. Only the lightest elements hydrogen and helium were created in the early universe, all heavier elements were brewed in stars.

2.2.3 The Galaxies

We can recognize many more stars with a telescope, and with it we also see that besides the Milky Way there are many fuzzy luminous spots in the sky, which turn out to be stellar systems like our Milky Way, "galaxies" as they are named.

Galaxies appear in a great variety of shapes (Figs. 2.5– 2.8):

Systems with spiral arms of similar size as our own galaxy (such as M31, the galaxy named "Andromeda" at a distance of two million light-years) or elliptical galaxies without spiral arms resembling an elliptical or nearly circular small disc are frequent. Elliptical galaxies can be very massive (some contain 10^{13} solar masses, a hundred times the mass of the Milky Way), but there exist also very small dwarf galaxies of similar appearance, but with a mass of only a few million solar masses. Light needs about 100,000 years to cross the Milky Way or the Andromeda galaxy, and the separation between galaxies is typically ten times larger, about a million light-years.

The huge distances between galaxies have an interesting consequence: The farther away a galaxy is, the earlier in its history we can see it. When we observe Andromeda, we do not see what happens there right now but we see what has happened 2 million years ago. It is thus impossible for us to observe the universe

Fig. 2.5 The Andromeda galaxy is the nearest large spiral galaxy, a neighbor of the Milky Way at a distance of 2 million light-years. Viewed from far away our own Milky Way would probably look quite similar. This image obtained from data of NASA's satellite WISE (Wide field Infrared Survey Explorer) shows Andromeda in infrared light at different wavelengths (from 4 μm (*blue*) to 22 μm (*red*)). Mature stars show up in *blue*, *yellow*, and *red* colors indicate regions where dust is heated by newborn, massive stars (*courtesy of NASA*)

Fig. 2.6 The spiral galaxy NGC4622 rotates clockwise – from the image one might draw the opposite conclusion (*courtesy of STScI*)

Fig. 2.7 The galaxy cluster A2218 is a gravitational lens. It deforms the images of distant galaxies to elliptical shapes (*courtesy of STScI*)

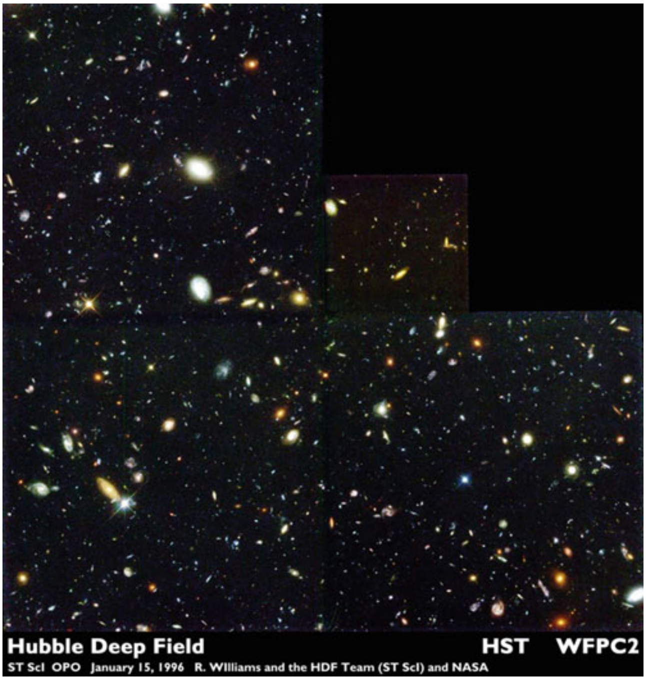

Fig. 2.8 This image of a small area of the sky (about 1 arcmin × 1 arcmin wide), the so-called Hubble Deep Field, is the deepest look into the universe in optical light reached so far. The picture does not only belong to the treasures of astronomy, but it is also a treasure of mankind. The small area in the constellation Ursa Major contains about 2,500 galaxies of all types – disclike elliptical, spiral, and irregular galaxies. The HST was pointed at this location for 10 consecutive days (*courtesy of STScI*)

in its present state. Astronomers investigate the world similar to archaeologists, who dig into deeper and deeper layers as well as earlier and earlier times. The advantage is that one can look directly at the evolution in time, the disadvantage is, of course, that one can never see the whole at one particular instant of time.

These aspects are due to the finite velocity of light, and they remain valid for small distances. But their consequences in that case are insignificant: Light takes 8 min from the Sun to the Earth, but in 8 min nothing much changes in the solar system even though the Bavarian Broadcasting Corporation (BBC) claims that "in 15 minutes the world may change."

Figure 2.8 shows a picture of many galaxies taken by the Hubble Space Telescope (HST), about 2,500 in a narrow region of the sky. Extrapolating to the whole sky astronomers can safely estimate that the volume of space accessible to observation contains about ten billion galaxies. Each individual galaxy with its billions of stars is in itself an interesting system, but in cosmology it is regarded as a test particle useful for exhibiting some, perhaps really existing, global properties.

2.2.4 The Expansion of the World, and the Cosmic Microwave Background

The modern view of the cosmos derives from insights into a fundamental property of the galaxies. In the 1920s the American astronomer Edwin P. Hubble found that the spectral lines of atoms measured in galaxies do not coincide with those measured in laboratories on Earth. Instead almost all galaxies (with the exception of Andromeda, and some small companions of the Milky Way) exhibit a shift of their spectral lines toward longer wavelengths (toward the red end of the spectrum) by a factor $(1 + z)$. For each galaxy this factor is a characteristic quantity, all its spectral lines are stretched in wavelength by the same factor. z itself is called the "redshift" of the galaxy.

The redshift z increases with the distance of the galaxy.

The Doppler effect explains this phenomenon quite naturally: For a source of light moving away from us, the wavelength of the signal received is longer than the emitted one. We may conclude that almost all galaxies are moving away from us, the farther away they are the faster they move. In fact, the velocity is proportional to the distance according to the famous relation discovered by Edwin Hubble

$$cz = v = H_0 d.$$

The quantity H_0 was named "Hubble Constant" to honor Edwin Hubble's discovery.

In this equation which describes the increase of the velocity $v = cz$ with distance d, we also find c – the velocity of light.

To measure the cosmic expansion, i.e., the flight of the galaxies, you need to measure the distance d and the redshift z of only one galaxy, at least in principle. In practice you meet a few difficulties: The astronomers know precise distances only to relatively close galaxies, and those have proper motions – induced by local mass concentrations – superimposed on the cosmic expansion motion.

Andromeda's proper motion even dominates over its cosmic motion. It approaches the Milky Way, and its spectral lines are therefore blue-shifted.

Hubble, and many astronomers after him, have used pulsating stars for cosmic distance measurements. "Cepheids" (named after the star δ-Cephei) change their brightness rhythmically; they pulsate with periods of hours to days. Slow pulsation signifies high luminosity, and two stars with the same pulsation period have the same luminosity. Therefore the measurement of the pulsation period and the brightness of a Cepheid is sufficient to determine its distance, if a few stars are known with precisely determined distances to calibrate the relation between pulsation period and luminosity. Cepheids can provide very precise distance determinations, but unfortunately for many years Cepheid stars could not be measured at the cosmic distances, where the expansion velocity dominates over local, peculiar velocities.

It was expected that the situation would improve decisively, when new telescopes, especially the space telescope "Hubble," would extend the classical Cepheid method out to a distance of 20 Mpc (the unit Mpc – "Megaparsec" – is about 3.26 million light-years). This is the distance to the center of the Virgo cluster of galaxies. At the edge of this huge system of thousands of galaxies lies the Milky Way. Unfortunately the Virgo cluster has shown itself as a relatively complex structure, where the center of mass is difficult to determine. Thus a spread-out range of values for the Hubble constant had to be accepted, namely

$$H_0 = 80 \pm 22,$$

in units of velocity (given in kilometers per second) per megaparsec. These units are favored by astronomers, and they have an easy interpretation: At 1 Mpc distance a galaxy recedes with a velocity of 80 km/s.

In view of the uncertainties involved in the Virgo distance, a new method has been developed which allows us to reach out to far greater distances without intermediate steps.

This approach makes use of the high luminosity of certain types of stellar explosions, the supernovae of type Ia (SNIa). Their spectra do not contain lines of hydrogen, only higher elements such as helium or carbon are present. Stars which end their existence as SNIa evidently have gone through a long time of evolution. They have burnt their hydrogen supply and the stellar material is essentially carbon and oxygen. Very probably these are white dwarfs, compact stars with a radius like the Earth and a mass like the Sun. The luminosity of such a supernova increases rapidly, reaches a maximum within a few days, and then decreases.

The explosion produces radioactive nickel (^{56}Ni) which decays via cobalt (^{56}Co) to iron (^{56}Fe), and thereby supplies the energy of the luminous phenomenon. According to theory the optical luminosity of a SNIa is due to the thermalization of high-energy gamma rays produced during the decay of nickel and cobalt.

SNIa are very bright. They can be observed at great distances far beyond Virgo. In addition they are good distance indicators, although they do not all have the same peak luminosity. One must expect some variation, because the luminosity depends on the amount of nickel produced, and this can vary according to the conditions in the star when it explodes. There is, however, a very helpful property: The observers found that there is a strong correlation between the maximum luminosity and the shape of the supernova light curve, especially the decline of the brightness. Rapidly decaying light curves belong to less luminous supernovae, while slowly decaying ones are more luminous at maximum. This empirical relation can be quantitatively fixed, and thus the maximum luminosity can be calibrated making supernovae Ia a precise indicator for cosmic distances.

During the past 12 years astronomers have succeeded to detect very distant type Ia supernovae systematically, and to measure the rapid increase as well as the decline after maximum of their brightness.

The collaboration of many observing stations around the world had to be organized such that each supernova could be traced by a big telescope immediately after its detection. Two large groups of observers, the "High-z Supernova Search Team" and the "Supernova Cosmology Project," have independently pioneered this research.

Figure 2.9 displays a Hubble diagram for a large number of SNIa. The data points below a redshift of $z = 0.1$ agree very well with the linear-Hubble relation. This leads to a determination of the Hubble constant

$$H_0 = 70 \pm 10.$$

The Hubble constant given in these astronomical units defines a characteristic time by its inverse $1/H_0$. This "expansion time" amounts to about 14 billion years with an observational uncertainty of about 10%. The expansion of the system of galaxies which we observe today has started 14 billion years ago provided the galaxies have moved with constant velocity. At that time all the galaxies we can see now must have been very close together.

The measurements of the cosmic expansion gain special importance, if we take another cosmological discovery into account:

Two scientists, Arno Penzias and Robert Wilson, working at the Bell laboratories discovered rather accidentally in the year 1964 a radiation signal while they were calibrating a special antenna for microwave transmissions. The radiation at a wavelength of 7.15 cm was apparently of cosmic origin, because typical temporal variations as they are shown by individual sources were absent. Further measurements gave evidence that the radiation with wavelength between 1 mm and 10 cm arrives from all directions with nearly equal intensity, and that its spectral distribution

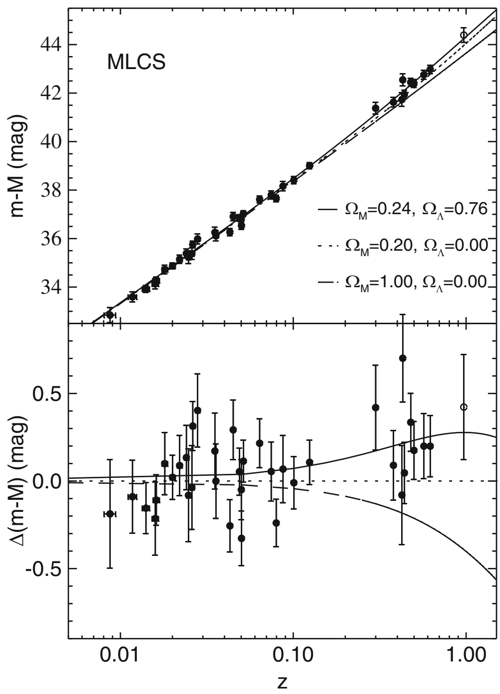

Fig. 2.9 The Hubble diagram for supernovae Ia shows the data points as they depend on distance and redshift z. Along the vertical axis the distances are given in a logarithmic unit loved by astronomers, the so-called distance modulus. The redshift z is displayed on a logarithmic scale along the horizontal axis. In the *top panel* you can see, how well the Hubble relation holds for SNIa at small redshifts (for z less than 0.1), while at large redshifts deviations from the linear Hubble relation occur. Supernovae at $z \approx 1$ clearly lie above the straight line of the linear relation indicating that they are more distant than their redshift would tell. The astronomers regard this as evidence for an accelerating cosmic expansion. Cosmological models with a positive cosmological constant have such a property. The graphs for three different cosmological models in the *upper panel* differ significantly at large z. The *lower panel* displays these differences referred to the model without a cosmological constant. At high redshift a model with a positive cosmological constant gives the best fit (after Riess et al., 1998, Astrophys. J. **504**, 935)

follows the law found by Max Planck around 1900 for the radiation emitted by a body in thermal equilibrium with its surroundings. Penzias and Wilson have received the Nobel prize for physics a few years later, since it immediately became clear that their discovery had a great impact on our knowledge of the cosmos.

Since it obeys Planck's formula, the cosmic microwave background (CMB) can be characterized simply by a temperature. The satellite COBE (Cosmic Background Explorer) has yielded measurements of the CMB spectrum over 2 years which determine this temperature precisely as:

$$T = 2.728 \pm 0.002 Kelvin.$$

(Kelvin is a temperature scale like degrees Celsius, shifted such that zero Kelvin corresponds to the absolute zero of −273.2 degrees Celsius.)

Within the measurement errors no deviations from an ideal Planckian spectrum could be found. Thus, the CMB defines the present temperature of the universe (see Fig. 2.10).

CMB and Hubble expansion taken together point at an interesting aspect of the history of the universe: If the galaxies now

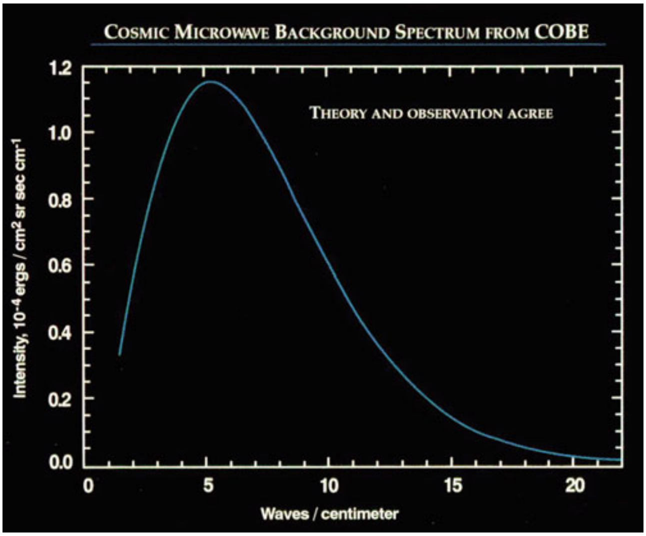

Fig. 2.10 The spectrum of the cosmic microwave radiation (CMB) as it has been registered by the satellite COBE fits perfectly the formula for thermal radiation with a temperature of 2.728 K, i.e., about 2.7 degrees above the absolute zero point of temperature. Measurement uncertainties are less than 2 mK (±0.002 K). This radiation is a natural consequence of the "hot big bang" model: It is the relic radiation of an early phase, where an almost uniform hot plasma was in thermal equilibrium with the radiation field. This cooled down because of the cosmic expansion (with permission of the COBE collaboration; Mather et al., 1990, Astrophys. J. **354**, L37; Fixsen et al., 1996, Astrophys. J. **437**, 576)

fly away from each other, they must have been closer together at earlier times. Then also the radiation must have been denser, more compressed, and hotter in the past. The conclusion seems inescapable that there has been a hot and dense early state of the universe. In the hot early universe galaxies and stars could not survive, and all that existed was a hot and dense mixture of matter and radiation.

The expansion time of 14 billion years derived from the Hubble diagram of type Ia supernovae defines the time in our past, when galaxies appeared out of the "primeval soup," and began their flight in space.

Even if this interpretation of the CMB and the general expansion sounds very plausible, we must be aware of the fact that it is not just a consequence of the observations. Theoretical conceptions are inextricably mixed into it. The universe as a whole is actually a theoretical construction, and a very special object of research, unique and unreproducible. Every physicist would be unhappy if he had to build his theories on a single experiment which could not be repeated.

But the situation is even more difficult, because we, the observers, are part of this object "universe," and living inside it we can only perceive a section limited in space and time. We assume that the part we can observe is typical for the whole – if that exists at all. But this is by no means sure. In cosmology we must work in the context of a given theory, and try to sketch a model of the cosmos using this theory, and the observations and measurement results. Only with the help of the model observations can be interpreted, and new observations and tests of the model can be suggested.

2.2.5 The Cosmological Model

The search for a simple model of the flight of the galaxies will focus on an easy mathematical representation of the uniform expansion. It seems reasonable to avoid the point of view which would put us into the center of the universe with all the galaxies moving away from us. There is no compelling cause for that, and therefore a better description would be to assume that the cosmic expansion looked the same observed from any galaxy, similar to what terrestrial astronomers observe.

Fortunately the uniform spreading out of the celestial bodies can be modeled by simple solutions of Einstein's theory of gravitation, the "theory of general relativity" (GR). Within the models the distribution of matter is taken into account only approximately, as an average matter density, and not exactly as a large system of galaxies and stars.

Let Albert Einstein himself comment on that: "The metric character (the curvature) of the four-dimensional space–time continuum is determined according to General Relativity at each point by the matter and its state in that point. The metric structure of the continuum must therefore be extremely tangled up due to the non-uniformity of the matter distribution. But if we care only about the structure on large scales, we may imagine the matter uniformly distributed over huge volumes, such that the distribution of the density becomes an enormously slowly changing function. We thus proceed similar to geographers, who approximate the Earth's surface which in small details is shaped extremely complex by an ellipsoid."

("Der metrische Charakter (Krümmung) des vierdimensionalen raumzeitlichen Kontinuums wird nach der allgemeinen Relativitätstheorie in jedem Punkt durch die daselbst befindliche Materie und deren Zustand bestimmt. Die metrische Struktur dieses Kontinuums muss daher wegen der Ungleichmäßigkeit der Verteilung der Materie notwendig eine äußerst verwickelte sein. Wenn es uns aber nur auf die Struktur im Großen ankommt, dürfen wir uns die Materie als über ungeheure Räume gleichmäßig ausgebreitet vorstellen, so dass die Verteilungsdichte eine ungeheuer langsam veränderliche Funktion wird. Wir gehen damit ähnlich vor wie etwa die Geographen, welche die im Kleinen äußerst kompliziert gestaltete Erdoberfläche durch ein Ellipsoid approximieren.")

It is favorable for model-building that the cosmic expansion does not depend on the way matter is distributed in a certain volume of space. Quite inhomogeneously condensed in galaxies and stars or uniformly spread out – it makes no difference; only the mean density, i.e., the mean mass per unit volume, counts. Therefore we can in a first approximation neglect all the structures, and regard the total mass in a certain volume of space, as e.g., in a giant sphere enclosing many galaxies, as a tenuous gas.

In fact, this gas has such a tiny density that it almost represents an ideal vacuum – only about one atom is contained within 1 m^3. Cosmologists find such a small density, when they add up the masses of the galaxies. In addition, there is strong evidence, as we shall see later, of the existence of nonluminous, so-called dark matter, and of a quite different component named dark energy. All these various types of matter and energy form a "cosmic substrate," as we might call it. This appears in the cosmological models only as a uniform density, i.e., as matter or energy averaged over large volumes of space. That is an approximation, but a very good one, as many computations have demonstrated.

A convenient simplification can be introduced, as is commonly done, by characterizing the various density components by non-dimensional numbers, i.e., by their ratio to a reference density which can be constructed from the gravitational constant G, and the Hubble constant H_0. Both quantities can be combined such that a term with the dimension of a mass density (grams per cubic centimeter) results:

$$\rho_c \equiv \frac{3H_0^2}{8\pi G}.$$

This reference density ρ_c is often called "critical density." Inserting the measured value of H_0 one finds that this critical density corresponds to a matter content of about ten hydrogen atoms per cubic meter. This is an excellent "vacuum" not yet achieved so far in terrestrial laboratories.

Following these approximations cosmologists use simple cosmological models, so-called Friedmann–Lemaître models (FL models for short; named after Alexander Friedmann (1922) and Georges Lemaître (1927), who were the first to derive and interpret these special solutions of Einstein's theory of gravitation): The expansion is thought of as the spreading out flow of an idealized uniform matter, comparable to a fluid with homogeneous density $\rho(t)$ and pressure $p(t)$ which change with time. The fluid particles can be imagined as representations of the galaxies in this picture.

Their separation increases with time as they follow the general flow pattern in the expanding cosmic material. This expansion can continue without end, or it can reach a maximum and then turn into a contraction (see Fig. 2.11). The difference in

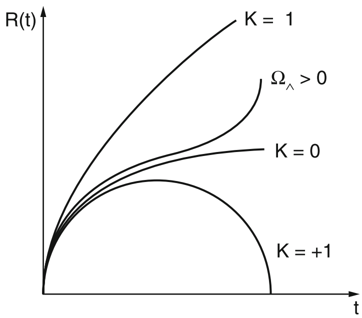

Fig. 2.11 In simple cosmological models, the Friedmann–Lemaître models, the separation of two particles of the cosmic medium changes proportional to a function of the time $R(t)$ in the way shown schematically in this figure. The number K characterizes the curvature of space ($K = +1$: spherical; $K = 0$: Euclidean; $K = -1$: hyperbolic). The curve labelled $\Omega_\Lambda > 0$ describes a model with a positive cosmological constant (see text) which seems to fit the observations very well. All models have the property that there are only changes with time. There are no variations in space. In all models there occurs a zero point of time, where all distances between objects go to zero, density and temperature become infinite. This singular point therefore lies outside the range of validity of the models depicted here

behavior is caused by the amount of matter, radiation, and other possible forms of energy in the cosmos.

The cosmic density ρ_0 is generally replaced by its ratio to the critical density, and thus written as a pure number, the "density parameter"

$$\Omega_0 \equiv \frac{8\pi G}{3H_0^2}\rho_0 \equiv \frac{\rho_0}{\rho_c}.$$

When the density is given in this way, as a dimensionless number, the fact is nicely illustrated that in these cosmological models there is no other dimension than the Hubble constant.

Not only the massive objects contribute to the total density, but any other form of energy. All the different components can be added up to a total density parameter Ω which is the sum of individual contributions each given as a fraction of the critical density.

If Ω is less than 1, i.e., if the density is below the critical one, the expansion will go on forever, but for Ω greater than 1 the expansion may turn over into a contraction, leading to the collapse of everything into a final singularity, a big "crunch."

These possibilities can also be seen in the graphs of Fig. 2.11. Which case corresponds to the real universe? Astronomers try to find out by measuring the cosmic density.

2.2.6 Accelerated Expansion

The supernovae plotted in Fig. 2.9 seem to be more distant at large redshifts than would correspond to the linear Hubble relation. Apparently the distance between us and these supernovae has grown faster than it would have, if these objects had moved with constant velocity. The expansion of the cosmos is accelerating, whereas in fact a slowly decelerating motion might be expected, if all the moving galaxies attracted each other gravitationally.

This accelerated expansion might be caused by a constant, positive energy density which would act like a repulsive gravitational force on cosmic scales. There is nothing new to a quantity of this kind. Albert Einstein already had introduced it in the equations of his theory of GR with the aim of deriving a world model for a uniform and infinite distribution of stars. Such an infinite, static system was the general view of the cosmos around 1915. Einstein defined a "cosmological constant Λ," a quantity which at present is generally written as Ω_Λ, a cosmological constant density parameter ("ccd" for short), where

$$\Omega_\Lambda \equiv \frac{\Lambda}{3H_0}.$$

As we have said already, a positive cosmological constant acts like a repulsive force which may, if it has the right magnitude, completely balance the attractive force of gravity.

When Edwin Hubble discovered the expansion of the universe, and when Alexander Friedmann showed that GR has solutions corresponding to expanding cosmological models, Einstein wanted to erase the cosmological constant from his theory. He felt sorry for having introduced it, his "biggest folly" ("größte Eselei" in German) as he said. But now this quantity has been finally established again due to the astronomical measurements of the Hubble expansion, albeit with a smaller value than the one postulated by Einstein. The equations of GR demonstrate that Ω_Λ accelerates the expansion, if it is bigger than half the matter density ($\Omega_m/2$).

The best fit to the data in Fig. 2.9 is achieved, if one chooses values of $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ for the cosmological model which clearly meets the conditions for accelerated expansion.

We will bring up further evidence for a positive cosmological constant, when we discuss the anisotropies of the CMB. In spite of the impressive observational indications, theoreticians feel somewhat uneasy about the existence of a cosmological constant. This component of the cosmic substrate is not really some "stuff" filling space like a gas, it is rather a property of empty space, a kind of inner tension which is relaxed and balanced by the expansion of space. Later on we will consider in detail the attempts to explain this mysterious quantity, especially the interpretation favored at present as the energy density of a field. Anyway, the name "dark energy" appears well chosen, since it hints at hidden action without accompanying luminous phenomena and at the darkness surrounding the true nature of this quantity.

2.2.7 Curved Space

In Friedmann–Lemaître models there are three different theoretically possible types of curved space: At any fixed time three-dimensional space is either the space well known from everyday experience, flat with Euclidean geometry, or a space with constant positive curvature, or a space with constant negative curvature. The conception of "curved spaces" is difficult, and without a recourse to mathematical expressions not easily understood. We might try to obtain a picture of those spaces in our imagination, if we think about the two-dimensional counterparts reducing the real spaces by one dimension. The three different types of space correspond then to the geometrical picture of a plane (this is the Euclidean space with curvature zero), the surface of a sphere (positive curvature), or a saddle-like surface (negative curvature) (Fig. 2.12).

Spherical and saddle-like space are surely more difficult to imagine, than the flat, infinite space. The spherically curved space is closed like the surface of a sphere: One returns to the starting point, if one continues to go straight ahead. "Straight ahead" means, of course, to follow a great circle (i.e., a circle with its center at the sphere's center) on the surface of the sphere.

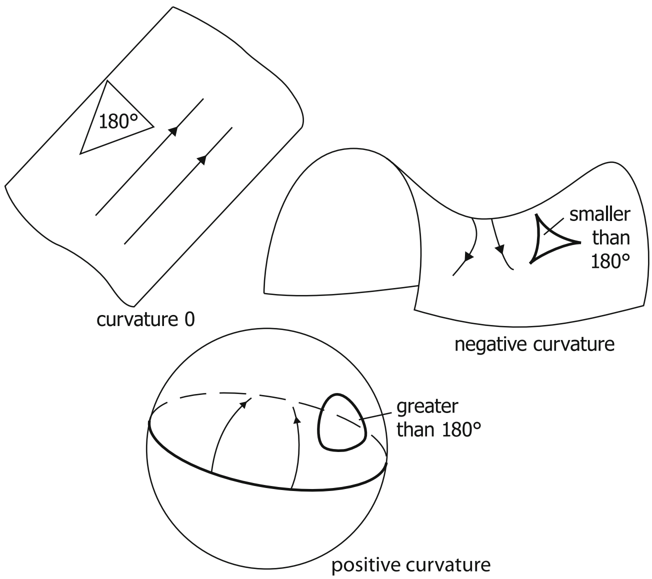

Fig. 2.12 Curved spaces can be illustrated as surfaces in 3-space, if one dimension is cut out. Three types of surfaces with constant curvature can be discerned: The plane corresponds to Euclidean space with curvature zero, where the sum of interior angles in a triangle is 180°, and where parallel lines never intersect. The spherical surface gives a picture of a space with positive curvature, where the sum of the angles in a triangle is greater than 180°, and where "parallel" lines meet the poles. Similar to the saddle-like surface is a space of negative curvature, where parallel lines diverge and the sum of angles in a triangle is less than 180°. The type of 3-space and the way the expansion develops are closely connected according to Einstein's theory

One never meets a boundary, because no boundaries exist on the surface of the sphere. The two-dimensional analogy is unfortunately somewhat unconvincing, because we must completely forget about the space outside of the surface of the sphere – only the surface itself exists and forms all of space. For three-dimensional space we have to imagine a spherical surface in four-dimensional space – not easy, even after long training.

The total volume of a spherical space is finite, just as the surface of a sphere has a definite, finite area. Flat spaces and saddle-like spaces are infinite and open, and by going straight you will never return to the starting point. In FL models it is the matter and energy density which curves space. Larger density leads to larger curvature, e.g., to a smaller sphere in the case of positive curvature.

The idea that space may be curved is a basic aspect of Albert Einstein's theory of GR: Space and time are not fixed and absolute, but defined by the masses and energies present. A massive body distorts the space–time measure in its environment, that is it influences the way clocks run and it changes the measuring rods. Vice versa the space–time geometry acts on the dynamics of the bodies. The interaction between all the masses and energies finally results in the cosmological model.

From this point of view it is absolutely astonishing that the interaction of all things in the cosmos leads to the smooth geometry of a space of constant curvature, or even of a Euclidean space.

An intuitive picture of the expansion might be given by imagining the stretching of an elastic plane, spherical, or saddle-like surface. Let us look for example at the spherical surface: The expansion is illustrated as a uniform inflation of the closed, finite surface, similar to the puffing up of a rubber balloon. "Galaxies" can be represented by marking points on the balloon. When the balloon inflates, the marked points move away from each other. The distances between points grow with the inflating balloon, although their positions (longitude and latitude) on the spherical surface remain the same. The distances change because the elastic material is stretched. This appears to be quite a useful intuitive illustration of the conditions as they are described by Einstein's theory: Distances grow because the space–time structure changes, not because the galaxies themselves move. Thus for an imagined two-dimensional observer in one of the "galaxies" on the balloon surface, the impression arises that all the other galaxies move away from him.

You can imagine such an observer on any "galaxy," and from any point of the rubber balloon he will obtain the same view of the expansion, if the points are distributed homogeneously on the surface.

For galaxies close to us the linear Hubble law holds, while for distant galaxies the curvature of the space–time becomes significant. The redshift can no longer be explained by just the Doppler effect of galaxies moving away from us, but in reality the properties of light propagation in FL models must be taken fully into account: Changes of distances by the cosmic expansion are proportional to a function of time $R(t)$ which is in our intuitive picture just the radius of the balloon.

Light propagates in the space–time geometry in a way such that one plus the redshift z is equal to the ratio of the radius $R(t_0)$ at the present time t_0 to the radius $R(t_e)$ at the time t_e of emission of the signal. (In mathematical terms $1+z = R(t_0)/R(t_e)$; for times t_e, close to the present time t_0, i.e., for close galaxies, the Hubble relation can be derived from this expression with $H_0 \equiv (dR(t_0)/dt)$ at t_0.)

If we look to the past, we see the balloon shrink. Toward the big bang all the points marked on the surface move closer and closer to each other. On the surface which represents our world there is no special point which marks the location of the beginning of the expansion, of the big bang. All points on the surface are always there, even arbitrarily close to the big bang and even on an arbitrarily small balloon. In the intuitive two-dimensional model one might think that the center of the spherical balloon is the point, where the big bang happened, but this point outside of the two-dimensional surface of the balloon does not belong to our two-dimensional world.

Moving back in time toward the big bang any finite separation of two particles goes to zero. As the particles pile up more and more, density and pressure grow beyond any limit, and become infinite at the initial state which is generally designated as the "big bang." Even theoretically we cannot follow the run of events further into the past, because the conceptions of the theory, even of time and space lose their meaning. This initial "singularity" marks the beginning of the world: Everything we observe now has come into being in a primeval explosion about 14 billion years ago. In the beginning there was infinite density, infinite temperature, and an infinite rate of expansion!

2.2.8 Redshift and Evolution in Time

The situation of the astronomers in a world described by an FL model is as follows: Light signals from distant galaxies arriving here and now have been sent by the source a long time ago. The galaxies are not observed in their present state, but as they were in a previous epoch. The astronomical observations yield a cross section through the history of the cosmos, and its present status can only be derived in connection with an appropriate model.

In our two-dimensional balloon illustration we can mark the observable region by a circle around our position. Objects within the circle can be observed, because light signals emitted by them can be received by us. The circle designates our "horizon," beyond are regions inaccessible to our observations. But our horizon grows with the velocity of light, its radius proportional to time, because light signals traveling with the velocity of light can reach us from more and more distant territories. On the other hand the balloon itself inflates – this expansion depends on the matter and energy densities. As long as matter and radiation are the dominant components of the cosmic substrate, the balloon grows more slowly than the circle representing the horizon, and new areas continuously come into the horizon. For matter the distance between two particles changes with time t as the power $t^{2/3}$, for radiation as the square root $t^{1/2}$, while the size of the horizon grows as t.

If the expansion is dominated by a cosmological constant, the rubber balloon stretches more rapidly than the horizon grows, and gradually individual galaxies disappear from our field of vision. Correspondingly our vision loses in range, when we follow the expansion back into the past. In the cosmos dominated by matter and radiation the horizon shrinks much faster than the universe contracts. This leads to the curious conclusion that as we approach the big bang there is less and less of the world within our horizon.

The redshift of the light of a distant galaxy is a direct measure of the cosmic expansion, because the universe has grown by the factor $(1 + z)$, since the time, when light has been emitted by the galaxy with redshift z.

Observations of galaxies with a redshift of $z = 6$ tell us that the universe had one seventh of its present size when that light had been emitted. The CMB tells us of an epoch with a redshift of 1100. The cosmos now is 1100 times as big as it was then. Clearly this implies that matter and radiation were much denser when the CMB originated than today.

2.2.9 A Time-Lapse Picture

Let us compress the history of the universe into 1 year. Each month then equals a bit more than a billion years in reality. Let us imagine that as the bells are ringing to welcome the new year our world starts with a big bang. The primeval substance, a radiation filling all of space homogeneously with enormous density and temperature, was without structure, but by the momentum of the mysterious initial explosion it expanded and cooled. Already in a tiny fraction of the first second of the 1st of January, matter was created: Elementary particles and soon after the simplest atomic nuclei, hydrogen and helium, were formed. Before the end of January radiation and matter decoupled and the galaxies formed. The first generations of stars in the galaxies brewed the higher chemical elements in their interiors, and hurled them – partly in the form of dust – during the final supernova explosion into the surrounding gas. Carbon was formed most abundantly; this was the basis for the formation of complex organic molecules on dust grains in the vicinity of stars.

In the middle of August our solar system formed out of a collapsing cloud of cool gas and dust. A day later the Sun was more or less in its present state supplying the planets with a pretty steady flow of radiation from its hot surface of 6,000 degrees. The hot solar radiation could be radiated away at a much lower temperature by the Earth, since the interstellar and interplanetary sky was dark and cold. These conditions on Earth permitted the build-up of complex chemical, and then biological structures. The middle of September saw the formation of the first solid rocks on the Earth's surface, and in those oldest rocks we find nowadays first traces of life: fossil one-cell organisms. Already in early October fossil algae developed, and in the course of the next 2 months a huge variety of plants and animals arose, at first in water. The first vertebrate fossils date from the 16th of December. On the 19th of December plants settled on land. On December 20 the landmasses of the continents were covered with forests. Life generated an oxygen-rich atmosphere for itself which shielded it from ultraviolet light, and thus created favorable conditions for even more complex and sensitive forms of life. Eventually, on December 22 and 23, fish evolved into amphibians which could live on dry land. On December 25 the first mammals arrived. The Alps started their folding up during the night before December 30. During the night before December 31 the human primates originated from the branch which also carried a twig leading to the present apes. Human evolution carried on with about 20 generations per second. Five minutes before midnight Neanderthal man lived on the Earth, 15 seconds before 12 o'clock Jesus Christ was born, half a second before the first sound of the bell the age of technology began. Here comes the New Year: How will the story continue?

2.3 Formation of Structures in the Universe

2.3.1 Deuterium, Helium, and Lithium

Within the first second after the big bang protons and neutrons formed out of the cosmic primeval soup. From these basic building blocks a chain of nuclear reactions led after further cooling to the atomic nuclei of the light elements deuterium, helium, and lithium. The nucleus of deuterium consists of one proton and one neutron. Below temperatures of 800 million degrees they are bound in a stable configuration. So the temperature of the cosmic structure must have decreased below that threshold – this can be computed to happen after about 3 min – before deuterium could exist as a stable nucleus. Then further protons and neutrons attached themselves to it, and built the nuclei of helium and in smaller number lithium. This attachment of protons and neutrons does not proceed further, because atomic nuclei with five or eight nucleons (protons and neutrons) are unstable. Therefore heavier elements like carbon or oxygen with 12 or 16 nucleons respectively could not build up. All these elements are produced in massive stars at a later stage in the cosmic evolution.

The big-bang model predicts that the atoms of helium and hydrogen ought to be present with a ratio of their numbers of 1–13, and this agrees well with astronomical observations. Additional assumptions are not necessary to obtain this result. It is a natural consequence of the simple hot big bang, i.e., the FL models.

We may even venture to state that during its first few seconds the universe follows especially well the rules of an FL model. Any small deviation from the expansion law of such a model would lead to a change in the production of helium. The precise measurements presently available to the abundance of helium exclude any significant effect of this kind.

The explanation of the synthesis of helium and deuterium is a big success of the standard big-bang scenario. It is also of great importance, because the production of these elements in stars is not enough: The helium abundance generated by stars is too small in comparison to the measured value of 24% and deuterium is not made in stars at all.

2.3.2 Structure Formation

The explanation of galaxy formation is more difficult, because an obvious discrepancy exists between the uniform, homogeneous cosmological models, and the astronomical observations showing the luminous matter to be arranged in discrete building blocks, the galaxies. Galaxy formation is, in fact, still in many details not understood. This is at present the most active field of research in cosmology.

One basic assumption is to consider galaxy formation as an evolutionary process which leads from initially very small fluctuations of the matter and radiation densities to the structures observed today. Small deviations from uniformity must have existed in the cosmos from the beginning, because nothing complex could evolve from a purely symmetric state.

During this process the initially small inhomogeneities in the cosmic primeval soup are intensified due to their own gravity. Eventually they separate from the general expansion and collapse to dense clumps which follow the expansion as whole objects. This appealing idea meets the following difficulty: Only after the decoupling of radiation and matter, about 400,000 years after the big bang, was it possible for small density contrasts to increase. At earlier times the condensation of matter was prevented by the radiation pressure on the free electrons. When the electrons combined with the atomic nuclei to form hydrogen and helium atoms, the radiation could propagate freely, and the matter could follow its tendency to collapse. The temperature at that epoch was about 3,000 K.

At this time, however, the density contrast of the inhomogeneities, i.e., the ratio of the overdensity of a region to the mean cosmic density, was very small, comparable to the relative amplitude of the fluctuations in the microwave background of about one hundred thousandth (10^{-5}). The density contrast of the matter can grow only by a factor thousand up to now, because the amplitudes increase proportional to the redshift. Thus they could reach only values of a few percent, but not the values characteristic for the density contrast of real galaxies. The conclusion would be that the universe had remained quite homogeneous, that galaxies and stars would not exist. This dilemma motivated cosmologists to investigate nonbaryonic dark matter as a way out for the following reasons: A background of particles of nonbaryonic dark matter does not interact directly with radiation, and is therefore not subject to the strict limit by the CMB anisotropies. Therefore the initial density fluctuations can be bigger than those in normal matter, and they can grow over a larger time span. Finally the dark matter particles would form mass concentrations which attracted and collected the normal matter. The luminous matter, that is to say the galaxies, was like the tip of an iceberg of dark matter which could not be seen itself, but which would determine by its gravity the distribution and velocities of the galaxies.

There is more in these considerations than a well thought-out scheme, because the astronomical evidence for the existence of dark matter is very strong. I will briefly describe some of it in the following.

2.3.3 The Luminous Matter

Visible light is emitted by stars. In the Milky Way and in a few neighboring galaxies stars can be discerned as single objects, more distant galaxies appear as a diffuse spot of light only. But the big telescopes catch every bit of this light down to very faint sources. Now the astronomers do what they like best: they count. They count all these galaxies down to the tiniest speck of light and add up the radiation energy. Then they try to estimate the volume of space which contains the sources they have counted.

The positions of the galaxies on the sky, and their distances have to be known for that.

The distances are estimated from the Hubble relation, and the redshifts and positions can easily be measured. Thus the spatial volume emitting the radiation is known, and therefore the radiation energy per volume, called luminosity density, can be computed.

One step is still missing to find the mass density of the luminous objects: Radiation must be connected to mass.

The theory of stellar evolution tells us how much light a star of a certain mass emits, and from precise observations in the solar neighborhood we know how the stars are distributed according to mass. There are very many stars with a small mass, and only a few with a big mass, because the small ones live long, the big ones live a short time. This fact can be expressed quantitatively as the mean ratio of mass and luminosity for stars.

Multiplication of this ratio with the luminosity density results in a value for the mean mass density of the luminous matter. About half a percent of the critical density is the estimate to date. Expressed in terms of a density parameter Ω_* ($*$ stands for star)

$$\Omega_* = 0.005.$$

There are, however, various possibilities for errors in this estimation: The galaxies chosen may not have been the most typical objects representing luminous matter, and also the Hubble constant itself is measured with some uncertainties. But the observers have counted galaxies in many different volumes – with somewhat different results – but nevertheless found that this value for Ω_* is quite reliable. It could be twice as big, but there is little doubt that the luminous matter reaches at most 1% of the critical density of the cosmic substance.

2.3.4 Dark Matter in Galaxies

In spiral galaxies the stars are arranged in a flat disk which rotates around the center. Astronomers have succeeded to measure rotational velocities at large distances from the center far outside of the luminous disk. They achieve this by observing the radio emission of clouds of neutral hydrogen. It turns out that mass is not concentrated near the central region, but that there is a nonluminous component of matter extending much further out than the visible light. Elliptical galaxies, which appear as luminous small disks without spiral arms do not rotate as a whole, but they also show evidence for dark matter, if the irregular velocities of their stars are analyzed. The mass in galaxies thus contributes somewhat more to the overall density than just the mass in stars. It reaches about 1.5% of the critical density,

$$\Omega_{Gal} = 0.015.$$

2.3.5 Dark Matter in Clusters of Galaxies

The galaxies are mostly bound in larger structures, especially dense assemblies of many hundred galaxies, so-called clusters. Their typical size is about ten million light-years (3 Mpc). These clusters are considered to be objects held together by their own gravitational force. The velocities of galaxies in clusters are, however, so high that the clusters would fly apart, if not additional dark masses existed which held them bound together.

Measuring the velocities of the galaxies, and applying Kepler's law to clusters, enables one to write down a mass balance. This forces one to accept a high fraction of dark matter in clusters. The nonluminous matter in a spherical halo around the galaxies cannot account for that. About ten times as much dark matter is necessary. This result is supported by further observations, such as the X-ray emission of galaxy clusters. A hot intracluster gas of a temperature of about 100 million degrees probably is responsible for the X-ray emission. A hot gas like that would simply evaporate from the cluster, if it were not bound by the gravity of additional, nonluminous masses. The quantitative estimates give a value for the density in agreement with the density derived from the motion of galaxies in clusters.

Many galaxy clusters act like a gravitational lens, that is they deflect light rays passing through the cluster which come from galaxies farther away from us to the cluster. The mass distribution in the cluster distorts the image of the source galaxy, and the analysis of the distortion allows us to reconstruct the mass distribution. These measurements indicate the same high fraction of dark matter in galaxy clusters.

All these data indicate that matter clumped on the scale of galaxy clusters adds up to a contribution to the total density of 15%,

$$\Omega_{cl} = 0.15.$$

The uncertainties still are considerable, and we should not exclude values higher by a factor 2. Thus a cautious estimate is

$$\Omega_{cl} = 0.3.$$

The astonishing result in any case is that dark matter is the dominant form of matter. There is about 30 times more dark matter than luminous matter. The normal matter, the chemical elements known to us, the "baryonic" matter as the physicists say, accounts for only 5% of the critical density, as we shall see below. There must be dark matter which is of a kind yet unknown. What could this unknown dark matter be?

2.3.6 Nonbaryonic Dark Matter

Astronomical measurements, and especially the analysis of the cosmic microwave background which will be discussed below, furnish many indications for the existence of dark matter which consists, apart from a small contribution of normal matter, largely of nonbaryonic matter. The elementary particles forming this dominating component of the matter are not yet known. We are familiar with neutrinos as representations of that species, but their mass is too small to contribute the required fraction of dark matter, although they originated in large number during the early epochs of the cosmos. In the Sun's interior neutrinos are produced continuously, and we meet them all the time without noticing it: They reach Earth in a steady flow and pass right through, also through us – about 100 billions of neutrinos per square centimeter, and per second! We do not feel them, because neutrinos interact only very weakly with matter. Even passing the big detectors in the underground mines of Kamioka (Japan) or Homestake (America) in tons of water they suffer only one collision per day on average.

Cosmologists take the neutrinos as examples, and postulate the existence of hypothetical particles which react weakly with normal matter like neutrinos, but which are much more massive. Up to now such particles have not been detected, although several experiments in underground laboratories have been set up to look for them. There are, on the other hand, a number of theoretical candidates. A favorite among them is the "neutralino," a particle without electric charge and with a mass of a few times the mass of the proton.

2.3.7 Galaxy Formation

The strategy in the theoretical modeling of galaxy formation has been to compute first of all the structures forming in the dark matter. This seems reasonable since there is evidence for about 10–100 times more dark than luminous matter in the cosmic structures. In a second step then the normal baryonic matter is distributed in the gravitational potential wells of the dark matter. The simulation of gas and dark matter together requires enormous computing power, and is only carried out in specially selected cases.

Such numerical and analytical investigations of cosmic structure formation are a major research topic of groups all over the world.

2.3.8 Dark Halos and Luminous Galaxies

Theoreticians have gained a lot of insights into the properties of structures formed by dark matter particles.

Although dark matter particles experience only their mutual gravitational force, the computation of their possible configurations is not quite easy, because the scientists want to follow the evolution of millions of particles to see what kind of structures are forming. This requires extensive computer simulations and numerical skills.

Some of the principal aspects can be clarified without too much mathematics. Consider a spatial volume in the expanding universe which contains a bit more mass than the average.

Under the influence of its own gravity this volume lags a bit behind the general cosmic expansion. Therefore matter is becoming less dense also in this region, but not to the same extent as outside. The contrast to the exterior region will increase in the course of the cosmic expansion and at a certain time become so large that within the volume considered the self-gravitation dominates. Then this clump of material separates, does not expand any longer, but collapses, and participates in the cosmic expansion as one whole object. This condensation of dark matter is called a "halo." The dark matter halo collects some normal matter which forms stars, galaxies, and galaxy clusters.

Let us assume for simplicity that the halo was spherical. Then after separation the density in the halo is about 180 times larger than the average cosmic density ($18\pi^2$ in a $K = 0$ model). Actually halos should rather be elliptical as numerical simulations have shown.

Figure 2.13 shows a section of a numerical simulation containing 16,777,216 particles of dark matter in a cube with an edge of 300 million light-years. The brightly colored areas are those with a very high density, and here you would expect the formation of luminous objects. There can be discerned various large-scale structures of high density like sheets or filaments, and also extended almost empty regions. All these qualitative features agree completely with astronomical data.

In Fig. 2.14 some results of the Las Campanas Redshift Survey are displayed, about 30,000 galaxies with redshifts up to 0.2. According to the Hubble law of expansion these galaxies have flight velocities of up to 60,000 $\mathrm{km\,s^{-1}}$. Their proper velocities of a few 100 $\mathrm{km\,s^{-1}}$ are in comparison quite insignificant. Thus one may use Hubble's law to estimate the distances to the galaxies of the survey, and taking the positions on the sky into account, one arrives at a three-dimensional picture of their distribution. Figure 2.14 contains galaxies selected from three bands across the sky of 6° latitude extent each and covering about 120° in longitude (the so-called right ascension). In this wedge diagram the galaxies are plotted according to their longitude and their redshift, while the latitude coordinate is compressed. The observer is situated at the tip of the wedge.

The spatial distribution appears extremely inhomogeneous. Almost all galaxies are in extended thin sheets which surround like a skin large, empty volumes (voids). The picture

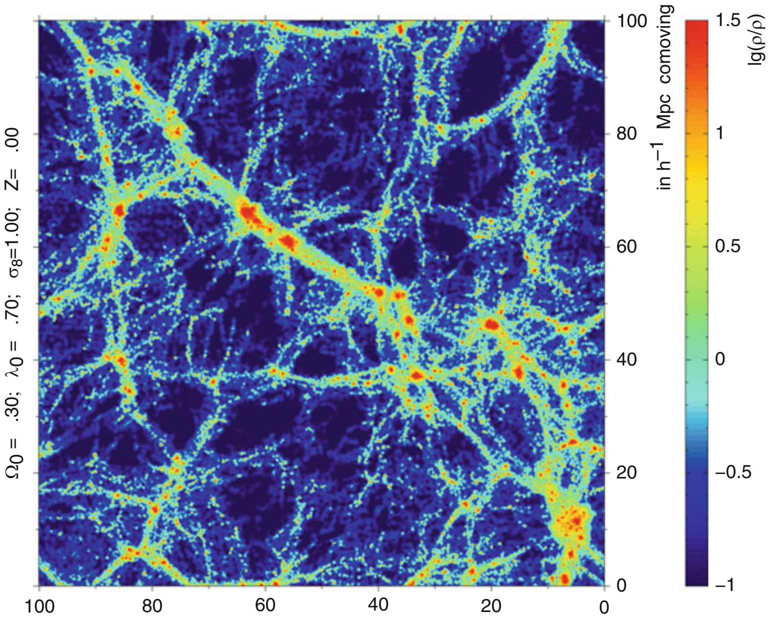

Fig. 2.13 A cross section through a cubic volume of a numerical simulation with dark matter particles shows similar condensations and voids as the observed galaxy distribution. *Bright (red)* areas mark a high concentration of particles, that is a big mass with a strong gravitational attraction. *Dark areas* do not contain particles, and therefore also no galaxies. The real volume represented by this simulation has a typical dimension of 300 million light-years. Not only by eye impression, but also in quantitative statistical measurements these simulations agree well with astronomical observations

of a spongelike pattern with galaxies situated in the thin walls of almost spherical voids seems adequate. Rich clusters of galaxies are located in places, where several walls come together. Quantitative comparisons must be done by employing a detailed model of galaxy formation. The crucial point is how to place galaxies in halos of dark matter. This is, of course, fully determined by the basic physical processes, but it is not yet possible to carry out the full-scale computations necessary to describe the complex behavior involved in the heating and cooling of the gas, the formation of stars, and the stellar explosions. Therefore the cosmologists test various recipes of how to populate halos with galaxies. Depending on the mass and history of a halo it may contain massive or very small, a few or many galaxies. The models are compared to the data in extensive

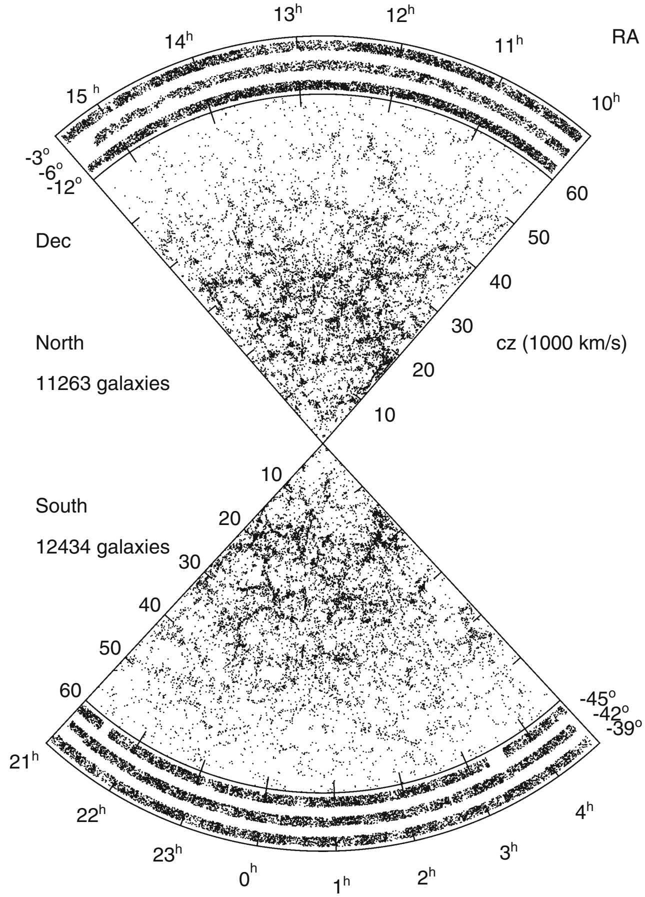

Fig. 2.14 Modern fast methods of measuring redshifts make it possible to undertake a cartography of the spatial distribution of the galaxies. To that end a galaxy catalog which lists positions on the sky of all the galaxies in a certain section of the sky and down to a limiting apparent brightness is used, and the redshifts of all the galaxies in it are measured. In this figure all the approximately 30,000 galaxies of the "Las Campanas Redshift Survey" are plotted in a wedge diagram of redshift against a sky coordinate. Only galaxies from three small bands across the sky are plotted, and the second positional coordinate is suppressed. One can clearly see characteristic features of the distribution: In a cell- or spongelike structure galaxies are localized in "walls" which surround large, almost empty volumes. The observer in this diagram sits at the tip of the wedge and surveys an angular section to the north as well as to the south

quantitative, statistical analysis. It turns out that the distribution of the galaxies in space, as well as their mean velocities, is well reproduced in the models, if the initial density fluctuations and the cosmological model are chosen adequately. Best fits are achieved for models which are at the critical density $\Omega = 1$, with 30% contributed by matter, and 70% by a cosmological constant. Similar values are obtained from an analysis of the anisotropies of the CMB (see below).

Not only in its present state, but also at earlier epochs can the galaxy formation model be tested, because meanwhile even at large redshifts many galaxies have been detected. All these tests show that the theoretical scenarios provide a reliable description of structure formation, even if not all details are correctly implemented as yet.

The earliest condensed hydrogen clouds are observed with redshift between 6 and 10. Such an early epoch can be reached only with the biggest telescopes available, and even then only a few spectral lines can be registered, no images. But in these spectra one finds not only the lines of hydrogen and helium, but also signatures of heavier elements. Even in these early epochs there must have existed stars which had after their explosion enriched the cosmic material with traces of carbon, oxygen, and magnesium. At redshifts around 3 astronomers find fully evolved galaxies shining in starlight in large numbers. Each galaxy is thought to lie inside a halo of dark matter.

For many years now the halo of the Milky Way has been investigated in large surveys. Astronomers are looking for a special phenomenon, the "microlensing" effect: The bending of light due to gravity can lead to a significant increase in the brightness of a distant star exactly in the case, when the straight line from the observer to the star just grazes the edge of a massive dark body in the halo. If the halo consisted of such objects which have received the pretty name "MACHOs" (massive compact halo objects), then some distant stars would occasionally brighten for a short time. The halo objects are not visible, but the effect of their gravitational potential on the light rays coming from a star outside. The light rays are deflected and bundled such that the passage of the MACHO leads to a brightening and subsequent completely symmetrical darkening of the star.

Millions of stars in the Large Magellanic Cloud have been surveyed now for about a decade. Several tens of microlensing effects have been observed. The conclusion is that about 30% of the halo mass lies in small, nonluminous celestial bodies. The remaining 70% of the dark matter of the halo of the Milky Way are supposed to be nonclumped exotic elementary particles.

2.3.9 Stars and Elements

The first stars formed in the condensing clouds of hydrogen and helium which we find as the predecessors of galaxies in the universe at redshift 6 and larger, i.e., when the universe had about one-tenth of its present size, and a density about thousand times bigger than now. In the interiors of these first massive stars the chemical elements heavier than helium – carbon, oxygen, and iron – were brewed. Every carbon or oxygen atom in our body has gone through several generations of stars, expelled into interstellar space in supernova explosions, recycled in the evolution of a new generation of stars, until it finally ended up on the Earth, when the solar system was formed. We consist literally of "stardust." The generations of normal stars which formed in a medium, where the heavy elements had been available already, with planetary systems around them, are a consequence of evolutionary processes which began in the early universe.

Why does this take billions of years? Well, the force of gravity is very weak, and thus it needs a long time to condense massive objects out of the cosmic matter which is blown apart by the tremendous momentum of the original cosmic explosion. The steady flow of energy from a star like our Sun, and the solid surface of a planet like Earth with its concentration of heavy elements finally provide favorable conditions for the origin of complex biological structures.

2.4 The Cosmic Microwave Background (CMB)

The big-bang model provides a simple and obvious explanation for the CMB as the relic radiation from a hot early phase of the cosmos. Therefore the CMB is considered as one of the important supporting pillars of this cosmological model. Apart from the more indirect arguments connected to the synthesis and present-day abundance of the light elements, there is no further experimental evidence of the early cosmic history.

Alternative cosmological models are sometimes being brought into the discussion, but they all fail to reproduce the uniformity and the ideal black-body spectrum of the CMB. Because of its enormous impact on our knowledge of the universe, the properties of the CMB should be discussed in some detail. I want to do this in the following.

The CMB is important, because its smoothness supports the idea of the uniform and homogeneous cosmological models, and also, because the small anisotropies of the CMB allow us to determine precisely the parameters of the models, such as the energy and matter density. Thus the CMB presents us with an independent approach to cosmic data besides the astronomical observations of stars and galaxies.

Within the framework of the FL models, we can trace the history of the cosmos to the past. As we reach earlier and earlier times, we find that the cosmic radiation field contained sufficient numbers of energetic photons to ionize all hydrogen atoms, i.e., to prevent the hydrogen nuclei, the protons, from forming an atom by binding an electron. This was still the state of affairs, when the average CMB temperature was about 3,000 K. At that time, about 400,000 years after the big bang, about one out of every billion photons in the CMB had an energy greater than the energy of ionization of a hydrogen atom, of 13.6 eV. That was just what was needed to keep the hydrogen nuclei separated from the electrons. Matter was composed of a rather uniform hot plasma. Stars and galaxies did not yet exist in that early epoch.

But due to the expansion the system cooled, and gradually first forms appeared in the primeval soup. At temperatures below 3,000 K the free electrons started to combine with the atomic nuclei to form hydrogen and helium. During this stage of "recombination" – as it is called inappropriately, because in fact hydrogen and helium atoms formed for the first time in cosmic history – the universe became transparent, the scattering of photons on electrons was strongly reduced. This happened within a short time span, but not suddenly – the process of "recombination" took about 40,000 years. The spectrum of the CMB does not show any features from this phase. No deviation from a Planckian spectrum (Fig. 2.10) with a temperature of

$$T_\gamma = 2.728 \pm 0.002 Kelvin$$

was found, and this is another, very beautiful fact in favor of the simple cosmological big-bang models: Even during the 40,000 years of recombination the temperature of the radiation and the photon energy must have followed perfectly the equations describing the FL models. Thus the shape of the Planckian spectrum has remained unchanged, while the intensity of the CMB (its energy density) decreased in proportion to the fourth power of the temperature.

2.4.1 Acoustic Oscillations in the Early Universe

Much more can be read out from the CMB. Mass concentrations of the dark matter had been forming already before the recombination epoch albeit with a very weak density contrast. The tightly coupled plasma of photons and baryons (essentially hydrogen and helium nuclei) followed these condensations, but the desire of the baryons to clump together was counteracted by the photon pressure which drove these plasma clouds apart. The competition of these two forces caused the plasma condensations to oscillate – a behavior analogous to sound waves. The largest oscillating plasma cloud had been crossed just once by a sound wave during the time interval from the big bang to the recombination time. Bigger clouds did not have enough time to develop a pressure counteracting gravity and just followed gravity by contracting slowly. Smaller clouds oscillated with higher frequency. All the oscillations were perfectly synchronized by the big bang. Contraction of the plasma condensations increased the density and heated up the photon gas, expansion decreased the density and cooled the photon gas. At the epoch of recombination the photons left the plasma clouds. Now they appear with slightly different temperatures in the detectors of the astronomers. The temperature fluctuations show up as hot and cold spots in the CMB sky maps.

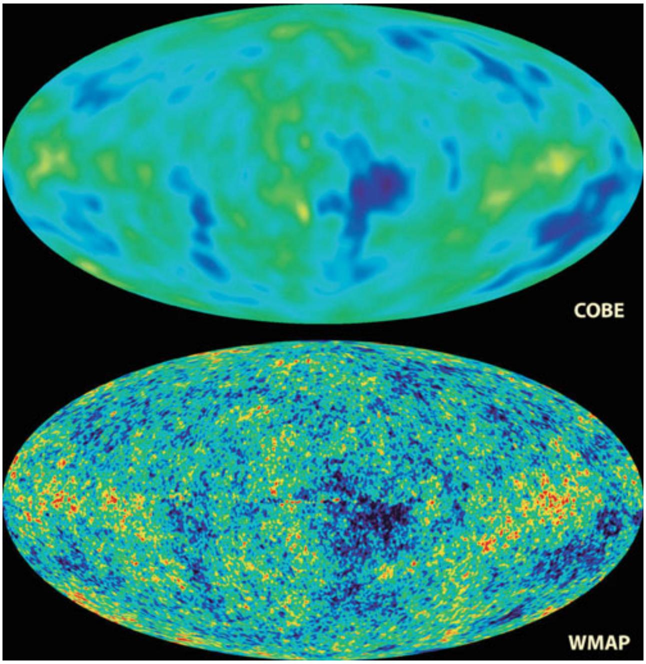

Fig. 2.15 A comparison of the sky maps obtained from the measurements of temperature fluctuations of the CMB by the satellites COBE and WMAP clearly demonstrates the improved resolution of the WMAP instruments. One also sees that specific spots of higher temperature in the COBE image (colored *yellow*) have corresponding spots in the WMAP map (*courtesy of the WMAP collaboration*)

In 1992 the first successful measurements of structure in the CMB have been carried out with NASA's satellite COBE. The sky maps obtained showed hot and cold spots on the sky with relative amplitudes of $\frac{\Delta T}{T} \simeq 10^{-5}$ (cf. Fig. 2.15).

The instruments aboard COBE had rather low angular resolution, the satellite was too "short-sighted" to recognize small structures, the angular extent had to be about 7° before a spot on the sky would be identified as a measuring point. If COBE had been looking down onto the Earth, then the whole of Bavaria would just have been one measuring point (cf. Fig. 2.16). The variations in intensity which would mirror the seeds of galaxies and galaxy clusters are expected to be on scales well below 1°.

In 2001 the satellite MAP was launched by NASA. MAP surveys the CMB sky with an angular resolution of about 15 arcmin

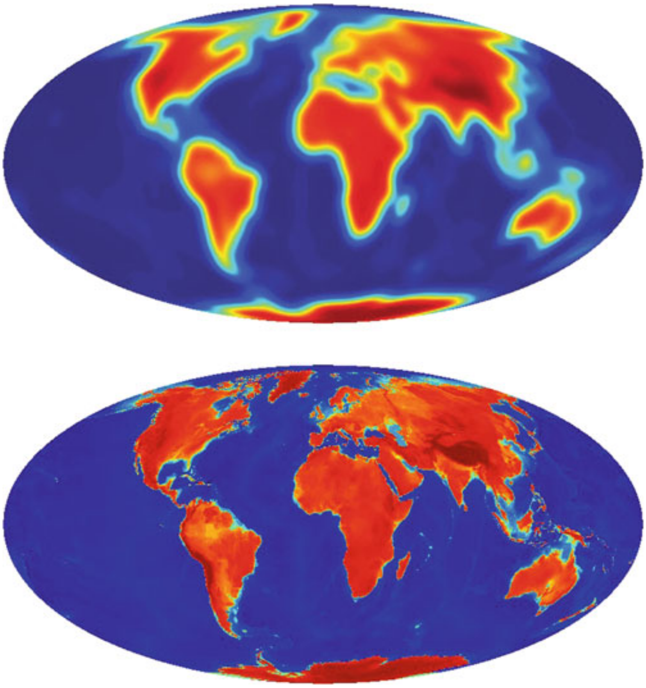

Fig. 2.16 The better resolution of WMAP observations can be illustrated by a fictitious view of the Earth by COBE (*left*) and WMAP (*right*). Bavaria would be one pixel for COBE, while for WMAP Munich would be one measurement pixel (*courtesy of M. Bartelmann, University of Heidelberg*)

in a range of wavelengths from 3 mm to 1.5 cm. The satellite was later renamed WMAP to honor David T. Wilkinson, a pioneer of CMB research, who passed away in September 2002.

The European satellite PLANCK has been launched in 2009. It has an angular resolution of 5 arcmin and covers a significantly wider range of wavelengths from 0.3 mm to 1 cm. The angular resolution of PLANCK is good enough to retrieve a major fraction of the spectrum of acoustic oscillations. Temperature fluctuations of the order of a microkelvin can be registered.

Both satellites measure besides the intensity of the CMB also its polarization properties which opens an additional window on cosmological parameters. Such measurements have been made possible by the development of new radiation detectors which are cooled down to temperatures of about 100 mK.

WMAP and PLANCK are located at a point outside of the Earth's orbit around the Sun, where the centrifugal and the gravitational forces acting on the satellites just cancel each other. At this "outer Lagrangian point" it is possible to orient the satellites such that they always look away both from the Earth and the Sun. In that way disturbing radiation is minimized.

Meanwhile the observational data gained with WMAP for the first 5 years of observation have been analyzed. The sky map of the CMB agrees well with previous experiments (cf. Fig. 2.15). As a result of these measurements astronomers can construct a power spectrum of the temperature fluctuations (Fig. 2.17).

The graph shows a sequence of maxima and minima of the temperature fluctuations depending on the angular scale in the sky over which the temperature has been averaged. The first

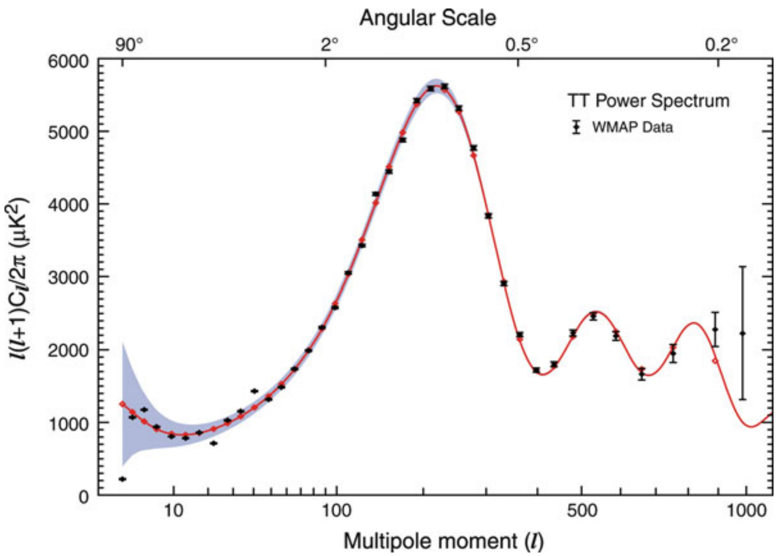

Fig. 2.17 The graph in this figure is the power spectrum of CMB anisotropies. It shows the square of the temperature fluctuations expanded in terms of multipoles. In a more intuitive way one might describe this as the square of the temperature difference between two small pixels on the sky separated by a certain angle, and then averaged over all pixel pairs. Many cosmological parameters can be read off from the shape of the curve, and its dependence on multipole index (ℓ) or angle ($\sim 200/\ell$ degrees). The regular sequence of maxima is as expected from the theoretical models of structure formation. The location of the first maximum at $\ell = 200$, and an angle $\simeq 1^\circ$ shows that the spatial curvature of the cosmos is zero (courtesy of WMAP collaboration)

maximum corresponds to the largest acoustic oscillation – its wavelength is the distance covered by a sound-wave in the time span between the big bang and the era of recombination. This distance appears on the CMB sky as a prominent signal with an angular extent of about 1°. This result tells us interesting facts about the structure of space: The viewing angle of a given length is determined by the curvature of space. The same length viewed in a space with positive curvature appears at a larger angle than in a zero curvature space, and at a smaller angle, when the curvature is negative:

The measured value of 1° means that the spatial curvature is zero, i.e., the Universe obeys Euclidean geometry – it is as simple as possible, geometrically. Curvature zero also means that the total mass and energy density Ω_{tot} reaches the critical value. Exact analysis results in

$$\Omega_{tot} = 1.00 \pm 0.03.$$

Only a small positive or negative curvature (a 3% deviation of the density parameter) is still acceptable within the limits of accuracy of the measurements.

The acoustic oscillations are a sequence of expansions and contractions, and a higher fraction of baryons causes a deeper contraction. The ratio of the amplitudes permits to derive (for a Hubble constant of 70)

$$\Omega_B = 0.044 \pm 0.003$$

for baryonic matter, and

$$\Omega_{CDM} = 0.21 \pm 0.03$$

for dark non-baryonic matter. These values are in excellent agreement with other astronomical measurements.

Baryonic and dark matter together reach only 26% of the critical density $\Omega_{tot} = 1$. Therefore there must be a further component of the cosmic energy density which balances this deficit. This component must be distributed uniformly; it must not show clumping on scales of galaxy clusters or below. It seems necessary to postulate a uniform cosmic energy density Ω_Λ with a range of values around 74%:

$$\Omega_\Lambda = 0.74 \pm 0.03.$$

A best fit to the CMB data yields values for the cosmic energy density components of

$$\Omega_{tot} = 1, \Omega_\Lambda = 0.74, \Omega_{CDM} = 0.21, \Omega_B = 0.05,$$

(see also Fig. 2.18).

2.4.2 Dark Matter and Dark Energy

Although the physicists have no direct experimental evidence yet of the nature of dark matter, there are many indications from astronomical observations that it resides in galaxies and in clusters of galaxies. Supposedly, it consists of elementary particles which have not yet been detected, but which are sought after in several experiments.

Even with dark matter there remains a gap of about 74% in the cosmic energy balance. Physicists are inclined to balance the deficit by the energy of a suitable field or by the energy of the vacuum, the ground state of the world. This reminds us of

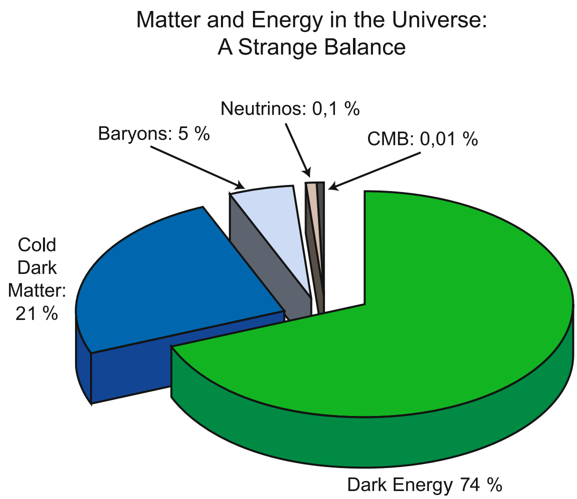

Fig. 2.18 The remarkable composition of the cosmic substrate is displayed in this diagram. Only about 5% of the cosmic matter and energy density are known. The sector inscribed "baryons" designates the fraction of matter known to us, the elements of the Periodic system. The small amounts contributed by the CMB (marked "CMB") and by neutrinos are also shown. The big majority is unknown: dark matter (21%) and dark energy (74%)

the futile attempt of Einstein to construct a static universe by the introduction of a cosmological constant. Similar to such a quantity, an almost constant field energy would accelerate the cosmic expansion, in contrast to the massive bodies in the cosmos which would decelerate the expansion due to their mutual gravitational attraction. The missing 74% of the cosmic energy density have been named "Dark Energy" (commonly written in capital letters, a custom I will not follow, because I consider the name "dark energy" a bit misleading). The dark energy would grow proportional to the spatial volume during expansion – its density would remain constant. A gas of particles on the other hand has an energy which stayed constant in an expanding volume; its energy density would shrink inversely proportional to the volume. Such a different behavior also has the consequence that the dark energy, small as it has been initially, will dominate in the course of time.

What then is dark energy? Quantum theory might help us in understanding this quantity as the energy of the vacuum. From the point of view of quantum theory empty space is a complex structure of interwoven fluctuating fields which cannot be observed, but which contribute to the energy of the ground state nevertheless. Some of these contributions can be estimated by theorists quite well, but they compute values which exceed the observational number by 60–120 powers of ten. Other contributions which cannot be computed (so far) might perhaps balance this value, but the balance must be incredibly accurate: down to the 120 first digits after the comma. It is one of the great mysteries of physics how this might be achieved.

It is absolutely remarkable that here a fundamental problem of quantum theory has become apparent through astronomical measurements. In all approaches to theories of elementary particles vacuum energies arise, but obviously they do not have a gravitational effect.

A side remark by the brilliant theoretical physicist and Nobel prize winner Wolfgang Pauli illustrates the problem nicely: A few years after the proposal of the theory of general relativity by Albert Einstein, Pauli calculated the radius of the universe under the assumption that the zero point energy of the electromagnetic field determines the value of the cosmological constant. He found that the radius of this universe would be smaller than the distance from the Earth to the Moon, in other words light rays in this cosmos would be deflected so strongly that we could not even see the Moon. This demonstrates the big discrepancy between theoretical predictions and the real situation.

The final theory, if it will ever be found, must also explain why the energy density of the vacuum is gravitationally inactive, contrary to all other kinds of energy densities. There is, of course, hope that a theory of everything, especially a unification of quantum theory and the theory of gravitation, will improve our understanding of these questions decisively. At the moment we just have to acknowledge the problem. We may also take note of the fact that for experiments in the laboratory only differences of energy count, and therefore this difficulty does not arise. Only when we consider the universe as a whole, the absolute value of the energy density plays an important role.

Thus we can only attempt to give a more precise mathematical description of our ignorance, perhaps by describing the dark energy as the energy of a field with the right properties. The beautiful name "quintessence" has been coined for such a designer-made field. But it remains actually a mystery why dark energy exists at all, and why it determines just now the cosmic expansion. If the dark energy remains constant, the cosmic expansion will continue forever and forever accelerate. But to link dark energy to the idea of field energy offers the interesting possibility that in the future the field will change with time and surprising new turns in the cosmic evolution may occur.

2.4.3 An Effect of Five Percent

Several remarkable insights follow from the study of the expanding cosmos. Evidently the big-bang model is a convenient framework to accommodate the cosmologically relevant observations in a model of cosmic evolution. The synthesis of the elements, the formation of structures in the universe can be explained without any great effort by such a model. To be sure we have to swallow a bitter pill with all that – or empty a whole glass of vermouth – because 95% of the cosmic substrate are unknown to us. We ourselves, the things around us, the planets and stars, are only a marginal phenomenon, a five percent effect, in the universe. Why is that so? Can we try to find some deeper explanation for it?

It seems to me that to this end we have to leave the area of secure knowledge, and to look at some speculative ideas about the earliest epochs in the universe.

2.5 The First Second

About one second after the big bang we can describe with some confidence the physical processes in the early universe, because then the conditions are not too different from those known from terrestrial laboratory experiments, and the known laws of physics should hold. But the first fractions of a second after the big bang are the area of more or less well-founded speculation. Close to the big bang in the standard model thermal energies are far above the energies reached in terrestrial particle accelerators. Finally, in the initial fireball temperature and density grow beyond any limit. Right at the big bang temperature, density, and curvature become infinite. The cosmological model loses its ability to describe the situation in terms of acceptable physics. Even Einstein's theory of gravitation fails at the singular big bang. It is admirable nevertheless that the theory exhibits its limits of validity on its own.

The popular question "What was there before the big bang?" leads beyond this singular boundary, and is by physicists often felt to be "not allowed," since time originated with the big bang itself, and therefore an earlier moment of time cannot exist, at least not in this model. But it seems legitimate to ask, whether for the big-bang model preliminary conditions of some kind can be imagined.

Very likely, the description of the cosmos as a classical space–time must be given up, if one wants to find out more about the beginning. Very close to the big-bang singularity the whole universe becomes in a (somewhat fuzzy) sense a quantum object. Without a unified theory which encompasses gravitation and quantum mechanics, all attempts at a more detailed description of the beginning must therefore be counted as speculative exercises. As long as such a theory is not available, one may try a more modest approach, and investigate the consequences of connecting a quantum description of matter and radiation with the classical space–time of the cosmological standard model.

It is fascinating to play around with the possibilities of cosmology and particle physics, and ask what kind of minimal structure had to be imprinted on the big bang itself, and which properties might have evolved out of physical processes.

The conceptions of elementary particle physics which come into play here will be discussed in detail in Chap. 3. Here only some basic characteristic features will be mentioned. Nevertheless some important connections between cosmology and particle physics will be pointed out. A typical example is the problem of how to explain the ratio of the number of matter to radiation particles: A ratio of about 10 billion quanta of radiation per one particle of matter characterizes the present state of the cosmos.

This ratio means that in the early phases of the cosmos the hot primeval plasma consisted primarily of particles and antiparticles (same mass, but opposite sign of charge as the corresponding particle) in almost equal numbers, but with a tiny surplus of 10 billion plus one particles versus 10 billion antiparticles. In the course of the cosmic expansion the primeval plasma cooled, particles and antiparticles annihilated into radiation, and the small surplus of particles of one in a billion remained.

We owe our existence to that tiny effect! Now one investigates the question whether this small asymmetry can be produced by the interactions of elementary particles from a completely symmetric initial state. Some more recent theoretical considerations make it quite plausible that this could happen during the phase transition, when the electroweak force splits up into the weak and the electromagnetic force, about 10^{-10} s after the big bang.

2.5.1 The Inflationary Model

The inflationary universe model has been the most popular scenario during the past 25 years, whenever cosmologists tried to describe the situation as close as possible to the singular big bang. What would happen, if right at the beginning not radiation and matter, but the energy of a field determined the dynamics of the cosmos? Physicists in Japan, the Soviet Union, and the USA asked themselves this question independently in 1981. They all found that in this case a dramatic change of the cosmic expansion would take place, an extreme acceleration of the expansion, where the separation of two particles would double every 10^{-35} s. During the tiny time interval between 10^{-35} and 10^{-33} s after the big bang – characterizing such models – the distance between particles would have been increased by the factor 10^{29}, while in a radiation-dominated FL model only a growth by a factor 100 would have occurred. The driving power behind such an "inflation" might be the energy of a scalar field. The existence of such a field was proposed originally in analogy to the designs of a unified theory of elementary particles, known under the acronym GUT ("Grand Unified Theory," cf. Chap. 3). In GUTs there are fields of this kind, so-called Higgs fields, which are introduced to describe the symmetry breaking responsible for the transition from a single fundamental force to the hierarchy of weak, electromagnetic, and strong forces observed today. Such designs suggest that the early universe was full of scalar fields, although up to now a scalar field has not been found in any experiment. The universe might have evolved from an initial phase of high symmetry with a high energy density of the scalar field in the course of continuing expansion and cooling to a state of low field energy density, and lower symmetry. If the phase transition from the symmetric to the asymmetric state occurs not immediately, but gradually and delayed, then the energy difference between the states of the scalar field may influence the expansion. It may even dominate over other thermal energies. The high-energy, highly symmetric initial state has been named "false vacuum," to indicate that it is not permanent, since the field will at last settle into the favored configuration of lower energy.

In a schematic and intuitive way we may illustrate this transition, the "symmetry breaking" as in Fig. 2.19.

Initially a small sphere lies on top of an ideal Mexican hat in the gravitational field of the Earth. The gravitational force is directed parallel to the axis of symmetry of the hat (a good approximation to the real situation on the Earth's surface). Therefore a rotation around this axis does not change anything, the system is rotationally symmetric. If the sphere rolls down into the brim of the hat and lies there at some specific location, the rotational

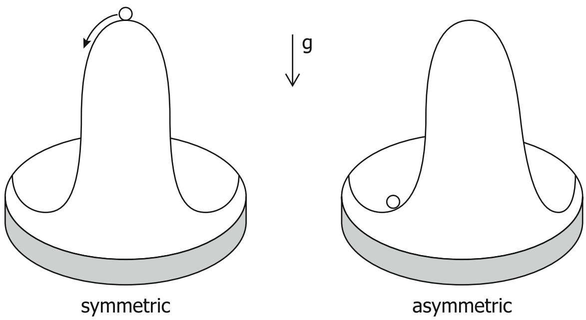

Fig. 2.19 A mechanical example for a symmetric state is a small spherical ball on top of a Mexican hat in the gravitational field on the Earth. The force of gravity points along the axis of the hat, and therefore this configuration is rotationally symmetric. Rotations around the axis of the hat do not change anything. But if the ball rolls down, and comes to rest somewhere inside the brim of the hat, the rotational symmetry is gone. In the inflationary universe model the symmetric state is interpreted as a "false vacuum" with a high energy density. The asymmetric state is the "right vacuum," where the scalar field has its lowest energy

symmetry is lost. Between the brim and the top of the hat there is a gravitational potential difference, the analog to the energy of the scalar field in the inflationary model.

But the analogy cannot be carried any further, because the energy density of the false vacuum has a remarkable property which is quite different from the behavior of normal matter. While the energy density decreases in an expanding volume filled with matter, the false vacuum keeps a constant energy density, does not thin out in an expanding volume. In fact, it is the ground state of the world, the "vacuum," which is determined by the value of the Higgs fields.

Even during the cosmic expansion the desire remains to stay in this ground state. This property is caused by the strange relation between pressure and density which holds for the vacuum, for which the pressure is equal to the negative energy density. Enlargement of a volume is then equivalent to work against a negative pressure, i.e., to gain energy. This special property of the false vacuum leads to the effect that in epochs when the energy density of the false vacuum dominates, the cosmic expansion is accelerated. This energy density acts like a repulsive force on the masses!

During the short time span of inflation all the objects that were there before are thinned out dramatically, their density becomes negligible. The temperature also goes down by the inflationary factor. The curvature of space–time is smoothed out like the wrinkles in a balloon, if it is blown up.

The inflationary phase persisted, as long as the scalar field stayed in the state of the false vacuum. It ended, when the field had reached its minimal energy. In the final phase the energy density of the false vacuum was transformed into a gas of hot radiation and particles. From this time onward the universe evolved as described by the standard model of cosmology, but with initial conditions which had been determined at least partly by physical processes. This "second beginning" requires that all matter, energy, and entropy of the observable part of the universe were created by the decay of the false vacuum.

You have every right to ask whether these doubtlessly extraordinary aspects of the first few fractions of a second have any effect on the present state of the world. Yes, they do, quite astonishingly in several respects: First of all, there is the problem of fine-tuning of the standard big-bang model which cannot be easily reduced to more fundamental, simpler properties.

Thus, the mean density must be close to the critical value, $\Omega = 1$, for the universe to exist for a sufficiently long time, and to build up enough structure. At very early times, close to the big bang, the density must be very precisely close to the critical value. Small differences would lead to early collapse, if the density was larger than critical. On the other hand, a density smaller than critical, would lead to rapid expansion, a rapid thinning out of matter, such that structures like galaxies or stars could not form.

Another difficulty of FL models is the existence of causally disconnected regions: Different space–time points can be connected by light signals only, if their separation is small compared to the size of the universe. We may try to understand this in the illustrative analogy of the balloon surface: The region connected causally to a given point, the horizon of that point, can be represented by a circle on the balloon surface. The radius of this circle grows with the square of the radius of the balloon, when the case of the radiation-dominated universe is considered. Looking to the past in this picture, we see the balloon shrink, and the horizon shrinking even faster. Any length scale on the surface of the balloon changes just proportional to the radius, and therefore any particular point which is now inside the horizon has been outside at sufficiently early times. Therefore the causal structure at early times is weird: Less and less of space is contained within the horizon of each point, the space–time splits up into a growing number of causally disconnected regions, until right at the big bang each point is completely on its own. This "horizon problem" has an especially awkward significance, when we apply it to the CMB. Observing the sky in opposite directions, we see the same CMB temperature. But at the time of recombination such regions have been separated by about 70 horizon lengths. How then, could the temperature be the same, so precisely?

The inflationary model solves all these problems by the huge expansion of space–time. The universe undergoing inflation has a curvature approaching zero, and a density close to the critical value. The horizon problem is solved, because the whole observable universe could have grown out of a tiny initial seed, a space–time bubble which was just a small part of one horizon. How big must this initial space have been? We can compute back from the present state with a temperature of 2.7 K, and a typical extent of 10^{28} cm to the epoch just at the end of inflation.

At that time the observable cosmos had a size of about 10 cm. Since inflation stretches all length scales by at least a factor 10^{29}, a dimension of about 10^{-28} cm for the initial seed of our universe would be sufficient. This is about one thousand times smaller than the causally connected volume at the beginning of inflation, at a time of 10^{-35} s. The causal length at that time is $ct = 10^{-25}$ cm.

The Russian physicist Andrej Linde, who now lives in Stanford (CA), has adorned this picture with a lot of imagination, and sketched a grand view of the universe consisting of disconnected, continuously emerging and decaying cosmic bubbles. According to Linde we are in one of these bubbles, a special one, because it provides acceptable living conditions.

This universe of bubbles is continuously changing, some parts experience inflation, others remain in the false vacuum with fluctuating scalar fields, but in total it is an eternal state without beginning and end. There is no problem with the origin of the universe, since it is not even clear whether an overall conception of time can be found for this bubbling chaos. The hypothetical inflationary model of Linde has the amusing property that the initial mass of the universe is tiny, of the order of the Planck mass, i.e., 10^{-5} g, like the mass of a small bacterium. Thus to create a complete universe like ours, only a small investment of mass or energy is necessary, at least according to this speculation. So much for Andrej Linde's scenario of "chaotic inflation."

Another very important success of the inflationary model is its prediction of small fluctuations of the energy density, a necessary ingredient for cosmic structure formation. The quantum fluctuations of the scalar field which are always present are stretched by inflation such that they attain astronomically relevant dimensions. Within detailed models a spectrum of inhomogeneities can be derived with the property that the mass excess in a given volume is decreasing proportionally to the length dimension of that volume. Data from the satellites COBE and WMAP confirm this prediction for the spectrum of CMB anisotropies.

Besides these points in favor of the inflation model, we must also mention some of its drawbacks. Especially the attempts to transform the scenario into a more precise mathematical model have met with difficulties again and again. I do not want to consider here in detail these more technical questions, but at least point out one fundamental problem: The inflationary expansion is driven by the energy density of fields acting in the early universe. We know, however, from our experience that the vacuum energy densities of the actual strong, weak, or electromagnetic interactions must not be gravitationally active, because typical energy densities are so large that contradictions to the astronomical observations would be obvious. Only a modest contribution of the order of the critical density can be tolerated – such as the dark energy derived from CMB observations. The energies of the inflation fields are larger by about 120 orders of magnitude.

Why should vacuum energies have dominated the evolution in an early cosmic phase, if now they cannot be allowed to act gravitationally at all? A good idea would be very desirable.

2.5.2 The Beginning

If you are not satisfied with the explanation given by the inflationary model, you have to investigate the initial conditions for the universe. My feeling is that this is not a question within the scope of physics, but rather a metaphysical one. Restricting physics to the explanation of phenomena within the universe saves us from a lot of difficult problems. We may, however, ask whether we should stop at the simple classical picture of the big bang, or whether we could not use arguments from physics to approach the origin a bit further.

A nonphysical answer has been given by St. Augustine in his "Confessions" (vol. 11): "To the question 'What did God do, before he created the world?' some might be tempted to answer: 'Then he created Hell for people, who ask such questions'."

A singular event like the origin of the world evidently makes the distinction between initial conditions and laws of physics obsolete. Even though, we would like to know in more detail why and how the big bang happened. Is there perhaps a quantum state, a kind of primeval vacuum, out of which the universe rises, like a bubble from the "primeval foam"? This definitely sounds metaphysical, at least in our present state of knowledge, where a theory unifying quantum physics and gravity is still missing.

A name for such a theory has already been proposed, however: "quantum gravity."

Even while quantum gravity is not yet here – or exactly then – one may indulge in speculations as to how a quantum state of the universe might be described. The English physicist Stephen Hawking has followed such inquiries intensely. He proposes to consider as possible models for the quantum cosmos only simply structured, smooth space–times; thinking in terms of our balloon analogy only a smooth balloon without wrinkles. Time does not exist in such a quantum universe. There is only a sequence of simple four-dimensional spaces – the four-dimensional surfaces of five-dimensional spheres. For illustration we can look at our balloon picture, where the surface is two-dimensional. Now try to add two more dimensions in your imagination! That is not easy, but worth trying. From this quantum cosmos our universe suddenly jumps out, and enters its temporal evolution with a finite volume from the start.

Those considerations are of principal interest, even though a nut-sized universe seems no less fantastic than a singular big bang. The question "what was before the 'primeval nut'?" cannot be asked because normal space and time categories do not exist in the quantum cosmos. Quite similarly, it makes no sense to ask for the longitude and latitude of a point outside of the Earth.

Following a different line of arguments the British mathematician Roger Penrose also argues that at the beginning the universe must have been a space–time of extraordinary smoothness and uniformity.

His starting point is an experience, we all have made every now and then: Most everyday occurrences are not reversible. A glass of water falling down from the table to the floor, splintering and spilling water, shows the normal and expected run of events. The reverse behavior, when a broken glass on its own became whole again, and jumped up onto the table, as in a backward running movie, would certainly leave us perplexed. The laws of mechanics allow this reversal in time. But actually things always happen by themselves such that an ordered state changes to a less ordered one. The notion of "entropy" is very helpful to understand this property of nature. Entropy is defined as a quantity which measures the amount of disorder in a system. An ordered system, like a crystal, has a low entropy, a gas of molecules bouncing around irregularly has a high entropy.

The everyday experience of growing disorder corresponds to the law of increasing entropy (the "second law of thermodynamics"). The numerical values for the entropy of a system result from the possible different positions and velocities for each particle subject to the fixed total energy and the volume occupied by the particles.

Penrose attempts to characterize quantitatively the entropy of the universe, rather of its observable part. The numerical estimates reach gigantic values, if besides radiation and matter the possibilities to produce entropy hidden in the gravitational field, in the wrinkles and curvatures of space–time, especially in black holes, are included.

The initial conditions for the universe, as we know it, represent just one out of $10^{10^{120}}$ possible configurations of the cosmos. Can we postulate a selection principle of such precision within the scope of physics? A strict smoothness condition as suggested by Penrose might be a possible approach. But can this be derived from the basic equations? This is still hidden in the darkness of the unknown and unexplained.

Physicists will be engaged for some time to come in explaining the big bang. The fun involved in speculations and the enthusiasm for conceivable scenarios makes cosmologists prone to believe that what is conceivable is already real. To quote Albert Einstein: "To the inventor the products of his imagination appear so necessary and natural that he sees them and wants them to be seen not as structures of his thinking, but as given reality."

All considerations about the first moments of the universe, about its initial state and conditions, belong to the empire of metaphysical speculation.

2.6 The Anthropic Principle

In a situation, where the explanations of physics for the origin of the world reach their limit, a chain of arguments has found widespread interest which is called "anthropic principle." The fact that intelligent life exists on the Earth means that the conditions for the origin of intelligent life must be fulfilled in the universe. This rather trivial, logical statement of a necessary consistency has led to remarkable, nontrivial insights.

Life as we know it, could not have originated, if the constants of nature were slightly different from their actual values. The strength of the attractive nuclear force is just enough to overcome the electrical repulsion between the positively charged protons in the nuclei of common atoms like oxygen or carbon. But the nuclear force is not quite strong enough to bind two protons together. The diproton does not exist. But, if the attractive nuclear forces were a bit stronger, the diproton could have been formed, and then almost all the hydrogen in the cosmos would have ended up as diprotons or higher elements. Hydrogen in that case would be a rare element, and stars like the Sun generating energy over a long period of time by the slow fusion of hydrogen into helium would not exist. On the other hand, with a weaker nuclear force it would be impossible to have larger atomic nuclei. If a star like the

Sun generating energy at a constant rate over billions of years is necessary for the evolution of life, then the strength of the nuclear forces must be within narrow bounds.

A similar, but independent numerical fine-tuning can be found with the weak interaction which in reality steers the fusion of hydrogen in the Sun. The weak interaction is about a million times weaker than the strong interaction responsible for the nuclear force. It is just so weak as to ensure a slow and uniform burning of hydrogen in the Sun. Stellar lifetimes would change dramatically, if the weak interaction was somewhat stronger or weaker, and this would make it difficult for life depending on sun-like stars.

Another numerical agreement concerns the mean distance between the stars which in our galactic environment amounts to a few light-years. Maintaining the view that the stars can have a decisive influence on human life is not necessarily an argument from astrology. We would not have any great chance of survival, if the mean distance between stars were ten times smaller, for example. In that case another star would have come close to the Sun with high probability during the past 4 billion years. If it came close enough to disturb the planetary orbits, the effect might be disastrous. It would be sufficient to push the Earth into a slightly more eccentric, elliptical orbit to make life impossible.

One could enumerate many more happy constellations of this kind: A sensitive balance between electromagnetic and quantum mechanical forces causes the variety of organic chemistry. Because of these fine-tunings water is liquid, chains of carbon atoms form complex molecules, hydrogen atoms build links between molecules. But a small change of the constants of nature can destroy all that.

These numerical coincidences are statements of the "weak anthropic principle" which generally expresses the opinion that our existence is only possible under specific conditions. The scientists, who proposed this principle, want to draw attention to the remarkable harmony between the structure of the universe and the necessary requirements of life and intelligence. Our universe satisfies these conditions, but the reason why it does so cannot be explained within the scope of present-day physics.

There are also those, who advocate a "strong anthropic principle," stating that the laws of nature are such as to lead finally to the evolution of human life. This way of arguing, setting a final goal as a cause for evolution, is forbidden in science, it is frankly theological in a scientific disguise. Now theology and science are different, and it would be a mistake to force theology to be a branch of physics. Therefore we shall tolerate theological principles like the "strong anthropic principle" as of metaphysical or theological value, and discuss possible clashes or concurrences with scientific reasoning.

Obviously these aspects provoke speculations of all kinds, also of theological importance. It is very suggestive to suppose a divine plan behind such a tailor-made universe. The more we investigate the connections between forces and constants of nature, the more we find evidence for a precise fine-tuning which enables life of our type to evolve. To that we may add arguments from the theory of evolution which indicate that the biological evolution toward human beings has moved along a narrow, precarious path.

But these arguments cannot serve as proofs in a scientific sense, just as the theological or cosmic proof of God's existence put forward in the Middle Ages cannot be accepted in science. Whenever a specific, actual situation is described, it becomes more and more improbable, as more of its characteristic properties are considered.

There are proponents of the anthropic principle, who understand it as a kind of selection principle. The chaotic inflationary model put forward by Andrej Linde with its multitude of bubbles, causally disjoint and perhaps even equipped with different laws of physics, different constants of nature, would be a "many-world" model which may be at least a logically acceptable possibility. Among the many bubbles there is at least one which is equipped with a combination of constants and laws of nature suitable for us. Just as you will probably find in a department store a suit that fits well, if there are many choices, there will be among the many "universes" one which allows the evolution of life. The mysterious fine-tunings are thus no longer mysterious, they are a trivial consequence of life finding the right one among the multiple-choice universes.

A number of theoretical physicists, among them the Nobel Prize laureate Steven Weinberg, find life in the "multiverse" – one of those ugly names coined by people, who apparently have no sense for classical Latin or Greek – apparently quite attractive. Apart from the inflationary model there are other speculations which also propose a multitude of worlds. If our universe has jumped out of a quantum vacuum as a random fluctuation, such processes of creation could happen again and again. A fundamental theory, like "string theory" with its complex vacuum structure may – as some speculations say – produce a variety of worlds quite naturally. Anything possible according to logic might in some sense exist in the multiverse. It appears to me that these considerations try to evade an answer to the question: "Why is our universe as it is?" I also feel that they are not very economical: To have billions of universes inflating and decaying to finally create the possibility for life on an insignificant planet at the edge of a galaxy seems a high investment. We do not even know which connections between the quantities of physics might not have been discovered yet. Therefore it seems premature for the physicists to argue for parallel universes, before the urgently sought after theory of everything has been formulated. We should not throw the towel into the ring too early.

The founder of the theory of gravity, Isaac Newton, has supposed that the fact that all planets move around the Sun in a plane is due to the will of the Creator. According to Newton's theory every planet might move in its own orbital plane, with an orientation different from the others. Today we think that the rotation of the primeval solar nebula made it collapse into a disk, and for this reason the planets move in the plane of that disk. An obvious astrophysical explanation. I expect that many at present inexplicable fine-tunings will be resolved in a similar way in the future.

The anthropic principle is of some importance for physics, because it points out relations which need to be explained. It cannot pass as a principle of physics, however. We may let it pass as a metaphysical argument, and as an indication that the world is made hospitable for us – by whomever or by whatever means.

2.7 How Will It End?

The dark energy accelerating the expansion of the universe will dominate the cosmic evolution for some time. If it is really a cosmological constant, the expansion of the universe will go on without end. It might also be the case, that the dark energy is the energy of the ground state of some field, and that it only appears to be a constant at present, but can change over large cosmic time spans. Then it depends on the time development of the field whether a new big bang will occur, or new particles will be created continuously from the stock of dark energy. Let us not dwell on these speculations now, because there is no experimental indication of a deviation of the dark energy from a constant value.

The continuing cosmic expansion in the case of a constant energy density is of decisive importance for the final state of the cosmos. When there is no end of time in the future, every physical process, even the slowest one, can run to its end.

The biosphere of the Earth will perish in 5 billion years, when the Sun will blow up to become a red giant extending beyond the Earth's orbit.

After that it will gradually get darker, because the stars will be extinguished after they have used up their nuclear fuel and the last supernova explosions fade away.

Systems bound by gravity will radiate away their energy in the form of gravitational waves according to Einstein's theory of gravity. Since 1978 we have learned from observations of the binary pulsar 1913+16 that its orbit changes exactly as the formula for the energy loss by gravitational radiation predicts. Thus over tremendously long periods of time, much longer than the actual age of the universe, all gravitationally bound systems will radiate away their energy of motion, and the bodies will crash into each other, finally ending in black holes. Gigantic black holes, each one made from all the stars in a galaxy, then move away from each other in a dark cosmos. At the same time it gets "colder" and "colder," because the temperature of the microwave background radiation keeps falling. Such is the dreary picture of the end of the universe as the cosmologists draw it.

The drawing is not yet complete, if Stephen Hawking is right with his hypothesis put forward in 1974 that black holes can evaporate. After about 10^{70} years all the black holes should have evaporated, and only very long-wavelength radiation would remain. As in the beginning the universe is filled with radiation, at the end of a temperature which is gradually approaching absolute zero.

The predictions for the end of the cosmos are extrapolations into a very distant future, and they depend on our present state of knowledge. May things turn out quite differently? Probably not, because physical processes relentlessly drive the evolution toward the final state, where everything is dead and cold, and swallowed by black holes. Only if, as we have said, the field responsible for dark energy developed in an interesting way, then quite a different story might unfold.

The picture is incomplete in yet another aspect. The fact that intelligent life has evolved in the cosmos has not been taken into consideration. How far and in what direction can a technically oriented culture evolve in a few billion years? This will be beyond our imagination. But we can speculate about the question whether the boundary conditions of the expanding universe imply that intelligent life has to come to an end. The physicist Freeman Dyson living in Princeton has indulged in such speculations. He reaches the optimistic conclusion that intelligent life can survive forever, if it learns to adapt itself to any type of environment. Even the cosmos consisting mainly of black holes would not be completely dead. Between long and quiet phases there would occur every now and then a burst of gamma radiation, when one of the black holes evaporated. During quiet phases life would be in hibernation, and become active only during short intervals to make use of the newly produced energy. There would be no end to such activities.

For mankind on the Earth, as we have mentioned already, there is an end to comfortable living, when the Sun blows up to become a red giant star. This event in about 5 billion years will destroy the Earth's biosphere. It is the task of future generations to survive that. Five billion years is a substantial time span available for the evolution of human intelligence. I do not doubt that we will learn during this time to control simple, astronomical events or to escape from them. Then mankind has a long future ahead.

2.8 Extremes of Space and Time: Big Bang and Black Holes

The universe begins with the big bang, a singular state out of which space and time, matter and radiation arise, and where there is no before. We have discussed its remarkable features in the last section, and to do that we had to look far into the past. But even at present there are celestial objects in our Milky Way which have similar properties, namely the black holes. Their existence is predicted by Einstein's theory of general relativity as an extreme state of matter: The black hole which lets neither matter nor radiation escape, but swallows everything which gets within its reach of attraction, is like a mirror image in time of the big-bang singularity from which space, time, and matter escape, but nothing is swallowed.

These singular states of a physical theory seemed so strange at first that it was a long time before physicists accepted them as real things to be taken seriously. Meanwhile these fascinating structures are mentioned like commonplace objects in movies, journals, and books of all kinds.

It is a fundamental belief of physicists that all quantities in nature are finite and exactly measurable. Singularities in physical theories are seen as the consequence of a faulty mathematical formulation, or an expression of the intrinsic incompleteness of the theory. In this sense GR theory predicts its own failure, the limits of its validity. It must be replaced by a more general theory which overcomes these limitations. Many physicists are convinced that a theory linking quantum mechanical conceptions with properties of Einstein's GRT is needed. One might think of quantizing GRT, but despite zealous efforts there has been no success. On the other hand one could imagine a more fundamental quantum theory encompassing GRT such that the classical theory would result as a well-defined approximation. String theory claims to have achieved just that, but this approach as far as it has been developed now gives us no more than a vague guess, as to how the problem of singularities might be resolved.

Intuitively it is obvious that big masses must suffer a catastrophic fate by the action of gravity, because the gravitational force is attractive for all massive particles in the same way. In addition it is long-ranged, i.e., it decreases slowly with distance (proportional to the inverse square of the distance, to be exact).

If we add more and more particles to a given mass, the gravitational force pulling the particles together grows proportionally to the number of particles. It will finally dominate over all kinds of pressure forces which might oppose its contracting power. The thermonuclear pressure in the interior of a star, the Fermi pressure of a cold gas of electrons or neutrons, and the repulsive force between nucleons very close to each other, all will be overcome by the gravitational force, if the mass of the body is sufficiently big. In addition not only matter, but also antimatter, as well as all forms of energy act gravitationally. Furthermore any kind of energy feels the pull of gravity. Therefore a huge inner pressure can balance a big mass up to a certain limit, but since the pressure also contributes to the gravitational force eventually it becomes itself responsible for the collapse.

At first these singularities were thought to be a consequence of the strong symmetry conditions which had to be imposed to find solutions of the complex equations of general relativity. The hope was that less symmetric, slightly changed solutions would not possess such singularities. In the period from 1965 to 1970 the British mathematicians and physicists Roger Penrose, Stephen Hawking, and Brendan Carter showed that singularities of space–time occur in general nonsymmetric cases (they are "generic") and are essentially stable against small perturbations.

The singularities themselves cannot be investigated without a theory of quantum gravity, but we can try to describe the space–time structure in the environment of these infinities. What happens close to black holes? The conceptions behind words like "black hole," "space–time," and "gravitational collapse" involve the intellectual power of Einstein's theory of gravitation (we often abbreviate "theory of general relativity" as "GRT"). We should make ourselves familiar with a few fundamental features of this theory, before we discuss the extreme aspects.

2.8.1 Space and Time

The conceptions of space and time are intuitively well known to everybody: Space has three dimensions, is infinite in all dimensions, and basically nothing else but the stage for all kinds of physical processes. According to the classical picture each localized event occurs at a definite time. We know from our everyday life quite well that a rendezvous can only be successful, if we agree upon both place and time.

In a train timetable, e.g., place and time of the stations along a route are given, and we can display this graphically in a space–time diagram as in Fig. 2.20.

This drawing shows the "world line" of a passenger, who travels from Munich to Stuttgart. Already this trivial example shows the possibility how to survey motions completely in a space–time diagram. More complex situations like the motion of masses in a plane or in space can be represented in the same graphic way: Kinematics is just space–time geometry. Classical mechanics established by Newton can be displayed in this graphic way which is nothing but a translation of the usual notions of space and time.

Newton assumed that space was "absolute" with geometrical properties which already Euclid had derived from a few assumptions as, e.g., the axiom of parallels. These assumptions were

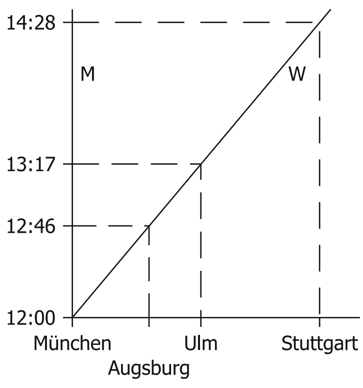

Fig. 2.20 A train timetable can be represented schematically as a space—time diagram. The graph can be interpreted as the world line of a passenger travelling from Munich to Stuttgart

considered by Newton self-evident, and he did not even mention them explicitly. In addition Newton supposed that a universal "absolute time" existed fit for all processes and measurable. His idea was that absolute space with its fixed metric was the unchanging background, where bodies moved according to a given absolute time measure. In Fig. 2.20 the axis for localities is an illustration of the absolute space, and the time axis of the absolute time of Newton. In classical, nonrelativistic physics it is simply supposed that of two events one can always determine their relative position and the time interval between them, although it is not really clear what is meant, when two events far apart from each other are said to happen simultaneously.

The motions actually occurring were explained by Newton in his famous law which states that each body, because of its inertia, is either at rest or in uniform linear motion. Forces acting on the body deflect it from this undisturbed motion.

One of these forces is gravity. Newton describes it as the mutual attraction between two massive bodies proportional to the product of the two masses and inversely proportional to the square of the distance between them.

Albert Einstein has replaced Newton's assumptions on space and time in two steps by new assumptions which are somewhat more general, and which fit the real situation even better. The desire to find a better theory developed in the wake of Michael Faraday's experiments and James Clerc Maxwell's theory of electromagnetic phenomena. It seemed that electric fields had to be considered as real physical objects. It follows from the field equations that electromagnetic waves in empty space propagate with the velocity of light. According to Maxwell the propagation does not depend on the motion of the light source, but only on the emission event. How can this be understood? Wouldn't we expect the velocity of light to be higher or lower depending on whether the source approaches us or moves away, respectively? These difficulties led Einstein in 1905 to abolish the conceptions of absolute space and absolute time, because they were not appropriate to describe such processes.

He began with two basic assumptions: The velocity of light ought to be independent of the motion of the source, as in the theory of Maxwell, and it ought to be an upper limit for the velocity of signal propagation. Both assumptions have meanwhile been confirmed many times in experiments.

The theory of special relativity (SRT) developed from this starting point has several remarkable and surprising consequences. Time no longer flows uniformly at all points of space, like Newton's absolute time, but the flow of time depends on the motion of the clock or of the observer, who by some means measures time. Moving clocks have a slower rate than clocks at rest. We do not realize this normally, because the effects are tiny, as long as the velocity of the clock is small compared to the speed of light. The slowing down of moving clocks has been demonstrated several times by comparing an atomic clock at rest on the Earth with another one transported in an airplane. The moving clock was really slower by a few billionth of a second, when both clocks came together again.

The change of the flow of time becomes distinctly noticeable, if the motion is very close to the speed of light, such as for particles in one of the big accelerators. It has been shown that particles which at rest would decay within fractions of a second, survive several seconds while racing around in the accelerator. There is also the famous "twin paradox" as another consequence of this effect:

One of the twins stays on the Earth, while the other one travels for several years with high velocity in space. When they meet again, the twin on the Earth has grown older by just so many years, while the traveler has remained young. The faster he has traveled, the younger he has stayed. How can this be reconciled with the principle of relativity? Each of the twins moves relative to the other, and it should not matter, who is considered to be at rest or to be moving. That is true, but when the twin traveling away from the Earth returns home, he has to change his direction of motion, and he must decelerate such that the relative speed between the twins becomes zero. These changes of the straight, linear motion (called "accelerations" in general by physicists) distinguish the traveling twin from the one at rest, and therefore they find a real difference in proper time, when they compare their clocks. In Fig. 2.21 we see the space–time triangle illustrating the twin paradox. At U the direction of motion of the traveler changes, and at R it again changes, when he comes to rest.

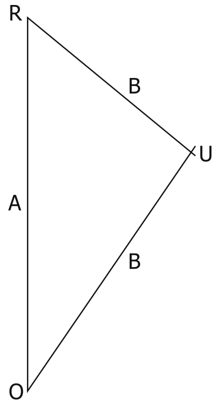

Fig. 2.21 In space–time the triangle inequality is valid in the unusual form that the distance OR (the proper time interval from O to R) is greater than the sum of the space–time distances OU plus UR. This is the so-called twin-paradox The twin at rest (path A) experiences a longer time interval between events O and R than the travelling twin, who moves from O to R along path B (Time is vertical, space horizontal in this diagram)

This experiment has not been carried out yet, but there are no doubts about the outcome, according to all we know about the space–time of SRT.

Very significant is the fact that in general we cannot decide whether spatially separated events are simultaneous or not. This depends on the state of motion of the observer. The classification of events into "earlier" or "later," the distinction between "before" and "after" is in general not unique, but different for different observers. If we see two events A and B such that B follows A, then an observer moving relative to us with the appropriate velocity might conclude that A follows B, i.e., exactly the opposite temporal ordering of events.

In a space–time diagram we would plot the world line of an observer, who moves with constant velocity compared to an observer at rest, as a straight line tilted against the vertical time axis. Events which are judged as simultaneous by the moving observer lie on a straight line which is tilted toward the world line away from the vertical direction. Moving and nonmoving observers have therefore quite a different view of the order in time of events (Fig. 2.22).

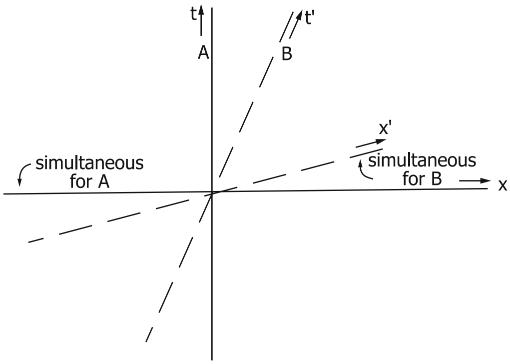

Fig. 2.22 All events considered as simultaneous by a moving observer B lie on a straight line which is tilted against the horizontal axis. The horizontal axis is the line containing all events simultaneous for an observer at rest A. Any event in the area between the two lines of simultaneity is considered by A to happen in time after the event at the origin of the coordinate system, by B to occur before

These consequences of Einstein's theory have found great public interest about a hundred years ago, when they were put forward. Probably the feeling was that these insights pointed to our redemption from the inflexible law of temporal order, of the "before" and "after" which now depended on the state of motion and had thereby lost some of its importance.

The light rays moving out from a certain space–time point form a surface in the space–time, the "light cone" of that point (Fig. 2.23). The totality of light cones has a deep significance: Our experience tells us that no signal can propagate faster than light. Therefore inside the light cone of a space–time point are all the events which can be reached by signals from the tip of the light-cone. We can also draw a "past light-cone" for each point, i.e., the surface consisting of all the light rays reaching that point. Inside the past light cone are all events which can reach the point at the tip of the light cone by signals (see Fig. 2.24).

As we can see in Fig. 2.24 this divides the environment of each space–time point into separate regions. This "causal structure" of space–time can be described mathematically by the consideration of the four-dimensional space–time distances of events: For two events the distance between the two points in space and the time interval are combined in the following way:

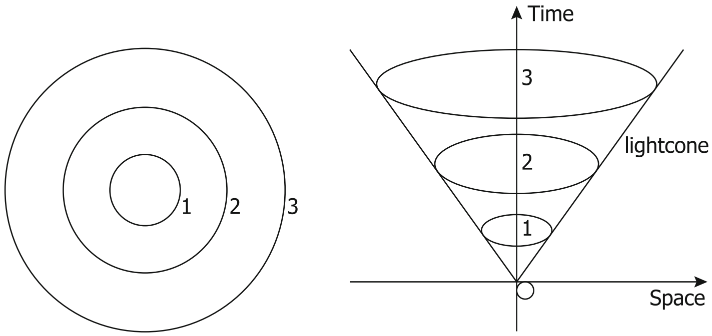

Fig. 2.23 Taking instead of 3D space a 2D surface we can graphically present the propagation of light rays. The signals of a source of light at the coordinate origin 0 reach larger and larger circles around 0 with increasing time. In a space–time diagram we can draw these circles above one another along the time axis. They form a cone with the tip at 0. The opening angle of the cone depends on the measure used for spatial distances. One can choose the unit such that this angle is a right one, i.e., 90° (this is achieved, for example, by choosing seconds as time units, and light-seconds as length units). Light rays then follow straight lines bent by 45° against the vertical time axis, and also by 45° against the horizontal space plane

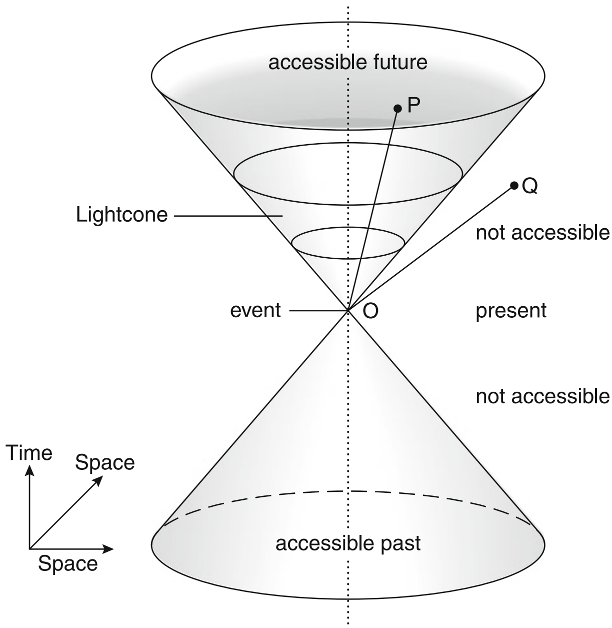

Fig. 2.24 The light cone of an event 0 is the set of events which can be reached by light signals from 0, or from which 0 can be reached

Squaring the time interval after multiplication by the velocity of light (i.e., squaring the distance a light ray would reach during that time interval), and subtracting from it the square of the spatial separation gives the square of a quantity which is commonly called "space–time distance," and designated by s. This quantity is zero along light rays, i.e., on the light cone (by construction); the square of s is positive for events inside, and negative for events outside of the light cone. A series of experiments have shown that s is the time measured by clocks in space–time.

The "twin paradox" is seen to be just the somewhat unusual inequality which holds for triangles in space–time (Fig. 2.21): The twin at rest moves along path A from O to R, and measures a proper time interval $s(OR)$. The twin along path B travels first from O to U and measures $s(OU) + s(UR)$ as his proper time interval. Now $s(OR)$ is greater than the sum $s(OU) + s(UR)$, as you can easily check. This is the opposite to the relation for a triangle OUR in Euclidean geometry.

The special theory of relativity is established firmly as the foundation of all parts of physics where gravity can be neglected. Especially for high-energy particle physics the theory is indispensable.

Einstein suggested to take gravity into account by a second refinement of the space–time structure. His theory of gravity, the theory of general relativity of 1915 (GRT), is based on the idea that the structure of space–time is not fixed, but determined by the masses and energies present. Each massive object distorts the space–time metric in its surroundings. The dynamics of the massive bodies, on the other hand, is determined by the geometry of space–time. Thus different objects act on each other: This is gravity. The metric field between the bodies acts also on light which is deflected near massive objects. Consequently the light cone structure is not fixed, but results from the distribution of the masses. This tight connection of space, time, and matter causes problems for computations in GRT. But this theory was a big step toward the unification of physics. Since gravity is woven into the geometric properties of space–time, there is no need to treat it as in Newton's theory like an additional independent structure. In that sense GRT is the most beautiful part of physics.

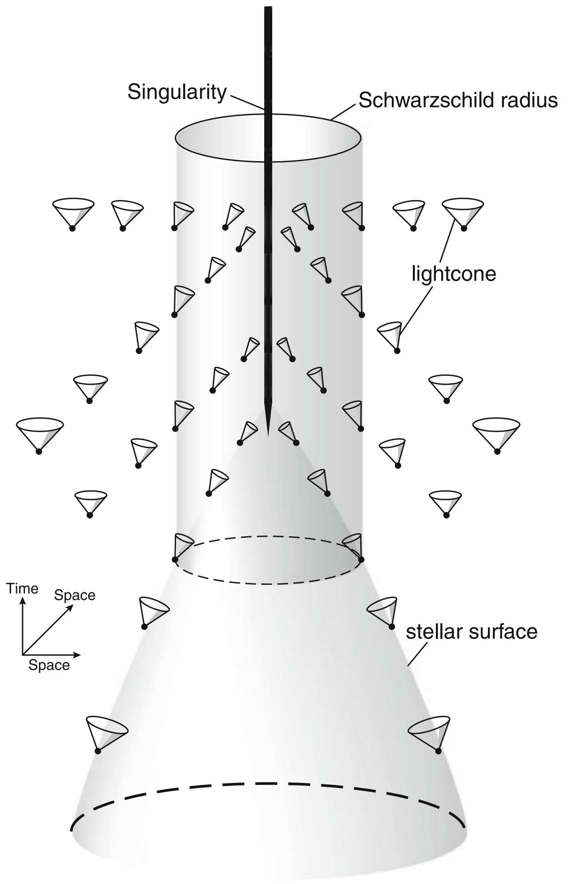

Fig. 2.25 A star collapsing to a black hole shrinks to a point, where the whole mass is concentrated with infinite density. In this space–time diagram the star is a circle contracting to zero radius. The light cones bend toward the mass. One sees that before the final singularity is formed a surface in space–time exists, where light can no longer propagate outward. The light cones at this surface all point inward toward the singularity. This surface is the Schwarzschild horizon, its radius is the Schwarzschild radius ($r_s = 2GM/c^2$; G: constant of gravity, M: mass, c: velocity of light). The Schwarzschild horizon is a null surface; its existence prohibits for any outside observer to see the black hole

In Fig. 2.25 the change of the light cones in the vicinity of a mass is shown. Close to the mass the light cones bend inward, that is the light rays are curved. This deviation from a straight line propagation is due to the "curvature of space" caused by a massive body according to GRT.

We may try to grasp this somewhat difficult idea by looking at the orbits of freely moving particles. In the four-dimensional, flat space–time, the Minkowski space, the particles move along straight lines, if no forces act on them. The straight lines in Minkowski space are "geodesics," and there are also geodesics in curved space–times. The four-dimensional distance measured along geodesic curves becomes maximal. (In Euclidean space the shortest distance between two points is given by the geodesic connecting them). As in Minkowski space the geodesic distance in curved space–times is equal to the proper time interval s measured by a clock carried along this curve.

In the case of positive curvature two "parallel" geodesics converge toward each other, like the big circles on the surface of a sphere which are parallel at the equator, and intersect at the poles. Negative curvature lets initially parallel geodesics diverge (cf. Fig. 2.12).

Light propagates along geodesics, along which any proper time interval is zero, so-called null geodesics. A photon does not experience time at all, even if it covers cosmic distances of billions of light-years between emission and reabsorption.

In a curved space–time the null geodesics form the light cone just as in a flat one. Since signals cannot propagate faster than light, the paths of particles of positive mass lie inside the light cones defined at every point of the path. Thus the light cones are the boundaries for the propagation of signals, they mark the causal structure of the space–time.

2.8.2 Gravity

Gravity by itself is no more in Einstein's theory. It is built into the space–time texture and has become part of the geometry. But how then can it determine the mutual attraction of masses and the motion of heavenly bodies?

Let us look for a moment at the solar system: According to GRT the orbit of the Earth around the Sun is determined by the distortion of the space–time in the solar neighborhood. Imagine the space–time structure like a sheet of elastic material. A big massive body like the Sun produces a deep dent. Bodies which come too close to this funnel fall down toward the Sun. This is the

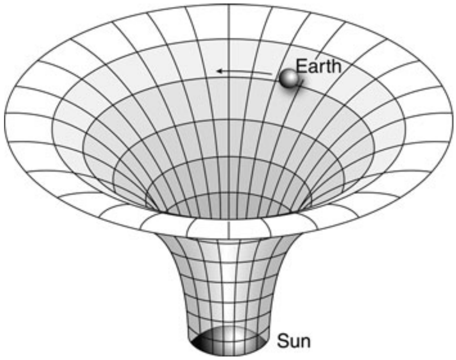

Fig. 2.26 Gravity is transformed to a geometric property of space–time in Einstein's GRT. The figure displays schematically how the Sun with its big mass is at the bottom of a deep funnel in space–time which we illustrate as an elastic rubber sheet. The Earth rolls along the funnel wall, prevented from crashing down into the Sun by its own velocity

attraction between masses which looks a bit one-sided in the case of the solar system, because the mass of the Sun is so dominant (Fig. 2.26).

The Earth can move on its path at the wall of this funnel, because its velocity is just right to keep it there – as Kepler's law requires.

The Earth has its own mass, and it forms a smaller funnel with satellites moving along its walls. Since the mass of the satellites is negligible, they do not experience any gravitational force inside, everything is weightless, at zero-G. The gravitational attraction of the Earth is compensated by the orbital motion. Practically everybody has seen TV transmissions of astronauts floating around in the space station, where the effects of weightlessness have been demonstrated impressively. These effects are examples for one fundamental principle of Einstein's theory, the "principle of equivalence." It states that at any one point the action of gravity cannot be distinguished from an appropriate acceleration. Along its ellipsoidal orbit around the Sun the Earth is exactly in this balance between acceleration and gravitational attraction.

It is astonishing that despite its very different approach GRT transforms into the Newtonian theory of gravitation, if gravitational fields are weak and velocities are small.

On several points, however, GRT corrects the predictions of Newtonian theory. Up to now GRT has successfully met all critical experimental examinations, in contrast to a number of rivaling theories of gravity proposed over the years. It explains the anomalous shift of the perihelion of the planet Mercury which has been known since 1859. In addition many observations have confirmed the prediction of the deflection of light rays by the Sun's gravitational field, a prediction which was first found true during observations of a solar eclipse in 1919. Many other and much more precise tests have meanwhile succeeded to establish the combined influence of gravity and motion on the rate of clocks, and the delay of radar signals due to the gravitational potentials they cross. All these observations and measurements constitute solid proofs for Einstein's supposition of the curvature of space–time, and the dynamical properties of the metric.

For atoms gravity is completely insignificant, but it gains importance as large masses come into play. The other known long-range force, the electromagnetic force, acts differently on positive and negative charges. Particles with the same charge repel each other, with opposite charge they attract each other. This leads to a screening of the electromagnetic force at large distances. There seems to be no screening effect for gravitation, because negative masses are not known.

2.8.3 Black Hole Basics

Two months after Albert Einstein had published his fundamental paper on GR, the brilliant German physicist Karl Schwarzschild derived at the end of the year 1915 the solution later named after him. The "Schwarzschild Geometry" has become famous as the prototype of a "black hole," although it gives a general description of the curved space–time outside of a spherically symmetric distribution of masses. The space around the Sun is also approximately described by the Schwarzschild geometry, as far as the Sun can be considered as a perfect sphere.

Let us imagine compressing a spherical mass, like the Sun, into a smaller and smaller radius. In Newtonian theory this is possible (at least in our imagination) until the idealization of a point mass with zero radius. Outside of the mass the Schwarzschild solution is valid at all times. But for radii smaller than

$$r_S = \frac{2GM}{c^2}$$

(M is the mass within r_S, G the gravitational constant, c the velocity of light.) the static coordinates, i.e., the coordinates independent of time, which Karl Schwarzschild had used lose their validity. In this thought experiment therefore we do not reach the equivalent of the Newtonian point mass concentrated at $r = 0$, but we are stopped at the "Schwarzschild radius" r_S, and we do not know what happens inside of r_S, until we find a better description. In the course of time theoreticians have explored different coordinate frames which are better suited to describe the inner part at radii less than r_S.

The easiest approach to study this space–time is to analyze the propagation of light. Far away from the Schwarzschild radius the situation is the same as in Minkowski space, but close to the Schwarzschild radius the light cones are being deformed. A light ray emitted at the Schwarzschild radius in any arbitrary direction will be bent so much that it cannot escape to the outside, but must remain caught within this sphere of radius r_S. Light rays emitted within this radius end necessarily in the singularity at $r = 0$ (Fig. 2.27).

Light from the outside can fall into the Schwarzschild radius without any problem, but no light signal can escape.

The Schwarzschild radius thus designates a structure in the space–time, a "horizon" which separates inside and outside inescapably. Since information cannot be transmitted faster than with the velocity of light, the horizon acts like a membrane letting energy and information pass only in one direction, from the outside to the inside.

Inside at $r = 0$ is the singularity, just like the Newtonian point mass. Every mass inside the horizon must necessarily end as a singular point mass according to GRT. Inside the horizon the light cones turn by 90°. It is as if space and time had changed their properties. Inside the horizon space "passes," just as does time in the real world. Nothing can stop the passing away of the distance to the point $r = 0$, everything ends up in this singularity.

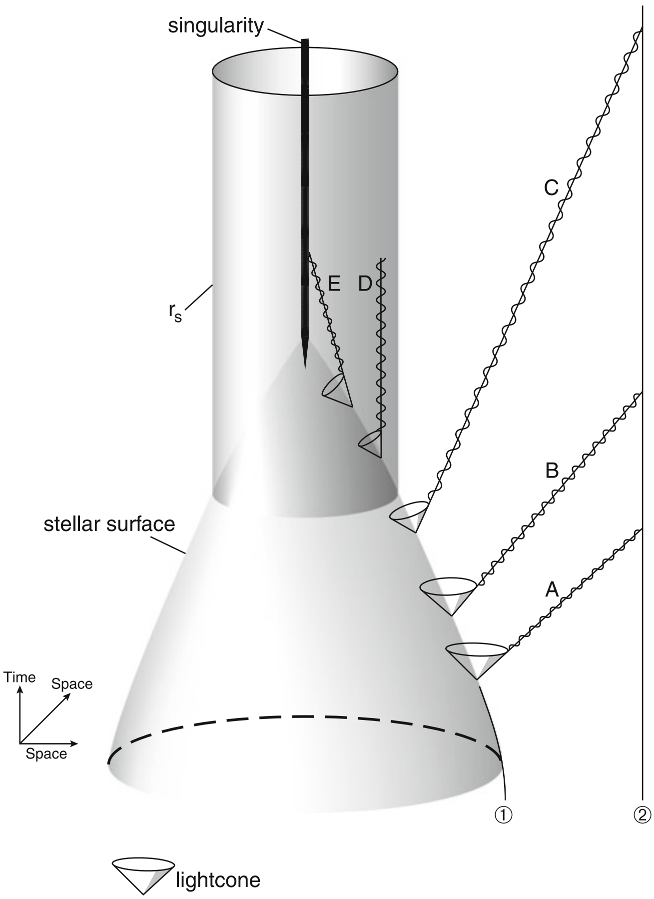

Fig. 2.27 A black hole separates space–time into two regions, inside and outside of the horizon, such that signals cannot propagate from the inside to the outside. Signals emitted at regular time intervals by a source falling into the black hole will be received by an observer outside with ever increasing delays. This time delay grows beyond any finite limit as the Schwarzschild radius is approached

The space–time of the Schwarzschild solution thus is empty, outside and inside the horizon there is no matter except for the point mass *M* in the center. An outside observer measures a gravitational force corresponding to a mass *M* inside. This structure, the point mass *M* surrounded by an horizon r_S is called "black hole," a name invented by the American physicist John Wheeler in a lecture in the year 1968.

(Actually there is a short novel written in 1905 by Hubert von Meyrinck with the title "The black sphere," where the emergence of this all-devouring nothingness is depicted. Meyrinck, however, ascribes the origin of the black hole to the materialization of the thoughts of an officer of the Habsburg K. and K. army.)

A black hole thus is neither a material body, nor does it consist of radiation; it is literally a hole in the space–time. The singularity inside does not have any possibility for a causal connection to an observer outside. Such horizons or causality boundaries are a remarkable result of GRT which may be characteristic for all realistic singularities: A "cosmic censor" covers it, such that no observer can see it.

Stars, planets, or other physical objects are, of course, much more extended than their Schwarzschild radius. The Schwarzschild radius for the Sun is 3 km, for the Earth 0.9 cm. The fact that Sun and Earth are much bigger, shows that the deviation from an Euclidean space is very small in the vicinity of these celestial bodies. Subatomic particles like protons and neutrons are larger than their Schwarzschild radius by a factor of about 10^{39}. This demonstrates that gravity is totally unimportant in the world of elementary particles. The situation is quite different when we look at a neutron star, an object with a mass like the Sun, but with a radius of only 10 km, about three times larger than its Schwarzschild radius.

The structure of neutron stars is definitely deviating from a purely Newtonian gravitational equilibrium. You need to squeeze a neutron star just to about one-third of its size to transform it into a black hole.

Karl Schwarzschild himself was worried by the singular behavior of his solution at the Schwarzschild radius. Therefore he investigated the properties of spheres of constant density within GRT. He was able to show that the radius of such a sphere must always be larger than $9/8\ r_S$. A sphere of this kind is always outside of the Schwarzschild radius.

Because of the enormous concentration of mass in a small volume, the tidal forces in the vicinity of a black hole are also larger than close to a normal star. Thus, a researcher falling into a solar mass black hole will be distorted by the tidal forces, outside of the horizon. Since his feet are more strongly attracted than his head, he will be stretched along the body axis, and compressed from the sides. The tensions occurring are very big – 10,000 to 100,000 g for a black hole of solar mass (g is the acceleration on the Earth's surface) – but decrease in inverse proportion to the square of the mass. The tidal forces near big black holes can be well tolerated, and the fall into the horizon proceeds quite uneventfully, even when the horizon is crossed, until it meets a horrible end at the singularity $r = 0$.

2.8.4 Gravitational Collapse

A big, massive star which has used up the nuclear fuel in its interior contracts more and more under the influence of its own gravitational field, and disappears finally in a singularity after it has crossed the horizon.

For an observer on the stellar surface this happens in the free fall time needed to cross the distance to the horizon, that is for a star of a few solar masses in a few seconds. For an observer outside, however, the stellar surface appears to come closer and closer to the horizon, but never reaches it. The star seems to "freeze" at the Schwarzschild radius. These impressions depend on the propagation of radiation in the vicinity of the horizon. Signals emitted at regular intervals by the bold researcher falling into the black hole are received by the cautious observer outside at steadily growing time intervals. Finally the last flash of light sent out at the horizon is not received by the outside observer – "it reaches him after an infinite time has passed" as relativity theorists like to phrase it (see Figs. 2.25 and 2.27).

All this is correct in principle, but in practice the star becomes invisible quite suddenly. This is due to the fact that the wavelength of the light emitted close to the horizon is received far away strongly red-shifted. The shift to longer wavelengths increases exponentially with the source of light approaching the horizon. The luminosity decays rapidly too. During the time span of one hundred thousandth of a second for a solar mass star ($T \sim r_{s/c} - 10^{-5}(M/M_{\odot})$ seconds is the time needed by light to cross the Schwarzschild radius r_s), the collapsing star becomes invisible.

The Schwarzschild solution depends only on the mass, but almost all stars rotate, and therefore we expect that black holes resulting from the collapse of a star will in general rotate. Compared to the Schwarzschild solution the space–time of a rotating black hole has quite a complex horizon structure. It is also possible, at least in principle, that black holes carry an electric charge. Mass, angular momentum, and charge completely identify a black hole for the outside world. It is amazing that stars despite their rich variety of forms and structures end up in such a simple state. It reminds one of Platon's idea of "ideal bodies."

Can Black Holes Be Observed?

Black holes do not emit radiation, but they can be observed indirectly, if they accrete matter which heats up while falling into the black hole. The radiation emitted by this in-falling matter can be registered.

Black Holes in X-Ray Binary Stars

Observations of the sky in X-rays by satellites uncovered the existence of binary stars, where a star visible in optical light is orbited by an invisible, compact X-ray source. In several cases the astronomers have concluded that the X-ray source must be a black hole. This conclusion rests on estimates of the mass made possible by using the periodic fluctuations of the optical and the X-ray light to obtain accurate measurements of the system parameters. If the mass of the compact X-ray source turns out to be significantly larger than the maximal mass of a neutron star, then it must be a black hole.

The most famous candidate is the X-ray star Cyg X-1 (Cygnus X-1), with a mass between 9 and 16 solar masses, at least three times the mass limit for neutron stars. Cygnus X-1 is very probably a black hole. There are similar arguments for a number of further candidates. Astronomers would be much happier, if they did not have to rely on such indirect proofs, but were able to identify a characteristic property of the radiation – such as a typical time variability – as the unique signature of a black hole. At present such a property is not known.

Black Holes in the Centers of Galaxies

Detailed observations of the central regions of active galaxies and quasars showed evidence for high concentrations of mass and energy, for high velocities in relatively small volumes. Thus the active galaxy M87 has been searched in detail by the Hubble Space Telescope, to name but one example. Its inner region of 500 light-years contains a mass of gas which rotates with a velocity of about 750 km s^{-1}. Such a fast rotation can best be explained as the motion around a black hole of about one billion (10^9) solar masses.

Even at the center of our Milky Way a black hole has been tracked down. Measurements in the infrared part of the spectrum disclosed stellar motions in a small region of 0.3 light-years extent influenced by a central mass of about one million (10^6) solar masses. The assumption that this must be a black hole seems plausible, because any other configuration – a dense star cluster, a giant star – would not be stable, and would anyhow evolve into a black hole within a few million years.

2.8.5 Quantum Theory and Black Holes

The Singularity

The gravitational collapse of a large mass proceeds relentlessly inside the horizon until everything is concentrated in a point-like singularity. Such a singularity, in this case a point mass of infinite density, should not occur in a well-defined theory. GRT leads necessarily to such a state, as a consequence of the overwhelming gravitational force which overcomes all counteracting forces. Thus the gravitational collapse of large masses is a fascinating prediction, but marks also the limits of validity of this theory. In Newton's theory the idealization of the point mass is also used, and the gravitational field of a spherically symmetric mass distribution is described exactly as if all the mass were assembled in the central point. This then would also be a singularity, but in Newton's theory the situation is quite different: Far away the field is like that of a point mass, but close to the masses the finite density of the real distribution is relevant. In GRT, however, the point mass is real, collapse to arbitrarily small volumes is the real fate of large masses. The only redeeming feature is the formation of a horizon which shields the outside world from the singularity.

Is there a way to stop the collapse? When the matter is compressed to larger and larger densities, the classical description must become invalid at some point. Even a very massive, large star will eventually become a small quantum object. Finally we can no longer describe space–time as a classical continuum, where the separation between two points can be arbitrarily small. Space–time must finally transform into a kind of quantized structure, perhaps with a fundamental, smallest length. A theory encompassing such a unification of GRT and quantum theory is not yet in sight, although, as we have already mentioned, it has a name already: "quantum gravity." There is a lot of activity in this field of research.

Supporters of the string theory believe that this theory, or rather the fundamental, somewhat mysterious M-theory, will contain solutions which might correspond to a quantum gravity theory. The basic building blocks of this theory are "strings" or membranes, subatomic small pieces of strings or surfaces, whose vibrations create the world from the vacuum. These vibrations take place in spaces of ten dimensions at least, but in the real four-dimensional space–time six of these dimensions are wound up somehow in tiny manifolds, such that they are not perceived. An illustrative picture might be a piece of straw which from afar looks like a piece of a line, a one-dimensional structure. Closer inspection clearly shows that it is tube-like, a two-dimensional cylindrical surface, if the thickness of the walls is neglected. Looking even closer we recognize the thickness of the walls, that is the complete three-dimensional structure.

Very similar – according to string theory – is the way in which the wound-up extra dimensions show up, when a singularity is approached. Thus, during gravitational collapse for example, the singularity does not occur. Instead the ten-dimensional vibrating string unfolds. It might be like that, and right now we cannot say more about it. The M-theory has obviously an enormous number of solutions, and up to now the ones describing our real world have not been identified.

Hawking Radiation

Instead of waiting for the theory of everything one can try to look for connections between quantum and gravitational physics by investigating quantum fields in a fixed space–time. In 1974 Stephen Hawking found an exciting result: Black holes are not really black, but they emit radiation just as if there were a body with a certain temperature at the Schwarzschild radius.

He examined the behavior of the vacuum state, the state without particles, in the space–time of a black hole. The quantum mechanical vacuum is not a quiet, empty thing, but rather full of activity with pairs of particles and antiparticles being continuously created and annihilated. Now, close to the Schwarzschild horizon there is the possibility that virtual antiparticles pass through this one-way membrane, and disappear. The corresponding particles are left behind, and gain a real existence. A distant observer registers these escaping particles as a thermal radiation.

The temperature is very low for black holes of stellar mass

$$T \sim 10^{-7}(M/M_{\odot})^{-1} Kelvin.$$

But small black holes, if they existed, could reach very high temperatures, and end in a burst of gamma rays. A spectacular event which might even be observable.

By the emission of such a thermal radiation, the so-called Hawking radiation, the black hole loses energy, its mass becomes smaller. The total mass would be radiated away in a time t, with

$$t = 10^{71}(M/M_{\odot})^{3} sec.$$

This is for solar mass black holes many orders of magnitude larger than the age of the universe.

What remains at the end? Only radiation, or some kind of scar in the space–time texture? We do not know, because the real problem, to compute the feedback of the energy loss on the black hole space–time, has not been solved.

2.8.6 Space and Time Arise and Decay

During this short walk through cosmology and astrophysics we have met many curious and remarkable phenomena. In my opinion special attention should be paid to the insight that our commonsense conceptions of space and time are shaken up: Space is not simply fixed and eternal, time is not infinite and uniformly flowing.

In the vicinity of a black hole a part of the space–time is sealed off. The Schwarzschild radius encircles a region which can no longer have contact with the space–time outside. The fall into a black hole ends in a final crunch at the singularity, and time ends there too. Viewed from the outside all clocks falling toward the black hole stop ticking at the Schwarzschild radius – it seems, as if time did not continue there.

The reverse process seems to take place at the big bang. Nothing can fall into this singularity, everything comes out of it. Space and time also have this origin in the initial singularity about 14 billion years ago.

The German philosopher Immanuel Kant has taught us that space and time are categories of our experience, they determine the principal way according to which we order our experiences. We cannot imagine otherwise, than to order them in space and time. Kant obviously had in mind the Newtonian conceptions of absolute space and uniformly flowing, eternal time. In his time these ideas were considered to be self-evident. Modern science has brought a great change. Although it is still true that we order our knowledge in the categories of space and time, the categories themselves have lost their absolute validity. If space–time itself has been created 14 billion years ago, it cannot be an eternally durable category. Thus the categories of our mind are not something given, fixed, and absolute, but they have arisen during the long evolution of the cosmos. These insights gained from physics contradict, as it seems, our intuition, and go beyond our everyday experience: If space and time arise and perish, then there might even be structures beyond space and time. This indication of a transcendent element in the world around us is like a confirmation of Kant. There may be things beyond space and time, but they are not accessible by our experience.

It is difficult for us to believe that there may be an aspect of reality not bound up with time. Our life is dominated by time: Beautiful and awful experiences passing, the before and after, the course of days and years, aging and death are inescapable reality for us.

The changes in our view of space and time do not imply with certainty that our existence reaches beyond time and that there exists really something not encased by space and time. But we can see clearly that there are restrictions imposed on our knowledge, and therefore we can also see the possibility to remove these boundaries in our thoughts.

Thus with the fall of time from its absolute throne, with the fall of the tyrant, who determines all things, and its translation to a quantity which itself must experience changes, the hope grows that – to use the words of the Austrian physicist Erwin Schrödinger "the whole time-table is not meant as seriously, as it appears at first sight" ("dass der ganze Zeitplan doch nicht so ernst gemeint ist, wie es zunächst scheinen mag").

http://www.springer.com/978-3-642-20103-5

The Wondrous Universe
Creation without Creator?
Börner, G.
2011, VII, 184 p. 37 illus., 13 in color., Softcover
ISBN: 978-3-642-20103-5