2 The Vanishing Stars

Were the succession of stars endless, then the background of the sky would present us a uniform luminosity—since there could be absolutely no point, in all that background, at which there would not exist a star.

Edgar Allan Poe, Eureka, 1848

In spite of Giordano Bruno's fate, the limits of the universe continued to occupy the minds of many scientists and philosophers. Is there indeed some ultimate celestial sphere? And if so, what is in that forbidden "room" beyond it? The existence of a final firmament, to which the fixed stars are attached, did in fact answer one rather curious question. Why is the sky dark at night? If there were no such sphere, if instead a world of stars continues on and on, homogeneously, with the same density, forever outward, then every spot in the sky will be filled with shining stars, some closer, some further out, and further yet. Copernicus insisted on a fixed outer sphere with a finite number of stars and thus avoided the problem. Kepler had realized the difficulty and therefore also ruled out the possibility of an infinite universe. Still the question kept reappearing and is today known as Olbers's paradox, after the German astronomer Heinrich Olbers, who formulated it most succinctly in 1823. It is an excellent illustration of how a well-posed question can lead to progress in thinking and understanding. To answer it, however, we first have to address one of the basic issues of physics: what is light?

2.1 The Speed of Light

But what and how great should we take the speed of light to be? Is it instantaneous perhaps, or momentary? Or does it require time, like other movements? Could we assure ourselves by experiment which it may be?

H. Satz, *Ultimate Horizons*, The Frontiers Collection,
DOI: 10.1007/978-3-642-41657-6_2, © Springer-Verlag Berlin Heidelberg 2013

The question had been around for quite a while when Galilei, in his Renaissance treatise on the *Two New Sciences*, made his *alter ego* Salviati ask it. Already Aristotle had complained more than 300 years before Christ that

> Empedocles says that the light from the Sun arrives first in the intervening space before it comes to the eye or reaches the Earth.

He, Aristotle, was sure that this was completely wrong, that “light is not a movement”, and his belief dominated western thinking for almost 2,000 years. The speed of light is infinite—even great scientists and philosophers like Johannes Kepler and René Descartes were more than convinced of that. Descartes said that “it is so certain, that if one could prove it false, I am ready to confess that I know nothing at all of philosophy”.

Galilei, of course, proposed the right way to resolve also this issue: experiment. He even tried it himself, but at that time terrestrial techniques were not up to the task. A distant assistant had to cover and uncover a lamp, and Galilei tried to measure the time it took him to see that. He correctly noted that light did travel faster than sound. But to determine its speed, one needed longer times and hence larger distances, and these were then to be found only in astronomical domains. The problem was, in fact, twofold. Is the speed of light finite, and if so, what is its value?

The first question was answered several decades later by Ole Rømer, a truly multi-talented man from Aarhus in Denmark. His real name would have been Ole Pedersen, but with so many Pedersens around, his father had started to call himself Rømer, after the island of Rømø, where they came from. Ole had studied physics, mathematics and astronomy in Copenhagen and eventually married the daughter of his professor there. In between, he had worked for King Louis XIV in Paris and took part in the design of the fountains of Versailles. After this interlude, he returned to Denmark for an appointment as “royal mathematician”, where he introduced the first national system of weights and measures, as well as the Gregorian calender. And besides all this, he became Chief of the Copenhagen Police, responsible for the installation of the first street lights there. In Paris, he had worked as an assistant for the astronomer Giovanni Domenico Cassini, and Cassini had made a remarkable observation. The planet Jupiter, fifth around the Sun and largest of all, had a Moon, called Io (named after a nymph seduced by the Roman god Jupiter, in his Greek avatar form of Zeus), which circled around it approximately once every 42 h, in contrast to the 28 days our earthly Moon takes for its orbit. That meant that seen from the Earth, there would be many “eclispses” of Io at any stage of the Earth ’s orbit around the Sun; the geometry is shown in Fig. 2.1. One could thus measure the time at which Io disappears behind Jupiter, and do this for a series of eclipses. This provided a determination of the time between successive eclipses, giving a prediction for the next.

And the striking observation first made by Cassini was that the onset of an eclipse fell more and more *behind schedule* the further away the Earth was from Jupiter. Cassini was not sure, but thought that perhaps *light takes some time to reach us*. Eventually, he seems to have rejected this conclusion. Rømer, instead, combined a number of different measurements, extrapolated them to eliminate interference

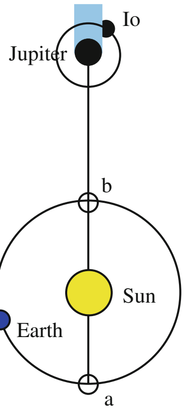

Fig. 2.1 Ole Rømer's basis for the determination of the speed of light

effects, and found that the delay in time of eclipse onsets seen from the point of greatest Earth—Jupiter separation (point a) compared to those seen from the smallest distance (point b) was about 22 min. From this he now concluded that the speed of light is indeed finite and that the 22 min is the time it needs to traverse the diameter of the orbit of the Earth around the Sun.

To obtain the actual value of the speed of light from these measurements, the size of the orbit of the Earth around the Sun had to be known. How far did light have to travel in these 22 min it took between the two extreme points? This distance, divided by 22 min, would then be the speed of light. The relevant information to determine the distance from Earth to Sun was actually available at that time, due mainly to the studies of Cassini. The first numerical value for the speed of light, however, was apparently obtained by the Dutch physicist Christiaan Huygens in 1678, two years after Ole Rømer had announced his conclusions.

Kepler had, in this "third law" of celestial motion, concluded that the time for a planet to orbit the Sun was related to the distance between this planet and the Sun; from this, the relative distances of all planets from the Sun were known. In particular, the distance between Mars and the Sun was found to be about 1.5 times that of the Earth and the Sun. To arrive at an actual value for the Earth–Sun distance, some astronomical distance had to be measured in terrestrial units, and this "calibration" had in fact been carried out by Cassini and his collaborator Jean Richer. They measured simultaneously the position of Mars relative to the fixed star background, Cassini in Paris and Richer in French Guiana. This gave them an angle and a known terrestrial distance, the 4,000 km between Paris and Guiana, and geometry then determined the distance between Mars and Earth. They found it to be about 73 million km. At the point of closest approach of Mars and Earth, that led to 146 million km for the distance between Earth and Sun. Since light travelled, according to Rømer, twice that distance in 22 min, Huygens noted that its speed must be about 220,000 km/s. This result, obtained over 300 years ago by a combination of logical thinking, abstraction and rudimentary measurements, is certainly one of the great achievements of the

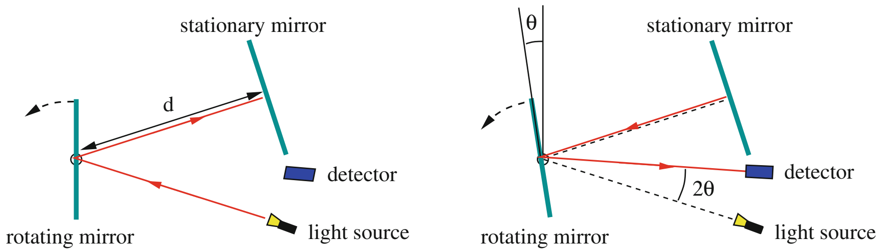

Fig. 2.2 The mirror arrangement used by Fizeau and Foucault for a terrestrial determination of the speed of light

human mind; it is only about 25 % too low according to today's precision value, measured using radio signals between space craft positioned in the solar system.

The first terrestrial measurements were carried out in Paris by Hippolyte Fizeau and Léon Foucault around 1850, improving the attempt of Galileo by reflecting light in a clever arrangement of mirrors. Foucault, with his celebrated pendulum, had in fact also provided for the first time direct proof of the rotation of the Earth around its axis. But he now modified an older apparatus devised by Fizeau to measure on Earth the time light needs to go from one point to another. The set-up is illustrated in Fig. 2.2. Two mirrors are placed as far apart as possible, at a distance d; they now play the role of Galileo and his assistant. One of the two mirrors is rotating at a speed ω, the other is stationary. A beam of light is directed at the rotating mirror, and that reflects it to the stationary one. When it now returns to the rotating mirror, it has travelled between the two mirrors a total distance $2d$. During the travel time, the rotating mirror has turned an angle θ, so it reflects the beam back not at the source of light, but at a detector placed at an angle 2θ away. Knowing d, θ and the rotation speed ω gives the speed of light as $c = 2d\omega/\theta$. The results of Fizeau and Foucault were within 1 % of the present value, 299,792,458 km/s.

So, the light from the Sun did have to travel through the intermediate space *before* reaching the Earth, as Empedokles had supposed 2,500 years ago. But what is this light travelling through what we think is empty space? What is it that is moving at 300,000 km/s?

This question led to another basic and universal phenomenon of the inanimate world: electromagnetism. Initially, electricity and magnetism entered as two quite separate and distinct features. The first appearance of electricity in the life of humans was lightning, for a long time thought to express the wrath of the gods in a frightful way, and beyond human understanding. A more mundane version was observed by the ancient Egyptians, more than 3,000 years ago; they were familiar with electric fish which could produce remarkable bolts of electricity to stun their prey. This source of electricity was supposedly used already in those days for the treatment of neural illnesses. In ancient Greece, it was noted that rubbing amber with a catskin made it attract feathers and other light objects—and it was this feature that gave the name to the mysterious force, with *elektron* as the word for amber in ancient Greek.

But it took still more than 1,500 years until these various and seemingly unrelated phenomena began to be understood, and only in the last 100 years has electricity dramatically changed human life.

Magnetism was more well-defined from the beginning. Several millennia ago it was noticed in China that a certain kind of stone attracts iron, and if suspended by a string, it would orient itself along a north–south axis. Making use of this, the ancient Chinese constructed the first magnetic compass for navigation. In ancient Greece, Thales of Milos described the effect, and since the stones showing such behavior there came from a province called Magnesia, he called it magnetic. In English, it became "leadstone" and finally "lodestone", presumably because it could be used to lead travellers in the desired direction.

Both electricity and magnetism became part of natural science only less than 300 years ago. It was discovered that there exist two different forms of electricity, arbitrarily denoted as *positive* and *negative*; each form could be produced by rubbing, for example, and each kind can exist on its own. If two metal balls were prepared to have different "charges", like and like repelled each other, while positive and negative showed attraction—both by invisible means across the distance of their separation. Charles Augustin de Coulomb in France showed in 1785 that these reactions followed a law very similar to that proposed by Newton for the equally invisible action at a distance provided by gravity (Fig. 2.3). Coulomb's law gives for the electric force

$$F = K\frac{q_1 q_2}{r^2},$$

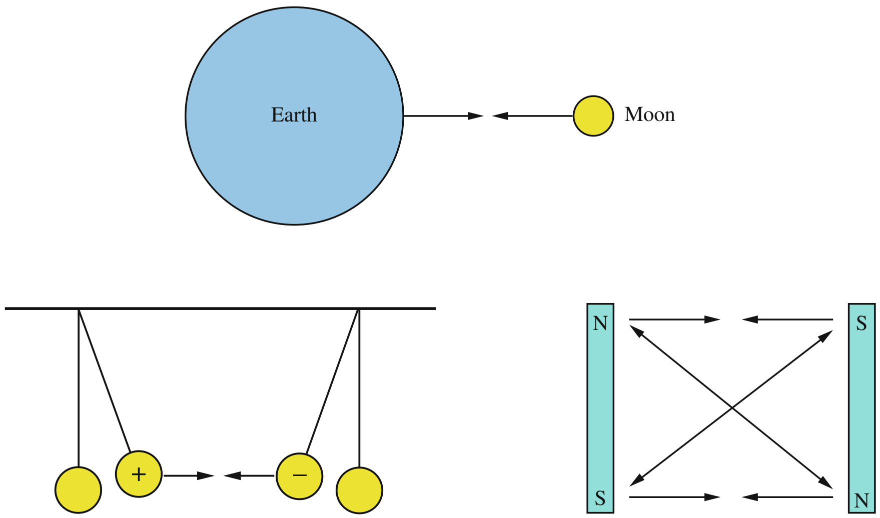

Fig. 2.3 Three forms of action at a distance: the gravitational attraction between the Earth and the Moon, the electric attraction between positive and negative charges, and the magnetic attraction between opposite poles, accompanied by the repulsion between like poles

where q_1 and q_2 measure the amount of charge on each ball and r their separation; the constant K plays the role of Newton's universal constant of gravitation, except that it is now positive (repulsion) for like and negative (attraction) for unlike charges.

While positive and negative electric charges could exist independently and could be produced separately, magnets were curious animals. They had a north pole and a south pole, and given two magnets, north and south attracted each other, while north/north or south/south meant repulsion. But there was no way to get just one pole. Cut a magnet in two in the middle, and you had two new magnets, each with its north and its south pole. And until today, physicists are still wondering if there isn't some way to create a monopole. The magnetic force was not quite of the inverse square form encountered in Coulomb's law of electric interaction or Newton's law of gravity, since each pair of magnets experienced both attraction, between the opposite poles, and repulsion, between the equal poles. Nevertheless, the interaction between two magnets, as well as that between metals and magnets, was again by some invisible means over the distance of separation.

So both electric and magnetic interactions showed a mysterious feature already encountered in the case of gravitation: an interaction over a distance, without any apparent connection between the interacting objects. How such an interaction could arise was something that had puzzled people at all times. Was there some invisible medium filling all of space to provide a connection? The beginning of an answer was provided by the British physicist Michael Faraday, who proposed that each charge would be surrounded by an electric field, radiating out starlike *lines of force* emerging from the source in all directions (Fig. 2.4). And this field would "feel" the presence of other charges and react accordingly: the lines of force would bend either towards the other charge or away from it, depending on the sign.

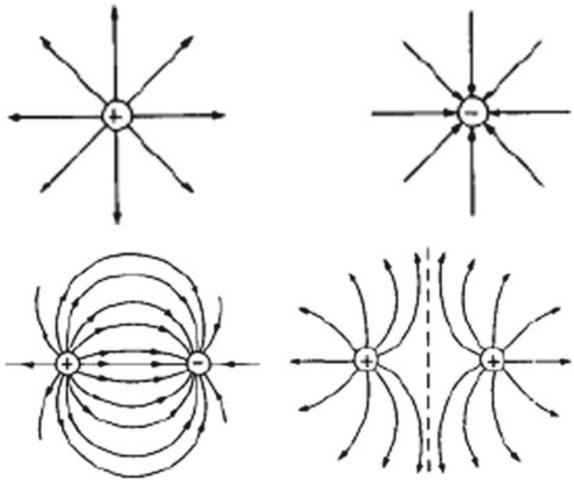

Fig. 2.4 Lines of force emerging from isolated sources of positive and negative electricity (*top*) and from neighboring like and unlike sources (*bottom*)

Moreover, in the early 1800s, Hans Christian Oersted in Copenhagen discovered that there was a strange connection between electricity and magnetism. It was known that certain materials—today's *conductors* —allow a rapid spreading of electric charge: they result in the flow of an electric current between opposite charges, forming an electric circuit. Now Oersted observed that a magnet would align itself in a direction orthogonal to the line of current flow, as if the current had created magnetic lines of force around its flow axis. So one could imagine unending lines of force corresponding to magnetic fields, closed loops having neither beginning nor end. This would explain why cutting a magnet in two simply produced two magnets, and did not yield an isolated pole.

In the course of the nineteenth century, extensive studies showed that electric and magnetic forces are indeed closely intertwined: electric currents produced magnetic fields and moving magnets *induce* electric currents. This suggested a unified theory of electromagnetic fields, and it was the great British physicist James Clerk Maxwell who created it, with his famous equations. Through Maxwell, electricity and magnetism were unified to electromagnetism. And in addition, he provided the basis for an understanding of how the interaction of electromagnetic sources could occur over distances. Maxwell showed that a changing electric field generates a magnetic field, just as a changing magnetic field would through induction create an electric field. So the combination of the two, electromagnetic fields, now gained an independent existence, without the need of currents or magnets. And one simple solution of Maxwell's equations was that of travelling waves, like an excitation travelling down a string, or a wave travelling across a pool of water. The action over a distance could thus occur through the exchange of electromagnetic signals in the form of such waves. They propagate through space at a fixed speed, which can be measured and was found to be the familiar speed of light. The fundamental question *what is light?* was therefore now answered: it is an electromagnetic wave travelling through space, and the different colors of light simply correspond to different possible wavelengths. Beyond the range of visible light, we recognize today electromagnetic radiation on both sides, with radio waves of longer wavelength (beyond the infrared) and X-rays of shorter wavelength (beyond the ultraviolet). And in a way, it also answered the question of how distant charges could interact: through the exchange of an electromagnetic signal.

But the answer was not really complete. If distant charges communicated by electromagnetic waves travelling between them: *what was being excited to form such waves?* In our everyday world, it can be a string, the surface of water, the density of air. But what is it in empty space that is vibrating? And so the *ether* entered the world of physics, an invisible medium filling all of the so-called empty space. This satisfied those who thought that truly empty space was "unnatural", such as the French philosopher Blaise Pascal, who believed that "nature abhors a vacuum". When Evangelista Torricelli in Italy succeeded in removing all the air from a vessel, Pascal noted that the absence of air does not mean empty. For light, the ether was first introduced by Robert Hooke, in 1665; he pictured a pulse of light like a stone thrown into a pool of water, with concentric waves spreading out. Just as a tsunami wave is formed by an earthquake at the bottom of the sea far out in the ocean and then travels towards some shore, so a change in the electromagnetic state somewhere would be communicated across space to a distant receiver in the form of an electromagnetic tsunami wave in the ether. This ether turned out to be one of the most-travelled dead-end roads of physics. From the time of Hooke to the time of Einstein, a great number of well-known physicists tried their hand at it, and always with rather limited success. Is the ether stationary, or is it comoving with stars? Is there an ether-wind due to the Earth moving through it? Is matter perhaps only a form of vortices in the ether? The presence of an ether resolved the puzzle of an action at a distance, but to do so, it had to be a material substance and yet, at the same time, not seriously affect the motion of the stars. One of the most celebrated experiments to find it was carried out in the 1880s by the American physicists Albert Michelson and Edward Morley. If light was travelling through the ether everywhere at its fixed speed, then it would have to be slower if measured in the direction of the Earth's motion than if perpendicular to it. They devised an interferometer constructed such as to have two beams of light, one along and one perpendicular to the motion of the Earth, travel the same distance and by means of a mirror arrangement meet again at a given point (see Fig. 2.5) The slowing effect of the Earth's motion would throw them out of phase, so that a valley in the wave of one would hit a peak in that of the other beam, causing interference. Much to their frustration, Michelson and Morley found no effect whatsoever; all waves arrived completely in phase. No matter how they positioned their apparatus, the speed of light seemed always to be exactly the same. So there was no evidence for any form of ether, and after numerous attempts to find a way out, it was finally banned from physics by Albert Einstein, almost 20 years later. It is now definitely ruled out, at least as far as electromagnetism is concerned.

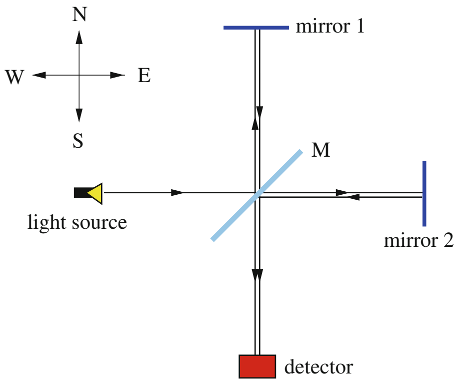

Fig. 2.5 The Michelson–Morley experiment to detect the presence of an ether. A beam of light is directed at a partially transmitting mirror M, from where part of it is reflected to mirror 1 and then on to the detector, another part to mirror 2 and then to the detector. The direction from mirror 1 to the detector is chosen to be north-south, that from the light source to mirror 2 east-west, and both mirrors 1 and 2 are equidistant from the central mirror M. The motion of the Earth (*east-west*) relative to the ether was predicted to modify the speed of the corresponding light beam and thereby lead to interference patterns between the two beams arriving at the detector

However, even today it is not so clear what the role of a cosmological constant or dark energy is; we shall return to these somewhat ether-like ideas later on.

Maxwell's equations implied a unique speed for electromagnetic waves travelling through empty space, the universal speed of light. This is in fact much more dramatic than it seems at first sight: such a behavior is simply not in accord with our everyday experience. A car moving at 100 km/h, as seen by a stationary observer, has a relative speed of only 70 km/h for someone moving in the same direction at 30 km/h. And two cars, both travelling at 100 km/h in the same direction, are not moving at all relative to each other. If someone in the compartment of a moving train drops a coin, it falls straight down: train, passenger and coin, though all are travelling at high speed for an observer on the ground, are at rest relative to each other. Light is not like that. If a stationary and a moving observer measure the same beam of light, they both find the same value for its speed. No matter how fast you move, the speed of light you measure is always that 300,000 km/s. By moving faster, you can neither start to catch up with a light beam, nor run away from it. And ten different observers, all moving at different speeds, find that, although their relative speeds differ, that of a given light beam is always the same universal value. In the framework in which Newton formulated his laws, this was simply impossible. In a fixed space with a universal time, the speed of light would change for observers moving at different speeds. To make a constant speed of light possible, the ideas of space and time had to be fundamentally modified. To keep a universal speed of light, the scales for distance and time must become dependent on the observer. Let me measure the speed of light in a laboratory here on Earth, and an astronaut measures it in a space ship moving at high speed relative to the Earth: if we both get the same result, than his standard meter and his standard second, as seen by me here on Earth, must have taken on different values than mine—and they do. The resulting milestone in physics was Albert Einstein's theory of relativity, more exactly, the *special* theory of relativity. The "special" is an *a posteriori* modification, indicating that it holds in a restricted spatial region of the universe only. The extension to the entire cosmos, including the role of gravity, followed 10 years later with the *general* theory of relativity, and again it was Einstein who did it.

To formulate his special theory of relativity, Einstein combined a principle proposed by Galileo Galilei 400 years earlier with the recently discovered universal speed of light. Galileo had insisted that the laws of physics be the same for all observers in uniform motion relative to each other. In other words, if I measure the time it takes a stone to fall to the ground from a height of one meter, once in the laboratory and once on a high speed train, the results should be identical. Einstein realized that if this was to hold and at the same time a universal speed of light was to be maintained for all observers in uniform relative motion, our ideas of space and time would have to be modified, space and time would have to be related, and their scales have to depend on the speed of the observer (see Box 1). In Newton's world, there was a unique time, the same everywhere, and one could talk about two events occurring at the same time. In a relativistic world, synchronization over large distances is not possible, and what is first for one observer, may be later for another.

Another striking result of relativity theory was the conclusion that no material body could ever move at the speed of light. According to Newton's law of force, an increase of force must increase the acceleration of a mass and hence eventually bring its speed to arbitrarily high values, faster than the speed of light. Einstein showed that in the regime in which relativistic effects cannot be neglected, that is, at speeds lower but comparable to that of light, Newton's law becomes modified. Only part of the force serves to increase the speed; an ever larger fraction goes into increasing the mass, the inertia of the accelerated body. In our everyday world, the speeds encountered are so far below that of light that we can safely ignore the speed corrections and work with a speed-independent inertial mass. But in modern high-energy particle accelerators, such as the Large Hadron Collider at the European Laboratory for Nuclear Research CERN in Geneva, Switzerland, one brings protons to speeds 95 % of the speed of light, and then the effective mass of these particles is more than three times their mass at rest. And so it is evident that we can never bring a material body to move at the speed of light—it would require an infinite force to do that. No massive object can ever catch up with a beam of light in empty space; light remains the fastest agent in the universe.

Box 1. Relativistic Motion

If an observer moving in a spaceship at a high speed v with respect to a laboratory on Earth finds that the speed of light is the same as ours, it must mean that from our point of view his length measure is shorter than ours, or his clock runs slower than ours, or both. Actually, it is indeed both: a given length d_0, a standard meter, has that value for us here as well as for the observer in his moving space ship. But his moving meter stick, as seen by us, becomes shortened to the length d,

$$d = d_0\sqrt{1-(v/c)^2},$$

where c denotes again the speed of light. And a fixed time interval t_0 on the spaceship clock, if we measure it from here on Earth, appears dilated to become to a longer interval t,

$$t = \frac{t_0}{\sqrt{1-(v/c)^2}}.$$

Evidently, the faster the space ship moves, the greater is the effect, both in the contraction of the length scales and the dilation of the time scales.

As a consequence, Newton's law of force becomes modified as well; it now reads

$$F = \frac{m_0}{\sqrt{1-(v/c)^2}}a,$$

so that the inertial mass m_0 of a body at rest is at speed v increased to

$$m = \frac{m_0}{\sqrt{1-(v/c)^2}}.$$

At low speed, as long as we can ignore the $(v/c)^2$, we recover both the speed-independent inertial mass m_0 and Newton's force law $F = m_0 a$.

If we consider the force F to be gravity, we see from the relativistic form of Newton's law that the inertial mass of a body, i.e., its resistance to a force, is not its rest mass, but rather a mass including the energy of motion. Einstein formulated this in his celebrated relation between mass and energy,

$$E = mc^2,$$

which means in particular that energy offers an inertial resistance to any force. Even photons, which have no rest mass, will thus be affected by gravity as if they had a mass determined by their energy. So we can weigh the photons trapped in a container: an empty container is lighter than one containing a gas of photons.

So we now know that the light from the stars we see today has been travelling for many years, waves of electromagnetic energy moving through an empty space containing no ether, at a speed of some 300,000 km/s, no matter who measured it. We are therefore prepared to return to the puzzle we had started with.

2.2 Why Is the Sky Dark at Night?

The paradox is today named after Heinrich Olbers; he was not the first to realize it, Kepler did earlier and concluded that the succession of stars is not endless. With Edgar Allan Poe, the problem entered the literary world, leading to pictures that a century later became science, such as an expanding universe starting from a Big Bang. As an earthly illustration of the problem, one can consider an infinite forest: wherever you look horizontally, your line of vision hits a tree. Olbers, in 1823, did state most clearly the assumptions which had led to the paradox:

- The universe is infinite in all directions and has existed forever as it is now.
- The stars are distributed with the same density throughout the universe, they have existed forever, and they have a finite size and brightness.

Given these conditions, the whole sky should be as bright as a typical star; it should never get dark at night. So something must be wrong somewhere, and that something leads us directly to the forefront of modern cosmology and its view of the origin of the universe.

If the age of the universe is finite, if there was a Big Bang starting everything a certain number of years ago, then the universe we can see today will also be of finite size, because light has only had those years to travel. To be sure, the numbers are huge, but they are not infinite. Moreover, the stars had to form sometime after the Big Bang, so their number is also finite. In other words, a finite age of the universe allows us to see only a finite spatial part of it, and in that part only a finite number of stars can have appeared since the Big Bang. That is why the sky is dark at night—a late answer to Heinrich Olbers, requiring both a finite speed of light and a Big Bang origin of the universe. A simple question can lead you a long way…

But how can we be sure that this view of things is really correct? The origin of the universe, in fact the question whether it has an origin, has been the subject of much dispute, scientific, philosophical and religious. There are two main reasons why today most scientists tend to believe in the Big Bang theory—but let us approach them slowly and step by step.

A well-known effect in the physics of everyday phenomena is that the pitch of a sound you hear is modified if the source of the sound is moving. The sound of a race car engine seems higher pitched as the car approaches and lower as it moves away, leading to a characteristic tonal flip as it moves past you. In earlier days, the change in tone of the whistle of a passing railroad engine was the typically cited example. The phenomenon is known as the Doppler effect, after the Austrian physicist Christian Doppler. The tone you hear is caused by sound waves of a certain wavelength, and when the source of the sound approaches you, the distance between wave peaks, the wavelength, is shortened, giving a higher sound, and when it moves away, it becomes longer and hence results in a lower sound. The same "Doppler effect" also occurs for light waves, so that one can in fact check if a given far-away star is stationary or moving. Stars emit light of certain characteristic wavelengths ("spectral lines"), and if this light is Doppler-shifted when it arrives at the telescope on Earth, its source must be moving. Let's say a star is emitting light of a fixed wavelength λ_0, as measured by an observer stationed on that star. For an observer moving away from the star with a speed v, that light will appear to have a longer wavelength $\lambda = \lambda_0/\sqrt{1-(v/c)^2}$, i.e., it will be shifted in the direction from blue towards red, it will experience a *redshift*.

The American astronomer Edwin Hubble, working in the 1920s at the Mount Wilson Observatory in California, had studied the light from very distant stars. From measurements of redshifts it was already known that they all seem to be moving away from us at different speeds. Hubble made the striking observation that the further away they are, the faster they recede. The Doppler shift, and hence the speed of the stars' motion, was rather well measurable—the crucial factor for reaching Hubble's conclusion was the determination of the distance of the stars in question. To measure the distance of fairly nearby objects in the sky, such as planets, one could use the parallax method employed by Cassini and Richer to determine the distance between Mars and the Earth. Howevever, for the very remote stars Hubble was after, the parallax angle became for too minute to be measurable. The solution came through the extension of a very simple phenomenon. The brightness of a given light source decreases the further one is away from it. Since light is emitted spherically from its source, the light incident on a given surface becomes less and less with distance. The size of the spheres grows as d^2, with d denoting that distance, and therefore the light per area decreases as $1/d^2$. So if we know the original brightness of the source and its apparent brightness at some distance, then the difference between the two measurements determines d. Now it so happened that the inherent brightness of the stars Hubble was studying, the so-called Cepheid variables, had recently been determined; they were what astronomers today call *standard candles*. Measuring their apparent luminosity as observed at Mount Wilson, Hubble had at least a good estimate of their distance, enough to show him that their speed of recession v became ever greater, the further they were from Earth, with d measuring that distance. The law $v = H_0 d$ was named after him, as was the crucial constant H_0. By today's measurement, his value of H_0 was off a bit, but the idea was right and changed our view of the universe. In fact, no matter where he looked, the stars appeared to move away in every direction, so it seemed that the whole universe was expanding. Could that be the case? In Box 2 we look in a little more detail why one might think that.

Box 2. The Expansion of Space

To simplify matters, we take space to have only two dimensions instead of three, a "flat" world. Consider three stars in this world, numbered 1–3, positioned at an arbitrary starting time $t = 0$ as shown in Fig. 2.6, with a separation distance d_0 between 1 and 2, as well as between 2 and 3.

Now let us assume that the space in this world expands with time t by a factor R_t in each direction, so that any distance s_0 at $t = 0$ becomes $s_t = R_t s_0$ at time t. The separation between stars 1 and 2 thus becomes $d_t = R_t d_0$, and so their speed of separation is

$$v_t(12) = \frac{d_t - d_0}{t} = \frac{(R_t - 1)}{t} d_0 = H_t d_0,$$

defining $H_t = (R_t - 1)/t$ as our "Hubble" constant. The relation tells us that the rate of separation grows with the initial separation distance d_0. To check that this is really true, we can look at the speed of separation of points 1 and 3, which are initially further apart, namely $r_0 = \sqrt{2} d_0$, as obtained from the triangle relation $r_0^2 = d_0^2 + d_0^2$. The rate of separation of 1 and 3 thus becomes

$$v_t(13) = \frac{r_t - r_0}{t} = \frac{(R_t - 1)}{t} r_0 = H_t r_0 = \sqrt{2} H_t d_0 \simeq 1.4 H_t d_0.$$

The separation velocity is thus a factor 1.4 larger than that between the closer stars 1 and 2. We have so far not said how the expansion of space takes place. If it happens at a constant rate, with $H_t = H_0 t + 1$, we get the time-independent form

$$v = H_0 d$$

of what is now known as *Hubble's law*, with H_0 for the *Hubble constant*. From Fig. 2.6 it is also directly evident that stars 1 and 3, compared to 1 and 2, have to separate by a larger distance in the same time interval and hence must have a higher speed of separation.

At this point, we can also clarify a little what is meant by the *acceleration* of the expansion. The crucial feature is the *scale factor* R_t, defining how much a meter stick expands in a given time t. For $R_t = H_0 t + 1$, the expansion rate is constant: the stick expands in one minute the same amount now as next year.

If the expansion increases with time, for an *accelerating expansion*, the meter stick will grow more in one minute next year than it does now—or less, for a decelerating expansion.

The same forms as discussed here in two dimensions hold of course as well in a three-dimensional space.

Hubble's discovery came really at a very opportune moment. The most up-to-date theory of the universe had just appeared at that time, in 1916: Albert Einstein's general theory of relativity, linking the effect of the force of gravity to the nature of space and time. A ball tied to a string will fly in a circle—but if you only look at its motion, it could just as well be rolling freely in a curved container. The role of the force can thus be replaced by force-free motion in a curved space. Near massive stellar objects, such as the Sun, the force of gravity would in this way distort the surrounding space to such an extent that even a ray of light passing near it would be deflected from its straight-line path. Einstein's theory was tested in celebrated observations during a solar eclipse in 1919, carried out by the British astronomer Arthur Eddington and his collaborators, and these showed that the positions of stars whose light passed close to the Sun appeared in fact shifted by the amount predicted by Einstein, bringing him world-wide aclaim. However, at the time he formulated his theory of gravitation, the general belief was that the universe was static, neither expanding nor contracting, and so Einstein needed some force to counteract the attractive force of gravity acting on all the matter in the universe. For this, there was no immediate candidate, and the problem has remained somewhat enigmatic until today. Einstein reluctantly solved it by introducing a "cosmological fluid", filling the entire universe uniformly and providing the pressure needed to balance gravity. It had to have rather strange properties—not affecting any phenomena in the universe, other than gravity, so that it remained undetectable in all other ways. And it had to be tuned very precisely in order to just balance gravity. In a sense, it was a late counterpart of the ether introduced earlier to provide a medium for electromagnetic waves, and this presumably made it particularly undesirable to Einstein. And when Hubble discovered that the universe was in fact expanding, Einstein called his introduction of a cosmological constant, as the fluid is now generally denoted, his biggest blunder. Had he stuck to his original equations, without such a constant, he

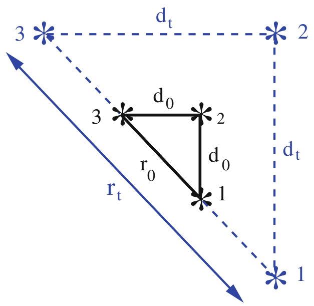

Fig. 2.6 The separation of stars due to the expansion of space, starting from a given initial time $t = 0$ (*black*) to a final time t (*blue*)

could have in fact *predicted* the expansion of the universe before it was discovered. Today, cosmologists are not so sure if it really was a blunder—*dark energy*, which we will encounter later in the context of an inflation scenario for the Big Bang, this dark energy may well turn out to be the modern version of Einstein's cosmological constant, or even of the ether of still earlier times.

In any case, in 1922 a Russian theorist, Alexander Friedmann, presented a general solution of Einstein's equations and showed that they can readily accommodate expanding or contracting universes. And when Hubble a little later found his expansion, the scene was set.

2.3 The Big Bang

The theory itself was initiated in 1927 by Georges Lemaitre, who had studied mathematics and physics at the University of Louvain in Belgium and at the same time prepared for Catholic priesthood; with success on both counts: he received his doctorate in physics in 1920 and was ordained as a priest in 1923. In 1926, when Einstein's equations had just been seen to describe so well the forces in and the structure of the universe, Lemaitre independently derived Friedmann's expanding solution and used it to account for the observations of Hubble: he concluded that our visible universe is continuously expanding. Looking the other direction in time, it must then have originated in a very dense, hot, energetic "primordial medium", which led to the creation of our world. For the Catholic priest Lemaitre, such a creation must have seemed very natural, even though it was a long way away from the dogma applied to Giordano Bruno or Galileo Galilei. But Einstein apparently was not so happy with the results of Lemaitre; "your calculations are correct, but your physics is abominable", he was supposed to have written to him.

Nevetheless, over the years the Big Bang theory continued to gain support, and the perhaps decisive step came in 1964, when the American astronomers Arno Penzias and Robert Wilson discovered what is now known as the cosmic background radiation. It is present throughout the universe as a direct relic of the Big Bang, and it can be measured in the different regions of the sky. Its discovery is one of the truly serendipitous findings of science. Penzias and Wilson were working for the Bell Telephone Company, and they were trying to establish a viable method of microwave communication, by reflecting such signals off high-up balloon satellites. This required the elimination of all other interfering sources of radiation, up to a remarkable precision. Even the detector was cooled to a temperature of a few kelvin, to prevent its "heat" from producing radiation. And when they had eliminated all known sources, including bird droppings on the antennas, there still remained a mysterious background radiation of some three kelvin. It was there day and night, and in all directions. From some friends they heard that in nearby Princeton University, Robert Dicke and collaborators were finishing work on background radiation produced by and remaining from the Big Bang. Penzias and Wilson got in touch with them, discussed their findings and concluded that they had indeed found this left-over flash of the Big Bang. Their work was published in 1965 in *Astrophysical Journal Letters*, in the same issue as the theoretical work of Robert Dicke, Jim Peebles and David Wilkinson, predicting that a form of primordial light should still exist today.

So there is more to consider than just the light from the stars. While the Big Bang, in the absence of air, could of course not really "bang", it did "flash", leading to the emission of light, and this light is still there as the microwave background radiation observed by Penzias and Wilson. The primordial matter initially was a medium of interacting constituents, a plasma of quarks, electrons, photons and more. Eventually, as the medium expanded and cooled, the quarks combined to form protons and neutrons, and these in turn combined with electrons to form electrically neutral atoms. From this time on, from the decoupling era, about 300,000 years after the Big Bang, the photons were "on their own"; in the absence of any charged constituents, they no longer interacted with the medium, and they don't interact with each other. From their point of view, the universe contained nothing but light passing freely into the expanding space. From our point of view, the photons of the microwave background radiation are the most primordial signals of the Big Bang we can ever get. Before decoupling, the plasma of charged constituents was opaque to light, so from this plasma we cannot get any direct information. The time of decoupling, of the formation of electrically neutral atoms, is thus for us an ultimate horizon in time—there is no way we can get any direct information from earlier times.

When the microwave background radiation was emitted, that is, when the photons became decoupled from any matter, they formed a gas of an effective temperature of about 3,000 K. As a result, the wavelength of the radiation was in the yellow part of our spectrum, so that then the sky was not dark at all—it was in fact bright yellow. But the universe kept on expanding, by about a factor 1,000 since the age of formation of atoms. Since its volume increased, its density of energy became lower and lower, and this in turn meant that its effective average temperature also decreased. Through the expansion, the hot universe of the decoupling era has by today cooled down to about 3 K. As a result, the wavelength of the radiation became longer and longer, so that with about 7 cm it is now in the microwave region, far below the visible range. In a way then, the sky is dark at night for us only because we cannot see this microwave radiation remaining from the Big Bang. If we could put on the right kind of glasses, we could see the glow of the sky at night…a glow not of stars everywhere, but the afterglow of the Big Bang itself.

At this point we should note that the light of the stars is, of course, also affected by the expansion of the universe. The Doppler effect that we mentioned above will "redshift" that light, move it to ever longer wavelengths. So up there, in addition to the cosmic background radiation, there is more light than just that of the stars we see. The stars that are moving away from us emit light of wavelengths beyond our visibility range—again, we would have to put on special glasses to see the light from all those stars pushed away from us at an increasing rate by the expansion of space. This redshift is thus an additional reason for the darkness of the night sky.

However, the afterglow of the Big Bang also leads to a striking problem. The microwave radiation we receive today from different regions of the sky was emitted in the decoupling era from regions of the universe that had no causal communication, which were outside each other's event horizon. The reason for this is that decoupling occurred so early in the evolution of the universe and hence so long ago, with immense expansion since then. Two markers separated by a distance of 1 km appear to an observer 1 km away from each marker to subtend an angle of 60° (see Fig. 2.7). For an observer 10 km away from each marker, the angle has decreased to only a little more than 10°. At decoupling, only regions separated by no more than 300,000 light-years could communicate with each other, and if they are now 10^{10} light-years away, they appear to us only some fraction of a degree apart in the sky. We have here for the moment neglected the expansion of the universe, which additionally enhances the effect. In other words, if we measure the microwave background radiation at a certain angle in the sky, and then at another angle only a few degrees away, the sources of the two radiation measurements had no chance to communicate at the emission time. So why do both show the same temperature? The microwave radiation we observe was emitted from millions of sources, of spatial regions, which up to decoupling had no way to "tune" their radiation. It is like a gigantic orchestra, without a conductor and with many, many musicians who have no possibility of getting in tune—yet they all end up playing the same melody. If the decoupling of photons and matter, due to the formation of electrically neutral atoms, had occurred at different times in different regions, the temperature of the background radiation should be correspondingly different. But all regions behaved as if some imaginary omnipotent conductor had lowered his baton and indicated "decouple now". This *horizon problem* is one of the big puzzles of today's cosmology, and it is not really resolved to everyone's satisfaction, in spite of some very interesting proposals. We will soon have a look at one of them in a little more detail.

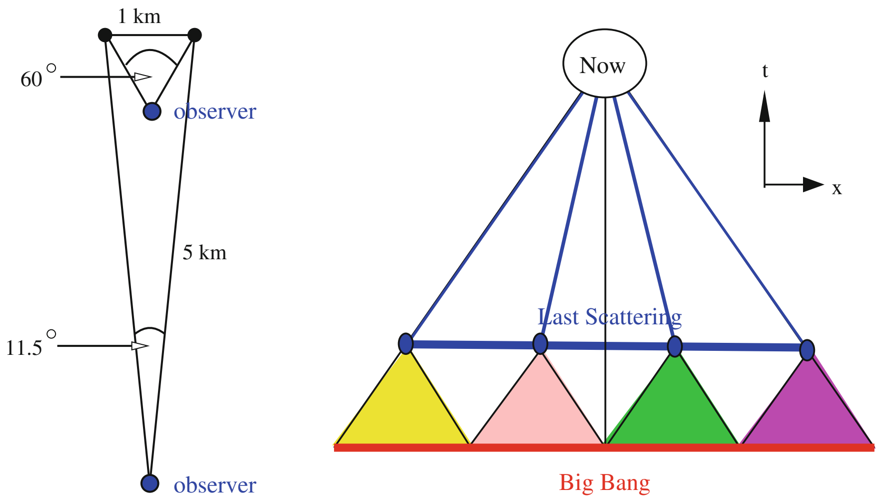

Fig. 2.7 The radiation emitted from a fixed spatial region covers an ever smaller angle of observation with time (*left*). As a result, the microwave background radiation we receive today comes from regions that were causally disconnected at decoupling time (*right*)

But first let us dwell a little more on the cooling process, which since decoupling has brought the temperature of the microwave background radiation from its initial 3,000 K to today's 3 K. The frequency of light emitted from a hot body decreases as its temperature is lowered. This occurs through the interaction of the light with the atoms of the system, which maintain the temperature of the medium. They exist in various states of excitation, and correspondingly emit and absorb photons on moving from one state to another. As the medium is cooled, the atoms absorb more high frequency photons and emit more low frequency photons, leading to an overall shift towards lower frequencies, i.e., longer wavelengths, for the radiation. How then can a "cosmic redshift" occur in the universe, where there is so much empty space and so few atoms to regulate the temperature? The origin of the cosmic cooling is a little bit like the Doppler effect we encountered earlier for waves emitted by moving sources. We see those waves "stretched" in wavelength as the source is moving away from us. A similar thing happens to a solitary wave travelling through an expanding space—the distance between crest and valley in the wave becomes stretched, the wavelength longer, the more the space expands. And if the space has expanded by a factor thousand since the emission of the cosmic light, the frequency of the light has decreased by this factor and the wavelength increased. So the cosmic redshift does not tell us that the source of the radiation is locally moving, but rather that the space through which it travels is expanding.

The expansion of the universe is encoded in "Hubble's law", stating that the velocity of a distant galaxy, relative to Earth, is proportional to its distance from Earth. The crucial scale factor is the "Hubble constant" H_0, for which the best present value is about 22 km/s per million light-years. So a galaxy one million light-years away is receding from us at a velocity of 22 km/s, while one two million light-years is receding at 44 km/s. If the expansion of the universe takes place at constant acceleration, the inverse of the Hubble constant gives us the age of the universe: 13.8 billion years; the details are shown in Box 3.

Box 3. The Age of the Universe

Hubble's law, $v = H_0 d$, gives us the recession velocity v of a distant star, with d specifying its distance from Earth and H_0 the Hubble constant.

For constant acceleration, i.e., for a constant rate of expansion of space, $v = d/t_0$, where t_0 is the time since the Big Bang, assuming both the star and the Earth were effectively born shortly afterwards.

Compared to the present distance, the separation of star and Earth at their birth are negligible, $d = 0$ shortly after the Big Bang. So it follows that $v = d/t_0 = H_0 d$, and from this that $t_0 = 1/H_0$ is the time since the Big Bang, the age of the Universe.

Many aspects, both observational and theoretical, enter the determination of the expansion and its time dependence. One crucial feature is the overall mass of the universe. If it is large enough, gravity could eventually stop the expansion and the universe will start to contract again. The result would then be a final "Big Crunch". If the mass of the universe is sufficiently small, the expansion can overcome gravity and the acceleration will increase with time. The critical boundary between the two extremes results in constant acceleration. The overall mass contained in the universe is not easily determined, since in addition to the visible content there is a large amount of invisible *dark matter*, which manifests itself only through gravity. And then, even more elusive, there is most likely an overall background of *dark energy*, which permeates the entire universe and hence affects its expansion rate. According to recent results that led to the award of the 2011 Nobel prize in physics, the vote goes to an acceleration increasing with time and hence assigns an important role to the mysterious dark energy. In any case, when all is said and done, the best value for the age of the universe today remains at about 14 billion years.

It is perhaps interesting to elaborate here a little on the nature of the expansion following the Big Bang. First, we should, however, note that the "reason", the initial cause of the bang, is not really known. One very impressive attempt to describe the very early stages of the universe was first proposed by the American cosmologist Alan Guth in 1980.

2.4 Cosmic Inflation

Whenever we measure something, we need a reference, a "zero". The height of a mountain is measured "above sea level", the depth of the ocean floor "below sea level". Mount Everest is the highest mountain on Earth only if we use the average sea level as reference, giving it a height of some 8,800 m. The volcano Mauna Kea on Hawaii rises more than 10,000 m above the floor of the ocean at its position—so it is indeed the *tallest* mountain on Earth. But let us now imagine a dammed river: on the high side, upstream, the level of water is quite different than on the low side, downstream. And this difference in water levels corresponds to a difference in potential energy that can be used, for example, to create electricity by the water rushing down the damm. So the transition from one level to the other can happen very abruptly, and it can liberate energy. In cosmic inflation, the entire universe we can see today was a small bubble of extremely hot matter just after the Big Bang, small enough to be causally connected and in uniform thermal equilibrium; its ground state, the reference point, was far above ours today. The bubble expanded, cooled and thereby was driven to a critical point, over the dam, down the waterfall. In this process, the space of the medium expanded dramatically in an extremely short time, and its new reference point became our physical vacuum of today. Since the medium had been in equilibrium before, it remained uniform even after the expansion of space had broken it up into causally disconnected regions. So that is why, according to inflation theory, we measure the same microwave radiation from all parts of the sky: before inflation, the sources not able to communicate with each other at decoupling time were originally all in the same pot, in which they could adjust to each other's tune, and this information was conserved in the transition. Moreover, in descending from the upper to the lower level, energy was liberated, and this energy, "dark energy" in today's terminology, permeates the entire universe; it drove and continues to drive the expansion of the universe.

But even cosmic inflation can only show that, given certain conditions, a hot expanding early universe can be formed in an extremely short time. It does not explain the origin of these conditions, so that, for the time being, the beginning of the world seems well beyond our science. The subsequent evolution depends, as we mentioned, on the strength of gravity, on the overall mass of the universe. The Big Bang provided the expansion, gravity counteracts this, dark energy may modify it, and, whatever the final verdict on the role of the different components, the universe continues to expand. This expansion is not an "explosion", throwing debris into some empty space. Rather, space was *made* in the Big Bang, and it is space itself that is expanding. So a better analogy for our present universe is that of raisins in a cake dough, after some time in the oven. As the cake "rises", any given raisin notes that all its neighboring raisins are moving further and further away. And the dough between the raisins, that is "space". For the concept of expansion, it does not matter how much dough there is or if there is an end to it. Similarly, in the Big Bang, the primordial matter as such was not localized at some point in space. We can only see that part of the universe from which light has been able to reach us in 14 billion years, and that part was indeed localized. Whatever more there was (and now is), we simply cannot tell. But we can speculate that there is more; we can't see it now, but it seems that if we, mankind, wait long enough, light from there will arrive, so that we, taken generically, should be able to see it then, at sufficiently much later time.

Unfortunately (or fortunately, depending on your point of view), that is not true. We can use Hubble's law to see how far away a distant star has to be at present so that for us it is moving away at the speed of light. Using the value of the Hubble constant given above, we find that the critical distance is 14 billion light-years. A star further away from us than that is now moving, relative to us, faster than the speed of light, and any signal it may send will never be able to reach us. So there is an absolute cosmic horizon.

2.5 The Absolute Elsewhere

For us and all our descendents, the universe presently further away than 14 billion light-years is forever beyond any communication; we cannot send "them" a signal, nor ever receive one from "there". Our world thus remains in principle bounded by the "Hubble sphere" with a radius of 14 billion light-years. But this specific limit applies only to us here on Earth. A distant star will have its own Hubble sphere, and that will cover a different region of space—which may or may not have an overlap with ours. There is more "out there", but our capability to communicate with it has an absolute limit.

But, you may say, how can something move faster than the speed of light? And indeed, nothing can "outrun" a light beam. The new feature entering in cosmic dimensions is theexpansion of space. The far-away star will emit a light beam, and, measured on that star, it will start on its path towards us with exactly the universal speed of light. The problem is that while it is travelling, the space of the universe expands, and if this expansion is fast enough, the light beam will never reach the Earth. So Hubble's law is not saying that the distant star is "running away" from us; from the point of view of other stars near it, it is stationary. And whatever region the light beam passes on its way, any observer there will see it moving with the universal speed of light. The light beam is thus a little like a worm crawling through the expanding cake, from one raisin towards another. Any observer it passes will see it crawling with its standard worm speed, but in the meantime, the rising cake stretches the space it has to traverse, and if the cake rises fast enough, the poor worm will never reach the next raisin. Even during inflation, it was space that was undergoing the abrupt expansion—on a sufficiently small local level, nothing was moving faster than the speed of light.

If the Hubble constant were really a constant, our Hubble sphere would have been the limit of our universe since the Big Bang. Slight time variations of H_0 even now are under discussion by the experts, and immediately after the Big Bang, as we saw, there may well have been a very short period of a much more rapid "inflationary" expansion. For our overall picture, we will skip over the evolution of the very early universe and assume that our Hubble radius has been "in effect" almost since the Big Bang. That means that any part of the universe beyond our Hubble horizon shortly after the Big Bang was then and ever afterwards outside our world, unreachable for us. It was then expanding away from us faster than the speed of light, and has continued to accelerate since then.

What about a star formed just *inside* our Hubble sphere not long after the Big Bang? The expansion of the space environment of that star proceeded, as seen by us, with an effective speed slightly less than that of light, and so the light of the star could still eventually reach the Earth. But the expansion rate continued to increase, and shortly afterwards became greater than that of light. From our point of view, at that instant the light of the star went off, it disappeared from our world. But the light it had emitted before crossing our Hubble limit continued to travel through the expanding space. And when it finally reaches us, its source star is far, far away outside our world, in our absolute elsewhere.

To find out how far, we ask if a light signal was sent out from Earth shortly after the Big Bang, how far has it travelled in the time since then, until today? For a static universe, that distance would be the speed of light times the age of the universe: about 14 billion light-years. But the expansion makes the distance much larger, as our worm discovered above inside the expanding cake dough. Taking the expansion rate to be that of constant acceleration, the light beam has travelled three times the static distance since the Big Bang: 42 billion light-years; the calculation is shown in Box 4.

Box 4. How Far Has Light Travelled Since the Big Bang?

For a static universe, the distance travelled by a light beam between an initial time t_i and a final time t_f would be $d = c(t_f - t_i)$, where c is the speed of light. But the universe expands in that time interval by a factor $(t_f/t_i)^{2/3}$, as is predicted by an acceleration not changing with time. In this case, the stretching of space makes the travel time longer, so that the distance now becomes

$$d = c \int_{t_i}^{t_f} \mathrm{d}t \, (t_f/t_i)^{2/3} = 3ct_f$$

if we take the initial time $t_i = 0$ to be that of the Big Bang. So when the light reaches us, it has travelled three times the distance it would traverse in a stationary world; one unit ct_f for "local" travel, two units thanks to the expansion of space.

As a result, the most distant stars we see are *now* much further away from us than the speed of light times the age of the universe. When the light we receive from them today was emitted by them, they were much closer to us than they are today, just inside our Hubble sphere. But during the time of travel of the light beam, the universe expanded, and so our distance to them today is, as we just saw, a combination of the time of travel of the light and the expansion of the universe during that time. The most distant star whose light we see today is therefore *now* 42 billion light-years away from us. Provided it still exists, of course…this we can never find out.

From a philosophical point of view, this form of an ultimate spatial limit, of an ultimate horizon of our universe, is really quite satisfying. The "last outer sphere" in older cosmologies always led to a number of unanswerable questions. What is the origin of such an ultimate sphere? What is it made of, what happens if a signal sent by us hits it? And finally, the forbidden question: what is behind this last limit, this end of the universe? In today's cosmology, the limit exists only in the eye of the beholder. At that imagined surface in space 14 billion light-years away from us, there is nothing special, no discontinuity, no great wall of any kind, and there is no reason to expect that beyond this limit, things are different. Only we can no longer check that. The limit exists for us, for our eyes only, not for other observers in far away parts of the universe. The world according to Thomas Digges, some 450 years ago, is also ours today, except that we now know it had a beginning and that our probing must reach an end.

Box 5. The Doll in the Doll

In the mechanics of Newton, instantaneous interactions over large distances were implicitily considered possible—and from our present view, this means that effectively the speed of light was taken to be infinite. For a vast range of natural phenomena, this assumption is satisfied to high precision: as long as things move with a velocity much less than that of light, Newtonian mechanics remains correct. It is only when particles move with velocities close to that of light, as they do for example in today's large particle accelerators, that relativity theory, more specifically, Einstein's special theory of relativity, becomes the correct description. The resulting relativistic mechanics contains Newton's non-relativistic mechanics as the limiting form obtained for small velocities. The mechanics of special relativity in turn remain correct only as long as the force of gravity remains comparatively weak. A light beam on Earth is not measurably bent downward, and for the motion of a particle in one of the mentioned accelerators, the effect of gravity can also be totally ignored. It is only on a cosmic scale, for forces between galaxies or light passing massive stars, that the deformation of space through interaction plays a role. At that point, Einstein's general theory of relativity gives the relevant explanation. In the limit of small scales and weak gravity, it gives the special theory as an excellent approximation. So in a way, it's like the Russian babushka dolls: the biggest, general relativity, contains a smaller one, special relativity, and this turn contains a still smaller one, Newtonian mechanics.

http://www.springer.com/978-3-642-41656-9

Ultimate Horizons
Probing the Limits of the Universe
Satz, H.
2013, XII, 172 p. 68 illus., 53 illus. in color., Hardcover
ISBN: 978-3-642-41656-9