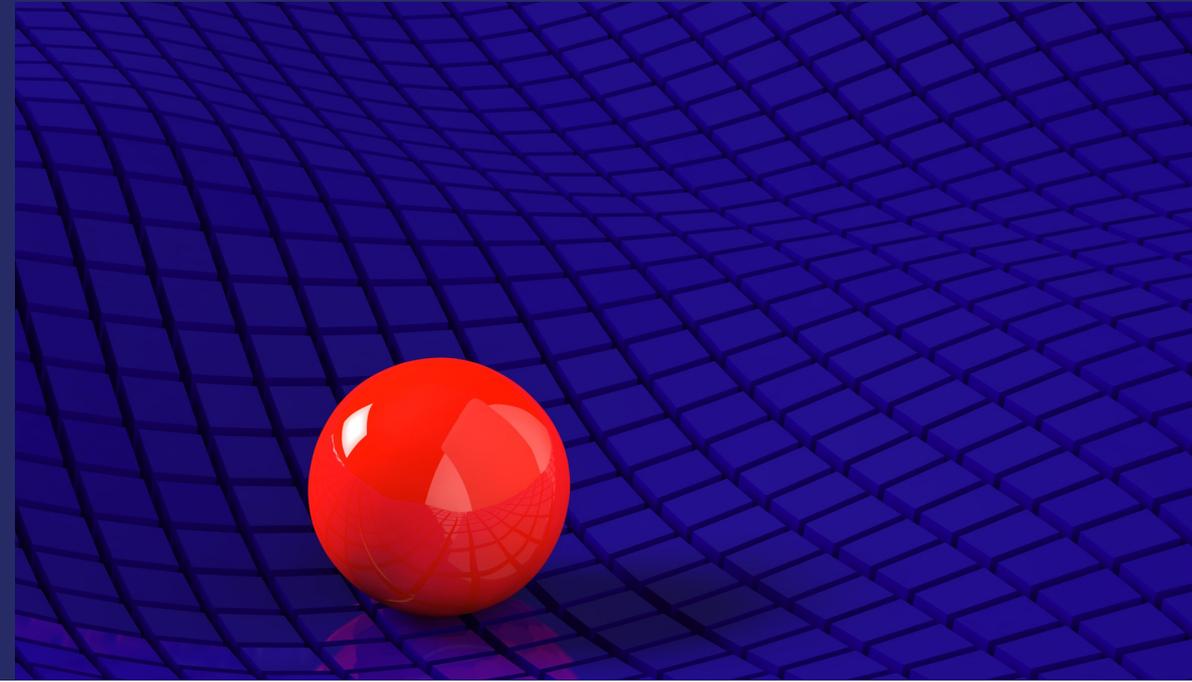


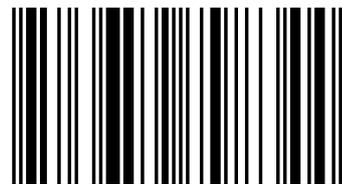
One hundred years ago, Albert Einstein came up with a new theory of gravitation. Later he added to his field equations the cosmological constant which is accountable for the accelerated expansion of the universe. Thus a natural question arises concerning what is the energy source for this expansion. This monograph is about the novel concept of antigravity and antigravitational forces. We present various observational arguments showing the expansion on small cosmological scales. In particular, the Solar system and single galaxies expand at a rate comparable to the Hubble constant. The novelty of our ideas is that traditionally the expansion of the universe according to Edwin Hubble has been assumed to take place only between galaxies. The boldness of our approach is that it points to a weak violation of the law of conservation of energy. We claim that any system of free bodies that interact gravitationally with delays expands on average. We suggest that this is due to gravitational aberration effects resulting from a finite speed of gravity. Our book should be especially useful for scientists who look for the origin of dark matter and dark energy.



Michal Křížek
Filip Křížek
Lawrence Somer

Michal Křížek is a senior researcher at the Institute of Mathematics of the Czech Academy of Sciences and Professor at Charles University. Filip Křížek is a junior researcher at the Nuclear Physics Institute of the Czech Academy of Sciences. Lawrence Somer is Professor of Mathematics at the Catholic University of America in Washington, D.C.

Antigravity - Its Origin and Manifestations



978-3-659-79834-4

*Dedicated to those who look for
the origin of dark matter and dark energy.*

*The effort to understand the universe is
one of the very few things that lifts human
life a little above the level of farce,
and gives it some of the grace of tragedy.*

STEVEN WEINBERG

Contents

List of symbols and constants	x
Foreword	xii

Part 1: Newton's theory of gravity and the problem of dark matter

1. Astronomy and mathematics: cross-fertilization over the millennia	3
1.1. Introduction	3
1.2. Kepler's laws	7
1.3. Some useful relations	10
1.4. Consequences of Kepler's second law	11
2. The role of the protractor in understanding the universe	14
2.1. Angle measurement devices	14
2.2. Measurement of relative distances in the Solar system ..	16
2.3. Establishment of absolute distances	18
2.4. Establishment of relative distances of inner planets	20
2.5. Improvement of the accuracy of the Earth-Sun distance	21

2.6.	Further improvement of the accuracy of the Earth-Sun distance.....	23
2.7.	Slowing-down of Earth's rotation	24
2.8.	Annual parallax of the nearest stars	26
2.9.	Measurement of the speed of light	27
2.10.	Spherical trigonometry	29
2.11.	Deflection of light in a gravitational field	32
3.	On Kepler's equation	36
3.1.	True and eccentric anomaly	36
3.2.	The relationship between the true and eccentric anomaly	38
3.3.	Kepler's equation for the eccentric anomaly	38
3.4.	Keplerian orbital elements	40
4.	The law of gravity — discovery of the millennium.....	42
4.1.	Newton's theorems	42
4.2.	The most important discoveries and applications	45
4.3.	The size of a constant in Kepler's third law	47
4.4.	Mass of the Sun	49
4.5.	Mass of Mars	49
4.6.	Falling into the Sun	50
4.7.	The size of escape velocities	51
4.8.	Flight altitude of geostationary satellites	53
4.9.	The flight time on Mars	53
4.10.	Mean mass density of the Sun	55
4.11.	Speed of Halley's comet	56
4.12.	The validity of the gravitational law outside the Solar system	57

4.13.	Determination of the distances of exoplanets from their mother stars	58
4.14.	Mass of the black hole at our Galaxy center	58
4.15.	Physical characteristics of the planets	62
5.	The N -body problem	65
5.1.	Introduction	65
5.2.	The two-body problem	66
5.3.	The three-body problem	69
5.4.	The N -body problem	73
5.5.	Total approximation error	75
6.	Eclipses and the aberration of light	81
6.1.	The importance of eclipses in exploring the universe ...	81
6.2.	A brief history of eclipses	83
6.3.	The origin and periodicity of eclipses	83
6.4.	Why lunar eclipses are less frequent than solar eclipses.	86
6.5.	Consequences of light aberration during total eclipses ..	87
7.	Zwicky's postulation of dark matter	90
7.1.	Fritz Zwicky	90
7.2.	The Virial theorem	93
7.3.	Zwicky's application of the Virial theorem to the Coma cluster	95
8.	The problem of missing matter	100
8.1.	Analysis of Zwicky's method	100

8.2.	Analysis of current data	108
8.3.	Reduction of the virial mass of the Coma cluster	112
8.4.	What is the mass of dark matter inside the Coma cluster?	120
9.	Vera Rubin and rotational curves of spiral galaxies	122
9.1.	Vera Rubin	122
9.2.	Spiral galaxies do not rotate according to Kepler's laws	123
9.3.	Orbital velocity around a spherically symmetric body .	127
9.4.	Orbital velocity around a flat disk	130
9.5.	Orbital velocity around galaxy bulges and halo	134
9.6.	Arguments against dark matter	136

Part 2: Antigravity and dark energy

10.	The accelerating expansion of the universe	143
10.1.	The 2011 Nobel Prize for Physics	143
10.2.	An expanding universe and the Hubble constant	145
10.3.	Type Ia supernovae — standard candles	150
10.4.	Measurements of cosmological parameters	153
10.5.	Historical notes	161
11.	Recession of Mars from the Sun	163
11.1.	Antigravity and the law of conservation of energy	163
11.2.	The rate of expansion of the Solar system	165
11.3.	Rivers on Mars	167
11.4.	Mars from the perspective of the Stefan–Boltzmann law	173

12. Recession of the Moon from the Earth	178
12.1. Measurement of the Earth-Moon distance	178
12.2. The paradox of tidal forces of the Moon	180
12.3. A remarkable coincidence	181
12.4. Recession speed of the Moon from the Earth due to tides	182
12.5. A time dependent momentum of inertia of the Earth .	188
12.6. The paradox of the large orbital angular momentum of the Moon	189
13. Recession of the Earth from the Sun	191
13.1. The faint young Sun paradox	191
13.2. The expansion of the ecosphere	192
13.3. Analysis of growth patterns on fossil corals from solar data	195
13.4. Analysis of growth patterns on fossil corals from lunar data	197
13.5. Consequences for the variable Earth-Moon distance ..	200
13.6. Prolongation of the sidereal year of the Earth	200
13.7. Elimination of other possibilities for the large recession speed	202
13.8. Why other authors obtained much smaller values for recession speeds	206
13.9. Generation of energy by the Earth-Sun system	209
14. Antigravity and the anthropic principle	211
14.1. The anthropic principle	211
14.2. Two-sided estimates	213
14.3. Will antigravity protect the Earth from an expanding Sun?	219
14.4. The probability of the appearance of life	220

15. Expansion of the Solar system	224
15.1. Fast satellites	224
15.2. Where was Larissa billions of years ago?	226
15.3. Satellites of Uranus	228
15.4. Falling Phobos	229
15.5. Delaying Neptune	233
15.6. Neptune-Triton system	235
15.7. Local Hubble expansion of the Solar system	237
16. Expansion of single galaxies	239
16.1. Do single galaxies expand due to antigravity?	239
16.2. Galactic expansion	239
16.3. Expansion of the Milky Way	241
16.4. Distribution of galaxies in the past	242
16.5. Star-formation rate	244
16.6. Activity of galactic nuclei	245
16.7. Gravitermal catastrophe	245
16.8. Exoplanet WASP-18b	246
17. What is the mysterious source of dark energy?	249
17.1. Gravitational aberration	249
17.2. Postnewtonian model or how energy is generated	253
17.3. The speed of gravitational interaction	257
17.4. Does the law of conservation of energy hold?	260
18. What the universe is	262
18.1. Non-Euclidean models of the universe	262
18.2. Isotropy and homogeneity of space	267

18.3. The ambiguity of the notion universe	268
18.4. Hyperbolic space	272
18.5. Maximally symmetric manifolds	277
19. A critique of the standard cosmological model	282
19.1. The standard mathematical cosmological model	282
19.2. Strange behavior of cosmological parameters	285
19.3. Excessive extrapolations	288
19.4. Expansion function	294
19.5. Dark matter versus baryonic matter	297
19.6. Dark energy versus cosmological constant	299
19.7. Conclusions	304
20. Apparent superluminal speeds in the universe	307
20.1. Observation of superluminal speeds	307
20.2. Mathematical justification of the observed paradox ...	309
20.3. Superluminal velocities at cosmological distances	312
20.4. The time-lens principle	314
20.5. The Big Bang singularity	316
Literature	318
Name index	340
Subject index	342

List of symbols and constants

$[0, 1)$	semi-closed interval
$\mathbb{N} = \{1, 2, 3, \dots\}$	set of natural numbers
\mathbb{E}^n	n -dimensional Euclidean space
\mathbb{S}^n	unit n -dimensional sphere
\mathbb{H}^n	unit n -dimensional pseudosphere
\mathbb{C}	set of complex numbers
$ F $	absolute size (norm)
\dot{a}	time derivative of the function $a = a(t)$
\times	multiplication or Cartesian product
\cdot	multiplication or scalar product
$:=$	assignment
\approx	approximate equality
\Leftrightarrow	equivalence
$x \in A$	x is an element of the set A
$B \subset A$	B is a subset of A
\hookrightarrow	isometric imbedding
\ll	much smaller
\gg	much larger
max	maximum
min	minimum
\log_b	logarithm to the base b
ln	natural logarithm
exp	exponential function $\exp(x) = e^x$
e	Euler number 2.718281828...
e	eccentricity
π	Ludolf number 3.14159265...
i	imaginary unit
i, j, k	integer indices
$\binom{m}{n}$	binomial coefficient m over n
\forall	for all
$o(\cdot)$	$f(\alpha) = o(g(\alpha))$, if $f(\alpha)/g(\alpha) \rightarrow 0$ for $\alpha \rightarrow 0$

Σ	sum
$\{x \in A \mid \mathcal{P}(x)\}$	set of all elements x from A with property $\mathcal{P}(x)$
$f : A \rightarrow B$	function f from A to B
$x \mapsto f(x)$	function which assigns value $f(x)$ to x
\square	Halmos symbol marking the end of a proof
\odot	Sun
M_{\odot}	mass of the Sun $1.988\,547 \cdot 10^{30}$ kg
m_{proton}	mass of the proton $1.67 \cdot 10^{-27}$ kg
z	redshift
c	speed of light in vacuum $299\,792\,458$ m/s
c_{G}	speed of gravitational interaction
G	gravitational constant $6.674 \cdot 10^{-11}$ m ³ kg ⁻¹ s ⁻²
Λ	cosmological constant $\approx 10^{-52}$ m ⁻²
σ	Stefan–Boltzmann constant $5.669 \cdot 10^{-8}$ Wm ⁻² K ⁻⁴
h	Planck’s constant $6.626\,069\,3 \cdot 10^{-34}$ Js
H_0	Hubble constant ≈ 70 km/(s Mpc)
$H(t)$	Hubble parameter
au	astronomical unit $149\,597\,870\,700$ m
AU	original notation of the astronomical units
pc	parsec 3.262 ly = $206\,265$ au = $3.086 \cdot 10^{16}$ m
yr	sidereal year $365.256\,36$ days = $31\,558\,149.54$ s
ly	light-year $63\,242$ au = $9.46 \cdot 10^{15}$ m
eV	electronvolt = $1.602 \cdot 10^{-19}$ J

Foreword

*The only law is that
there is no law.*

JOHN ARCHIBALD WHEELER

Revolutionary developments in physics such as Newton's theory of gravity, Einstein's special and general relativity, and quantum mechanics came at a time when some researchers found the courage to get out of a scientific rut, and looked at natural phenomena and measured data in an entirely new way. The prevailing idea of contemporary cosmology is that 68% of the universe is composed of some exotic dark energy, 27% of an unknown dark matter, and only less than 5% of normal baryonic matter. However, all models (with no exceptions!) which are used to describe the evolution of the universe were tested only on much smaller spatial and time scales. When they are applied to the whole universe, we necessarily make a considerable extrapolation without any guarantee that the result obtained is correct. The combination of measured data and model prediction is then interpreted as the presence of dark matter and dark energy. The goal of this publication is to point out some treacherous pitfalls that cosmologists encounter if they identify the results of simple mathematical models with reality.

This monograph about antigravity appears just 100 years after the formulation of Einstein's theory of gravity. It is organized in the following way. Before we start with the analysis of many open questions in contemporary cosmology, we recall some important milestones that illustrate how the current concept arrived. As such, the initial chapters are devoted mainly to the history of exploring the universe and

the steps leading up to the discovery of Newton’s law of gravitation. Next, we take a detailed look at calculations made by Fritz Zwicky and Vera Rubin, who came up with the idea that when describing the dynamics of large gravitationally bound systems — galaxy clusters and spiral galaxies — it is necessary to consider the existence of dark matter. We point out phenomena in their estimates that were omitted and why they had to postulate dark matter. Nevertheless, the scientific results should be independently verifiable at any time. We will use recent measured data, which essentially reduces the perceived amount of dark matter and increases the amount of baryonic matter.

In the second part of this book we will discuss the influence of a finite speed of gravitational interaction in an arbitrary bound system of gravitationally interacting free bodies, which theoretically leads to its gradual secular expansion. Thus, a question arises; whether such expansions are manifested in our universe. In Chapters 11–16 we present a wide range of observational arguments suggesting that the Solar system, as well as single galaxies slightly expand over long time intervals. This is caused by an ubiquitous, repulsive force which is called *antigravity*. It is locally very tiny but globally large on cosmological scales. Antigravity could be a consequence of causality and the finite speed of propagation of gravitational interaction. The existence of such a force would mean that the energy conservation law is slightly disturbed. This suggests a new alternative to the origin of dark energy.

We are well aware that the astronomical data may be fairly imprecise (of fuzzy type), e.g. when it comes to the weight and measure of the size and distances of galaxies. Therefore, many of the equalities “=” in this book should rather be replaced by the symbol “ \approx ” if they are not definitions or equalities in a mathematical model. Relations, which are supplemented by physical units in parentheses, are to be understood in the sense that all terms between dimensionless equations or inequalities are in these units.

This book arose from our works [126]–[148] that we had published between the years 1990–2015. Most of the chapters can be read inde-

pendently of the previous exposition. There is no problem in skipping sections if the material is too difficult. The reader himself can recalculate whether the presented values and methods are correct or not.

The contents of this book were improved by often very polemical debates with Jan Brandts, Miroslav Brož, Jan Chleboun, Yurii Dumin, Soňa Ehlerová, Jiří Grygar, Helena Holovská, Bruno Jungwiert, Marian Karlický, Josip Kleczek, Jaroslav Klokočník, Oldřich Kowalski, Pavel Kroupa, Petr Kulhánek, František Lomoz, Martin Markl, Ctirad Matyska, Attila Mészáros, Jan Novotný, Oldřich Novotný, Vladimír Novotný, Marcello Ortaggio, Jan Palouš, Alena and Vojtěch Pravda, Petr Preuss, Vojtech Rušin, Petr Sadílek, Alena, Jakub, and Martin Šolc, Ladislav Šubr, Michal Švanda, Marie Větrovcová, David Vokrouhlický, Jan Vondrák, Vladimír Wagner, Marek Wolf, Richard Wunsch, and Weijia Zhang. Their help is greatly appreciated.

Furthermore, we are deeply grateful to Jan Brandts, Jan Maršák, and Karel Segeth for careful reading of the manuscript and improving the presentation of the text, Caroline Griffis and Jana Grünerová for proofreading, Pavel Křížek for drawing most of the figures, Eva Ritterová for help with English translation, Jarmila Štruncová for ordering expert literature, and Hana Bílková for technical assistance in the final typesetting of the manuscript, and drawing some figures.

Finally, we are indebted to Sarah Cole, the acquisition editor of the LAMBERT Academic Publishing, for her helpful cooperation in the preparation of this book. Our great thanks go to the referees for valuable suggestions and also to our families for patience and understanding.

The work on this book was supported by RVO 67985840 of the Czech Republic. This support is gratefully acknowledged. We appreciate all of your comments.

Prague, March 8, 2015

Michal Křížek
Filip Křížek
Lawrence Somer

Part 1

Newton's Theory of Gravity and the Problem of Dark Matter

1. Astronomy and mathematics: cross-fertilization over the millennia

*Mathematics is a language,
spoken by all exact science.*

NIKOLAI I. LOBACHEVSKY

1.1. Introduction

Astronomy and mathematics are among the oldest scientific disciplines. They coexist and have mutually enriched one another for thousands of years. The determination of astronomical distances uses trigonometric methods. By means of numerical methods we may calculate the trajectories of space probes, which have allowed us among other things to visit the Moon, to capture unique pictures of the planets and their moons, to compute tracks of objects that threaten the Earth and to launch telecommunication and meteorological satellites among other things. Data from these devices are first compressed through sophisticated mathematical algorithms and are, during transmission, protected by error-correcting codes which enhance their reliability. Fourier analysis is then applied for processing the incoming signal. Without careful numerical calculations it would at present be impossible to construct and install many astronomical instruments, such as telescopes, hyperbolic mirrors, laser retroreflector arrays, interferometers, CCD cameras, GPS, and many others. Modern astrophysics extensively profits from mathematical software, supercomputers, and computer networks, too.

On the other hand, mathematics is indebted to astronomy for the invention of various numerical methods to approximate solutions of differential equations and values of integrals, for the theory of interpolation and extrapolation methods. Astronomers have partly developed also least squares and optimization methods, statistical methods, group theory, theory of chaos, the theory of series, mathematical modeling, stereometry, non-Euclidean geometry, tensor calculus, and so on.

Since ancient times, several cultures have been fascinated by the night sky, and made their own astronomical observations. The celestial sphere served mainly for estimating orientation and time, proof of which are numerous megalithic structures preserved at various places on the Earth. One of the oldest known astronomical observatories is Stonehenge, which is located in southern England. It served for the introduction of calendars based on accurate determination of the positions of celestial bodies as well as solstices, sunrises, and sunsets. It was constructed about 5 000 years ago. In Britain, there are also other very old observatories, such as Callanish and Castlerigg (see Fig. 1.1) whose stones are aligned with midwinter sunrise and various lunar positions. A similar semi-circular stone structure called Taosi dating back to 2100 BC is located in China's Shanxi Province. In the Czech Republic it was demonstrated that the rectangle at Makotřasy was also used for astronomical purposes approximately 3500 BC. As evidenced by the remarkable mathematical structures of the original Mayan or Chinese calendars, ancient astronomers were good observers and calculators.



Fig. 1.1. About 5 200 years ago Castlerigg Stone Circle east of Keswick in England served as an astronomical observatory (photo by Pavel Křížek).

In other civilizations, astronomy also developed successfully. The greatest ancient Greek observer Hipparchos (190–125 BC) compiled a catalog of the positions of more than 800 stars. He also introduced the concept of magnitude of stars and was a supporter of geocentrism, which assumes that the Earth is at the center of the universe.

Another Greek mathematician and astronomer, Claudius Ptolemy (ca. 100–170 AD), accepted Hipparchos' data about the stars and his opinions about the immobility of the Earth and its location in the middle of the universe. He created the so-called *Ptolemaic geocentric system* to explain the movements of all celestial bodies. His theory was later adopted by the Catholic Church. As a consequence in the Middle Ages it was not accepted to have a different opinion. Modern astronomy thus began to develop 13 centuries later, when the Polish astronomer Nicolas Copernicus (1473–1543) in his seminal work *De revolutionibus orbium coelestium* (On the Revolutions of the Heavenly Spheres) created the *heliocentric system*, in which all the planets revolve around the Sun.

Supporting arguments in favor of the heliocentric system, and hence against the geocentric one, were given by the Italian astronomer Galileo Galilei (1564–1642), when he discovered the phases of Venus and the moons of Jupiter. He was the first person who used a telescope to observe the celestial sphere. At that time he also discovered craters on the Moon, sunspots, stars in the Milky Way, and the rings of Saturn. He was one of the founders of modern physics. Galileo also attempted to measure the speed of light and was the first to come up with the idea that all bodies fall towards the Earth with the same speed, if not hampered by the atmosphere.¹

Along with the development of observation techniques, laws governing the seemingly irregular of movements planets were found. The key to this mystery was proposed by the famous German mathematician and astronomer Johannes Kepler (1571–1630). During his stay in

¹A practical experiment was realized, among others, in 1969 on the Moon with a feather and a hammer.

Prague he empirically derived two laws of planetary motions around the Sun (see Fig. 1.2). Later he added a third law, which is perhaps the most widely used relationship in astronomy at all (cf. Subject index). All three laws were formulated as mathematical statements. Kepler is therefore rightfully regarded as the founder of celestial mechanics. Kepler's discoveries and their uses in astronomy will be discussed throughout Chapters 1–3.



Fig. 1.2. The discovery of Kepler's first two laws in Prague is commemorated by a plaque on Charles Street that states: *From 1607 to 1612 Johannes Kepler lived here. In this period he discovered the first two laws about the motion of planets around the Sun.*

The culmination of these efforts was the creation of Newton's theory of gravity. Kepler's ideas were further developed by the English scholar Sir Isaac Newton (1643–1727) in the work *Philosophiae naturalis principia mathematica* (Mathematical Principles of Natural Philosophy). In it he formulated his three laws of motion and the law of gravity, which are indeed powerful tools to explore nature, as we shall see in Chapters 4 and 5), and which supported the heliocentric view of our universe. Newton was also the founder of infinitesimal calculus, which was used among other things to derive Kepler's laws from Newtonian mechanics.



1.2. Kepler's laws

Already since the 7th century AD Chinese astronomers were assembling detailed tables of positions of the planets, a thousand years before Kepler. They even knew that the trajectories of the planets form peculiar loops on the celestial sphere.² However, they did not discover Kepler's laws. The probable reason is that rather than concerning themselves with an explanation for the movements of planets, ancient Chinese astronomers were too much concentrated on predictions of solar and lunar eclipses, which were associated with disasters. They also did not have a good geometric model of the functioning of the Solar system and their measurements of positions of planets were one order of magnitude less accurate than the measurements of the Danish astronomer Tycho Brahe (1546–1601), who used a mural quadrant with a very fine scale similar to a nonius (vernier), see Fig. 2.1. It must also be taken into account that Kepler knew the Copernican heliocentric model of the Solar system very well and certainly Kepler's ingenuity³ played a big role as well.

During his stay in Prague from 1600 to 1612, Johannes Kepler analyzed the very accurate data of Tycho Brahe on the movements of the planets which were collected from observatories in Uraniborg and Benátky nad Jizerou (see e.g. [115, p. 54]). He found that the planets move in elliptic orbits and that the areal velocity of the radius vector that originates at the Sun is constant for each planet. In this way the first and second laws of Kepler were discovered. They were first published in 1609 in one of Kepler's most important works,

²The main reason for these loops is the fact that each planet orbits in another plane and at a different speed.

³For example, Kepler found all regular periodic tilings of the plane by regular polygons. He also constructed three regular star-shaped polyhedra and the rhombic triacontahedron with 30 congruent rhombic faces, which is the intersection of five cubes. He also invented the refractor telescope, in which the eyepiece and objective are formed by a converging lens, etc. Kepler was considerably ahead of his time. He even thought about cities on the Moon (see Ioh. Kepleri *Mathematico Olimimperatorii*, which was issued by his son Ludovico Kepler in 1634). While walking on Charles bridge in Prague, he asked a profound question: *why does every snow flake have a different shape from the other flakes, but each one has a six-fold symmetry?* Therefore, he is rightly considered to be one of the founders of crystallography.

Astronomia nova. Circumstances leading to this significant discovery are described, e.g., in [268].

Kepler's first law: *The orbit of each planet is elliptic with the Sun at one of its foci.*

Kepler's second law: *The radius vector connecting a planet and the Sun sweeps out equal areas of the ellipse during equal intervals of time.*

We will recall now, how Johannes Kepler discovered these laws. Kepler knew that the orbital period of Mars is 687 days⁴, therefore, after this time period Mars returns to the same place, while the Earth has orbited the Sun almost twice. This actually defines two different directions in which Mars is projected on the celestial sphere and Kepler could therefore determine its location in the orbital plane (see Fig. 1.3). Repeating this procedure for different time instants, with the help of coordinates of Mars measured by Tycho Brahe, Kepler could draw the entire orbit of Mars and find that its trajectory is elliptic (see Kepler's first law). When Kepler ascribed relevant time data to the respective positions of Mars, he discovered the second law.

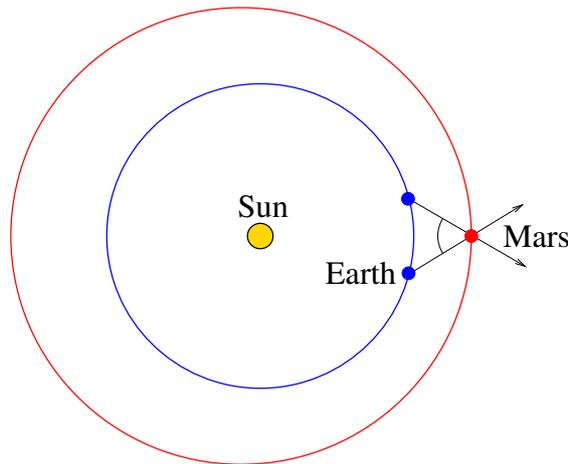


Fig. 1.3. Illustration of Kepler's method for the determination of the elliptic orbit of Mars, which after one Martian year projects to a different part of the sky. Mars is located at the intersection of the two directions.

⁴Today we know that the mean orbital period of Mars is 686.971 days.

Let $a \geq b$ stand for the lengths of the semiaxes of an elliptic orbit of the planet. For simplicity, we will also denote these semiaxes by the same symbols a and b . Here a is called the *semimajor axis* and b the *semiminor axis* provided $a > b$.

The distance ε of the focus of the ellipse to its center is called the *linear eccentricity* and is defined by the relation (Fig. 1.4)

$$\varepsilon = \sqrt{a^2 - b^2}.$$

Similarly

$$e = \frac{\varepsilon}{a}$$

is the *numerical eccentricity*. For simplicity, we will call this dimensionless number the *eccentricity*. The case $e = 0$ corresponds to a circle. The eccentricity of an ellipse is always less than unity. Note that a for parabola $e = 1$ and for a hyperbola $e > 1$.

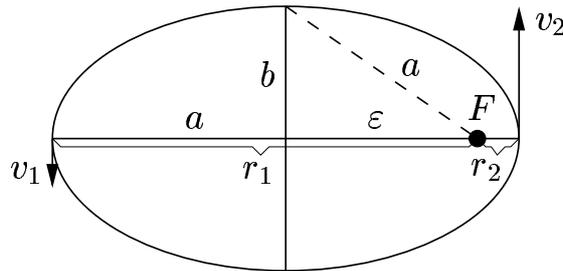


Fig. 1.4. Keplerian trajectory

Kepler was actually very lucky that he concentrated his attention mainly to Mars, because its orbit⁵ has a relatively large eccentricity $e = 0.0935 \approx 0.1$ (see Fig. 1.3). A consequence of Kepler's first and second law is (see [127]):

Kepler's third law: *The square of the orbital period of a planet is proportional to the cube of the semimajor axis of its orbit.*

⁵On the other hand, the Earth elliptic orbit has a very small eccentricity $e = 0.0167$. The semimajor axis $a = 149.598 \cdot 10^6$ km and the semiminor axis $b = a\sqrt{1 - e^2} = 149.577 \cdot 10^6$ km differ not before the fifth significant digit.

Kepler discovered the third law due to consistent and systematic work. Essentially, he found it empirically, when he left Prague. Kepler’s third law can be formulated as

$$\boxed{T^2 = Ca^3.}$$

Here, T is the orbital period of the planets and $C > 0$ is a constant. At that time the multiplication of numbers was accomplished by means of the sum of logarithms and after using the inverse procedure, Kepler obtained the product. For this purpose he used Bürgi’s tables⁶. Kepler noticed a simple dependence

$$2 \log T - 3 \log a = \text{const.}$$

between the decimal logarithms of the measured value of the orbital period T and the length of the semimajor axis a . It was then a short step to the formulation of the third “harmonic law”, which Kepler published in 1619 in his work *Harmonices mundi libri V*. In 1621, Kepler found that the new law also applied to the four large moons of Jupiter.



1.3. Some useful relations

Assume that the mass of a planet is negligible relative to the mass of the Sun. If r_1 and r_2 are the respective distances of the planet at aphelion and perihelion from the focus F , where the Sun is located, then $r_1 = a + \varepsilon$ and $r_2 = a - \varepsilon$ (see Fig. 1.4).

From the relations

$$2a = r_1 + r_2, \quad 2\varepsilon = r_1 - r_2,$$

and $b = \sqrt{a^2 - \varepsilon^2}$ we see that the length of the semimajor axis a is equal to the arithmetic mean of distances r_1 and r_2 , i.e.,

$$a = \frac{r_1 + r_2}{2}, \tag{1.1}$$

⁶Joost Bürgi (1555–1632), Swiss watchmaker and mathematician who developed the logarithms independently of John Napier.

while the length of the semiminor axis b is equal to their geometric mean, i.e.

$$b = \sqrt{\left(\frac{r_1 + r_2}{2}\right)^2 - \left(\frac{r_1 - r_2}{2}\right)^2} = \sqrt{r_1 r_2}. \quad (1.2)$$

Denote by v_1 and v_2 the respective velocities of a planet at aphelion and perihelion. By the law of conservation of angular momentum ($mr v = \text{const.}$) it follows that

$$r_1 v_1 = r_2 v_2. \quad (1.3)$$

From the above equalities we obtain

$$\frac{v_2}{v_1} = \frac{r_1}{r_2} = \frac{a + \varepsilon}{a - \varepsilon} = \frac{1 + e}{1 - e}. \quad (1.4)$$

For a fixed e , the ratio v_2/v_1 is thus a constant independent of the size of the elliptic orbit. For the eccentricity of Mars' orbit, $e = 0.0934$, we get a fairly high ratio $v_2/v_1 = 1.206$, which actually helped Kepler to discover his second law. For the eccentricity of Mercury's orbit, $e = 0.2056$, the relatively large ratio is given by $v_2/v_1 > 1.2/0.8 = 1.5$. Moreover, according to (1.4) and (1.3), we find that

$$e = \frac{v_2 - v_1}{v_2 + v_1} = \frac{r_1 - r_2}{r_1 + r_2}, \quad (1.5)$$

where $r_1 = a(1 + e)$ and $r_2 = a(1 - e)$.



1.4. Consequences of Kepler's second law

Relation (1.3) is actually the result of Kepler's second law according to which the radius vector sweeps at equal time intervals a surface of the same area (see Fig. 1.5) within the ellipse. The so-called areal velocity is thus constant, leading to the equality

$$\frac{1}{2} r_1 v_1 \cdot T = \frac{1}{2} r_2 v_2 \cdot T = \pi ab, \quad (1.6)$$

where the expression on the right-hand side is equal to the *area of the ellipse* and T is the orbital period of a planet. Hence, the speed v_1 is minimal and v_2 maximal.

Equation (1.6) can be proved by infinitesimal calculus. Here we only briefly outline its derivation.

Divide the orbital period T into n equally long time intervals $\Delta t = T/n$. If Δt approaches zero (i.e. n tends to infinity), then the individual segments of Fig. 1.5 taper and their curved side “straightens”. They resemble more and more triangles all having the same area. It is equal to the area $\frac{1}{2}r_1v_1\Delta t$ of the crosshatched triangle of Fig. 1.5 with height r_1 and base $v_1\Delta t$. The sum of the areas of all these infinitesimal triangles is then equal to $\frac{1}{2}r_1v_1T$, which is the value of the left-hand side of equation (1.6).

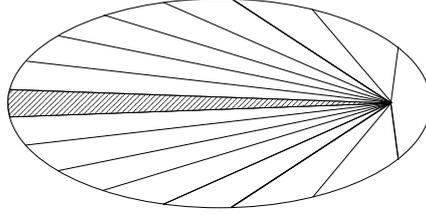


Fig. 1.5. Schematic illustration of Kepler’s second law stating that the areal velocity is constant.

Using Kepler’s laws, we now derive some useful relations which will be used later. From equalities (1.1), (1.2), and (1.6) it follows that

$$\frac{1}{2}r_1v_1 \cdot T = \pi \frac{r_1 + r_2}{2} \sqrt{r_1r_2}.$$

Substituting the expression for r_2 resulting from the law (1.3), we get

$$r_1v_1 \cdot T = \pi r_1^2 \frac{v_1 + v_2}{v_2} \sqrt{\frac{v_1}{v_2}}.$$

Hence, for r_1 (and similarly for r_2) we obtain

$$r_1 = \frac{T}{\pi} \frac{v_2}{v_1 + v_2} \sqrt{v_1v_2}, \quad r_2 = \frac{T}{\pi} \frac{v_1}{v_1 + v_2} \sqrt{v_1v_2}, \quad (1.7)$$

which according to (1.1) and (1.2) leads to

$$a = T \sqrt{\frac{v_1v_2}{2\pi}}, \quad b = \frac{Tv_1v_2}{\pi(v_1 + v_2)}. \quad (1.8)$$

Thus, we can determine the length of both the semiaxes from the orbital period and the minimal and maximal speed.

Finally, let us introduce a relationship between the angles corresponding to two radius vectors of a planet. Suppose that the Sun is located at the origin and the second focus is at the point $(-2\varepsilon, 0)$, where $\varepsilon = \sqrt{a^2 - b^2}$ and $a > b$. The second radius vector connects the second focus with the planet. Denote by (\bar{x}, \bar{y}) the coordinates of a planet (see Fig. 1.6) and by k_1 and k_2 the corresponding line slopes.

Substituting $\bar{y} = k_1\bar{x}$ into the equation $b^2(\bar{x} - \varepsilon)^2 + a^2\bar{y}^2 = a^2b^2$ of the ellipse, we obtain a quadratic equation for $\bar{x} > 0$. The slope of the second radius vector can then be expressed as (see Fig. 1.6)

$$k_2 = \frac{k_1\bar{x}}{\bar{x} + 2\varepsilon}. \quad (1.9)$$

For the angle φ between two radius vectors of the planet we have (see [224, p. 175])

$$\varphi = \arctan \left| \frac{k_1 - k_2}{1 + k_1k_2} \right|, \quad k_1k_2 \neq -1, \quad (1.10)$$

where after substituting the expression for k_2 obtained from (1.9) we get the expression for the angle φ as a function of the slope k_1 only. According to [224, p. 116], the normal at the point (\bar{x}, \bar{y}) bisects the angle φ . From this we can deduce how the normal direction differs from that of the slope k_1 .

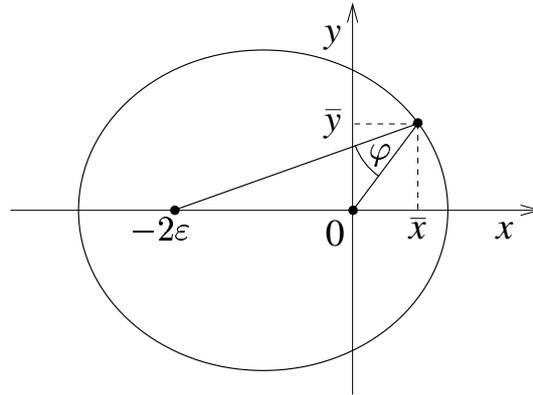


Fig. 1.6. Radius vectors of a planet include the angle φ .



2. The role of the protractor in understanding the universe

Mathematics, which governs our physical world, is extremely fruitful and powerful. I consider this relationship to be a deep secret.

ROGER PENROSE

2.1. Angle measurement devices

It took thousands of years until humankind attained our current picture and knowledge about the structure and functioning of the universe. By means of a series of examples we show that original geometrical ideas and an ordinary protractor played an essential role in this process. More precisely, various angle measuring instruments gradually developed in the past, such as the gnomon, the triquetrum, the armillary sphere, the astrolabe, the quadrant (cf. Fig. 2.1), the sextant, and the octant. Many of these instruments were already used in ancient civilizations to record various celestial phenomena.

For example, the use of sundials has a long tradition in China, Mesopotamia, and also in Greece. Time measurements are converted to measuring angles of shadows cast by a stone monolith or a rod (gnomon). In the year 545 BC, Anaximander measured the noon altitude of the Sun during the summer and winter solstice and divided their difference by two. In this way, he obtained the angle 23.5° between the plane of Earth's equator and the ecliptic plane.¹

¹The *ecliptic* is the plane in which the Earth orbits around the Sun.

A mural quadrant of Tycho Brahe (see Fig. 2.1) from the late 16th century allowed measurement of the azimuth to an accuracy of around one arc minute (which is on the verge of the resolving capabilities of the human eye), and was more than one order of magnitude higher than the other angle measuring devices at that time. From a careful analysis of angular measurements done by Tycho Brahe, Johannes Kepler then discovered his three famous laws that are nowadays derived from Newtonian mechanics.



Fig. 2.1. Each graduation of Tycho's quadrant is diagonally divided into ten segments each having six arc minutes, which in turn allows interpolating the measured angle with an accuracy of about one minute.

At present, the most common astronomical angle measuring device is a telescope. The resolution capability of the Hubble Space Telescope

is around one hundredth of an arc second. Optical and radio interferometers are among the most accurate devices. They allow us to measure extremely small angles less than $0.001''$.



2.2. Measurement of relative distances in the Solar system

The ancient Greek astronomer Aristarchus of Samos (3rd century BC) was probably the first scholar who proposed that planets orbit about the Sun. Therefore, he is considered to be the founder of the heliocentric model of the Solar system. He had several ingenious ideas and showed that seemingly complicated astronomical problems can be solved by means of elementary geometrical tools. When the Moon was in the first (or last) quarter, he realized that the angle SME is a right angle, where S, M, E stand for the Sun, Moon, and Earth, respectively (see Fig. 2.2). Using an ancient protractor, he measured the angle SEM and found that the hypotenuse SE of the right triangle SME is 19 times longer than the leg ME . In the modern notation, his idea may be written as follows:

$$\cos \alpha = \frac{|ME|}{|SE|} = \frac{1}{19}, \quad (2.1)$$

where $\alpha \approx 87^\circ$ is the angle SEM . In this way, Aristarchus deduced that the Sun is roughly 19 times further from the Earth than the Moon.

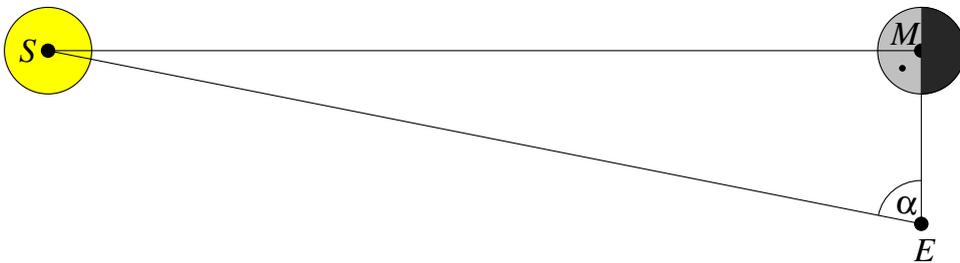


Fig. 2.2. When the Moon is in the first quarter, the angle SME is right, where S denotes the Sun, M the Moon, and E the Earth.

Of course, it was very difficult to determine the exact instant of the first quarter and then to measure the angle α by instruments of that time. Today we know that the Sun is approximately $389\times$ farther from the Earth than the Moon, which corresponds to an almost right angle $\alpha = 89.8527^\circ$. The large difference in the relative distances is due to the fact that $(\cos 87^\circ)^{-1} \ll (\cos 89.8527^\circ)^{-1}$, even though the respective angles have practically the same size.

Aristotle (ca. 384–322 BC) in his treatise *On the heaven* [10] argued that the Earth is a sphere, because its shadow on the Moon during lunar eclipses is always circular (see Fig. 2.3), independently of the position of the Earth. Later Aristarchus measured the angular size of this shadow $\approx 1.5^\circ$ and found that it is 3 times larger than the angular radius of the Moon $\approx 0.5^\circ$. He stated that the Earth freely moves in space and its radius is 3 times larger than the radius of the Moon (at present we know that it is 3.67 times). Since both bodies are close to each other and very far from the Sun, the rays of the Sun are almost parallel. From this, Aristarchus calculated that the Moon is 70 Earth’s radii from the Earth, which can be expressed in modern notation as follows:

$$\tan(3 \cdot \frac{1}{2} \cdot 0.5^\circ) = \tan 0.75^\circ \approx \frac{R}{70R}, \quad (2.2)$$

where R is Earth’s radius. According to present knowledge, the Earth-Moon distance is about 60 Earth’s radii.² Moreover, Aristarchus formulated the revolutionary hypothesis that the Earth moves around the Sun and not vice versa. He proposed that the Sun and the Moon have the same angular size, but the Sun is much bigger than the Earth, since it is 19 times further away than the Moon, whereas the Earth is only three times larger than the Moon (see Figs. 2.2 and 2.3). Almost none of the original writings by Aristarchus have been preserved (cf. [9]). His reasoning, however, is mentioned in the treatises *The Sand Reckoner* by Archimedes.

²Since the angular size of the Moon is approximately $31.1'$, it then immediately follows that $\tan(3.67 \cdot \frac{1}{2} \cdot 31.1') \approx R/(60R)$.

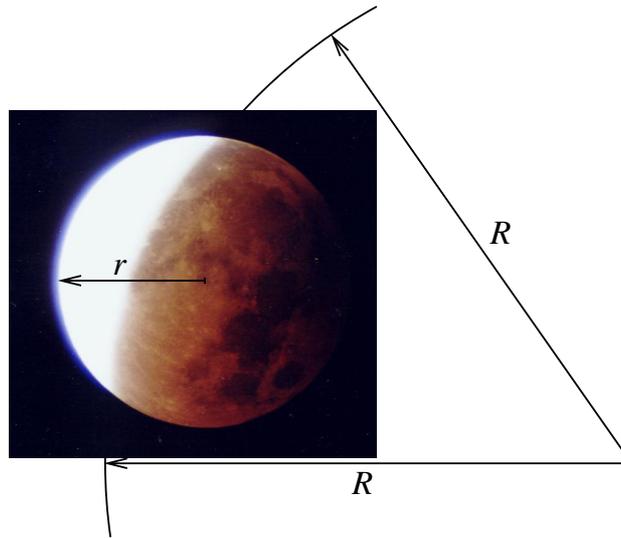


Fig. 2.3. A proof of Earth's sphericity: The shadow of the Earth on the Moon during a lunar eclipse is circular. Its radius R is more than 3 times larger than the radius of the Moon r .



2.3. Establishment of absolute distances

Aristarchus' determination of relative distances in the Solar system was supplemented in a more sophisticated manner by another ancient Greek astronomer and scholar, Eratosthenes of Cyrene (ca. 276–194 BC). He is famous not only for his prime number sieve, but also for the first trustworthy and ingenious calculation of the Earth's circumference (see [89]). Here again, angular measurements played an important role. At that time it was known that the Sun's height varies at distinct latitudes. Eratosthenes used the simplest astronomical instrument — the gnomon — which is just a straight stick perpendicularly raised to the Earth's surface. He knew that the Sun shines on the bottom of deep wells in Syene (at the tropic of Cancer near today's Aswan) at midday of the summer solstice. This means that the Sun is at its zenith and that the gnomon does not throw any shadow there. At the same time, in Alexandria (which lies on almost the same meridian as Syene) Eratosthenes found that the angle between the vertically placed gnomon and the Sun's rays is $\beta = 7.2^\circ$ (see Fig. 2.4), i.e. $\frac{1}{50}$ of

the full angle of 360° subtended by the entire circle. The distance $d = 5\,000$ stadia ≈ 920 km between Alexandria and Syene was found by travelling on carts, as stated in several standard references. By the relation

$$\frac{d}{o} = \frac{\beta}{360^\circ},$$

Eratosthenes derived that the Earth's circumference³ is

$$o = 250\,000 \text{ stadia} \approx 46\,000 \text{ km.} \quad (2.3)$$

It is not essential to know how large the Greek distance unit “stadium” was exactly (most probably its value was in the interval 148–210 m). More important was the invention of an elegant method for measuring the Earth's circumference.⁴

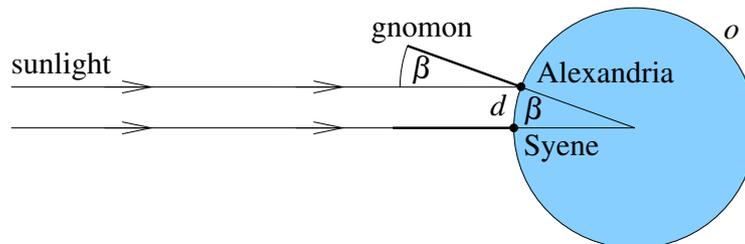


Fig. 2.4. The Earth's circumference o was calculated from the known distance d between Alexandria and Syene and the angle β was determined by gnomon at the midday of the summer solstice in Alexandria.

According to Aristarchus' and Eratosthenes' considerations (compare with (2.1), (2.2), and (2.3)), the Earth-Sun distance was about $19 \cdot 70 \cdot 46\,000 / (2\pi)$ km, which is less than 10 million kilometers. Measuring distances in the Solar system by the ancient Greeks is described in more detail e.g. in [108].

We can estimate the mean absolute distance of the Moon using knowledge of the equatorial radius of the Earth $R = 6378$ km and the relation (2.2) derived from lunar eclipses. In this way we would get too rough estimate, because the boundary of Earth's shadow on the Moon is not sharp. Let us therefore introduce a more sophisticated

³At present we know that $o = 40\,000$ km.

⁴In the 17th century, the value of the Earth's circumference was refined using measurements of angles and distances in a given triangulation (see [19]).

method based on measuring the angular diameter of the full Moon from the Earth's equator.

Denote by r_1 (resp. r_2) the unknown distance of the Moon at apogee (resp. perigee) when the Moon has the smallest (resp. largest) angular size. When the Moon is near the apogee and it rises⁵ above the horizon, its measured angular diameter is $\zeta = 29.469'$. About 6 hours later, when the Moon is at its highest position due to the rotation of the Earth, it will be $R = 6378$ km closer and its angular size will increase to $\xi = 29.940'$. The radius of the Moon r can be expressed in the following two ways

$$r = r_1 \sin(\zeta/2), \quad r = (r_1 - R) \sin(\xi/2).$$

From this we get $r_1 = 405\,500$ km and $r = 1738$ km. In a similar way we deduce that $r_2 = 363\,300$ km, when the Moon is at the perigee.⁶ According to equation (1.1) the semimajor axis of the elliptical orbit of the Moon is $a = 384\,400$ km.



2.4. Establishment of relative distances of inner planets

At the beginning of the 16th century, Nicolas Copernicus used angle measurements to determine the relative distances of the known planets in the Solar system at that time, including Saturn.⁷ On that basis, he came to the conclusion that the planetary orbits are circular and have the Sun as their common center [56]. For instance, he established (see Fig. 2.5) that the radius of the orbit of Venus is about 72% of the radius of the Earth's orbit by measuring the maximum separation angle of Venus from the Sun (see [253, p. 39 and 44]). The

⁵Or it descends.

⁶Distances r_1 and r_2 fluctuate with time. For instance, on February 3, 2125, the distance to the Moon will be $r_1 = 406\,720$ km, while in January 4, 1912 it was $r_2 = 356\,375$ km.

⁷Uranus was discovered by William Herschel only in 1781. Due to irregularities in its orbit found by angular measurements, Johann Gottfried Galle discovered the last planet Neptune in 1846 (see Section 4.2).

distances a_1 and a_2 to the Sun of both the inner planets, Mercury and Venus, were estimated by the relation

$$a_i = a_3 \sin \alpha_i,$$

where α_i is the maximum angle of elongation, i.e. the largest angular distance between the Sun and a planet on the celestial sphere⁸ (see Fig. 2.5). Copernicus' method of measuring the relative distances of outer planets to the Sun is described e.g. in [19, p. 265] and [56]. It is similar to Kepler's method of Fig. 1.3. Besides measuring the angles it also requires measurement of time.

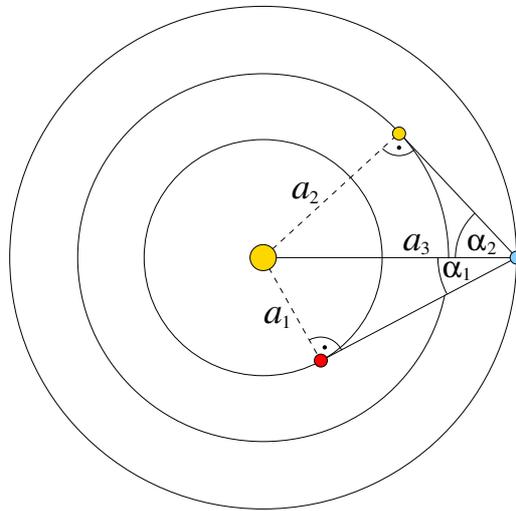


Fig. 2.5. Copernicus' method for establishing relative distances of the inner planets by means of the maximum elongation $\alpha_1 = 28^\circ$ for Mercury and $\alpha_2 = 47^\circ$ for Venus.



2.5. Improvement of the accuracy of the Earth-Sun distance

The estimate of the Earth-Sun distance dramatically improved in 1672, when G. D. Cassini⁹ measured the distance from the Earth to

⁸In fact, the orbit of Mercury is not circular, and thus the maximum elongation varies between values 18° and 28° .

⁹Giovanni Domenico Cassini (1625–1712) also discovered a large gap in the rings of Saturn.

Mars by using a protractor. In Paris (P) he measured the position of Mars on the celestial sphere when Mars was at its nearest point to Earth, i.e. in opposition with respect to the Sun (see [119, Chap. 1]). At the same instant, his colleague Jean F. Richer in Cayenne (C) in French Guyana also measured the position of Mars (M) on the celestial sphere. From the corresponding parallax¹⁰ $\angle CMP = 18''$ and the known distance $d = 7280$ km between Paris and Cayenne it was found by the law of sines and standard trigonometric methods that Mars is 73 million km from the Earth.¹¹ Then Kepler's third law was applied

$$\frac{T_i^2}{T_j^2} = \frac{a_i^3}{a_j^3}, \quad i, j = 1, 2, 3, \dots, \quad (2.4)$$

where T_i is the sidereal period of i th planet and a_i is the length of the semimajor axis of its elliptic orbit. For the Earth and Mars we have $T_3 = 1$ and $T_4 = 1.88$ years. Hence,

$$a_4 = 1.88^{2/3} a_3. \quad (2.5)$$

Another equation for the unknowns a_3 and a_4 follows from the fact that planetary orbits are almost circular together with the above-mentioned angular measurement, i.e., $a_4 - a_3 = 73 \cdot 10^6$ km. From this and (2.5) we immediately get $a_3 \approx 140 \cdot 10^6$ km which is a quite good estimate of the modern value $a_3 = 149.6 \cdot 10^6$ km.

The semimajor axis of Earth's orbit around the Sun was chosen by astronomers as the basic length unit and called the *astronomical unit*¹². The distances a_i of all the other known planets were then calculated from Kepler's third law (2.4) and the observed periods T_i .



¹⁰The angle by which the body is shifted compared to the distant background when observed from two different locations.

¹¹This result was obtained, in fact, in French miles, 1 French mile is 1.949 km.

¹²At present the *astronomical unit* is defined as follows: 1 au = 149 597 870 700 m. It is approximately equal to the present mean distance of the Earth from Sun.

2.6. Further improvement of the accuracy of the Earth-Sun distance

Another ingenious geometric method (see [95], [167]) to establish the astronomical unit more precisely was suggested by the famous astronomer Edmond Halley (1656–1742). He developed a practical method based on transits of Venus, observed from different places of known latitude. Halley’s method was applied later by another generation of astronomers¹³ in 1769. According to [253, p. 133], the transit of Venus over the Sun’s disc was observed by more than 120 astronomers from about 60 stations. In particular, one group of astronomers headed by Maximilian Hell was at the island Vardø (at present in Norway) and another one with Captain James Cook and Charles Green travelled to Tahiti (see [19, p. 267]). In Fig. 2.6 we see a sketch of two trajectories AB and CD of Venus observed from these two places. The angular distance between the two line segments AB and CD was found to be approximately $\alpha = 40''$. To refine the parallactic angle α , also the transit time was measured. Note that the angular diameter of the Sun is about $32'$, which is almost fifty times larger than α .

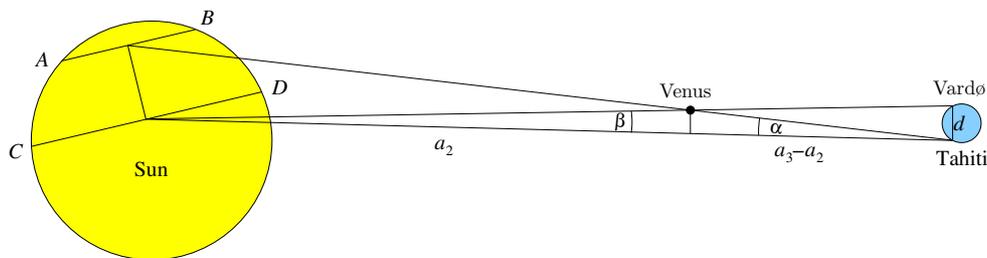


Fig. 2.6. A schematic illustration of two different trajectories AB and CD of Venus passing over the Sun’s disc observed from Vardø and Tahiti in 1769. The real angular distance between AB and CD is much smaller than in this figure.

¹³Also in 1761. The first known prediction of Venus’ transit across the solar disk dates from J. Kepler. In 1626 he calculated that the transit will appear in 1631 and 1639. This very rare phenomenon happens only several times per millennium, since the inclination $i = 3.4^\circ$ of Venus’ orbit to the ecliptic is relatively large. The last two transits appeared in June 8, 2004 and June 6, 2012 (see Fig. 6.1).

Since $T_2 = 0.615$ years, by (2.4) we get that $a_2 = 0.723 a_3$. Moreover, from Fig. 2.6 we find that $a_2 \tan \beta = (a_3 - a_2) \tan \alpha$. For simplicity, assume that the Earth-Sun line was perpendicular to the segment Vardø–Tahiti at a certain instant during the time of the transit. Then we obtain

$$a_3 \approx \frac{d}{\tan \beta} = \frac{a_2}{a_3 - a_2} \cdot \frac{d}{\tan \alpha} = \frac{0.723 d}{(1 - 0.723) \tan \alpha},$$

where the distance $d = 11\,425$ km can be calculated from the latitudes and longitudes of Vardø ($70^\circ 21', 31^\circ 02'$) and Tahiti ($-17.5^\circ, 149^\circ$) by means of ellipsoidal coordinates as in Section 2.10. In this way the estimation of the distance a_3 between the Earth and Sun was improved to $153 \cdot 10^6$ km.

At present there exist, of course, more sophisticated methods (see e.g. [253]), which take into account the motion of the Earth and Venus during the time of the transit, as well as other circumstances.



2.7. Slowing-down of Earth's rotation

The delay of the Earth's rotation is given by the difference between the terrestrial time, which is derived from the most accurate atomic clock and the world (Greenwich) time defined by the Earth's rotation. The Earth's angular speed of rotation slowed down, such that the length of a day increased by $1.7 \cdot 10^{-3}$ s per century (see [247, p. 270]). This value has been obtained by a thorough data analysis of ancient records of angular measurements of occultations of stars by the Moon and solar and lunar eclipses [271] during the last 2700 years. Just for illustration, we restrict ourselves to only one simple example taken from a clay tablet of ancient Babylonians containing a record of the total eclipse of the Sun on April 15, 136 BC.¹⁴ The tablet is preserved in the British Museum in London (see [266, p. 340]). Its English translation can be found in [269].

¹⁴This date was derived with respect to the present calendar.

At that time, a day was about 0.036 312 seconds (≈ 21.36 century $\times 1.7$ ms/century) shorter than in the year 2000. The time period from 136 BC to 2000 AD contains approximately $N = 780\,000$ days. Due to cumulative effects the rotation of the Earth has slowed down about 4 hours more than if it would rotate uniformly (see Fig. 2.7). This corresponds to an angle of $60^\circ (= 360^\circ \cdot 4/24)$. Let us describe now in detail, how these values can be derived.

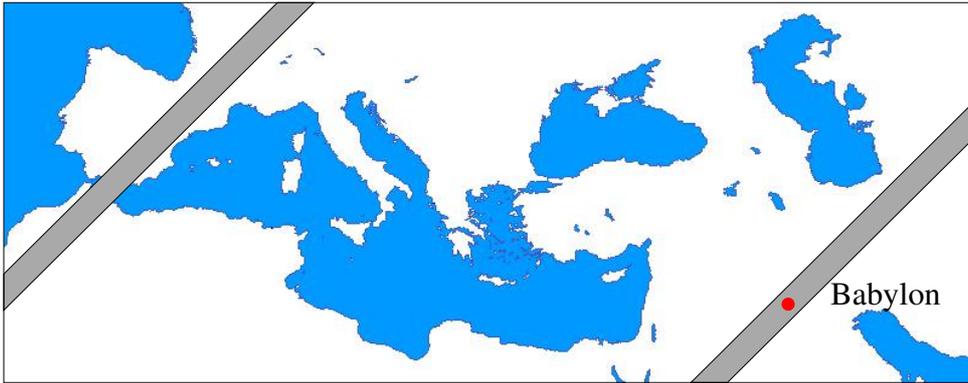


Fig. 2.7. The position of the band of totality during the solar eclipse observed by ancient Babylonians is on the right. On the left is the calculated position of the belt, if the rotation of the Earth would not have slowed down.

If the Earth's rotation were constant, then the ancient Babylonians could not have observed the total eclipse at the place where they actually described it, but 4 time zones shifted to the west of Babylon on the same latitude. Now we can exactly establish their local time of 8:45 a.m. during the eclipse from the height of the Sun over the horizon, which was measured by protractor and carefully recorded. From the shift $\Delta T = 4$ hours (four time zones) and the known number of days N , we can calculate the corresponding slow-down of the Earth's rotation.

Assume for simplicity that the length of each day increased linearly by the value Δt , i.e., the n th day is about $n\Delta t$ longer than the day on which the total eclipse initially appeared corresponding to $n = 0$. For the entire delay ΔT we obtain

$$\Delta T = \Delta t(1 + 2 + \dots + N) = \Delta t \frac{N(N + 1)}{2} = 4 \cdot 3600 \text{ s.}$$

Substituting for N the entire number of days, we can derive that $\Delta t = 4.734 \cdot 10^{-8}$ s. During one year, the length of a day then increased on average about

$$\mathcal{T} = 365.25 \cdot \Delta t = 1.7 \cdot 10^{-5} \text{ s.} \quad (2.6)$$

This value is in agreement with measured data from the satellite Lageos (see [57] and [300]). Today's precise radio measurements of the Earth's rotation slow-down, using distant quasars, confirm the average value (2.6) as well (see [287]).

Finally, note that angle measurement devices were also used in the discovery and measurement of the precession and nutation of the Earth's rotational axis.



2.8. Annual parallax of the nearest stars

The Earth's orbital motion around the Sun causes stars to orbit on very small ellipses on the celestial sphere that are called *parallactic ellipses*. Their semimajor axis is greater the closer the star is to us. This enables us to find distances to the nearest stars. The angular size of the semimajor axis of the parallactic ellipse is equal to the so-called annual parallax. Let us introduce now its definition.

Let C stand for a relatively nearby star. Suppose for simplicity that the Earth's orbit is a circle with radius r and center S (Sun). Then there exist two opposite points A and B on this orbit in the plane passing through the center S that is perpendicular to the line CS . The triangle ABC is thus isosceles with base AB (see Fig. 2.8). Then the distance from C to AB is

$$d = \frac{r}{\tan \gamma},$$

where one half of the angle ACB is called the *annual parallax* γ . In other words, γ is the angle at which a hypothetical observer at point C would see the radius r of the Earth's orbit.

The first measurements of annual parallaxes of the nearest stars were carried out by F. W. Bessel in 1838. At present we know that the nearest star to the Sun is Proxima Centauri. Its annual parallax is only $0.76''$ and the distance about $d \approx 4 \cdot 10^{13}$ km ≈ 4.243 ly.

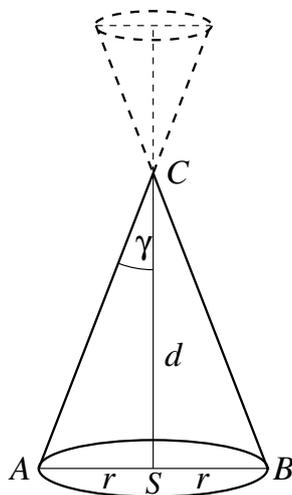


Fig. 2.8. The distance d of a close star at the point C can be determined from the annual parallax γ and the radius r of the Earth's orbit. The line segment AB is parallel to the semimajor axis of the parallactic ellipse (depicted by the dashed line). The angle γ is greatly exaggerated.

The discovery of parallactic ellipses was an important evidence for the orbiting of the Earth around the Sun. Already Tycho Brahe had been looking for these ellipses when trying to decide whether the Ptolemaic or Copernican model of the Solar system was correct. But he did not find any parallax ellipses, because he was not able to measure such small angles by his instruments. The astrometric satellite Hipparcos measured these parallaxes (and hence distances) of more than 100 000 stars in our Galaxy with a very high accuracy of $0.001''$. Another satellite, Gaia, launched at the end of 2013, will measure the parallax of billions of stars.



2.9. Measurement of the speed of light

The aberration of light is generally referred to as the apparent change in the position of a celestial body caused by the movement

of the observer and the finite speed of light. Stars observed perpendicularly¹⁵ with respect to the direction of movement of an observer with speed v seem to be deflected by an *aberration angle* α (see [232]), which satisfies

$$\tan \alpha \approx \frac{v}{c},$$

where c is the speed of light in vacuum. Around 1727 James Bradley discovered the so-called annual aberration. As a result of the revolution of the Earth around the Sun, stars on the celestial sphere circumscribe apparent ellipses (*aberration ellipses*) whose semimajor axes have length $\alpha \approx 20''$ and do not depend on the distance to the star. This phenomenon, which is an order of magnitude larger than the effect of the parallax from Section 2.8, significantly helped to determine an accurate value of the speed of light,¹⁶ and contributed to the confirmation of the heliocentric model. For the average speed of the Earth with respect to the Sun

$$v = \frac{2\pi r}{T} = 29.8 \text{ km/s}, \quad (2.7)$$

where $r = 149\,597\,871$ km and

$$T = 31\,558\,149.5 \text{ s} \quad (2.8)$$

is the sidereal year,¹⁷ we get $c \approx 300\,000$ km/s. The aberration angle

¹⁵In a general case, we have $\tan \alpha \approx (v \sin \delta)/c$, where δ is the angle between the direction of Earth's movement and the direction of an observed star.

¹⁶Already in 1676 the Danish astronomer Olaf Rømer (1644–1710) suggested another elegant method to measure the speed of light, which was later realized by Christiaan Huygens. When the Earth moved toward Jupiter, the time interval between successive eclipses of Jupiter's moon Io became steadily shorter with respect to the terrestrial time. When the Earth moved away from Jupiter these eclipses became steadily longer, i.e., they were behind the expected values. Rømer actually found a phenomenon, which was later named after Christian Doppler. Io's orbital period is 1.769 days, which corresponds to an extremely low Doppler frequency. However, minor deviations in the change of frequency were accumulated so that when Jupiter was in opposition to the Sun, the light rays arrived from Io to the Earth about 22 minutes earlier than when Jupiter was in conjunction. Based on these observations, extrapolation techniques, and the knowledge of the diameter of Earth's orbit, Huygens estimated that the speed of light is approximately $c \approx 2 \cdot 150\,000\,000 / (22 \cdot 60) = 227\,000$ km/s.

¹⁷The *sidereal year* (365.25636 days) is the time period during which the Earth makes one complete revolution around the Sun, i.e. 360° . As a result of the precession of the Earth's axis the *calendar year* (365.2425 days) is shorter.

α is very small, and therefore in radian measure it is almost equal to $\tan \alpha$ (the relative error is less than 10^{-8}). We will thus write only

$$\alpha = \frac{v}{c}. \quad (2.9)$$

The huge gravitational force between the Sun and the Earth (of about $354 \cdot 10^{20}$ N) curves the Earth's trajectory and the direction of its motion around the Sun changes every day by about 1° ($\approx 360^\circ/365.25$ per day). Sunlight photons travel approximately 8.3 minutes from the Sun to the Earth. During this time interval the Sun moves with respect to the stars about an angle of

$$\alpha' \approx \frac{8.3}{60 \cdot 24 \cdot 365.25} 360^\circ \approx 20''. \quad (2.10)$$

Thus we do not see the Sun in its actual position, but shifted by $\alpha' \approx 20''$ (cf. Fig. 6.3). The fact that this angle of a circular path coincides with the above aberration angle α is not a coincidence, but follows from (2.10), (2.7), and (2.9). In the numerator of the fraction in (2.10) there is $r/c = 8.3$ minutes, the denominator is T , and 360° is 2π in radian measure. From this we see that

$$\boxed{\alpha' = \frac{2\pi r}{cT} = \frac{v}{c} = \alpha.}$$



2.10. Spherical trigonometry

When solving problems in celestial mechanics we should have in mind that sometimes we cannot use results from classical Euclidean geometry. For instance, in the plane, the sum of angles of a triangle is 180° . However, on the celestial sphere, the Riemannian spherical geometry is valid in which the sum of the angles α, β , and γ of a spherical triangle is larger than 180° , i.e.

$$\alpha + \beta + \gamma > 180^\circ. \quad (2.11)$$

Let us consider a concrete example.

During summer evenings the southern horizon of the sky is dominated by the so-called *Summer triangle* formed by the stars *Altair* (A) of the constellation Aquila, *Deneb* (B) of Cygnus, and *Vega* (C) of the constellation Lyra (see Fig. 2.9). The French call this striking trio *les Trois belles d'été* (The three beauties of summer). Let us denote the angular length of the opposite sides of the summer triangle successively by a , b , and c . They correspond to the shortest connection paths (*geodesic*) between the considered stars on the celestial sphere, i.e. parts of great circles.

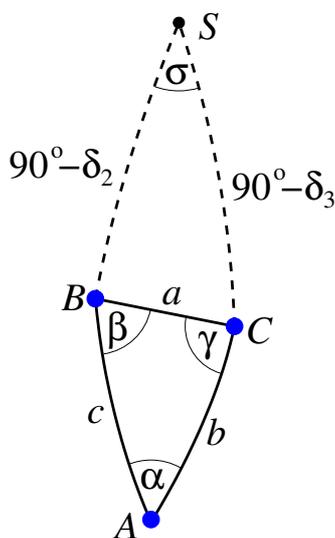


Fig. 2.9. The Summer triangle: A denotes Altair, Deneb B , and C Vega. The dashed lines indicate the meridians and S the North pole of the celestial sphere.

The positions of celestial objects are usually defined by coordinates called the right ascension and declination. The *right ascension* is the hour angle between the plane passing through the two poles and the considered object and the plane passing through both poles and the vernal equinox¹⁸ (where $24 \text{ h} \cong 360^\circ$). The *declination* is the angular separation between the object and the plane passing through the celestial equator. The values of right ascension and declination of stars forming the Summer triangle are:

¹⁸ *Spring* (vernal) and *autumn points* are the intersections of the ecliptic with the celestial equator.

$$\begin{array}{ll}
\text{Altair} & r_1 = 19 \text{ h } 50 \text{ min } 47 \text{ s}, \quad \delta_1 = 8^\circ 52', \\
\text{Deneb} & r_2 = 20 \text{ h } 41 \text{ min } 26 \text{ s}, \quad \delta_2 = 45^\circ 16', \\
\text{Vega} & r_3 = 18 \text{ h } 36 \text{ min } 56 \text{ s}, \quad \delta_3 = 38^\circ 47'.
\end{array}$$

The angular separation between two objects on the celestial sphere is usually expressed in degrees. From the known values of r_i and δ_i we can determine the angular length of the side a corresponding to the arc Deneb–Vega as follows. Two meridians that intersect at the North pole S of the celestial sphere are indicated by dashed lines in Fig. 2.9. The angle $\sigma = \sphericalangle BSC$ between them is obviously equal to the difference of the corresponding right ascensions

$$\sigma \cong r_2 - r_3 = 2 \text{ h } 4 \text{ min } 30 \text{ s} \cong 31.125^\circ. \quad (2.12)$$

Because Vega's declination is δ_3 , the arc length CS equals $90^\circ - \delta_3$. Similar considerations can be made for the arc BS . Then by the law of cosines for the angular length of the side a of the spherical triangle BSC we get (see [224, p. 85])

$$\begin{aligned}
\cos a &= \cos(90^\circ - \delta_2) \cos(90^\circ - \delta_3) + \sin(90^\circ - \delta_2) \sin(90^\circ - \delta_3) \cos \sigma \\
&= \sin \delta_2 \sin \delta_3 + \cos \delta_2 \cos \delta_3 \cos \sigma = 0.91462447.
\end{aligned}$$

The angular arc length of Deneb–Vega is therefore

$$a = 23.848^\circ.$$

Analogously, we may derive that the arcs Altair–Vega and Altair–Deneb have angular length $b = 34.197^\circ$ and $c = 38.003^\circ$. Hence, using again the law of cosines for the spherical side of the triangle ABC ,

$$\cos a = \cos b \cos c + \sin b \sin c \cos \alpha, \quad (2.13)$$

we obtain the angle $\alpha = 40.566^\circ$ at the star Altair. Similarly, we find the angle $\beta = 64.695^\circ$ at Deneb and the angle $\gamma = 82.036^\circ$ at Vega. We see that for the sum of the angles in the Summer triangle $\alpha + \beta + \gamma = 187.297^\circ$, i.e., the inequality (2.11) is valid.

Because we use the non-Euclidean spherical geometry, a triangle may have two (or even three) right angles. For example, the North pole S in the vicinity of Polaris forms with two other stars on the celestial equator such a spherical triangle. Finally, note that the whole celestial sphere has 4π sr (steradians) which is

$$\frac{4\pi}{(\pi/180^\circ)^2} = \frac{(360^\circ)^2}{\pi} \approx 41\,253 \text{ square degrees}, \quad (2.14)$$

where one square degree is $(\pi/180)^2$ sr.



2.11. Deflection of light in a gravitational field

In his pioneering paper [69] from 1911, Albert Einstein derived that light deviates from rectilinear motion in the gravitational field of massive objects. This remarkable effect was first photographed during the total eclipse of the Sun in 1919, when the light rays of stars in a close neighbourhood of the Sun's disc were deflected from their original direction. By comparison of this photo with pictures of the same night part of the celestial sphere, the deflection was found to be in good agreement with the value $1.75''$ predicted by Einstein. In this way, the angular measurements of the deflection contributed to accepting the validity of Einstein's General theory of relativity and to the explanation of the principle of gravitational lenses.

Thus each mass object makes spacetime locally curved. Light photons travel along shortest paths (geodesics). In Fig. 2.10, we observe two examples of the bending of light in a close neighbourhood of stars. The three trajectories in Fig. 2.10 a) form a curved triangle. Notice that the sum of its angles satisfies the inequality $\alpha + \beta + \gamma > 180^\circ$, which corresponds to relation (2.11) of Riemannian elliptic geometry. On the other hand, in Fig. 2.10 b) we see two stars of equal masses and three trajectories forming another curved triangle whose sum of angles satisfies $\alpha + \beta + \gamma < 180^\circ$, which is a manifestation of Lobachevskian hyperbolic geometry.

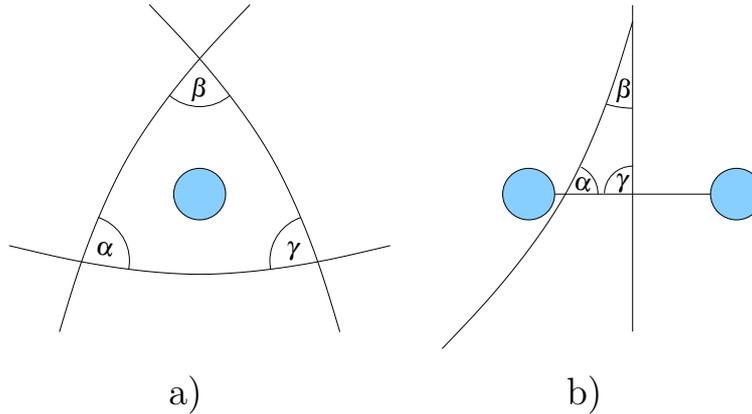


Fig. 2.10. Bent trajectories of light near massive objects show that the geometry of the universe can be locally a) elliptic (Riemannian) or b) hyperbolic (Lobachevskian).

These two examples show that the universe has locally different kinds of geometries with various curvatures. However, to find a “global curvature” of the universe for a fixed time (see Chapter 18), we have to consider very large scales, on which all local curvatures are averaged. This is like Earth’s surface, whose curvature locally changes very much (due to mountains, valleys, saddle-points, etc.), but whose global curvature is positive, and almost constant. According to Einstein’s cosmological principle, our universe on each isochrone is homogeneous and isotropic on large scales, i.e., its curvature is constant at any point and in any direction. This assumption is continually verified by astronomers. For instance, it is confirmed by the homogeneity and isotropy of the cosmic microwave background radiation¹⁹. Also the well-known γ -ray bursts show almost uniform distribution on the celestial sphere. Pictures of the Hubble Deep Field and Hubble Deep Field-South taken by the Hubble space telescope illustrate this homogeneity and isotropy, too.

On the other hand, it is known that the universe on scales of about 100 million light-years consists of such large-scale structures as giant sheets and long fibers²⁰. Therefore, the density of galaxies from

¹⁹The cosmic microwave background (CMB) radiation possesses tiny fluctuations of order 10^{-4} K from its mean temperature 2.725 K.

²⁰For example, a fiber consisting of many thousands galaxies with a total length of 1.37 billion light years was discovered, the so-called Sloan Great Wall.

equally distant regions of the universe from us in different parts of the sky is systematically compared. The global curvature of the universe significantly depends on so-called nonbaryonic dark matter whose distribution is the subject of intense study (see Chapters 7–9). When would it be demonstrated that Einstein’s cosmological principle holds on very large scales, we would get very restrictive conditions on the global topology of the universe.

Angle measurements helped to discover another effect of Einstein’s General theory of relativity in a vicinity of the brightest star of the night sky — Sirius. In 1844, F.W. Bessel used angle measurements to find that the path of Sirius (α CMa) is affected by an invisible companion. Only after the death of Bessel, this companion was observed by A.G. Clark. It was calculated that the newly discovered body has approximately the mass of the Sun. However, at that time nobody was surprised that the absolute luminosity is about five orders of magnitude lower than that of the Sun. In 1914, W. Adams proved that Sirius B, as this body was called, is a white dwarf with an incredible density of several hundreds kilograms per cubic centimeter (see Section 4.2). The precise angle measurements thus actually helped in the discovery of the first white dwarf. In 1924, Adams discovered a gravitational redshift²¹ of spectral lines that was predicted by A. Einstein for all material objects.

Angle measuring instruments played an important role also in other effects of Einstein’s General theory of relativity, e.g. in determining the perihelion advance of Mercury or the Lense-Thirring precession effect of a gyroscope moving in curved spacetime near the rotating Earth. In Chapter 20, we shall see how to interpret the measurement of angles, which apparently lead to the observation of superluminal ve-

²¹The *redshift* z is defined by $z = (\lambda - \lambda_0)/\lambda_0$, where λ_0 denotes the wavelength of a spectral line of a particular atom or molecule in a terrestrial laboratory and λ is the corresponding measured wavelength of the observed object. It is not hard to check that z is defined independently of the selected spectral line. For instance, the spectral H_α line corresponds to an electron that jumps from the third to the second energy level of the hydrogen atom, where the radius of the second and third energy level is $2^2 a_0$ and $3^2 a_0$, respectively, and $a_0 = 5.3 \cdot 10^{-11}$ m is the Bohr radius. In this case a photon of wavelength $\lambda_0 = 656.3$ nm is emitted.

locities in the universe (cf. also [183]). In Chapter 4, we show how to use angular measurements to determine the mass of the black hole in the center of the Milky Way (also see [127]). Angle measurements thus significantly contributed to shaping our modern view of the universe.



3. On Kepler's equation

*Where there is matter, there is geometry.
(Ubi materia, ibi geometria.)*

JOHANNES KEPLER

3.1. True and eccentric anomaly

Kepler's equation is one of the most amazing of Kepler's mathematical results. It nicely illustrates his immense genius. By Kepler's first law, orbits of the planets are elliptic and the Sun is located at one of the two foci. Each planet is generally moving along its orbit nonuniformly, and it is therefore important to be able to determine its coordinates at a given time instant in the plane of its elliptic path, and those are what is determined by the Kepler equation. The actual position on the ellipse can be described by the angle called the eccentric anomaly. In Section 3.3 we derive Kepler's equation that relates the eccentric anomaly with uniformly passing time, also see [20, p. 304].



Fig. 3.1. Johannes Kepler (1571–1630)

Let $a \geq b$ be the semimajor and semiminor axis of an elliptic orbit, respectively. Then for its eccentricity we have

$$e = \frac{\sqrt{a^2 - b^2}}{a}. \quad (3.1)$$

It will be helpful to consider a circle of radius a with the same center as the ellipse. Further, denote by $r > 0$ the heliocentric distance of the planet P from the Sun S , with points A , B , C , and angles α and φ as depicted in Fig. 3.2. Then we observe that $|AS| = ae$ and

$$a \cos \alpha = |AC| = ae + r \cos \varphi, \quad (3.2)$$

$$\frac{r \sin \varphi}{a \sin \alpha} = \frac{|PC|}{|BC|} = \frac{b}{a},$$

where the angle α is called the *eccentric anomaly* and the angle φ the *true anomaly*. After the use of (3.1) from both equalities we get

$$\begin{aligned} r^2 \cos^2 \varphi &= a^2 \cos^2 \alpha - 2a^2 e \cos \alpha + a^2 e^2, \\ r^2 \sin^2 \varphi &= b^2 \sin^2 \alpha = a^2(1 - e^2) \sin^2 \alpha. \end{aligned}$$

If we sum these equalities, we obtain $r^2 = a^2 - 2a^2 e \cos \alpha + a^2 e^2 \cos^2 \alpha$. From this we get the dependence of the distance r on the eccentric anomaly α ,

$$r = a(1 - e \cos \alpha). \quad (3.3)$$

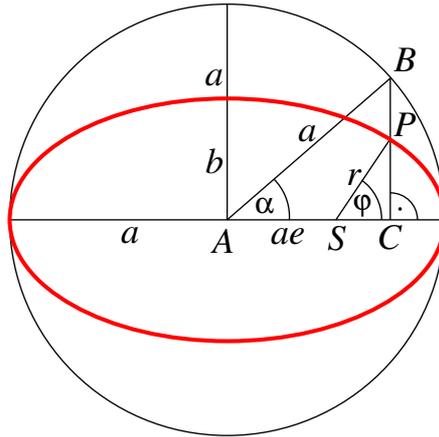


Fig. 3.2. The Sun S is at the focus of the ellipse along which the planet P orbits. The eccentric anomaly α can be determined from Kepler's equation (3.6) and the true anomaly φ can then be derived from equation (3.5).



3.2. The relationship between the true and eccentric anomaly

Let us now express the true anomaly φ using α . Substituting $\cos \alpha$ from relation (3.2) into (3.3), we see that $r = a - e(ae + r \cos \varphi)$, i.e.

$$\frac{r}{a}(1 + e \cos \varphi) = 1 - e^2. \quad (3.4)$$

In this way, sometimes Kepler's first law is expressed in polar coordinates (r, φ) . From (3.2) divided by a and from (3.4) we obtain

$$\cos \alpha = e + \frac{1 - e^2}{1 + e \cos \varphi} \cos \varphi = \frac{e + \cos \varphi}{1 + e \cos \varphi}.$$

Substituting this into the relation for the tangent of the half angle we have

$$\begin{aligned} \tan^2 \frac{\alpha}{2} &= \frac{1 - \cos \alpha}{1 + \cos \alpha} = \frac{1 - \frac{e + \cos \varphi}{1 + e \cos \varphi}}{1 + \frac{e + \cos \varphi}{1 + e \cos \varphi}} = \frac{1 + e \cos \varphi - e - \cos \varphi}{1 + e \cos \varphi + e + \cos \varphi} \\ &= \frac{1 - e}{1 + e} \cdot \frac{1 - \cos \varphi}{1 + \cos \varphi} = \frac{1 - e}{1 + e} \operatorname{tg}^2 \frac{\varphi}{2}. \end{aligned}$$

In this way we get the desired expression for the true anomaly in terms of the eccentric anomaly

$$\varphi = 2 \arctan \left(\sqrt{\frac{1 + e}{1 - e}} \tan \frac{\alpha}{2} \right). \quad (3.5)$$



3.3. Kepler's equation for the eccentric anomaly

Now we derive an equation to calculate α . Consider the planet's perihelion passage at $t = 0$ and let T be the orbital period. According to Kepler's second law, the planet's areal speed is constant and is equal to $\pi ab/T$ (see Fig. 1.5 and [127]). The radius vector SP circumscribes the area of $\pi abt/T$ during the time interval $[0, t] \subset [0, T]$. If the ellipse is linearly transformed onto a circle along the vertical axis, then the

segment SB circumscribes during the same time interval the area of $\pi a^2 t/T$ (i.e. a/b times greater) and the segment AB circumscribes the area of $\pi a^2 \alpha(t)/(2\pi)$. Their difference is equal to the area of triangle ASB (see Fig. 3.2),

$$\frac{a^2 \alpha(t)}{2} - \frac{\pi a^2 \pi t}{T} = \frac{ae}{2} a \sin \alpha(t),$$

where $\alpha(t)$ expresses the fact that the angle α depends on time t . After division by the number $-a^2/2$ we finally arrive at *Kepler's equation*, sometimes called *Kepler's time equation*, for the eccentric anomaly

$$\boxed{\frac{2\pi}{T}t = \alpha(t) - e \sin \alpha(t).} \quad (3.6)$$

The left-hand side $M(t) = 2\pi t/T$ is called the *mean anomaly*, because it is a linear function of time.

For a fixed time instant t , we can determine the angle $\alpha = \alpha(t)$ from Kepler's equation (3.6) by means of a suitable iterative method (method of successive approximations, Newton's method, etc.). The sine in the (transcendental) equation (3.6) can also be approximated by a polynomial from its associated Taylor series, and then we solve only the algebraic equation, e.g. $M(t) = \alpha - e\alpha + e\alpha^3/3! - e\alpha^5/5!$. From equation (3.3) we then determine the distance $r(t)$ and from (3.5) we calculate $\varphi(t)$. In this way, the position of the planet is uniquely determined.

Kepler's equation (3.6) can also be derived by integral calculus. For the angle $\varphi(t) \in [0, 2\pi]$ at time t we get the equality for areas (cf. also [20], p. 304)

$$\frac{\pi ab}{T}t = \frac{1}{2} \int_0^{\varphi(t)} r^2(\varphi) d\varphi, \quad (3.7)$$

where $\varphi \in [0, \varphi(t)]$ is the integration variable and $r(\varphi)$ expresses the fact that the distance r depends on the angle φ . This integral can be calculated using substitution of (3.5). By differentiation it can be deduced that

$$d\varphi = \frac{\sqrt{1-e^2}}{1-e\cos\alpha}d\alpha = \frac{b}{a(1-e\cos\alpha)}d\alpha. \quad (3.8)$$

Relations (3.7), (3.3), and (3.8) then yield

$$\frac{2\pi}{T}t = \frac{1}{ab} \int_0^{\alpha(t)} a^2(1-e\cos\alpha)^2 \frac{b}{a(1-e\cos\alpha)}d\alpha = \int_0^{\alpha(t)} (1-e\cos\alpha)d\alpha,$$

where $0 \leq \alpha \leq \alpha(t)$. From this Kepler's equation (3.6) immediately follows.



3.4. Keplerian orbital elements

Consider again an elliptic orbit with semimajor axis a and eccentricity e . To describe the motion of a body outside the ecliptic plane, four further elements are added. The *inclination* i to the ecliptic plane and the *longitude of the ascending node* Ω determines the orbital plane.¹ The *argument of perihelion* ω defines the orientation of the elliptic orbit in three-dimensional space. The angle ω is measured from the ascending node to the perihelion (see Fig. 3.3). The five Keplerian elements $(a, e, i, \Omega, \omega)$ for a unique determination of the position of the body should be supplemented by a sixth parameter making the orbit uniquely determined. This can be either the instant of perihelion passage or the mean anomaly, see (3.6). For more details about these six Keplerian orbital elements (i.e. ephemerides) see e.g. [20].

At present we admire how Kepler derived equation (3.6) without knowledge of integral calculus. He must have had a huge geometric imagination and physical intuition. He had only a large amount of data on the positions of the planets seen from a moving Earth and he had to transform them to a heliocentric coordinate system. Therefore, it was quite complicated.

¹The longitude is measured from the vernal equinox which moves about 50.27'' per year along the ecliptic. Thus the Keplerian elements are related to some year, e.g. the epoch 2000.

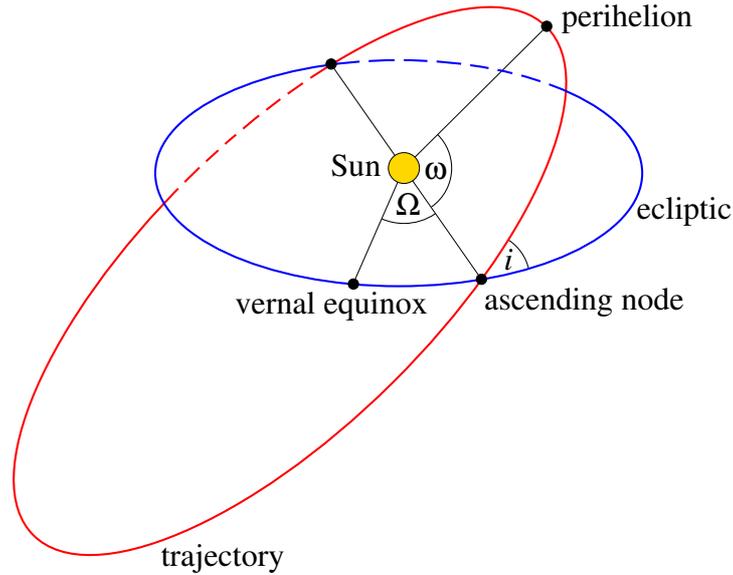


Fig. 3.3. Elliptical orbit elements that determine its orientation in space, the inclination i , longitude of the ascending node Ω and the argument of perihelion ω .

Kepler also dealt with the calculation of the length of the elliptic orbit. In 1609, he proved a guaranteed lower bound using the geometric mean

$$2\pi\sqrt{ab} \leq L(a, b),$$

where $L(a, b)$ is the *circumference of an ellipse* with semiaxes a and b . Later Leonhard Euler introduced the two-sided bound²

$$\pi(a + b) \leq L(a, b) < \pi\sqrt{2(a^2 + b^2)}. \quad (3.9)$$

Unfortunately, for $a \neq b$ the circumference of an ellipse cannot be determined exactly. However, an expression from Colin MacLaurin is known in the form of an infinite series

$$\begin{aligned} L(a, b) &= \int_0^{2\pi} \sqrt{a^2 \sin^2 s + b^2 \cos^2 s} \, ds \\ &= 2\pi a \left[1 - \underbrace{\left(\frac{1}{2}\right)^2}_{\odot} e^2 - \underbrace{\left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2}_{\odot} \frac{e^4}{3} - \underbrace{\left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2}_{\odot} \frac{e^6}{5} - \dots \right]. \end{aligned}$$

²The first two-sided bound was derived by Archimedes (287–212 BC) for the number π . He inscribed and circumscribed regular polygons to the unit circle.

4. The law of gravity — the discovery of the millennium

An apple fall down on the heads of thousands of people, but only one guessed why.

A paraphrased quotation

4.1. Newton's theorems

Newton's law of universal gravitation played a crucial role in the development of physics. It significantly contributed to understanding the structure and evolution of the universe. Sometimes it is called the discovery of the last millennium, as will be indeed seen from the following paragraphs. It was first formulated by Sir Isaac Newton (see Fig. 4.1) in his treatise *Principia* from 1687. In particular, he was inspired by Kepler's third law.

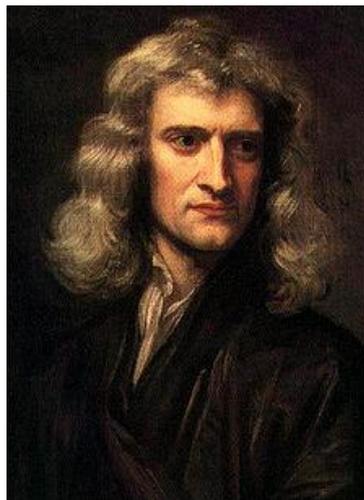


Fig. 4.1. Sir Isaac Newton (1643–1727) showed that Kepler's laws are just consequences of the gravitational law.

According to Newton's law of gravitation, the size of the gravitational force between two point masses is equal to

$$\boxed{F = G \frac{mM}{r^2}}, \quad (4.1)$$

where m and M are their masses, r is their distance, and

$$G = 6.674 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2} \quad (4.2)$$

is the *gravitational constant*.

In some textbooks it is incorrectly stated that F in (4.1) is the size of gravitational force between two bodies. In doing so, the reader will not know how to define precisely their distance. For example, what is the distance r between a homogeneous ring and a point mass at its center? That is: \odot

1. If the distance r is equal to the radius of the ring, we do not get the correct answer by (4.1), since the resulting total force (and thus also its size) is zero.
2. However, for $F = 0$ relation (4.1) yields $r = \infty$, which certainly has nothing to do with reality.
3. If r is the distance between the centers of gravity of both bodies, then we surprisingly obtain $F = \infty$, because the denominator in (4.1) is $r = 0$. Again, we get into trouble.

Thus we see that a mechanical use of the law of gravity can lead to unexpected paradoxes (see also Remark 4.1 below). In what follows we shall, therefore, often restrict ourselves only to idealized point masses, which in the real world do not exist. However, relation (4.1) remains unchanged if instead of point masses we consider balls with special density distributions as follows from the next theorem.

Theorem 4.1 (Newton's first theorem). *If the density distribution of a ball of mass M is spherically symmetric, then the size of the force between the ball and a point mass m , that lies outside the interior of the ball, is given by (4.1), where r is the distance between the point and the center of the ball.*

The proof is based on the specific shape of the gravitational potential of mass uniformly distributed on a spherical surface. The resulting equation (4.1) is then obtained by integration. Theorem 4.1 can evidently be generalized to the interaction between two balls with spherically symmetric mass density distribution (see Fig. 4.2). In actual bodies with diameter over 1 000 km gravity automatically arranges approximately spherical shape and also spherically symmetric mass density distribution. This process is called *gravitational differentiation*.

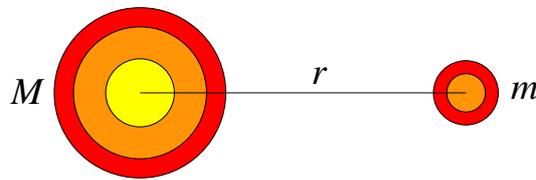


Fig. 4.2. Illustration of Newton's first theorem for two spherically symmetric bodies

Remark 4.1. In general, bodies that do not have a spherically symmetric mass density distribution cannot be replaced by point masses at their centers of gravity.¹ To see this, it is enough to consider how a dumbbell-shaped body with mass $M = M_1 + M_2$ acts on the second body, which is a homogeneous ball of mass m . The mass of the straight middle part of the dumbbell is neglected. Set $M_1 = M_2 = m = 1$ kg. The scale on the horizontal axis in Fig. 4.3 is indicated in meters. Then the magnitude of the total force between the two bodies is equal to

$$F = G \frac{mM_1}{3^2} + G \frac{mM_2}{1^2} = \frac{10G}{9}.$$

However, if we concentrate the mass of the dumbbell at its center of gravity at 0, then the resulting force \bar{F} will be essentially different from F , i.e.

$$\bar{F} = G \frac{m(M_1 + M_2)}{2^2} = \frac{G}{2} \quad \text{and} \quad 2\bar{F} < F.$$

¹This concerns e.g. the elongated asteroid Ida which is orbited by the satellite Dactyl.

The quadratic nonlinearity r^2 in equation (4.1) thus causes that the force F is more than two times bigger than the force \bar{F} corresponding to the mass of the first body concentrated at its center of gravity.

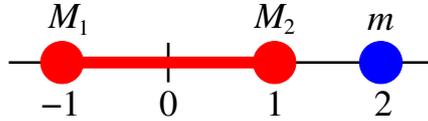


Fig. 4.3. The force between two bodies cannot generally be substituted by the force between two point masses located at their centers of gravity.

Another important assertion for a spherical layer states (see Fig. 4.4):

Theorem 4.2 (Newton's second theorem = Shell theorem).
A spherical layer with spherically symmetric mass density distribution exerts no force on a point mass located inside.

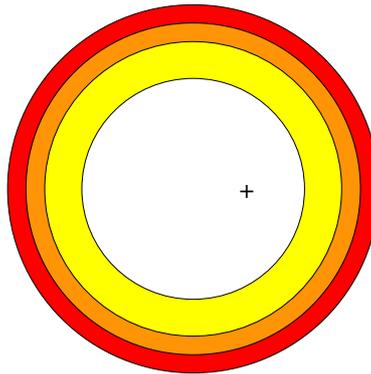


Fig. 4.4. Illustration of Newton's second theorem in a cross-section of the spherical layer. The force acting on a point mass inside the cavity marked by + is zero for a spherically symmetric mass density distribution. The gravitational potential inside the cavity is constant.



4.2. The most important discoveries and applications

In 1798, the British physicist and chemist Lord Henry Cavendish estimated the mean density and mass of the Earth using a torsion balance and large lead balls [51]. A hundred years later his method

led to the value $G \approx 6.75 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$. Therefore, the gravitational constant is sometimes also called the *Newton–Cavendish constant*. Cavendish is also known for the discovery of hydrogen and the chemical composition of water.

A great triumph of the Newtonian gravitational law was the discovery of the planet Neptune due to observed perturbations in the orbit of Uranus which speeded up to $7''$ per year until 1831 and then it slowed down. A supposed position of Neptune was initially calculated by the English mathematician and astronomer John Couch Adams and independently by the French astronomer Urbain Jean Joseph Le Verrier. In 1846, the German astronomer Johann Gottfried Galle discovered Neptune (see Fig. 4.5) less than one degree from the position calculated by Le Verrier by the least squares method for ephemerides from the O–C diagram (Observed minus Calculated).

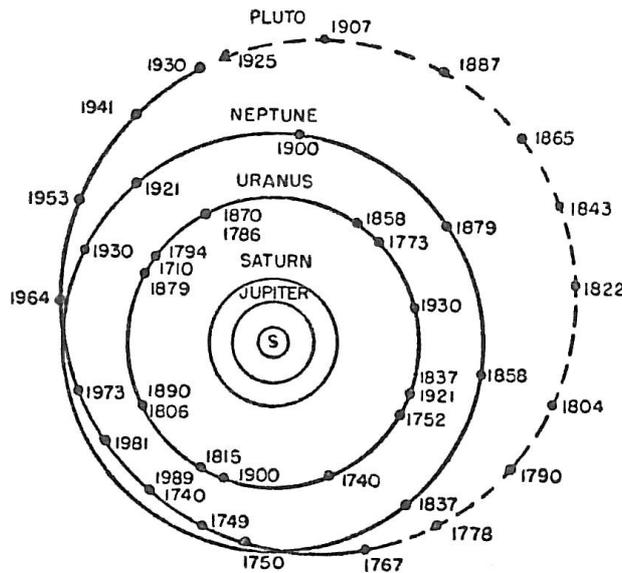


Fig. 4.5. The positions of the planets Uranus and Neptune

Using (4.1), several unusual types of celestial bodies were found. For instance, in 1862 the American optician Alvan G. Clark observed a barely visible companion of the star Sirius, which later led to the discovery of superdense objects — white dwarfs whose densities are of order 10^7 to 10^{11} kg/m^3 . Later, densities of order 10^{18} kg/m^3 were established in neutron stars. The relation (4.1) is also applied to determine the mass of black holes (see Section 4.14).

The Newtonian gravitational law played an important role in the postulation of dark matter. In 1933, the American astrophysicist Fritz Zwicky found that the galaxy cluster in the constellation Coma Berenices contains about one thousand galaxies that orbit much faster around its center than should result from the gravitational law (more precisely from the Virial theorem). He predicted the existence of a mysterious dark matter in the universe. We will examine this topic in Chapters 7 and 8.

Newton's universal law of gravitation enables us to design and calculate trajectories of spacecrafts, to get extraordinary data about the universe, to predict a collision of an asteroid or comet with the Earth, to estimate how much we would weigh on Mars, and so on. In this chapter we will introduce some simple and interesting applications of Newton's law of gravitation. We shall see how to use it to identify some seemingly undetectable values of physical quantities at large distances (e.g. the mass of Mars or the mean density of the Sun). The gravitational law has helped us to push the boundaries of human knowledge far ahead in unexpected directions.



4.3. The size of a constant in Kepler's third law

Kepler's third law in its simplest form says (see Section 1.2) that the ratio of the third power of the semimajor axis a of an elliptic orbit of a planet to the square of its orbital period T is constant, i.e. $a^3/T^2 = C$. From this we may calculate e.g. distances of all the planets from the Sun knowing their orbital periods and the Earth–Sun distance. However, Kepler did not know the actual value of the constant C .

For a body orbiting the Sun it can easily be deduced from (4.1), e.g. for a circular planetary orbit² of radius $r = a$. If m is the mass of a planet and $M = M_\odot$ the mass of the Sun, then according to another

²Derivation for an elliptic orbit can be found e.g. in [127].

law of Newton, the law of action and reaction, the gravitational force is equal to the centripetal force, i.e.,

$$G\frac{mM}{a^2} = \frac{mv^2}{a}, \quad (4.3)$$

where v is the orbital velocity of the planet. Substituting for $v = 2\pi a/T$ into (4.3), we obtain as a result Kepler's third law in the form

$$\frac{a^3}{T^2} = \frac{GM}{4\pi^2} (= C), \quad (4.4)$$

where it is implicitly assumed that $m \ll M$. From this we can get other important information, as we shall see later. Otherwise the (generalized) third law of Kepler can be written as³

$$\frac{a^3}{T^2} = \frac{G(M+m)}{4\pi^2}, \quad (4.5)$$

where $a = a_1 + a_2$ and a_1, a_2 are the semimajor axes of elliptic orbits of both components with respect to their common center of gravity.

The determination of the length of the semimajor axis of Earth's orbit was gradually refined by various methods [19], some of which are based just on Kepler's third law. It is practically equal to the astronomical unit, i.e. the median Earth–Sun distance,

$$1 \text{ au} = 149\,597\,870\,700 \text{ m} \approx 149.6 \cdot 10^6 \text{ km}. \quad (4.6)$$

By relation (4.6), the astronomical unit au is connected with the International System of Units SI (the metric system).

Expressing the semimajor axis a of a planetary orbit in astronomical units and the period T in years, Kepler's third law (4.4) can be written more easily as

$$a^3 \cong T^2, \quad (4.7)$$

where the symbol \cong denotes the equality of numerical values when expressed in appropriate units. Multiplying the right-hand side of (4.7)

³Thus two bodies of unequal masses orbiting around the Sun at the same radius have different speeds, which does not follow from the classical third law of Kepler (4.4).

by the constant $C = 1 \text{ au}^3/\text{yr}^2$, where yr indicates the year, we moreover get dimensional equality. The above notation has several advantages, e.g., we can easily calculate the distances of all planets from the Sun from mere knowledge of their orbital periods. Let us illustrate it on the planet Mars, whose orbital period around the Sun is $T = 1.881$ yr. Therefore, according to (4.7), we have

$$a = 1.881^{2/3} = 1.524 \text{ (au)}.$$



4.4. Mass of the Sun

First observe that equation (4.4) connects three important quantities with the physical dimensions of length, mass, and time. In the system of units SI, they are expressed in

$$\text{m, kg, s.}$$

Knowing two of them, we can calculate the third one. Therein lies the beauty of Kepler's third law. Substituting $a = 1 \text{ au}$ from (4.6) into equation (4.4), we immediately get the mass of the Sun

$$M_{\odot} = \left(\frac{2\pi a}{T}\right)^2 \frac{a}{G} = 1.99 \cdot 10^{30} \text{ kg}, \quad (4.8)$$

where $T = 31\,558\,149.54 \text{ s}$ is the orbital period of the Earth around the Sun (the so-called *sidereal year*).



4.5. Mass of Mars

In 1877, the American astronomer Asaph Hall discovered the Martian moon Phobos, whose orbital period is

$$P = 27\,554 \text{ s,}$$

i.e. 0.3189 days. This fact has allowed us to refine significantly our knowledge of Mars' mass m . Using angular measurements, we can

estimate the length of the semimajor axis of the path of Phobos as $r \approx 9.377 \cdot 10^6$ m. Therefore, by (4.4)

$$\frac{r^3}{P^2} = \frac{Gm}{4\pi^2} \quad (4.9)$$

and from this we get

$$m = 6.42 \cdot 10^{23} \text{ kg.}$$

The mass of Mars can be estimated without knowledge of the gravitational constant G , if we know the Sun's mass M_\odot . From relations (4.4) and (4.9) it follows that

$$m = \frac{r^3 T^2}{a^3 P^2} M_\odot,$$

where $T = 59\,355\,072$ s (i.e. 686.971 days) is the orbital period of Mars, and $a = 227.94 \cdot 10^9$ m is the length of the semimajor axis of its elliptic orbit (which can also be found by (4.4)).

Similarly, we can calculate the mass of the outer planets and of the Earth, too. To specify the mass of Venus, the Magellan satellite orbiting around it was used in an analogous manner, and the Messenger satellite was used in a similar way to determine the mass of Mercury.



4.6. Falling into the Sun

From equation (4.8) for a circular orbit with radius a we get the orbital velocity of a planet

$$v = \frac{2\pi a}{T} = \sqrt{\frac{GM_\odot}{a}}. \quad (4.10)$$

Substituting for a the Earth–Sun distance from (4.6) and M_\odot mass of the Sun (4.8), we find the Earth's mean orbital velocity

$$v = 29.8 \text{ km/s.} \quad (4.11)$$

Assume for a moment that someone stops an object on this orbit. Then it actually will fall towards the Sun. The question is how long will this fall last.

If the Earth were to orbit at half of its distance from the Sun, then by Kepler's third law (4.7), the period would be $0.5^{3/2} = \sqrt{2}/4$ years. However, if the Earth's path were an elongated ellipse with semimajor axis 0.5 au which in the limiting case is equal to the Earth–Sun line segment, then it would also orbit the Sun in $\sqrt{2}/4$ years. The Earth's descent into the Sun would consequently take half of this time, i.e., $\sqrt{2}/8$ years = 64.6 days.

Analogously, we can derive that Neptune would fall onto the Sun within $\frac{1}{2}15^{3/2} \approx 29$ years and our Moon would fall onto the Earth during 4.83 days, if they would also be halted on their orbits.



4.7. The size of escape velocities

A body orbiting the Earth along a circular trajectory with a radius of about $r = 6550$ km (i.e. just above the dense layers of the atmosphere, where it can perform at least one orbit), would need to have *escape velocity* v_I to achieve a stable orbit. Similarly to (4.10) we find⁴ that

$$v_I = \sqrt{\frac{GM}{r}} = 7.9 \text{ km/s}, \quad (4.12)$$

where

$$M = 5.9736 \cdot 10^{24} \text{ kg} \quad (4.13)$$

is the mass of Earth. According to [20, p. 303], the corresponding instantaneous orbital velocity on an elliptic orbit with semimajor axis a is

$$v^2 = GM \left(\frac{2}{r} - \frac{1}{a} \right), \quad (4.14)$$

where r denotes the distance of the body from the center of the Earth.

In order to escape a body with a negligible mass $m \ll M$ and orbiting at a distance r from Earth's gravitational field, it is necessary

⁴Rockets are usually launched from locations close to the equator in the direction of Earth's rotation, where they obtain for free initial speeds of up to $40000 \text{ km} / (24 \cdot 3600 \text{ s}) = 0.46 \text{ km/s}$.

to travel with at least the *escape velocity* v_{II} from the Earth. If the potential and kinetic energy of a body at infinity is zero, then at a distance r from the Earth the sum of potential and kinetic energy also has to be also equal to zero, i.e.

$$-G\frac{mM}{r} + \frac{1}{2}mv_{\text{II}}^2 = 0.$$

From this and (4.12) we get

$$v_{\text{II}} = \sqrt{\frac{2GM}{r}} = \sqrt{2}v_{\text{I}} = 11.2 \text{ km/s}, \quad (4.15)$$

where again $r = 6450 \text{ km}$. Analogously, for a body orbiting the Moon the escape velocity is 2.3 km/s and for the Sun 617.3 km/s for the corresponding radii of both bodies.

For the *escape velocity from the Solar system* from the distance of 1 au we get by (4.15)

$$v_{\text{III}} = \sqrt{2}v = 42.1 \text{ km/s},$$

where v is given by (4.11). To give to some earthly body the speed v_{III} relative to the Sun, we should use the fact that the body has already velocity (4.11). However, it is not enough to increase its speed about $v_{\text{III}} - v = 12.3 \text{ km/s}$ in the tangential direction, because the body still needs to leave Earth's gravitational field, which requires at least the speed v_{II} . Hence, the total kinetic energy needed for to escape the Solar system is at least

$$\frac{1}{2}m\tilde{v}_{\text{III}}^2 = \frac{1}{2}mv_{\text{II}}^2 + \frac{1}{2}m(v_{\text{III}} - v)^2,$$

where the resulting speed is

$$\tilde{v}_{\text{III}} = \sqrt{11.2^2 + 12.3^2} \approx 16.6 \text{ (km/s)}$$

and the body should travel in the tangential direction of the motion of the Earth.



4.8. Flight altitude of geostationary satellites

A geostationary orbit was described by the Slovenian physicist Herman Potočnik already in 1928. For a body orbiting the Earth above the equator and remaining above the same point of the Earth's surface, it is necessary that its orbital period is the same as the sidereal rotation period (i.e. with respect to the stars) of the Earth around its axis, namely $T = 23 \text{ h } 56 \text{ min } 4 \text{ s} = 86\,164 \text{ s}$. This idea was used in 1945 by the science-fiction writer Arthur C. Clarke for designing radio communication satellites. Therefore, the geostationary orbit is sometimes called the *Clarke orbit*. According to Kepler's third law (4.4), the height of geostationary satellites above the equator is equal to

$$h = \sqrt[3]{\frac{GMT^2}{4\pi^2}} - R \approx 35\,786 \text{ km},$$

where $R = 6\,378 \text{ km}$ denotes the Earth's equatorial radius. Geostationary satellites thus orbit at an altitude five times larger than the radius R above the Earth's surface. From there one can see 42.4 % of Earth's surface.



4.9. The flight time on Mars

Set $r = 1.524$ and assume, for simplicity, that the Earth and Mars orbit the Sun on circular trajectories with radii of 1 au and r au, respectively. The economic path⁵ of a probe from the Earth to Mars and back (the so-called Hohmann⁶ transfer orbit or trajectory) is depicted in Fig. 6.4. It is essentially elliptic, because after launching, this probe switches off its power engines and for most of the flight time to Mars its trajectory is determined by the gravitational field of the Sun. The

⁵Some probes, however, do not use this economic path for various reasons involving decisions regarding the time for the probe to reach its destination. For example, Mariner 7 reached Mars in just 128 days.

⁶The German engineer Walter Hohmann (1880–1945) determined optimal trajectories of spacecrafts.

real situation is actually much more complicated.⁷ The semimajor axis of the elliptic orbit of the probe is $a = \frac{1}{2}(r + 1)$ au (see (1.1)) and the Sun is at one of its focal points. According to Kepler's third law (4.7), the time required for the flight to Mars is equal to

$$\frac{1}{2}T \cong \frac{1}{2} \left(\frac{r + 1}{2} \right)^{3/2} \cong 0.7 \text{ yr},$$

which corresponds to 259 days.

The semiminor axis of an elliptic orbit is equal to $b = \sqrt{r}$ au, as follows immediately from (1.2). The total length of the Hohmann transfer orbit is by (3.9) equal to 7.843 au. Let us also note that the linear eccentricity equals the distance between the center and the focus, i.e. $\varepsilon = \frac{1}{2}(r - 1)$ au, and its numerical eccentricity is

$$e = \frac{r - 1}{r + 1} = 0.208.$$

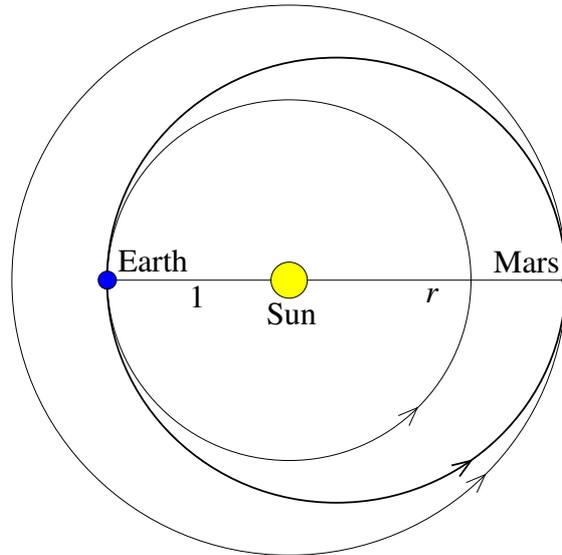


Fig. 4.6. The economic Hohmann transfer orbit of a probe from the Earth to Mars uses the fact that the Earth moves around the Sun at speed (4.11). It should be launched during the so-called launch window to reach the orbit of Mars in the region, where Mars will appear within 0.7 years.

⁷Planetary orbits are not circular and they are not in the same plane. The probe has first to obtain the escape velocity from the Earth. Sometimes it is necessary to correct its path, etc.

In a similar manner a flight to Pluto from the Earth would take 46 years. Therefore, for the probe New Horizons launched in 2006 a different trajectory was selected. Using the large attraction force of Jupiter (i.e. a gravitational slingshot) allowed one to reach Pluto in just nine years from launch with an initial speed of 16.26 km/s.



4.10. Mean mass density of the Sun

Let us show now how to calculate the mean mass density of the Sun by protractor. To solve this seemingly absurd problem assume for simplicity that the Earth's orbit is circular. By Newton's law of gravitation, the second and third law (of action and reaction), we get

$$G \frac{M_{\odot} m}{r^2} = \frac{mv^2}{r}, \quad (4.16)$$

where M_{\odot} is the mass of the Sun, m is the mass of the Earth, r is their mutual distance, and v is the speed of the Earth. Using the protractor,⁸ we can find that the viewing angle of the Sun is approximately $\delta = 32'$. Hence, $R_{\odot} = r \sin \frac{1}{2}\delta$ is the radius of the Sun (see Fig. 4.7). Clearly (cf. (4.10)),

$$v = \frac{2\pi r}{T}, \quad (4.17)$$

where $T = 31\,558\,149.5$ s (=365.25636 days), is the Earth's orbital period (the sidereal year). Denoting by $V = \frac{4}{3}\pi R_{\odot}^3$ the volume of the Sun and substituting M_{\odot} from (4.16), we get by (4.17) that the mean mass density is

$$\rho = \frac{M_{\odot}}{V} = \frac{v^2 r}{GV} = \frac{(2\pi r)^2 \cdot r}{T^2 G \cdot \frac{4}{3}\pi (r \sin \frac{1}{2}\delta)^3} = \frac{3\pi}{T^2 G \sin^3 \frac{1}{2}\delta} = 1409 \text{ kg/m}^3, \quad (4.18)$$

i.e., it is about 40% higher than the density of water.

⁸It is good to have a sunscreen filter. The Sun's angular diameter can also be measured using a pinhole camera, as suggested by Leonardo di ser Piero da Vinci.

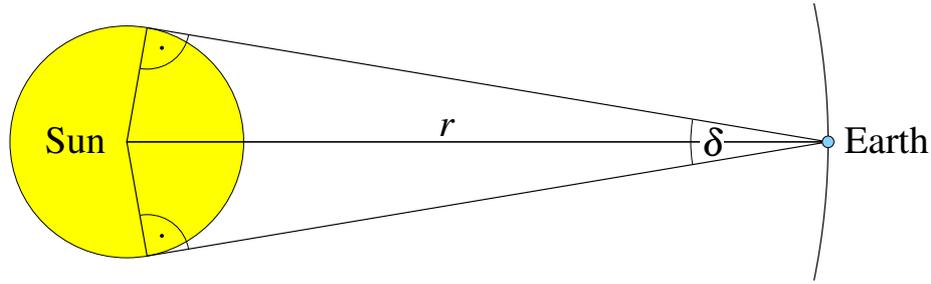


Fig. 4.7. The mean mass density of the Sun can be determined from its viewing angle δ and the Earth's orbital period (see (4.18)).

The total mass of the Sun can be now calculated as follows: From the known mean distance $r = 149.6 \cdot 10^9$ m and the measured angle $\delta/2$, we find⁹ the radius

$$R_{\odot} = 696 \cdot 10^6 \text{ m}$$

and volume $V = 1.413 \cdot 10^{27} \text{ m}^3$. Then by (4.18) we obtain

$$M_{\odot} = \rho V = 1.99 \cdot 10^{30} \text{ kg}.$$

Compare with (4.8).



4.11. Speed of Halley's comet

Halley's comet is named after Edmond Halley, who first predicted its return. Its elliptic trajectory is *retrograde*, i.e., it orbits the Sun in the opposite direction with respect to all the planets. From Kepler's third law (4.4) and the present orbital period $T = 75.7$ years¹⁰ we find that the semimajor axis is $a = 17.9$ au. The comet approaches the Sun at a distance $r_2 = a - ae = a(1 - e) = 0.585$ au (see Fig. 1.4). From this we find the eccentricity $e = 0.9673$ of its orbit. The distance in aphelion is equal to $r_1 = a + ae = 35.21$ au, i.e., the comet recedes

⁹More precise values are $M_{\odot} = 1.988\,547 \cdot 10^{30}$ kg and $R_{\odot} = 695\,508$ km.

¹⁰The orbital period is not constant but varies between 74 and 79 yr due to the influence of other planets (cf. Fig. 5.5).

beyond the orbit of Neptune. From equation (4.14) used for the Sun we get the speed of Halley's comet near aphelion and perihelion

$$v_1 = \sqrt{\frac{GM_\odot(1-e)}{a(1+e)}} = 908 \text{ m/s}, \quad v_2 = \sqrt{\frac{GM_\odot(1+e)}{a(1-e)}} = 54.6 \text{ km/s},$$

respectively. The same results can be obtained from Kepler's second law (1.6).



4.12. The validity of the gravitational law outside the Solar system

In 1780, Frederick William Herschel¹¹ found that the star ξ UMa in Ursa Major is a visual binary star. Today we know that ξ UMa is at least a quadruple star. It is about 27 light-years away from us. Its two main stars are similar in mass to the Sun. Around 1827 the French mathematician and astronomer Felix Savary found by long-term systematic observations that the two main components have elliptic orbits, which can be described by Newton's law of gravitation. He was actually the first person who confirmed the validity of this law outside the Solar system. Later this led to the belief that all other physical laws apply locally throughout the whole universe.

Coincidentally, the plane of the orbits of the two main components is almost tangent to the celestial sphere [99]. The Keplerian parameters of the actual elliptic orbits projected on the celestial sphere thus remain almost unchanged. In this case, Felix Savary was therefore able to determine the total mass of this binary system using a modified form of Kepler's third law (cf. (4.5) and (4.7)) as follows:

$$M + m \cong \frac{\alpha^3}{\gamma^3 T^2} \cong 1.84 M_\odot,$$

¹¹The famous English astronomer, discoverer of many comets, the planet Uranus, infrared (heat) radiation, and so on. He estimated also Sun's power output with an error of a few percent.

where $T = 59.8$ yr is the orbital period, $\alpha = 2.53''$ is the angular size of the semimajor axis of the relative distance¹², $\gamma = 0.135''$ is the annual parallax of the system (see Chapter 2) and is also expressed in arc seconds, and the sum of $M + m$ is then given in masses of the Sun (4.8).



4.13. Determining the distances of exoplanets from their mother stars

The mass of the star 51 Peg¹³ is $M = 1.05 M_{\odot}$. In 1995, slight periodic shifts of spectral lines caused by the Doppler effect were observed in its spectrum, since some orbiting body continually interacted with this star (see [299, p. 47]). In this way the first *exoplanet* was discovered. Its orbital period is $T = 4.231/365.256$ years. However, since its distance from us is 15.6 pc, we cannot determine the size of the axes of its orbit by angular measurements and standard trigonometric relationships. Nevertheless, combining (4.4) and (4.7), we find that

$$a \cong \sqrt[3]{1.05 T^2} \cong 0.052 \text{ au.}$$

Therefore, the exoplanet orbits its mother star closer than Mercury orbits our Sun.



4.14. Mass of the black hole at our Galaxy center

At the center of our Galaxy, which is about 26 000 ly from us, a massive black hole Sgr A* is located. The star S2 orbits about this

¹²The relative distance is the distance of one component with respect to another, which is considered to be fixed. This path is an ellipse in which the second component appears at one of the foci.

¹³One *parsec* (pc) is the distance at which a straight line segment of length 1 au would be perpendicularly observed at the angle $1''$, i.e. $1 \text{ pc} = 3.0857 \cdot 10^{16} \text{ m}$. From the known distance $d = 15.6 \text{ pc}$ and the observed magnitude $\mu = 5.49 \text{ mag}$ we may find by the Pogson relation the absolute magnitude $\bar{\mu} = \mu + 5 - 5 \log d = 4.55 \text{ mag}$, and hence it is possible to estimate the mass of the star M . For comparison, the absolute magnitude of the Sun is only slightly larger: 4.71 mag.

curious object with a period of 15.56 years (see [250]). Denote by a the semimajor axis of its elongated elliptic orbit. By means of angular measurements it was found that the observed trajectory (i.e. the projection of the actual trajectory on the celestial sphere) has a semimajor axis of length $\bar{a} = 107 \cdot 10^{12}$ m. From (4.4) and (4.8) we immediately get a lower bound of the mass of the massive black hole

$$M_{\bullet} = \frac{4\pi^2 a^3}{GT^2} \geq \frac{4\pi^2 \bar{a}^3}{GT^2} = 3 \cdot 10^{36} \text{ kg} = 1\,500\,000 M_{\odot},$$

where $T = 15.56 \text{ yr} = 4.91 \cdot 10^8 \text{ s}$.

To improve this estimate, let us now show how to calculate the unique length of a from the Pythagorean theorem and the solution of a quadratic equation (see [127]). The eccentricity e of the actual elliptic orbit may be directly established from the observed projected elliptical trajectory (see Fig. 4.8).

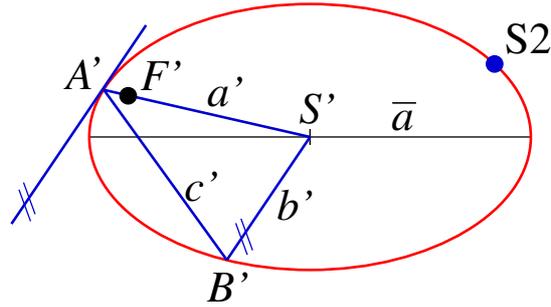


Fig. 4.8. The projection of the trajectory of S2 on the celestial sphere according to [250]. The eccentricity of its actual trajectory is equal to $e = |F'S'|/|A'S'|$, where F' denotes the observed position of the black hole.

For simplicity denote the point corresponding to the strong X-ray source Sgr A* by F' , which is the projection of the focus F of the actual trajectory. Consider the half line $S'F'$ originating at the center S' of the observed elliptic trajectory and let A' be the intersection of the half line $S'F'$ with this trajectory. The semimajor axis a containing the focus F is then projected on the line segment $A'S'$. Therefore, we obtain (see Fig. 4.9)

$$e = \frac{\varepsilon}{a} = \frac{|FS|}{|AS|} = \frac{|F'S'|}{|A'S'|},$$

where the fraction on the right-hand side can be evaluated, $|\cdot|$ stands for the length of the line segment, and $\varepsilon = |FS|$. In particular, for the situation from Fig. 4.8 we have $e = 0.875$.

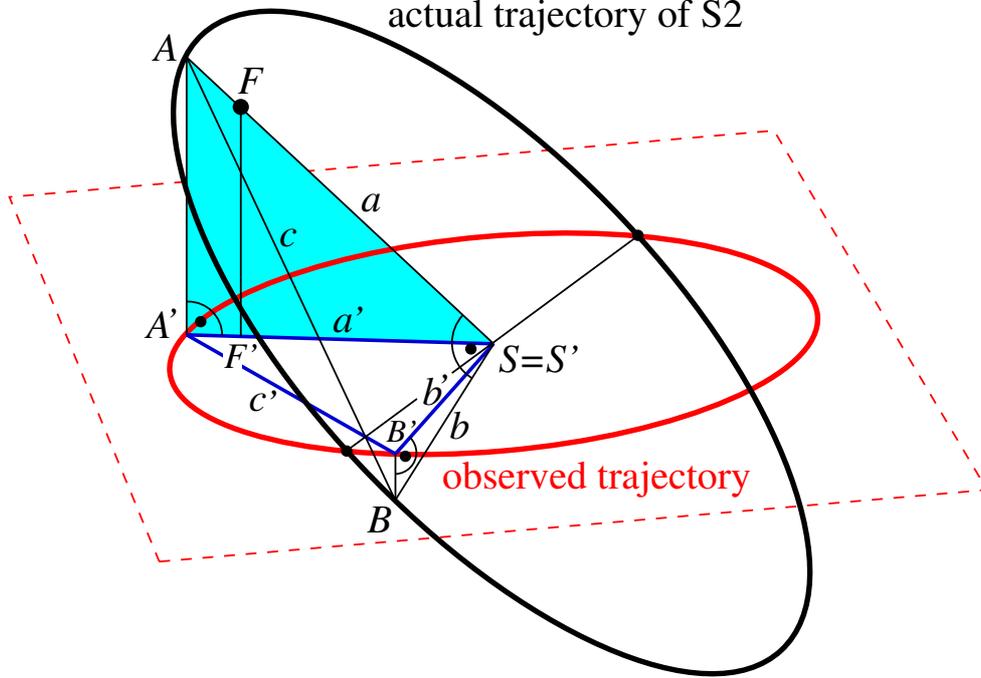


Fig. 4.9. Actual and observed trajectory of the star S2. Line segments having lengths a , b , and $c = \sqrt{a^2 + b^2}$ are projected on line segments having lengths a' , b' , and c' . The intersection of the planes ABS and $A'B'S'$ is called the *line of nodes*.

Now we establish the projection b' of the semiminor axis b of the actual trajectory. Let us construct the point B' at the observed elliptical trajectory such that the straight line $B'S'$ is parallel to the tangent¹⁴ at the point A' and the angle $A'S'B'$ is not obtuse. Applying the Pythagorean theorem to the right triangles ABS , $AA'S$, and $BB'S$ from Fig. 4.9, we find that

$$a^2 + b^2 = c^2 = c'^2 + (\sqrt{a^2 - a'^2} + \sqrt{b^2 - b'^2})^2,$$

where $a' = |A'S'|$, $b' = |B'S'|$, and $c' = |A'B'|$. From this it follows that

$$a'^2 + b'^2 - c'^2 = 2\sqrt{a^2 - a'^2}\sqrt{b^2 - b'^2}.$$

¹⁴The tangent at point A' is perpendicular to a normal which bisects the angle between the two lines connecting A' with both foci. In this case, relation (1.10) can be used.

Squaring this equation and substituting $b^2 = (1 - e^2)a^2$, we obtain a quadratic equation¹⁵ for the unknown a^2 ,

$$(1 - e^2)(a^2)^2 - [(1 - e^2)a'^2 + b'^2]a^2 + a'^2b'^2 - \frac{1}{4}(a'^2 + b'^2 - c'^2)^2 = 0.$$

The situation described in Fig. 4.8 corresponds to $a' = 4$ ld, $b' = 2.5$ ld, and $c' = 4$ ld, where ld denotes the light day. The only physically feasible solution is

$$a = 5.45 \text{ ld} = 943.6 \text{ au} = 141.166 \cdot 10^{12} \text{ m}.$$

The second positive solution is not relevant, since it is smaller than a' . Inserting a and T into Kepler's third law (4.4), we get by (4.8) that¹⁶

$$M_{\bullet} \approx 6.9 \cdot 10^{36} \text{ kg} = 3.47 \cdot 10^6 M_{\odot}.$$

We will introduce another method for determining M_{\bullet} by means of velocity measurements of the star S2. Since the sum of the kinetic and potential energy in Newtonian mechanics is preserved, we have

$$\frac{1}{2}v_1^2 - \frac{GM_{\bullet}}{r_1} = \frac{1}{2}v_2^2 - \frac{GM_{\bullet}}{r_2},$$

where v_1 is the speed at the apocenter and v_2 in the pericenter. Hence,

$$v_2^2 - v_1^2 = 2GM_{\bullet} \left(\frac{1}{r_2} - \frac{1}{r_1} \right).$$

From this and the law of conservation of momentum $r_1v_1 = r_2v_2$ (cf. (1.3)) we may finally express the unknown mass M_{\bullet} as

$$M_{\bullet} = \frac{1}{2G}(v_2^2 - v_1^2) \frac{r_1r_2}{r_1 - r_2} = \frac{1}{2G}(v_2^2 - v_1^2) \frac{r_1^2v_1/v_2}{r_1(1 - \frac{v_1}{v_2})} = \frac{1}{2G}r_1v_1(v_1 + v_2).$$

¹⁵For the mirror image with respect to the plane $A'B'S'$ from Fig. 4.9 the same quadratic equation holds. The correct sign of the inclination angle can be determined from the Doppler effect.

¹⁶The corresponding Schwarzschild radius $R_{\bullet} = 2GM_{\bullet}/c^2$ is approximately $10^{10} \text{ m} \approx 0.07 \text{ au}$.

Inserting r_1 from (1.7) into the above relation, we obtain that

$$M_{\bullet} = \frac{T}{2\pi G} (v_1 v_2)^{3/2}. \quad (4.19)$$

Relation (1.4) enables us to get the speed v_1 of the star S2 at the apocenter from its measured velocity

$$v_2 = 7 \cdot 10^6 \text{ m/s} \quad (4.20)$$

at the pericenter and the eccentricity $e = 0.875$. Substituting now these values and the orbital period $T = 4.91 \cdot 10^8$ s of S2 into (4.19), we get by (4.8) the approximate mass of the central black hole

$$M_{\bullet} \approx 6.91 \cdot 10^{36} \text{ kg} = 3.47 \cdot 10^6 M_{\odot}. \quad (4.21)$$

The resulting mass depends, of course, on the accuracy of input data T , v_2 , and e . These values will be known more precisely after several revolutions.

From (1.7) we may deduce that the distance of the pericenter of the star S2 from Sgr A* is more than three times the distance of Neptune from the Sun. In the experimentally established values v_1 during the period 1994/95, no radial components of velocities were measured, and therefore the ratio v_1/v_2 given in [250, p.695] essentially differs from (1.4). The true value of M_{\bullet} is continually being refined also by the examination of trajectories of other stars S1, S3, S4, ... From this it seems that $M_{\bullet} \approx 4 \cdot 10^6 M_{\odot}$, cf. (4.21).



4.15. Physical characteristics of the planets

Although Newton's universal law of gravitation describes the behavior of planets on short time scales quite accurately, we have to remember that it is only a mathematical model. It describes physical reality only approximately.

According to the Committee on Data for Science and Technology, the gravitational constant in SI units is equal to

$$G = 6.67428 \pm 0.00067 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}. \quad (4.22)$$

The uncertainty in the determination of the gravitational constant is considerable (already in the fourth significant digit). It is the fundamental constant of nature whose measurement is the most imprecise at present (see Section 14.1). Thus, it has a negative impact on most calculations and their long-term predictions in celestial mechanics. For example, the semimajor axis of an elliptic orbit by a body is equal by (4.4) to

$$a = \left(\frac{T^2 GM}{4\pi^2} \right)^{1/3}$$

and the estimate of its actual size thus depends on the exact determination of the constant G , respectively the product GM which is usually known with better accuracy involving more significant digits. The uncertainty in establishing G has an influence on the resulting masses of planets in Table 4.1. On the other hand, the orbital period can usually be measured quite accurately.

Using relations (4.1), (4.9), and (4.15), we can obtain Table 4.1. Based on our knowledge of the distance of i th planet and its angular diameter, we may determine its radius r_i . The gravitational acceleration g_i on a planetary surface can then be easily calculated from Newton's second law and the law of gravitation (4.1), i.e.

$$mg_i = G \frac{m_i m}{r_i^2}$$

after canceling out the testing particle mass $m > 0$. Finally, let us concentrate on the last column in Table 4.1. If Uranus had a solid surface, then we would weigh less than on the Earth, even though Uranus is 15 times more massive than Earth, since Uranus' density is low. Note also that the gravitational accelerations on Mars and Mercury are almost the same.

Table 4.1. The symbol i denotes the i th planet, m_i its mass in kg, a_i is the semimajor axis of its orbit in meters, T_i is the orbital period in years, v_i is the mean orbital velocity of the planet around the Sun in km/s, v_{II} is the escape velocity in km/s from the surface (the thickness of atmosphere is neglected for simplicity), r_i is the mean radius of the planet in km, and g_i is the gravitational acceleration in m/s^2 on the planet surface.

i	planet	$m_i/10^{24}$	$a_i/10^9$	T_i	v_i	v_{II}	r_i	g_i
1	Mercury	0.33022	57.9	0.241	47.9	4.25	2440	3.697
2	Venus	4.8676	108.2	0.615	35.0	10.36	6052	8.867
3	Earth	5.97219	149.6	1	29.8	11.18	6371	9.807
4	Mars	0.64185	229.97	1.881	24.0	5.03	3390	3.726
5	Jupiter	1898.6	778.4	11.861	13.1	60.19	69911	25.91
6	Saturn	568.46	1427.0	29.457	9.6	36.09	58232	11.182
7	Uranus	86.81	2869.6	87.011	6.8	21.37	25362	9.004
8	Neptune	102.43	4496.6	164.79	5.4	23.56	24624	11.268



5. The N -body problem

*If I have seen further, it is by
standing on the shoulders of giants.*

ISAAC NEWTON

5.1. Introduction

In this chapter we shall investigate the N -body problem. The main task will be to determine the equation for the trajectories of N mass particles that interact only by Newton's law of gravitation. Let us stress right from the beginning that in the classical Newtonian theory of gravity the size of forces, which act among the bodies themselves, depends only on the immediate positions and masses of the bodies. In other words, an infinite speed of propagation of gravitational interaction is assumed, which in the real world is certainly finite. This artificial assumption causes the so-called *modeling error*. Nonetheless, at short time intervals Newton's theory is quite precise in the Solar system, and is therefore perfectly suited to calculate trajectories of interplanetary probes or near-Earth asteroids. In these cases, the modeling error is relatively small indeed. Newton's theory, however, is not suitable for long-term simulations on time scales of millions of years, when the modeling error accumulates in such extent that it devalues the resulting numerical solution. More details about these problems will be discussed in Section 5.4 and in the second half of this book.



5.2. The two-body problem

If only two bodies interact, we can analytically determine their orbits by a closed formula. A general solution in the form of conic section curves is given e.g. in [20]. In the previous Chapters 3 and 4, we assumed that one of the bodies has a negligible mass with respect to the other one. In this special case, we may define like in (4.12) and (4.15) the *circular velocity* and *escape velocity*, respectively, by the relations

$$v_1 = \sqrt{\frac{Gm}{r}} \quad \text{and} \quad v_2 = \sqrt{\frac{2Gm}{r}},$$

where G is the gravitational constant, $m > 0$ is the mass of the central body, and r is the radius of a circular trajectory. Depending on initial conditions, namely, the size of tangential velocity at the pericenter or apocenter P , we obtain various shapes of trajectories as shown in Fig. 5.1.

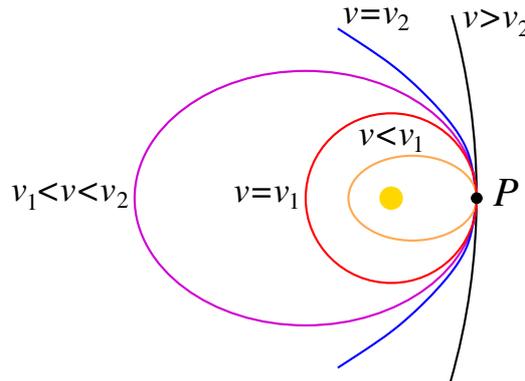


Fig. 5.1. All trajectories have a common focus and pass through the point P , in which the following conditions on the speed are assumed: for $v < v_1$ the path of a body is elliptic, for $v = v_1$ the path is circular, for $v_2 > v > v_1$ the path is again elliptic, for $v = v_2$ the path is parabolic, and for $v > v_2$ hyperbolic.

Further, we shall assume that the masses m_1 and m_2 of the two considered bodies are nonzero. Without loss of generality we may assume that the center of gravity of this system is fixed at the origin of the coordinate system. Then for every time instant we have¹

$$m_1 r_1 = m_2 r_2,$$

¹This relation is similar to the balancing of two objects on a two-armed lever.

where r_1 and r_2 are the respective distances of the bodies from their common center of gravity. From this and Table 4.1 we can find, for example, that for $m_1 = M_\odot$ (see (4.8)) the center of gravity of the system Sun–Jupiter lies outside the Sun, since the distance

$$r_1 = \frac{m_2 r_2}{m_1} = \frac{1898.6 \cdot 10^{24} \cdot 777.7 \cdot 10^9}{1.99 \cdot 10^{30}} = 742 \cdot 10^6 \text{ (m)} \quad (5.1)$$

is greater than the Sun’s radius $R_\odot = 696 \cdot 10^6$ m. In this case, both orbits are basically circular,² neglecting the influence of other planets (for the opposite case see Fig. 13.1).

Fig. 5.2 shows the general case of two trajectories of unequal mass points orbiting C which is located at the origin of the coordinate system. Their elliptic trajectories have the same eccentricity e . The fact that the distances r_i , $i = 1, 2$, depend on time t will be denoted as follows: $r_i = r_i(t)$. The orbital period T satisfies the generalized form of Kepler’s third law (4.5)

$$\frac{a^3}{T^2} = \frac{G(m_1 + m_2)}{4\pi^2},$$

where $a = (r_1(t^*) + r_2(t^*))(1 + e)$ is, in fact, the sum of the semimajor axes of both ellipses and t^* corresponds to the instant when the two bodies have the largest possible distance from each other (i.e., when they are located at their apocenters [156, p. 118]).

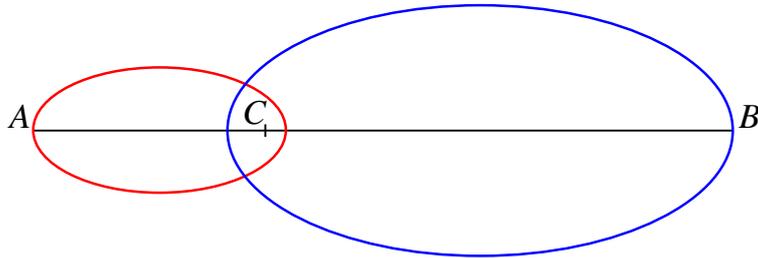


Fig. 5.2. Two mass points A and B orbiting about a common center of gravity C have elliptic trajectories. The point C is also one of the foci of each ellipse. The ratio of distances AC/BC remains constant in time.

²The eccentricity of Jupiter’s orbit is only 0.048.

Before we derive a differential equation describing the movement of three gravitationally interacting bodies, we derive a simple differential equation for two bodies in the one-dimensional case from Fig. 5.3.

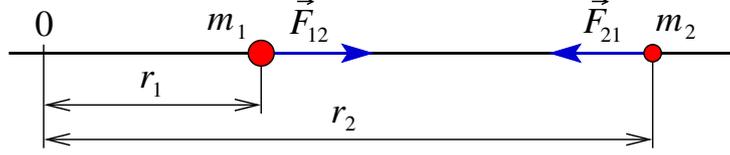


Fig. 5.3. One-dimensional two-body problem

Let $r_1 < r_2$ be the coordinates of two point particles with positive masses $m_1 > 0$ and $m_2 > 0$ whose center of gravity need not be at the origin. According to Newton's law of gravitation, their interaction produces oppositely directed forces \vec{F}_{12} and \vec{F}_{21} ,

$$|\vec{F}_{12}| = |\vec{F}_{21}| = G \frac{m_1 m_2}{(r_2 - r_1)^2}, \quad (5.2)$$

where $|\cdot|$ stands for the length of a vector. Assume that the two particles are not influenced by any other forces except for the gravitational forces \vec{F}_{12} and \vec{F}_{21} . It is known that the velocity of the i th body is $\dot{r}_i = \dot{r}_i(t)$, where the dot denotes the time derivative. The acceleration in time t is then given by $\ddot{r}_i(t)$. According to Newton's second law of motion, the force acting between them is linearly proportional to the acceleration and we have

$$|\vec{F}_{12}| = m_1 \ddot{r}_1(t), \quad |\vec{F}_{21}| = -m_2 \ddot{r}_2(t). \quad (5.3)$$

Further on, for simplicity we shall usually not indicate the dependence of r_i on time t . From (5.2) and (5.3) we obtain

$$\ddot{r}_1 = G \frac{m_2}{(r_2 - r_1)^2}, \quad -\ddot{r}_2 = G \frac{m_1}{(r_2 - r_1)^2}.$$

Writing $r = r_2 - r_1$ and summing the above two equations, we get the following relation for the unknown function $r = r(t)$

$$\ddot{r} = -G \frac{m_1 + m_2}{r^2}. \quad (5.4)$$

Differentiating twice with respect to time t , is easy to see that this second order differential equation is satisfied by the function

$$r(t) = c(\tau - t)^{2/3} \quad \text{for } t < \tau,$$

where $c = (\frac{9}{2}G(m_1 + m_2))^{1/3}$ is a constant and τ is the instant of collision. We see that the speed $|\dot{r}(t)| = \frac{2}{3}c(\tau - t)^{-1/3}$ converges to ∞ , when t tends to τ from the left. This illustrates that the Newtonian theory of gravitation allows superluminal velocities.

However, there also exist other solutions of equation (5.4). To get a unique solution locally, it is enough to prescribe the values r and \dot{r} at some time instant (e.g. for $t = 0$).



5.3. The three-body problem

We may succeed only in a few special cases to derive a formula describing trajectories of three bodies (see e.g. [168]). Such solutions were found e.g. by J. d'Alembert, L. Euler, W.R. Hamilton, W.W. Heinrich, C.G.J. Jacobi, J. Kepler, J.L. Lagrange, P. Laplace, V. Nechvíle, K. Petr (viz [168]). The three-body problem is governed by a system of 18 nonlinear differential equations of the first order for the three components of positions and the three components of velocities of each of the three bodies. Isaac Newton once declared about this problem:

An exact solution exceeds, if I am not mistaken, the force of any human mind.

Therefore, the so-called restricted three-body problem is often investigated (see [20]). For example, if one of the bodies has a negligible mass, then the problem can be greatly simplified (e.g. a satellite in the system Sun–Earth–satellite). We can derive only one second order differential equation for the angular speed of the two heavier bodies that orbit about the common center of gravity along elliptic trajectories. From this we may calculate the trajectories of all three bodies.

When two of the three bodies are relatively close to each other (e.g. the Earth and Moon in the Sun–Earth–Moon system), the problem is often reduced to two much simpler two-body problems, e.g. the Earth–Moon system is replaced by its center of gravity. However, this is always only an approximation (cf. Fig. 4.3). Special cases of the three-body problem are investigated in e.g. [20], [128], [168].

Henri Poincaré [216] proved however that for the three-body problem there is, in general, no analytical solution in the form of a closed formula. Therefore, its approximate solution is mostly calculated by numerical methods.

Now let us derive general equations for the trajectories of three mass points that interact only gravitationally. For simplicity we shall restrict ourselves to the planar case (the three-dimensional case can be treated analogously). The force will be a two component vector. Consider three point particles with masses m_1 , m_2 , and m_3 . Their position in the coordinate system (x, y) will be characterized by the vector $\vec{r}_i = \vec{r}_i(t)$, $i \in \{1, 2, 3\}$, which is called the *radius-vector* (see Fig. 5.4). Set

$$\vec{r}_{ij} = \vec{r}_j - \vec{r}_i, \quad i, j \in \{1, 2, 3\}, \quad i \neq j, \quad (5.5)$$

and let $r_{ij} = |\vec{r}_{ij}|$ be the length of vector \vec{r}_{ij} . The size of the force acting between the first and second body equals (similarly to (5.2))

$$|\vec{F}_{12}| = G \frac{m_1 m_2}{r_{12}^2}.$$

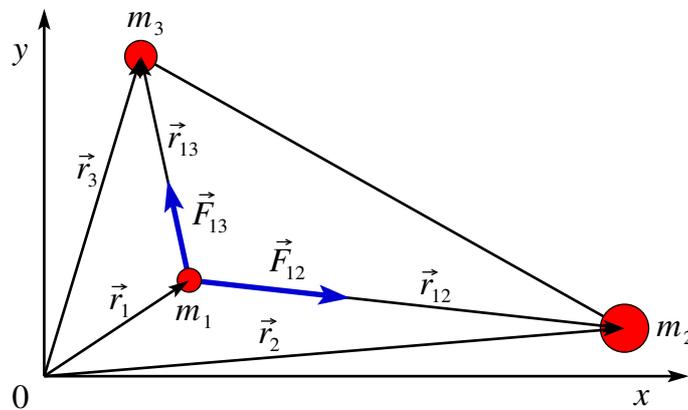


Fig. 5.4. Two-dimensional three-body problem

The force \vec{F}_{12} acts in the direction of the unit vector \vec{r}_{12}/r_{12} , and thus

$$\vec{F}_{12} = G \frac{m_1 m_2 \vec{r}_{12}}{r_{12}^3}.$$

Therefore, the total force acting on the first body is

$$\vec{F}_{12} + \vec{F}_{13} = G m_1 \left(\frac{m_2 \vec{r}_{12}}{r_{12}^3} + \frac{m_3 \vec{r}_{13}}{r_{13}^3} \right)$$

and equals $m_1 \ddot{\vec{r}}_1$ by Newton's second law. For the acceleration of the first body we obtain the equation

$$\ddot{\vec{r}}_1 = G \left(\frac{m_2 \vec{r}_{12}}{r_{12}^3} + \frac{m_3 \vec{r}_{13}}{r_{13}^3} \right).$$

Similarly, for all three bodies we get a system of three vector differential equations for the unknowns \vec{r}_1 , \vec{r}_2 , and \vec{r}_3 ,

$$\ddot{\vec{r}}_i = G \left(\frac{m_j \vec{r}_{ij}}{r_{ij}^3} + \frac{m_k \vec{r}_{ik}}{r_{ik}^3} \right) \quad (5.6)$$

for all i, j, k such that $\{i, j, k\} = \{1, 2, 3\}$ and $j < k$. These equations are not independent, as follows from relation (5.5). Since the \vec{r}_i are vectors of two components, in (5.6) there are, in fact, six unknown functions of time. However, the system (5.6) does not have a unique solution. Therefore, at time $t = 0$ we impose *initial conditions* on positions \vec{p}_i on all three bodies and their velocities \vec{v}_i

$$\vec{r}_i(0) = \vec{p}_i, \quad \dot{\vec{r}}_i(0) = \vec{v}_i \quad \text{for } i = 1, 2, 3, \quad (5.7)$$

which, of course, have an essential influence on the resulting motion of the bodies.

From the theory of differential equations, we know that the solution of problem (5.6)–(5.7) exists uniquely on an arbitrarily long time interval, where all three distances r_{ij} for $i < j$ are nonzero, i.e., the bodies do not collide. Note that the probability that two or more mass points collide within a fixed time interval is equal to zero, even though

this case may theoretically occur. Although there are numerous collisions in the real world, we should not forget that in the classical multiple body problem only mass points are considered, while actual bodies always have positive sizes.

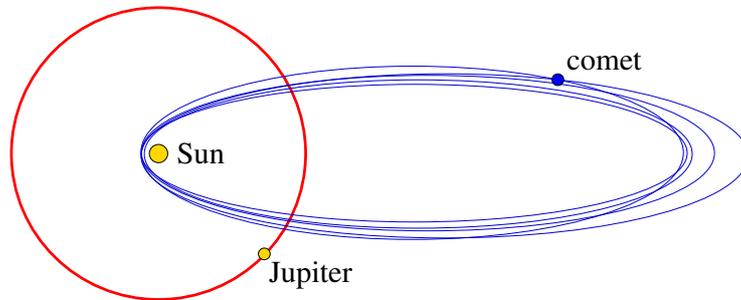


Fig. 5.5. The influence of Jupiter on a comet's trajectory

There exists a vast literature on numerical solutions of a system of ordinary differential equations, see e.g. [225]. Numerical solution of problem (5.6)–(5.7) enables us to investigate many real-life practical problems. For instance, we can simulate the influence of Jupiter on a comet's trajectory³ (cf. Fig. 5.5).

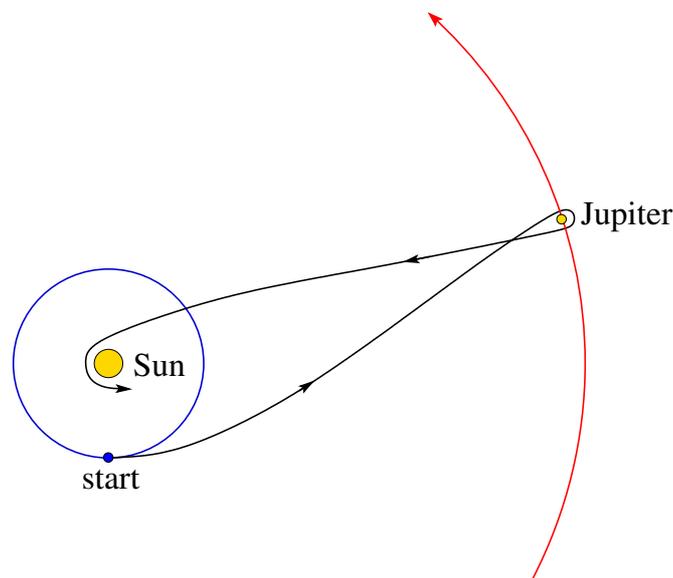


Fig. 5.6. Economic way of a probe sent from the Earth to the Sun.

³The great red spot on Jupiter could be a consequence of impact of such a comet or asteroid, since a large amount of heat was produced. The spot is perpetually at the same location and slowly decreases like the spots caused by the comet Shoemaker-Levy 9.

Let us introduce another typical example which cannot be solved analytically for which we thus need to look for its numerical solution. The Earth orbits around the Sun at a speed of 29.8 km/s (see (4.11)). To reach a close neighborhood of the Sun, we may send a probe in the opposite direction also at speed 29.8 km/s, and then the probe would freely fall into the Sun. Because the kinetic energy is directly proportional to the square of the speed, it is much more energetically convenient to first send the probe to Jupiter in an appropriate direction. This requires only to achieve the escape velocity from the Earth of 11.2 km/s (see (4.15)). The strong gravitational field of Jupiter will then direct the probe to the Sun,⁴ see Fig. 5.6. This maneuver is sometimes called *gravitational ping-pong*.⁵



5.4. The N -body problem

For more bodies one can derive in a quite similar manner as in the previous section a system of differential equations describing their trajectories. Suppose there are N point particles with masses $m_i \geq 0$, $i = 1, 2, \dots, N$, which interact only gravitationally. For simplicity, let us put $r_i = \vec{r}_i$, i.e., the arrow will be removed from now on. Then the vector trajectories r_i of the considered bodies are described by the system of nonlinear differential equations

$$\ddot{r}_i = G \sum_{j \neq i}^N \frac{m_j (r_j - r_i)}{|r_j - r_i|^3} \quad (5.8)$$

for $i = 1, \dots, N$ with given initial conditions on the positions $r_i(0)$ and velocities $\dot{r}_i(0)$ of all N bodies. Here $|\cdot|$ stands again for the length of a vector.

⁴Luna 3, which photographed the opposite side of the Moon as early as 1959, had a similar shape for its trajectory if Jupiter is replaced by the Moon. A similar return path was also used by Apollo 13 after an accident in 1970.

⁵The idea of accelerating probes by planetary gravitational fields was developed around 1960 during numerical simulations of the multiple body problem. It is a conception of the American mathematician Michael Andrew Minovitch.

This problem has many real-world technical applications. As an example let us recall the complex gravitational ping-pong of the journey of Voyager 2 to Jupiter, Saturn, Uranus, and Neptune, when they all were on the same side of the Sun. This orbit was designed by Gary Arnold Flandro in 1965. The actual trajectory was then obtained by the numerical solution of a system of differential equations similar to (5.8). Each planet greatly accelerated⁵ the probe by means of its gravitational field and also changed its direction.⁶ The probe actually fell into the gravitational potential well of each planet. Sometimes we describe this manipulation of the orbit as *a gravitational slingshot*. The speed of the probe after a close flyby about a planet decreases, because every planetary orbit is curved. Higher kinetic energy is gained by the probe at the expense of the total energy of the planet.



Fig. 5.7. Collision of two galaxies called Antennae. Behind each galaxy there is a clear “tidal tail” showing their original nearly Keplerian elliptic trajectories. The two galaxies will merge over time as a result of tidal forces. This process cannot be treated as a classical two-body problem. (Photo NASA)

The system of differential equations (5.8) is sometimes also used to simulate the evolution of galaxies or even galaxy clusters (see e.g. (7.2) and Section 7.3). Note that galaxies have huge speeds reaching

⁶Any change of the direction is very energy consuming. It is actually also an acceleration. A probe flying at speed v needs for a directional change of 60° the same energy as to achieve the speed v . To see this, we may rotate the probe about 120° and then give it the speed of size v .

up to a few percent of the speed of light. However, system (5.8) does not describe relativistic phenomena nor effects of tides (see Fig. 5.7), rotation of galaxies, the finite speed of propagation of the gravitational interaction, and so on. It is again only an approximate model.

System (5.8) does not satisfy the well-known Carathéodory sufficient conditions for the existence of the solution, since the right-hand side of (5.8) is not continuous. In a spite of that, the solution exists locally, as long as the point-bodies do not collide. We may also obtain a large number of nonrealistic solutions of (5.8). For instance, in [242] five bodies are considered such that all of them reach infinity in a finite time for appropriate initial conditions. In this case the modeling error is infinite. The Newtonian theory of gravitation thus may describe completely absurd situations. Paper [188] presents also a solution of (5.8) when three equally massive particles orbit along a trajectory that looks like ∞ which has not yet been observed in the universe.



5.5. Total approximation error

As already mentioned, H. Poincaré knew that the solution of system (5.8) can be analytically written only in a few special cases, and he proved that it generally cannot be explicitly expressed analytically by a closed formula. Therefore, we have to look for an approximate solution. A continuous mathematical model is approximated by a discrete finite dimensional problem whose solution is based on numerical algorithms (e.g. Runge–Kutta methods, symplectic methods, multistep methods) that can be implemented on a computer.

In numerical simulations of the Solar system for billions of years⁷ forward or backward, some programmers do not care about the size of errors that were committed. Let us therefore look at the risks, and whether one can believe what they have actually calculated without them having detailed knowledge of the theory of numerical solutions of differential equations.

⁷Such simulations reminiscent calculation of weather or earthquake prediction for months ahead.

In Fig. 5.8 there is a general computational scheme of numerical solutions of problems of mathematical physics on a computer [135], where we always commit three basic errors: modeling error $e_0 = e_0(t)$, discretization error $e_1 = e_1(t)$, and rounding errors $e_2 = e_2(t)$. We try to make them simultaneously minimal. If just one of them is large, then the total error $e = e_0 + e_1 + e_2$ is also large. Large errors are generally not canceled. In addition, it makes no sense to try to keep the discretization error e_1 very small, if the other errors will be large. The error e_1 usually consists of several other errors: error of numerical integration, various interpolation, approximation, and extrapolation errors, and so on. To test the size of actual numerical errors e_1 and e_2 of a computer program it is useful to compare the exact analytical solution of a special N -body problem with its numerical solution.

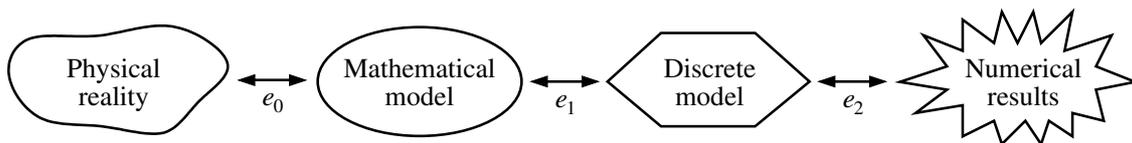


Fig. 5.8. Modeling error $e_0(t)$ is the difference between physical reality and the solution of its mathematical model (description). The discretization error $e_1(t)$ is the difference between solutions of the mathematical model and the discrete model. Finally, in $e_2(t)$ rounding errors (or iteration errors) are included.

If at least two bodies are close to each other at some time t (i.e., $r_i(t) \approx r_j(t)$ for $i \neq j$), then in the denominator on the right-hand side of (5.8) almost equally large vectors are subtracted. In a numerical solution of such a problem not only a large discretization error may arise, but also considerably large rounding errors may appear. Therefore, during any calculation one should monitor whether two numbers of almost the same size are subtracted. The examples given in [33] should be warning enough. A catastrophic loss of accuracy due to rounding errors e_2 is given e.g. in [148].

For $N \geq 3$, system (5.8) is *not stable with respect to continual perturbations*, i.e., small perturbations in the right-hand side can cause large changes in the solution over a long time interval. System (5.8)

is also *not stable with respect to initial conditions*, i.e. a slight change in the initial conditions may cause a large change in the solution. For this reason, a correct numerical calculation is extremely difficult. For example, the asteroid no. 99942 Apophis about 270 m in diameter will on Friday, April 13, 2029, be at a distance of only 30 000 km from the Earth's surface. But so far, the minimum distance of Apophis during its next approach to Earth in the year 2036 cannot be reliably estimated. The total error of numerical methods for solving a system of (5.8) usually grows exponentially with time. Therefore, it is necessary to correct the trajectory a few times during the flights of probes through the Solar system.

To try to make sure that the discretization error is small the following trick is often used. If an integration step h gives almost the same numerical results as $h/2$, the integration of the system (5.8) is usually stopped, see [225, p.487]. However, this heuristic criterion need not to work properly. Now we are going to present another way to check whether our numerical results are good. It is based on forward and then backward integration of (5.8) as sketched in Fig. 5.9.

Let $r = (r_1, \dots, r_N)$ denote a vector whose components are three-dimensional vectors and let $f = f(r)$ stand for the right-hand hand side of (5.8). The following theorem enables us to use instead of the backward integration with a negative integration step forward integration on the interval $[0, T]$, where the bodies do not collide. In this case we start from the calculated value $r(T)$ and we change only the sign of velocities $(\dot{r}_1(T), \dots, \dot{r}_N(T))$ in (5.12).

Theorem 5.1. *Let $r = r(t)$ be the unique solution of the system*

$$\ddot{r} = f(r) \tag{5.9}$$

on the interval $[0, T]$ with given initial conditions

$$r(0) = r_0 \quad \text{and} \quad \dot{r}(0) = v_0. \tag{5.10}$$

Then the function $s = s(t)$ defined by

$$s(t) = r(T - t) \tag{5.11}$$

solves the same system (5.9) with initial conditions

$$s(0) = r(T) \quad \text{and} \quad \dot{s}(0) = -\dot{r}(T), \quad (5.12)$$

and we have

$$s(T) = r_0 \quad \text{and} \quad \dot{s}(T) = -v_0. \quad (5.13)$$

P r o o f . According to (5.11) and (5.9), we obtain

$$\ddot{s}(t) = (-\dot{r}(T-t))' = \ddot{r}(T-t) = f(r(T-t)) = f(s(t)).$$

We see that s satisfies the same system (5.9) as r . For the final conditions by (5.11) and (5.10) we get relations (5.13),

$$s(T) = r(0) = r_0 \quad \text{and} \quad \dot{s}(T) = -\dot{r}(T-T) = -\dot{r}(0) = -v_0. \quad \square$$

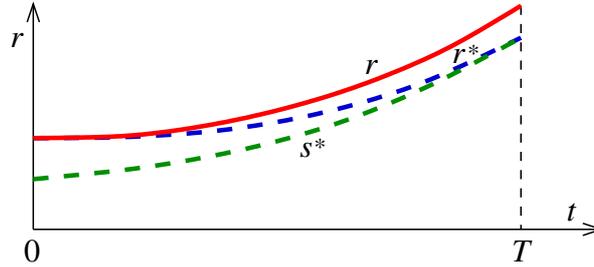


Fig. 5.9. Application of Theorem 5.1 to numerical solution of the N body problem. The symbol r stands for the true solution, r^* is a numerical solution, and s^* is a control backward solution.

The above theorem can be applied to long-term intervals as follows. Denote by r^* and s^* a numerical solution of system (5.9) with initial conditions (5.10) and

$$s(0) = r^*(T) \quad \text{and} \quad \dot{s}(0) = -\dot{r}^*(T),$$

respectively. If $\delta > 0$ is a given tolerance and

$$|s^*(T) - r_0| + C|\dot{s}^*(T) + v_0| \gg \delta,$$

then it is not likely that $|r(T) - r^*(T)| + C|\dot{r}(T) - \dot{r}^*(T)| < \delta$, where $C > 0$ is an appropriate constant and r is the exact solution, i.e., the numerical error of the original problem (5.9)–(5.10) would also be large, as is schematically depicted in Fig. 5.9.

The proposed mathematical model is often identified with reality which is a fundamental drawback. For example, the truth speed of gravitational interaction is finite, and therefore the system of equations (5.8) cannot describe reality exactly. In numerical simulations the modeling error e_0 continually accumulates while rounding errors are statistically partially canceled. The modeling error includes e.g. the error in establishing various physical and geometrical data, initial and final conditions on positions and velocities, and so on. Moreover, no real physical process is strictly deterministic, while the solution of system (5.8) is deterministic.

Remark 5.1. As already mentioned, Newton’s theory assumes an infinite speed of propagation of gravitational interaction. Therefore, we should consider planets (or other objects) where they are right now, and not where we observe them. For example, by (4.10) the mean orbital speed of Jupiter is 11.861 km/s. Its light reaches us in about 45 minutes and during this time the position of Jupiter has shifted by more than 30 000 km. Similarly we find that the position of Mercury at its closest approach to the Earth (e.g. while eclipsing of the solar disk) has shifted more than three times of its diameters during the time in which the information about its position comes to the Earth (which can be deduced from Table 4.1). The deviation is over half arc minute! Similarly, the position of Neptune has shifted by an amount approximately equal to its diameter. These tiny bugs in modeling must be carefully removed [34], [232], otherwise they continually accumulate which may cause large errors during long time integrations. Therefore, if we perform numerical simulations for millions or even billions of years forward (or backward), then it makes no sense to reach definitive conclusions as is sometimes done. Numerical results depend not only on the numerical methods used, but also on rounding errors,

the specific computer used, the programming language, the way of programming, and so on.

A large modeling error can also occur when the N -body problem is applied to simulate the evolution of galaxies in curved spacetime. In doing so, it often happens that the main emphasis is on the aesthetic impression and not on estimation of modeling or discretization errors. Moreover, astronomical observations show that gravity on galactic scales manifests itself differently than in the Solar system. For instance, consider a miniaturized model of a spiral galaxy with a diameter of 10 au, where stars are replaced by asteroids with masses linearly proportional to those of stars. Such an object would not be spinning like spiral galaxies for no initial conditions. The main reason is that the speed of gravitational interaction is finite and thus it is not a scale invariant. Therefore, we should not overestimate the utility of numerical simulations and interpret them in a way that fits in with their preconceived conclusions.

Finally, we note that it is not clear how to define the actual center of gravity of two or more bodies of unequal weights, since the speed of gravitational interaction has not yet been measured.



6. Eclipses and the aberration of light

Imagination is more important than knowledge.

ALBERT EINSTEIN

6.1. The importance of eclipses in exploring the universe

By observing an eclipse of the Sun, Moon, and other bodies, we can gather important data. When observing Earth's shadow during lunar eclipses, ancient scholars already discovered that our Earth is round and that it freely flows in space. From the shape of Earth's shadow they estimated that the Earth is three times larger than the Moon, see Fig. 2.3.

Using the records of solar and lunar eclipses of ancient Babylonian astronomers, we can now re-calculate the irregularities in the Earth's rotation and the rate of its deceleration (see Section 2.7). Various records of eclipses also helped to uncover several calendars of ancient civilizations, which now allows us to precisely determine the dates of some significant events.

The eclipse of the solar disk by Venus helped to further refine the Earth–Sun distance (see Section 2.6, Fig. 6.1, and [258]). In 1761, Mikhail V. Lomonosov discovered a luminating atmosphere of Venus during such a transit.

In 1911, during his stay in Prague, Albert Einstein deduced that light in a solar neighborhood is bent by gravity [69]. This effect of the general theory of relativity was first photographed during the total solar eclipse of 1919. The starlight deviated from its original straight trajectory by less than two arc seconds, compared to nocturnal photographs of the same part of the sky.

Total solar eclipses offer a unique opportunity for observations of solar protuberances, the solar chromosphere, and its corona. This is important for understanding the mechanism of processes taking place inside our nearest star — the Sun. In 1983, during the total solar eclipse, a dust ring orbiting the Sun was discovered in its vicinity almost in the ecliptic plane.

The Moon on its trajectory across the sky also occasionally overshadows some invisible source of radio waves. After some time the waves appear again. The location of this radio source can be easily determined as one of the intersection points of two circles representing the circumference of the Moon at the beginning and at the end of alignment. This trick was used particularly at a time when direct detections of positions of radio sources were not accurate.

Slight eclipses of distant stars by exoplanets and their moons allow to use Kepler's laws to determine the masses of these stars together with other parameters of this system (see [299]). We can also study the atmospheres of extrasolar planets, expose biogenic elements (H, C, N, O, P) and molecules (e.g. CH_4) in their spectrum, and search for traces of extraterrestrial life.



Fig. 6.1. The third of four contacts of Venus when leaving the solar disk on June 6, 2012 (photo Jozef Leško)



6.2. A brief history of eclipses

From ancient times the story of the unfortunate fate of two Chinese imperial astronomers Si and Che is recorded. In 2137 BC, they neglected their obligations during a total solar eclipse. They drank together instead of shooting arrows at the monster devouring the Sun. The emperor's order was clear: Execution! However, this event is not supported by any documents, because at that time, Chinese was not a written language (Chinese characters originated after 2000 BC).

The oldest text with a picture of an eclipse was found in Anyang in the Henan Province in China, and appeared on a bone. According to [93, p. 82], its origin dates back to around 1300 BC. The classical Chinese book *Chūnqiū* (The Spring and Autumn Annals) contains a list of 36 solar eclipses between 720–495 BC.

Further description of the certified solar eclipse from around 750 BC is recorded by cuneiform script on a clay plate from Mesopotamia stored in the British Museum in London (see [266]). In Europe, the first prediction of a solar eclipse was made by the Greek philosopher Thales of Miletus (ca. 620–555 BC) concerning the eclipse of 585 BC. He probably knew the periodicity of eclipses (see [270]). Let us also mention another article [247] which contains literal translations of eclipse records on clay plates by Babylonian astronomers from the period 829–1019 AD.



6.3. The origin and periodicity of eclipses

If the Moon orbited the Earth in the ecliptic plane, we would observe a lunar eclipse at every full moon and a solar eclipse at every new moon. However, the inclination of the Moon's orbit to the ecliptic is 5.145° . The intersection points are called the *nodes*. Therefore, a solar eclipse may occur somewhere on Earth if and only if the distance between the Sun and one of the nodes is less than 10° in the ecliptic plane and simultaneously we have a new moon. Lunar eclipses appear under similar conditions.

The Earth's elliptic orbit around the Sun has a small eccentricity $e = 0.0167$, and thus the angular diameter of the Sun changes from 0.52° in summer to 0.54° in winter. Also, the angular diameter of the Moon due to its elliptic orbit with eccentricity $e = 0.0554$ varies from 0.49° up to 0.56° . These angles happen to be almost the same as for the Sun, see Fig. 6.2. If the centers of the Sun, Moon, and Earth in this order are approximately on one line (cf. Section 6.5), we observe a solar eclipse at some region on Earth. Additionally, if the angular diameter of the Moon is less than the angular diameter of the Sun, an annular eclipse may occur. Otherwise the eclipse is total.

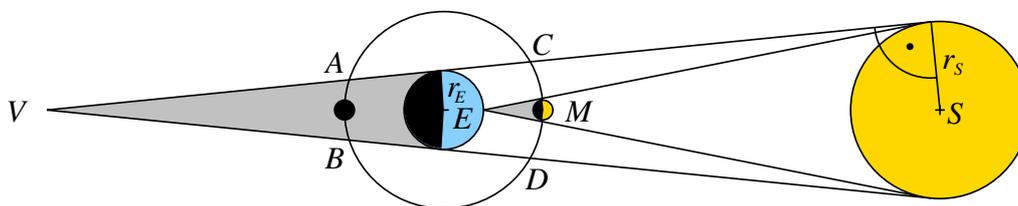


Fig. 6.2. Shadows of the Earth and Moon in the ecliptic plane

If the Earth lies near the segment Sun–Moon, then a lunar eclipse may appear. It can be seen from across the unlit hemisphere, while a solar eclipse can be seen only from a narrow strip on the Earth's surface.

Already the ancient Babylonians noticed that eclipses show some periodicity. Knowledge of the occurrences of past eclipses then allowed one to predict both lunar and solar eclipses (see [265, p.133]). To clarify this, let us first introduce two definitions.

Draconic month = 27.21222 days is the time since the passage of the Moon through a node to the same node.

Synodic month = 29.53059 days is the time from one phase of the Moon to the same phase.¹

Now we can easily find that

$$242 \text{ draconic months} \approx 223 \text{ synodic months (= cycle period)}. \quad (6.1)$$

¹A synodic month corresponds to the periodicity in the arrangement of the Earth, Sun, and Moon in their projection to the ecliptic. It is not periodical with respect to the stars.

Relation (6.1) is obtained similarly to that of the lowest common multiple, where rational numbers are considered instead of integers. From the 6th century BC this period of eclipses was called *saros* by the Chaldeans. Its length is 18 years 9–11 days, 7 hours and 43 minutes². The number of days 9, 10, or 11 given above corresponds to 3, 4, or 5 leap years during saros. During this period 40 solar eclipses occur (14 ring, 12 total, and 14 partial eclipses) and 26 lunar eclipses also occur (17 total and 9 partial). The previous numbers have to be understood only as approximate, because they change with time. In addition, these predicted solar eclipses may in fact not occur millions of years in the future, because the mean Earth–Moon distance continually increases (even though only 38 mm per year). On the other hand, the mean Earth–Sun distance may increase, too, as we shall see in Chapter 13. Finally note that the lunar eclipse period has the same length as the solar one.

The difference between the right and left sides of relation (6.1) is only 0.03567 of the day. By accumulation of these deviations every total eclipse after several periods will become partial and later it will not appear at all and will be replaced by a new eclipse. This is the reason why these periods are numbered. Their numbering was introduced by G. van den Bergh 1955. In particular, we have at present the following table for the solar eclipses (see [96]):

saros	last eclipse	next eclipse
130	February 26, 1998	March 9, 2016
135	August 22, 1998	September 1, 2016
140	February 16, 1999	February 26, 2017
145	August 11, 1999	August 21, 2017



²During 7 hours and 43 minutes the Earth rotates nearly 120° to the west. Therefore, after three periods of saros an eclipse occurs almost at the same location.

6.4. Why lunar eclipses are less frequent than solar eclipses

The answer to this question is given in Fig. 6.2. The Earth, illuminated by the Sun, casts a shadow in the form of a pointed cone, which is roughly $d = 1\,356\,000$ kilometers long. If the Moon enters into this cone a lunar eclipse will occur. Denoting by V the vertex of the cone of Earth's shadow, S the center of Sun, and E the center of Earth, the value of $d = |VE|$ can be readily determined from the similarity of right-angled triangles (cf. Fig. 6.2). Indeed, we have

$$\frac{r_E}{r_S} = \frac{|VE|}{|VS|} = \frac{d}{d + R_E},$$

where $r_E = 6378$ km, $r_S = 695\,990$ km, and $R_E = 1.496 \cdot 10^8$ km are the mean radius of the Earth, the radius of the Sun, and the Earth-Sun distance, respectively.

The mean distance of the Moon from the Earth is approximately $R_M = 384\,402$ km (i.e. the length of the line segment AE on Fig. 6.2). As already mentioned, an eclipse may occur only if the Moon is near the ecliptic. If the Moon is close to the arc \widehat{AB} (see Fig. 6.2), a lunar eclipse occurs, and if the Moon is close to the arc \widehat{CD} , a solar eclipse occurs somewhere on Earth.

However, notice that the arc \widehat{AB} is shorter than the arc \widehat{CD} . The actual angle at vertex V is very small (about 0.5°), and therefore, the ratio of their lengths is roughly equal to the ratio of the lengths of the line segments AB and CD . Hence,

$$\frac{|AB|}{|CD|} = \frac{|AV|}{|CV|} \approx \frac{d - R_M}{d + R_M} \approx 0.56.$$

We see that this ratio is roughly equal to the observed ratio

$$\frac{26 \text{ lunar eclipses}}{40 \text{ solar eclipses}} = 0.65$$

during the saros period (see Section 6.3).

Consequently, there are more solar than lunar eclipses. A lunar eclipse may occur up to three times during one year, but in some years

it may not occur at all. On the other hand, we can see 2–5 solar eclipses of various types every year. On average, there are two total solar eclipses within three years. The belt dimensions (max. width 270 km and length of thousands or ten-thousands of km) of the Moon’s shadow on the Earth’s surface do not allow frequent observations of a total solar eclipse. For one particular place on Earth, on average there is a total solar eclipse only once every 360 years.

The duration of a solar eclipse depends primarily on the observer’s distance from the central line of the belt. A partial solar eclipse can last up to two and a half hours, while the duration of a total eclipse is only a few minutes (on the central line usually 1–4 minutes to a maximum of 7.6 minutes). The Moon’s shadow engulfs the Earth for a maximum of 6 hours and its surface speed across (always from west to east) is 600–1000 m/s.

An interesting phenomenon occurred on November 3, 2013 in Africa, where an annular solar eclipse changed into a total eclipse.



6.5. Consequences of light aberration during total eclipses

In this section we show that during a total solar eclipse the centers of the Earth, Moon, and Sun are paradoxically not exactly aligned, as it might seem. This is due to an inconspicuous phenomenon called aberration of light (see Section 2.9).

The Earth orbits the Sun at a speed of $v \approx 30$ km/s. Stars observed perpendicularly to the direction of the Earth’s instantaneous movement seem to be deflected by an angle of aberration α . To get an observed star to the middle of the range of vision of a telescope, the telescope should be just slightly shifted about this aberration angle α (see Fig. 6.3). The aberration of light is due to the finite speed of propagation of electromagnetic waves. The aberration angle α is relatively small, and thus it is almost equal to its tangent in radian measure. Therefore, for the Earth we can write (compare with (2.9))

$$\alpha = \frac{v}{c} = \frac{30}{300\,000} = 0.0001.$$

The aberration angle is therefore equal to $\alpha = 0.0001 \text{ rad} \approx 20''$.

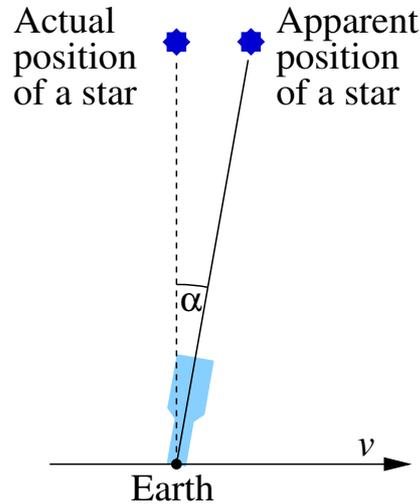


Fig. 6.3. The angle of aberration of light α appears during the movement of an observer with respect to the light source. The telescope has to be inclined by an angle α , which is in reality much smaller than shown on the picture.

Now let us ask the question, when does a solar eclipse actually occur, since the phenomenon of light aberration has indeed some quite unexpected consequences. It is generally believed that during a total solar eclipse, the centers of the Earth, Moon, and Sun are exactly aligned. Due to the aberration of light, this is not true. All three bodies will be in a straight line only 40 seconds after the middle of the eclipse, which might be after the total eclipse if it takes e.g. only one minute (see Section 6.4). This is a little known fact.

To understand this surprising statement we recall a few basic facts. First, the Moon moves each day along the ecliptic with respect to the Sun by about 12° ($= 360^\circ/30$) eastward, i.e. about half a degree per hour and $30''$ per minute. Hence, to move about the aberration angle $\alpha = 20''$ the Moon needs 40 seconds. In fact, we should also take into account the intrinsic light aberration of the Moon. However, the speed of the Moon around the Earth is about 1 km/s only, which is

almost negligible when compared with the Earth's speed v around the Sun. The angle of aberration of light of the Moon with respect to the Earth is just

$$\alpha' = 0.7'' . \tag{6.2}$$

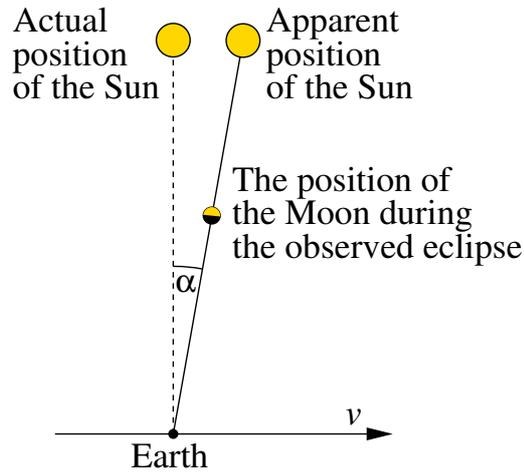


Fig. 6.4. The centers of the Earth, Moon, and Sun are not aligned during a total solar eclipse. This paradox is caused by the aberration of light.

For better understanding, the entire situation is illustrated in Fig. 6.4. Due to the aberration, caused by the finite speed of light, a special phenomenon arises that actually occurs in a way different from how it is observed.



7. Zwicky's postulation of dark matter

*Never identify any model
with physical reality.*

Basic physical rule

7.1. Fritz Zwicky

In this chapter we describe in detail how Fritz Zwicky postulated the existence of dark matter. In 1933, he published the groundbreaking paper [304] that essentially changed the development of astronomy and cosmology for many decades. Using the Virial theorem, he found that in order to explain fast movements of galaxies in the giant galaxy cluster Abell 1656 in Coma Berenices (the so-called *Coma galaxy cluster*) he had to assume the existence of a 400 times larger amount of nonluminous than luminous matter to keep the cluster gravitationally together. Sinclair Smith [262, p.27] independently reached a similar conclusion for the Virgo cluster whose distance is about 15 to 22 Mpc. Later the huge factor 400 was reduced to approximately 10 (see e.g. [196]).

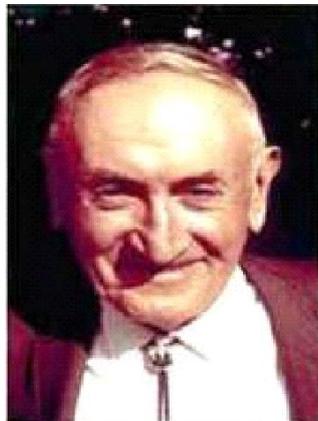


Fig. 7.1. Fritz Zwicky (1898–1974)

The term *dark matter* (in German *dunkle Materie*) was used in Zwicky's paper [304, p. 125] from 1933. This term first appeared in Jan Oort's paper [198, p. 285] from 1932. Dark matter should decelerate the expansion of the universe. However, later an accelerating expansion of the universe was predicted (see [94], [160], [277]). It started to be verified also experimentally at the end of the 20th century [207] and [226]. Therefore, physicists introduced another term *dark energy*, which on the contrary accelerates the expansion of the universe and thus acts against gravity (see an extensive survey of literature on this topic in [222]).

Fritz Zwicky was born on February 14 in 1898 in Varna (Bulgaria). His mother Františka born Vrčková was Czech and his father Fridolin Zwicky was Swiss. During the period 1916–1925 young Fritz studied mathematics and experimental physics on the Eidgenössische Technische Hochschule in Zurich. Then he emigrated to the USA. He worked at the well-known observatories at Mt. Wilson, Mt. Palomar, and also at the California Institute of Technology in Pasadena, where he received a professorship in astronomy in 1942.

In 1934, Zwicky together with Walter Baade (1893–1960) postulated the existence of neutron stars¹ as remainders of supernova explosions and stated the hypothesis that this could be a source of cosmic rays (see [11]). Later they suggested that supernovae could be promising candidates for measuring cosmic expansion, since they are observable from the farthest depths of the universe. Their idea was later used by the 2011 Nobel Prize laureates for Physics to discover the accelerated expansion of the universe (see [207], [226], [206]). In [306, p. 237] Zwicky introduced his new method of gravitational lensing by an intervening galaxy. He noticed that the probability of one galaxy overlying another galaxy is much larger than the probability of one star covering another star [164]. In [306] he even considered grav-

¹Already in 1932, Lev D. Landau in [154, p. 228] briefly mentions stars which could have the density of an atomic nucleus. At present, we assume that the core of neutron stars consists of cold quark-gluon plasma and that their mean density could be up to three times higher than the density of a neutron itself.

itational lensing by galaxy clusters. Illustration of this phenomenon is shown in Fig. 7.2.

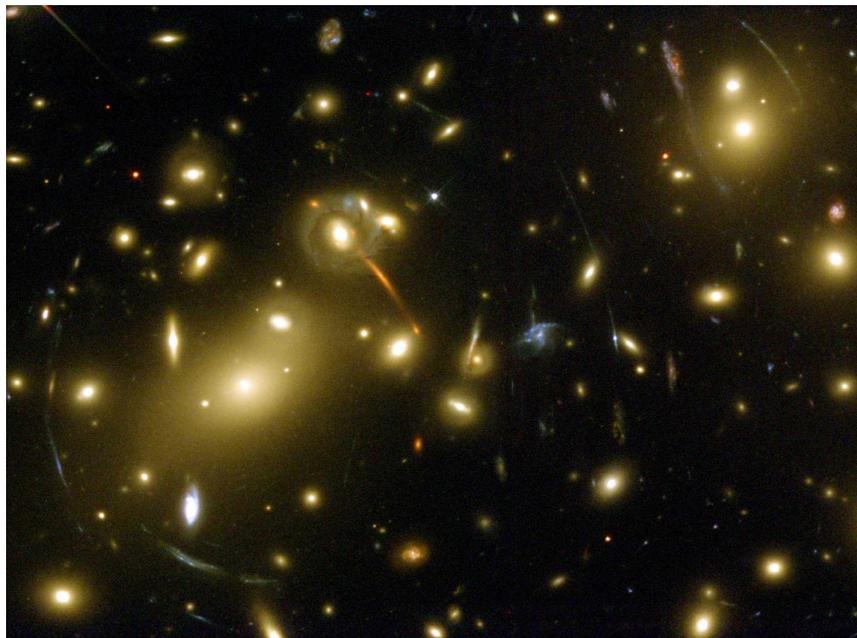


Fig. 7.2. The deformation of spacetime by the galaxy cluster A2218 is revealed by gravitational lensing predicted by Fritz Zwicky (photo NASA).

In 1929, the astronomers Edwin P. Hubble and Milton L. Humason found that the universe expands.² At the same time Zwicky drew attention by his article [303] to an alternative explanation of cosmological redshift³ of galaxies other than due to cosmic expansion. He proposed his own theory of the so-called tired light. He believed that the redshift is caused by the loss of energy of photons that give a part of their momentum to neighbouring matter. Fritz Zwicky was right only partly. A photon from a star may give a part of its momentum to some atom, but then, in general, a rescattered photon changes its direction and thus cannot be detected. The intergalactic medium has such a small density (only a few protons per m^3 , where $m_{\text{proton}} \approx 1.67 \times 10^{-27}$ kg), that most of the photons reach our detectors without any interactions along their intervening path, which is proved by the detection of sharp spectral lines. Later, Zwicky admit-

²A detailed chronological description of this discovery is given in the monograph [196].

³The term cosmological will, for simplicity, often be omitted.

ted his mistake and wrote another paper [305] on the interpretation of the variable width of spectral lines from rotated galaxies. The redshift of spectral lines also played an important role in Zwicky's prediction of dark matter.



7.2. The Virial theorem

Zwicky's postulation of dark matter is based on the Virial theorem which provides an approximate estimation of the mean kinetic energy of bound systems. First of all we will formulate this theorem. In three-dimensional space \mathbb{E}^3 , consider a system of N mass points with masses m_1, \dots, m_N that mutually interact gravitationally and that are not influenced by any other forces. Denote their position vectors by r_1, \dots, r_N with respect to a chosen coordinate system (the arrow above r_i will, for simplicity, again be omitted), i.e., for each time instant t the vector $r_i(t) \in \mathbb{E}^3$ is a point of the trajectory of the i th mass point. Thus, it is the standard N -body problem from Chapter 5. Then the kinetic and potential energy of this system are given by

$$T = \frac{1}{2} \sum_{i=1}^N m_i \dot{r}_i \cdot \dot{r}_i, \quad V = - \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Gm_i m_j}{|r_j - r_i|}, \quad (7.1)$$

where $\dot{r}_i = dr_i/dt$ stands for the time derivative, $G = 6.674 \times 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$ is the gravitational constant, $|\cdot|$ is the norm in \mathbb{E}^3 , and \cdot is the scalar product in \mathbb{E}^3 . The usual product in \mathbb{E}^1 will be denoted by \times in this and the next chapter. For simplicity, we do not indicate the dependence of T , V , r_i , \dots on time t .

Applying Newton's second law $F_i = m_i \ddot{r}_i$ and the gravitational law, we obtain for the acceleration of the i th body the differential equation (5.8),

$$\ddot{r}_i = \sum_{j \neq i}^N \frac{Gm_j (r_j - r_i)}{|r_j - r_i|^3}. \quad (7.2)$$

From this and (7.1) we have

$$\begin{aligned}
V &= -\frac{1}{2} \sum_{i=1}^N m_i \sum_{j \neq i}^N \frac{Gm_j (r_j - r_i) \cdot (r_j - r_i)}{|r_j - r_i|^3} \\
&= \frac{1}{2} \sum_{i=1}^N m_i \sum_{j \neq i}^N \frac{Gm_j (r_j - r_i) \cdot r_i}{|r_j - r_i|^3} + \frac{1}{2} \sum_{i=1}^N m_i \sum_{j \neq i}^N \frac{Gm_j (r_i - r_j) \cdot r_j}{|r_i - r_j|^3} \\
&= \frac{1}{2} \sum_{i=1}^N m_i \ddot{r}_i \cdot r_i + \frac{1}{2} \sum_{j=1}^N m_j \ddot{r}_j \cdot r_j = \sum_{i=1}^N F_i \cdot r_i, \tag{7.3}
\end{aligned}$$

where the indices i and j are interchanged in the last double sum at the end of the second line of (7.3).

Let $I = \sum_i m_i r_i \cdot r_i$ be the polar moment of inertia (see [306, p. 229]). Then from (7.1) and (7.3) we get

$$\ddot{I} = 2 \sum_{i=1}^N m_i (\dot{r}_i \cdot \dot{r}_i + \ddot{r}_i \cdot r_i) = 4T + 2V. \tag{7.4}$$

For stabilized systems with a large amount of particles the value I is almost constant in time. Also the total kinetic energy T and the total potential energy V practically do not change. One version of the Virial theorem for mass particles states that if $\ddot{I} = 0$, then

$$V = -2T. \tag{7.5}$$

Under the assumption that the total mechanical energy $E = T + V$ of the system does not change (cf. [145]), for gravitationally stabilized systems we moreover have by (7.5) that

$$E = \frac{1}{2}V. \tag{7.6}$$

It is difficult to define what a gravitationally stabilized system is. Consider, e.g., only two bodies orbiting along elongated elliptic trajectories (cf. Fig. 5.2). Then equality (7.6) does not hold, since on the left-hand side there is a constant, whereas the right-hand side

oscillates. Thus we should choose $N \gg 1$ and instead of E , T , and V , we take only average values over long time intervals.

Hence, the Virial theorem is not a mathematical statement with exactly formulated assumptions, but a certain proposition that was mainly verified experimentally. It was already known at the beginning of the 19th century. The term *virial* is the original denotation for the potential energy $V = \sum_i F_i \cdot r_i$ (see (7.3)). In 1870, Rudolf Clausius derived the **Virial theorem** (cf. (7.1), (7.3), and (7.5)) for $E < 0$ in the form

$$\left\langle \sum_{i=1}^N m_i v_i^2 \right\rangle + \left\langle \sum_{i=1}^N F_i \cdot r_i \right\rangle = 0,$$

where the angular parentheses denote mean values over a very long time interval and $v_i = |\dot{r}_i|$.



7.3. Zwicky's application of the Virial theorem to the Coma cluster

Around 1915 Vesto Merlin Slipher [261] discovered that spectra of the majority of galaxies show an unexpected redshift. Later, Fritz Zwicky also dealt with this problem. He was wondering why redshifts of spectral lines of individual galaxies from the Coma cluster (see Fig. 7.3) have such a large dispersion from the average redshift of the whole cluster. In the observed region, this dispersion is so large that about 15 galaxies exhibit blueshift,⁴ even though the whole cluster moves away (at 2% of the speed of light) from us due to the expansion given by the present value of the Hubble constant⁵, see [212],

$$H_0 \approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}. \quad (7.7)$$

⁴These galaxies are observed only in projection and reach velocities approaching up to 350 km/s. However, by the left part of the histogram of Fig. 8.4 from the next chapter they do not belong to the Coma cluster at present, even though they may have belonged to it in the past.

⁵It is not easy to establish the present value of the Hubble constant H_0 , since we always look into the past. It is difficult to reliably extrapolate the present value of H_0 from velocities in a distant universe. Also, in our close neighborhood, the measurement of H_0 is biased by the local movements of galaxies.

In another words, an object at a distance of one million parsecs will recede from us at the mean speed 70 km/s. However, the expansion rate of the universe depends on time as we shall see in Chapter 8 (cf. Fig. 8.7).



Fig. 7.3. Giant galaxy cluster Abell 1656 in the constellation Coma Berenices. In the middle there are two supergiant elliptic galaxies NGC 4889 and NGC 4874 that grew larger due to galactic cannibalism (photo NASA).

Galactic clusters represent large cosmic laboratories for testing theories of gravitation. In [304] Zwicky found that some galaxies orbit the center of A1656 much faster than would follow from the Virial theorem. He assumed $N = 800$ galaxies in the cluster. From measured luminosities he deduced that the mass of each galaxy is on average 10^9 Sun's masses. In this way he got the following estimate of the total mass (see [304, p. 124])

$$\mathcal{M} = 800 \times 10^9 \times M_{\odot} = 1.6 \times 10^{42} \text{ kg}, \quad (7.8)$$

where $M_{\odot} = 2 \times 10^{30}$ kg is the mass of Sun. Nevertheless, from the Virial theorem he obtained a 400 times higher mass \mathcal{M} of the cluster (see his paper [304, p. 125] from 1933). Four years later he published a more detailed analysis [306], where a similar factor is reduced to 150

(cf. (7.8) and (7.17)). To explain this discrepancy he assumed that there exists a large amount of some invisible dark matter that has gravitational influence.

Let us take a closer look at Zwicky's method for establishing the total mass of galaxy clusters by means of the Virial theorem. We try to follow his original notation from the papers [304] and [306]. Denote the total mass of the Coma galaxy cluster by

$$M = \sum_{i=1}^N m_i, \quad (7.9)$$

where m_i is the mass of the i th galaxy, and let \bar{v}_i be the time independent mean speed of the i th galaxy with respect to the Sun.⁶ Then the center of gravity of the cluster recedes with the mean speed

$$\bar{v} = \frac{1}{M} \sum_{i=1}^N m_i \bar{v}_i. \quad (7.10)$$

Zwicky then approximated the total kinetic energy of galaxies with respect to the center of the cluster as follows

$$\bar{T} = \frac{1}{2} M \bar{v}^2 = \frac{1}{2} \sum_{i=1}^N m_i \bar{v}^2 := \frac{1}{2} \sum_{i=1}^N m_i (\bar{v} - \bar{v}_i)^2, \quad (7.11)$$

where the root-mean-square speed \bar{v} of all galaxies with respect to the center of mass of the cluster is defined by the last equality in (7.11).

To estimate the potential energy of the cluster, Zwicky assumed that galaxies are uniformly distributed in a sphere with radius R . The corresponding constant density is

$$\rho = \frac{3M}{4\pi R^3}.$$

The position of the i th galaxy is given by the radius-vector r_i originating at the center of the Coma cluster. The force acting on this

⁶By chance the Coma cluster is located near the north Galactic pole. Therefore, \bar{v}_i is practically equal to the recession speed of the i th galaxy from the Milky Way, even though the orbital speed $v_{\odot} = 230$ km/s of the Sun about the Galactic center is relatively high.

galaxy can thus be approximated by means of Newton's first and second theorem from Section 4.1 as follows

$$F_i \approx -\frac{4\pi|r_i|^3\rho m_i r_i}{3|r_i|^3} = -\frac{GMm_i r_i}{R^3},$$

taking into account that $M \approx M - m_i$. The corresponding potential energy of the i th galaxy is then

$$V_i = F_i \cdot r_i \approx -\frac{GMm_i|r_i|^2}{R^3}. \quad (7.12)$$

Then Zwicky estimated the root-mean-square distance \bar{r} from the cluster center for a typical galaxy (see [306, p. 230]) as

$$\bar{r}^2 = \frac{1}{M} \sum_{i=1}^N m_i |r_i|^2 \approx \frac{3}{4\pi R^3 \rho} \int_0^R r^2 \times 4\pi r^2 \rho dr = \frac{3R^2}{5}, \quad (7.13)$$

where the density ρ of the cluster is assumed to be constant. By (7.3), (7.12), (7.9), and (7.13) we have

$$V = \sum_{i=1}^N F_i \cdot r_i \approx -\frac{GM}{R^3} \sum_{i=1}^N m_i |r_i|^2 \approx -\frac{GM\bar{r}^2}{R^3} \sum_{i=1}^N m_i = -\frac{3GM^2}{5R}.$$

From this, the Virial theorem (7.5), and estimate (7.11), we get the resulting approximate relation

$$\boxed{M = \frac{5R\bar{v}^2}{3G}} \quad (7.14)$$

for the total mass (see [304, p. 124] and also [306, p. 230]). We shall call M the *virial mass*.

Zwicky used the following data to establish the radius R . He assumed that the Coma cluster in the celestial sphere subtends the angle $\beta = 1.7^\circ$ (for comparison, our Moon subtends an angle of only 0.5°). At that time, Hubble along with Humason estimated the distance to the cluster to be 13.8 Mpc. Since $1 \text{ pc} = 3.086 \times 10^{16} \text{ m}$, we have

$$R = 13.8 \times 10^6 \times 3.086 \times 10^{16} \times \sin \frac{1}{2}\beta = 6.318 \times 10^{21} \text{ (m)}, \quad (7.15)$$

i.e. $R \approx 0.2 \text{ Mpc}$.

Radial components⁷ of velocities of individual galaxies can be well-determined from the Doppler effect. From data available at that time, Zwicky found that their redshifts have a large dispersion from the mean value of the whole cluster, even though by [248, p. 14] he considered only the eight largest galaxies. From this he calculated the square of the mean radial velocity $\overline{v}_{\text{radial}}^2 = 5 \times 10^{11} \text{ m}^2\text{s}^{-2}$ with respect to the cluster center. Due to the expected isotropic distribution of velocities and spherical symmetry, for the mean value \overline{v} he obtained

$$\overline{v}^2 = 3 \overline{v}_{\text{radial}}^2 = 1.5 \times 10^{12} (\text{m/s})^2 \quad \text{and} \quad \overline{v} = 1.22 \times 10^6 \text{ m/s}. \quad (7.16)$$

The first equality in (7.16) follows from the Pythagorean theorem if the vector of velocities is expressed by three mutually orthogonal components (see also [121, p. 2692]).

Substituting (7.15) and (7.16) into (7.14), we find the amount of the virial mass

$$M = 2.367 \times 10^{44} \text{ kg}, \quad (7.17)$$

which is almost 150 times higher value than found in (7.8). This discovery by Zwicky was ignored for many decades. When he compared the measured luminosity of the whole cluster with its theoretical value, he got a similar factor 500/3 (see [306, p. 232]).

In [306] Zwicky suggested two other methods for revealing dark matter. The first method is based on gravitational lensing by means of an intervening galaxy (cf. Fig. 7.2). The second method uses statistical evaluation of luminosities. In the introduction of [306] Zwicky examines rotational curves of galaxies, which is another tool for detecting the influence of dark matter (see Chapter 9).



⁷Annual variations of radial velocities caused by the revolution of the Earth around the Sun by the speed (2.7) are only $\pm 26 \text{ km/s}$, since the inclination of ecliptic to the Galactic plane is 62° .

8. The problem of missing matter

The important thing is not to stop questioning.

ALBERT EINSTEIN

8.1. Analysis of Zwicky's method

Fritz Zwicky made many groundbreaking discoveries as we know from the previous chapter. For instance, in his paper [304] from 1933 he derived the formula (see (7.14))

$$\boxed{M = \frac{5R\bar{v}^2}{3G}} \quad (8.1)$$

for the virial mass of the Coma galaxy cluster A1656, where R is its radius and \bar{v} is the mean quadratic speed with respect to the center of gravity of the cluster. Zwicky's approach that led him to propose the existence of dark matter from relation (8.1) requires, however, a more detailed discussion.

Let us first recall the main trick of Zwicky's method from Section 7.3. By means of the Virial theorem $V = -2T$ (see (7.5)) we may relate the total potential and kinetic energy of the cluster, i.e.

$$V \approx -\frac{3GM^2}{5R}, \quad T \approx \bar{T} = \frac{1}{2}M\bar{v}^2$$

(see (7.11)–(7.14)). Notice that M appearing in the potential energy is squared, whereas in the kinetic energy it appears in the first power. This enables us to express the virial mass M by (8.1), where we should rather write \approx instead of the equality $=$. Nevertheless, can we reliably

assert on the basis of such a simple algebraic formula as (8.1) that the Coma cluster from Fig. 7.3 contains a great amount of some mysterious dark matter of unknown composition?

Zwicky had to make many simplifying assumptions to estimate the total mass of the Coma cluster. In this section we collect a number of facts which should be taken into account in any attempt to decide whether Zwicky's approach can prove the existence of dark matter.

1. According to current measurements of velocities and the Hubble law (see (8.13) later), the distance to the Coma cluster is not 13.8 Mpc as stated in (7.15) but about $d = 100$ Mpc. From this we get for the angular diameter $\beta = 1.7^\circ$ of the Coma cluster a much bigger radius

$$R = d \sin \frac{\beta}{2} = 1.48 \text{ Mpc} = 4.58 \times 10^{22} \text{ m}, \quad (8.2)$$

and thus also a larger mass than in (7.17). Altogether from (8.1), (7.16), and (8.2) the Virial theorem yields

$$M = 1.71 \times 10^{45} \text{ kg},$$

which is about 1000 times larger than the value given in (7.8). For an intuitive illustration about the size of the radius R of the Coma cluster, we point out that the distance between the centers of our Galaxy and the Andromeda galaxy M31 is about 0.778 Mpc (cf. (8.2)).

2. The angular diameter $\beta = 1.7^\circ$ from (8.2) is by more recent measurements somewhat greater. By [24] the Coma cluster is in a region of dimensions $2.7^\circ \times 2.5^\circ$ with a boundary that is not well defined. Some other sources give smaller values.

3. Zwicky in (7.8) supposes that the mass of each galaxy is on average 10^9 solar masses. These data are, on the other hand, considerably underestimated. A great amount of light from stars is absorbed by interstellar gas. For comparison, our Galaxy has about 400×10^9 stars and its total mass is approximately $M_G = 10^{12} M_\odot$ (see [156, p. 127]), which is even more than the total mass \mathcal{M} of all 800 galaxies from (7.8) estimated by Zwicky. This is one of several arguments why

cosmologists think at present that there is one order of magnitude more dark matter than luminous baryonic matter formed mainly by protons and neutrons.¹ Note that dark matter if it exists does not shine in any range of the electromagnetic spectrum.

Zwicky made a number of further approximations which have an essential influence on the resulting calculation of the mass:

4. The estimate of the total number of galaxies $N = 800$ (see (7.8)) is slightly underestimated, although Zwicky in [306, p.244] admits $N \geq 1500$. At present we know more than one thousand galaxies in the Coma cluster. Each Mpc^3 thus contains at least 70 galaxies on average. In 1933, Zwicky could not observe from the 18 inch telescope at Mt. Palomar² many dwarf galaxies in the cluster that contribute to the total mass of the cluster. Their number may be large. For instance, about ten dwarf galaxies were recently discovered in a close neighborhood of our Galaxy.

5. Relation (7.9) does not consider intracluster matter inside the cluster. Its density is quite high in the central region of the cluster. Current data show that there exist gas, dust, rogue planets, and solitary stars that were ejected from galaxies by various gravitational collisions [282]. From X-ray analysis (see [2], [27], and [106]) we know that the intracluster medium contains at least five times more nonluminous baryonic matter than luminous matter in galaxies.

6. Zwicky considered a uniform distribution of galaxies inside the cluster A1656 (see [306, p.229]). However, the central region is much denser than the region around the boundary (see Fig. 8.6) and larger galaxies are closer to the center, in general. In another words, a galaxy cluster has a higher density towards the center (as in globular clusters) than for a uniform distribution. This is the reason why the coefficient $\frac{5}{3}$ from (8.1) is overestimated for the observed nonuniform, but almost radially symmetric mass distribution, as we shall show in

¹Astronomers include to baryonic matter all particles that are contained in the Standard model of elementary particles (protons, neutrons, but also electrons, neutrinos, etc.) and their interactions, and degenerate baryonic matter hidden in black holes, too.

²The famous Mt. Palomar 5 m telescope was not finished until 1949.

Section 8.3. Since the cluster is seen only in projection, it can also be elongated or flattened.

7. The kinetic energy \overline{T} with respect to the center of gravity and averaged over long-time intervals is not expressed exactly, since the mean recession velocity \bar{v}_i from (7.11) over long time intervals is replaced by the current value $v_i(t) = |\dot{r}_i(t)|$ for a given $i \in \{1, \dots, N\}$. At present we know the values of the heliocentric radial components of velocities of individual galaxies with respect to the Sun more precisely, but we cannot establish their mean values over long time intervals as required by the Virial theorem. From the data presented in [1], [24], and [54] for galaxies belonging to the Coma cluster we obtain radial components of velocities (see Fig. 8.1) and

$$\bar{\bar{v}} \approx 1.686 \times 10^6 \text{ m/s}, \quad (8.3)$$

which is even more than \bar{v} from (7.16) derived by Zwicky. The reason is that he considered only the brightest galaxies in his calculation.

The potential energy (7.12) is also not stated exactly due to the approximation $M \approx M - m_i$.

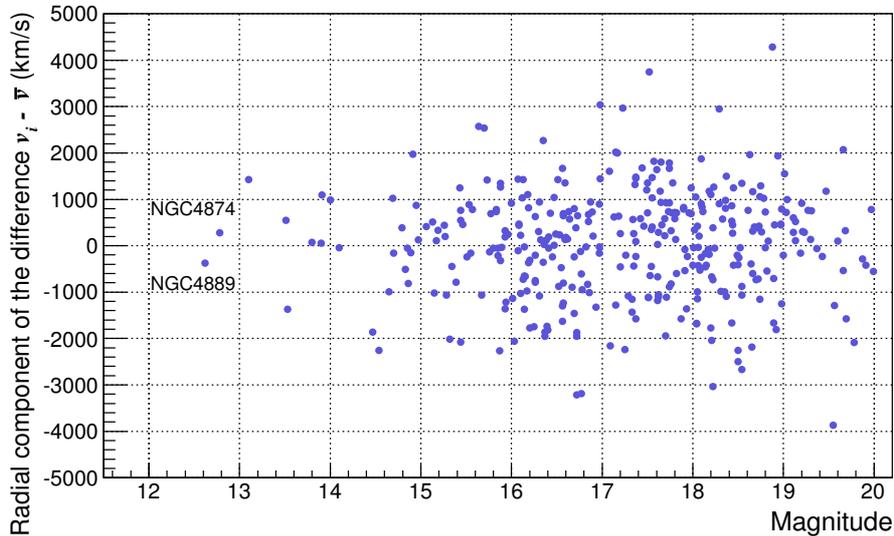


Fig. 8.1. The dependence of the radial component of the difference $v_i - \bar{v}$ on magnitude for galaxies from the Coma cluster, where $v_i = v_i(t)$ corresponds to the present time.

8. Zwicky assumed an isotropic distribution of velocities. He approximated their root-mean-square speed by means of radial velocities and relation (7.16). The measured values of radial velocities³ are sketched in Fig. 8.1. In the histogram of Fig. 8.5, we can see a slight asymmetry with respect to the Gaussian curve (see also [191]).

9. Dark energy might also contribute to larger velocities of galaxies inside the Coma cluster. There is no reason to assume that dark energy would somehow avoid the Coma cluster which is a part of the expanding “cosmic web”. The linear growth rate of large cosmic structures was observed by [221]. Moreover, as shown in [64], [65], [131], [133], [137], [145], [302], dark energy also acts locally due to repulsive antigravitational forces. This could also modify the estimated mass as we shall see in Section 8.3. In addition, the speed in (8.1) is squared! Therefore, a reliable knowledge of radial velocities is essential.

10. Zwicky restricted himself to the case in which all galaxies have the same time independent mass [306, p. 231]. Nevertheless, galaxies exchange mass with the intracluster environment and their magnitudes⁴ differ by eight orders of magnitude (see Fig. 8.1).

Let us put forward some other facts which should be taken into account for a thorough error analysis.

11. Zwicky assumed that the Coma cluster is in equilibrium and that the Virial theorem holds exactly. Since their formation, galaxies could orbit the cluster center only a few times by the velocity \bar{v} from (8.3), since one period takes about

$$2\pi\bar{r}/\bar{v} = 4.11 \times 10^9 \text{ years}, \quad (8.4)$$

where $\bar{r} = \sqrt{3}R/\sqrt{5}$ is the mean distance from (7.13) and R is given

³Stars in the Sun’s neighborhood revolve about the center of our Galaxy with a speed about 230 km/s. Sometimes a star with speed over 1000 km/s appears in the Galaxy. It is assumed that these stars were ejected from globular clusters or from a close vicinity of black holes. For instance, when a binary star approaches another very massive object, then the more massive component can be caught on an elliptic orbit, while the lighter one is flung away along a hyperbolic trajectory. Similar collisions of galaxies also appear in the Coma cluster.

⁴The weaker a light source is, the higher is its magnitude. The difference of one magnitude corresponds to the ratio $100^{1/5} : 1 = 2.512 : 1$ of luminosity fluxes.

by (8.2). Although \bar{r} and \bar{v} are slightly overestimated with respect to the mean values, we probably cannot talk about the relaxed (stabilized) system and thus one has to defend whether the mechanical use of the Virial theorem is well-founded.

12. Zwicky used Newtonian mechanics with an infinite speed of gravitational interaction, whereas the real speed is surely finite. From Fig. 8.5 we shall see that the velocities of some galaxies with respect to the cluster center are more than 1 % of the speed of light, i.e. long-term relativistic effects influence the evolution of the system. Thus also gravitational aberration effects are not negligible (see [145]). We know well the functioning of Newtonian mechanics on short time scales and low velocities in the Solar system. However, by (8.2) the cluster has a diameter of approximately $3 \text{ Mpc} > 6 \times 10^{11} \text{ au}$, where $1 \text{ au} \approx 150 \times 10^9 \text{ m}$ is the mean Earth-Sun distance. It is not clear whether we are allowed to employ Newton's laws to objects many orders of magnitude larger.⁵

13. Zwicky replaced galaxies of diameter 10^{10} au by mass points. This does not allow one to consider spin angular momenta of rotating galaxies which surely contribute to the total angular momentum, even though orbital angular momenta are much larger than spin angular momenta, in general. We also cannot include tidal forces that influence the dynamics of the system. For instance, an "isolated binary" system of two galaxies that orbit closely to each other is not stable, since galaxies will merge due to tidal friction (cf. Fig. 5.7), whereas the classical two-body problem has a periodic solution on an infinitely long time interval (see Fig. 5.2).

Zwicky restricted himself to the case in which N is constant. However, sometimes galaxies merge due to tidal forces or various collisions in the densely populated space (see Fig. 7.3).

⁵This is similar to the application of the laws of quantum mechanics (e.g. discrete electron orbits of atoms, tunnel phenomenon, dispersion of electrons on a slot) to objects of scale on the order of meters.

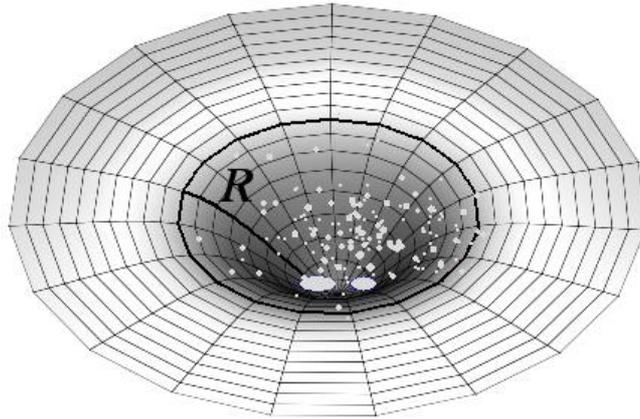


Fig. 8.2. Deformation of spacetime due to the high density of galaxies in a cluster of radius R . The circumference of the circle with radius R is smaller than $2\pi R$.

14. Zwicky substituted the spacetime curved by a thousand of galaxies (see Fig. 8.2) with a total mass about 10^{45} kg by the Euclidean space \mathbb{E}^3 . Deformation of the spacetime containing a galaxy cluster causes gravitational lensing (see Figs. 7.2 and 8.3). In fact, we need not consider an intervening galaxy as Zwicky suggested in [306], since the cluster itself magnifies the perception of distances between its own objects (see Fig. 8.3) and the observed angular distances of individual galaxies from the center of Coma cluster. Note that the “index of refraction” of a gravitational lens represented by a galaxy cluster increases toward the center, where all angular sizes seem larger. The selflensing effect (see Fig. 8.3) modifies the observed density profile so that the cluster seems to be more sparse. It also enlarges the observed radius R in (8.1), (7.15), and (8.2) which then yields a greater calculated virial mass M than the cluster has in reality.

Moreover, the volume of a sphere in such a deformed space is not $4\pi R^3/3$ (cf. e.g. (7.13) and Fig. 8.2), but is smaller due to the Bishop-Gromov inequality [209], (see also [184, p.1099]). Analogously, the surface of a sphere integrated in (7.13) is smaller than $4\pi r^2$.

The curved spacetime brings other effects which we analyze in detail in Section 8.3. For instance, the total redshift is caused not only by the cosmological expansion of the universe, but also partly by a gravitational redshift. Photons have to overcome the potential well

of a star,⁶ as well as a much deeper potential well of the corresponding galaxy and the potential well of the whole cluster of galaxies. The relation for change of frequency in a field of central force is given in [184].

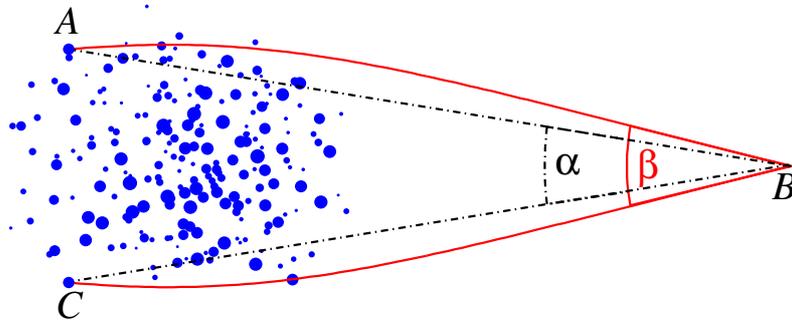


Fig. 8.3. A schematic illustration of the selflensing effect. The observation angle $\beta = \widetilde{\sphericalangle}ABC$ is larger than the angle $\alpha = \sphericalangle ABC$ due to the bending of light caused by the gravitation of the galaxy cluster itself, where $\widetilde{\sphericalangle}ABC$ denotes a curved angle.

15. Further sources of uncertainties are in input data and rounding errors (e.g. (7.11) represents a large sum with many terms of various orders of magnitude). The proposed distance to the center of the Coma cluster varies from 99 Mpc to 103 Mpc in the literature. According to [24], the right ascension of the cluster center is $\alpha = 12 \text{ h } 57.3 \text{ min}$ and the declination $\delta = 28^\circ 14.4'$. Other sources give different data, for instance, $\alpha = 13 \text{ h } 00 \text{ min } 00.7 \text{ s}$ and $\delta = 27^\circ 56' 51''$ by [230]. It is even not clear how to define the center, since the cluster is seen only in projection, its boundary is fuzzy and the speed of gravitational interaction is unknown (see the very end of Chapter 5).

A large amount of small errors of various origins may essentially distort the resulting calculation of the mass M . Zwicky became well aware that he committed a lot of simplifications and rough approximations that are treated in points **2**, **4**, **6**, **11**, **15** (see [306, p. 230, 231, 233, 242, 244]), otherwise he could not compute anything. However,

⁶The gravitational redshift of photons from neutron stars corresponds by its size to a cosmological redshift up to $z \approx 0.4$. However, for ordinary stars it is almost negligible.

he did not consider several important facts that are surveyed e.g. in points **9, 12, 13, 14**.

There exists a large number of different methods that have been applied to determine the total mass of the Coma cluster. They are based on the Virial theorem [165], gravitational lensing [84], [152], and X-ray emitting gas [2], [27], [35], [74], [106], [290].



8.2. Analysis of current data

The Coma cluster is located at the beginning of the Sloan Great Wall — a long fiber of galaxies. The cluster contains in its center two supergiant elliptic galaxies NGC 4889 and NGC 4874 which are 10 times larger than the Milky Way and contribute essentially to the dynamics of the whole cluster (see Fig. 7.3). Their magnitudes are 12.62 and 12.78. An extremely large black hole with mass $10^{10}M_{\odot}$ sits in the center of galaxy NGC 4889. On the other hand, Galaxy NGC 4874 contains a record number of 30 000 globular clusters (our Galaxy has about 150 globular clusters). At present, the Coma cluster has well measured redshifts and magnitudes of all its galaxies.

Unfortunately, Zwicky in [304] and [306] does not present any input data on velocities and magnitudes of individual galaxies from the Coma cluster. He only gives R and \bar{v} from (7.15) and (7.16), respectively.

Now let us show what Zwicky would get by his method for the present data. To reconstruct Zwicky's calculation, we employ data given in [1], [24], and [54], that partly overlap. They contain some galaxies that do not belong to the Coma cluster, although they are in the considered region of the celestial sphere [158]. There are for instance more than 50 galaxies whose radial speeds exceed 40 000 km/s. One galaxy (see the right part of Fig. 8.4) has radial speed 114 990 km/s, which is more than one third of the speed of light! By the special relativity formula [61]

$$z = \sqrt{\frac{c+v}{c-v}} - 1$$

its redshift is roughly $z \approx 0.5$. Note also that a galaxy whose radial speed is 40 000 km/s would fly through the distance corresponding to the radius (8.2) of the cluster within less than 50 million years. Thus, galaxies from the right part of Fig. 8.4 cannot be in the cluster.

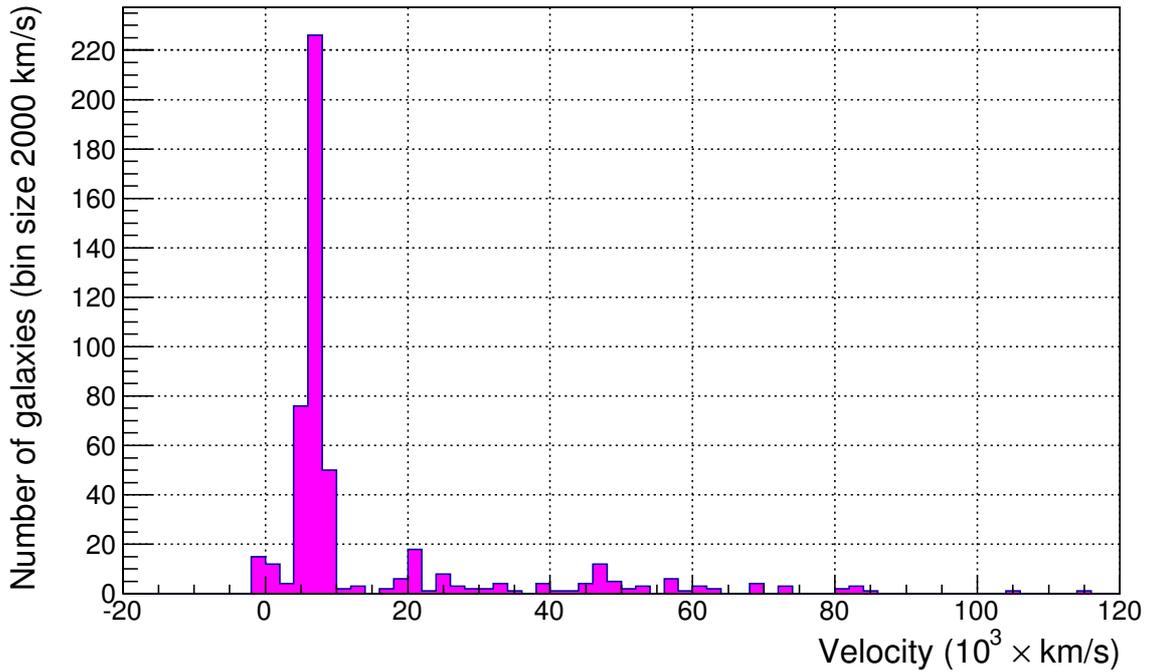


Fig. 8.4. Histogram of radial velocities of galaxies whose magnitude does not exceed 20 which are projected to the region about the Coma cluster.

The nonuniform distribution of velocities with respect to the cluster center is clearly seen in the histograms in Figs. 8.4 and 8.5. Note that some galaxies with blueshift or very small redshift form another independent cluster⁷ in the left part of Fig. 8.5. Thus to calculate the mean value (8.3) (and also (8.5) further on), we restrict ourselves only to galaxies having their velocities from the interval 2 000 – 12 000 km/s. A slightly different choice is given in [114], [165], and [230].

⁷This small cluster of about 30 galaxies acts as a weak supplementary converging lens.

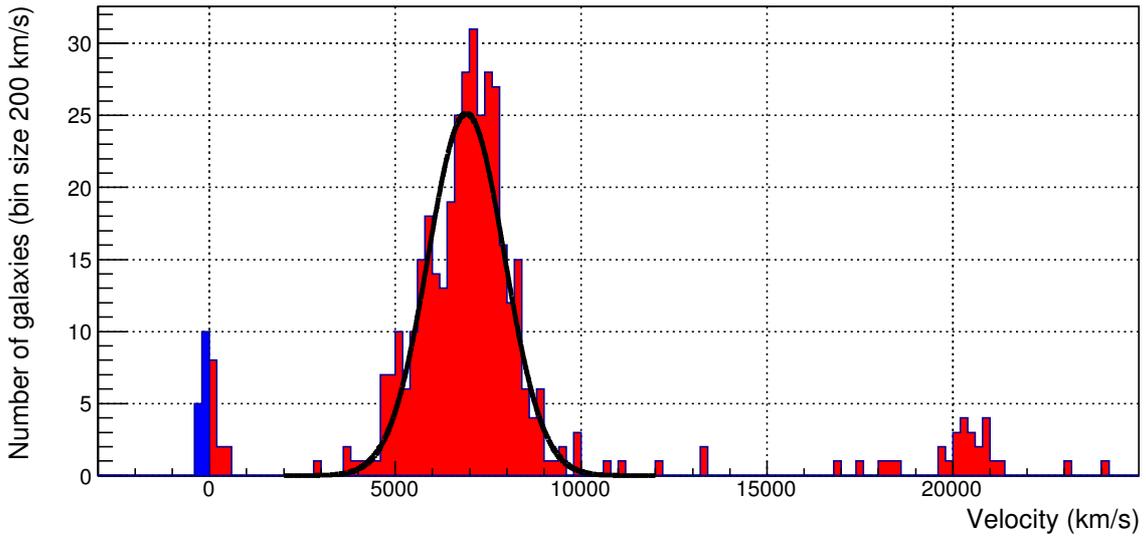


Fig. 8.5. A detail of the histogram from Fig. 8.4 for radial velocities less than 25 000 km/s. Galaxies possessing blueshift are on the left. The dark line represents the Gaussian curve fitted to the considered data.

Since the cluster has a relatively small angular radius, all its components are approximately at the same distance from us. We shall therefore assume that the mass of each galaxy is linearly proportional to its observed luminosity. Denote by I the density of light flux of some reference galaxy whose mass is m and whose magnitude is μ . Then the flux density I_i of the i th galaxy with mass m_i and the measured magnitude μ_i satisfy *Pogson's equation*⁸

$$\mu - \mu_i = 2.5 \log_{10} \frac{I_i}{I}.$$

Dividing this equality by 2.5 and using the inverse logarithm procedure, we get

$$10^{0.4(\mu - \mu_i)} = \frac{I_i}{I} = \frac{m_i}{m}.$$

We may therefore assume, for simplicity, that the masses m_i are by Pogson's relation proportional to $10^{-0.4\mu_i}$. This trick enables us to

⁸Pogson's equation shows a relationship between the magnitudes of the observed difference $\mu - \mu_i$ of two light sources and the ratio I_i/I of densities of the luminosity flux. If the ratio is $I_i/I = 100$, we see that the difference between the sources is of 5 magnitudes. For example, the observed magnitude of Polaris is $\mu = 2.2$ mag. Then the ratio is $I_i/I = \sqrt[5]{100} = 2.512$ for the source with a difference in magnitude equal to one.

calculate the mean speed \bar{v} (resp. $\bar{\bar{v}}$) from relation (7.10) (resp. (7.11) and (8.3)) without knowing the real masses m_i . It is enough to know only magnitudes of individual galaxies. In this way we obtain the mean recession speed of the whole cluster

$$\bar{v} \approx \frac{1}{M} \sum_{i=1}^N m_i v_i \approx \frac{\sum_i 10^{-0.4\mu_i} v_i}{\sum_i 10^{-0.4\mu_i}},$$

where the sum is taken over the 352 most luminous galaxies with known reference magnitudes not exceeding 20. Individual terms differ by several orders of magnitude. To reduce rounding errors, we have to sum from the smallest to the largest terms (see [33]). Inserting the measured magnitudes and velocities, we obtain

$$\bar{v} \approx 6877 \text{ km/s.} \quad (8.5)$$

By (8.1), (8.2), and (8.3) the total virial mass of the cluster is⁹

$$M = 3.25 \times 10^{45} \text{ kg.} \quad (8.6)$$

For comparison (see also (7.8)) let us give a lower bound of the total mass of the cluster based on Pogson's relation and the measured luminosities of individual galaxies

$$\mathcal{M} > C \sum_{i=1}^N 10^{-0.4\mu_i} = 3.3 \times 10^{44} \text{ kg,}$$

where $C = m 10^{0.4\mu}$ is the scaling constant and $\mu = 12.78$ mag is the magnitude of the comparative reference galaxy NGC 4874, which is according to [307] ten times more massive than our Galaxy, i.e.,

$$m = 10M_G = 10^{13}M_\odot = 2 \times 10^{43} \text{ kg,} \quad (8.7)$$

⁹From this and (8.2) we get the associated mean density $\rho = 8 \times 10^{-24} \text{ kg/m}^3$ of the cluster, which is substantially bigger than the present mean density $\approx 10^{-27} \text{ kg/m}^3$ of space. For comparison, by [31] the density of dark matter in our Galaxy is $0.008 M_\odot \text{pc}^{-3} = 5.444 \times 10^{-22} \text{ kg/m}^3$. However, by [187] this density is at least one order of magnitude smaller.

where the total mass of our Galaxy $M_G = 10^{12}M_\odot$ (see the point **3**). We observe that the total mass M derived from the Virial theorem is one order of magnitude greater than the lower bound for \mathcal{M} . Nevertheless, natural questions arise concerning how much additional mass is formed by approximately one thousand existing dwarf galaxies which are not included in the above sum, what is the contribution of nonluminous intracluster baryonic matter, and so on. A giant tidal tail of stars with total mass equal to 20% of the mass of galaxy NGC 4874 was detected in [91, p. 551]. According to [282], the intracluster medium contains 30–50% of stars from all the stars of the cluster. All these arguments contribute to a larger value of \mathcal{M} and help us to explain high observed velocities of galaxies.

When claiming that dark matter exists, we should first reliably estimate all possible errors from points **1–15** in Section 8.1 and perhaps some others. In particular, errors in **6**, **9**, and **11–15** may be quite large.



8.3. Reduction of the virial mass of the Coma cluster

Now, let us take a closer look on various errors in determining the virial mass (8.1) caused by some phenomena from Section 8.1.

Nonuniformity of mass distribution. We show that the coefficient $\frac{5}{3}$ appearing in (8.1) should be smaller. From Fig. 8.6 we observe that the distribution of galaxies in the Coma cluster is far from being uniform as was assumed by Zwicky. Let us suppose that the mass density $\rho = \rho(r)$ is spherically symmetric in the sphere of radius R given by (8.2). Then the whole mass of the Comma cluster can be expressed as

$$M = \int_0^R \rho(r) 4\pi r^2 dr. \quad (8.8)$$

Zwicky assumed that ρ is independent on r . Consider a more general density distribution which enables us to find a closer approximation

of the observed density profile, namely,

$$\rho(r) = a(R^b - r^b), \quad 0 \leq r \leq R, \quad (8.9)$$

where $b > 0$ and the parameter

$$a = \frac{3(b+3)M}{4\pi b R^{b+3}} \quad (8.10)$$

is chosen so that condition (8.8) is satisfied, i.e.,

$$\int_0^R a(R^b - r^b)4\pi r^2 dr = 4\pi a \left(\frac{R^{b+3}}{3} - \frac{R^{b+3}}{b+3} \right) = \frac{4\pi ab R^{b+3}}{3(b+3)} = M.$$

Now we will modify equation (7.12) derived for a constant density to the case of spherically symmetric mass density distribution ρ . For the force F_i acting on the i th galaxy with mass m_i , given by the radius-vector r_i , we get by Newton's first and second theorems from Section 4.1 and (8.9) that

$$\begin{aligned} F_i &= -\frac{Gm_i r_i \int_0^{|r_i|} 4\pi \rho(r) r^2 dr}{|r_i|^3} = -\frac{4\pi G a m_i r_i}{|r_i|^3} \left(\frac{R^b |r_i|^3}{3} - \frac{|r_i|^{b+3}}{b+3} \right) \\ &= -4\pi G a m_i r_i \left(\frac{R^b}{3} - \frac{|r_i|^b}{b+3} \right). \end{aligned}$$

Hence, the total potential energy is

$$V = \sum_{i=1}^N F_i \cdot r_i = -4\pi G a \sum_{i=1}^N m_i \left(\frac{R^b |r_i|^2}{3} - \frac{|r_i|^{b+2}}{b+3} \right). \quad (8.11)$$

Further, we need to evaluate the mean value (cf. (7.13)) of the power r^e for $e = 2$ and $e = b + 2$. By (8.8) and (8.9) we obtain

$$\begin{aligned} \langle r^e \rangle &= \frac{\int_0^R r^e \rho(r) 4\pi r^2 dr}{\int_0^R \rho(r) 4\pi r^2 dr} = \frac{4\pi a}{M} \int_0^R (R^b - r^b) r^{e+2} dr \\ &= \frac{4\pi a}{M} \left(\frac{R^{b+e+3}}{e+3} - \frac{R^{b+e+3}}{b+e+3} \right) = \frac{4\pi ab R^{b+e+3}}{M(b+e+3)(e+3)}. \end{aligned}$$

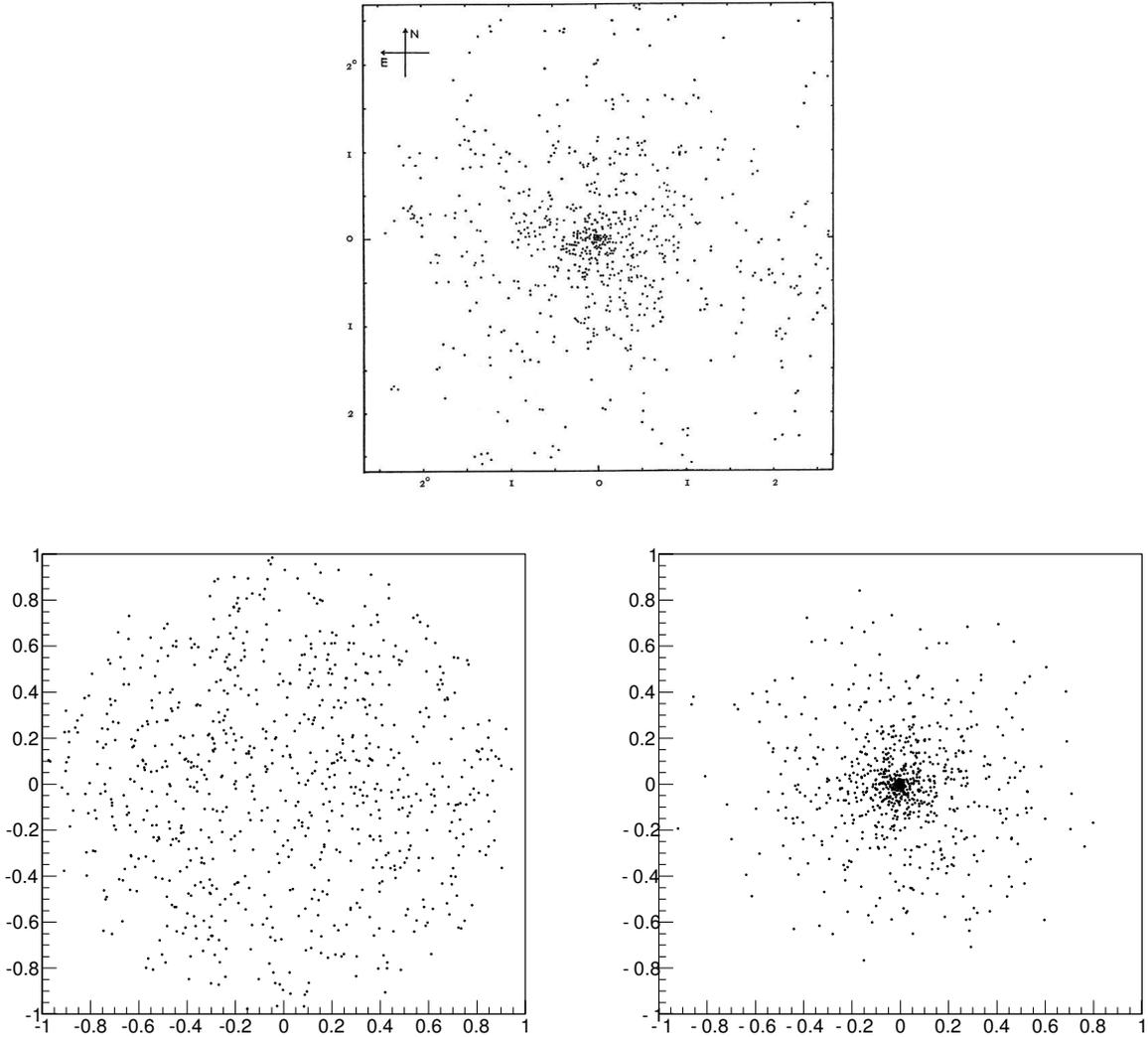


Fig. 8.6. The upper figure shows the distribution of galaxies in the Coma cluster from Zwicky's original paper [306, p. 227]. The bottom left figure illustrates a randomly generated uniform distribution of the same number of points inside a three-dimensional sphere projected to the plane and the right lower figure shows projected mass distribution (8.9) with $b = 1$ which is similar to the actual distribution from the upper figure.

Using this equality for the exponents $e = 2$ and $e = b + 2$, and relations (8.11), $M = \sum_i m_i$, and (8.10), we find that

$$\begin{aligned}
V &\approx -4\pi Ga \left(\frac{R^b 4\pi ab R^{b+5}}{3M 5(b+5)} - \frac{4\pi ab R^{2b+5}}{M(b+3)(b+5)(2b+5)} \right) \sum_{i=1}^N m_i \\
&= -(4\pi a)^2 GR^{2b+5} \left(\frac{b}{15(b+5)} - \frac{b}{(b+3)(b+5)(2b+5)} \right) \\
&= - \left(4\pi \frac{3(b+3)M}{4\pi b R^{b+3}} \right)^2 GR^{2b+5} \frac{b^2(2b+11)}{15(b+3)(b+5)(2b+5)} \\
&= - \frac{3GM^2 (b+3)(2b+11)}{5R (b+5)(2b+5)}.
\end{aligned}$$

From this, the Virial theorem (7.5), and (7.11) we obtain a new relation for the reduced virial mass

$$M = \frac{5R\bar{v}^2}{3G} \frac{(b+5)(2b+5)}{(b+3)(2b+11)}, \quad (8.12)$$

which converges to Zwicky's original estimate (8.1) for $b \rightarrow \infty$. The best fit of the parameter b of mass density distribution to Zwicky's data from the upper part of Fig. 8.6 seems to be close to the value $b \approx 1$. The corresponding coefficient $\frac{35}{26}$ from (8.12) is only 80 % of the value $\frac{5}{3}$ from (8.1) proposed by Zwicky. However, since larger galaxies are closer to the center (see Figs. 7.3 and 8.1), formula (8.1) may overestimate the total virial mass by about 20–25 %.

Relativistic effects of high velocities. From the Hubble law

$$v = H_0 d \quad (8.13)$$

with $v = \bar{v}$, and from formulae (7.7) and (8.5), we may estimate that the Coma cluster is at a distance of $d \approx 100$ Mpc from us. The corresponding measured redshift $z = 0.023$ is assumed to be linearly proportional to the distance d , i.e.,

$$z = \frac{H_0}{c} d = 0.023, \quad (8.14)$$

where $c = 299\,792\,458$ m/s is the speed of light in the vacuum.¹⁰ However, the distance d above is slightly overestimated, since relativistic effects of the large speed (8.5) have to be taken into account. The linear dependence

$$v = cz$$

has to be replaced by the nonlinear relativistic relation [61]

$$v = cf(z),$$

where

$$f(z) = \frac{z(z+2)}{z^2 + 2z + 2}.$$

From this for $z = 0.023$ we get $v = cf(z) = 6820$ km/s, which is 1% smaller than the speed in (8.5). Hence by (8.13), the distance d is about 1% smaller and thus also the radius R from (8.1) should be about 1% smaller.

For any redshift $z \in (-1, \infty)$ we obviously get velocities less than c . We easily find that $f(-1) = -1$, $f(0) = 0$, $f'(0) = 1$, $f(z) < z$ for $z > 0$, and $f(z) \rightarrow 1$ for $z \rightarrow \infty$.

Furthermore, note that the classical formula accounting for the increase in wavelength λ of electromagnetic radiation

$$z = \frac{\lambda - \lambda_0}{\lambda_0} = \frac{v}{c},$$

where λ_0 is the wavelength for the case in which the source is stationary with respect to the observer, has to be replaced by the relativistic relation

$$\lambda = \lambda_0 \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}.$$

¹⁰When observed light photons left the galaxy cluster, the universe was $(z+1)$ times smaller. Therefore, it is very difficult to determine distances in an expanding universe and moreover, such a procedure is not unique. Thus in cosmology several different distances are defined (angular, comoving, Euclidean, Hubble, light-year, luminosity, Minkowski, parallax, proper motion, redshift, ... distance), see e.g. [211]. By Weinberg [293, pp. 423, 485] the proper motion distance is $(z+1)$ times smaller than the luminosity distance (see also [17] and [185] for the discussion about the luminosity function of the Coma cluster in an expanding universe).

Gravitational redshift. The above distance d is also slightly overestimated due to the gravitational redshift of the Coma cluster, which has to be subtracted from the total measured redshift. The product cz for small z (i.e. $z \ll 1$) is sometimes also called the *redshift* and it is measured in km/s. According to [43, p. 10], the total gravitational redshift of the two large central galaxies of the Coma cluster is about 61 km/s, which is cca 1 % of the speed given in (8.5). Although the redshifts of galaxies near the boundary of A1656 are about 20 km/s, this also leads to the overestimation of the distance of A1656 from us and thus also of its radius (8.2), speed (8.3), and total virial mass (8.1). Similar gravitational redshifts of galaxy clusters are given also in [35], [118], and [298].

The selflensing effect. The selflensing effect of the Coma cluster may be estimated by means of (8.2), (8.6), and the famous relativistic formula for the bending angle¹¹ (see e.g. [186], [267, pp. 91, 104, 109], [20, p. 544])

$$\phi = \frac{4GM}{c^2 R} \approx 2 \times 10^{-4} \text{ rad} \approx 0.7', \quad (8.15)$$

where $\phi = (\beta - \alpha)/2$ (see Fig. 8.3). This value represents about 1 % of 1° corresponding to the angular radius $\beta/2$ of the Coma cluster (see (8.2)). Hence, R in (8.1) should be again about 1 % smaller.

Decreasing Hubble parameter. The expansion speed of the universe, which is characterized by the Hubble parameter $H = H(t)$, depends essentially on the total mass density and the density of dark energy. We will define it in (10.3). It decreases with time (see Fig. 8.7). Therefore, formula (8.14) has to be replaced by

$$z = \frac{H(t)}{c} d = 0.023.$$

According to [211], the value $H(t)$ for $z = 0.023$ is more than 1 %

¹¹For a photon flying just above the Sun's surface, i.e. at the distance $R_\odot = 6.955 \cdot 10^8$ m, we get by (4.8) the bending angle equals $\phi = 0.852 \cdot 10^{-5}$ rad $\approx 1.75''$ which is in perfect agreement with the measured value.

larger than the present value H_0 . Hence, the distance d in (8.2) is again slightly overestimated.

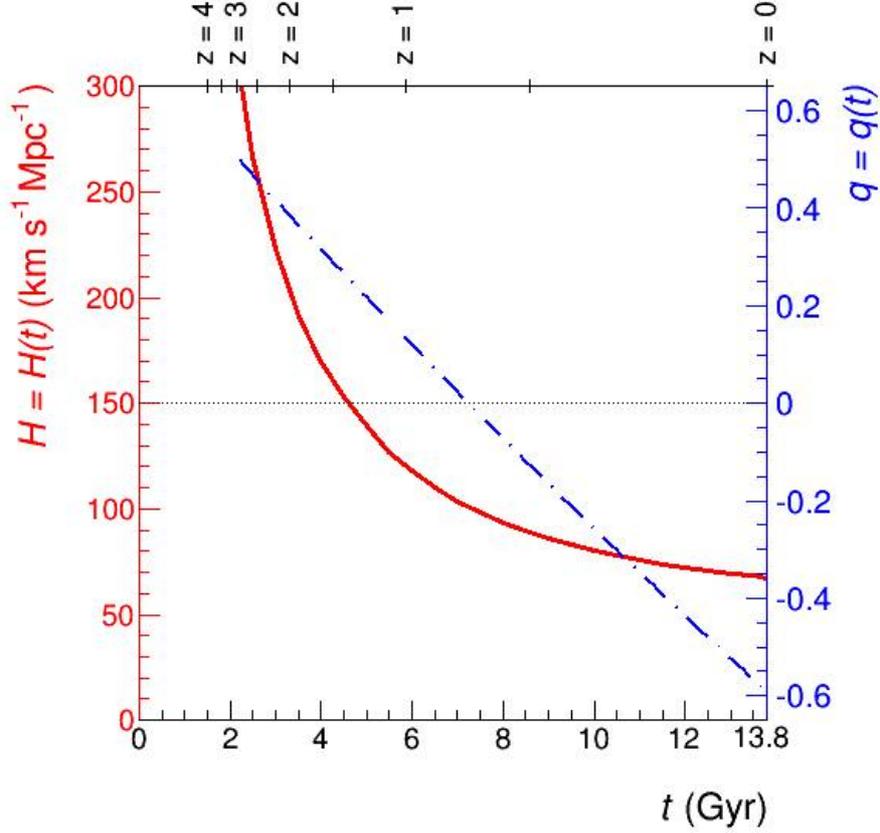


Fig. 8.7. The behavior of the Hubble parameter $H = H(t)$ is sketched by the solid line. The dashed-dotted line stands for the corresponding deceleration parameter $q = -1 - \dot{H}/H^2$ that was evaluated by means of numerical differentiation of the function $H = H(t)$. The lower horizontal axis shows time in Gyr since the Big Bang. In the upper horizontal axis we see the corresponding redshift z according to [211].

Contribution of dark energy. The effect of local expansion of the universe has a long history dating back to the 1933 paper [175] by McVittie. Such an expansion at a rate similar to the Hubble constant has been observed even on scales of the Solar system, see e.g. [64], [65], [133], [302], and of galaxies or even larger structures, see [137], [145], [221], as we shall discuss in the second half of this book. The Hubble constant by (7.7) rescaled on the radius (8.2) of the cluster

yields the expansion speed $RH_0 \approx 10^5$ m/s, which is more than 5% of the speed from (8.5). This, of course, yields a larger root-mean-square speed \bar{v} than would occur if dark energy did not act on the cluster. This seems to be another reason why Zwicky observed a large dispersion of velocities in the Coma cluster. Since the speed \bar{v} in (8.12) is squared, the contribution of dark energy could seemingly increase M by about 10% (see Table 8.1).

Table 8.1. Effects and the corresponding percentage that reduces the virial mass M , radius of the Comma cluster R , and the root-mean-square speed \bar{v} .

	Effect	M	R	\bar{v}
1	Nonuniformity of galaxy distribution	20–25	0	0
2	Relativistic effects of high velocities	3	1	1
3	Gravitational redshift	3	1	1
4	The selflensing effect	1	1	0
5	Decreasing Hubble parameter	3	1	1
6	Contribution of dark energy	10	0	5

Reduction of the root-mean-square speed. Now, let us introduce one more quadratically nonlinear effect which has a nonnegligible influence on the total mass estimate [138]. Above we saw that the mean recession speeds \bar{v} and \bar{v}_i were overestimated by several percent. If it were, say 8%, then the square \bar{v}^2 defined by formula (7.11) would be overestimated by about $100(1 - 0.92^2)\% \approx 15\%$. This essentially reduces the estimated mass (8.6) with respect to the virial mass (8.1) or (8.12).

In summary, the seven effects analyzed above can essentially reduce the total mass (8.1) obtained from the Virial theorem by a factor of about two (cf. (8.6)). Hence, the total mass of the cluster could be at most five times larger than its luminous mass.



8.4. What is the mass of dark matter inside the Coma cluster?

According to [114], the possible distribution of dark matter in the Coma cluster cannot be significantly different from that of the galaxies inside the cluster. Thus, the distribution of dark matter approximately follows the distribution of galaxies.

Finally, we shall present a “back of the envelope” calculation illustrating whether it is necessary to assume some extragalactic dark matter in the center of the Coma cluster.

For simplicity assume that the two supergiant elliptic galaxies NGC 4889 and NGC 4874 (see Fig. 7.3) have the same mass m defined in (8.7) and that they orbit along a circular trajectory with center O , radius r , and velocity v . If one of these two galaxies were to be smaller, it would orbit the larger one by a higher velocity and along a longer path. Then it would absorb more additional galaxies than the larger one. By this mechanism the masses of both galaxies are balanced.

Since the gravitational potential inside a homogeneous spherical layer is constant, (see Newton’s second theorem 4.2), external galaxies and possible dark matter outside the sphere with center O and radius r have almost no influence on this motion. From Newton’s laws and the relation for centripetal force we get

$$\frac{Gm^2}{4r^2} = \frac{mv^2}{r}. \quad (8.16)$$

The distance of both galaxies on the celestial sphere is $8.15'$ which in projection on the distance 100 Mpc gives 7.32×10^{21} m. Thus for the radius r we have

$$r \geq 3.66 \times 10^{21} \text{ m}. \quad (8.17)$$

According to [1, p. 19], the measured radial velocities of both supergiant galaxies are 6472 km/s and 7189 km/s. Their average velocity $\tilde{v} = 6830.5$ km/s nicely corresponds to the mean recession speed of the whole cluster (8.5). For the radial velocity v_{radial} with respect to \tilde{v} we get by (8.7), (8.16), and (8.17)

$$\begin{aligned}
3.585 \times 10^5 &= \frac{7\,189\,000 - 6\,472\,000}{2} = v_{\text{radial}} \leq v = \sqrt{\frac{Gm}{4r}} \\
&\leq \sqrt{\frac{6.673 \times 10^{-11} \times 2 \times 10^{43}}{4 \times 3.66 \times 10^{21}}} = 3.02 \times 10^5 \text{ (m/s)}.
\end{aligned}
\tag{8.18}$$

Comparing the left-hand and the right-hand sides, we find a small discrepancy.¹² This simplified example shows that Newtonian mechanics does not describe reality correctly or the masses or radial velocities of both the elliptic galaxies are wrongly estimated or we have to assume the existence of invisible matter between galaxies, even though it does not seem that there should be 10 times more invisible matter than visible.

Moreover, considering the gravitational influence of other matter (small galaxies and a large amount of solitary stars [282]) that are inside the sphere with center O and radius r , the right-hand side of (8.18) should be larger. Also the lower bound in (8.17) is smaller, since the cluster magnifies the angular distances due to the abovementioned effects. This is another reason why the right-hand side of (8.18) should be larger.

By Newton's second theorem 4.2, the velocity of the two giant galaxies is mainly influenced by matter situated inside the ball of radius r . As was already mentioned, inside galaxy clusters there is at least five times more baryonic matter in the form of hot gas emitting X-rays than baryonic matter contained in galaxies (see [2], [27], [106], [290]). Zwicky's paradox of large observed velocities vanishes, since it may have a quite natural explanation without dark matter.



¹²By older data [24] from 1995 the radial velocities of both the galaxies are 6505 km/s and 7108 km/s, which yields on the left-hand side of (8.18) a smaller value 3.015×10^5 m/s.

9. Vera Rubin and rotational curves of spiral galaxies

The mysterious dark matter is only a modeling error.

AUTHORS

9.1. Vera Rubin

Vera Rubin born Cooper has devoted her entire scientific career to the promotion of revolutionary ideas in astrophysics. Her thoughts significantly affected the development of contemporary astronomy. Her father came from Lithuania and her mother from Moldova. In her dissertation at Cornell University she looked at the fundamental question of whether the universe could rotate as a whole.¹ The supervisor of her doctoral dissertation was George Gamow, who published the article [83] on this topic. Rubin was the first woman who used devices on Mt. Palomar Observatory in California. Her life is described in detail e.g. in [199].



Fig. 9.1. Vera Rubin born Cooper (b. 1928), photo Robert Rubin

¹Also Kurt Gödel worked on a theory of a rotating universe [88]. Note that for any odd dimension $n = 1, 3, 5, \dots$ the sphere \mathbb{S}^n may rotate around its center such that all points have the same speed (because \mathbb{S}^n can be “combed” if n is odd). It is not a rotation around an axis as e.g. for $n = 2$. For $n = 3$ the corresponding centrifugal force could then contribute to the expansion of the universe (as in case of a rotating circle for $n = 1$).

At the end of the seventies of the last century, Rubin claimed that spiral galaxies do not have enough mass to explain their rapid rotation. But she did not believe that the universe should contain some dark matter, although her measurements suggested that. She said about this discrepancy [36]:

If I could have my pick, I would like to learn that Newton's laws must be modified in order to correctly describe gravitational interaction at large distances. That's more appealing than a universe filled with a new kind of sub-nuclear particle.



9.2. Spiral galaxies do not rotate according to Kepler's laws

Vera Rubin's greatest discovery was the fact that spiral galaxies have “flat” rotational curves (see [239]). On that basis, in the 70's of the last century she developed her own theory of rotational curves of galaxies. From the high orbital speed of stars she concluded that galaxies should contain much more nonluminous than luminous matter to be kept together by gravity — see e.g. her review articles [237] and [238] on dark matter.

Now let us look more closely at her hypothesis. Consider a test particle of mass m (typically this will be a star) and let $M \gg m$ be the mass of some body generating the central force field. Assume that the test particle revolves about the center along a circular orbit with radius r and speed v . Then from Newton's law of gravitation and the relation for centripetal force Rubin [237] easily obtained that

$$G \frac{Mm}{r^2} = \frac{mv^2}{r}, \quad \text{i.e.,} \quad v = \sqrt{\frac{GM}{r}}. \quad (9.1)$$

The velocity v of a particle on a circular orbit is thus proportional to $r^{-1/2}$. Such orbits are called *Keplerian* (see Fig. 9.2).

Vera Rubin states (see [239, p. 491]) that *the stellar curve does not decrease as is expected for Keplerian orbits.*

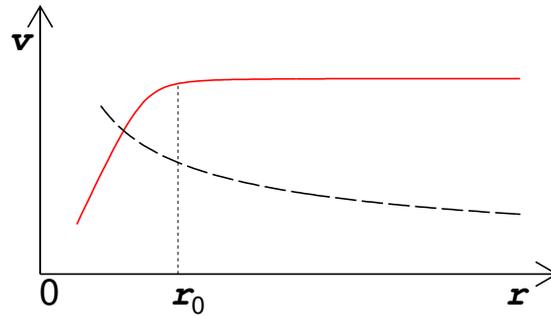


Fig. 9.2. The dashed line shows the rate of decrease of velocities for Keplerian orbits that depend on the distance r from the center of a spiral galaxy. The solid line shows an idealized rotational curve whose shape was derived by Vera Rubin by means of a variety of measurements.

To explain this discrepancy, it is important to realize that spiral galaxies do not have a central force field except within a close neighborhood of the center, where e.g. in the Milky Way stars S1, S2, ... orbit the central black hole according to Kepler's laws with velocities up to 7 000 km/s (see (4.20)). By (4.21) the mass of this black hole is roughly 3.5 million solar masses, which is only 0.01 % of the total mass of our Galaxy (cf. (9.5)). In the Solar system, on the contrary, 99.85 % of the mass is concentrated at the Sun. The planets barely interact gravitationally among themselves and their movements are determined mainly by the central force of the Sun. On the other hand, trajectories of stars in a galactic disk are substantially influenced mainly by neighboring stars, because the central bulge contains only about 10 % of all mass of a galaxy (cf. Fig. 9.5).

In Remark 9.1 we outline why the force of a disk-shaped galaxy acting on a test particle is much larger than it would be if its whole mass were to be concentrated at one central point (this will be elaborated in detail in Section 9.4). Therefore, the speed v of stars on circular orbits in a spiral galaxy should be higher than for Keplerian orbits (see Fig. 9.2).

In nearby spiral galaxies, Vera Rubin [238] (see also [240, p. 480]) found that all stars of these galaxies move at almost the same constant

speed² of order $v \approx 200$ km/s for $r > r_0$, where $r_0 > 0$ roughly corresponds to the radius of the central bulge and is typically equal to a few light years (see Figs. 9.2 and 9.3). On the other hand, she observed that for $r \leq r_0$ the inside of the spiral galaxy (including a possible bar) rotates with roughly constant angular speed in a manner like that of a DVD record, i.e., the speeds of these stars are approximately linearly proportional to their distance from the center (see Fig. 9.2). An exception is a close neighborhood of the central black hole (cf. (4.20)).



Fig. 9.3. Large spiral galaxy M31 Andromeda occupies an area in the celestial sphere six times larger than the full Moon. It has a visible central bulge.

The average thickness of the disk (except for the bulge) of spiral galaxies varies from 300 pc to 1 kpc. It is therefore about 30 to 100 times thinner than the diameter of the visible part of the galaxy. This is easily seen when galaxies are observed edge on (see Fig. 9.5). Moreover, the gas and dust are mainly found close to the plane of symmetry of the disk. Consequently, in Section 9.4 we will treat the disk just as a two-dimensional body, which is obviously a better approximation than a central mass point. The gravitational field of spiral galaxies will therefore be approximated by the gravitational field of a flat disk with rotationally symmetric mass density distribution.

²In the Solar system a similar phenomenon would correspond to Mercury orbiting the Sun at the same speed as Neptune.

Remark 9.1 Equation on the right of (9.1) provides only a rough estimate for expressing orbital velocities of stars in a spiral galaxy. Let us therefore show now that a test particle (star) orbiting a ball of radius r with arbitrary spherically symmetric mass density distribution (cf. Newton's first theorem 4.1) has a lower speed than when orbiting a disk of the same radius r and the same mass. In doing so, we will consider a special distribution of the density of the disk, which arises as projection of the original ball perpendicularly to the horizontal xy plane of the disk.

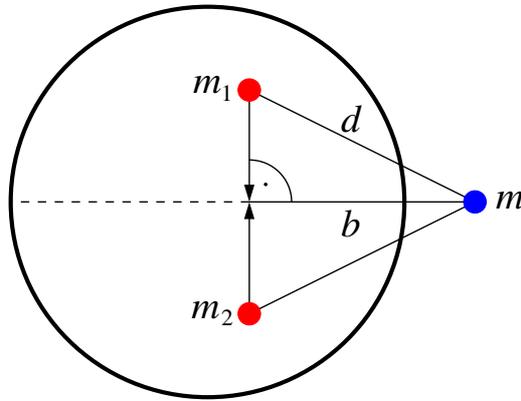


Fig. 9.4. A ball with symmetrically distributed mass with respect to the horizontal plane acts on a test particle by a smaller force than the mass projected perpendicularly to the horizontal plane of the disk — dashed.

To be convinced of this assertion, just consider two arbitrary mass points with masses $m_1 = m_2$ located inside a ball placed symmetrically with respect to the horizontal xy plane (see Fig. 9.4). Then the total force F of both mass points acting on the test particle of mass m , will be less than the force \overline{F} of both mass points projected perpendicularly to the disk and acting on m . Let d be the distance between m_1 and m . Denoting by b its orthogonal projection on the horizontal xy plane, we find that

$$F = G \frac{2m_1 m}{d^2} \cdot \frac{b}{d} \quad \text{and} \quad \overline{F} = G \frac{2m_1 m}{b^2}.$$

Thus we see that the ratio of forces \overline{F} and F is equal to the third power of the fraction d/b

$$\bar{F} = \left(\frac{d}{b}\right)^3 F \geq F. \quad (9.2)$$

By (9.1) this cubic nonlinearity causes a greater attractive gravitational force by the disk than by the ball, and thus also a higher orbital speed around the disk.³



9.3. Orbital velocity around a spherically symmetric body

In this section we introduce a rough conservative estimate for the orbital velocities of stars for the case in which all baryonic matter (i.e. mainly protons and neutrons) of the Milky Way is replaced by a ball with spherically symmetric mass density distribution. In the next section we will focus on a flat disk with arbitrary rotationally symmetric mass density distribution.

The radius of the visible part of the disk of our Galaxy is estimated by

$$r_G = 16 \text{ kpc} = 4.938 \cdot 10^{20} \text{ m}. \quad (9.3)$$

Our Sun orbits the center of the Milky Way with the speed⁴

$$v_\odot = 230 \text{ km/s} \quad (9.4)$$

on a path of radius $r_\odot = 8.3 \text{ kpc}$, i.e. it is found about halfway out from the center of the Galaxy, where the density of stars is relatively small. Stars orbiting the center of our Galaxy at a distance $r > r_0 \approx 3 \text{ kpc}$ should have a speed similar to v_\odot due to the expected flat rotational curve (see Fig. 9.2).

Denote by $\mathcal{M}(r)$ the *mass of baryonic matter* within the ball of radius r and center in the middle of our Galaxy. To estimate $\mathcal{M}(r_G)$ for r_G given by (9.3) we will use the distribution of stars from Table 9.1 (see e.g. [177, p. 394]). It is based on Hipparcos' data taken from our close neighborhood up to a distance of several hundreds parsecs.

³An analytical expression of the gravitational influence of the entire disk on an outer test particle leads to elliptic integrals (see [23, p. 73]).

⁴Most sources give the speed of the Sun v_\odot as being in the range of 220 to 240 km/s.

Table 9.1. Distribution of stars in our Galaxy according to their spectral classes.⁵ The second line shows the corresponding mass of a typical star in units of the solar mass M_{\odot} . The third line indicates the number of stars of a particular spectral class⁶ divided by 10^9 . The last line presents the calculated mass of all the stars in a particular spectral class in billions of solar masses. The last column corresponds to white dwarfs (WD) which belong the luminosity class V.

Spectral class	O	B	A	F	G	K	M	WD
Mass in M_{\odot}	25	5	1.7	1.2	0.9	0.5	0.25	0.7
Number in billions	10^{-5}	0.3	3	12	26	52	270	35
Product	≈ 0	1.5	5.1	14.4	23.4	26	67.5	24.5

From the third line of Table 9.1, we see that our Galaxy contains approximately 400 billion stars. While at the end of the last century it was thought that red dwarfs of spectral class M form only 3% of the total number of stars (see [22, p. 93]), at present it is estimated from Table 9.1 that they are in the vast majority — about 70%. To support this statement it should be noted that among the 20 nearest stars to our Sun, 13 red dwarfs are currently known. Note that the mass of a red dwarf ranges from $0.08M_{\odot}$ to $0.45M_{\odot}$. From the last line of Table 9.1 it is evident that the spectral class M contributes the most to the total mass of our Galaxy of all the spectral classes. Rubin, of course, could not know about the existence of so many red dwarfs in this smallest weight category. This growth is due to the continual improvements of the sensitivity of space telescopes. In this way, the estimated mass of the baryonic matter in our Galaxy has considerably increased. Summing up the numbers in the last row of Table 9.1, we get

$$\mathcal{M}(r_G) \geq 162.4 \cdot 10^9 M_{\odot} = 3.25 \cdot 10^{41} \text{ kg}.$$

⁵The Harvard Spectral Classification (en.wikipedia.org/wiki/Stellar_classification) shows a similar relative representation of stars that will be further improved by data from the Gaia satellite. Gaia is able to look at the center of our Galaxy and in the opposite direction also at its boundary. However, the accuracy of measurements depends essentially on the magnitude and extinction.

⁶A mnemonic to remember says: *Oh, Be A Fine Girl, Kiss Me*. The spectral class K contains the so-called *orange dwarfs*.

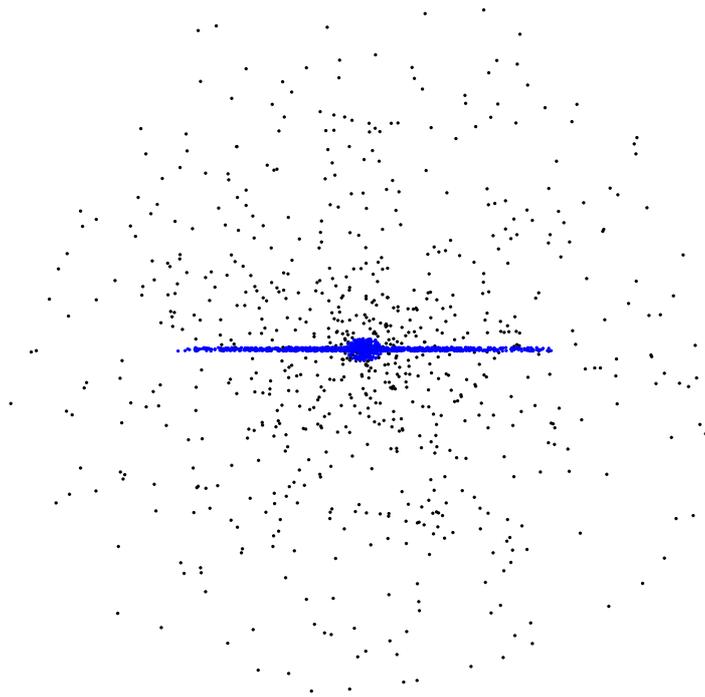


Fig. 9.5. A schematic view of a spiral galaxy as seen from the side. A spherical central bulge is surrounded by a flat disk and a sparse spherically symmetric halo filled mainly by neutral hydrogen and helium, old stars, and globular clusters.

Unfortunately, we cannot so far reliably determine the contribution to $M(r_G)$ from black holes, neutron stars⁷, infrared dwarfs⁸, exoplanets, etc., whose luminosity is small. According to [177, p.393], the mass of the baryonic matter of all the stars in the Galaxy is about

$$175 \cdot 10^9 M_\odot = 3.5 \cdot 10^{41} \text{ kg},$$

including further stars of the luminosity classes I–IV (i.e. supergiants, giants, and subgiants, cf. Fig. 11.1). The disk and bulge contains also a large amount of non-luminous baryonic matter in the form of dust, gas, and plasma. In [177, p.353], the amount of interstellar matter

⁷The amount of stars in the left part of Table 9.1 is so small, because they live very briefly. On the other hand, there may exist many superdense compact remnants left by these stars in the Galaxy.

⁸Three new spectral classes for small cold stars were recently introduced: L (red-brown dwarfs), T (brown dwarfs), and Y (black dwarfs). For instance, in 2013 Kevin Luhman discovered a pair brown dwarfs only 6.5 ly from the Sun. Another brown dwarf WISE J085510.83-071442.5 is located 7.2 ly from us.

(without hypothetical dark matter) is estimated at about 10% of the total mass of the Milky Way's stars. Sparse non-luminous baryonic matter is also spread at a spherical galactic halo (see Fig. 9.5), as can be determined from radio waves of 21 cm, corresponding to spin flip in the hydrogen atom (see [240, p. 485]). Therefore, for the total mass of baryonic matter inside the considered ball of radius r_G we have a lower estimate⁹

$$\mathcal{M}(r_G) \geq 3.85 \cdot 10^{41} \text{ kg.} \quad (9.5)$$

According to [177], the mass density distribution $\rho = \rho(r)$ beyond the visible edge decreases faster than r^{-2} ; otherwise the integral $\int_{r_G}^{\infty} \rho(r) 4\pi r^2 dr$ would diverge. However, the Shell theorem 4.2 indicates that this matter (including possible dark matter) has no effect on the movement of stars, if the mass distribution is spherically symmetric. Let us concentrate all baryonic matter inside the ball of radius r_G to one central point. Then from relations (9.1), (9.3), and (9.5) we find that the orbital velocity of stars on the radius r_G of the visible disk is

$$v = \sqrt{\frac{GM(r_G)}{r_G}} \geq \sqrt{\frac{6.674 \cdot 10^{-11} \cdot 3.85 \cdot 10^{41}}{4.938 \cdot 10^{20}}} = 228 \cdot 10^3 \text{ m/s,} \quad (9.6)$$

This value is indeed comparable to the measured speed (9.4). Although relation (9.6) is only approximate, to postulate the existence of 5–6 times more dark matter than baryonic matter (see e.g. [29], [212], and Fig. 9.9) to hold the Galaxy together by gravity seems to be somewhat overestimated.



9.4. Orbital velocity around a flat disk

Of course, one can raise the objection that relationship (9.6) was derived just in the case of a central force for a given mass point (that is by Theorem 4.1 equivalent to a ball with a spherically symmetric

⁹By astronomical tables [156, p. 127], the total mass of the Galaxy is $M_G = 10^{12} M_{\odot} = 2 \cdot 10^{42} \text{ kg}$. Another source [109] even reports a three times greater value amounting to 200 kpc from the center.

mass density distribution) which may lead to a large modeling error. In this section we will therefore approximate the gravitational field of a spiral galaxy by the gravitational field of a flat disk with rotationally symmetric mass density distribution.

Theorem 9.1. *A particle orbiting a central mass point along a circular trajectory of radius R has a smaller speed than if it were to orbit a flat disk of radius R and the same mass with an arbitrary rotationally symmetric density distribution.*

P r o o f . A greater attractive force has to be balanced by a larger orbital speed if the testing particle should stay on a circular trajectory. Therefore, we only need to compare the force of the central mass point with the force of a disk of the same mass. Under the assumptions of Theorem 9.1 the areal density of the disk $\rho = \rho(r) \geq 0$ depends only on the distance from its center. First, we will investigate the gravitational influence of a fixed one-dimensional homogeneous ring of radius $r \in (0, R)$ on a test particle of mass m , whose distance from the center is R . The total mass of the ring equals $M = 2\pi r\rho$, where ρ is the length density (i.e. 1D mass density). Concentrating the mass of the ring to its center, the corresponding force acting on a test particle is equal to

$$\overline{F} = G \frac{2\pi r \rho m}{R^2}. \quad (9.7)$$

Our goal will be to show that \overline{F} is smaller than the force of the ring acting on the test particle. The statement of Theorem 9.1 will then follow by integration along r .

In polar coordinates (r, φ) , consider two equal length elements of the ring

$$dl = r d\varphi \quad (9.8)$$

located symmetrically with respect to the horizontal axis at a distance s from the test particle as shown in Fig. 9.6. Then according to the law of cosines, we have

$$s^2 = r^2 + R^2 - 2rR \cos \varphi \quad (9.9)$$

and the force with which this pair acts on the test particle equals

$$dF = G \frac{2dl \rho m}{s^2} \cos \alpha. \quad (9.10)$$

From the law of sines $r \sin \varphi = s \sin \alpha$ it follows that

$$\cos \alpha = \sqrt{1 - \sin^2 \alpha} = \frac{1}{s} \sqrt{s^2 - r^2 \sin^2 \varphi}. \quad (9.11)$$

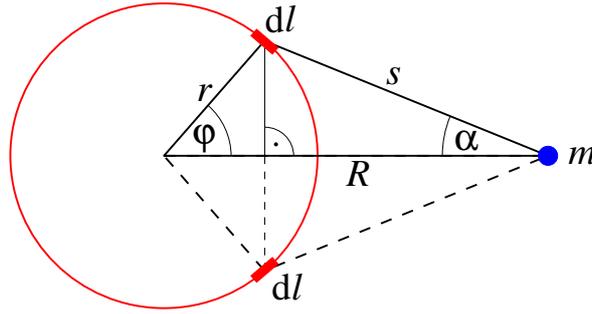


Fig. 9.6. A homogeneous ring acts gravitationally on an outer particle by a larger force than if the total mass were to be concentrated to the center of the ring.

Without loss of generality we may further assume that the gravitational constant $G = 1$, $R = 1$, $m = 1$, and that the length density of the ring is $\rho = 1$. Then for $r \in (0, 1)$ and $\varphi \in [0, \pi]$, by substituting (9.8), (9.9), and (9.11) into (9.10), we get

$$\begin{aligned} dF &= \frac{2 dl}{s^2} \frac{1}{s} \sqrt{r^2 + 1 - 2r \cos \varphi - r^2 \sin^2 \varphi} = \frac{2r d\varphi}{s^3} \sqrt{(1 - r \cos \varphi)^2} \\ &= 2r \frac{1 - r \cos \varphi}{(r^2 + 1 - 2r \cos \varphi)^{3/2}} d\varphi, \end{aligned}$$

because $1 > r \cos \varphi$. Thus the total gravitational force of the ring of radius r that acts on the test particle is

$$F(r) = 2r \int_0^\pi f(r, \varphi) d\varphi = 2r \int_0^\pi \frac{1 - r \cos \varphi}{(r^2 + 1 - 2r \cos \varphi)^{3/2}} d\varphi, \quad (9.12)$$

where for a fixed $r \in (0, 1)$ the integrated function

$$\varphi \mapsto f(r, \varphi) = \frac{1 - r \cos \varphi}{(r^2 + 1 - 2r \cos \varphi)^{3/2}}$$

is positive, continuous, and decreasing. Since the values at the end-points $f(r, 0) = (1 - r)^{-2}$ and $f(r, \pi) = (1 + r)^{-2}$ are finite numbers, the integral in (9.12) is finite (see Fig. 9.7).

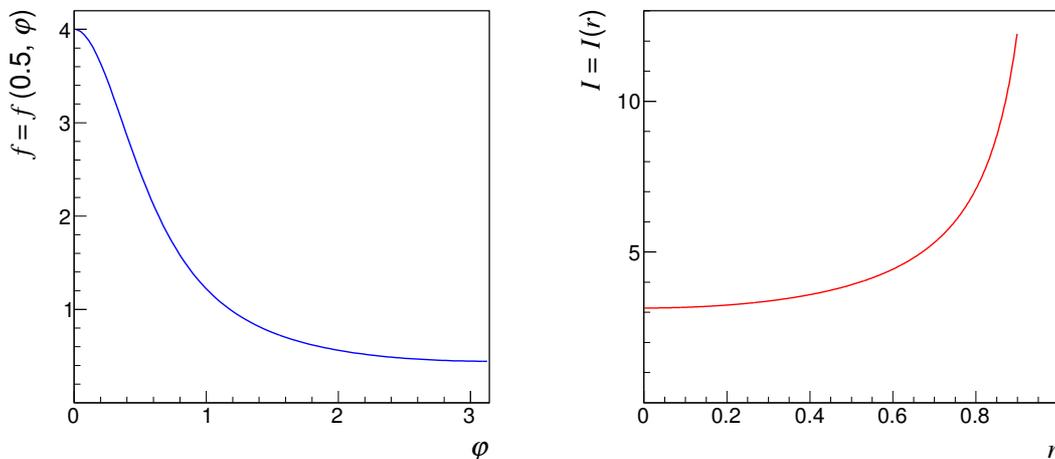


Fig. 9.7. On the left there is a graph of the integrated function (9.13) for $r = 0.5$ on the interval $[0, \pi]$. On the right there are numerically calculated values of the integral $I(r)$ for $r \in [0, 1]$.

The integral

$$I(r) = \int_0^\pi \frac{1 - r \cos \varphi}{(r^2 + 1 - 2r \cos \varphi)^{3/2}} d\varphi \quad (9.13)$$

appearing in equation (9.12) unfortunately has no known analytical expression for $r \in (0, 1)$. However, we can find that $I = I(r)$ is an increasing function¹⁰ and may analytically evaluate its limits. For $r = 0$ we see that the integrated function is equal to one, and thus (see Fig. 9.7)

$$I(0) = \pi. \quad (9.14)$$

Consider now the point $r = 1$. By the Taylor expansion we get

$$\cos \varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \dots \geq 1 - \frac{\varphi^2}{2}.$$

Hence,

$$\varphi^2 \geq 2 - 2 \cos \varphi, \quad (9.15)$$

¹⁰The function I is even strictly convex and $\dot{I}(0) = 0$.

and therefore we obtain (see Fig. 9.7)

$$2I(1) = \int_0^\pi \frac{2 - 2 \cos \varphi}{(2 - 2 \cos \varphi)^{3/2}} d\varphi = \int_0^\pi \frac{d\varphi}{\sqrt{2 - 2 \cos \varphi}} \geq \int_0^\pi \frac{d\varphi}{\varphi} = \infty, \quad (9.16)$$

that is

$$I(1) = \infty. \quad (9.17)$$

A similar trick with the lower estimate can be used to replace the cosine function in (9.13) by quadratic polynomials in φ , which can already be calculated analytically and leads to the required inequality

$$F(r) = 2rI(r) > \bar{F} = 2rI(0) \quad \text{for } r \in (0, 1], \quad (9.18)$$

where the forces are defined in (9.12) and (9.7). \square



9.5. Orbital velocity around galaxy bulges and halo

The resulting force acting on a galaxy is approximately the sum of the gravitational effects of the bulge, flat disk, and halo for $r \leq r_G$ if the outside of the galaxy is spherically symmetric (see Theorem 4.2). The bulge of spiral galaxies is usually spherically symmetric. For instance, the neighboring Andromeda galaxy M31 in Fig. 9.3 has a clear bulge making up to 20–25 % of its radius.¹¹ By Newton's first theorem 4.1, the gravitational influence of the spherical bulge on outer stars can be approximated by the central force of a mass point, into which the whole mass of the bulge is concentrated. By Newton's second theorem 4.2 we may neglect the gravitational influence of the halo $r > r_G$.

Theorem 9.1 can be modified to a ring with inner radius 20–25 % of R and outer radius R , since the mass density function $\rho = \rho(r) \geq 0$

¹¹The observed orbital velocity of stars which are not close to the center of M31 is again about 230 km/s according to Vera Rubin (see [238, p. 7]). The radius of M31 is $r_A \approx 2r_G$ and the total mass is estimated to be $M_A \approx 3M_G$. Then by (9.6) we even get a bigger disagreement with the postulation of dark matter in M31 than for our Galaxy. Namely, the calculated velocity by (9.6) will be $\sqrt{3/2}$ times larger than the observed orbital velocity. In this case, not a single gram of dark matter is needed, since we even have surplus of luminous matter.

is arbitrary. The gravitational force of the ring on the test particle on the outer edge of the ring is again larger than when the total mass is concentrated at the center. The velocity of stars at the edge of the Galaxy at distance R from the center is therefore larger than in (9.6).

Remark 9.2. The fact that rotational curves of spiral galaxies are flat does not imply that there must be some dark matter concentrated around galaxies. From Theorem 4.1 it follows that the gravitational force between a two-dimensional homogeneous sphere (shell) and a mass point lying on it is finite. On the other hand, the force of a one-dimensional ring acting on a fixed point, which lies on it, is infinite¹² according to (9.12)–(9.17), since the function $f = f(r, \varphi)$ from (9.12) has a singularity for $r \rightarrow 1$ and $\varphi \rightarrow 0$. Thus we see that there is a substantial difference between the two-dimensional and three-dimensional model. From the two-dimensional model it is also evident why the stars at the edge of the spiral galaxy orbit so rapidly. Therefore, we should not exchange a gravitational field of a galaxy with the gravitational field of a central mass point. The discrepancy in large velocities of stars observed by Rubin thus may have a natural explanation.

Remark 9.3. Spiral galaxies, of course, do not have exactly rotationally symmetric distribution of mass. For instance, the shape of a barred spiral galaxy of type Sbc resembles an open letter S. It is noteworthy that the stars of these spiral galaxies are measured to move at almost constant speed (see Rubin [238, p. 7]), but these galaxies are not winding up and do not show an expected tightening of arms even though they have already gone through many revolutions. With difficulty, it can be assumed that the arms of type Sbc galaxies are formed by some kind of density waves as suggested in [22, p. 544].



¹²However, if the ring had a positive constant thickness, then the force on a mass point would be finite.

9.6. Arguments against dark matter

In previous chapters (as well as in [138]) we introduced several arguments showing that the amount of dark matter seems to be considerably overestimated. It is very probable that Newton's law of gravity on large cosmological scales approximates reality only very roughly, since it assumes an infinite speed of gravitational interaction. Therefore, we should not accept results of various Newtonian numerical simulations (like, for example, the Millennium simulation) which seek to prove that without dark matter galaxies could not form after the Big Bang.

Several modifications of Newtonian theory, e.g., MOND (Modified Newtonian Dynamics) [173], [180], [249] and its relativistic generalization Teves (Tensor-Vector-Scalar) [18] are at present being developed and studied. Effects that are attributed to dark matter are explained by a different form of the gravitational law (see also [8]). Furthermore, Pavel Kroupa in [150] state altogether several arguments that point to the absence of dark matter around our Galaxy. A number of other papers ([13], [76], [77], [82], [111], [151], [187], [192], and [257]) also confirm that on scales of galactic disks, Newton's theory of gravitation is still a fairly good approximation of reality and it is not necessary to modify it, or to assume the existence of dark matter.

The observed oscillations of stars perpendicularly to the galactic plane can be explained by classical Newtonian mechanics without dark matter (see [187]). In other words, dark matter may be referred to as a modeling error resulting from an incorrect model and misinterpretation of measured data on large scales.

The influence of dark matter in the Solar system has not been observed [187], even though our Sun is a large gravitational attractor. Thus it seems that dark matter, if it exists, is not able to dissipate its inner energy, and therefore cannot be concentrated in the Sun's neighborhood.



Fig. 9.8. The collision of two galaxy clusters MACS J0025.4-1222. Hypothetical dark matter is artificially colored on the left and right. The central region produces X-rays due to the collision of gas from both the clusters [52].

On the other hand, Douglas Clowe in his paper [52]: *A direct empirical proof of the existence of dark matter* proposes an example of the collision of two galaxy clusters, where the intergalactic gas is stopped, while the galaxies continue in an unchanged direction together with dark matter which is “detected” by gravitational lensing. However, we are not able to measure tangential components of the velocities of these clusters to prove that the collision really happened. Moreover, there are several strange circumstances:

1. The clusters MACS J0025.4-1222 from [52] have almost the same size and they lie on one line together with clouds of dark matter (see also the Bullet cluster or the Musketball cluster). This is very unlikely from a statistical point of view. The clusters should have different sizes and their positions together with gas should not lie on one line, since their initial velocities were not in one line, in general.

2. Due to the large density of galaxies, tidal tails (cf. Fig. 5.7) and the effect of dynamical friction should be observed among galaxies, but it is not.

3. The proposed infall velocity $v \approx 3000\text{--}4500$ km/s for this collision is at least 1% of the speed of light and has the opposite sign to the overall expansion speed of the universe. How could these two

galaxy clusters get such unlikely large velocities and thus also kinetic energy proportional to v^2 in an isotropic and homogeneous universe, where the local peculiar speed of galaxies is usually only several hundreds km/s.

4. The regions with dark matter are artificially colored (see Fig. 9.8) on the basis of some numerical, not exactly explained simulations based on gravitational lensing.

Now we shall investigate another argument. The Fourier analysis of fluctuations of the cosmic microwave background (CMB) radiation, detected by the Planck satellite (see Fig. 18.4 and [212]), was interpreted so that our universe is composed of 27 % dark matter, 5 % baryonic matter (of which less than 1 % is luminous) and the rest is dark energy (see Fig. 9.9). However, as we shall see in Section 19.3, these values are based on incorrect extrapolations. Thus it seems that ratio 27 : 5 between the amount of dark matter and baryonic matter is again greatly exaggerated.

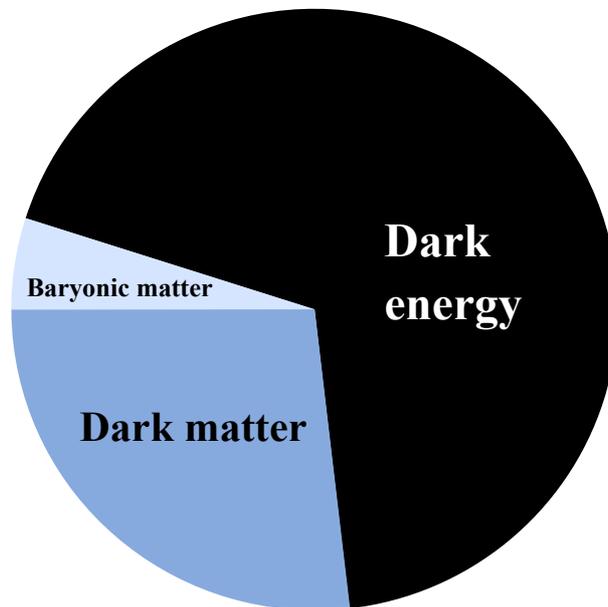


Fig. 9.9. According to the scientific results of the Planck satellite [213], our universe is composed of 68 % dark energy, 27 % dark matter, and about 5 % baryonic matter. However, as presented in Chapter 19, these data were obtained from the Friedmann cosmological model which is based on excessive extrapolations.

Moreover, for more than 13 Gyr the CMB has been biased by gravitational lensing of billions of galaxies and their clusters. This phenomenon is called *weak lensing*. However, it is weak only for a relatively small redshift z , say $z < 1$. On the other hand, we will illustrate in the next example that it is relatively strong for $z \gg 1$. A large amount of noise in the CMB was thus produced especially from the distribution of protogalaxies and their clusters when our universe was young.

Example 9.1. According to [73], the redshift of the CMB is $z = 1089$. Assume now that some relic photon was gravitationally bent at the distance $z = 9$ about the angle $\phi = 1'$, which by (8.15) is a quite acceptable value. Moreover, at that time the universe was $(z + 1)^3 = 1000$ times more dense than it is at present which makes bending effects even larger. According to [211], a photon with $z = 9$ has travelled more than 13 Gyr. If the universe had not expanded then the distance d of the photon from its original straight trajectory would satisfy

$$d > \tan \phi \cdot c \cdot 13 \cdot 10^9 \text{ yr} = 3.78 \cdot 10^6 \text{ ly}$$

as indicated in Fig. 9.10. Nevertheless, because the universe expanded $z + 1 = 10$ times in each direction (see [293, p. 453]), the relic photon will have deviated more than 37.8 million light years from its original straight trajectory. For $z > 9$ such a magnification is, of course, larger. Consequently, the larger z is, the larger is the smearing.



Fig. 9.10. Bent trajectory of a photon in a nonexpanding universe.

The most typical diameter of fluctuations in the CMB is about one angular degree [212]. From (2.14) we find that the area of the sky with such a typical circular fluctuation contains on average $\frac{\pi}{4} \cdot 10^{12} / 41\,253 \approx 19 \cdot 10^6$ galaxies. Many of them have $z > 9$ and thus their gravity produces a considerable smearing in the CMB radiation. This fact is not taken into account in [214].

Thus a natural question arises whether fluctuations in CMB are mostly random noise due to gravitational lensing – see [214]. Note that relict photons that travelled along more curved trajectories have slightly larger wavelength than those with straighter trajectories. Another source of bias is the inverse Compton scattering effect of the CMB, called the *Sunyaev–Zeldovich effect* (see [215], [273]).

Nowadays there is a large discussion on what dark matter is. The discrepancy of some model with reality does not mean that dark matter really exists, since the model can be wrong. Therefore, direct proofs of the existence of dark matter are searched for. For this purpose many sophisticated detectors (CDMS, DAMA/LIBRA, ADMX, ...) were constructed, but for the time being no particle of dark matter has been detected. Also the Large Hadron Collider in CERN has not found any signs of new physics that could explain dark matter.

In the middle ages, the geocentric model of our universe was abandoned,¹³ since it could not explain all observed phenomena. Later the heliocentric model was as rejected. However, it seems that the current cosmological model involving dark matter and dark energy (cf. Fig. 9.9) possesses a lot of discrepancies and paradoxes. Thus it is probably also incorrect as we shall learn in Chapter 19.



¹³Note that the Earth is in the center of the observable universe.

Part 2

Antigravity and Dark Energy

10. The accelerating expansion of the universe

*There should be doubt about everything.
(De omnibus dubitandum est.)*

ARISTOTLE

10.1. The 2011 Nobel Prize for Physics

Cosmology is a branch of physics dealing with the largest spatial distances and time intervals and issues concerning the formation and evolution of the universe as a whole. In 2006, John C. Mather and George F. Smoot won the Nobel Prize for Physics for cosmology for demonstrating the Planck spectrum and diminutive anisotropy of the cosmic microwave background radiation¹ by means of the satellite COBE (cf. Fig. 18.4 obtained by the satellite Planck).



Fig. 10.1. Saul Perlmutter, Adam Riess, Brian Schmidt

¹This radiation comes from a time period called *recombination*. As a result of temperature decrease to about 3000 K, free ions and electrons combined into atoms and the universe became transparent to photons. At present, the temperature of the cosmic microwave background corresponds to the black body radiation at a temperature of 2.73 K. Note that when the universe was about 100 times smaller than at present, the temperature of the cosmic microwave background radiation was around 0 °C, i.e. 273.15 K.

In 2011, another Nobel Prize for cosmology was awarded to three astronomers for their discovery of the accelerating expansion of the universe. The Royal Swedish Academy of Sciences decided to divide the sum of 10 million Swedish crowns between leaders of the two competing teams especially for their works [207], [226], and [206] from the years 1997–1999, where they published the measured values of some cosmological parameters and the discovery that the expansion of our universe accelerates due to dark energy.

The first laureate is the American Saul Perlmutter (b. 1959) who won one half of the Nobel Prize (see Fig. 10.1). Perlmutter headed the *Supernova Cosmology Project* at the University of California at Berkeley. He studied physics at Harvard University graduating in 1981 and received his PhD degree also in physics at the University of California at Berkeley in 1986.

The second Nobel Prize Winner is Adam Guy Riess (b. 1969), who is a professor of astronomy at Johns Hopkins University and the Space Telescope Science Institute in Baltimore, Maryland. This well-known American cosmologist graduated from the Massachusetts Institute of Technology in 1992 and defended his PhD at Harvard University in 1996.

Finally, the third laureate is the American-Australian astronomer Brian Schmidt (b. 1967) who was the team leader of the *High- z Supernova Search*. He studied astronomy at the University of Arizona graduating in 1989 and received his PhD from Harvard University in 1993. The latter two laureates shared the second half of the Nobel Prize.

Festive lectures by all three Nobel Prize laureates were held on December 8, 2011, at the Aula Magna at Stockholm University. The audience learned about the theoretical background of their research, how the observational program was organized, which cosmological models can now be ruled out a priori, and so on. The Nobel Prizes for Physics were then awarded on December 10, the anniversary of Alfred Nobel's death (he died in 1896). Recall that the Nobel Prize for Physics has

been awarded since 1901, with Wilhelm Conrad Röntgen being the first person to receive this award.



10.2. An expanding universe and the Hubble constant

Let us first recall some important milestones in the development of cosmology. At the end of the 16th century, the Italian astronomer Giordano Bruno (1548–1600) in his treatise *De l'Infinito, universo e mondi* [38] conjectured that the universe is infinite and that each star is similar to our Sun. These hypotheses are often considered as the beginning of modern cosmology [184]. For his revolutionary ideas Bruno was burned at the stake at Rome's square Campo de' Fiori² (Field of Flowers) on February 17, 1600. Since then many discoveries in cosmology have been made.

In 1900, the German physicist Karl Schwarzschild (1873–1916) conjectured in his work [252] that the universe could have a finite volume and that it can be described as a huge three-dimensional hypersphere in the Euclidean space \mathbb{E}^4

$$\mathbb{S}_r^3 = \{(x, y, z, w) \in \mathbb{E}^4 \mid x^2 + y^2 + z^2 + w^2 = r^2\},$$

where r is independent of time. He even derived a lower estimate for the radius $r \geq 100\,000\,000$ AU and studied its non-Euclidean structure using parallaxes of the nearest stars. In 1924, Sir Arthur Eddington established a lower estimate of the radius of the universe modeled by \mathbb{S}_r^3 from the distances of some globular clusters (see [196, p. 76]) which was about two orders of magnitude larger than that derived by Schwarzschild.

If some space object recedes from us, its characteristic spectral lines in the visible range show redshift due to the Doppler effect. On the other hand, if the object approaches us, the spectral lines are shifted towards the blue end of spectrum. For example, our neighboring galaxy M31 in Andromeda exhibits a blueshift, because the radial

²Today there is a statue that commemorates this execution.

component of its velocity (i.e. speed toward an earthly observer) is about 300 km/s (see [260]). For completeness, recall that *redshift* z is defined by

$$z = \frac{\lambda}{\lambda_0} - 1,$$

where λ_0 is the wavelength of a spectral line when the source and the observer are stationary relative to one another, and λ is the corresponding measured wavelength of the light source. Objects with a negative value of z therefore approach the observer, while objects with positive value of z recede from the observer. If $z = 1$, then $\lambda = 2\lambda_0$. Galaxies that are more than 7 billion light years away from us have such a high redshift [211]. The number $1 + z$ obtained from the spectrum of a very distant galaxy actually shows the degree to which the universe³ has expanded since the light reached us. Because photons propagate through vacuum at the same speed c for all wavelengths, z does not depend on the choice of λ_0 . Using the Special relativity formula $z = \sqrt{(c + v)/(c - v)} - 1$, we may estimate the recession speed v of monitored objects for z not too large by (see e.g. [61])

$$v = \frac{(z + 1)^2 - 1}{(z + 1)^2 + 1}c.$$

The idea that the universe might expand started to evolve already in 1915. At that time, the American astronomer Vesto Mevlin Slipher (1875–1969) investigated spectra of 15 well-observed spiral nebulae, see [261]. To his surprise he found that 11 of them showed redshifts of the spectral lines of iron and vanadium, while only three showed blueshift, and one object had approximately zero redshift. However, Slipher had no idea that these objects were galaxies. Three nebulae (NGC 1068, 4565, and 4594) were even receding from us with a speed greater than 1000 km/s, and the average radial velocity of all 15 spiral

³The *universe* will be modeled by an isochrone in spacetime, which corresponds to a given time instant after the Big Bang. Its expansion is then modeled by a three-dimensional expanding hypersurface in a four-dimensional spacetime. Attention: a completely different three-dimensional hypersurface in spacetime is called the *observable universe*, which in addition is seen only in projection on the celestial sphere, see Chapter 18.

nebulae from the Earth was about 400 km/s. The universe in our neighborhood was therefore likely to be expanding on average.

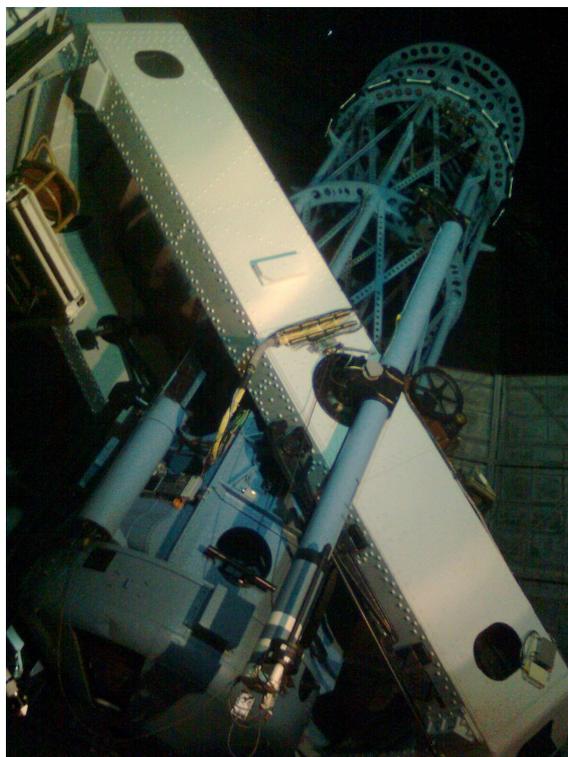


Fig. 10.2. The Hooker 100-inch telescope at the Mount Wilson Observatory in California. This allowed E. Hubble to discover other galaxies and the expansion of the universe (photo M. Krížek).

Another important discovery was made by Edwin Powell Hubble (1889–1953) at the Mount Wilson Observatory in California, where the largest telescope in the world at that time with a mirror of diameter 2.5 m was placed (see Fig. 10.2). When observing the nebula M31 in Andromeda (see Fig. 9.3) he found that it is composed of a large number of stars, like our Milky Way. In the years 1922–1924, using pulsating variable stars — Cepheids⁴, he found that M31 and other nebulae do not belong to our Galaxy,⁵ but that they are very distant stellar islands [104]. Hubble thus changed the classification of many

⁴Cepheids are stars whose brightness varies periodically. They were discovered by Henrietta Swan Leavitt. Around 1912, she noticed that the average luminosity of Cepheids and their period are directly proportional. Cepheids thus belong to the class of so-called *standard candles*, which are astronomical objects of any class with known luminosity.

⁵A similar conclusion was reached by Heber Curtis already in 1917, see [58].

objects in the vicinity of the Milky Way. The nebula M31 therefore lies beyond the border of the Milky Way and is a single galaxy, which is even larger than ours. According to current measurements, it is over 2 million light years away, while the diameter of our Galaxy is about 100 000 light years.

In 1925, Gustav Strömberg (1882–1962) published a review article [272]. He compared the radial velocities of 43 galaxies that were mostly measured by V. Slipher. Only five of them showed a blueshift, while 38 showed a redshift.⁶ This was a very significant fact that required a deeper analysis. From a statistical point of view it is almost impossible that this was only a coincidence if the universe is assumed stationary on average. Indeed, if the probability of a galaxy having blueshift and redshift were equal to 0.5, then the probability P that 43 randomly selected galaxies would have at most five blueshifts is by the binomial theorem equal to

$$P = \sum_{j=0}^5 \binom{43}{j} 0.5^j (1 - 0.5)^{43-j} = 2^{-43} \sum_{j=0}^5 \binom{43}{j} < 10^{-6}.$$

In this way, it was again confirmed that the universe is expanding on average in our neighborhood with probability almost equal to one.

In 1927, the Belgian cosmologist Georges Henri Édouard Lemaître⁷ (1894–1966), inspired by Strömberg’s article [267], came up with the idea of the Big Bang⁸ (see [159]). Two years later, the expansion

⁶Later Hubble’s collaborator Milton L. Humason (1891–1972) discovered from the spectrum of the elliptical galaxy NGC 7619 in the constellation Pegasus that it is receding from us at a speed of 3800 km/s, which is already more than one percent of the speed of light!

⁷Even today his theory is in accordance with redshift of galaxies and their perceived evolution at cosmological distances, with the character of the cosmic microwave background radiation, and the primordial nucleosynthesis of light elements (especially helium and lithium) that occurred just after the Big Bang (cf. [3]). Note that the half-life of a free neutron, which is not bounded in a nucleus, is 611 seconds only. The lifelong work of G. Lemaître is described in the comprehensive monograph [103].

⁸The English term *Big Bang* was first used by Fred Hoyle in 1949. Nevertheless, the idea that the universe could have “zero radius” in the very distant past was formulated by A. Friedmann in 1922. In the English translation of his article [80], footnote 11 states: *The time since the creation of the world is the time that has flowed from that instant when the space was one point ($R = 0$) until the present state ($R = R_0$); this time may also be infinite.*

of the universe was independently confirmed by E. Hubble. In his article [105] he published a chart (see Fig. 10.3) which shows that the radial component v of velocity of a galaxy depends approximately linearly on the distance d from us, that is

$$v \approx H_0 d \quad (10.1)$$

now known as the *Hubble law*. He used a uniform upper estimate of the absolute luminosity of stars and Cepheids from over 20 galaxies.⁹

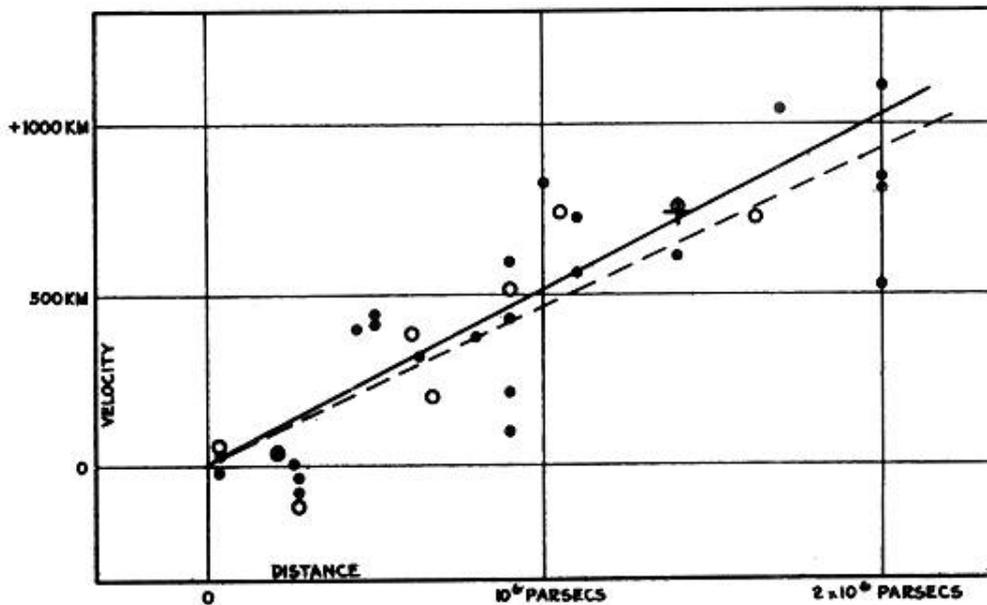


Fig. 10.3. The original graph characterizing the expansion of the universe from Hubble's article [105]. The horizontal axis shows the distance of a particular galaxy from us in parsecs and the vertical axis corresponds to the radial component of velocity of this galaxy in km/s, which is corrected for the movement of the Sun in our Galaxy. Black dots stand for galaxies and the solid line illustrates the relationship (10.1). Circles and the dashed line correspond to smaller groups of galaxies.

From Fig. 10.3 we may deduce the constant of proportionality in (10.1),

$$H_0 \approx 500 \text{ km s}^{-1} \text{Mpc}^{-1} \approx 1.62 \cdot 10^{-17} \text{ s}^{-1},$$

⁹It is remarkable that Hubble [105] does not cite Slipher's articles [260] and [261], nor Lemaître's article [159].

which is now called the *Hubble constant*.¹⁰ However, such a large value contradicted certain facts. For example, the approximate derived value of the Hubble age of the universe $T_0 = 1/H_0$ is then less than 2 billion years, which clearly contradicts the age 4.6 billion years of the Solar system. That is to say, Hubble greatly underestimated the distances of galaxies. Later, the value of the Hubble constant was refined. Current measuring technologies lead to a much smaller value (see e.g. [169], [229], [288])

$$H_0 \approx 72 \text{ km s}^{-1}\text{Mpc}^{-1} \approx \frac{1}{13.6 \text{ Gyr}} \approx 2.33 \cdot 10^{-18} \text{ s}^{-1} \quad (10.2)$$

in the neighborhood of our Galaxy with redshift $z \approx 0$ and it is assumed that elsewhere in the universe (cf. Fig. 18.5) it is about the same value at present. The value of the Hubble parameter is larger for galaxies in the observable universe, since we always look into the past (see Fig. 8.7).



10.3. Type Ia supernovae — standard candles

For cosmology, only the gravitational interaction is essential. The effects of the other three physical interactions, strong, weak, and electromagnetic¹¹ can in case of large spatial distances be neglected. Mathematical formulations of basic cosmological principles are given e.g. in [293], [184], [202]. For a popular interpretation, we refer to the famous Weinberg publication [294].

According to *Einstein's cosmological principle*¹² the universe is homogeneous and isotropic at all its points. We believe that as observers

¹⁰The Hubble constant was not first introduced by Hubble as it is often incorrectly claimed. Already two years earlier, Lemaître states in [159, p. 56] its value to be equal to 625 km/(s Mpc). He derived it from Strömberg's list (see [272, p. 200]) of redshifts and blueshifts of extragalactic nebulae, after subtracting the speed of the Solar system relative to the Milky Way.

¹¹The question arises which of the four fundamental interactions played some role during the Big Bang.

¹²The term *Einstein's cosmological principle* was introduced in 1935 by E. A. Milne [179]. Nevertheless, already in 1922, Carl V. L. Charlier in [49] writes that Albert Einstein introduced the cosmological principle, even though Einstein did not call it in this way.

we do not live at a privileged place in the universe. *Homogeneity* is the assumed property of the universe that at every fixed time instant and on large spatial scales¹³ the universe appears the same to all observers, wherever they are. In other words, for each fixed time translation symmetry of the universe is required.

Similarly, *isotropy* is the assumed property of the universe in which the universe at large spatial scales would seem to an observer at any point¹⁴ to be the same in all directions, i.e., we require the rotational symmetry of the universe.¹⁵ It can be derived (see e.g. [184, p. 714], [294, p. 29]) that isotropy at each point implies homogeneity.¹⁶

Given the isotropy of the universe, we will now deal with determining its rate of expansion. For this purpose, we introduce several additional terms. *Supernova* means a star that has exploded and whose intensity (luminosity) has increased many billion times as a result of gravitational collapse and subsequent explosive nucleosynthesis. Already in 1938, Walter Baade (1893–1960), who collaborated with Fritz Zwicky, said that supernovae could be promising candidates for measuring cosmic expansion. Also Charles Thomas Kowal [122] in 1968 referred to the possibility of measuring distances using supernovae of type Ia.

A supernova is able to produce as much light for several weeks as a small galaxy. The light from such an object travels through the expanding universe, which continually extends the wavelength of every photon (this is the so-called cosmological Doppler effect). The mea-

¹³Usually we consider distance scales on the order of billions of light years, since there are very large structures in the universe, e.g., the *Sloan Great Wall of galaxies* is a fiber of length 1.37 billion light years.

¹⁴If the universe had the shape of the surface of an egg and an observer would be at its tip, then he would see the universe as being isotropic, but it is then in fact not isotropic at all points.

¹⁵The possibility of some kind of rotating universe, which would be homogeneous and anisotropic, was considered by Kurt Gödel [88] (see also [202, p. 15]) — e.g., in all points we would observe half of the sky with blueshift and the other half with redshift. This can be achieved for any odd dimension n on the sphere \mathbb{S}^n .

¹⁶Assuming the homogeneity of an isotropic universe is therefore redundant, although it is usually done. It is as if we would say, let a constant function be continuous, or let the identity matrix be symmetric, etc.

sured wavelength and brightness of supernovae contain information about the history of the expansion of the universe. Such measurements were extensively used by the 2011 Nobel Prize laureates.

Supernovae can be classified into several types. If they do not have spectral lines of hydrogen¹⁷, they belong to class I, which is further divided into the two types Ia and Ib according to whether or not its spectrum contains the characteristic 615 nm absorption line of silicon. Significant similarities when comparing individual explosions of type Ia supernovae show that they probably have the same trigger mechanism. Type Ia supernovae reach their maximum luminosity within about 20 days. Then after many weeks they have essentially the same continually decreasing luminescence after the main peak on the light curve. In particular, the rate of decrease of intensity proved to be crucial for the calibration and determination of distances, because the maximum of their absolute luminosity power varies.

In a generally accepted model,¹⁸ it is assumed that the mechanism of type Ia supernovae is as follows: Consider a close binary system of which one component is a white dwarf (with a high density up to 10^5 kg/cm^3) and the second component is a red giant which continually increases in size. Once the so-called Roche lobe¹⁹ is filled, there appears an overflowing mass from the red giant over the point L_1 onto the white dwarf (see [112]). The matter overflows until the mass of the white dwarf reaches the so-called *Chandrasekhar limit of instability* $1.4 M_\odot$, where $M_\odot = 1.989 \cdot 10^{30} \text{ kg}$ is the Sun's mass. After exceeding this limit the dwarf will gravitationally collapse. First, the inner core collapses into a neutron star (with incredibly high density about 10^{12} kg/cm^3) on to which the outer parts of the dwarf then begin to fall. The energy released causes a huge explosion. Type Ia super-

¹⁷Supernovae of class II have lines of hydrogen in their spectrum and arise by the explosion of large stars called *supergiants* when they are at the end of their development. The inner burned part collapses under its own gravity towards the center, where a neutron star or black hole is then formed.

¹⁸There are several other models. For instance, merging of two white dwarfs could also lead to type Ia supernova.

¹⁹*Roche lobe* is defined as an equipotential surface passing through the Lagrangian point L_1 .

novae exhibit at maximum intensity spectral lines of silicon (Si II) and sulfur (S II), but no lines of hydrogen.²⁰

In 1983, the Indian astrophysicist Subrahmanyan Chandrasekhar (1910–1995) obtained the Nobel Prize in Physics for the discovery of the stability bounds and other related results. The Chandrasekhar limit of instability actually defines a standard candle in the universe from which we can estimate the distance using luminosity. However, a difficulty lies in the fact that the luminosities of type Ia supernovae differ from each other along the axis of rotation of an exploding dwarf and in the direction perpendicular to it.

In a typical galaxy, only a few supernovae per millennium appear completely haphazardly. Therefore, both winning teams every few weeks compared images from a particular region of the sky containing many thousands of galaxies. After subtraction of the two images they occasionally found tiny light points — candidates for type Ia supernovae and their light curves were subsequently analyzed [205]. Each monitored region of sky, in which Brian Schmidt and his colleagues looked for type Ia supernovae, contained about 5 000 galaxies. In this way they discovered in particular very distant supernovae with redshifts of $z \geq 0.2$. The greater z is, the more galaxies are contained in the monitored region, because their number grows approximately with the square of distance (for large z it is not so). Therefore, no supernova was discovered for small z .



10.4. Measurements of cosmological parameters

The rate of expansion of the universe is not constant in time, since, among other way, it is affected gravitationally by mass whose mean density decreases. Therefore, instead of a constant H_0 from equation (10.2) we will consider a function $H = H(t)$ for which $H(t_0) = H_0$, where $t_0 = 13.82$ Gyr is an estimated age of the universe according to

²⁰Supernovae type Ib have the maximum luminous intensity of spectral lines of helium (He I) and type Ic supernovae contain lines of calcium (Ca II) and oxygen (O I).

currently accepted cosmological models. For a fixed time t , its value does not depend on spatial variables, because of the assumed homogeneity and isotropy of the universe. The function $H = H(t)$ is called the *Hubble parameter* and it is defined as the ratio

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (10.3)$$

where the dot denotes the time derivative and $a = a(t)$ is a nonnegative continuously differentiable *expansion function* (*cosmological scale parameter*). For a fixed time instant t the value of $a(t)$ denotes the radius of the universe, provided it has a positive curvature. Then it is modeled²¹ by the three-dimensional sphere $x^2 + y^2 + z^2 + w^2 = a^2(t)$ in \mathbb{E}^4 , see [159]. For example, in simple classical cosmological models without dark energy there exists a constant $C > 0$ such that (see e.g. [184] and [294])

$$a(t) = Ct^{2/3}, \quad \text{i.e.} \quad H(t) = \frac{2}{3t} \quad (10.4)$$

for the last $t_0 - 380\,000$ years when matter dominated over radiation. Hence, the function $H = H(t)$ is decreasing and we may establish that

$$H(t_0) = 1.53 \cdot 10^{-18} \text{ s}^{-1}.$$

Notice, however, that the measured value in (10.2) is in fact about 52% larger. The main reason for this discrepancy is the fact that relations (10.4) do not consider the effect of dark energy. If we knew the exact behavior of the Hubble parameter, then by integration of (10.3) we obtain the relation (cf. a model situation in Fig. 8.7)

$$a(t) = a(t') \exp \int_{t'}^t H(\tau) d\tau \quad \text{for } 0 < t' \leq t$$

²¹For a negative curvature the universe is sometimes modeled by a hyperbolic hypersurface $x^2 + y^2 + z^2 - w^2 = -a^2(t)$ with Minkowski metric (see (18.5)). For zero curvature the value of $a(t)$ denotes the distance of two “typical” galaxies. The function $a = a(t)$ appears in the Friedmann-Lemaître-Robertson-Walker metric that defines the spacetime manifold (see [46], [202]).

with $a(0) = 0$. Nonetheless, this initial condition does not allow us to consider e.g. an exponential expansion function of the form

$$a(t) = C_1 e^{C_2 t},$$

where C_1 and C_2 are positive constants, which leads to a constant Hubble parameter.

At the end of the 20th century, cosmologists thought that the expansion function is concave everywhere, i.e. \dot{a} is a decreasing function of time, since the expansion of the universe is slowed down by an attractive gravitational force. Then a big surprise came. In the late nineties, the Supernova Cosmology Project and the High- z Supernova Search teams focused on supernovae at large distances corresponding to redshifts of 0.2 to 1. They independently discovered that type Ia supernovae have up to 15 % lower intensity (see [86], [206], [226]) than they should have if the expansion of the universe were to be decelerating. Using differently colored filters they realized that the attenuation of intensity is almost certainly not caused by the absorption of light in the matter. But this means that the light of a supernova propagates into a larger volume than if the universe expansion slowed down only by gravity. To explain this paradox, it was necessary to introduce, in addition to dark matter, *dark energy* that on the contrary accelerates the expansion of the universe. Thus, it was found that the derivative $\dot{a} = \dot{a}(t)$ is increasing (i.e. a is strictly convex) in the time interval of about the last five billion years, which corresponds approximately to the redshift $0 \leq z \leq 0.5$.

In Fig. 10.4 an overview of the results of the two competing teams led by Saul Perlmutter and Brian Schmidt is given. The horizontal axis indicates the redshift and the vertical axis contains observed magnitudes of supernovae, which is a unit for measuring the luminosity of celestial objects — see [293, pp. 421–426].

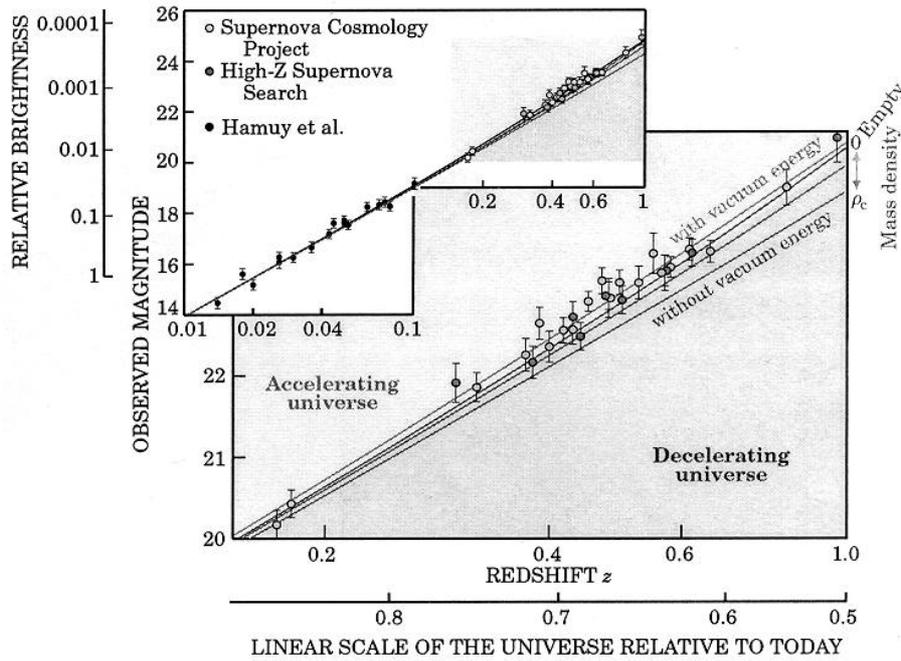


Fig. 10.4. Dependence of magnitude of type Ia supernovae on redshift z showing the accelerated expansion of the universe adapted from Perlmutter's article [205]. For comparison results of Hamuy's team measured before 1996 are also shown for $z < 0.1$.

For very distant objects with $z > 1$ the measured data suggest that there was a period of deceleration in cosmic expansion, i.e., the period when the derivative of the function expansion \dot{a} is decreasing with time, see [227], [226], and [229]. Some observed supernovae were located even more than 10 billion light years from us, which allowed one to determine that the slowing expansion of the universe had changed into an accelerated expansion after about 8–9 billion years from the Big Bang.

Both teams focused on the determination of several other important cosmological parameters. The present value of the Hubble parameter was identified as being close to (10.2). Hence, the so-called *Hubble age of the universe* is

$$T_0 = \frac{1}{H_0} = \frac{\text{Mpc}}{72 \text{ km/s}} = \frac{3.086 \cdot 10^{19} \text{ km}}{72 \text{ km/s}} = 4.286 \cdot 10^{17} \text{ s} \approx 13.6 \text{ Gyr}$$

which might be just a very rough estimate of its true age t_0 .

In the early twenties of the last century, Alexander A. Friedmann (1888–1925) derived from the system of ten Einstein’s equations²² applied to a perfectly symmetric universe, which is homogeneous and isotropic for every fixed time instant, a nonlinear differential equation (see [80], [81]) for the expansion rate of the universe

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}, \quad (10.5)$$

where G denotes the gravitational constant, $c = 299\,792\,458$ km/s the speed of light in a vacuum, $\rho = \rho(t)$ the average mass density in the universe, k/a^2 the spatial curvature (see Chapter 18),

$$k \in \{-1, 0, 1\}$$

the normalized curvature (curvature index), and Λ the cosmological constant²³ that appears in the equations (19.8) of General relativity (see e.g. [46], [184], [202]). The penultimate term in equation (10.5) containing Λ plays a dominant role as $t \rightarrow \infty$, because the density $\rho(t)$ is proportional to $a^{-3}(t)$. Dividing equation (10.5) by the square $H^2 \neq 0$, we get for all t by (10.3) the following equation for three dimensionless parameters

$$\boxed{1 = \Omega_M(t) + \Omega_\Lambda(t) + \Omega_K(t).} \quad (10.6)$$

Here Ω_M is the *density of dark and baryonic matter*, Ω_Λ is the *density of dark energy* (see [202]), Ω_K is the *density of spatial curvature*,

²²Einstein’s equations of General relativity (19.8) are based on tensor calculus, which Einstein acquired during discussions with George Pick in Prague during the period 1911–1912.

²³A. Einstein assumed that the universe has a finite volume with $k = 1$ and is stationary, and initially did not believe in its expansion. In order to prevent the gravitational collapse of the universe, he introduced in 1917 the cosmological constant in his equations of General relativity [72]. Its repulsive character enabled him to consider a non-expanding and also stationary universe for $k \leq 0$. In 1917, however, Willem de Sitter (1872–1934) found a very special solution to Einstein’s equations [62], which describes an isotropic expansion of the universe with zero mass density and $\Lambda > 0$. When in 1929 Hubble published his article [105] on the expansion of the universe, Einstein renounced the cosmological constant and said that this was the biggest blunder of his scientific career [79].

and

$$\Omega_{\text{M}}(t) := \frac{8\pi G\rho(t)}{3H^2(t)}, \quad \Omega_{\Lambda}(t) := \frac{\Lambda c^2}{3H^2(t)}, \quad \Omega_{\text{K}}(t) := -\frac{kc^2}{H^2(t)a^2(t)}. \quad (10.7)$$

The validity of the normalized Friedmann equation (10.6) is discussed in Chapter 19. This equation shows the relationship between the density parameters of baryonic and dark matter, dark energy, and of the curvature of the universe. For a flat universe with $k = 0$ we have $\Omega_{\text{M}} + \Omega_{\Lambda} = 1$. However, this equality can never be proved by measurements which will always show some uncertainty. Note that only sharp inequalities may be verified by measurements.

An important aim was therefore to determine the present values of the parameters Ω_{M} and Ω_{Λ} . The remaining parameter Ω_{K} can then be calculated from (10.6). Articles [207], [226], and [206] from the period 1997–1999 state that the density parameter of dark energy Ω_{Λ} is positive with probability greater than 99 % and the density parameter of mass $\Omega_{\text{M}} \approx 0.2$. At present over a thousand supernovae of type Ia with redshift of $z > 0.1$ are known. Thus, the values of the parameters Ω_{M} and Ω_{Λ} are continually improved. The corresponding Ω_{M} versus Ω_{Λ} diagram was published in [251, p. 16]. For example, according to the article by Riess and his collaborators [229], the present values of cosmological parameters are approximately equal to $\Omega_{\text{M}} = 0.3$ and $\Omega_{\Lambda} = 0.71$. Therefore, by (10.6) and (10.7) the following relation should hold²⁴

$$\Omega_{\text{K}}(t_0) = -0.01 = -\frac{kc^2}{H^2(t_0)a^2(t_0)},$$

which corresponds to a positive curvature index $k = 1$, that is associated with a three-dimensional hypersphere embedded into four-dimensional Euclidean space. From this and (10.2) we get an unimag-

²⁴In article [254], the following values of the parameters $\Omega_{\text{K}} = -0.014$ and $\Omega_{\Lambda} = 0.716$ were obtained by studying the fluctuations of the cosmic microwave background radiation. Also the standard Λ CDM model (i.e. Lambda–Cold Dark Matter) considers similar values. Cosmological parameters measured by the Planck satellite [213] are presented in (19.11).

inably large radius of the present universe²⁵

$$a(t_0) = \frac{10c}{H_0} \approx 1.3 \cdot 10^{27} \text{ m} \approx 140 \text{ Gly}, \quad (10.8)$$

where ly stands for the light year, and consequently it seems to us that the universe is almost flat. However, let us stress that the relation $\Omega_M + \Omega_\Lambda \approx 1$ does not imply the equality $\Omega_K = 0$.

Let us now focus our attention on the cosmological constant Λ . An important task of both the teams led by Perlmutter and Schmidt was to determine its true value. It was not known whether its value is positive, zero, or negative.²⁶ Perlmutter and his team derived from the first 42 supernovae observed a positive value of Λ with a probability of 99.8 % (see [207, p. 580]). The present literature suggests a number of upper and lower bounds that are usually roughly about 10^{-52} m^{-2} (see e.g. [116], [124]). According to (10.7), (10.2), and the measured value $\Omega_\Lambda(t) = 0.71$, we indeed have

$$\Lambda = \frac{3H_0^2\Omega_\Lambda(t)}{c^2} \approx 1.22 \cdot 10^{-52} \text{ m}^{-2}.$$

From (10.5) we see that for $\Lambda > 0$ and $a \rightarrow \infty$ the derivative \dot{a} increases from a certain time instant t_1 , i.e., the function $a = a(t)$ is convex for $t \geq t_1$.

Next we define the dimensionless *deceleration parameter*²⁷

$$q := -\frac{\ddot{a}a}{\dot{a}^2} = -\frac{\ddot{a}}{a}H^{-2} = -\dot{H}H^{-2} - 1, \quad (10.9)$$

where the second equality follows from (10.3). From this we observe that the current value of the deceleration parameter $q_0 = q(t_0)$ appears at the quadratic term in the Taylor expansion (see e.g. [184, p. 781], [202, p. 313], [224, p. 652])

²⁵All values presented here, however, should be taken “with a great reserve”, because only approximate models loaded with a number of various errors are examined.

²⁶From equation (10.5) for Einstein’s stationary universe with $\dot{a} = 0$ we obtain an upper estimate of $\Lambda < 3k/a^2$.

²⁷Observe that $q \equiv 0$ when (10.4) is true.

$$\begin{aligned}
a(t) &= a(t_0) + \dot{a}(t_0)(t - t_0) + \frac{1}{2}\ddot{a}(t_0)(t - t_0)^2 + \dots \\
&= a(t_0)\left(1 + H_0(t - t_0) - \frac{1}{2}q_0H_0^2(t - t_0)^2 + \dots\right). \quad (10.10)
\end{aligned}$$

Therefore, it was necessary to obtain spectra of type Ia supernovae, which are very far ($z \approx 1.7$), see [79]. In the paper [229, p. 110] was found a negative value of the parameter (cf. Fig. 8.7)

$$q_0 \approx -0.6, \quad (10.11)$$

i.e., a is strictly convex in a neighborhood of t_0 .

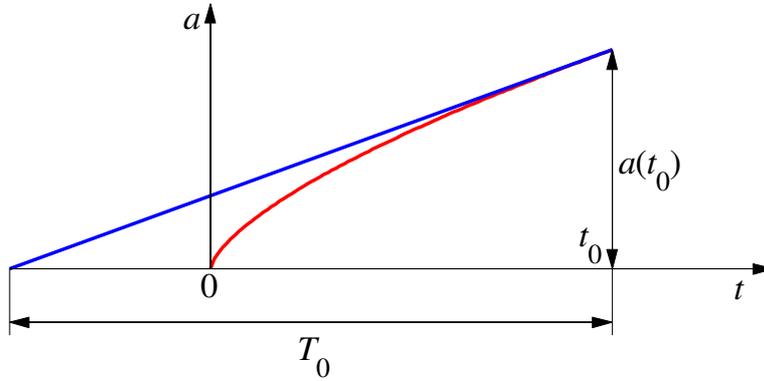


Fig. 10.5. If the expansion function were to be concave, then its graph would be below the tangent line through the point $(t_0, a(t_0))$. Thus for $\dot{a}(t_0) > 0$ by (10.12) the age of the universe t_0 would not exceed the Hubble age $T_0 = 1/H_0 = 13.6$ billion years, which is not in accordance with the measured data.

Nevertheless, a negative value of $q_0 < -1$ was also predicted (see [277]) by the New Zealander Beatrice Tinsley (1941–1981) at the end of the seventies.²⁸ Indeed, if the expansion function $a = a(t)$ in its definition domain were to be concave everywhere (see Fig. 10.5) and $\dot{a}(t_0) > 0$, then by (10.3) and (10.2) the following would hold for the truth age t_0 of the universe

$$t_0 \leq T_0 = \frac{a(t_0)}{\dot{a}(t_0)} = H_0^{-1} \approx 4.29 \cdot 10^{17} \text{ s} \approx 13.6 \text{ billions of years.} \quad (10.12)$$

²⁸Tinsley did not examine supernovae. Her hypothesis was based on observations that the highest concentration of quasars (i.e. quasi-stellar objects) is at $z \approx 2$.

But the inequality (10.12) is contradicted by some observations, because there are stars which were judged to be older than 13.6 Gyr independently of cosmological models (see e.g. [28]). Consequently, in 1978, Tinsley found [277] that the expansion function must be strictly convex²⁹ in some interval, which corresponds to the accelerating expansion of the universe (see also [94]).

Let us further note that if $q \geq -1$ and a is strictly convex (i.e. \dot{a} is increasing) on some subinterval, then $H = H(t)$ is not increasing there by (10.9). On the other hand, the case $q < -1$ already implies that $\dot{H} > 0$, i.e., $H = H(t)$ would be an increasing function on the respective subinterval. For instance, if we linearly extrapolate the deceleration parameter $q = q(t)$ in Fig. 8.7 to 5 billion years into the future, then its value would be less than -1 , which would lead to the increasing Hubble parameter due to (10.9). But we should not forget that it is only a model.³⁰



10.5. Historical notes

Great discoveries usually do not arise from nothing, but are based on results of many other researchers. K. Schwarzschild unquestionably deserves enormous credit for applying results of non-Euclidean geometry to our universe. Already in 1900 he realized that the universe might have a finite volume [252]. V. M. Slipher (see [261], [267]) studied the redshift of extragalactic nebulae much earlier than E. P. Hubble.

A. Friedmann proposed a model of an expanding universe, but not the expansion of the actual universe. Already in 1922 he assumed [80] that the universe in the past could have a “zero radius”, i.e. five years earlier than a similar conclusion based on astronomical observations was reached by G. Lemaître [159]. Clear arguments for the accelerating expansion of the universe were given by B. Tinsley in late seventies

²⁹The strict convexity of the expansion function at present time has already been predicted by G. Lemaître, see [103, p. 30] and [160].

³⁰Fig. 8.7 corresponds to values $H_0 = 67.15$ km/(s Mpc), $\Omega_\Lambda = 0.683$, and $\Omega_M = 0.317$, cf. (19.11).

of the last century.³¹ Although her groundbreaking article [277] was published in *Nature*, it is not cited in the works of Nobel laureates [205]–[207], [226]–[229].

Currently there is a great discussion about what is the source of the mysterious dark energy which by its antigravity effects causes the accelerating expansion of the universe, see [5], [86]. It is speculated that the fundamental physical constants may depend on time. For example, if the value of the gravitational constant were to appropriately decrease, we would observe an accelerated expansion [259], which could also be explained by the vacuum energy. Sometimes the existence of a dynamical scalar field (quintessence — a hypothetical fifth fundamental force) is considered that causes the accelerated expansion. By means of a positive gravitational aberration causing antigravitational forces we try to explain in Chapter 17 from where at least part of the dark energy causing the accelerated expansion of the universe might come.



³¹Already in 1934 Lemaître [160] proposed that dark energy acts like matter with negative gravitational pressure.

11. Recession of Mars from the Sun

*Mars is the only known planet
inhabited only by robots.*

In Chapter 10, we examined the global expansion of the universe. In the following Chapters 11–15 we will survey a wide range of arguments showing that the Solar system expands at a rate comparable to the expansion of the universe, which is determined by the Hubble constant. This obviously contradicts the law of conservation of energy from classical mechanics. The main objective of the second part of this book is to explain why this fundamental law of physics does not apply exactly in the reality, but only approximately. We present other arguments showing that even single galaxies slowly expand (see Chapter 16), which we attribute to antigravitational forces. In Chapter 17 we demonstrate that their existence may come from the little known and often neglected phenomenon of gravitational aberration. We shall see that energy in the universe is gradually and spontaneously generated. This kind of energy will be grouped with dark energy.

11.1. Antigravity and the law of conservation of energy

The “validity” of physical laws is empirically verified through measurements. However, absolutely precise measuring instruments cannot be realized. Thus, in principle we cannot check that generally accepted laws, such as the laws of conservation of energy, momentum, and angular momentum apply to any number of decimal places. The law of

conservation of energy is one of the cornerstones upon which the current physical model is based. Newton's theory of gravity is formulated so that this law holds exactly. But what about the real world that is only modeled by Newton's theory or the General theory of relativity? To answer this question we will use a broad interdisciplinary approach. We will present more than 10 concrete examples that illustrate a slight increase in the total mechanical energy of a system of actual bodies that mutually interact gravitationally.

First we present some astrophysical, astronomical, geometrical, geophysical, geochronometrical, heliophysical, climatological, paleontological, and observational arguments to support the conjecture that the Solar system slowly expands with a speed of about $5 \text{ m yr}^{-1} \text{ au}^{-1}$ and that such a substantial expansion cannot be explained by the loss of solar mass, or the solar wind or tidal forces. This is obviously inconsistent with Kepler's laws, and therefore also with the law of conservation of energy, since planets of the Solar system are sufficiently isolated from the gravitational influence of nearby stars. For instance, by (4.1) the gravitational force between the Earth and Alpha Centauri (whose mass is $1.1M_{\odot}$ and distance 4.37 ly) is about one million times smaller than the maximal gravitational force between the Earth and Venus. To illustrate the sparsity of our Galaxy, let us mention that if we reduce the Sun with diameter $1.391016 \cdot 10^9 \text{ m}$ to the size of a ping-pong ball with a diameter of 3.7 cm, then Alpha Centauri would be

$$d = \frac{0.037 \cdot 4.37 \cdot 9.46 \cdot 10^{15}}{1.391016 \cdot 10^9} = 1.01 \cdot 10^6 \text{ (m)}$$

from us, i.e. slightly more than 1000 km. This shows how sparse our Galaxy is in our neighborhood. The diameter of the Galaxy in this scale would be 100 times larger than the Earth-Moon distance.

Some authors argue (see e.g. [45], [55]) that the dark energy does not manifest itself at all in the Solar system. In Section 13.7 we demonstrate where their reasoning is erroneous. We also explain where we obtain the energy (at least partly) that is necessary for this acceler-

ated expansion. In Section 17.1 we hypothesize that one of the possible sources of this energy is the so-called *gravitational aberration*, which is a consequence of causality and the finite speed of propagation of gravitational interaction (see also [131] and [133]).

An apparent force which causes a gradual expansion of the Solar system and other gravitationally bound systems is called *antigravity*. We shall see its manifestations on large and small spatial and time scales, unless it is disturbed by other phenomena (resonances, tides, strong electromagnetic fields, and so on). Antigravity is not a new fifth force, but only a side-effect of gravitational forces caused by the finite speed of propagation of the gravitational interaction. Similarly, a side-effect of the strong interaction acting between quarks is that it holds atomic nuclei together.



11.2. The rate of expansion of the Solar system

At the end of the last century, astronomers discovered that the universe should be filled with some mysterious dark energy, which is spread fairly uniformly and whose antigravity effects cause the accelerating expansion of the universe (see Chapter 10). The rate of expansion is given by the Hubble parameter whose size depends substantially on the amount of dark energy.

Rescaling the present value of the Hubble parameter H_0 to the mean Sun–Earth distance, i.e. one astronomical unit¹

$$1 \text{ au} = 149\,597\,870\,700 \text{ m} \approx 150 \cdot 10^9 \text{ m}, \quad (11.1)$$

and taking into account that $1 \text{ pc} \approx 206\,265 \text{ au}$ and that one sidereal year has about $31\,558\,150$ seconds, we get

$$H_0 \approx 70 \text{ km s}^{-1} \text{Mpc}^{-1} = 70 \text{ m s}^{-1} \text{kpc}^{-1} = \frac{70 \cdot 31\,558\,150}{206\,265\,000} \text{ m yr}^{-1} \text{au}^{-1}.$$

¹This definition of the astronomical unit has been accepted at the 28th General Assembly of the International Astronomical Union in Beijing in August 2012. The originally stated value $1 \text{ AU} = 149\,597\,870\,691 \text{ m}$ was increased by 9 meters.

Therefore,

$$\boxed{H_0 \approx 10 \text{ m yr}^{-1} \text{ au}^{-1}} \quad (11.2)$$

From this we see that 1 m^3 of the space will increase by an average of 0.2 mm^3 per year, since

$$\left(1 + \frac{10}{150 \cdot 10^9}\right)^3 \approx 1 + 3 \frac{10}{150 \cdot 10^9} = 1 + 0.2 \cdot 10^{-9}. \quad (11.3)$$

Hence, each cubic meter of the universe has increased its volume approximately two times on average since the origin of the Solar system $4.6 \cdot 10^9$ yr ago. From our point of view, there is no reason that manifestations of dark energy have somehow avoided our Galaxy or Solar system. Since values given in (11.2) and (11.3) are relatively large, the influence of dark energy should be found in the Solar system.

Admitting the manifestation of dark energy in the Solar system, we can easily explain a wide range of puzzles such as the faint young Sun paradox [155], the formation of Neptune and the Kuiper belt of comets [20], the existence of rivers on Mars and its satellite Phobos, the paradox of tidal forces of the Moon [287], the paradox of the large orbital angular momentum of the Moon and Triton, migration of planets, the slow rotation of Mercury and the absence of its moons.

In the following sections we show that the Solar system is expanding at a rate comparable to the order of the Hubble constant (11.2), although usually a little bit smaller. For example, for the Earth this local expansion can be written as (see Chapter 13)

$$H_0^{(\text{loc})} \approx 0.5H_0 \approx \frac{\dot{R}}{R}, \quad (11.4)$$

where $R = R(t)$ is the mean Earth–Sun distance at the time t and the dot denotes the time derivative.



11.3. Rivers on Mars

The total solar power incident per unit area of 1 m^2 perpendicular to the Sun's rays at the distance of 1 au is equal to the *solar constant*

$$L_0 = 1.361 \text{ kWm}^{-2}. \quad (11.5)$$

The value of this constant currently varies by less than 0.1 % depending mainly on the number and size of sunspots. Hence, the total solar power is about

$$L_{\odot} = 4\pi R^2 L_0 = 3.828 \cdot 10^{26} \text{ W},$$

where $R = 1 \text{ au}$.

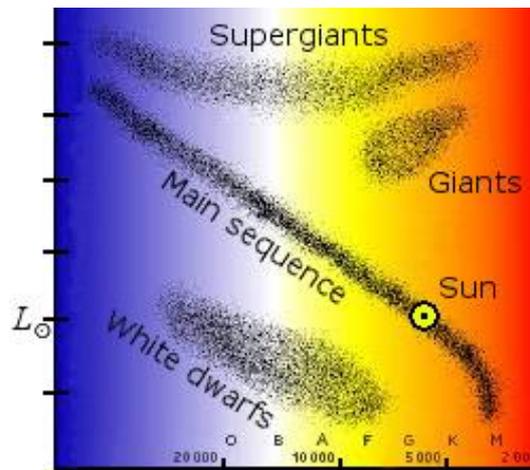


Fig. 11.1. The Hertzsprung–Russell diagram shows the relation between the temperature and luminosity of stars. The Sun is located on the main sequence and slowly moves along it upward to the left. Both axes have logarithmic scales. Due to historical reasons, on the horizontal axis the temperature in Kelvins increases from right to left instead of left to right as is usual. The vertical axis shows a relative power of the stars with respect to L_{\odot} and one segment corresponds to a one hundred times larger power.

Since the Sun is a star on the main sequence of the well-known Hertzsprung–Russell diagram, its position raises to the left in Fig. 11.1. In doing so, the Sun travels only a very short distance on the main

sequence. According to [243, p. 461], the Sun had a surface temperature of 5586 K about 4.5 Gyr ago and its luminosity was only 70 % of the current value (see also [153], [85, p. 48]). Its luminosity increased approximately linearly with time up to the present (see Fig. 11.2). The current effective temperature is 5770 K and after 3 Gyr it will rise to 5843 K, see [243, p. 461]. At that time the Sun will already produce 133 % of the current power.² Thus we see that the Sun's temperature increases relatively slowly, while its power grows much faster (cf. (11.8)).

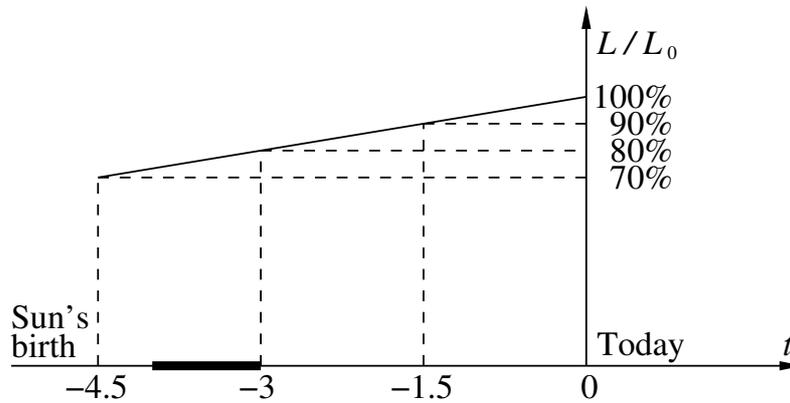


Fig. 11.2. Relative luminosity L/L_0 of the Sun from the origin of the Solar system up to the present. The time t is given in Gyr. The bold interval on the time axis indicates the period when Mars had liquid water on its surface.

Hydrologists estimated from the number of craters in dry river valleys (cf. Fig. 11.3) that Mars had liquid water on its surface 3–4 Gyr ago (see [98] and Fig. 11.2). At that time the luminosity of the Sun was only 75 % of its present value (see e.g. [245], [300]). The solar power decreases with the square of the distance from the Sun. Therefore, by (11.1) the corresponding solar constant for Mars would only be

$$L_{\text{Mars}} = 0.75L_0 \left(\frac{150}{225} \right)^2 = \frac{L_0}{3} \quad (11.6)$$

²The Sun will become a red giant after 12 Gyr from its origin. It is estimated that its radius will increase about 165 times, i.e. to 0.77 au, and its power will be enormous (see [243] for details).

provided that Mars had been on average

$$r = 225 \cdot 10^9 \text{ m}, \quad (11.7)$$

farther away from the Sun than it is now. However, a value for L_{Mars} that is three times smaller than L_0 is not able to explain a mean temperature necessary for the existence of hundreds of rivers. Now we observe their dry riverbeds and many lakes between -50° and 50° of Martian latitude (see Google Mars Maps). Note that the sidereal period of Mars' rotation of 24.623 hours is very similar to Earth's rotation.



Fig. 11.3. Evidence of rivers on Mars 3–4 Gyr ago. Neither wind nor lava can create such sinuous formations. The bottom of the former sea is at the bottom right. The center of the image (whose dimensions are $175 \times 125 \text{ km}^2$) is at 42.3° south Martian latitude and 92.7° west longitude (photo NASA).

Imagine that we were to have each day a permanent two-thirds eclipse of the Sun on the Earth. Then the Earth's surface would not have warmed too much. A long term decrease of solar luminosity of about only 2% caused ice ages in the past, even though there was also

the greenhouse effect. The huge decrease of 66.6 % as given in (11.6) thus excludes the existence of rivers on Mars, if it were to have the same orbit as now (see (11.7)). Its surface would be completely frozen. Higher concentrations of CO₂ (as suggested by [20, p.177]) surely contributed to a higher surface temperature, but cannot fully explain the presence of liquid water. In Section 13.2 we show that radioactive isotopes only contributed negligibly to a higher temperature 3.5 Gyr ago.

According to [208], on the Northern hemisphere of Mars there was a large ancient ocean which might have been frozen, and which covered about one third of the Mars' surface. The proof of this claim is based on the fact that there are almost no craters in the large neighborhood of the North Pole. On the other hand, there are many craters around the South Pole. Due to asteroid impacts a large amount of water certainly escaped from the area into space due to low gravity,³ but some water sublimated as well. The number of asteroids falling to the North and South hemispheres would have been similar at that time. The so-called *late heavy bombardment* occurred approximately 4.1 to 3.8 billion years ago.

Northeast from the four large shield volcanoes on Mars there is a gigantic stream up to 100 km wide, which developed due to fast melting of glaciers during volcanic eruptions. However, hundreds of its other riverbeds have dimensions comparable to earthly riverbeds (see Fig. 11.3). Water flowed here when life on Earth began to develop. Riverbeds of small brooks and streams have been smoothed by wind erosion.

Automatic probes sent us other evidence to prove the existence of liquid water on Mars (cf. Fig. 11.4). There also exists a large quantity of limonite, jarosite, gypsum, and sandy clay, which need liquid water for their formation. This proves that rivers on Mars were not formed by methane, as we know they were on Titan, Saturn's largest moon.

³It is not excluded that some water on Earth came from Mars, since also Martian meteorites are present on Earth.



Fig. 11.4. Sand and pebbles of diameter 2 – 40 mm represent a further proof of liquid water on Mars given by the Curiosity mission. The robot landed in the Gale crater, where it is assumed that there was flowing water (photo NASA).

At present there is an ice age on Mars. The current annual average temperature of -60°C is well below the freezing point of water, as seen by the measurements of the missions Viking⁴, Pathfinder, Spirit, Opportunity, and so on. It is true that the temperature can climb up to 20°C on dark rocks in the equatorial regions at noon, if it is calm and with Mars is at perihelion.⁵ However, for the existence of rivers it is necessary that the mean daily temperature including the night time is not be too far below the freezing point.

There are dozens of climate models on the original Martian atmosphere, which try to explain the existence of liquid water on Mars in the distant past (see e.g. [244], [245], [246]). If the temperature⁶ of water is 273.16 K and the pressure 611.7 Pa, then water can exist simultaneously in gaseous, liquid, and solid state. This is called a *triple point* (see Fig. 11.5). Automatic probes measured the pressure

⁴For example, the measured temperature by Viking 2 mission was in the range -100°C to -24°C .

⁵The orbit of Mars currently has a relatively large eccentricity nearly equal to 0.1.

⁶The freezing point 0°C corresponds to a slightly lower temperature 273.15 K.

on Mars' surface to be in the range of 600–900 Pa, which is comparable with the necessary 611.7 Pa. The highest pressure on Mars (about 1050 Pa) is supposed to be in the great depression Hellas Planitia (see Fig. 15.3), while on the top of the Olympus Mons it is about 30 Pa. Thus liquid water may exist there only when the temperature has a slight difference from the freezing point. At present, therefore, most of the water is frozen (see Fig. 11.6) and a small amount exists as vapor or salt water.

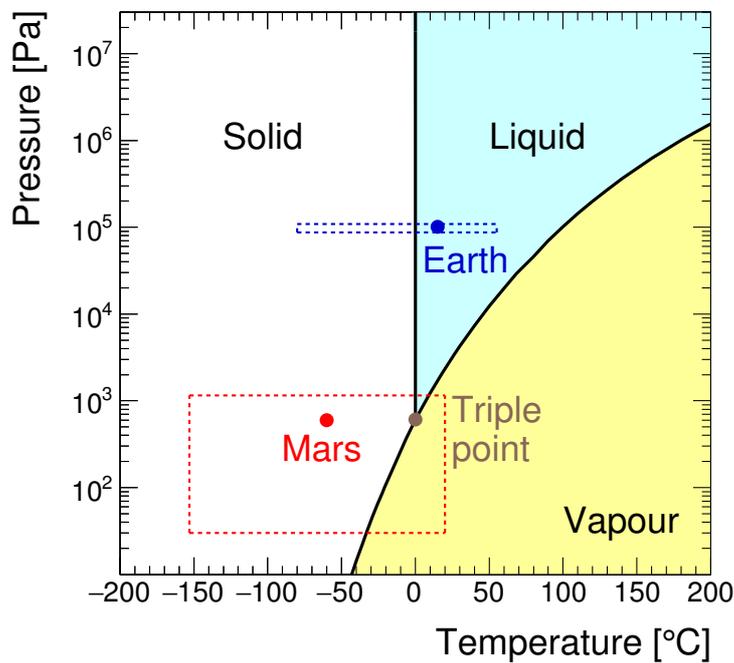


Fig. 11.5. The triple point in the phase diagram of water. The average temperature and pressure on the Earth and Mars is indicated by a small bullet. Boxes show approximate ranges of temperature and pressure that can be achieved on the Earth and Mars. The vertical axis has the logarithmic scale.

Mars had a higher atmospheric temperature and pressure 3–4 Gyr ago, otherwise we could not observe traces of flowing water. However, according to [98] Mars did not have a denser atmosphere than is now on the Earth, since its gravitational field is too weak in comparison with other planets (except for Mercury, which does not have an atmosphere).



Fig. 11.6. A mixture of water ice H_2O and dry ice CO_2 on Mars (photo NASA)



11.4. Mars from the perspective of the Stefan–Boltzmann law

Let us take a closer look at the temperature situation on Mars with respect to the Stefan–Boltzmann law. Assume for a moment that the Sun and Mars are absolute black bodies. In this case the absorbed solar energy is equal to the emitted energy. If Mars were to absorb the total luminosity the Sun coming to Mars’ surface, then its temperature would be given by calculating

$$4P \cdot \sigma T^4 = L_{\odot} \frac{P}{4\pi r^2}, \quad (11.8)$$

where $\sigma = 5.669 \cdot 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$ is the Stefan–Boltzmann constant, r is the Mars–Sun distance (11.7), P is the area of the maximal cross-section of Mars, and $4P$ is its surface area, T is the equilibrium temperature at the distance r for the total solar power $L_{\odot} = 3.828 \cdot 10^{26} \text{ W}$ and albedo $A = 0$.

However, Mars is not a black body and the present value of the Bond albedo is

$$A = 0.25.$$

The equilibrium temperature on Mars' surface is thus given by the relation

$$T_{\text{Mars}} \approx \sqrt[4]{\frac{(1-A)L_{\odot}}{16\sigma\pi r^2}} = 211 \text{ K} \approx -62^{\circ}\text{C}. \quad (11.9)$$

We see that the theoretical temperature from (11.9) is indeed very close to the yearly planetwide average temperature $\approx -60^{\circ}\text{C}$ measured on several stations.

At the time that the Sun's luminosity was 75 % of the present value (see Fig. 11.2), then we get from (11.9) only

$$T_{\text{Mars}} = 197 \text{ K} \approx -76^{\circ}\text{C}.$$

For such a low value the greenhouse effect could hardly guarantee that the mean temperature would reach the freezing point of water 273.15 K ($= 0^{\circ}\text{C}$), even though Mars had a denser atmosphere in the past, which was lost due to low gravity and the solar wind.⁷ For example, by Fig. 11.7 the average temperature on the Earth without the greenhouse effect would be $(-14)^{\circ}\text{C}$. However, the current average temperature is around 15°C , and thus the greenhouse effect produces about 29°C higher temperature.

Moreover, the Bond albedo of Mars' surface 3–4 Gyr ago was higher than $A = 0.25$, since there were water clouds causing snowing or raining which fed many rivers (see Fig. 11.3). Therefore, Mars could not have been a completely reddish planet in the past. Ice and snow were present not only at the polar caps but also at other regions that increased the albedo and decreased the temperature (11.9), too. On the other hand, there were active volcanoes which increased the average surface temperature of Mars.

⁷Mars has a very weak magnetic field, which does not act as a barrier to the solar wind as in the case of Earth.

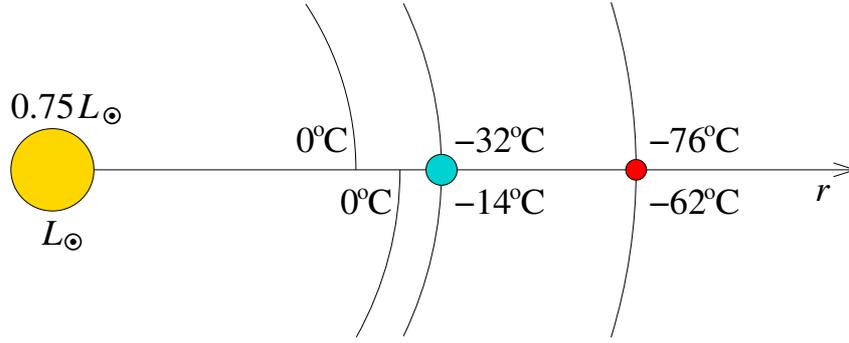


Fig. 11.7. The equilibrium temperature by the Stefan-Boltzmann law (11.9) for albedo $A = 0.25$. The temperatures in the upper part correspond to the luminosity $0.75 L_{\odot}$ and in the lower part to L_{\odot} . The radii of the above circles are 117, 134, 150, and 225 million km. The Sun is on the left, Mars is on the right, and the Earth is in between.

When the Sun's luminosity was $0.75L_{\odot}$, then by (11.9) we would only need for the radius r to be $r = 117 \cdot 10^6$ km to reach the freezing point of water $T_{\text{equilibrium}} = 273.15$ K (see Fig. 11.7). However, this distance is more than 100 million km less than the current average Mars-Sun distance (11.7). Along with all the above arguments we see that Mars had to be tens of millions of kilometers closer to the Sun than it is now to allow liquid water to be available to form rivers (compare with Figs. 11.2, 11.8, 11.9) over a period of one billion years. But this corresponds to a long-term average recession speed of Mars from the Sun in the order of 10 meters per year, which is comparable to the Hubble expansion (11.2).

Let us now examine in detail how to achieve such an estimate. If Mars were e.g. 180 million kilometers away from the Sun at its formation, then its average recession speed would be just 10 m per year during the 4.5 Gyr of its existence to reach the present distance $r = 225$ million km. By (11.1) this would correspond to the current local Hubble expansion, which is comparable to the size of the Hubble constant (11.2) rescaled to the distance r

$$H_0^{(\text{loc})} = \frac{150}{225} H_0 \approx 0.67 H_0. \quad (11.10)$$

In this model example the corresponding value of the solar constant for Mars would be equal to (cf. (11.6))

$$L_{\text{Mars}} = 0.75L_0 \left(\frac{150}{180} \right)^2 = 0.52L_0.$$

For the total solar power $0.75L_\odot$ and L_\odot we would get by (11.9) the values -53°C and -37°C , respectively, for $r = 180$ million km in Figure 11.7. Equation (11.10) obviously represents only a guess. Here we present it only to illustrate a possible recession speed of Mars from the Sun to explain liquid water on the Martian surface. The actual recession speed could be even higher.

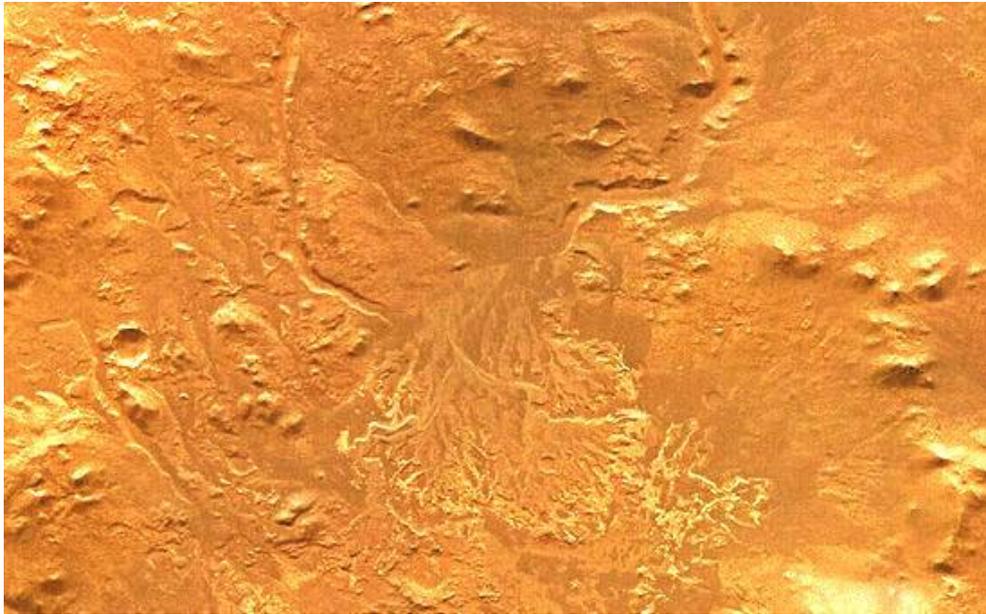


Fig. 11.8. In the western part of the Eberswalde crater the delta of a river is located in a region of about $25 \times 40 \text{ km}^2$. Its center has the coordinates -23.8953° S and 326.7426° E (photo NASA).

In conclusion, let us note that tidal forces, solar wind, solar mass losses, etc. can explain the expansion velocity of only a few centimeters per year, as we will show in more detail for the Earth in Section 13.6. The magnetic field of Mars is almost zero and thus the Sun's magnetic field also has a negligible effect on the Mars–Sun distance. Decreasing temperature conditions on Mars accompanied by an increasing heat flux from the Sun suggest a recession speed of Mars comparable

to (11.10), which may be caused by the antigravity effects. In the next chapter we introduce a similar estimate as in (11.10) for the Moon, but with a higher accuracy of several centimeters per year.



Fig. 11.9. Sediments in the Gale crater taken by the Curiosity mission represent a further proof of liquid water on Mars during a long time period (photo NASA).



12. Recession of the Moon from the Earth

*The universe is expanding almost as rapidly
as the Moon recedes from the Earth.*

AUTHORS

12.1. Measurement of the Earth–Moon distance

In this chapter we show that the Moon moves away from the Earth more rapidly than would be gathered from classical mechanics. The hypothesis about antigravitational forces and the local expansion of the Solar system can be very well tested just by means of a precise measurement of the Earth–Moon distance. The Moon's orbit has the small eccentricity $e = 0.0554$. Thus the Moon slowly approaches the Earth and then moves away during each period. Current technologies allow us to establish long-term changes in the parameters of its orbit very precisely.

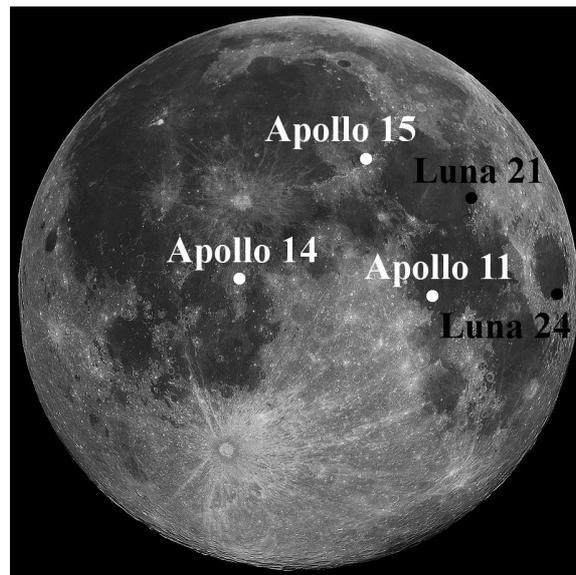


Fig. 12.1. Positions of retroreflectors on the Moon

Since the seventies of the last century, changes in the mean Earth–Moon distance

$$D \approx 384\,402 \text{ km} \quad (12.1)$$

have been carefully measured by laser retroreflectors installed on the Moon by the Apollo 11 mission in 1969 and later by the Apollo 14 and 15 missions, as well as by the Luna 21 probe with Lunokhod 2. In 2012, the US spacecraft Lunar Reconnaissance Orbiter managed to find the lost probe Luna 24 with another retroreflector. Thus at present we have five functioning retroreflectors on the Moon (see Fig. 12.1).

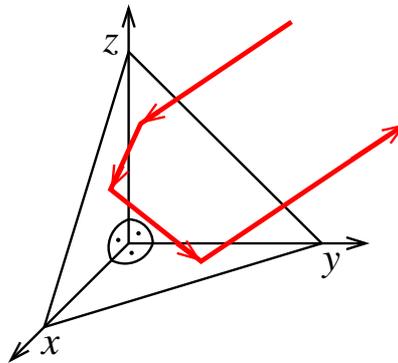


Fig. 12.2. The principle of an idealized retroreflector. Its faces in the Cartesian system (x, y, z) are given by the equations $x = 0$, $y = 0$, and $z = 0$. The laser beam transmitted in the picture in the direction of (a, b, c) moves after reflection from the face $x = 0$ such that a changes into $-a$ and the other two components of the direction vector remain unchanged. This property is a consequence of the law of reflection stating that the angle of incidence equals the angle of reflection. A similar situation takes place at the other faces. Thus the beam in turn moves through the directions (a, b, c) , $(-a, b, c)$, $(-a, -b, c)$, and $(-a, -b, -c)$.

Retroreflectors are composed of many cube-corner tetrahedra that arise “by cutting” a corner from a homogeneous hyaline quartz cube. An impulse laser beam sent to the retroreflector after passage through the front face¹ into an optically denser environment is refracted towards the vertical. Then it proceeds as illustrated in Fig. 12.2, where a simplified retroreflector is sketched. It consists of three mutually

¹The front face in Fig. 12.2 is the equilateral triangle parallel to the plane $x + y + z = 1$.

perpendicular reflective surfaces. Total reflections appear at the faces which are described by the relations $x = 0$, $y = 0$, and $z = 0$. After the exit from the front face, the beam returns back to the original direction independent of the inclination of the retroreflector.² Since the time between sending and receiving the beam can be measured very precisely, retroreflectors can be used to determine changes in the mean Earth–Moon distance with a precision up to one millimeter (but not the distance itself as is often claimed).



12.2. The paradox of tidal forces of the Moon

The nearer side of the Earth to the Moon is attracted by a greater force than the opposite side. As a result of a nonconstant gravitational potential, *tidal forces*³ are formed. Due to the Earth’s rotation, they give rise not only to ebbs and flows of the seas and oceans, but also to the lifting of the Earth’s mantle (see Fig. 12.3). Continual deformations of the Earth cause its rotation to slow down, and thus its spin (rotational) angular momentum decreases over time. Then from the law of conservation of total angular momentum it follows that the orbital angular momentum of the Earth–Moon system has to increase.

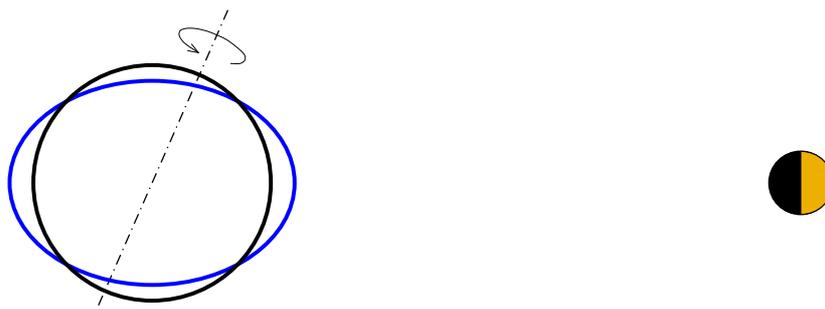


Fig. 12.3. Reduction of the angular velocity of the Earth’s rotation is caused by friction when creating tidal bulges, which are formed by the gravitational influence of the Moon. Deformations of the Earth are converted into heat.

²An ordinary mirror does not have this property!

³Tidal forces played an important role in the deposit of sediments. They provide a record of the lengths of the year and month over time as discovered by G. E. Williams [297].

Already in 1975, T. C. van Flandern [284] (cf. also [64], [68, Chap. 6]) noticed that the Moon has a somewhat anomalous orbit, because it moves away from the Earth faster than can be explained by tidal forces and Newtonian mechanics. He examined whether it could be due to a slight decrease in the value of the gravitational constant. Long-term measurements show (see [63]) that the mean distance D continually increases on average⁴

$$v = 3.84 \text{ cm per year.} \quad (12.2)$$

However, the action of tidal forces on the Earth's mantle, hydrosphere, atmosphere, etc., can explain a much smaller expansion velocity — only about 55 % of the value given in (12.2), see (12.20) and also [195, p. 67], [203, Sec. 9.10.4]). This is usually called the *Paradox of tidal forces of the Moon*, see [284]. Verbunt [287] even writes about the *tidal catastrophe*.

Therefore, the Moon is receding from us probably not only due to tidal forces. For the remaining increment of almost the same order of about 1.71 cm per year we should look for other explanations. For instance, one possibility is that antigravity also contributes to the observed recession of the Moon.



12.3. A remarkable coincidence

The current expansion rate of the universe can be characterized by the Hubble constant (see (11.2))

$$H_0 \approx 70 \text{ km s}^{-1}\text{Mpc}^{-1} \approx 10 \text{ m yr}^{-1}\text{au}^{-1}.$$

It is not difficult to recalculate the speed of this expansion only to the changing distance between the Earth and Moon. Denoting yr to be the unit of time corresponding to one sidereal year and $R = 1 \text{ au}$, then by (12.1) we have

⁴Hence, the trajectory of the Moon is not an ellipse, but a very dense spiral.

$$H_0 \approx 10 \text{ m yr}^{-1} \text{ au}^{-1} = 10 \frac{D}{R} \text{ m yr}^{-1} D^{-1} = 2.57 \text{ cm yr}^{-1} D^{-1}. \quad (12.3)$$

We see that this value is surprisingly quite close to the mean value (12.2) of the recession speed of the Moon from the Earth. At the same time, antigravity can explain a larger measured recession speed (12.2) than that resulting from the tidal forces.⁵ In the next Section 12.4, we show that from the law of conservation of the total angular momentum one can derive a recession speed of the Moon from the Earth due to tidal forces of only about 2.13 cm per year (cf. (12.20)) assuming a constant moment of inertia of the Earth.

If we want to demonstrate to someone how fast the current expansion rate of the universe is, then it is enough to recall relations (12.2) and (12.3):

The universe is expanding almost as rapidly as the Moon is moving away from the Earth.



12.4. Recession speed of the Moon from the Earth due to tides

Now we will estimate in detail the contribution of tidal forces to the speed (12.2) by the method introduced in [131]. Consider the isolated binary system Earth–Moon with masses

$$m_1 = 5.97219 \cdot 10^{24} \text{ kg}, \quad m_2 = 7.3477 \cdot 10^{22} \text{ kg} \quad (12.4)$$

and assume, for simplicity, that their orbits are circular (estimation of the tidal influence of the Sun on the change in the Earth's rotation

⁵On the other hand, the orbital period of close binary pulsars is shortened (cf. [107], [275]). In this case, the binary system creates a strong and rapidly changing gravitational field in a curved spacetime. According to the General theory of relativity, the system loses energy in the form of gravitational waves. Since pulsars have extremely strong magnetic fields, the energy also decreases in the form of electromagnetic waves. These phenomena and several other effects predominate over antigravity.

will be discussed later). Then the corresponding distance (see (12.1)) can be expressed as follows:

$$D = R_1 + R_2, \quad (12.5)$$

where

$$R_1 = \frac{Dm_2}{m_1 + m_2} \quad \text{and} \quad R_2 = \frac{Dm_1}{m_1 + m_2} \quad (12.6)$$

are the distances of the Earth and the Moon from their Newtonian centre of gravity, respectively.

By the conservation of the total angular momentum of this system (i.e. the sum of the spin angular momenta and the orbital angular momenta), the value

$$J = I_1\omega_1 + I_2\omega_2 + m_1R_1v_1 + m_2R_2v_2 \quad (12.7)$$

has to be constant, i.e., its time derivative is zero,

$$\dot{J} = \frac{dJ}{dt}.$$

Here v_1 and v_2 are the orbital speeds of the Earth and Moon, respectively, relative to their center of gravity, I_1 and I_2 are the momenta of inertia of the Earth and Moon,

$$\omega_1 = \frac{2\pi}{T_1} = 7.292 \cdot 10^{-5} \text{ s}^{-1}, \quad \omega_2 = \frac{2\pi}{T_2} = 2.669 \cdot 10^{-6} \text{ s}^{-1} \quad (12.8)$$

are the angular frequencies of the Earth's and Moon's rotations about their axes, $T_1 = 86\,164.1$ s is the sidereal day, and $T_2 = 27.32166 T_1$. By [41] we have

$$I_1 = (8.036 \pm 0.008) \cdot 10^{37} \text{ kg m}^2. \quad (12.9)$$

We shall examine the sizes of the terms $I_1\dot{\omega}_1$ and $I_2\dot{\omega}_2$ appearing in the equation $\dot{J} = 0$ (see (12.14) and (12.17) below). Consider the following three steps.

1. Using the formula for the momentum of inertia of a homogeneous ball [224, p.109], we find for the momentum of inertia of the Moon (whose density $\rho(r)$ increases towards its center) the inequality

$$I_2 < \frac{4}{3}\pi r_2^3 \cdot \rho_2 \cdot \frac{2}{5}r_2^2 = 8.87 \cdot 10^{34} \text{ kg m}^2, \quad (12.10)$$

where

$$\rho_2 = 3348 \text{ kg/m}^3 \quad (12.11)$$

denotes the mean density of the Moon and $r_2 = 1737 \text{ km}$ is its radius. Note that the term $I_2\omega_2 < 2.37 \cdot 10^{29} \text{ kg m}^2\text{s}^{-1}$ corresponding to the Moon is much smaller than $I_1\omega_1 = 5.86 \cdot 10^{33} \text{ kg m}^2\text{s}^{-1}$. However, we have to compare their time derivatives.

According to Section 2.7, the Earth's rotation has slowed down so much during the last 2700 years that the length of the day increased on average about 1.7 ms per century, i.e.

$$\mathcal{T}_1 = 1.7 \cdot 10^{-5} \text{ s per year.} \quad (12.12)$$

Estimations of the size of these long-term changes in Earth's rotation caused by tidal forces were obtained by a thorough analysis of ancient Babylonian (see e.g. [195, p.62], [247, p.270]), Arab, Greek, and Chinese [203] records concerning the angular heights of the Sun during observed eclipses. This is in accordance with present measurements. Contemporary increased melting of glaciers, internal processes in the Earth (cf. (12.23)–(12.25)), mass transfer in the atmosphere, hydrosphere, etc., cannot explain such a large value as given in (12.12). For simplicity, we will first assume that the momentum of inertia of the Earth I_1 is independent of time. In the next section we shall allow the time dependence $I_1 = I_1(t)$.

The Sun and Moon have practically the same angular diameter (see Section 6.3). Since tidal forces decrease with the third power of the distance from the Earth and the volume increases also with the third power, the tidal influence of each of the two bodies on the Earth is directly proportional to its mean density. By equalities (12.11)

and (4.18) we find that the ratio⁶ of the densities of the Moon and Sun equals $3348 : 1409 = 2.38$. The lunar tidal effect on the Earth is thus larger than the solar tidal effect.

According to [210], the Earth's rotation slows down about 68.5% due to the influence of the Moon's tidal forces, and 31.5% from those of the Sun. Hence, the increase $\mathcal{T}_1 = 0.685\mathcal{T}$ corresponds to the Moon and $0.315\mathcal{T}$ to the Sun. The angular frequency of the Earth's rotation caused by the tidal forces of the Moon during one year is

$$\bar{\omega}_1 = \frac{2\pi}{T_1 + \mathcal{T}_1}. \quad (12.13)$$

From relation (12.12) and (12.13) we see that the corresponding time change of the Earth's angular frequency is⁷

$$\dot{\omega}_1 = \frac{\bar{\omega}_1 - \omega_1}{T} = -\frac{2\pi}{T} \frac{\mathcal{T}_1}{T_1(T_1 + \mathcal{T}_1)} = -3.123 \cdot 10^{-22} \text{ s}^{-2},$$

where

$$T = 31\,558\,149.54 \text{ s}$$

is the sidereal year. By (12.9) the corresponding change of the spin angular momentum of the Earth is equal to

$$I_1 \dot{\omega}_1 = -2.509 \cdot 10^{16} \text{ kg m}^2 \text{ s}^{-2}. \quad (12.14)$$

2. The Moon also reduces its angular frequency about its rotational axis due to the recession speed (12.2) and the 1:1 resonance between the Moon's rotation and the orbital period T_2 of the Moon around the Earth. The Moon has fallen into the so-called *tidal trap*.

Now, we show that $I_2 \dot{\omega}_2$ is several orders of magnitude smaller than the value given in (12.14). Taking into account (12.5), we may

⁶The ratio of tidal forces is actually slightly smaller than 2.38, since the mean angular size of the Sun is $31.98'$, whereas for the Moon it is $31.07'$.

⁷The correct calculation of the derivative $\dot{\omega}_1$ requires the use of the measured value (12.12) of Earth's rotation slowdown, and not a theoretically derived value \mathcal{T} using the conservation law of the total angular momentum and relation (12.2) as done in [195, p. 65].

apply the generalized Kepler's third law,⁸ which perfectly approximates the actual situation on short time scales. From (4.5) the ratio D^3/T_2^2 is constant. Hence, the product $\omega_2^2 D^3$ is also constant by (12.8). Differentiating $\omega_2^2 D^3$ with respect to time, we obtain the differential equation

$$2\omega_2\dot{\omega}_2 D^3 + 3\omega_2^2 D^2 \dot{D} = 0,$$

i.e.,

$$\dot{\omega}_2 = -\frac{3\omega_2}{2} \frac{\dot{D}}{D}. \quad (12.15)$$

By (12.2) the mean long-term observed time change of D is given by

$$\dot{D}_{\text{observed}} = \frac{3.84 \text{ cm}}{T} = 1.2 \cdot 10^{-9} \text{ m/s}. \quad (12.16)$$

From this, (12.1), (12.8), (12.10), and (12.15) we obtain

$$|I_2 \dot{\omega}_2| < 1.1 \cdot 10^8 \text{ kg m}^2 \text{s}^{-2}. \quad (12.17)$$

Hence, the decrease of the Moon's spin angular momentum is negligible in comparison to the value given in (12.14).

3. The decrease of the spin angular momentum of the Earth in (12.14) has to be compensated by the increase of the orbital angular momentum $m_1 R_1 v_1 + m_2 R_2 v_2$ in (12.7). We will again employ the 1:1 resonance between the angular frequency of the Moon about its rotational axis and the angular speed of the Earth about their common center of gravity, i.e., $\omega_2 = v_2/R_2 = v_1/R_1$. Consider further the Earth–Moon system whose center of gravity is at rest. Then by the momentum conservation law $m_1 v_1 = m_2 v_2$, (12.5), and (12.6), we get

$$\begin{aligned} m_1 R_1 v_1 + m_2 R_2 v_2 &= (R_1 + R_2) m_1 v_1 = D m_1 v_1 \\ &= D m_1 R_1 \omega_2 = D^2 \frac{m_1 m_2}{m_1 + m_2} \omega_2. \end{aligned}$$

⁸By (12.6) we get $R_1 = 4672 \text{ km}$, and thus $R_2 \approx D$ and the center of gravity of the Earth–Moon system is located inside the Earth.

From this we get by differentiating (12.7) with respect to time and neglecting $I_2\dot{\omega}_2$ (cf. (12.17)) that

$$\begin{aligned} I_1\dot{\omega}_1 &= -\frac{m_1m_2}{m_1+m_2} \frac{d(D^2\omega_2)}{dt} = -\frac{m_1m_2}{m_1+m_2} (\dot{\omega}_2 D^2 + 2\omega_2 D\dot{D}) \\ &= -\frac{m_1m_2}{m_1+m_2} \frac{\omega_2 D}{2} \dot{D}, \end{aligned} \quad (12.18)$$

where the last equality follows from (12.15). Substituting from (12.1), (12.4), (12.8), and (12.14), we finally obtain

$$\dot{D} = 0.674 \cdot 10^{-9} \text{ m/s}, \quad (12.19)$$

which is only slightly more than one half of the measured value (12.16).

Multiplying (12.19) by the length of the sidereal year T , we find that the Earth–Moon distance should increase by only about the value

$$v_{\text{tides}} \approx 2.13 \text{ cm per year}. \quad (12.20)$$

Denote the difference between the measured value from (12.2) and the value (12.20) by

$$v_{\text{remainder}} \approx 1.71 \text{ cm per year}, \quad (12.21)$$

that is,

$$v = v_{\text{tides}} + v_{\text{remainder}}.$$

The origin of the term $v_{\text{remainder}}$ is unknown, but its magnitude is only a little bit smaller than the current value of the Hubble constant in (12.3). This supports the hypothesis concerning the local influence of antigravity. More precisely, the speed 1.71 cm per year from (12.21) equals 67% of the value 2.57 cm per year of the Hubble constant, as given in (12.3), i.e., the local Hubble expansion is

$$\boxed{H_0^{(\text{loc})} = 0.67 H_0.} \quad (12.22)$$

Such a large recession does not seem to be caused by tidal forces nor other nongravitational effects, as we shall see below and in further chapters.

In 2003, Yurii Dumin (see [64, p. 2463]) derived from astrometric measurements corrected to ancient eclipses a similar local expansion rate of $H_0^{(\text{loc})} \approx 0.5 H_0$ for the Earth–Moon system. In his paper [65] from 2008 this value was increased to $H_0^{(\text{loc})} \approx 0.85 H_0$ for the data from the last three centuries.



12.5. A time dependent momentum of inertia of the Earth

Geophysicists admit (see e.g. [195], [287]) that it is not easy to explain the large discrepancy between the observed value (12.16) and value (12.19) theoretically derived from tidal forces. Therefore, we shall look for some other source which would produce values on the same order of magnitude as tidal forces for the recession of the Moon from the Earth. For example, Oldřich Novotný [195, p. 67] considers a time variable momentum of inertia $I_1 = I_1(t)$ in which the left-hand side of equation (12.18) is replaced by the derivative $d(I_1\omega_1)/dt$. To compensate the difference between (12.16) and (12.19) during the last 2700 years from the first observations of ancient Babylonian astronomers (cf. (12.12)) there should exist a permanent mass flux to the center of the Earth which would guarantee that the size of $-\dot{I}_1$ is of the order 10^{20} to 10^{21} kg m²/s.

Let us now examine in detail how such a conclusion can be made. Assume that the observed time change (12.16) is caused by a variable momentum of inertia of the Earth. Then instead of (12.18) we shall consider the equation

$$\frac{d}{dt}(I_1\omega_1) = \dot{I}_1\omega_1 + I_1\dot{\omega}_1 = -\frac{m_1m_2}{m_1+m_2} \frac{\omega_2 D}{2} \dot{D}_{\text{observed}}. \quad (12.23)$$

Subtracting (12.18), we get

$$\dot{I}_1\omega_1 = -\frac{m_1m_2}{m_1+m_2} \frac{\omega_2 D}{2} (\dot{D}_{\text{observed}} - \dot{D}). \quad (12.24)$$

By dividing ω_1 and substituting from (12.1), (12.4), (12.8), (12.16), and (12.19), we obtain

$$\dot{I}_1 = -2.686 \cdot 10^{20} \text{ kg m}^2\text{s}^{-1}. \quad (12.25)$$

Such a big change in the momentum of inertia is not realistic and cannot be explained by any simple process. The change corresponds to an immense flow of mass to the center of the Earth lasting for at least 2700 years, which is quite unlikely. Therefore, considering a time-independent momentum of inertia of the Earth, we will not commit a large modeling error in deriving (12.22).

There are a variety of nongravitational forces that lead to a change in the Earth–Moon distance, such as solar wind, thermal radiation of the Earth and the Moon, the Yarkovsky effect, collisions with interplanetary dust and meteorites, and the presence of magnetic fields. Their influence, however, seems to be negligible in comparison to the effects of tides.

Let us mention several other hypotheses that try to explain the relatively large recession speed of the Moon from the Earth (see e.g. [64], [65], [67], [131]). This relatively rapid expansion is attributed to vacuum energy in [67], where it is assumed that the speed of gravitational interaction is infinite. Relativistic effects and also a possible dependence of the gravitational constant on time is investigated in [171]. Nevertheless, the orbital velocities $v_1 = 12.5 \text{ m/s}$ and $v_2 = 1020 \text{ m/s}$ from (12.7) are so small that relativistic effects cannot be clearly manifested.



12.6. The paradox of the large orbital angular momentum of the Moon

By tidal forces alone it is difficult to explain the current very large and paradoxical orbital angular momentum of the Earth–Moon system (see [20, p. 534], [120]). If we admit the action of antigravitational

forces, we obtain an additional shift (12.21) in the recession speed of the Moon from Earth. According to [120], our Moon formed at a distance of only $d = 20\,000$ km from the Earth approximately 4.5 billion years ago.⁹ This corresponds to the mean recession speed

$$\bar{v} = 8 \text{ cm per year}$$

to reach the current distance (12.1). Because tidal forces decrease with the third power of the distance, tidal forces played a significant role especially at the beginning. However, antigravity could slightly but permanently increase the recession speed of the Moon from the Earth within the whole 4.5 Gyr long interval.



⁹In this case, solar and lunar eclipses would have been much more frequent than at present, since the angular diameter of the Moon was about $10^\circ \approx 2 \arcsin(r_2/d)$. The Moon was shining extremely brightly at that time, since its area was 400 times larger than now. Its orbital period was less than 6 hours by Kepler's third law (4.4). Moreover, tidal forces were huge.

13. Recession of the Earth from the Sun

*Give me a lever and a place to stand
and I will move the Earth.*

ARCHIMEDES

13.1. The faint young Sun paradox

Let us first recall the faint young Sun paradox. Assume for a moment that the Earth was at its birth about 1 au away from the Sun as it is now. In order to not be completely frozen and to be able to develop life on its surface, the Sun at its origin had to be approximately as hot as it is now. However, this is not in agreement with the fact that the Sun as a star on the main sequence of the Hertzsprung–Russell diagram had a lower luminosity at its origin (cf. Figs. 11.1 and 11.2). This leads to the paradox usually referred as the *faint young Sun paradox* [155]. The mean temperature on the surface of Earth would have been much below the freezing point of water, in contrast with the absence of glaciation in the first 2.7 Gyr (see [20, p. 177]).

The faint young Sun paradox is, in fact, more severe due to ice-albedo feedback of the frozen ocean. To prevent the Earth from freezing over, a much higher concentration of CO₂ than today is assumed in [117]. This oxygen is at present deposited in CaCO₃.

If Mars were much closer to the Sun by tens of millions of kilometers (see Chapter 11), then our Earth would also have to be closer to the Sun. Otherwise these planets could have close encounters and their orbits would not be stable over time. In the following sections we present three further independent arguments suggesting that the

Earth was much closer to the Sun in the past and that their average distance from each other increases by a speed of several meters per year. Assuming that this is due to antigravity, the faint young Sun paradox ceases to be a mystery.



13.2. The expansion of the ecosphere

The Earth is located inside the so-called habitable zone, i.e. the area in which liquid water is present permanently. In this section and mainly in Section 14.2 we estimate such an expansion speed of the Earth from the Sun to ensure a constant solar energy flux for 3.5 billion years, which would provide favorable conditions for the development of life on our planet.

So far, unfortunately, we cannot measure the actual average expansion speed of the Earth from the Sun with an accuracy on the order of a meter per year, because the position of the center of gravity of the Solar system changes by thousands of kilometers each year as a result of the gravitational influence of the giant planets, see (5.1) and Fig. 13.1. Therefore, we will consider the very long time interval of 3.5 billion years of life on the Earth.

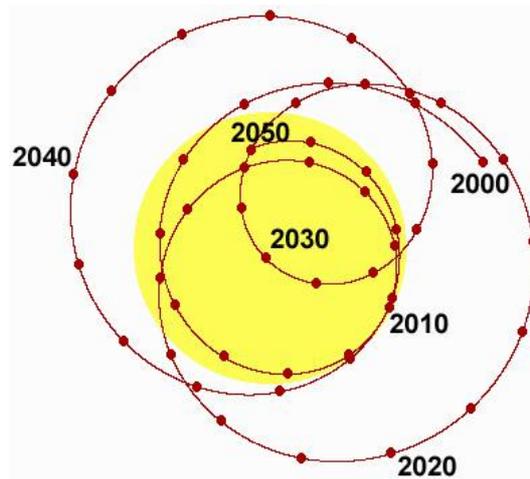


Fig. 13.1. The trajectory of the Solar system barycenter during the period 2000–2050. The diameter of the Sun is almost 1 400 000 km. The barycenter shifts about 1 000 km per day along a curved trajectory. Hence, the heliocentric coordinate system is not inertial.

To ensure favorable conditions for life on Earth it is necessary at present that the Sun's luminosity be at most 5% larger or smaller than the solar constant

$$L_0 = 1.36 \text{ kWm}^{-2}. \quad (13.1)$$

Such a ring (resp. spherical layer) is called an *ecosphere*. Since the energy flux from the Sun decreases with the square of the distance, its radii are $(0.95)^{1/2}$ au and $(1.05)^{1/2}$ au, which corresponds to a very narrow interval of 145.8–153.3 million km (see Fig. 13.2). If the Earth's elliptic orbit were to leave this ring for a long time period, it would have disastrous consequences for life on our planet. A permanent reduction of the Sun's luminosity of more than 5% would cause overall glaciation of the planet. On the other hand, at temperatures over 57°C DNA sequences of certain multicellular organisms decay. Even at temperatures above 45°C mitochondria stop producing adenosine triphosphate (ATP) that supplies chemical energy to eukaryotic cells.

When life appeared on Earth (i.e. 3.5 billion years ago), the Sun's luminosity was about 77% of the present value (see Fig. 11.2). To ensure a favorable climate for the long-term evolution of life, where liquid water is necessary, the Earth was probably tens of millions of kilometers closer to the Sun at that time. This claim is supported, for example, by data on the occurrence of fossil thermophilic bacteria, from which it is believed [163] that the temperature of oceans was about 80°C three and half Gyr ago. Due to higher volcanic activity at that time it is, however, not clear to what extent this may only be a selection effect. On the other hand, formula (11.9) yields a quite large discrepancy in temperatures, namely, for given a total solar luminosity of $0.77L_{\odot}$, $r = 1$ au, and a present value for the mean Earth's albedo of $A = 0.306$, we get the temperature -35°C .

Heat produced by decay of radioactive isotopes also contributed to a higher temperature during the first few hundred million years of the Earth's existence, when elements with a relatively short half-life decayed. However, the Earth's surface cooled significantly over the

first half billion years due to the radioactivity decay of these elements. According to current measurements, the heat flow from the Earth is less than 0.1 W/m^2 . It could not have been much larger 4 Gyr ago, because the half-lives of the contemporary natural radioactive isotopes ^{232}Th , ^{238}U , ^{40}K , and ^{235}U are 13.9, 4.468, 1.248, and 0.704 Gyr, respectively [60]. It is therefore estimated (see [85, p. 58]) that the heat flow from the Earth 4 Gyr ago was no more than 5 times larger than today. Nevertheless, this heat flow (mainly from various geothermal sources) is completely negligible when compared to the solar constant (13.1).

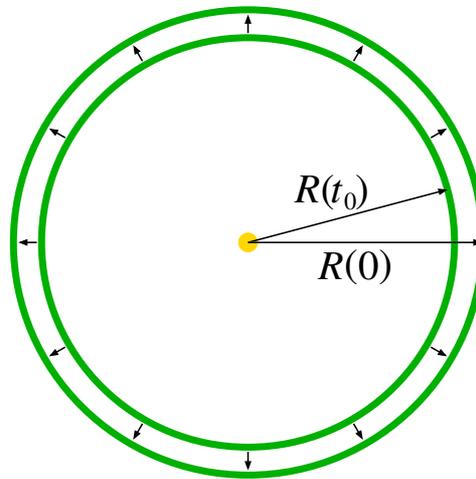


Fig. 13.2. Schematic illustration of the expansion of the ecosphere during the last 3.5 Gyr, where $R(t_0) = 1.3 \cdot 10^{11} \text{ m}$, $R(0) = 1.5 \cdot 10^{11} \text{ m}$, and $t_0 = -3.5 \text{ Gyr}$.

Although the Sun's power grew approximately as illustrated in Fig. 11.2, the Earth should have relatively stable conditions that are needed for the development of life during the period lasting 3.5 Gyr. In Section 14.2 we show that for the average speed¹ of the Earth moving away from the Sun

$$\bar{v} = 5.2 \text{ m per year}, \quad (13.2)$$

¹A large expansion velocity (13.2) would cause problems with the original definition of the astronomical unit. The newly-established definition of the astronomical unit au is independent on the Earth-Sun distance (see (4.6)).

the Earth would receive an almost constant density of flux of energy comparable with the solar constant (cf. (13.1))

$$L(t) = 1.36 \pm 0.005 \text{ kW m}^{-2}$$

for all t in the interval of length 3.5 Gyr. The speed (13.2) is optimal in this sense. Nevertheless, we do not claim that the Earth recedes from the Sun just by this speed, but with a speed whose size has order comparable with (13.2). The speed (13.2) by the relation

$$H_0 \approx 10 \text{ m yr}^{-1} \text{ au}^{-1}, \quad (13.3)$$

which was derived in Chapter 11 (cf. (11.2)), corresponds to the mean local Hubble expansion

$$\boxed{H_0^{(\text{loc})} = 0.52 H_0.} \quad (13.4)$$

Due to a similarity with the order of the Hubble constant, we propose that this recession originates from local antigravity effects of dark energy.

Antigravity forces thus cause secular migration of our planet of order meters per year, so that it stays within the expanding ecosphere (see Fig. 13.2). If antigravity were not to act, then conditions favorable for the development of life on Earth would exist for only about one billion years. Intelligent life would not have had enough time to develop due to a continual rise in temperatures (cf. Fig. 11.2, [133], and [146]).



13.3. Analysis of growth patterns on fossil corals from solar data

In this section we present the method proposed by Weijia Zhang et al. [302]. The present value of the sidereal year is

$$Y = Y(0) = 31\,558\,149.54 \text{ s} = 365.25636 \cdot 24 \cdot 3600 \text{ s}. \quad (13.5)$$

However, the length of the sidereal year in seconds in ancient time was

$$Y(t) = n(t)(24 \cdot 3600 - f(t)) \quad \text{for } t \leq 0, \quad (13.6)$$

where $(-t)$ is the geological age in years and an ancient day was shorter about $f = f(t) > 0$ seconds, $t = 0$ corresponds to the present time, $f(0) = 0$, and $n(t)$ is the number of ancient days per year which is known from paleontological data by means of calculating the number of layers deposited during one year in fossil corals. Namely, each coral increases during one day by a few microns, more in summer and less in winter. Examining data for several consecutive years (e.g. layers that arose during twelve years are investigated in [200]), allows us to minimize the error in determining the number of days in a year. Hundreds of patterns were examined by microscope e.g. in [302, pp. 4013–4014].

In particular, for the Devonian era Zhang et al. [302] found that $n(\tau) \approx 405$ days for $\tau = -370 \cdot 10^6$ years ago, but those days were shorter than at present. A similar value of about 400 days can be found in the seminal paper by J. W. Wells [296, p. 949] from the seventies.

Due to larger tidal forces when the Moon was closer to the Earth and the Earth was closer to the Sun, the function f is decreasing. Note that tidal forces decrease cubically with distance, see [20, p. 96]. According to [302, p. 4014], $\mathcal{T}(\tau) = 2.6 \cdot 10^{-5}$ s per year, whereas the present value is $\mathcal{T} = \mathcal{T}(0) = 1.7 \cdot 10^{-5}$ s per year (see (2.6)). It was measured with respect to some fixed quasars at cosmological distances. The Earth's rotational history (paleorotation) is examined also in [8], [200], and [297]. Substituting the above data into (13.6) as in [302], we get

$$Y(\tau) = 405(24 \cdot 3600 - 2.6 \cdot 10^{-5} \cdot 370 \cdot 10^6) = 405 \cdot 76780 = 31\,095\,900 \text{ (s)}, \quad (13.7)$$

i.e., the day in the Devonian era had about 76 780 seconds ≈ 21.327 hours (cf. (13.5)).

Now denote by $R(t)$ the semimajor axis Earth's orbit at time t . Note that Kepler's laws are not reliable over long time periods due to resonances, local manifestations of dark energy, and other effects.

Nevertheless, for a given time t Kepler's third law

$$\frac{R^3(t)}{Y^2(t)} = \frac{GM_\odot}{4\pi^2} \quad (13.8)$$

describes reality quite well. Here $G = 6.674 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$ is the gravitational constant and

$$M_\odot = 1.989 \cdot 10^{30} \text{ kg} \quad (13.9)$$

is the Sun's mass. The present value of M_\odot differs from the value corresponding to the Devonian era by an amount of only about 0.003 % (see (13.19) below). Inserting (13.7) into (13.8) for $t = \tau$, we obtain the length of the semimajor axis of Earth's orbit in the Devonian era

$$R(\tau) = \left(\frac{Y^2(\tau)GM_\odot}{4\pi^2} \right)^{1/3} = 148.1 \cdot 10^9 \text{ m.}$$

This yields the following average recession speed of the Earth from the Sun for $R(0) = 149.6 \cdot 10^9 \text{ m}$,

$$\bar{v} = \frac{R(\tau) - R(0)}{\tau} = \frac{(149.6 - 148.1) \cdot 10^9}{370 \cdot 10^6} = 4 \text{ (m/yr)},$$

which has the same order of magnitude as that in (13.2) and by (13.3) this leads to the local Hubble expansion

$$\boxed{H_0^{(\text{loc})} = 0.4 H_0.} \quad (13.10)$$



13.4. Analysis of growth patterns on fossil corals from lunar data

From Chapter 12 we know that the present mean Earth–Moon distance equals

$$D = 384.402 \cdot 10^6 \text{ m} \quad (13.11)$$

and that the present mean recession speed $v = v(t)$ of the Moon from the Earth is

$$v(0) = 3.84 \text{ cm/yr.} \quad (13.12)$$

Let $P = P(t)$ be the length of the sidereal month and $s = s(t)$ the number of sidereal months per year. At present it is $P(0) = 27.322$ days and $s(0) = 13.368$. The number $s(t)$ is known from paleontological data for many negative t 's, since s equals one plus the number of lunar months. The number of lunar months can be manually calculated from the study of many growth patterns on coral fossils from the full moon to the next full moon (see Fig. 13.3 and [302, p. 4012]).

In the Cambrian era 500 Myr ago, the Moon was about 20 000 km closer to the Earth than it is now for the mean recession speed 4 cm/yr extrapolated from [63] (see also Section 13.5). Its spatial angle was more than 10 % larger than it is now and thus lunar patterns are better visible on fossil corals. In particular,

$$s(\tau) \approx 14.2$$

of sidereal months for $\tau = -5 \cdot 10^8$ years according to [302, p. 4013].

Using the generalized Kepler's third law (4.5) for the Earth-Moon system, we obtain the following length of the year in time t :

$$Y(t) = s(t)P(t) = s(t) \left((D + v(t)t)^3 \frac{4\pi^2}{G(M + m)} \right)^{1/2}, \quad (13.13)$$

where

$$M = 5.9736 \cdot 10^{24} \text{ kg} \quad \text{and} \quad m = 7.349 \cdot 10^{22} \text{ kg} \quad (13.14)$$

are the masses of the Earth and Moon, respectively, and $v(t)$ is their recession speed. Since tidal forces decrease with the third power of distance, the function $v = v(t)$ is decreasing with time.

From this, relations (13.8), (13.13), (13.14), (13.11), and (13.12) we obtain for $t = \tau = -5 \cdot 10^8$ yr the following upper estimate for the Earth-Sun distance in the Cambrian era

$$\begin{aligned} R(\tau) &= \left(Y^2(\tau) \frac{GM_\odot}{4\pi^2} \right)^{1/3} = s(\tau)^{2/3} \left(\frac{M_\odot}{M + m} \right)^{1/3} (D + v(\tau)\tau) \\ &< 14.2^{2/3} \cdot 328919^{1/3} (384.402 \cdot 10^6 + v(0)\tau) = 147.8 \cdot 10^9 \text{ (m)}, \end{aligned}$$

since $v(\tau) > v(0)$. This yields the following guaranteed lower bound for the average recession speed of the Earth from the Sun:

$$\bar{v} = \frac{R(\tau) - R(0)}{\tau} > \frac{(149.6 - 147.8) \cdot 10^9}{5 \cdot 10^8} = 3.6 \text{ (m/yr)}.$$

In this case we get from (13.3) that

$$\boxed{H_0^{(\text{loc})} > 0.36 H_0.} \quad (13.15)$$

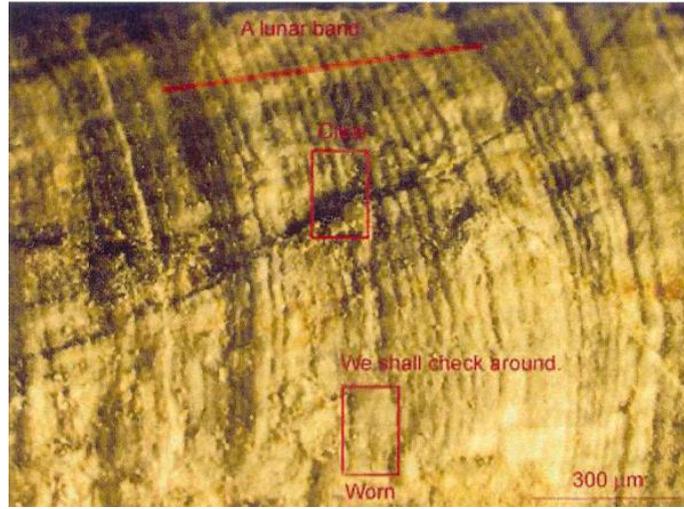


Fig. 13.3. The method by [302] for determining the length of the lunar month between two consecutive full moons by means of increments of Paleozoic corals. Instead of the lower rectangle, where the structure of daily increments is unclear, the strips are counted in the upper rectangle.

For calculation of the period $P(t)$ in (13.13) Zhang et al. [302] also used Kepler's third law. By a thorough analysis of growth patterns on fossil corals from lunar data (which are independent of solar data) Zhang et al. in [302, pp. 4013–4016] got further values of $s(t)$ for other time epochs t leading to the local Hubble expansion

$$\boxed{H_0^{(\text{loc})} = 0.57 H_0.} \quad (13.16)$$

Yurii Dumin in his paper [66] proposes an even larger value, namely $H_0^{(\text{loc})} = 0.8 H_0$, provided the solar flux has increased about 50 % from its initial value 4.5 Gyr ago.



13.5. Consequences for the variable Earth-Moon distance

Growth patterns on fossil corals from lunar data enable us also to detect long-term secular changes of the Earth-Moon distance. According to (13.13), the current orbital period of the Moon is

$$P(0) = \frac{Y(0)}{s(0)} = \frac{1}{13.368} \text{ yr},$$

i.e. the present sidereal years. However, the sidereal year $Y(\tau)$ in the Cambrian era for $\tau = -5 \cdot 10^8$ yr was slightly shorter than it is now. From the previous section we find by Kepler's third law that its length satisfies

$$Y(\tau) = Y(0) \left(\frac{R(\tau)}{R(0)} \right)^{3/2} < 1 \text{ yr} \cdot \left(\frac{147.8}{149.6} \right)^{3/2} = 0.982 \text{ yr}.$$

By (13.13) the orbital period of the Moon in the Cambrian era was $P(\tau) = Y(\tau)/s(\tau) < 0.982/14.2$ yr. From this, (13.11), and Kepler's third law again, we may estimate the radius of Moon's orbit $D(\tau)$ in the Cambrian era as follows:

$$D(\tau) = D \left(\frac{P(\tau)}{P(0)} \right)^{2/3} < 364.8 \cdot 10^6 \text{ m}.$$

Hence, the long-term average secular recession speed of the Moon from the Earth satisfies

$$\tilde{v} = \frac{D - D(\tau)}{|\tau|} > 3.92 \text{ cm/yr},$$

which is in good agreement with the present value given in (13.12).



13.6. Prolongation of the sidereal year of the Earth

In Sections 13.1–13.4 we presented several indirect but mutually independent arguments showing that the mean value of the semimajor axis of the Earth's orbit increases about several meters per year.

Another effective tool to determine the average recession speed of the Earth from the Sun could be systematic measurements of the length of the sidereal year. But it is necessary to make measurements over a long-term period and very accurately. During a few years such small changes cannot be reliably detected, since from Fig. 13.1 we know that the Newtonian barycenter of the Solar system travels hundreds of thousands km per year with respect to the Sun due to the influence of large planets. Moreover, the ellipticity of the orbit of the “double planet” Earth–Moon is slightly perturbed by the Sun, Jupiter, Venus, and other bodies.

Assume that the semimajor axis $R = 1$ au of Earth’s orbit increases on average about $\Delta R = 5.2$ m per sidereal year (cf. (13.2)). Then from Kepler’s third law

$$\frac{(R + \Delta R)^3}{(Y + \Delta Y)^2} = \frac{R^3}{Y^2}$$

we can easily find that the increase of the orbital period of the Earth after one year would be only $\Delta Y = 1.6$ ms. In particular,

$$Y^2(R^3 + 3R^2\Delta R + \dots) = R^3(Y^2 + 2Y\Delta Y + \dots).$$

Neglecting higher order terms in ΔR and ΔY , we observe by (11.1) and (13.5) that the sidereal year increases every year about

$$\Delta Y \approx \frac{3Y}{2R}\Delta R = 0.0016 \text{ s.} \quad (13.17)$$

Such a small time change also cannot be reliably detected, as explained in Fig. 13.1. The increase of the orbital period by about $\Delta Y = 1.6$ ms would require one additional second after 35 years, since after two years we have to add $2\Delta Y$ to the orbital period, after 3 years $3\Delta Y$, and so on. From this we get

$$(1 + 2 + \dots + 35)\Delta Y = (1 + 35)\frac{35}{2} \cdot 0.0016 \approx 1 \text{ (s)}.$$

This makes the evidence of a slightly increasing orbital period very difficult to obtain. The increase of the sidereal year will not be notice-

ably apparent until several decades have passed. We must not forget that this is only an average value.

Note that 1 or 2 leap seconds are added almost every year to the calendar year² to compensate for the slowing down of decreases in the Earth's rotation speed.³ The above value (13.17) was derived under the assumption (13.2). However, reality can be a little bit different. For example, relations (13.5) and (13.7) imply that the sidereal year increases by about

$$\Delta Y = \frac{Y(\tau) - Y(0)}{\tau} = \frac{31\,558\,149.54 - 31\,095\,900}{370 \cdot 10^6} = 0.00125 \text{ (s)}.$$

Another way to determine the recession speed of the Earth from the Sun could be to measure angular distances. Although we can currently measure angles with an accuracy of a millionth of a second arc, to derive the Earth's recession by measuring the angular diameter of the Sun is problematic. The boundary of the Sun is not sharp and the reduction of the angular diameter by an amount of about 0.000 001'' corresponds to an increase in the distance of 176 meters. Furthermore, the speed of the increase of the Sun's diameter in meters per year is not precisely known.



13.7. Elimination of other possibilities for the large recession speed

In this section, we show that the reduction of the Solar mass (as a result of nuclear reactions, ejection of plasma jets, and solar wind),

²The *Gregorian calendar year* has 365.2425 days. It is almost equal to the *tropical year* of 365.24219 days, which is the time between two successive passages of the Sun through the *vernal equinox*, i.e. one of the two intersections of the ecliptic and the celestial equator. The tropical year has 31 556 925.4 seconds and is therefore about 20 minutes shorter than the sidereal year (13.5). As a result of the precession of Earth's axis, the vernal equinox shifts along the ecliptic about 50.27'' relative to the stars during one tropical year.

³During earthquakes it may occasionally happen that the Earth's rotation is slightly accelerated, because the Earth reduces its moment of inertia or its axis of rotation is shifted.

tidal forces, magnetic fields, and so on, has only a negligible influence on the Earth's recession from the Sun. Without the presence of antigravitational forces, it is not easy to deduce the relatively high speed (13.2) of recession. Nevertheless, local effects of antigravity (see (13.4), (13.10), (13.15), and (13.16)) can again explain this paradox.

The nucleus of helium is 0.7 % lighter than the nuclei of 4 hydrogen atoms, each of which consists of a single proton. This means that at most 0.7 % of the Sun's mass changes into energy during 10 Gyr (the estimated life period of the Sun). When the Sun was born, it already contained about 30 % of helium. Hydrogen changes into helium only in the central parts of the Sun and by the end of its time on the main sequence of the Hertzsprung–Russell diagram (see Fig. 11.1) the Sun will still contain a lot of hydrogen. Thus it is assumed that only 0.07 % of the Sun's mass will change into energy [194]. Since matter changes into energy, the actual solar mass loss is by (13.5) and (13.9) on average only $0.0007M_{\odot}/(10^{10} \cdot Y(0)) = 4.46 \cdot 10^9$ kg per second. The nuclear reactions represent the main contribution to the total mass loss.

The Sun also loses its mass partly due to plasma outbursts. If the speed of a solar plasma outburst is larger than 613 (resp. 434) km/s, then by (13.9) and (4.15) (resp. (4.12)) plasma can escape the Solar system (resp. Sun) which reduces the Sun's mass as well. For smaller speeds plasma falls back onto the Sun. Occasionally, comets and other bodies fall into the Sun, which in turn increases its mass.

According to Noerdlinger [194], every second the Sun loses altogether $5.75 \cdot 10^9$ kg of its mass due to nuclear reactions, solar wind, electromagnetic radiation, neutrino losses, and large eruptions. Taking into account that the mass loss during one year (see (13.5)) is $1.815 \cdot 10^{17}$ kg/yr, we find by (13.9) that

$$\frac{\dot{M}_{\odot}}{M_{\odot}} = \frac{1.815 \cdot 10^{17}}{1.989 \cdot 10^{30}} = -9.13 \cdot 10^{-14} \text{ (yr}^{-1}\text{)}, \quad (13.18)$$

where dot stands for the time derivative. As theoretically derived in [194], the radii of planetary orbits expand at the same rate. Hence,

the average recession speed of the Earth from the Sun due to the radiative and particle loss of the Sun's mass is by (13.18) approximately

$$-\frac{\dot{M}_\odot}{M_\odot}R = 9.13 \cdot 10^{-14} \text{ yr}^{-1} \cdot 149.6 \cdot 10^9 \text{ m} = 0.014 \text{ m/yr},$$

which is a much smaller expansion speed than that given in Section 13.2 (cf. e.g. (13.2)). The solution of differential equation (13.18) has the form

$$M_\odot(t) = M_\odot e^{-Ct},$$

where $C = 9.13 \cdot 10^{-14} \text{ yr}^{-1}$. Changes of the Sun's mass are thus negligible. For instance, if $t = -370 \cdot 10^6 \text{ yr}$ (which corresponds to the Devonian era), we find that

$$M_\odot(\tau) = 1.989067 \cdot 10^{30} \text{ kg}, \quad (13.19)$$

if magnetohydrodynamic effects are ignored. This represents rather an upper estimate of the real mass due to the lower power of the Sun in the Devonian era.

The solar radiation pressure on the Earth also cannot explain a significant part of the recession of the Earth from the Sun. The energy that comes to us from the Sun per year⁴ is equal to

$$E = SYL_0 = 5.4 \cdot 10^{24} \text{ J},$$

where $S = \pi(6.371 \cdot 10^6)^2 \text{ m}^2 = 1.275 \cdot 10^{14} \text{ m}^2$ is the maximum area of the cross section of our Earth, Y is the sidereal year from (13.5), and L_0 is the solar constant from (13.1). Further, denote E_i , λ_i , ν_i , and p_i to be respectively the energy, wave length, frequency, and momentum of the i th photon. Then we have

$$p_i = \frac{h}{\lambda_i} = \frac{h\nu_i}{c} = \frac{E_i}{c},$$

⁴By (11.5) we know that the total power of the Sun is equal to $L_\odot = 4\pi R^2 L_0 = 3.8 \cdot 10^{26} \text{ W}$, where $R = 1 \text{ au}$. From this and Einstein's formula $E = mc^2$, the Sun's mass losses are $4.23 \cdot 10^9 \text{ kg/s}$ due to nuclear reactions. According to (13.9) it follows that $L_\odot/M_\odot = 0.0002 \text{ W/kg}$. Although the nuclear reactor is located only in the central part of the Sun, it has much less power per unit mass than e.g. the human body which produces about 1 W/kg .

where $h = 6.626\,0693 \cdot 10^{-34}$ Js is Planck's constant and c is the speed of light. Summing the above equation over all photons coming to the Earth from the Sun during one year, we get

$$p = \sum_i p_i = \frac{E}{c} = \frac{5.4 \cdot 10^{24}}{3 \cdot 10^8} = 1.8 \cdot 10^{16} \text{ (kg m/s)}.$$

If the Earth were to absorb all photons from the Sun intersecting the Earth's surface, then from (13.14) we get

$$v = \frac{p}{M} = 0.095 \text{ m/yr},$$

which is a much smaller speed than that given in (13.2).

Tidal forces also do not explain an essential part of the recession speed of the Earth from the Sun. From Chapter 12 we know that they have a small influence on the slowing down of the Earth's rotation — about 68.5 % is caused by the Moon and only 31.5 % by the Sun, see [41]. The influence of other planets is negligible. Tidal forces from the Sun per 1 kg of the Earth are approximately equal to $2GM_{\odot}r/R^3$, where the mass M_{\odot} of the Sun is given by (13.9), $R = 1$ au, and r is the Earth's radius. We see that they decrease with the third power of the Earth–Sun distance. From this it can be derived (see e.g. [20, p. 606], [195]) that the Earth–Sun distance increases by about only a few cm per year due to tidal forces.

The Earth moves in the Sun's magnetic field. Since the Earth has a large iron core, circulating eddy (Foucault) currents should appear and thus the Earth should descend onto lower orbits. By previous sections this is not observed. The reason is that the magnetic potential decreases⁵ as r^{-2} (whereas the gravitational potential decreases as r^{-1}). The interaction between the magnetic fields of the Earth and the Sun occurs mostly in the terrestrial magnetosphere (the plasma envelope around the Earth, whose size is on the order of 10 Earth's

⁵Therefore, close binary neutron stars can immediately collide after a gradual approach, since the attractive magnetic force may suddenly exceed the centrifugal force. This then triggers a strong gamma ray burst.

radii). Hence, the magnetic force has almost no influence on secular changes of Earth's orbit. Moreover, the Sun reverses its polarity every 11 years, so the possible errors due to the reversal of magnetic polarity are not accumulated but canceled.

For small bodies of the Solar system the Yarkovsky and YORP effect [20] change their trajectories and rotations, respectively. However, these effects are entirely negligible in the case of Earth. The same is true for interplanetary dust and other nongravitational forces.



13.8. Why other authors obtained much smaller values for recession speeds

There are some discrepancies between our results and the results of other authors. Now we will explain why.

G. A. Krasinsky and V. A. Brumberg [125] derived that the present recession speed of the Earth from the Sun is equal to $v = 15$ cm/yr. Their calculation is based on the assumption that the Newtonian theory of gravitation describes all motions in the Solar system absolutely exactly. They solve an algebraic system for 62 unknown Keplerian parameters (see Section 3.4) of all planets and some large asteroids and do not take into account small antigravitational forces. In other words they implicitly assume that modeling, discretization, and rounding errors are negligible. However, classical Newtonian theory assumes an infinite speed of gravitational interaction, whereas the real speed is surely finite. Hence, the modeling error is surely not zero.

Some authors (see e.g. [45], [55], [172]) claim that dark energy has no influence on the expansion of the Solar system. Now we show where these authors have used a faulty argumentation. The Hubble parameter $H = H(t)$ which describes the expansion rate is defined by the relation

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \tag{13.20}$$

where $a = a(t)$ is the expansion function and the dot denotes the time derivative. For instance, if a global model of the universe had a constant positive curvature at all points and in all directions at any given time t , then the value of the expansion functions $a(t)$ would be equal to the radius of the universe at time t (see e.g. G. Lemaître [159]). In this case, the universe is modeled by the three-dimensional surface of a four-dimensional inflating ball (see Section 10.4). Differentiating (13.20) with respect to time, we get

$$\dot{H} = \frac{\ddot{a}}{a} - H^2 = -qH^2 - H^2, \quad (13.21)$$

where $q = -\ddot{a}/\dot{a}^2$ is the dimensionless deceleration parameter which characterizes deceleration or acceleration of the expansion of the universe. Expressing the expansion function $a = a(t)$ as a Taylor series in time $t = 0$, which corresponds to the present time, we have (see Fig. 13.4)

$$a(t) = a(0) + \dot{a}(0)t + \frac{1}{2}\ddot{a}(0)t^2 + \dots = a(0)(1 + H_0t - \frac{1}{2}q_0H_0^2t^2 + \dots), \quad (13.22)$$

where $H_0 = H(0)$ and $q_0 = q(0) = -0.6$ is the usually accepted value of the deceleration parameter (see [229, p.110]) which is negative, since the expansion of the universe accelerates.

M. Carrera and D. Giulini [45, p. 175] correctly derive that at the distance of Pluto (i.e. about 40 au) the acceleration of the expansion of the universe is only $2 \cdot 10^{-23}$ m/s² which is indeed an entirely negligible quantity. Similar tiny values were derived by F. I. Cooperstock et al. [55, p.62] and B. Mashhoon et al. [172, p.5041]. However, all these authors concentrated only on the single quadratic term in expansion (13.22) and did not consider the large value of the Hubble constant (13.3) which stands at the linear term in (13.22). In other words, an accelerated expansion does not manifest itself on scales of the Solar system, but the expansion itself is observable. In particular, we have

$$|H_0t| \gg \frac{1}{2}|q_0|H_0^2t^2$$

for t close to 0. Consequently, the accelerated expansion given by the quadratic term only appears at cosmological distances. In spite of that, the single quadratic term is so small that the linear term $|H_0 t|$ from (13.22) essentially dominates not only in the neighborhood of 0, but for all t in the whole interval $(-1/H_0, 0)$, since we have

$$0.3 \cdot |H_0 t| > \frac{1}{2} |q_0| H_0^2 t^2,$$

where $\frac{1}{2} |q_0| = 0.3$.

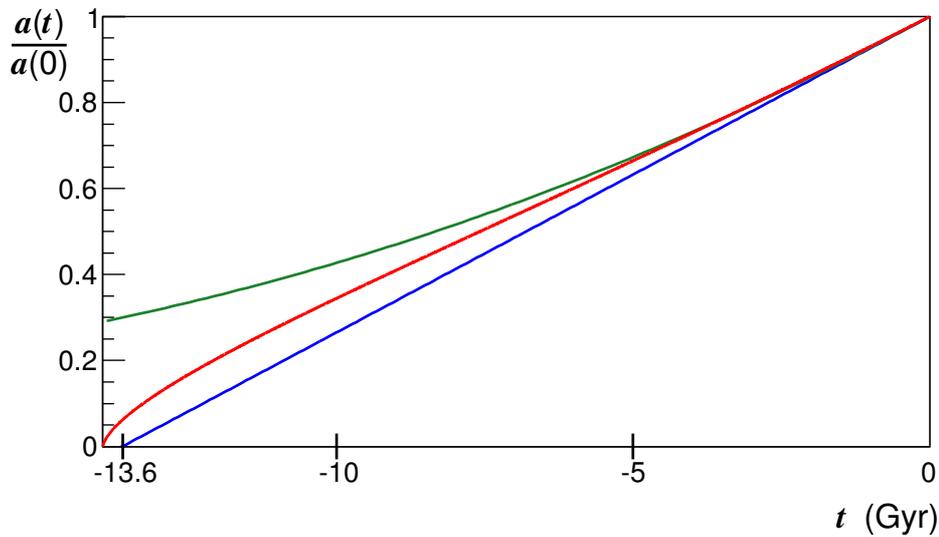


Fig. 13.4. The lower graph corresponds to the linear function $1 + H_0 t$ from the Taylor expansion (13.22) on the interval $[-1/H_0, 0]$, where $1/H_0 = 13.6$ Gyr is the Hubble time. The upper graph shows the quadratic function defined by the first three terms of the Taylor expansion $1 + H_0 t - \frac{1}{2} q_0 H_0^2 t^2$ with $q_0 = -0.6$. The middle graph illustrates the normalized expansion function $a(t)/a(0)$ obtained by integration of (13.20). The values on the horizontal axis are given in Gyr. The quantities on the vertical axis are relative with no physical dimensions. We observe that the accelerated expansion differs little from the linear expansion during the last few Gyr.

Without dark energy the expansion of our universe would slow down due to gravity (see (10.4)). Therefore, not only the quadratic term in (13.22), but also the linear term depends on dark energy.



13.9. Generation of energy by the Earth–Sun system

No model describes reality absolutely exactly. Therefore, an extremely small deviation $\varepsilon > 0$ of the actual position of a body from its position defined by Newton’s theory of gravity during one year may cause after a billion years a large and well-detectable error on the order of $10^9\varepsilon$. These tiny deviations from Newtonian theory of gravity do not cancel but accumulate (while e.g. rounding errors tend to nullify each other). The cumulative character causes secular changes, which means that these deviations always act towards one and the same “direction” and therefore are possibly detectable. From this one can see why antigravity acts locally. For example, for the Moon we have $\varepsilon = 1.71$ cm per year (see (12.21)) and for the Earth $\varepsilon \approx 5$ m per year (see (13.2)). As we have seen and as we shall also see in the next chapters, not only the Solar system but also other systems of free bodies “inflate” on average with increasing time. To prove this, we must be able to measure distances very accurately, as in the case of our Moon (see Chapter 12), or consider very long time and space scales, when the effect of antigravitational forces accumulates so much that it can be detected.

A significant recession of the Earth from the Sun described in Sections 13.1–13.4 cannot be explained by classical Newtonian mechanics, but we may estimate how much dark energy is generated by this system per year. For simplicity, assume that the Earth has a circular orbit of radius $R = 1$ au. From (13.5) and Kepler’s third law $R^3/Y^2 = GM_\odot/4\pi^2$ we find that the total (i.e. kinetic and potential) energy of the Earth is equal to

$$E(R) = \frac{1}{2}M\left(\frac{2\pi R}{Y}\right)^2 - \frac{GMM_\odot}{R} = -\frac{GMM_\odot}{2R}, \quad (13.23)$$

where the Earth’s mass M is given in (13.14) and the Sun’s mass M_\odot by (13.9). Now, for $\Delta R = 5.2$ m (cf. (13.2)), we obtain that the annual

increase of the total energy can be estimated as follows:

$$\begin{aligned} E(R + \Delta R) - E(R) &= -\frac{GMM_{\odot}}{2} \left(\frac{1}{R + \Delta R} - \frac{1}{R} \right) \\ &= \frac{GMM_{\odot}\Delta R}{2(R + \Delta R)R} = 9.4 \cdot 10^{22} \text{ J}, \end{aligned} \quad (13.24)$$

which is about 10 orders of magnitude smaller value than the kinetic energy of the Earth, i.e.

$$|E(R)| = \frac{1}{2}M \left(\frac{2\pi R}{Y} \right)^2 = \frac{GMM_{\odot}}{2R} = 2.69 \cdot 10^{32} \text{ J}.$$

The value (13.24) corresponds by (13.5) to a continual power of nearly 3000 TW. To shift the Earth weighing $5.9736 \cdot 10^{24}$ kg in the gravitational field of the Sun by only 5 meters thus requires a huge amount of energy (cf. (13.24)). For another annual increment ΔR , the energy or power is only linearly rescaled by the rule of three.

We see that the deviation from the energy conservation law of Newtonian mechanics during one year differs at the tenth significant digit. Hence, we should not trust too much long-term simulations (e.g. the evolution of the Solar system for hundreds of millions of years) based on Newtonian theory of gravity, because then the error of this model is relatively large due to the accumulation of errors caused by antigravity (cf. Section 5.5).



14. Antigravity and the anthropic principle

Concerning extraterrestrial civilizations:

Where are they?

ENRICO FERMI

14.1. The anthropic principle

In 1973 the Australian mathematician and theoretical physicist Brandon Carter introduced the term *anthropic principle*¹ in [48]. This term was later developed and extended in the 1986 monograph *The anthropic cosmological principle* [31] by John Barrow and Frank Tipler. The weak formulation of this principle states that all fundamental physical constants have just such values that they enabled the origin of life. Similarly, the strong formulation² of this principle postulates that evolution necessarily leads to the origin of humans (= *Anthropos* in Greek).

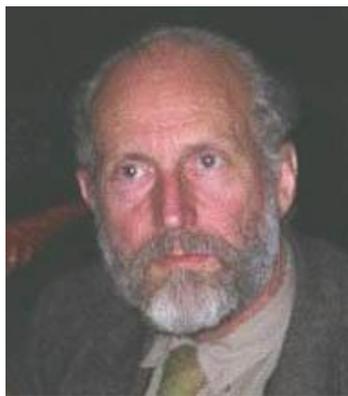


Fig. 14.1. Brandon Carter (b. 1942)

¹It was first introduced to the public in Krakow at the symposium dedicated to the 500th anniversary of the birth of Nicolaus Copernicus.

²There exist also other definitions of the strong and weak anthropic principles.

However, we may raise the following objection against the strong anthropic principle formulation. If the asteroid that caused the extinction of the dinosaurs had missed the Earth, then the human race would never have developed. It cannot be denied that there would possibly be other intelligent beings. Nevertheless, it is very surprising that fundamental physical constants have exactly the values that allow the birth and development of life in the universe.

Let us emphasize that no physical constant should be considered as a standard mathematical constant. For instance, the irrational numbers

$$\begin{aligned}\pi &= 3.1415926535 \dots \quad (\text{Ludolf number}), \\ e &= 2.7182818284 \dots \quad (\text{Euler number}), \\ \sqrt{2} &= 1.4142135623 \dots\end{aligned}$$

have infinitely many digits. On the other hand, the Newton gravitational constant $G = 6.674 \cdot 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$ is not a real number with an infinite decimal expansion. According to (4.22), its fourth significant digit is probably close to four, but the other digits are not known. In the future it will be impossible to find e.g. one thousand digits of G , since physical constants have a completely different character from real numbers. Physical constants should be rather treated as “fuzzy numbers” or “interval arithmetic numbers” or the “density of some probabilistic distribution function”. The reason is that the Heisenberg uncertainty principle prevents from obtaining infinitely many true decimal digits.

Note that the Newtonian gravitational law³ represents only a certain idealization of reality, since G is well defined only purely theoretically between two mass points, and no mass points exist in the real world. Similarly there are no exactly homogeneous balls (cf. Theorem 4.1), because the balls are composed of atoms. It is impossible

³Most plants and animals can detect gravity. For example, a person has in addition to the five basic senses (sight, hearing, smell, taste, and touch), the often neglected sixth sense of balance located in the middle ear. This sensory organ registers the direction in which the vector of gravity points.

to establish their masses and distances, e.g. to 20 significant digits. Therefore, the gravitational constant can never be measured and stated with absolute exactness.

The product GM , where M is the mass of a star, is proportional to the pressure inside the star. Hence, the value G has an influence on the central temperature, luminosity of the star, its age, and many other parameters. If G were to be only one part per million smaller or larger than its current value, then all stars and also galaxies would evolve in a completely different way, and hence the Earth could not come into being as it is.

The same is also true for other physical constants such as Avogadro's number, Planck's constant, the dimensionless constant of fine structure $\alpha \approx \frac{1}{137}$, and many others. The masses of the proton, neutron, and electron and the sizes of their interactions (weak, strong and electromagnetic) are very finely tuned so that stable atoms could be formed, which would then be the building blocks for the creation of complex molecules (see Fig. 14.3). The expansion rate of our universe is also an important parameter in the anthropic principle. If this rate were to be too large, galaxies and stars would not form. If it were too small, the universe would collapse and life would not have enough time to appear. In this chapter we will concentrate on the expansion rate of the universe and also the local expansion of the Solar system. We show how this contributed to the origin and evolution of life on our Earth.



14.2. Two-sided estimates

The faint young Sun paradox can well be explained by the anti-gravitational forces that generate dark energy locally, see [133]. As in Section 13.2 we first assume that during the past 3.5 billion years the Earth has receded from the Sun at the constant speed

$$\bar{v} = 5.2 \text{ m/yr}, \quad (14.1)$$

which is comparable in magnitude to the Hubble constant corrected to 1 au (see (13.3)).

Now, we will show in detail that \bar{v} is optimal in the sense that the Earth receives practically the same flux of energy from the Sun as the solar constant L_0 (defined by equality (13.1)) over a very long time period of 3.5 billion years. The optimal flux L_{opt} will correspond to the particular expansion rate (14.1).

In order to prove this (see Theorem 14.1 below), we put

$$\tau = -3.5 \text{ Gyr}$$

and $t = 0$ will again correspond to the present time. The main results of this section are formulated in the form of mathematical theorems to make it clear what is assumed and what is claimed. The symbol $\forall t$ means “for all t ”.

Since the luminosity of the Sun increases approximately linearly with time and it was only about 77% of its present value 3.5 Gyr ago (see Fig. 11.2), we set

$$L(t) = \left(1 - 0.23\frac{t}{\tau}\right)L_0 \quad \forall t \in [\tau, 0], \quad (14.2)$$

i.e. $L(\tau) = 0.77L_0$ and $L(0) = L_0$. As the luminosity decreases with the square of the distance, we can state the following quite surprising assertion.

Theorem 14.1 (Optimal recession speed of the Earth from the Sun). *Set*

$$L_{\text{opt}}(t) = \frac{L(t)R^2}{(R + \bar{v}t)^2}, \quad t \in [\tau, 0], \quad (14.3)$$

where $R = 1 \text{ au}$ and \bar{v} is given by (14.1). Then

$$|L_{\text{opt}}(t) - L_0| < 0.005 \text{ kW m}^{-2} \quad \forall t \in [\tau, 0]. \quad (14.4)$$

P r o o f . The very small dispersion of luminosity $\pm 0.005 \text{ kW m}^2$ on the right-hand side of formula (14.4) can be easily derived an-

alytically by investigating the rational function $L_{\text{opt}}(t)$. It is concave on the whole interval $[\tau, 0]$, By inspection we find that its minimum is attained in the left end point τ , $L_{\text{opt}}(\tau) = 1.357387\dots$ and the maximum of the function L_{opt} is attained in $t^* \approx -1.7 \cdot 10^9$ yr, $L_{\text{opt}}(t^*) = 1.364574\dots$ Hence, the following two-sided estimate holds (cf. Fig. 14.2):

$$1.355 < L_{\text{opt}}(\tau) = \min_{t \in [\tau, 0]} L_{\text{opt}}(t) < \max_{t \in [\tau, 0]} L_{\text{opt}}(t) < 1.365$$

and $L_{\text{opt}}(0) = L_0$. \square

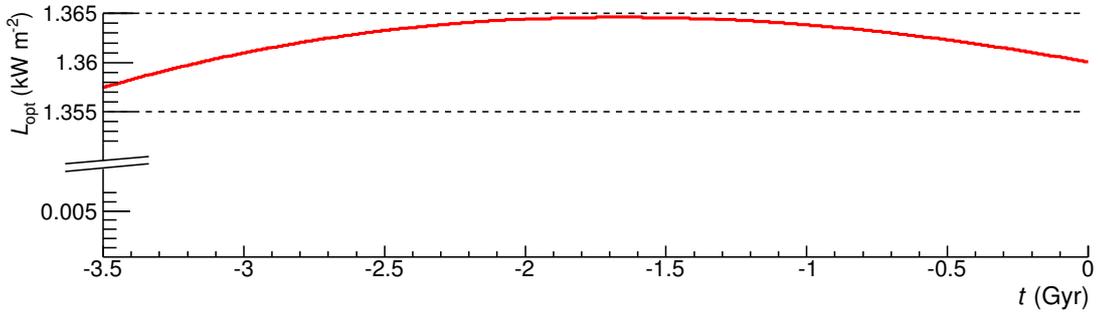


Fig. 14.2. Plot of the almost constant function $t \mapsto L_{\text{opt}}(\tau)$ over the interval $[\tau, 0]$. The vertical axis is substantially shortened for the clarity of the presentation.

The luminosity flux

$$1.36 \pm 0.005 \text{ kW m}^{-2} \quad \forall t \in [\tau, 0]$$

would, of course, guarantee very stable conditions (14.4) for the development of intelligent life on Earth over the very long period of 3.5 Gyr. In particular, the amount of dark energy seems to be just right for an almost constant influx of solar energy and thus also for the appearance of humankind.

Dark energy thus represents further support for the (weak) anthropic principle, which states that basic physical constants are favorable to the emergence of life only if they are in very narrow intervals [133]. Moreover, the speed in (14.1) is optimal in the sense that any other slightly different speed would not yield an almost constant

flux expressed by the rational function in (14.3) on the time interval 3.5 Gyr. Thus it is probable that the real average recession speed of the Earth from the Sun oscillates about the value 5.2 m/yr (see Theorems 14.2 and 14.3), e.g. due to the influence of other planets.

There is evidence supporting the hypothesis that the Earth during its history was a “snowball” several different times (see [85, p. 58]) but, on the other hand, sometimes the temperature of the oceans was much higher than today [163]. Of course, the composition of the atmosphere and also the greenhouse effect played a significant role. It is known that a decrease of luminosity of only a few percent caused ice ages in the past. A decrease larger than 5% would cause total glaciation of the whole planet. From Section 13.2 we know that a decrease or increase of the solar constant L_0 up to 5% corresponds to a ring — popularly called the *ecosphere* (*habitable zone*) — with radii $(0.95)^{1/2}$ au and $(1.05)^{1/2}$ au that represents a very narrow interval 145.8–153.3 million km from the Sun (see Fig. 13.2).

Now, for a variable continuous recession speed v on the interval $[\tau, 0]$ we define similarly to (14.3) the associated luminosity

$$\mathcal{L}(v, t) = \frac{L(t)R^2}{\left(R - \int_t^0 v(\theta)d\theta\right)^2}, \quad t \in [\tau, 0], \quad (14.5)$$

where $\tau = -3.5$ Gyr, $R = 1$ au, $L(t)$ is given by (14.2), and the integral of the velocity yields the distance. In the case that we are not able to establish a proper value of some quantity, the so-called *two-sided estimates* may be quite useful (see e.g. [144]).

Theorem 14.2 (Two-sided estimates). *If the recession speed $v = v(t)$ of the Earth from the Sun lies in the interval from 4.26 m/yr to 6.14 m/yr for every $t \in [\tau, 0]$, then the luminosity defined by (14.5) changes at most about 5% from L_0 , namely,*

$$0.95L_0 \leq \mathcal{L}(v, t) \leq 1.05L_0 \quad \forall t \in [\tau, 0].$$

P r o o f . By inspection we find that for the constant velocity $v_1 \equiv 4.26$ m/yr the rational function $t \mapsto \mathcal{L}(v_1, t)$ is increasing on $[\tau, 0]$ and thus by (14.5),

$$0.95L_0 = \frac{L(\tau)R^2}{(R + v_1\tau)^2} = \mathcal{L}(v_1, \tau) \leq \frac{L(t)R^2}{(R + v_1t)^2} = \mathcal{L}(v_1, t) \quad (14.6)$$

for any $t \in [\tau, 0]$. Analogously, for the constant velocity $v_2 \equiv 6.14$ m/yr, we get that $t \mapsto \mathcal{L}(v_2, t)$ is decreasing, and therefore,

$$\mathcal{L}(v_2, t) \leq \mathcal{L}(v_2, \tau) = \frac{L(\tau)R^2}{(R + v_2\tau)^2} = 1.05L_0. \quad (14.7)$$

Putting (14.6) and (14.7) together, we obtain by (14.5) and the proposed estimates $4.26 \leq v(t) \leq 6.14$ that for any $t \in [\tau, 0]$

$$0.95L_0 \leq \frac{L(t)R^2}{(R + 4.26t)^2} \leq \frac{L(t)R^2}{\left(R - \int_t^0 v(\theta)d\theta\right)^2} \leq \frac{L(t)R^2}{(R + 6.14t)^2} \leq 1.05L_0. \quad \square$$

A more important converse proposition has stronger assumptions on the velocities.

Theorem 14.3 (Additional two-sided estimates). *If the average recession speed \bar{v} lies outside the interval $[4.26, 6.14]$ m/yr, then there exists a nonempty subinterval $I \subset [\tau, 0]$ such that the luminosity $\mathcal{L}(\bar{v}, t)$ is less than 95 % or greater than 105 % of L_0 for all $t \in I$.*

P r o o f . If $\bar{v} < v_1 \equiv 4.26$ m/yr, then similarly to (14.6) we get

$$\mathcal{L}(\bar{v}, \tau) = \frac{L(\tau)R^2}{(R + \bar{v}\tau)^2} < \frac{L(\tau)R^2}{(R + v_1\tau)^2} = \mathcal{L}(v_1, \tau) = 0.95L_0.$$

From the continuity of the rational function $t \mapsto \mathcal{L}(\bar{v}, t)$ it follows that there exists a nonempty time interval I_1 such that $\mathcal{L}(\bar{v}, t) < 0.95L_0$ for all $t \in I_1$.

Analogously to (14.7) we find that for $\bar{v} > v_2 \equiv 6.14$ m/yr there exists a nonempty interval $I_2 \subset [\tau, 0]$ such that $\mathcal{L}(\bar{v}, t) > 1.05L_0$ for all $t \in I_2$. \square

An average speed \bar{v} lying outside the interval $[4.26, 6.14]$ m/yr corresponds to truly inhospitable conditions for the evolution of life into multicellular forms. On the Earth there would be either too strong a frost, or vice versa a tremendous heat, because the associated solar constant would differ by more than 5% from its current value of L_0 during some time period. The previous two theorems can also be easily modified for other values than just 5%.

Note that in some models the linear function $L(t) = (1 - 0.23t/\tau)L_0$ from (14.2) and Fig. 11.2 is replaced by the rational function (see e.g. [20, p. 177])

$$\hat{L}(t) = \frac{L_0}{1 + 0.3t/\tau_0}, \quad t \in [\tau, 0],$$

where $\tau_0 = -4.5$ Gyr. In this case the optimal average recession speed (guaranteeing an almost constant energy flux from the Sun) is

$$\bar{v} = 4.36 \text{ m/yr}$$

and the mean recession speed should be in the interval $[3.27, 5.21]$ m/yr to keep variations of the solar energy flux below 5% as in Theorems 14.2 and 14.3.

According to [6, p. 218], the universe expands exponentially due to dark energy. From Fig. 8.7 we observe that the Hubble parameter $H = H(t)$ has been almost constant during the last 4.5 Gyr when the Solar System was formed. It lies in the interval $[H_0, \frac{5}{4}H_0]$. Assuming that $H(t)$ is constant, i.e., $H(t) \equiv H_0$, we get by (10.3) that Theorem 14.1 can be modified for an exponential expansion as follows. If the average recession speed of the Earth from the Sun is $\tilde{v} = 5.014$ m/yr, then $L_{\text{opt}}(t) = 1.36 \pm 0.008 \text{ kW m}^{-2}$ for all $t \in [\tau, 0]$, which is analogous to (14.1) and (14.4). Also Theorems 14.2 and 14.3 can correspondingly be slightly modified. Moreover, from Fig. 13.4 we observe that the expansion function $a = a(t)$ is almost linear during the last 4.5 Gyr. Therefore, the expansion given by equation (14.5) approximates reality better than the exponential expansion proposed in [6].



14.3. Will antigravity protect the Earth from an expanding Sun?

The assumptions of Theorem 14.3 would not guarantee suitable conditions for the development of life. From Theorems 14.2 and 14.3 we find that the probable secular expansion rate of the radius of Earth's orbit lies in the interval

$$H_0^{(\text{loc})} \in [0.426 H_0, 0.614 H_0].$$

Such a local Hubble expansion is therefore perfectly tuned by antigravity (see [133]).

Therefore, not only the fundamental physical constants are finely tuned. For example, the average surface temperature of the Sun $T = 5770$ K is also perfectly balanced. If it were to decrease about 1 %, i.e. a mere 57.7°C , then by the Stefan–Boltzmann law $L_0 = \sigma T^4$ with a constant $\sigma > 0$ the solar constant L_0 would decrease about 4 % as $0.99^4 L_0 \approx 0.96 L_0$. The total luminosity of the Sun would also decrease by 4 %.

It is often argued⁴ that after half a billion years, the water in the oceans will evaporate, because the power of the Sun will be too high⁵ (see Fig. 11.2). However, the recession speed (14.1) would guarantee very stable conditions on the Earth for several Gyr in the future. For instance, in the next 3.5 Gyr the flux density of solar energy will be in the very favorable interval $1.33\text{--}1.36$ kW m⁻², if the luminosity would evolve as in (14.3). The function $t \mapsto L_{\text{opt}}(t)$ is in fact decreasing in the interval $[0, |\tau|]$ and from relations (14.2) and (14.3) it follows immediately that

$$L_{\text{opt}}(|\tau|) = 1.33 \text{ kW m}^{-2}.$$

Thus there would be pretty good prospects into the far future. Life

⁴These simplified conclusions do not take into account that the Earth can be inhabited by a technically advanced civilization, which can, for instance, place a giant reflection sheet at the Lagrangian point L_1 to control the incoming flux of variable solar energy.

⁵According to [243, p. 461], the Sun's luminosity will increase to $1.33 L_\odot$ after 3 Gyr. Thus, it may exceed $1.05 L_\odot$ over half a billion years.

on the Earth would have had a chance to develop during a long time period in very stable conditions.

On the other hand, dark energy and the reduction of atmospheric pressure on Mars caused Mars to leave its ecosphere.⁶ It is therefore possible that Mars could move away from the Sun faster than the ecosphere stretched around its orbit due to increased solar flux and antigravity.

After 5–7 billion years, when reserves of hydrogen in the central region of the Sun will run out, the Sun will begin to change to a red giant. Its radius will extend beyond the current orbit of Venus. At that time the Earth could be about 180 million miles away from the Sun, provided the receding velocity is (14.1). Hence, antigravity can keep the Earth sufficiently far away from the growing Sun.



14.4. The probability of the appearance of life

The Sun is relatively young compared to the oldest stars in our Galaxy, whose ages are estimated at more than 13 billion years. Nature thus has had much more time to perform experiments leading to the origin of life on some exoplanets in the habitable zone of the Milky Way. It is assumed that intelligent civilizations might gradually begin to colonize the Galaxy within a few tens of millions of years after their appearances. There exist about 10^{12} galaxies in the observable universe and each galaxy contains on average more than 10^{12} exoplanets. Nonetheless, we do not observe any signs of other civilizations. This is called the *Fermi paradox*. A popular story recounts that during a lunch with his colleagues at Los Alamos National Laboratory some time in 1950, Enrico Fermi ask the following question about extraterrestrial civilizations: *Where are they?*

⁶The popular concept that we can easily arrange on the Martian surface suitable temperature conditions for humanity by the greenhouse effect is probably too optimistic. Additionally, the cosmic radiation on Mars is several times higher than on Earth due to the weak magnetic field of Mars.

In 1961, Frank Drake⁷ introduced an equation for the number N of civilizations in our Galaxy⁸ with which it would be possible to establish radio contact. The number N is equal to the product of seven variables whose values are difficult to estimate. One of them is the probability of life on a planet in the habitable zone, see [85, p. 200]. Let us examine this in more detail.

While in space there are huge amounts of organic substances (for example, the characteristic spectra of the amino acid glycine have been detected in the interstellar space) and even more complex molecules including nucleotides have been discovered in meteorites falling onto the Earth, the probability of the appearance of a self-replicating molecule is very small. It is not easy to induce any organic molecule to create copies of itself. There is a certain minimum number of information bits that allows for self-replication. For instance, the smallest known viruses (also computer viruses) have about 1000 bits of information. However, such a sequence of 0s and 1s is certainly not random and to find it requires one to search through about 2^{1000} possibilities. This important fact is not taken into account in many optimistic forecasts (e.g. the Fermi paradox or the Drake equation).

The reader may ask himself/herself how such complex molecules such as nucleic acids RNA and DNA and the complicated information processes associated with them (see [140]) emerged in the universe. Nature on Earth experimented with an enormous amount of organic molecules over a period of about one billion years in a huge biochemical laboratory on the whole surface of the Earth (consisting of almost 500 000 000 000 000 m²), in various cracks, in the oceans at different temperatures and pressures and so on. So apparently life on the Earth could have appeared from an original prebiotic soup in one location, if it was not carried here from elsewhere.

At the very beginning of life there was almost certainly no molecule of RNA or DNA (see Fig. 14.3), but rather some primitive proteins.

⁷F. Drake is also the author of the first message to extraterrestrial civilizations in 1974.

⁸The diameter of our Galaxy is approximately 100 000 ly.

Experiments with projectiles contained inside amino acids show that a sharp impact may produce short protein chains containing up to 5 amino acids. These basic building blocks of life are among others contained in comets. Thus during their impacts on the Earth simple proteins could be synthesized, and these proteins could be further improved by mutations in exceptional circumstances.

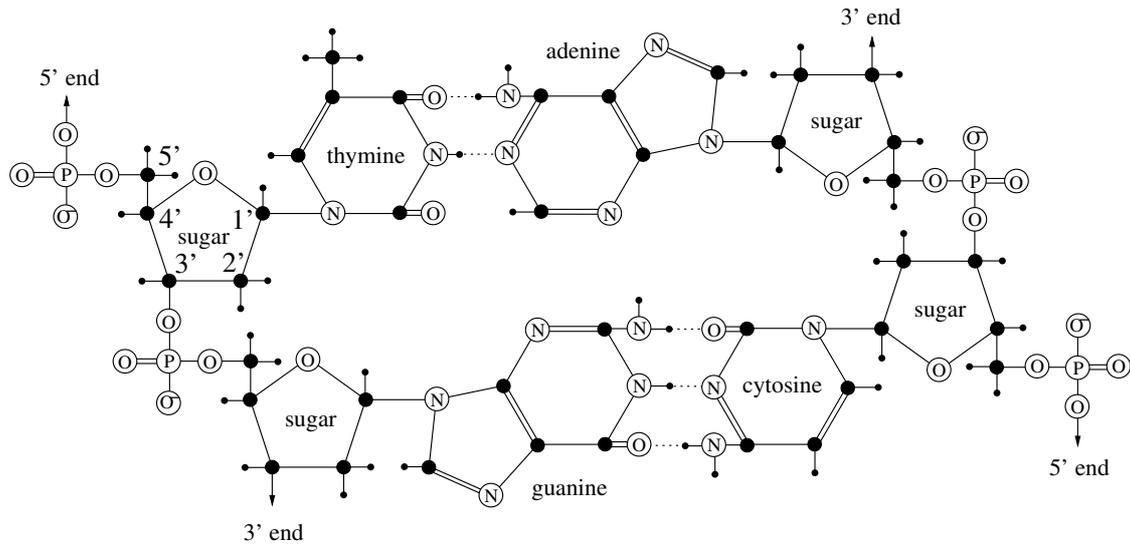


Fig. 14.3. Schematic illustration of the chemical structure of DNA: atoms of carbon are marked by large black dots, atoms of hydrogen by small black dots, covalent bonds are shown as solid lines, and hydrogen bonds as dotted lines. At both edges there is a solid skeleton of sugar phosphates, which protects the genetic information from damage. In these edges the deoxyribose sugar alternates with the phosphate group PO_4 .

In 1997, Stanley B. Prusiner won the Nobel Prize for the discovery of prions. This is a complex protein (known from mad cow disease), which does not contain any nucleic acid and it is not coded by them. It reproduces itself by changing similar proteins in any organism to itself. It is therefore not inconceivable that similar proteins could appear at the very beginning of life on our planet before RNA appeared. About one billion years after the formation of RNA the helix nucleic acid DNA evolved. Note that gluing the same type of molecule (see Fig. 14.3) always in the same way, a helical structure

is obtained [53] (which in limiting cases may degenerate to a circle or line). Since the standard genetic code is highly redundant, there is still space for some hidden secondary functions of this code [140]. Darwin's evolutionary theory of the origin of species by natural selection in a competitive environment explains how today's advanced human civilization appeared.



15. Expansion of the Solar system

*Our future discoveries must be looked
for in the sixth place of decimals.*

ALBERT ABRAHAM MICHELSON

15.1. Fast satellites

In this chapter we will present many examples suggesting that gradual expansion can also be observed in other smaller subsystems of the Solar system. This apparently contradicts the classical law of conservation of energy. Below we again present arguments supporting the hypothesis that antigravity significantly contributes to today's expanding trajectories of planets and their moons.



Fig. 15.1. Fast satellite Phobos with dimensions $27 \times 22 \times 19$ km orbits around Mars once every 7.65 hours while Mars rotates about its axis in 24.62 hours. The largest crater Stickney, which is 9 km in diameter, is located on the right (photo NASA).

In the Solar system we know of 19 satellites of Mars, Jupiter, Uranus, and Neptune that are below the so-called *stationary orbit*. In this stationary orbit the orbital period is the same as the rotation of the planet about its axis. By Kepler's third law (4.4) the radius of the stationary orbit is equal to

$$r_i = \left(\frac{Gm_i P_i^2}{4\pi^2} \right)^{1/3}, \quad (15.1)$$

where m_i is the mass of the i th planet and P_i is its sidereal rotation. We call them *fast*, since their orbital period is shorter than P_i . From a statistical point of view it is very unlikely that all these satellites were captured, since all of them move in the same direction on circular orbits with almost zero inclination.¹ Therefore, they have been in their orbits approximately 4.5 Gyr, even though some may be parts of larger disintegrated satellites. Some of them are larger than Phobos (see Fig. 15.1), some smaller.

By Newtonian mechanics, tidal bulges continually reduce potential energy and orbital periods of these fast satellites. Their speed increases and the rotation of the mother planet also slightly increases to keep the total angular momentum constant. According to [20, p. 96] the tidal forces corresponding to 1 kg of the satellite mass are proportional² to m_i/r^3 , where r is the radius of a particular satellite orbit and $r < r_i$. Notice that the ratio m_i/r^3 has the same order for all 19 known fast satellites (see the last column of Table 15.1).

In Section 15.4 we derive that the satellite Phobos (see Fig. 15.1) approaches Mars with an average speed of 1.9 cm per year. Assuming the other fast satellites are approaching by a similar speed of 1–2 cm per year, then they have moved 45 000–90 000 km closer towards their mother planets during the 4.5 Gyr of their existence. However, this

¹The ellipsoidal shape of planets causes the gravitational potential not to be spherically symmetric which forces all satellites and also rings to move to the plane of symmetry perpendicular to the rotational axis of a planet.

²The effect of tides is greater for gaseous planets than for elastoplastic planets which is greater than for rigid planets.

contradicts the fact that the radius of the stationary orbit of Uranus is

$$r_7 = 82\,684 \text{ km}$$

and for Neptune

$$r_8 = 83\,512 \text{ km},$$

since their fast satellites are on very high orbits with radii approximately 48 227–76 400 km (cf. also the penultimate column of Table 15.1). Moreover, by (15.1) the radii of stationary orbits were smaller in the past, since the planets rotated faster, leading to a smaller value for P_i . It might again be antigravity acting in the direction opposite to gravity and thus preventing the fast satellites from crashing into their mother planet.



15.2. Where was Larissa billions of years ago?

The satellite Larissa orbits Neptune at the distance

$$d = 73\,548 \text{ km},$$

which is a value very close to the radius r_8 of the stationary orbit. Neptune rotates around its axis once every 16.11 hours and Larissa orbits around it in 13.32 hours. Neptune's rotation is slowed mainly by the large moon Triton, and thus several billion years ago r_8 was smaller due to (15.1). So it is not clear where the satellite Larissa was several billion years ago, since it should be continually approaching Neptune according to the laws of classical mechanics. If it were above the stationary orbit, it would be moving away from Neptune. Antigravity again probably slows its gradual fall due to its repulsive forces that act in the opposite direction than the tidal forces, and Larissa seems to be carried above due to antigravity. In this case it seems that the effect of tidal forces is of the same order as antigravitational forces but has the opposite sign.

Table 15.1. Fast satellites in the Solar system. The symbol m_i stands for the mass of the i th planet in kg divided by 10^{24} , r_i is the radius of the stationary orbit (15.1) in km, r is the radius of the orbit of a particular satellite in km, and the corresponding values of m_i/r^3 proportional to tidal forces related to 1 kg of mass of a satellite are given in kg/m^3 .

i	planet	m_i	r_i	fast satellite	r	r/r_i	m_i/r^3
4	Mars	0.64185	20 429	Phobos	9 377	0.459	778.6
5	Jupiter	1898.6	160 020	Metis	127 690	0.798	911.9
				Adrastea	129 690	0.810	870.4
7	Uranus	86.81	82 684	Cordelia	49 770	0.602	704.3
				Ophelia	53 790	0.651	557.9
				Bianca	59 170	0.716	419.1
				Cressida	61 780	0.747	368.2
				Desdemona	62 680	0.758	352.6
				Juliet	64 350	0.778	325.9
				Portia	66 090	0.799	300.8
				Rosalind	69 940	0.846	253.8
				Cupid	74 800	0.905	207.5
				Belinda	75 260	0.910	203.7
8	Neptune	102.43	83 512	Pertida	76 400	0.924	194.7
				Naiad	48 227	0.577	913.2
				Thalassa	50 074	0.600	815.8
				Despina	52 526	0.629	706.8
				Galatea	61 953	0.742	430.8
				Larissa	73 548	0.881	257.5

For example, the Hubble constant (see (11.2))

$$H_0 \approx 10 \text{ m yr}^{-1} \text{ au}^{-1} \quad (15.2)$$

rescaled to the distance d of Larissa from Neptune is

$$\boxed{H_0 \approx 0.5 \text{ cm yr}^{-1} d^{-1},}$$

which is a value that is actually comparable with the effects of tides. The antigravitational force of our Moon is approximately of the same order of magnitude as the tidal forces (cf. (12.20) and (12.21)).

Similar considerations as for Larissa are valid for the satellites Pertida and Belinda of Uranus. However, the rotational axis of this planet is nearly in the plane of the ecliptic due to some ancient collision for which it is not known when it occurred.



15.3. Satellites of Uranus

Let us introduce another argument supporting the hypothesis of the existence of antigravitational forces in the Solar system. The average difference between radii of orbits of two neighboring satellites below the stationary orbit of Uranus is only 2 663 km. The distances of such neighbors above the stationary orbit are much bigger (cf. Fig. 15.2). The orbit of the satellite Puck has a radius of 86 010 km, followed by Mab with 97 700 km and Miranda with 129 390 km. This is because for quick satellites, antigravitational forces and tidal forces are mutually subtracted, while for the satellites above the stationary orbit they are summed. Therefore, it is possible that the satellites Puck, Mab, Miranda, etc., depart from Uranus also due to the action of antigravity (see Fig. 15.2).

In Fig. 15.2 we still notice that the satellite Pertida is located just below the stationary orbit and the satellite Puck just above the stationary orbit. The distance of their orbits is only 9610 km. Why did tidal forces not push them further apart over the course of 4.5 billion years? The main reason might again be antigravity, which gradually pushes all satellites above the stationary orbit. It is therefore not excluded as a possibility that the satellite Puck (and resp. Mab) was once under the stationary orbit of Uranus and that antigravitational forces pushed it above, because the tidal forces near the stationary orbit are small. After a billion years Pertida also could circulate above the stationary orbit, if its rate of recession is of order 1 cm/yr.

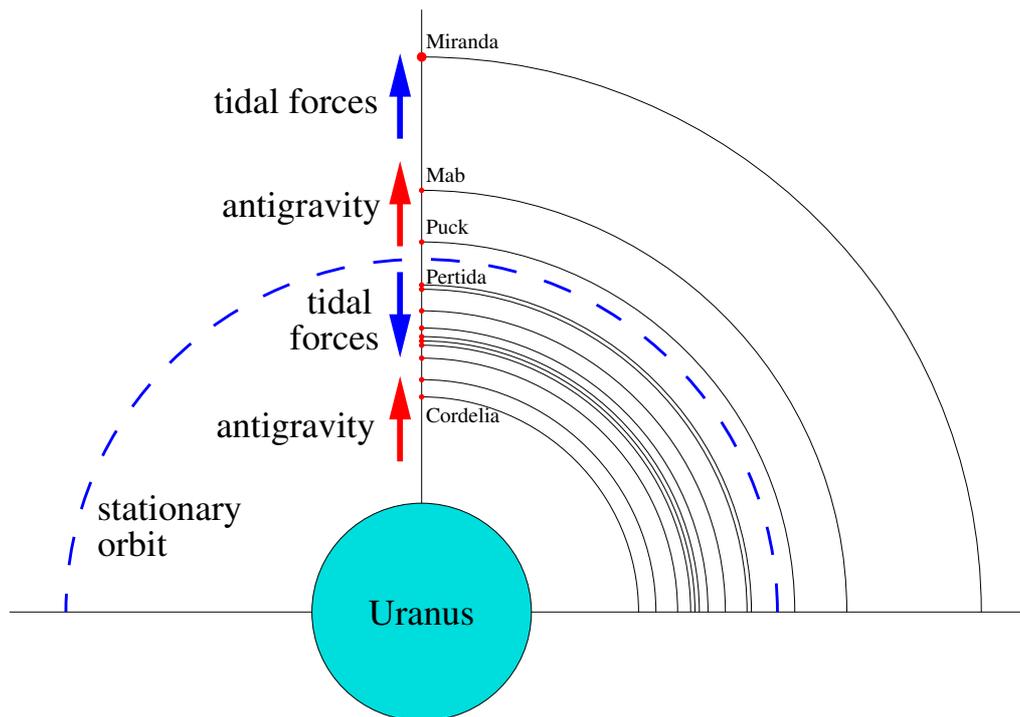


Fig. 15.2. Fast satellites of Uranus from Table 15.1 and their neighbors above the stationary orbit. Below this orbit the tidal forces and antigravitational forces are subtracted, since they have opposite directions, whereas above this orbit they are summed coherently. The distances between neighboring satellites above the stationary orbit are substantially greater than below it.

Analogous considerations can be stated for the planet Neptune. Above its stationary orbit circulates a moon tentatively named S/2004 N1. Its orbit has a radius of 105 283 km, followed by Proteus whose orbit has a radius of 117 646 km.



15.4. Falling Phobos

Apart from our Moon, Phobos is the most studied satellite in the Solar system, because it has a rapidly changing period of revolution. As a result of tidal forces it is continually spiraling closer to Mars, its orbital speed (about 2.13 km/s) is growing, and this satellite also slightly accelerates the rotation of Mars. Its angular orbital velocity is more than 3 times larger than the rotation of Mars around its own axis and is in the same direction.

Phobos has already orbited Mars for quite a long time, because its trajectory is almost circular and its orbital plane is perpendicular to the rotation axis of Mars. In addition, Phobos always faces Mars by its longest axis.

The radius of the planet Mars is 3 390 km. The radius of Phobos' trajectory is

$$a = 9377 \text{ km}, \quad (15.3)$$

and thus it orbits above the surface of Mars at a distance of just 5987 km, which is the lowest known value of all the satellites in the Solar system. It is a great mystery how Phobos could get on such a track (about 11 000 km below the stationary orbit). Since no large moons orbit around Mars, Phobos could not be captured only due to the gravitational field of Mars.

It is possible that Phobos was a component of a binary asteroid, the lighter component of which after a close encounter with Mars left its gravitational field. However, then Phobos would probably enter an elliptic orbit with a high eccentricity and close to the ecliptic plane, where almost all asteroids are situated [233]. An elongated elliptical orbit would be difficult to place into a very low stationary orbit with a radius of 20 429 km. Then Phobos could hardly reach an almost circular orbit with an inclination of 1° with respect to the plane of the equator of Mars. Note that the slope of the equator of Mars to its orbit is 25.19° (i.e. it is similar to the Earth's slope of 23.439°).

Another, more likely possibility is that Phobos originated from accreting fragments (like our Moon) after a big body hit the surface of Mars (see e.g. Fig. 15.3). All large craters and depressions on Mars are very old and Phobos also has an approximately 4 billion years old surface dotted with craters. Its trajectory in the past was certainly changed by impacting bodies (see Fig. 15.1), tidal forces, and probably also antigravity. It seems, therefore, that Phobos started to orbit Mars several billion years ago and thus it had plenty of time to take up its current circular path.

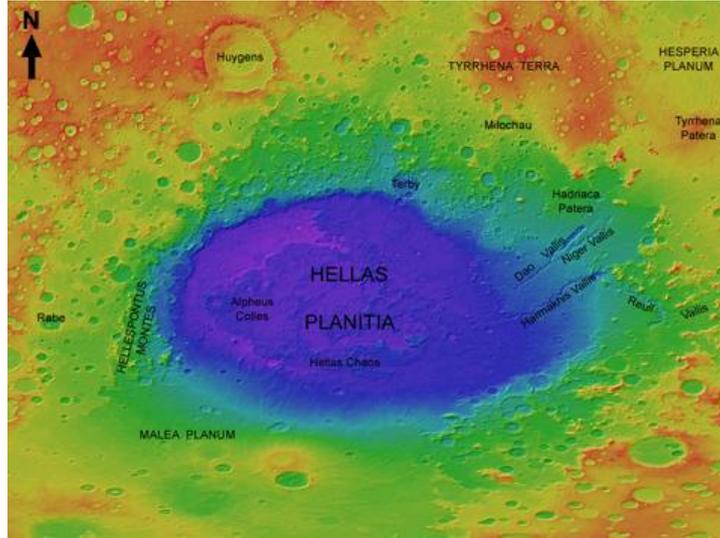


Fig. 15.3. The great depression Hellas Planitia on Mars is of impact origin. It has a diameter of over 2000 km, depth of 8.2 km, and is about 4 billion years old (photo NASA).

Initial estimates derived from the assumed effects of tidal forces asserted that Phobos approaches Mars with the speed over 5 cm per year. In [40, p. 170] this value was reduced to 2.68 cm per year. However, estimates of this speed further decreased to the currently accepted value

$$\dot{a} = -1.9 \text{ cm/yr}, \quad (15.4)$$

where the dot denotes the time derivative.

Now, let us examine in detail how this speed can be deduced. Positions of Phobos since its discovery in 1877 (see Section 4.5) have been carefully monitored — cf. the survey paper [189]. Since that time, Phobos orbited Mars more than 150 000 times. From this we can very accurately determine the current angular velocity of revolution. Each day Phobos travels an angle of 1128.844407° and thus

$$\omega = 412\,317.5991^\circ \text{ yr}^{-1}, \quad (15.5)$$

where yr stands for one sidereal year ($= 365.25636$ days).

Determination of the time derivative $\dot{\omega}$ is much less accurate and these estimates have decreased during the course of time. For example,

in 1945 it was believed that $\dot{\omega} = 0.001882^\circ \text{ yr}^{-2}$ (see [110, p. 674]). In 1988 Milan Burša [40, p. 168] showed that Phobos sped up by 17.7° during one century, i.e. $\dot{\omega} = 0.00177^\circ \text{ yr}^{-2}$. In 2010, R. A. Jacobson [110] obtained a similar, but again smaller value

$$\dot{\omega} = 0.00127^\circ \text{ yr}^{-2}. \quad (15.6)$$

According to Kepler's third law, a^3/T^2 is constant. Hence,

$$3a^2\dot{a}T^2 - 2T\dot{T}a^3 = 0,$$

where the period T satisfies the equality

$$\omega = \frac{360^\circ}{T}.$$

From this we deduce that $\dot{\omega}T = -\dot{T}\omega$, and thus by substitution from relations (15.3), (15.5), and (15.6) we get (cf. (13.17))

$$\dot{a} = \frac{2a\dot{T}}{3T} = -\frac{2a\dot{\omega}}{3\omega} = -1.9 \text{ cm/yr}.$$

Phobos is thus approaching Mars. It has too low an orbit, where tidal forces, which increase with the cube of the distance, seem to prevail over antigravity. Planetary scientists estimate that in approximately 30–80 million years Phobos will hit the surface of Mars, or it will be ripped apart by tidal forces as soon as it appears under the so-called Roche limit. The Newtonian models that were used led to a deviation of about 10 km kilometers from the observed positions of Phobos since 1950 (see e.g. [21], [110, p. 677], [197, p. 1149]). The reason is perhaps that effects of antigravity were not taken into account.

It is planned to place a laser device on Phobos capable of detecting a signal from the Earth to determine changes in its distance with an accuracy of about one millimeter [281].



15.5. Delaying Neptune

It is an open problem how Neptune could have been formed as far away as $R = 30$ au from the Sun, where the original protoplanetary disc was very sparse and where all motions are very slow (see [20, p. 534]). To reach Neptune's mass of about 10^{26} kg during its first 100 million years of existence, the proto-Neptune would have had to pick up an average of 30 billion kilograms of material per second in a very sparse environment.

If we would consider, for example, only half the rate of the expansion (15.2) similarly as in Chapters 12, 13, and 14, then $t = 4.5$ Gyr ago Neptune could have been formed several astronomical units closer to the Sun on an orbit with radius r . Due to relation (10.3) we will assume the exponential expansion $R = r \exp(\frac{1}{2}H_0t)$. From this and (15.2) it follows that

$$r = R \exp(-\frac{1}{2}H_0t) = R \exp\left(-\frac{5 \cdot 4.5 \cdot 10^9}{150 \cdot 10^9}\right) = R e^{-0.15} = 25.82 \text{ au.}$$

For a linear expansion rate we would get by (11.4) an analogous value

$$r = R - \frac{1}{2}H_0tR = 25.5 \text{ au.}$$

For a slight increase in distance ΔR during one orbit of Neptune around the Sun, we get by Kepler's third law

$$\frac{(R + \Delta R)^3}{(P + \Delta P)^2} = \frac{R^3}{P^2},$$

where ΔP is the associated increase of the orbital period $P = 164.79$ yr. Multiplying individual algebraic factors and neglecting terms of higher order in ΔP and ΔR , we find similarly to (13.17), that

$$\frac{2\Delta P}{P} \approx \frac{3\Delta R}{R}. \quad (15.7)$$

Considering again only half the rate of that of (15.2), we get from (15.7) that Neptune during one orbit around the Sun is delayed by the quite

tiny angle $\alpha \approx \tan \alpha$,

$$\alpha \approx \frac{\Delta P}{R} \frac{2\pi R}{P} = \frac{2\pi \Delta P}{P} \approx \frac{3\pi \Delta R}{R} = 0.01'', \quad (15.8)$$

where $\Delta R = RP \cdot 5 \text{ m yr}^{-1} \text{ au}^{-1} = 24718.5 \text{ m}$. Such a small unexplained shift (15.8) unfortunately cannot be confirmed by Galileo's observations in 1612 of Neptune which he sketched manually [123] (see also M. L. Lalande (1795)), but modern tools confirm a slight observed delay of Neptune of the same order [264] (see also [223]). For comparison, let us mention that the angular diameter of Neptune is $2.3''$.

Originally, astronomers thought that the delay is due to an additional outer planet, which, however, was not found. The mass of Pluto proved too small to change the orbit of Neptune. At the turn of the 20th century the distance between these two bodies was approximately 15 au (cf. Fig. 4.5). When Clyde Tombaugh discovered Pluto in 1930 (based on calculations of Percival Lowell from 1915), Neptune was 30 au away from it. At such a great distance the gravitational influence of Pluto is indeed negligible. So what is then the source of an unexplained delay in Neptune's orbit? One possibility for the gradual recession of Neptune from the Sun along a spiraling path is the effect of antigravitational forces. Equation (15.8) actually shows how extremely small effects occur on short time intervals (see also (12.21)).

The popular Newtonian Nice model [280] (developed at the University of Nice in France) tries to explain the migration of Neptune and other planets by using various resonances. This model predicts that Uranus and Neptune migrated away from the Sun for billions of years and in addition they swapped their orbits. In order to keep the total energy of the Solar system constant during the migration, the heavy planet Jupiter shifted closer to the Sun (see e.g. [280, p. 435]). However, this scenario has a number of drawbacks and suffers from several defects. Let us mention only the most important ones:

a) It is known that the classical N -body problem has a unique global solution for given initial conditions, if these bodies do not collide. However, the authors of [280] did not check that the Nice model

is potentially correct by showing that backward integration from the current position of Uranus and Neptune would cause their swapping in the past and that the original initial conditions could be reached. To this end Theorem 5.1 can be employed.

b) The solution of the classical N -body problem is unstable in the sense of Lyapunov. In other words, extremely small changes in the initial conditions lead after billions of years to a huge deviation in the final state. Long-term numerical integration is thus not justified.

c) The absolute validity of Newton's theory of gravitation is assumed on the long time interval of 4.5 billion years. The influence of dark energy, the finite speed of propagation of gravitational interaction, and modelling errors are ignored.

d) It is not explained how the rich families of satellites could survive the swapping of Uranus and Neptune.

e) There is no analysis of numerical or other errors that occur during the simulation. Note that numerical errors usually grow exponentially.

Various resonances played, of course, an important role in the development of the Solar system and they surely influenced the orbit of Neptune. Nevertheless, antigravity acted continually and produced the huge amount of energy required for the migration of all planets (cf. (13.24)).



15.6. Neptune-Triton system

The enormously large orbital angular momentum of the Neptune-Triton system (see Fig. 15.4) is a deep mystery. Triton is probably a trapped moon (due to some N -body gravitational collision or a crash with another body), because it orbits around Neptune in the opposite direction than Neptune rotates about its axis [50]. Such an orbit is called *retrograde*. Triton slows Neptune's rotation (as our Moon reduces Earth's rotation). However, since it circulates in the opposite

direction, the tidal forces cause it to fall onto lower tracks. The spin angular momentum of Neptune takes the opposite sign than the orbital angular momentum of the Neptune-Triton system. According to the law of total angular momentum conservation, the Neptune-Triton distance should decrease.

Nevertheless, it is a big mystery how such a huge body with a diameter of 2705 km could be captured at a distance greater³ than the radius 354 760 km of its current orbit. Triton has probably orbited Neptune for a very long time, because the eccentricity of its orbit is almost zero:

$$e = 0.000\,016.$$

This is the smallest eccentricity of all known bodies in the Solar system. When Triton was captured, its orbit was almost certainly an elongated ellipse and it took billions of years for Triton to reach a circular orbit.



Fig. 15.4. The large moon Triton with a diameter of 2705 km orbits Neptune (bottom left) in the direction opposite to the direction in which Neptune rotates (photo NASA).

There is again a quite simple explanation. Repulsive antigravitational forces continually act on Triton and, moreover, it is conceivable

³For comparison, our Moon was probably formed after a huge collision with a very massive body about the size of Mars leading to the Moon originally having an orbit at a distance of about 20 000 km above the Earth more than 4 billion years ago. Then the Moon has traveled to its present distance of 384 402 km (see Chapter 12).

that at present they are even larger than the tidal forces that push Triton to Neptune, depending on initial sizes of tidal and antigravitational forces. In this way, the Neptune-Triton system could obtain its huge observed orbital angular momentum.



15.7. Local Hubble expansion of the Solar system

The long-term influence of antigravity in the Solar system has left a number of other footprints which are recorded in various physical characteristics of the planets and other bodies [136]. We present further arguments supporting the expansion of the Solar system.

For instance, the rotation of Mercury is very slow (59 days) which could be caused by a very large impact in the very early stage of its development. However, Mercury has no tectonic activity that would renew its 4.5 billion years old surface which is now uniformly dotted with craters. Thus, another possibility is that its slow rotation is due to antigravity, because Mercury was once closer to the Sun, and then due to antigravitational forces it slowly moved to a higher orbit with a semimajor axis of about 57.9 million km. Because tidal forces are decreasing with the cube of the distance from the Sun, they act on Mercury by a $(149.6/57.9)^3 \approx 17$ times larger value per unit mass than on the Earth. Moreover, if Mercury were e.g. only 40 million kilometers from the Sun at the time of its creation, which is in accordance with (15.2), then the tidal forces from the Sun would be $(57.9/40)^3 \approx 3$ times greater than today. Altogether, we get a $3 \cdot 17 = 51$ times larger tidal influence per unit mass than of that acting on the Earth at present. This would significantly slow the rotation of Mercury, which has about 100 times smaller moment of inertia than the Earth.

If the Earth was once closer to the Sun (see Chapters 13 and 14), Venus could not be at its current distance of 108 million km from the Sun, since it would have an unstable orbit. Therefore, Venus was also closer to the Sun. Mercury and Venus have no moons, since the

corresponding lunar orbits would be unstable due to tidal forces when they were closer to the Sun.

If Mars was originally much closer to the Sun than it is now (see Chapter 11), Jupiter would also have to be closer to the Sun, since otherwise Mars would have grown to a larger size. Mars has only one-tenth of Earth's mass, because a more massive Jupiter has taken material which could otherwise have been used for building up the planet Mars. Not only Jupiter, but also Saturn, Uranus, and Neptune could also more easily collect their immense mass closer to the Sun. The increase of their orbital periods has destabilized the asteroid belts, which led to the bombardment of the inner planets. Another conjecture states that such a bombardment was caused by the passage of a close star. Both conjectures could have happened independently.

According to [20, pp. 409 and 534], there is strong evidence that the Kuiper belt of comets (at 30 to 50 au from the Sun) was formed much closer to the Sun in a region in which bodies orbit with larger velocities. Antigravity (cf. (15.2) or (11.4)) can explain a shift of at least 10 au during the last 4.5 Gyr. A similar argument applies to Sedna-like bodies and the Oort cloud, whose origins are not easily understood.

The Pioneer probes are delayed about half a day behind the position calculated by Newton's theory. This is the so-called "Pioneer effect". However, it is probably not caused by gravity, because it is only a very short-term phenomenon compared with the age of the Solar system. The Pioneers are slowed likely by the heat radiation from a radioactive source, which is asymmetrically located on the probes. Slowing of the probes by interplanetary dust has also to be taken into account.



16. Expansion of single galaxies

*No amount of experimentation can prove me right,
a single experiment can prove me wrong.*

ALBERT EINSTEIN

16.1. Do single galaxies expand due to antigravity?

The positive answer to this question is based upon several independent arguments that are presented in the following individual sections.

First, there is no reason to assume that antigravity would somehow not be present in the interior of galaxies, since its manifestations are observed not only at large cosmological distances (see Chapter 10), but also inside the Solar system (see Chapters 11–15 for the local expansion). In spite of that it is generally claimed that galaxies do not expand, because they are gravitationally bound and that only the space between them expands. Galaxies are usually included in clusters that should also not expand, because they are gravitationally bound as well (see Chapter 7). Galaxy clusters are gravitationally bound again in superclusters. So where does the universe expand?



16.2. Galactic expansion

The first catalog of hand-drawn shapes of celestial nebulae was created by William Herschel. Later Charles Messier made another famous catalog. When Hubble discovered that some of these nebulae are

galaxies (i.e. stellar islands very distant from the Milky Way), he classified them as elliptic, lenticular, spirals without bars, barred spirals,¹ and irregular galaxies. Small dwarf galaxies (such as the Magellanic Clouds) are generally irregular. Nevertheless, the strong gravitational potential of giant galaxies tends to neutralize any fluctuations and creates nice symmetric structures when no other galaxy is nearby. At present, it is not known why this is so specific for spiral galaxies.

The vast majority of large rotating galaxies have two spiral arms and they are approximately point symmetric with respect to their center. Exceptionally, however, there are galaxies with three spiral arms (see e.g. NGC 5054),² or four or more. The deeper we look into the early universe, the simpler are the shapes of the observed galaxies.

In recent times, we have obtained many astronomical observations documenting the expansion of galaxies themselves. For instance, R. J. Bouwers et al. [30] found that galaxies grow slowly. Also I. Trujillo with his team [279] discovered that the size of massive galaxies increases with time. This increase can be partially explained by intergalactic dust that falls into galaxies due to gravity, as well as galactic cannibalism. However, galaxies at cosmological distances have on average more stars per unit volume when compared with the present situation.

By [78] superdense galaxies were quite common in the early universe with redshift $z > 1.5$. At present they are quite sparse. Paper [283] also suggests that early galaxies were smaller and denser just after their formation. According to [39], the mass density of some galaxies for $z > 1$ is even comparable with the density of globular clusters themselves, i.e., on average several stars per pc^3 (at the center of a globular cluster about a hundred of stars per pc^3 , cf. Fig. 16.3).

¹Almost no galaxy has a bar for the redshift $z > 1$, while almost 80 % of spiral galaxies for $z \approx 0$ have bars. In addition, their bulges are slowly growing. It appears that the larger a spiral galaxy is, the greater its central black hole is.

²Note that the solution of the three-body problem of three equal masses moving along a circle at intervals of 120° is unstable. Gravity on large distances acts differently than that described by Newton's theory due to the finite speed of its propagation.

On the basis of this enumeration (see also [59], [90], [278], etc.) we may suppose that antigravity essentially contributed to the above-mentioned expansion and thus led the interiors of the galaxies themselves to become more tenuous. Similarly as in the Solar system, the rate of expansion of galaxies themselves may be smaller than the Hubble expansion, but it may also have the same order. This would support the picture that galaxies “swell up” very slowly like sea sponges or bread rising in the oven.



16.3. Expansion of the Milky Way

By [241] the measured density of stars in galaxies for a redshift $z \approx 3$ is on average approximately eight times larger than in galaxies in our neighborhood. Therefore, these galaxies are seen to have been roughly a factor of two times smaller in every direction than they are at present. By Fig. 8.7 they correspond to distances of about 11 Gly.

Let us apply this observation to our Galaxy, i.e., the Milky Way whose present diameter is about

$$D = 10^5 \text{ ly.} \quad (16.1)$$

Assume that the Galaxy has grown up from some smaller protogalaxy with diameter $d = D/2$ during the last 11 Gyr. Now we show that the present size D can be reached by the Hubble expansion. The present value of the Hubble constant is

$$H_0 \approx 70 \text{ km s}^{-1} \text{Mpc}^{-1} \approx 20 \text{ km s}^{-1} \text{Mly}^{-1},$$

since $1 \text{ pc} = 3.262 \text{ ly}$. Hence, its value rescaled on the size of our Galaxy is

$$\boxed{H_0 = 2 \text{ km s}^{-1} D^{-1}.} \quad (16.2)$$

Since the speed of light is about $c = 300\,000 \text{ km/s}$, for $t = 11 \cdot 10^9 \text{ yr}$ we get by (16.1) and (16.2) that the current extrapolated size of the Galaxy would be about

$$d \exp(H_0 t) = d \exp\left(\frac{2 \cdot 11 \cdot 10^9}{300\,000 \cdot 10^5}\right) = d e^{11/15} \approx 1.04D,$$

which is indeed a value comparable with the actual value D . Taking into account that the Hubble parameter was larger than H_0 (see Fig. 8.7), we may get the estimated diameter of our Galaxy close to the actual diameter D with a half-size of the Hubble expansion or a smaller size of the original protogalaxy. Therefore, the current expansion rate of our Galaxy seems to be comparable with H_0 , even though it is probably slightly smaller.

Using the measured metallicity of the Sun, the structure of the Oort Cloud, and several other arguments, the papers [113] and [235] conclude that since its creation 4.6 billion years ago the Sun gradually migrated several kpc from the center of the Galaxy to its current distance of 8.3 kpc. This again is comparable with the rate of the Hubble expansion. For instance, if the Sun were shifted about 2 kpc, then the corresponding rate of migration would be (cf. (10.2))

$$H_0^{(\text{loc})} = \frac{2 \text{ kpc}}{4.6 \text{ Gyr} \cdot 8.3 \text{ kpc}} = \frac{1}{19 \text{ Gyr}} = 0.71 H_0. \quad (16.3)$$

Even if the Solar system moved during its existence only about 1 kpc, or on the other hand about 5 kpc, it would still be of an order of magnitude comparable with H_0 .



16.4. Distribution of galaxies in the past

The density of galaxies in space 10–13 Gyr ago was much higher than at present, since the universe was smaller. For instance, for redshift $z \approx 3$, which corresponds to Hubble Deep Fields HDF, HDFS, HUDF, XDF, space was in any direction $(z + 1)$ times smaller and it contained on average $4 \cdot 4 \cdot 4 = 64$ times more galaxies per unit volume. However, since protogalaxies were smaller at that time, their larger packing is not observed (see Fig. 16.1).

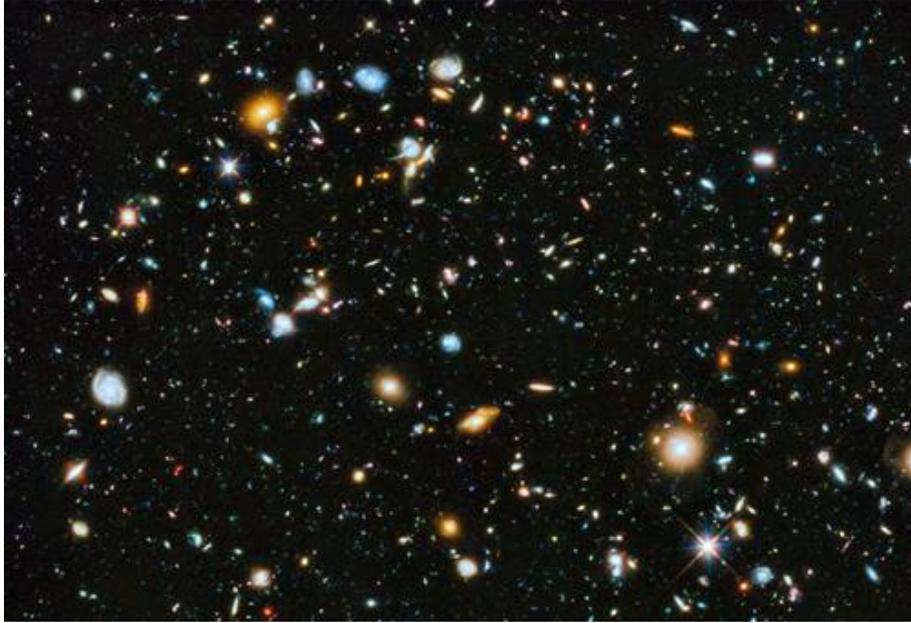


Fig. 16.1. The Hubble Deep Field spans $2.5'$ in the horizontal direction. It contains most galaxies with redshift $z \approx 3$ (see [25], photo NASA).

Further, we shall employ a geometrical argument based on proof by contradiction. Suppose for a moment that galaxies have constant volume (i.e., they do not expand with time), that they do not collide and that the universe is homogeneous and isotropic for every time instant. The right part of Fig. 16.2 illustrates what we would observe at cosmological distances, if galaxies would have the same size all the time. On the left, five galaxies are schematically depicted in the unit cube for $z = 0$. Thus for $z = 2$ there would be on average $5 \cdot (z + 1)^3 = 5 \cdot 3^3 = 135$ crowded galaxies in the unit cube, since the middle cube from Fig. 16.2 can be placed into the left unit cube $3^3 = 27$ times. Analogously we find that for $z = 4$ there would be $5 \cdot 5^3 = 625$ tightly packed galaxies in the unit cube. However, such a tight arrangement is not observed, since galaxies were smaller at that time. Moreover, there exist galaxies with $z \approx 10$. In this case the number of galaxies in the unit cube would be more than 1000 times greater than at present and galaxies could touch or even penetrate each other, if their sizes would be constant. This is in contradiction to observations.

We may also argue in the following way: The volume of our Galaxy is about $10^5 \times 10^5 \times 10^3 \text{ ly}^3$. The total volume of all 10^{12} observed galaxies can thus be approximated to 10^{25} ly^3 . Hence, if the radius of this part of the universe was smaller than 10^8 ly and protogalaxies were to have the same size as today's galaxies, they would necessarily interpenetrate each other, which would lead to a contradiction.

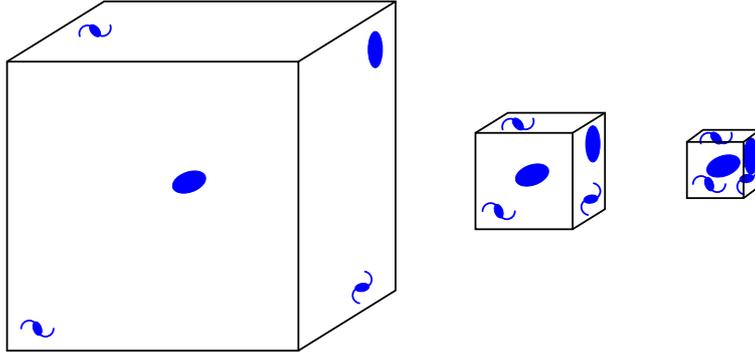


Fig. 16.2. A schematic illustration of nongrowing galaxies of constant size over time in an expanding universe. The unit cube is on the left. It contains several galaxies in our neighborhood for the redshift $z = 0$. The distribution of galaxies at cosmological distance $z = 2$ is in the middle and for $z = 4$ on the right. Such a picture of tightly crowded galaxies has not been observed by astronomers.



16.5. Star-formation rate

According [30] and [274] the star-formation rate in galaxies at cosmological distances is proportional to $(1 + z)^{1.9 \pm 0.1}$. For instance, for the redshift $z \approx 2.3$ the rate is 10 times larger than in our neighborhood. Such a large rate can again be explained by a higher mass density inside galaxies for large z than for $z \approx 0$ corresponding to the present.

One of the largest known galaxy clusters, SPT-CLJ 2344-4243 (Phoenix), is located at a distance of $z = 0.6$. It has an immense mass $2.5 \cdot 10^{15} M_{\odot}$, where $M_{\odot} = 1.989 \cdot 10^{30} \text{ kg}$ is the Sun's mass. In its central galaxy the ongoing star formation rate is about $740 M_{\odot}$

per year. The star formation rate is even higher in some very distant galaxies. For example, in the galaxy HFLS3 new stars are born 2000 times more frequently than in the Milky Way, which indicates a large interior mass density.



16.6. Activity of galactic nuclei

According to [256], the observed activity of galactic nuclei at cosmological distances is much greater than in our neighborhood, with $z \approx 0$. This may also be explained by the higher density of matter inside galaxies with large redshift z , although their central black holes have been smaller than at present.

A large number of quasars with redshift $z \geq 6$ are known (for instance, J1148+5251, J1319+0950). Most of them were found by the submillimeter interferometer system ALMA that works with the impressive accuracy of $0.6''$, which corresponds to 3 kpc at the distance $z = 6$ for the currently accepted values of the cosmological parameters. The quasar J2310+1855 has the largest known luminosity of $1.8 \cdot 10^{13} L_{\odot}$, where $L_{\odot} = 3.846 \cdot 10^{26}$ W is the Sun's luminosity, which is two orders of magnitude higher than that of a usual galaxy. Thus it seems that the interior density of distant galaxies was quite high and then gradually decreased as a result of antigravity, even though the luminosity of quasars could also decrease when they swept out the material in their surroundings. Both of these processes presumably took place simultaneously.



16.7. Gravitermal catastrophe

Almost all open and globular clusters seem to be unstable, as was shown by Pavel Kroupa [149]. Smaller stars tend to move away from the center of a cluster, while the massive ones gradually accumulate

around the center. This leads to mass segregation. During multiple collisions, light-mass stars are sometimes even ejected from the cluster which is usually called *evaporation of stars* in the clusters. The surrounding, heavier stars drift closer to the center due to loss of potential energy. This process ends by the so-called *gravithermal catastrophe* (see [87]).



Fig. 16.3. Globular cluster M13 in the constellation Hercules

However, the local expansion causes all the stars on average to move away from the center. Hence, it would act against the gravithermal catastrophe inside the cluster and it would slowdown the entire process. Some globular clusters³ have existed for more than 13 billion years and their gravithermal catastrophe has not been manifested (see Fig. 16.3).



16.8. Exoplanet WASP-18b

Another example of the local influence of antigravity inside our Galaxy is the exoplanet WASP-18b, which orbits its mother star with mass $1.25M_{\odot}$ along a nearly circular trajectory with a radius of 3 million km and a period of 0.94 days. Because the star rotates about

³There exist about 150 globular clusters in the Milky Way. They are very old systems and contain hundreds of thousands or up to millions of stars.

its axis once every 5.64 days (see [37]), by Kepler's third law the corresponding stationary orbit has a radius of about 10 million km, i.e., the exoplanet orbits the star under the stationary orbit (cf. relation (15.1)). According to [100], the exoplanet should spiral down onto its mother star in less than a million years due to tidal forces. The star, however, has already existed for about 700 million years [263]. It is a mystery how this exoplanet (weighing about 10 Jupiters) could ever get to its orbit and why it will spiral down onto its mother star in such a short geological time.

This paradox can again be explained by the fact that antigravity acts in the opposite direction than the tidal forces, and thus it protects the exoplanet against collision. From the evolution of all orbital parameters we will be able to estimate after some time how quickly the exoplanet approaches to its mother star and how much energy is dissipated by tides. Several other exoplanets discovered by the Kepler telescope are below the stationary orbit of their mother stars. Other examples are WASP-19, WASP-103, CoRoT-7, and so on.

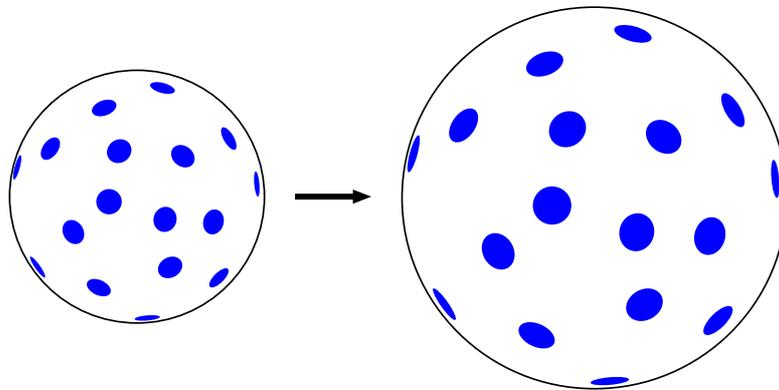


Fig. 16.4. A two-dimensional model of an expanding universe with positive curvature and swelling galaxies. In [184, p. 719], galaxies have constant sizes over time.

The arguments presented in previous sections show that galaxies themselves grow (see Fig. 16.4) although probably by a somewhat lower rate than the Hubble parameter. Antigravity, which very slowly but steadily increases the total energy (i.e. kinetic + potential) of each coupled system of gravitationally interacting bodies, causes each

cluster to slightly expand on average. Antigravity forces thus seem to act locally also in our Galaxy and in other star systems. They can contribute to the expansion of habitable zones (as in the case of the Earth in Fig. 13.2), while the luminosity of the mother star is gradually increasing. In this sense, the habitable zones are then more stable, since they may exist for a longer time period (see Chapter 14).



17. What is the mysterious source of dark energy?

All truth passes through three stages:

First, it is ridiculed.

Second, it is violently opposed.

Third, it is accepted as self-evident.

ARTHUR SCHOPENHAUER

17.1. Gravitational aberration

In Chapters 11–16 we have presented many examples that support the hypothesis that antigravity does not act only globally, but also locally. Based on the present state of knowledge, this suggests that the law of conservation of energy need not hold or we do not know where the energy comes from. Now we show that dark energy which is required for the accelerating expansion of the universe may (at least partially) come from a small but positive value of gravitational aberration, which is a consequence of causality and the finite speed of propagation of gravitational interaction.

Consider first only two bodies A and B of equal masses that orbit symmetrically with respect to their center of gravity. If A attracts B and B also attracts A at their instantaneous positions (i.e., when the *speed of gravitational interaction* c_G is infinite), then by the Newtonian theory of gravity these forces are in the same line and in balance.

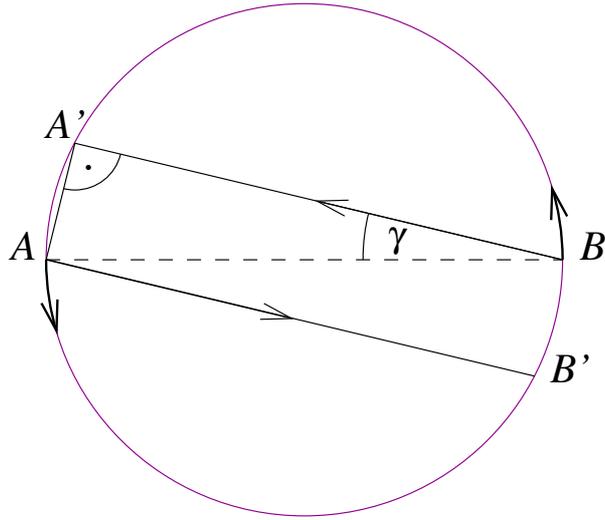


Fig. 17.1. Schematic illustration of two interacting bodies of equal masses. The gravitational aberration angle $\gamma = \angle ABA'$ is extremely small.

However, the speed of gravitational interaction c_G is actually finite. The body B is attracted by A towards its previous position A' as depicted in Fig. 17.1. Similarly A is attracted by B towards its previous position B' . Then a couple of non-equilibrium forces arise that permanently act on this system. It increases the total angular momentum and the total energy of the system.

The angle ABA' (resp. BAB') is called the *angle of gravitational aberration*. According to Thales' theorem, the triangle AAB' on Fig. 17.1 is a right triangle and

$$|A'B| < |AB|. \quad (17.1)$$

Hence, by (17.1) the attractive forces (in this postnewtonian model) are slightly larger than if they were to act along the hypotenuse AB .

Let us point out that Fig. 17.1 is slightly imprecise. An arbitrarily small positive value of the gravitational aberration angle γ of the considered binary system gradually also prolongs the orbital period. Hence, the corresponding trajectories form two very slowly expanding spirals (see Fig. 17.2).

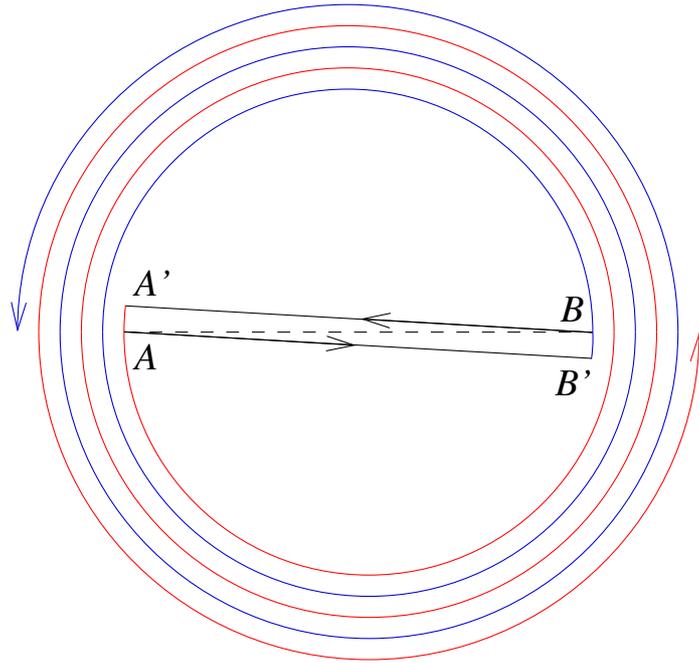


Fig. 17.2. Trajectories of two interacting bodies of equal masses form a double spiral. The distance between two neighboring arms is actually much smaller than in the picture. The gravitational aberration angles ABA' and BAB' are extremely small but positive.

This simple example (considered by Arthur Eddington [68, p. 94]) shows why the classical laws of conservation of energy and of momentum break down for a finite speed of gravitational interaction. If they were not to be disturbed, then the trajectories of the two bodies would be a fixed circle of constant radius for suitable initial conditions. For expanding spiral trajectories the total kinetic energy decreases, but the total potential energy increases two times more quickly (cf. (13.23)).

The above example can be modeled by a system of ordinary differential equations with delay (see (17.4)–(17.5) below). The considered binary system can be generalized to the case of more bodies of unequal masses. It can be again shown that the total energy (i.e. kinetic + potential) of such a system increases [129, p. 243].

A system of equations with nonnegative delays models reality better than the classical Newtonian system of ordinary differential equations (5.8), since it enables us to consider a finite speed of gravitational

interaction producing the gravitational aberration, and to generate expected spiral trajectories (see e.g. Chapters 12 and 13). Moreover, for vanishing gravitational aberration we get the classical Newtonian system. A positive gravitational aberration has a repulsive character like the positive cosmological constant. On the other hand, the proposed delay equations do not describe some effects of the General theory of relativity due to curved spacetime.

The described mechanism contributes also to the expansion of the Universe and can explain (at least partly) the mystery of dark energy. The actual angle of gravitational aberration has to be necessarily positive, since an angle of aberration equal to zero contradicts causality. Assume for a moment that the body A from Fig. 17.2 exploded. Then the second body B has to move some time along the unchanged trajectory until it receives information by means of the gravitational field about the change of the trajectory of A . Therefore, the aberration angles ABA' and BAB' in Fig. 17.2 have to be positive.

In the framework of General relativity Steven Carlip in [44] derived that the gravitational aberration angle γ of a body with speed v is bounded from above by the fraction v^3/c^3 , i.e.

$$\gamma = o\left(\frac{v^3}{c^3}\right), \quad (17.2)$$

while the angle of light aberration is approximately equal to

$$\alpha = \frac{v}{c} \quad (17.3)$$

by (2.12). In doing so, he assumes that

- a) the gravitational interaction has the same speed of propagation as light,
- b) the cosmological constant is zero,
- c) some nonlinear terms, which he cannot estimate, are vanishing,
- d) the laws of conservation of energy and momentum hold.

Therefore, he does not suggest spiraling trajectories as schematically depicted in Fig. 17.2.

The recession speed and the angle of gravitational aberration depend on masses, velocities, and positions (also in the past) of all free bodies that gravitationally interact among themselves [129]. Gravitational aberration thus has a nonnegligible influence on the expansion rate of the universe. Therefore, it is also necessary to look at the cosmological constant as just a quantity that is averaged over all mass objects and not as a fundamental physical constant (such as e.g. the gravitational constant G).



17.2. Postnewtonian model or how energy is generated

In this section we show how gravitational aberration can be modeled mathematically by modifying the system of differential equations (5.8). The resulting problem involving a finite speed of gravitational interaction c_G between two bodies will now be described by the following system (17.4)–(17.6) of ordinary differential equations with delay.

Consider only two mass points m_1 and m_2 in two-dimensional or three-dimensional Euclidean space with the standard norm $|\cdot|$ representing the distance. Introducing a delay into gravitational interactions, the classical Newtonian system (5.8) for $N = 2$ can be rewritten by the system of ordinary differential equations for two vector trajectories r_1 and r_2 :

$$\ddot{r}_1(t) = G \frac{m_2[r_2(t - d_2(t)) - r_1(t)]}{|r_2(t - d_2(t)) - r_1(t)|^3}, \quad (17.4)$$

$$\ddot{r}_2(t) = G \frac{m_1[r_1(t - d_1(t)) - r_2(t)]}{|r_1(t - d_1(t)) - r_2(t)|^3}, \quad (17.5)$$

where d_1 and d_2 are two time variable delays satisfying some natural conditions (see (17.7) below). The initial conditions are

$$r_i(t) = p_i(t), \quad \dot{r}_i(t) = v_i(t), \quad t \in [t_0, 0], \quad i = 1, 2. \quad (17.6)$$

Here $t_0 \leq 0$ is an appropriate given number and p_i and v_i are given vector functions characterizing previous positions and velocities.

This simple postnewtonian model does not take into account gravitational waves (which have not yet been directly detected), but involves a nonzero gravitational aberration and a finite speed of gravitational interaction not necessarily equal to c .

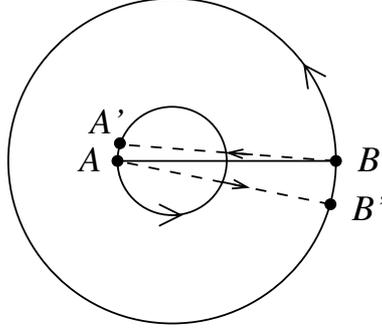


Fig. 17.3. Schematic illustration of gravitational interaction between two bodies of unequal masses $m_1 > m_2$

If $c_G = \infty$ then we obtain $t_0 = d_1 = d_2 = 0$ and the system (17.4)–(17.6) reduces to the classical Newtonian two-body problem (see Chapter 5). For $c_G < \infty$ the delay function satisfy the relations (cf. Fig. 17.3)

$$d_1(t) = \frac{|r_1(t - d_1(t)) - r_2(t)|}{c_G}, \quad d_2(t) = \frac{|r_2(t - d_2(t)) - r_1(t)|}{c_G}, \quad (17.7)$$

i.e., each d_i has to be calculated iteratively using e.g. the classical Banach fixed-point theorem [141].

Assume now that

$$m_1 R_1 = m_2 R_2, \quad (17.8)$$

where R_1 and R_2 are the distances from the Newtonian center of gravity. Let us define

$$p_1 = (R_1, 0), \quad p_2 = (-R_2, 0), \quad v_1 = \left(0, \frac{\sqrt{Gm_2 R_1}}{R_1 + R_2}\right), \quad v_2 = \left(0, -\frac{\sqrt{Gm_1 R_2}}{R_1 + R_2}\right).$$

These values yield exactly circular orbits for $t_0 = 0$ in (17.6) and $c_G = \infty$. They are employed to establish the initial conditions (17.6) for $c_G < \infty$. This, of course, requires one to store old values of r_1 and r_2 throughout the computation due to nonstandard initial conditions.

A big advantage of computer simulations is that we can easily perform many tests for various parameters appearing in (17.4)–(17.7).

For instance, we may arbitrarily change the speed of gravitational interaction c_G in the proposed postnewtonian model which need not agree with the real (yet unmeasured) value. It is only an input parameter.

Example 17.1. The analytical solution of problem (17.4)–(17.7) for $c_G < \infty$ is not known. Numerically calculated trajectories r_1 and r_2 for $m_1 = m_2 > 0$ and $c_G \leq c$ are depicted in Fig. 17.2. We see that they are quite unrealistic, since they form two quickly expanding spirals, which does not correspond to astronomical observations. However, model (17.5)–(17.7) yields quite satisfactory results for $c_G \gg c$. Such a solution, of course, differs from the solution of (17.4)–(17.5) with initial conditions $r_i(0) = p_i(0)$, $\dot{r}_i(0) = v_i(0)$, $i = 1, 2$, for $c_G = \infty$.

Example 17.2. The largest value of gravitational aberration is obtained when $m_1 \approx m_2$ (cf. Fig. 17.2). Nevertheless, in the Solar system such objects do not exist. For the Earth–Moon system the ratio $m_1 : m_2$ equals 81 : 1 and for the Pluto–Charon system 8 : 1. This is the smallest ratio greater than 1 in the Solar system among larger bodies.

Consider again the binary system Earth–Moon with the masses given by (12.4) and the mutual distance 384 402 km. To get the speed of recession derived in (12.21), we have to take $c_G = 4.287 \cdot 10^{15}$ m/s for the considered postnewtonian model. In this case the gravitational aberration angle at the point B representing the Moon in Fig. 17.3 is

$$\gamma = \frac{v}{c_G} \approx 2.424 \cdot 10^{-13} \text{ rad}, \quad (17.9)$$

where $v = |\dot{r}_2| \approx 1$ km/s and the two trajectories form two very slowly expanding spirals. Note that the aberration angle of light of the Moon is much bigger, $\alpha = v/c = 0.7''$ (see (6.2)), and the aberration angle of the Earth is 81 times smaller.

It is not difficult to generalize the previous problem (17.4)–(17.7) to any number $N \geq 2$ of interacting bodies. This can be done similarly as in Section 5.4.

Example 17.3. In numerical experiments we observed expanding trajectories for three bodies of equal masses that are at the vertices of an equilateral triangle and that orbit at the same speed about their center of gravity. A similar phenomenon was achieved for the case marked in Fig. 17.2, where the third body lies at the midpoint of AB . Expanding trajectories were also obtained for a system consisting of two double stars of equal masses.

Further examples are given in [129]. The finite speed of gravitational interaction slightly protects stars against collisions, since they are not able to react so quickly upon changing positions due to delay. The probability of their collision is smaller than if they would react immediately (see [129, p. 242]). All calculations were done in extended 10 byte precision by the standard fourth order explicit Runge-Kutta method which gives a surprisingly small discretization error when the orbits are circular. Popular symplectic methods are not suitable for these purposes, because they conserve energy, and in addition have a lower order of approximation.

Each dynamical system of free particles, which act only gravitationally, expands on “average” due to gravitational aberration, because the bodies interact mutually with delays. The term on “average” should be understood as a gradual long-term (i.e. secular) change in distances. For instance, for a finite speed of gravity the two bodies shown in Fig. 5.2 would be alternately approaching and receding, but their trajectories would take the shape of slightly evolving elliptical spirals.

On the other hand, our Earth does not expand, even though it is held together mainly by gravitational force. It should be seen as a stationary object and not as a system of free particles.



17.3. The speed of gravitational interaction

In 1805, Pierre Laplace decided on the basis of a detailed analysis of the motion of the Moon that the actual Newtonian velocity of propagation of gravitational interaction c_G must be at least $7 \cdot 10^6 c$ (see [157, Chap. VII, p. 642]). Van Flandern [285] have increased this estimate to $2 \cdot 10^{10} c$, since otherwise two separate bodies would be not orbiting around a common center of gravity on relatively stable orbits. Their total angular momentum would not be in equilibrium for $c_G = c$. This equality is suggested by the General theory of relativity (see (17.10)).

On the other hand, in 1905 Henri Poincaré¹. in [218, p. 1507] conjectured that for the speed of gravitational interaction c_G we have to use²

$$c = c_G, \quad (17.10)$$

where c is the speed of light in the vacuum. Note that Albert Einstein came to the same conclusion later. However, a natural question arises concerning what is the meaning of this equality. Is it valid e.g. to twenty significant digits? If the velocities c and c_G differ only slightly (even though only 0.001 %), it would be rather difficult to identify the source of gravitational waves (e.g. during a supernova explosion) with their optical counterpart. A superluminal speed of gravitational interaction may explain why some spiral galaxies are so perfectly symmetric. For the speed (17.10) communication through gravity between the ends of spiral arms would last hundreds of thousand of years. It is also not clear what is the speed of gravitational interaction in real matter (glass, the interior of the Sun, etc.), where light exhibits dispersion.

¹The well-known Einstein's formula $E = mc^2$ of 1905 was discovered by Poincaré [217] five years earlier. The radiation energy flux is $S = Ec$. Poincaré assumed that the emitted radiation carries a momentum p . By Poynting's theorem, this momentum is $p = mc = S/c^2$, i.e. $E = mc^2$.

²Note however that not all physical interactions have the same speed of propagation. For example, the weak interaction is transmitted by the intermediate bosons W^+ , W^- , and Z^0 , which are approximately 80–90 times heavier than the proton and thus cannot move at the speed of light.

The speed of gravitational interaction could be stated for binaries with a known distance d between the two components as follows. If one star explodes, then it is enough to just measure the time period τ in which the second star begins to change its original path. Contemporary Doppler techniques allow one to measure changes in the radial velocity up to 1 m/s. From this academic example we can then determine

$$c_G = \frac{d}{\tau}.$$

At present several costly projects are being prepared or underway (GEO, LIGO, NGO, VIRGO, ...) to measure the speed of gravitational waves and to determine the direction from where they come. However, so far they have not been detected, and thus their actual speed is not known.

Gravitational interaction behaves very differently than electromagnetic interaction does. It is generally believed that the gravitational force is only attractive, while the electromagnetic force is both attractive and repulsive. Nevertheless, antigravity also has a repulsive character even though very small. It is not a new fifth physical force but only a side effect of gravitational forces due to the finite speed of gravitational interaction. Another difference between gravitational and electromagnetic interactions, however, are the aberration phenomena.

Suppose for a moment that the star in Fig. 6.3 asymmetrically explodes and that electromagnetic and gravitational waves have the same speed of propagation (17.10) as stated by the General theory of relativity. Then the two types of the respective wavefronts from it will propagate at the same speed. The angle of light aberration α will at the same time be relatively large (see (17.3)).

Assume further that the telescope in Fig. 6.3 is replaced by an instrument that can detect the direction from which the gravitational waves come. Then we find that they will come from the same direction as electromagnetic waves provided (17.10) holds. The angle of gravitational aberration should be the same as α . The attractive

force of the star paradoxically acts from a slightly different direction, otherwise no system of two bodies would be relatively stable. Indeed, if $\alpha = \gamma$, then the Earth would move about 150 million km away from the Sun in 400 years (see [162, p. 350]).

According to Newtonian mechanics, the gravitational aberration angle γ is zero. If the causality principle is true, the actual angle of gravitational aberration γ should be positive, although extremely small (see (17.9)). Consider e.g. the Sun-Jupiter system, whose center of gravity according to (5.1) is located outside of the Sun. Both bodies continually deform the spacetime around them by their strong gravitational fields and the exchange of information about their positions is transmitted with delays. Because $c_G < \infty$, the Sun is attracted towards some previous position of Jupiter and Jupiter is in turn also attracted by the Sun towards some previous position of the Sun (see Fig. 17.3). According to Example 17.2, Jupiter should have the largest expansion velocity of its orbit among all planets recalculated to 1 au. This is also the reason why it could leave behind itself an uncleared asteroid belt between Mars and Jupiter (cf. [20, p. 513]).

Let's ask another question:

Is Earth attracted to the Sun exactly in the direction where we see it, or is the vector of the gravitational force pointing slightly off center?

As a result of the properties of gravitational aberration, light rays coming to us from the Sun are not parallel with the vector of the Newtonian gravitational force between the Sun and Earth. These paradoxical phenomena are caused by the fact that the actual angles of light and gravitational aberration differ. Earth has a 333 000 times smaller mass than the Sun, and thus moves in a nearly stationary gravitational field, even though Earth continually deforms it locally. In this case the gravitational aberration phenomenon manifests itself relatively little. Therefore, $\alpha \gg \gamma > 0$ and the Sun does not attract us in that direction, where we see it. How to interpret these paradoxical data and determine whether they are not in conflict with causality

is discussed in articles [44] and [170]. More details about aberration phenomena and their interpretation is also discussed in [285].



17.4. Does the law of conservation of energy hold?

In Chapters 11–15 we showed many examples which can be interpreted so that the Solar system is gradually expanding as a result of antigravity at a rate comparable to the order of the Hubble constant. This apparently contradicts the law of conservation of energy (cf. (13.24)) and the law of conservation of total angular momentum. Thus, the total energy of the universe very slowly but continually grows. This also indicates that the equivalence principle³ could be slightly violated.

It is said that the source of dark energy, which causes the accelerating expansion of the universe, is not known yet. There are many different hypotheses (e.g. variable fundamental physical constants, vacuum energy, the influence of quintessence) that attempt to explain the mystery of dark energy. Their overview is given e.g. in [5]. Also, Richard Panek [199] surveys nearly 50 models of dark energy. In this chapter we presented another hypothesis, which is based on the concept of gravitational aberration and which also suggests *where the energy needed for accelerated expansion of the universe comes from*.

Our Galaxy and the Solar system are unique astrophysical laboratories for testing whether or not the law of conservation of energy holds and whether a finite speed of propagation of the gravitational interaction generates as a byproduct the sought energy for general expansion. In a number of specific examples we have shown that this energy is generated not only globally, but also locally, e.g. the Sun–Earth system is continually creating energy. The idea of a local expansion of the universe first appeared already in 1933 in the article [175] and it should be further verified.

³The equivalence principle claims that we are not able to distinguish locally between a homogeneous gravitational field and a uniform acceleration of the coordinate system.

Recall (see Section 12.3) that there exist close binary pulsars whose orbital period, by contrast, decreases over time. In this case, the system loses kinetic energy due to the radiation of gravitational⁴ and electromagnetic waves, due to tides, friction by interstellar dust, and so on. The resulting forces then prevail over antigravity. Also, various resonances may cause much larger effects than antigravity, whose manifestations on short time scales are indeed negligible (see e.g. (12.21), (13.2), (15.4), and (15.8)). However, antigravity operates continually in any gravitationally bound system of asteroids, moons, planets, stars, galaxies, clusters of galaxies, or orbiting clusters of galaxies. It gradually increases its total (kinetic + potential) energy. In this way it contributes to migration of planets and their moons over long-term time periods, it causes that star clusters continually “evaporate” [149], it acts against gravithermal catastrophe of galactic clusters and star clusters [87], it reduces the frequency of collisions of stars [129] and galaxies, it slowly expands the “cosmic web”, it stabilizes the Solar system, and so on. Antigravity also helped create the conditions suitable for the emergence of life on Earth over a period of several billion years, during which solar power is growing (see Chapters 13 and 14).



⁴Note that the exact solution of Einstein’s equations of two-body problem is not known. The decrease in the orbital period is deduced only from some approximative formulae [275].

18. What the universe is

*Cosmologists are often wrong
but never in doubt.*

LEV LANDAU

18.1. Non-Euclidean models of the universe

In 1584 Giordano Bruno wrote the treatise [38], where among other things he conjectured that the universe is infinite. From that time, opinions of the shape of the universe have often changed. Isaac Newton and many others understood the universe as the Euclidean space \mathbb{E}^n for $n = 3$.



Fig. 18.1. Karl Schwarzschild (1873–1916)

In 1900, however, the German physicist Karl Schwarzschild (see [252, p. 66]) was probably the first who ever realized that the universe

might be non-Euclidean¹ and even finite, i.e. having a finite volume. He envisioned it as a three-dimensional manifold² (cf. (18.2))

$$\mathbb{S}_r^3 = \{(x, y, z, w) \in \mathbb{E}^4 \mid x^2 + y^2 + z^2 + w^2 = r^2\} \quad (18.1)$$

for $r > 0$, which is, in fact, a three-dimensional surface of the four-dimensional ball with time independent radius r (cf. Fig. 18.2). The manifold \mathbb{S}_r^3 has at any point and in any direction³ the same curvature $1/r$ (similarly \mathbb{E}^3 has at any point and in any direction zero curvature). This geometry enables us to model a universe with high homogeneity and isotropy on large spatial scales.

For every $n = 1, 2, \dots$ define a *sphere* (*hypersphere*) with radius $r > 0$ by

$$\mathbb{S}_r^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{E}^{n+1} \mid x_1^2 + \dots + x_{n+1}^2 = r^2\}.$$

For $r = 1$ we shall for simplicity write only \mathbb{S}^n . On the sphere \mathbb{S}^n we may consider a non-Euclidean geometry which is an analogue of the spherical geometry for the sphere \mathbb{S}^2 discussed in Section 2.10.

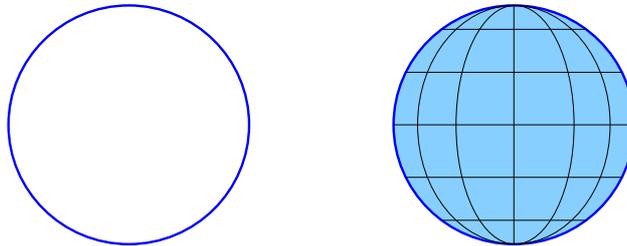


Fig. 18.2. The unit circle on the left is the sphere $\mathbb{S}^1 = \{(x, y) \in \mathbb{E}^2 \mid x^2 + y^2 = 1\}$. The surface of the unit ball is the sphere $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{E}^3 \mid x^2 + y^2 + z^2 = 1\}$.

¹Non-Euclidean geometry arose in the first half of the 19th century during many attempts to understand the axiomatic construction of Euclidean geometry — especially in proving the independence of Euclid’s fifth postulate of parallels. We include among its founders Carl Friedrich Gauss, Nikolai I. Lobachevskii, János Bolyai, Bernhard Riemann, Sophus Lie, Felix Klein, and many others. The history of the development of non-Euclidean geometries is described in detail in the review article [42].

²Recall that an *n-dimensional manifold* is a set of points for which there exists an open neighborhood that can be continuously mapped onto an open set in \mathbb{E}^n such that the inverse is continuous, too. An example of a manifold is the graph of a parabola, a hyperboloid of two sheets, the surface of a torus, and so on. On the other hand, the union of the hyperplane $x_1 = 0$ and axis x_1 in \mathbb{E}^n is not a manifold for $n > 1$. Also the set of rational numbers is not a manifold.

³The curvature of the sphere \mathbb{S}_r^3 at a given point and given direction is equal to the reciprocal value of the radius of the corresponding osculating circle.

The shortest connection of two points of a manifold is called a *geodesic*. It is a straight line in a Euclidean space. The shortest connections of two points on the sphere \mathbb{S}^n are arcs of great circles. Geodesics on a sphere are not uniquely determined, e.g., there are infinitely many shortest paths between the North and South Pole on the sphere \mathbb{S}^2 and they are represented by meridians.

Every two different great circles of the sphere \mathbb{S}^2 (i.e. “lines” in elliptic geometry) intersect in just two opposite points.⁴ Therefore, there are no parallels in elliptic geometry. The sum of the angles in a spherical triangle, whose sides are the shortest arcs of great circles, is larger than 180° . For instance, the sum of angles of the spherical triangle, which arises as the intersection of the sphere \mathbb{S}^2 and the octant in \mathbb{E}^3 , is 270° .

The circumference of a circle of radius $R \in (0, \pi)$ on the sphere \mathbb{S}^2 is smaller than $2\pi R$, since the radius of a circle is measured by the length of an arc of a great circle in \mathbb{S}^2 . For instance, from the right side of Fig. 18.2 it is evident that the length⁵ of the equator is 2π for the radius $R = \pi/2$ measured from the North or South Pole in the direction of the meridians of the sphere \mathbb{S}^2 . Also the area of the circle in \mathbb{S}^2 , the surface area and volume of the balls in \mathbb{S}^3, \dots are less than $\pi R^2, 4\pi R^2, 4\pi R^3/3, \dots$, respectively. Standard relations known from Euclidean geometry thus do not hold on the sphere \mathbb{S}^n (cf. Fig. 8.2).

Recall now the definition of a metric (distance).

Definition 18.1. The function $\rho : M \times M \rightarrow \mathbb{E}^1$ is said to be a *metric* on the manifold M if:

1. $\rho(A, B) \geq 0 \quad \forall A, B \in M$,
2. $\rho(A, B) = 0 \iff A = B$,
3. $\rho(A, B) = \rho(B, A) \quad \forall A, B \in M$,
4. $\rho(A, B) \leq \rho(A, C) + \rho(C, B) \quad \forall A, B, C \in M$ (triangle inequality).

⁴On the other hand, two great circles on the sphere \mathbb{S}^3 may not have common points, e.g., the circles $x^2 + y^2 = 1$ and $z^2 + w^2 = 1$. They can be interpreted as skew “lines”.

⁵In the Euclidean plane, the circumference of a circle of radius $\pi/2$ is $\pi^2 (> 2\pi)$.

For instance, in the Euclidean space \mathbb{E}^n the distance is defined by means of the generalized Pythagorean theorem

$$\rho(A, B) = \sqrt{\sum_{j=1}^n (a_j - b_j)^2} \quad \forall A = (a_1, \dots, a_n), B = (b_1, \dots, b_n) \in \mathbb{E}^n.$$

The distance between two points A and B on the sphere \mathbb{S}^n is given by the length of the associated geodesic connecting the points $A, B \in \mathbb{S}^n$,

$$\phi(A, B) = \arccos\left(\sum_{j=1}^n a_j b_j\right).$$

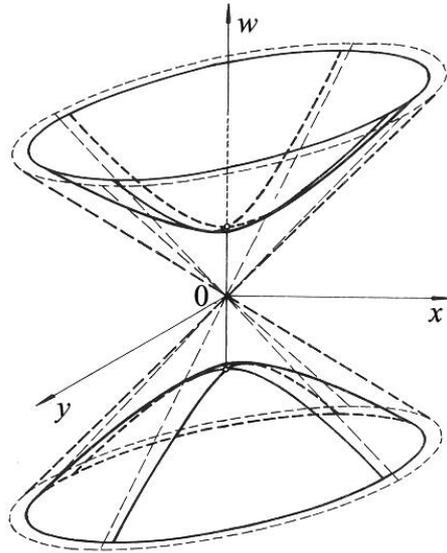


Fig. 18.3. The rotational hyperboloid of two sheets $x^2 + y^2 - w^2 = -1$ can be obtained from (18.2) for $z = 0$, $r = 1$, and $|w| \geq 1$.

Karl Schwarzschild [252, p.67] considered even the space with hyperbolic geometry on a *pseudosphere*, which for better visualization is usually represented by the hyperbolic hypersurface for $r > 0$ (cf. Fig. 18.3)

$$\tilde{\mathbb{H}}_r^3 = \{(x, y, z, w) \in \mathbb{E}^4 \mid x^2 + y^2 + z^2 - w^2 = -r^2\} \quad (18.2)$$

with the Minkowski pseudometric which will be defined in (18.6). Let us emphasize that w in (18.2) is a space coordinate and is **not time**

as it could seem from the often used confusing notation $t = w$ (see e.g. [291, p. 95]). If w were time, then the corresponding spatial manifold $w = \text{const.}$ would have the dimension only two. In the case $r = 1$ we will again for simplicity omit the lower index r .

Let us show now how the sphere \mathbb{S}_r^3 can be formally transformed into a part of the two-sheet hyperbolic hypersurface $\tilde{\mathbb{H}}_r^3$. Introducing hyperspherical coordinates (i.e. a natural generalization of the standard spherical coordinates)

$$\begin{aligned}x &= r \sin \chi \sin \theta \cos \phi, \\y &= r \sin \chi \sin \theta \sin \phi, \\z &= r \sin \chi \cos \theta, \\w &= r \cos \chi,\end{aligned}$$

we find for $\chi, \theta \in [-\pi/2, \pi/2]$ and $\phi \in [0, 2\pi]$ by (18.1) and the relation $\cos^2 \phi + \sin^2 \phi = 1$ that $(x, y, z, w) \in \mathbb{S}_r^3$. Applying the transformations $w \mapsto iw$, $r \mapsto ir$, and $\chi \mapsto -i\chi$ (see [202, p. 299], [231, p. 826]) and taking into account the relations $\cos \chi = \cosh i\chi$ and $\sin \chi = -i \sinh(i\chi)$, we get

$$iw = ir \cos(-i\chi) = ir \cosh(-i^2\chi)$$

which yields $(x, y, z, w) \in \tilde{\mathbb{H}}_r^3$, where

$$\begin{aligned}x &= r \sinh \chi \sin \theta \cos \phi, \\y &= r \sinh \chi \sin \theta \sin \phi, \\z &= r \sinh \chi \cos \theta, \\w &= r \cosh \chi.\end{aligned}$$

The points (x, y, z, w) determined by these relations lie on that part of the hyperbolic hypersurface $\tilde{\mathbb{H}}_r^3$ for which $\chi \in [-\pi/2, \pi/2]$. This can be verified by (18.2) and the well-known relations $\cos^2 \phi + \sin^2 \phi = 1$ and $\cosh^2 \chi - \sinh^2 \chi = 1$.

The manifold \mathbb{S}_r^3 can be divided into elementary spherical shells with areas $4\pi r^2 \sin^2 \chi$ and thickness $r d\chi$. Using the above-mentioned

hyperspherical coordinates it can be derived that the volume of \mathbb{S}_r^3 is (see [72, p. 152])

$$V = 4\pi r^3 \int_{-\pi/2}^{\pi/2} \sin^2 \chi \, d\chi = 2\pi^2 r^3.$$



18.2. Isotropy and homogeneity of space

According to the *Einstein cosmological principle* (see [179]) our “universe” is homogeneous and isotropic for large spatial scales and fixed time. The *homogeneity* is expressed by a translation symmetry (i.e. the space has at any point the same density, temperature, pressure, etc.), while *isotropy* is expressed by rotational symmetry (i.e. there are no preferred directions at any point and an observer is not able to distinguish a given direction from another direction by means of local physical measurements).

Gravity in a slightly inhomogeneous medium tends to create long fibers. For example, there exists a filament of galaxies called the *Sloan Great Wall* which is 1.37 billion light years long. This means that the real universe is not homogeneous on very large scales, where it rather resembles a cosmic web. The cosmic microwave background radiation also slightly differs from ideal isotropy which was observed by the satellites COBE, WMAP, and recently Planck (see Fig. 18.4). Moreover, it is slightly polarized. Nevertheless, in the standard cosmological model the homogeneity and isotropy of space are assumed; otherwise the model would be too complicated and we would not be able to compute too much.

There are only three mathematical models of space that satisfy the cosmological principle: the sphere \mathbb{S}_r^3 , the Euclidean space \mathbb{E}^3 , and the pseudosphere \mathbb{H}_r^3 . The corresponding curvature index 1, 0, and -1 appears in the Friedmann equation (10.5). These manifolds have the maximal group of symmetries that is defined by means of six linearly independent Killing’s fields (see [293, Chap. 13]). It is very difficult to

imagine the hyperbolic geometry of the pseudosphere \mathbb{H}_r^3 which is only for better visualization represented by various models, for instance the already mentioned hyperbolic hypersurface (18.2). In other words, $\tilde{\mathbb{H}}_r^3$ is only a model of the manifold \mathbb{H}_r^3 which is also only a model of our universe. Note that $\tilde{\mathbb{H}}_r^3$ has two components whereas \mathbb{H}_r^3 is connected. More specifically, we will focus on this topic in Sections 18.4 and 18.5.

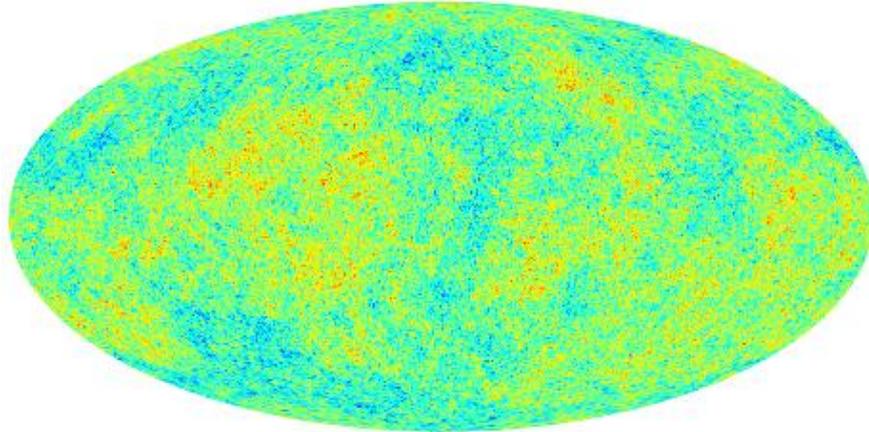


Fig. 18.4. Fluctuations in temperature ≈ 2.73 K of the cosmic microwave background (CMB) radiation that corresponds to the redshift $z = 1089$. This radiation arose when the space was 1090 times smaller than at present and its mean temperature was almost 3000 K (photo from the Planck satellite).



18.3. The ambiguity of the notion universe

The term “universe” is used in cosmology with various meanings: true spacetime, true space (i.e. spacetime for a fixed time), and the observable universe, which is seen as a projection on the celestial sphere. These are three different objects. Their mathematical models are also three completely different manifolds (see Fig. 18.5). Thus altogether we have six meanings of the problematic notion “universe” for which the terminology is not fixed yet. The first three contain real matter, whereas the other three are only abstract mathematical idealizations of reality.

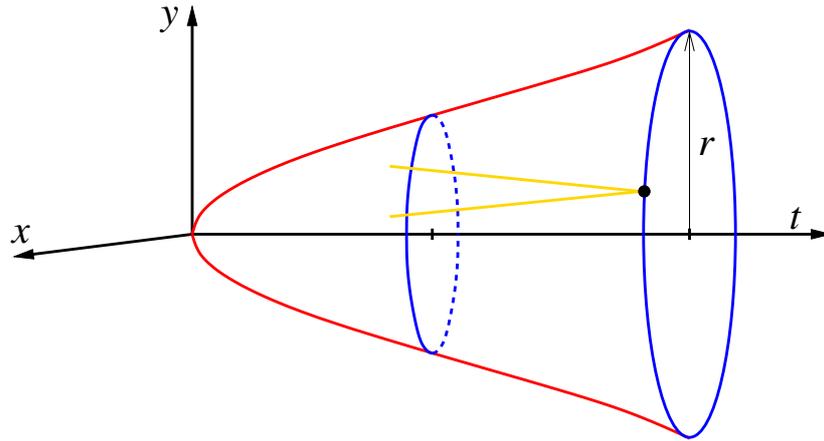


Fig. 18.5. Three different manifolds corresponding to the curvature index $k = 1$. For simplicity, the hypersphere (18.1) is replaced only by its great circle $\mathbb{S}_{r(t)}^1$ for $z = w = 0$ and for a fixed time instant t . This is the model of the space (universe). The model of spacetime can be obtained by rotation of the graph of the expansion function about the time axis t . The observable universe is marked by a light cone. The space dimensions are reduced by two.

In accordance with the Einstein cosmological principle, we shall understand by the *universe* a cross-section of spacetime for a fixed time instant, i.e., the universe will be an isochrone in spacetime for constant t . For instance, if the curvature index is positive, the corresponding model of the universe is the hypersphere \mathbb{S}_r^3 for some fixed radius $r = r(t) > 0$, which is a three-dimensional manifold in the four-dimensional Euclidean space \mathbb{E}^4 (cf. Fig. 18.2). The associated model of spacetime in \mathbb{E}^5 has dimension four and the model of the observable universe has dimension three (cf. Fig. 18.5).

All six above-mentioned objects have to be carefully distinguished; otherwise we may come to various confusions. The observable universe is not homogeneous, since for different cosmological redshifts z , it has a different mass density. Thus, it is an entirely different object than the universe as a space.

In the observable universe some cosmologists incorrectly try to measure angles α, β, γ in some large triangle to ascertain the spherical, Euclidean, or hyperbolic geometry of the universe by means of their sum $\alpha + \beta + \gamma$. Such a triangle has to be considered in the uni-

verse (space), in which we see only our close neighborhood (strictly speaking, only the one point at which we are situated). This limits our ability to perform such measurements.

Let us emphasize that the observable universe cannot be modeled by three-dimensional Euclidean space (nor for $k = 0$), since it is a cone whose interior does not belong to it. The whole situation is sketched in Fig. 18.5, where two spatial dimensions are removed. We see that the observable universe can be modeled by a light cone which is deformed towards the origin of the time coordinate. Instead of spheres $\mathbb{S}_{r(t)}^3$ we may consider only their great circles $\mathbb{S}_{r(t)}^1$ having centers on the time axis (cf. Theorems 18.2 and 18.3).

Now we show that the expansion speed in the early universe was much larger than at present. It is estimated that the relic radiation originated at the time $t_1 = 380\,000$ years after the Big Bang and that the age of the universe is $t_0 = 13.82$ Gyr (see e.g. [73], [213]). Then for the measured redshift $z = 1089$ we obtain⁶

$$\frac{t_1 a(t_0)}{t_0 a(t_1)} = \frac{t_1}{t_0} (z + 1) = \frac{380\,000 \cdot 1090}{13.82 \cdot 10^9} = \frac{1}{33.3},$$

where the first equality follows directly (see [184, p. 730], [202, p. 96]) from the definition of redshift. From this we get

$$33.3 \cdot \frac{a(t_0)}{t_0} = \frac{a(t_1)}{t_1}, \quad (18.3)$$

and thus the mean expansion rate of the space on the interval $(0, t_1)$ was 33.3 times larger than on the interval $(0, t_0)$. This indicates that the expansion function had much bigger first derivative in a close neighborhood of the origin than at present (see Figs. 13.4 and 18.5).

It would be a mistake to believe that the known map of the cosmic microwave background (CMB) radiation from Fig. 18.4 shows the entire universe, how it looked like 380 000 years after the Big Bang. This map shows only a two-dimensional slice of a three-dimensional

⁶Each current cubic meter of the space was in average squeezed in one cubic millimeter.

manifold corresponding to the universe for $z \approx 1089$. For the normalized curvature $k = 1$ the radius of the universe was 1090 times smaller than it is at present. Moreover, we observe everything only on the projection on the celestial sphere. For example, the relic radiation produced at that time in our neighborhood is not on the map of the cosmic microwave background radiation. There is also no relic radiation from the places where we have found to date all 10^{12} galaxies in the observable universe. At each of these galaxies we would observe at present completely different different maps of the cosmic microwave background fluctuations. So on the Earth, an observer may have an idea about how only a tiny part of the early universe looked like.

We also often hear that the universe has no center. This is similar to the statement that a circle has no center. The circle, of course, has its center even though it does not belong to it. Therefore, also the model⁷ of the universe \mathbb{S}_r^3 has its center at the origin of coordinates of the space \mathbb{E}^4 (see (18.1)). The center of an inflating balloon in the model of the expanding universe is then represented by the Big Bang at an initial time (see Figs. 18.5 and 20.4).

Let us point out that the sphere

$$\mathbb{S}_{r(t)}^3 = \{(x_1, x_2, y_1, y_2) \in \mathbb{E}^4 \mid x_1^2 + x_2^2 + y_1^2 + y_2^2 = r^2(t)\}$$

with increasing radius $r = r(t)$ can be identified by means of the substitutions $x = x_1 + ix_2$ and $y = y_1 + iy_2$ with an ordinary circle whose coordinates are complex variables:

$$\mathbb{K}_{r(t)}^1 = \{(x, y) \in \mathbb{C}^2 \mid |x|^2 + |y|^2 = r^2(t)\},$$

where \mathbb{C} is the set of complex numbers (the Gaussian plane). This is how the universe can be simply modeled at time t , if it has positive curvature.

⁷In 2002–2003 Grigoriy Jakovlevič Perelman proved the famous Poincaré conjecture that each simply connected compact three-dimensional manifold is homeomorphic to the sphere \mathbb{S}^3 . The largest problem was with knots on Jordan curves that exist only in three dimensions. In higher dimensions each knot can be untied.

Assume again that the universe can be modeled by the sphere \mathbb{S}_r^3 . Then the farthest point from the Earth in the space (i.e. the *horizon*) is located at the intersection of all the great circles of \mathbb{S}_r^3 that pass through the Earth. Its distance from us is $\pi r \approx 3r$, i.e., it is approximately $3\times$ further than the present radius of the space.

To close this section, let us emphasize that the Earth is at the center of the observable universe which is finite. The above-mentioned horizon of the space is completely different object than the horizon of the observable universe (see Fig. 18.4) which can be modeled by a two-dimensional sphere.



18.4. Hyperbolic space

In this section we shall study Lobachevskian hyperbolic geometry on pseudospheres. We have already mentioned that visualization of hyperbolic geometry is much more difficult than for elliptic geometry on spheres. The main reason is the fact that contrary to spheres, the maximally symmetric hyperbolic manifolds of dimension $n > 1$ cannot be isometrically⁸ imbedded in the Euclidean space \mathbb{E}^{n+1} (cf. (18.1)).

Already in 1901, David Hilbert proved (see [102]) that there does not exist a smooth (bounded or unbounded) surface in \mathbb{E}^3 without boundary with a constant negative Gaussian curvature defined as the product of the curvatures in two perpendicular principal directions (see [224]). Note that the sphere \mathbb{S}_r^2 has a constant positive Gaussian curvature $r^{-2} = r^{-1} \cdot r^{-1}$, since all osculating great circles have diameter r . A survey of two-dimensional surfaces in \mathbb{E}^3 with constant negative Gaussian curvature is given in [174] and [299]. However, all these surfaces have a singularity, such as an edge (cf. Fig. 18.7) or a cusp point, and thus they are not globally smooth, which violates the required smoothness of the model of the universe.

⁸*Isometry* is a continuous mapping $f : M \rightarrow M$, whose inverse exists and which is also continuous and preserves distances on the manifold M . In other words $\rho(f(A), f(B)) = \rho(A, B)$ for all $A, B \in M$, where ρ is a metric on M .

Hilbert in [102] in fact proved that there is no isometric imbedding of the hyperbolic plane \mathbb{H}^2 into the three-dimensional space \mathbb{E}^3 , while the sphere \mathbb{S}^2 can be isometrically imbedded into \mathbb{E}^3 (see Fig. 18.2). Similarly the famous Klein bottle (i.e. a two-dimensional nonoriented closed surface) cannot be imbedded into \mathbb{E}^3 , but it can be imbedded into \mathbb{E}^4 . Note that its three-dimensional model (usually made from glass) is not a manifold, since this surface intersects itself. Consequently, it is very difficult to develop an intuitive image concerning maximally symmetric hyperbolic manifolds.

Let us, therefore, start with a simple model of the hyperbolic plane \mathbb{H}^2 due to Poincaré. In the Euclidean plane let us consider a circle k with radius 1. The model of the hyperbolic plane (not the hyperbolic plane \mathbb{H}^2 itself) is located inside this boundary circle, which does not belong to the model (see Fig. 18.6). Geodesics (i.e. “lines” in hyperbolic geometry) are again represented similarly as in Section 18.1 by circular arcs which are moreover perpendicular to k at the endpoints. These arcs may degenerate to straight line segments like e.g. the horizontal diameter on the right part of Fig. 18.6. We easily find that there exists just one circular arc passing through two different arbitrary points A and B which is perpendicular to k at its endpoints $P \in k$ and $Q \in k$. The distance between A and B is then given by the relation (see [204, p. 36])

$$d(A, B) = \left| \ln \frac{|AQ| \cdot |BP|}{|AP| \cdot |BQ|} \right|, \quad (18.4)$$

where \ln is the natural logarithm and $|AP|$, $|AQ|$, $|BP|$, and $|BQ|$ denote the standard distances in the Euclidean plane.

It can be shown that the function d satisfies conditions 1–4 from Definition 18.1. We see that d is a nonnegative function and that $d(A, B) = 0$ if and only if $A = B$. The symmetry $d(A, B) = d(B, A)$ is obvious. To prove the triangle inequality $d(A, B) \leq d(A, C) + d(B, C)$ is technically somewhat more difficult.

The circle with radius R in the metric (18.4) has a larger circumference than $2\pi R$. Since the boundary circle k is unit in the stan-

dard Euclidean norm, the concentric circle k' with diameter $R'=1$ in the hyperbolic metric (18.4) has length 7.384... instead of the usual $2\pi = 6.283\dots$ (see Fig. 18.6).⁹ This is similar to measuring the length of a circle around a saddle point on some surface. The length of a unit circle with an arbitrary center in the hyperbolic plane is always the same in metric (18.4). A curve that has a constant Euclidean distance from the hyperbolic line is not a line (i.e. geodesic) in hyperbolic plane, see [42, p. 88].

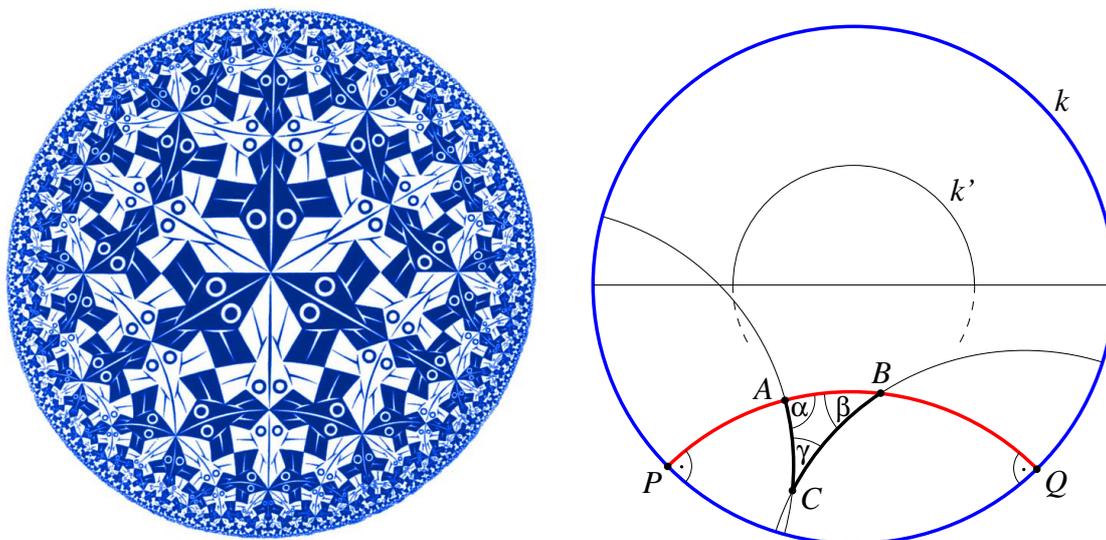


Fig. 18.6. All fish on the Escher painting of the hyperbolic plane on the left are of the same size in the metric (18.4). Figuratively speaking, if one were to swim in the hyperbolic plane, all fish in one's immediate vicinity would appear to have equal sizes. One can swim in any direction at any distance, because the boundary circle k is infinitely far away. The shortest connections represented by circular arcs are on the right picture. They are perpendicular to the boundary circle k at its endpoints. The sum of the angles in the triangle ABC is $\alpha + \beta + \gamma < 180^\circ$.

Let us point out that there are several other representations of the hyperbolic plane (see e.g. [204, p. 38]). The authors of [42] present six different representations of the manifold \mathbb{H}^2 illustrating how hyperbolic

⁹Choosing the points P, A, B, Q on the horizontal diameter of the circle k from Fig. 18.6 such that A is in the center, $P, Q \in k$, $B \in k'$, and $R' = |AB|$ is radius of the circle k' , then by (18.4) we have $1 = \ln(1 \cdot (1 + R') / (1 \cdot (1 - R')))$. From this it follows that the circle k' in the Euclidean metric has radius $R' = (e - 1) / (e + 1) = 0.462\dots$, where $e = 2.718\dots$ is the Euler number.

geometry can be modeled. Each model has a different metric. One of them is sketched in Fig. 18.3. Another famous model is the Poincaré disk (see [92, pp. 475–476]) which is similar to the situation depicted in Fig. 18.6.

Consider now an arbitrary dimension n . For $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n) \in \mathbb{E}^n$ denote the *scalar product* by

$$(a, b) = \sum_{j=1}^n a_j b_j$$

and let $\|a\| = \sqrt{(a, a)}$. Hyperbolic geometry is usually introduced on the manifold (see Fig. 18.3 for $n = 2$ and relation (18.2) for $n = 3$)

$$\tilde{\mathbb{H}}_r^n = \{(x_1, \dots, x_{n+1}) \in \mathbb{E}^{n+1} \mid \|x\|^2 - x_{n+1}^2 = -r^2\}, \quad (18.5)$$

where $r > 0$ and instead of the standard Euclidean metric in \mathbb{E}^{n+1} the *Minkowski pseudometric*

$$\mu(A, B) = (\|a - b\|^2 - (a_{n+1} - b_{n+1})^2)^{1/2} \quad (18.6)$$

is applied for $A = (a_1, \dots, a_{n+1})$, $B = (b_1, \dots, b_{n+1}) \in \tilde{\mathbb{H}}_r^n$. Although relation (18.5) contains the difference of two squares, the function μ is nonnegative, as will be shown in the next theorem. Recall again that the coordinate x_{n+1} is not time as is common in the General theory of relativity, but the usual space coordinate. The manifold $\tilde{\mathbb{H}}_r^n$ is obviously not connected. It consists of two mirror-image symmetrical hypersurfaces. For simplicity, they can be identified by the relation $x \equiv -x \in \tilde{\mathbb{H}}_r^n$, which enables us to deal with only one component for which $x_{n+1} > 0$. We will again write only $\tilde{\mathbb{H}}^n$ provided $r = 1$.

Theorem 18.1. *The function μ is a metric on $\tilde{\mathbb{H}}^n$.*

P r o o f . It is necessary to verify the conditions 1–4 from Definition 18.1. First we prove that for $A, B \in \tilde{\mathbb{H}}^n$ the value $\mu(A, B)$ is nonnegative. From relation (18.6) we get

$$\begin{aligned} (\mu(A, B))^2 &= \|a\|^2 - 2(a, b) + \|b\|^2 - a_{n+1}^2 + 2a_{n+1}b_{n+1} - b_{n+1}^2 \\ &= 2(a_{n+1}b_{n+1} - (a, b) - 1), \end{aligned} \quad (18.7)$$

where by (18.5) we employed the fact that

$$a_{n+1}^2 = \|a\|^2 + 1 \quad \text{and} \quad b_{n+1}^2 = \|b\|^2 + 1, \quad (18.8)$$

which is in fact the relation for the sum of two squares. Hence,

$$a_{n+1}^2 b_{n+1}^2 = (\|a\|^2 + 1)(\|b\|^2 + 1) = (\|a\|\|b\| + 1)^2 + (\|a\| - \|b\|)^2. \quad (18.9)$$

From this and the Cauchy–Schwarz inequality $|(a, b)| \leq \|a\|\|b\|$ for positive a_{n+1} and b_{n+1} it follows that

$$a_{n+1}b_{n+1} \geq \|a\|\|b\| + 1 \geq |(a, b)| + 1 \geq (a, b) + 1. \quad (18.10)$$

We observe that the right-hand side of (18.7) is nonnegative, and thus the square root in (18.6) is nonnegative, too.

Furthermore, we show that $\mu(A, B) = 0$ if and only if $A = B$. If $\mu(A, B) = 0$, then from (18.7) we get

$$(a, b) + 1 = a_{n+1}b_{n+1}.$$

Hence, by (18.10) we have

$$a_{n+1}b_{n+1} = \|a\|\|b\| + 1.$$

From this and (18.9) it follows that $\|a\| = \|b\|$, and by (18.8) we have $a_{n+1} = b_{n+1} > 0$. According to the definition of metric (18.6), it is obvious that $a = b$ which yields $A = B$.

From definition (18.6) we immediately also obtain the converse implication, i.e. $\mu(A, A) = 0$ is valid.

The symmetry relation $\mu(A, B) = \mu(B, A)$ from Definition 18.1 is trivial and the proof of the triangle inequality is carried through standard (even though a rather technically more difficult) way by relations (18.7)–(18.10). \square

Although equality (18.6) contains the difference of squares, we proved that the function μ is nonnegative on the manifold $\tilde{\mathbb{H}}^n$. The function μ can be naturally extended by (18.6) to the cone

$$x_1^2 + \cdots + x_n^2 \leq x_{n+1}^2$$

(cf. Fig. 18.3), where it is not metric. The value $\mu(A, B)$ can be zero for $A \neq B$ and the usual triangle inequality does not hold. For some triples even the converse triangle inequality holds¹⁰

$$\mu(A, B) \geq \mu(A, C) + \mu(B, C)$$

(see [204, p. 420]). Hence, we should use the term *pseudometric* instead of metric.

In the old literature the hyperboloid of two sheets $x^2 + y^2 - w^2 = -1$ from Fig. 18.3 is sometimes referred to as a pseudosphere with “imaginary radius” i . Note that in the General theory of relativity the cone $x^2 + y^2 = t^2$ and also the hypersurface $x^2 + y^2 + z^2 = t^2$ for time $t = w$ are both called the *light cone* for the unit speed of light. This is the set of points A for which $\mu(A, 0) = 0$.

Each of the two components $\tilde{\mathbb{H}}^n$ represents the graph of a convex (resp. concave) function in \mathbb{E}^{n+1} . To get hyperbolic geometry on the model $\tilde{\mathbb{H}}^n$ of the pseudosphere, the shortest lines (geodesics) have to be considered in the Minkowski pseudometric (18.6) and not in the Euclidean metric. Then the sum of the angles in a triangle, whose sides are geodesics, is less than 180° . The proof is given in [42, p. 88].



18.5. Maximally symmetric manifolds

In this section we give an overview of some mathematical results concerning the maximally symmetric manifolds \mathbb{S}^n , \mathbb{E}^n , and \mathbb{H}^n , i.e., those manifolds that have the maximum number of symmetries. According to [293, Chap. 13], there are no other maximally symmetric manifolds up to scaling. Obviously, the spherical surface \mathbb{S}^2 cannot be isometrically imbedded into the plane \mathbb{E}^2 . However, the sphere \mathbb{S}^n can be isometrically imbedded into \mathbb{E}^{n+1} . Now let us examine in more detail how to isometrically imbed spheres and pseudospheres.

¹⁰The famous twin paradox is closely connected with the sharp converse triangle inequality [204].

Theorem 18.2. *For $r \leq R$ the sphere \mathbb{S}_r^n can be isometrically imbedded into \mathbb{S}_R^{n+1} .*

The proof is constructive, but we will not present it, even though it is easy. For instance, on the right part of Fig. 18.2 there are two circles corresponding to parallels at $\pm 60^\circ$ isometrically imbedded into the two-dimensional sphere \mathbb{S}^2 . Thus, each sphere can be isometrically imbedded into a larger sphere if the dimension is increased by one. In other words, the sections of the sphere \mathbb{S}_r^n by the hyperplanes $x_{n+1} = C$, where C is constant and $|C| < r$, are spheres of dimension $n - 1$.

Analogously, the sections of the manifolds $\tilde{\mathbb{H}}_r^n$ by the hyperplanes $x_{n+1} = C$, where C is constant and $|C| > r$, are spheres of dimension $n - 1$. Theorem 18.2 can also be modified for pseudospheres:

Theorem 18.3. *For $r \leq R$ the pseudosphere \mathbb{H}_r^n can be isometrically imbedded into \mathbb{H}_R^{n+1} .*

Now we will present relations between Euclidean space¹¹ and the pseudosphere (see [32, p. 3]).

Theorem 18.4. *The space \mathbb{E}^n can be isometrically imbedded into \mathbb{H}^{n+1} .*

Converse imbeddings are not so simple. In 1955 Danilo Blanuša proved the hyperbolic plane \mathbb{H}^2 can be isometrically imbedded into the space \mathbb{E}^6 (see [26]). Therefore, if we would transform ourselves into six-dimensional beings in \mathbb{E}^6 (with the Euclidean distance), we could admire the beauty of the symmetry of the pseudosphere \mathbb{H}^2 , similarly to the way in which as three-dimensional beings we enjoy the beautiful symmetry of the sphere \mathbb{S}^2 . For the time being it is not known whether the dimension six can be reduced. Blanuša's assertion was generalized by David Brander [32] as follows:

Theorem 18.5. *For $n > 1$ the pseudosphere \mathbb{H}^n can be isometrically imbedded into \mathbb{E}^{6n-6} .*

¹¹For $n = 1$ the space \mathbb{E}^1 can be isometrically imbedded into \mathbb{H}^1 and vice versa. Nontrivial hyperbolic geometries thus exist only in dimensions $n > 1$.

It is again not known whether the dimension $6n - 6$ can be reduced. The manifold \mathbb{S}^3 , resp. \mathbb{H}^3 , which possibly model our universe at a fixed time, can by previous statements be isometrically imbedded to the Euclidean space \mathbb{E}^4 , resp. \mathbb{E}^{12} , i.e.

$$\boxed{\mathbb{S}^3 \hookrightarrow \mathbb{E}^4, \quad \mathbb{H}^3 \hookrightarrow \mathbb{E}^{12}.} \quad (18.11)$$

Here the symbol \hookrightarrow denotes the isometric imbedding. In 12-dimensional Euclidean space the distances in \mathbb{H}^3 will be nondeformed. The pseudosphere \mathbb{H}^3 is a rather exotic object and as 12-dimensional beings we could convince ourselves about its maximal symmetry. Visualization of the pseudosphere \mathbb{H}^3 without deformation of distances is therefore extremely difficult. On the other hand, it is easier to imagine the manifolds \mathbb{E}^3 and \mathbb{S}^3 which are additional candidates for the model of the universe and which do not contain any privileged points or directions.

In [143] we prove the theorem that any n -dimensional manifold in \mathbb{E}^{n+1} which has the same curvature at any point and any principal direction is necessarily the sphere \mathbb{S}_r^n or the hyperplane in \mathbb{E}^{n+1} . Since the pseudosphere \mathbb{H}^n cannot be imbedded into \mathbb{E}^{n+1} for $n > 1$, the assumptions of this theorem are not satisfied for the manifold \mathbb{H}^n .

The surface in Fig. 18.7 is often presented as a good model of the universe with hyperbolic geometry. Below we present several drawbacks of this model. This manifold, which looks like a trumpet, has negative constant Gaussian curvature (-1) everywhere and each point is a saddle point. This surface arises by rotation of the *tractrix*¹² about its asymptote incident with the axis x_3 . For $x_1 \in (0, 1]$ the tractrix is defined by the equation

$$x_3 = \ln \frac{1 - \sqrt{1 - x_1^2}}{x_1} + \sqrt{1 - x_1^2}.$$

There also exist other parametric expressions [143, p. 530].

The upper circular edge on Fig. 18.7 does not belong to this manifold, which cannot be smoothly extended beyond this edge so that

¹²The tractrix is the involute of the catenary function.

the Gaussian curvature would remain (-1) . Sometimes this surface is also called a pseudosphere like the hypersurface (18.5) that has an infinite volume. Contrary to the pseudosphere \mathbb{H}^2 , rotation of the tractrix generates a surface which by [174, p. 33] has a finite surface area and volume, as was already known to Christian Huygens in 1639 (see [166, p. 324]).

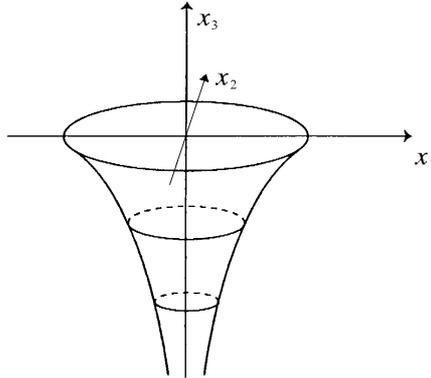


Fig. 18.7. The surface generated by the rotation of the tractrix has constant negative Gaussian curvature (-1) at all its points. Nevertheless, the curvatures in the principal directions are different, and therefore, an outside observer can find that this manifold is not isotropic. The inner observer will also detect an anisotropy near the top circular edge.

We cannot travel on the surface of Fig. 18.7 unrestrictedly on the surface in the vertical direction and therefore, this surface does not model an isotropic universe. We can also choose on this surface an extremely small circle which passes through its center¹³ for an appropriate $x_3 \ll 0$. Hence, two-dimensional beings living on this manifold would recognize that their space is not isotropic.

As candidates of a curved space describing our universe the hyperboloid of one sheet, torus, etc., are also proposed. However, these manifolds are not equally curved at all points and in all directions.

Let us add that the one-sheet hyperboloid $x^2 + y^2 - w^2 = 1$ has a negative Gaussian curvature everywhere while the hyperboloid from Fig. 18.3 has a positive curvature everywhere. Nevertheless, the

¹³To see this consider, for simplicity, an infinite cylinder of radius ρ . Using a rope of length $2\pi\rho$ an observer may sketch a circle on the cylinder surface which passes through its center.

Gaussian curvature does not distinguish among directions on a manifold, while the sectional curvature does.¹⁴ The maximally symmetric manifolds \mathbb{S}_r^n , \mathbb{E}^n , and \mathbb{H}_r^n have for $r > 0$ the constant sectional curvature $1/r^2, 0, -1/r^2$ which will be called the *spatial curvature*.

Let us present one more argument that favours \mathbb{S}_r^3 as a model of our universe. The manifolds \mathbb{E}^3 and \mathbb{H}_r^3 have infinite volume. However, the actual space could not first be finite after its origin and then change to infinite. Moreover, we can hardly imagine that the actual infinite universe would have everywhere on large scales the same density,¹⁵ temperature, pressure¹⁶ and so on, at a given time instant after the Big Bang as required by the Einstein cosmological principle. In this case, information would have to be transmitted at infinite speed. The popular theory of inflation cannot explain such a homogeneity and isotropy of an infinite universe.

It is not natural to equip the manifold $\tilde{\mathbb{H}}_r^n$ with the Minkowski pseudometric (18.6) for $n = 3$, since x_{n+1} is not time. There are also arguments against \mathbb{E}^3 being the correct model of our universe (see also Section 19.5). By Einstein's General theory of relativity, matter curves space. Nevertheless, the curvature of \mathbb{E}^3 is independent of the decreasing mean mass density $\rho = \rho(t) > 0$ with time.

Therefore, the most probable model of our universe seems to be the sphere \mathbb{S}_r^3 . Moreover, for the model with normalized curvature $k > 0$, the well-known horizon problem¹⁷ is not as severe as for $k \leq 0$, since all points were close together at early times.



¹⁴The sectional curvature of a manifold in a given direction is a function of two linearly independent vectors v and w and it expresses the Gaussian curvature of a two-dimensional submanifold with tangent vectors v and w .

¹⁵If the density function $\rho = \rho(x_1, x_2, x_3) > 0$ were to be constant on the entire Euclidean space \mathbb{E}^3 , then the Newtonian gravitational potential $V = V(x_1, x_2, x_3)$ would also be constant. Nevertheless, this contradicts the Poisson equation $\Delta V = 4\pi G\rho > 0$, where the left-hand side $\Delta V = \sum_i \partial^2 V / \partial x_i^2 = 0$. Thus a stationary solution is not allowed.

¹⁶Furthermore, these quantities should attain arbitrarily large values at infinitely many points just after the Big Bang.

¹⁷The horizon problem states that different regions of the universe have not been in contact with each other, but nevertheless they have the same physical properties.

19. A critique of the standard cosmological model

*To criticize is easy,
more difficult is to create something.*

Well-known motto

19.1. The standard mathematical cosmological model

The main purpose of this chapter is to show that the proposed amounts of mysterious dark energy and dark matter can be considerably influenced by the modeling error of the current mathematical Λ CDM (Lambda–Cold Dark Matter model) model of the expanding universe. This model is based on the Friedmann equation (10.5) which was derived by Alexander Friedmann [80] in 1922.



Fig. 19.1. Alexander Friedmann (1888–1925)

Recall that Friedmann applied Einstein's equations to a maximally symmetric universe which is homogeneous and isotropic for each fixed

time instant (see Chapters 10 and 18). In this way he obtained a non-linear differential equation of the first order for an unknown sufficiently smooth expansion function $a = a(t) > 0$, which describes the expansion of the whole universe:

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}, \quad (19.1)$$

where $\rho = \rho(t) > 0$ denotes the mean mass density of the universe at time t , G is the gravitational constant, Λ is the cosmological constant, c is the speed of light in the vacuum, k/a^2 is the spatial curvature, and k is the normalized curvature (curvature index). The value $k = 1$ corresponds to the hypersphere \mathbb{S}_r^3 with variable radius (see (18.1))

$$r = r(t) = a(t).$$

The case $k = 0$, which was not considered by Friedmann in [80], corresponds to \mathbb{E}^3 . Lemaître in [159] also derived the equation (19.1) without citing Friedmann's paper [80].

By means of equation (19.1) Friedmann described the dynamical behavior of the universe as an alternative to Einstein's stationary universe [72]. In 1924, Friedmann published another paper [81], where the negative normalized curvature $k = -1$ is considered. However, he derived equation (19.1) only for a negative mass density (see [81, p. 2006]) and it is not clear how to satisfy such a paradoxical assumption.¹ Fortunately, equation (19.1) may also be investigated for $k = -1$ and $\rho \geq 0$. If $k = -1$ then the space at a fixed time can be modeled by the maximally symmetric hyperbolic manifold \mathbb{H}_r^3 (see Chapter 18). In the standard cosmological model the curvature index may attain only three values

$$k \in \{-1, 0, 1\}.$$

To obtain (19.1), Friedmann employed only the first of ten Einstein's equations for the 00-component of the tensor of density of energy and momentum (see (19.8) below). By means of the trace of

¹For negative mass the sum of the angles in the triangle of Fig. 2.10 a) would be less than 180° . Such a triangle would have its curved sides bent inward.

components 11, 22, and 33 one may derive another differential equation of the second order for the expansion function (see e.g. [159])

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)a + \frac{\Lambda c^2}{3}a, \quad (19.2)$$

where p is the pressure.² Notice that this equation does not depend explicitly on the normalized curvature k . It is linear even though Einstein's equations are nonlinear.

From Section 10.4 we know that in 1917 Albert Einstein included a positive cosmological constant Λ to his equations of the General theory of relativity to avoid gravitational collapse and to save his model of the stationary universe [72]. However, the resulting stationary solution of equation (19.1) is not stable, i.e., any small deviation from constant $a = a(t)$ will cause either a gravitational collapse, or expansion (see [184, p. 746]).

Although the General theory of relativity was formulated to explain various shortcomings of the Newtonian theory of gravity for large velocities, masses, densities, etc., the Friedmann equation (19.1) for $\Lambda = 0$ can easily be formally derived from the Newtonian theory (cf. [180]).

Recall that for small redshifts $z \ll 1$ the recession speed of galaxies v from our Galaxy is approximately proportional to their distance d , that is

$$v \approx H_0 d \quad (\text{Hubble law}), \quad (19.3)$$

where H_0 is the Hubble constant whose physical dimension is s^{-1} . The Planck Collaboration report [213, p. 30] presents several of its possible current values, for instance,

$$H_0 = 67.3 \pm 1.2 \text{ km s}^{-1} \text{ Mpc}^{-1} \quad \text{and} \quad H_0 = 73.8 \pm 2.4 \text{ km s}^{-1} \text{ Mpc}^{-1} \quad (19.4)$$

that are probably influenced by large systematic errors.

²In the standard cosmological model, the equation of state is usually expressed in the form $p = w\rho c^2$, where the parameter w is constant. The case $w = 0$ corresponds to a matter-dominated universe (coherent dust) and $w = 1/3$ to a radiation-dominated universe (ultrarelativistic matter – the so-called photon gas).

Let us recall the definition of the Hubble parameter (cf. (10.3))

$$H(t) := \frac{\dot{a}(t)}{a(t)} \quad (19.5)$$

such that $H(t_0) = H_0$, where t_0 is the age of the universe. Notice that the Hubble law (19.3) was derived for the observable universe (see Fig. 18.5), whereas (19.5) describes the expansion of the universe (space). The expansion speed of the universe was larger in the past (cf. Fig. 8.7). By (19.5) the expansion function $a = a(t) > 0$ is increasing at present as $H_0 > 0$. It is not easy to establish the current value of the Hubble parameter $H(t)$, since we always observe only the past. In our close neighborhood, the measurement of $H_0 = H(t_0)$ is impaired by local movements of galaxies. On the other hand, it is very difficult to reliably extrapolate the current value of H_0 from long-distance objects (e.g. from the cosmic microwave background radiation which traveled to us for over 13 Gyr, see [212, p. 16]).



19.2. Strange behavior of cosmological parameters

In literature in cosmology, division of equation (19.1) by the square $H^2 = (\dot{a}/a)^2 \geq 0$ is usually done without any preliminary warning that we may possibly divide by zero which may lead to various paradoxes. Recall the *normalized Friedmann equation* (10.6) for three dimensionless parameters

$$1 = \Omega_M(t) + \Omega_\Lambda(t) + \Omega_K(t) \quad (19.6)$$

for all t , where

$$\Omega_M(t) := \frac{8\pi G\rho(t)}{3H^2(t)} > 0, \quad \Omega_\Lambda(t) := \frac{\Lambda c^2}{3H^2(t)}, \quad \Omega_K(t) := -\frac{kc^2}{\dot{a}^2(t)}, \quad (19.7)$$

and where Ω_M is the (mean) *density of dark and baryonic matter*, Ω_Λ is the *density of dark energy*, and Ω_K is the *curvature parameter*, see [202] and [213]. The function $\rho_c(t) = 3H^2(t)/(8\pi G)$ is called the *critical density* for historical reasons, since if $\Lambda = 0$, then

$$\begin{aligned}
k = 0 &\iff \rho = \rho_c, \\
k = 1 &\iff \rho > \rho_c, \\
k = -1 &\iff \rho < \rho_c.
\end{aligned}$$

By (10.2) we find that the current critical density $\rho_c(t_0) \approx 10^{-26} \text{ kg/m}^3$ which corresponds to 6 protons per cubic meter.

1. Let us first study the behavior of cosmological parameters in the case of Einstein's stationary universe \mathbb{S}_r^3 (see (18.1)), where $r = a$ is constant, i.e., $\dot{a}(t) = 0$ for all t (cf. the top left part of Fig. 19.2). Then from (19.5) we have $H(t) = 0$. Even though nothing dramatic happens, by (19.7) the density of baryonic and dark matter $\Omega_M(t) = \infty$ for all t . We should write more precisely that this parameter is not well defined. Reasonably defined physical quantities should not attain infinite values. Moreover, the true density of baryonic matter for any time instant is surely finite!

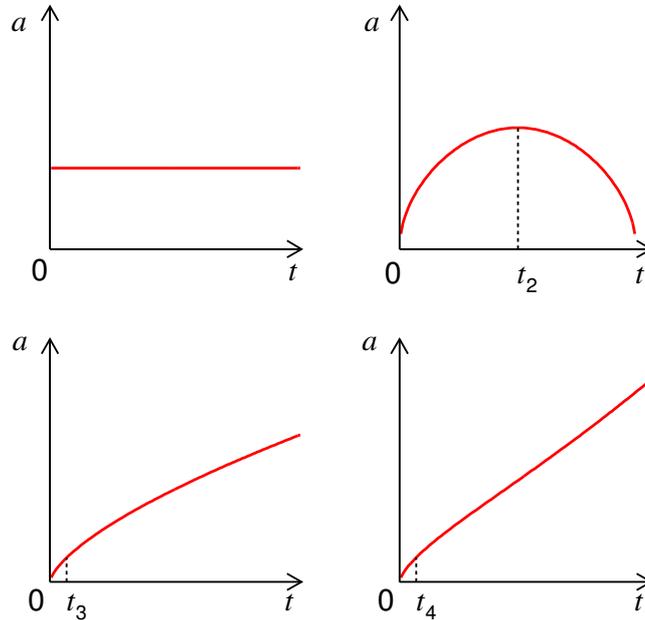


Fig. 19.2. The expansion function for the stationary universe, the cyclic universe, the universe with zero cosmological constant, and for the currently accepted expansion of the universe with a positive cosmological constant.

2. Consider now another classical model, the so-called cyclic or pulsating or oscillating universe. Assume for a moment that its expansion stops at some time $t_2 > 0$ and then starts to shrink (see the top right part of Fig. 19.2). Then $\dot{a}(t_2) = 0$ and by (19.7) for $\Lambda > 0$ the density of dark energy, which should accelerate the expansion of the universe, is equal to $\Omega_\Lambda(t_2) = \infty$. However, the universe starts to collapse. Even in a close neighborhood of the point t_2 , where we do not divide by zero, the behavior of cosmological parameters is bizarre, since their values rapidly grow beyond all bounds.³ Moreover, the ratio between the density of dark and baryonic matter is again infinite. Is this not strange?

3. In the model with zero cosmological constant and $k = -1$ it is derived that the expansion function tends to infinity for $t \rightarrow \infty$ and is strictly concave for $t > t_3 > 0$ (see the bottom left part of Fig. 19.2 and [184, p. 735]). Hence, the derivative \dot{a} and also its square are decreasing functions. By (19.7) the curvature parameter $\Omega_K > 0$ increases for $t \rightarrow \infty$, whereas the spatial curvature k/a^2 tends to zero. From points 1–3 we observe that all three cosmological density parameters (19.7) do not have appropriate names.

4. A somewhat more curious behavior of the parameter Ω_K is obtained for the currently accepted expansion function. Similarly as in the previous point 3 we shall consider only $t > t_4 > 0$, where t_4 denotes the time instant of the origin of the cosmic microwave background radiation.⁴ According to the measurements by the 2011 Nobel Prize Winners [229] the expansion function $a(t)$ is strictly concave over the interval circa $(t_4, 9)$ Gyr and then changes to a strictly convex function on the interval $(9, 14)$ Gyr. In other words, the function \dot{a} is first decreasing and then increasing (see the bottom right part of Fig. 19.2). From this it follows by (19.7) that the curvature parameter $\Omega_K(t)$ is not a monotonic function, even though the universe expands continually.

³The first derivative of the expansion function of de Sitter universe [92, p. 43] also attains zero value at one point.

⁴The cosmic microwave background radiation did not arise suddenly, but in a time interval several thousand years long.

The absolute value of the curvature parameter $|\Omega_K| > 0$ on the interval $(t_4, 9)$ Gyr increases for $k \neq 0$, but the spatial curvature tends to zero with increasing time. We again see that the name for Ω_K was not appropriately selected.

Let us further note that by the theory of inflation, the universe expanded exponentially during a very short time instant after the Big Bang, i.e., the expansion function $a = a(t)$ was strictly convex. Then it was strictly concave and then surprisingly it was again strictly convex by point 4.

The classical model without inflation assumes that $\dot{a}(0) = \infty$ (see Fig. 19.2 and [184, p. 735]). However, by (19.7) the curvature parameter $\Omega_K(0)$ was zero at the origin of our universe, i.e., almost the same as at the present time. For $k \neq 0$ its curvature would on the contrary be very large at time $t \approx 0$ and with increasing time it would converge to zero. For $k = 1$ the curvature also tends to infinity for $t \rightarrow 0$ (see the famous model of an inflating balloon from Fig. 16.4). In the next section, we present many severe drawbacks of the standard cosmological model.



19.3. Excessive extrapolations

The issue of whether the Friedmann equation (19.1) sufficiently exactly describes the expansion of the real universe is entirely essential. If it is not so, then cosmologists solve by various means the same questionable equation, which should not be identified with reality. We will support our examination of the unrestricted use of the Friedmann equation by several arguments.

Phenomena in our universe are usually modeled by equations of mathematical physics, such as linear elasticity equations, Maxwell's equations, semiconductor equations, Einstein's equations, and so on. However, no such equation describes reality absolutely exactly. Thus we always get a nonzero modeling error with respect to some criterion (maximum surface temperature, mean velocity, minimum pres-

sure, and so on). Every equation of mathematical physics has certain restrictions on the size of the investigated objects⁵ where reality is modeled well and, on the other hand, where its description fails, i.e., the modeling errors essentially depends on the size of these objects. We will demonstrate this by a few examples.

Consider a unit cube with edge $e = 1$ m (see Fig. 19.3) and solve a steady-state heat conduction problem⁶

$$-\Delta u = f$$

with some boundary conditions, where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplace operator, u is the temperature, and f is proportional to the density of heat forces. This elliptic problem approximates very well the true temperature in homogeneous isotropic solids which can be verified by direct measurements (see [142]). However, in applying the heat equation on the atomic level in the cube with edge $e = 10^{-10}$ m, we get nonsensical numbers, since it is not clear how to define the temperature on such a small scale. We also obtain nonsensical numbers for $e = 10^{10}$ m (see the left part of Fig. 19.4). Such a large cube would immediately collapse into a black hole, since the diameter of this cube is about ten times larger than the diameter of the Sun. We can, of course, solve the above steady-state heat conduction problem on an arbitrarily large cube. However, the question is for which e do we still get acceptable results, and when the resulting temperatures have nothing to do with reality. The criterion used can be e.g.

$$E = E(e) = \frac{|U_{\max} - u_{\max}|}{U_{\max}},$$

where U_{\max} (resp. u_{\max}) is the maximum real (resp. theoretical) surface temperature on the cube of Fig. 19.3.

⁵This is the main reason, why the General theory of relativity and quantum mechanics cannot be unified and why there will be no appropriate “theory of everything” valid on all spacetime scales.

⁶The distribution of the Newton gravitational potential for $f = -4\pi G\rho$ is also given by this equation.

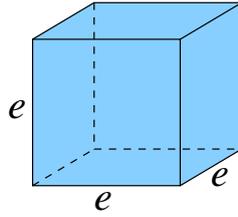


Fig. 19.3. A cube with edge of length e

Other equations of mathematical physics (such as supraconductivity equations, Navier–Stokes equations for fluids, Korteweg-de Vries equations, magneto-hydro-dynamic equations, and so on) with respect to given criteria are subjected to analogous restrictions, even though we get different graphs of the left part of Fig. 19.4. Analogously, we cannot apply Keplerian laws on scales of 10^{-10} m or the Schrödinger equation on objects that have the size of a cat.

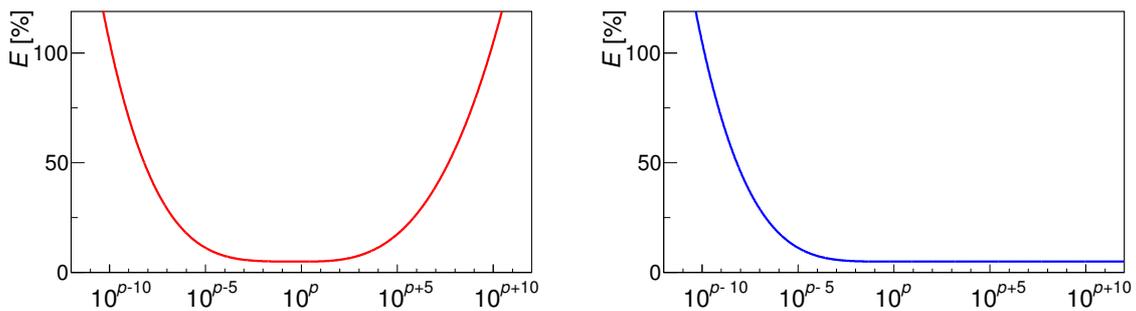


Fig. 19.4. Left: Schematic illustration of the general behavior of the relative extrapolation error E for equations of mathematical physics. The horizontal axis (in appropriate units) has a logarithmic scale and p is the exponent yielding the smallest relative extrapolation error. Right: Enforced behavior of relative extrapolation error in the case of Einstein’s equations.

Also time scales cannot be arbitrary. To see this, we present two more examples. The classical N -body problem yields very good predictions of the positions of the planets in the Solar system after one year. However, it produces nonsensical numbers for the period of 10^{10} years. Also the backward integration to the past has no sense, since the Solar system did not exist 10^{10} years ago. Therefore, long-term extrapolations in time are not reliable as well, even if we knew the exact solution of the N -body problem.

Current mathematical models of weather forecast yield quite acceptable results for one day in advance. But we cannot predict the weather forecast for 10^{10} days in advance, and so on.

Therefore, in any calculation we have to take account of the modeling error which is small only under some space-time limits. For all that, Alexander Friedmann when deriving (19.1) applied Einstein's gravitational field⁷ equations on cosmological scales even though they are "verified" on much smaller scales than the size of the whole universe. This is considered as a platitude and almost nobody deals with the question, whether it is justified to perform such a fearless extrapolation without any observational support. Einstein's equations are not scale invariant. For instance, they do not describe well processes on the atomic level. They are nonlinear, contain the fundamental physical constants G and c , and it is assumed that the speed of gravity is just c . This suggests that the equation (19.1) was obtained by incorrect extrapolations from Solar system scales to cosmological scales (see the right part of Fig. 19.4), even though it was derived by correct mathematical operations involving Einstein's equations on the maximally symmetric manifolds \mathbb{S}_r^3 , \mathbb{E}^3 , and \mathbb{H}_r^3 . Moreover, Einstein's equations are "verified" in the observable universe which is represented by a different manifold. The current cosmological model is thus based on a questionable Friedmann equation. Below we present further arguments to support this conjecture.

Before explaining why we use the term "verified" in quotation marks, we first recall the system of Einstein's 10 equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \quad (19.8)$$

where $\mu, \nu \in \{0, 1, 2, 3\}$, which contains on its left-hand side several thousands of terms (partial derivatives of scalar functions). The reason

⁷There are still many open problems with gravity. For the time being we do not know whether gravitational waves exist, what is the range of gravity, what are its aberration effects, what is its speed of propagation (e.g. is it the same inside and outside the Sun?), is gravity quantized, and so on.

is that there are 10 components of the unknown symmetric metric tensor $g_{\mu\nu}$,

$$R_{\mu\nu} = \sum_{\varkappa=0}^3 R_{\mu\varkappa\nu}^{\varkappa}$$

is the symmetric *Ricci tensor* of one time variable x_0 and three space variables x_1, x_2, x_3 , i.e., $R_{\mu\nu} = R_{\mu\nu}(x_0, x_1, x_2, x_3)$ (for simplicity the dependence of functions on these non-delay variables is nowhere indicated)

$$R = \sum_{\mu,\nu=0}^3 g^{\mu\nu} R_{\mu\nu}$$

is the *Ricci scalar*, $T_{\mu\nu}$ is the symmetric *tensor of density of energy and momentum*,

$$R_{\mu\rho\nu}^{\varkappa} = \frac{\partial \Gamma_{\mu\nu}^{\varkappa}}{\partial x^{\rho}} - \frac{\partial \Gamma_{\mu\rho}^{\varkappa}}{\partial x^{\nu}} + \sum_{\lambda=0}^3 \Gamma_{\mu\nu}^{\lambda} \Gamma_{\lambda\rho}^{\varkappa} - \sum_{\lambda=0}^3 \Gamma_{\mu\rho}^{\lambda} \Gamma_{\lambda\nu}^{\varkappa}$$

is the *Riemann curvature tensor* that has 20 independent components due to symmetries⁸

$$\Gamma_{\varkappa\rho}^{\mu} = \frac{1}{2} \sum_{\nu=0}^3 g^{\mu\nu} \left(\frac{\partial g_{\varkappa\nu}}{\partial x^{\rho}} + \frac{\partial g_{\rho\nu}}{\partial x^{\varkappa}} - \frac{\partial g_{\varkappa\rho}}{\partial x^{\nu}} \right)$$

are *Christoffel's symbols of the second kind*. From this and the relation $g_{\varkappa\rho} = g_{\rho\varkappa}$ we observe for $\mu = 0, 1, 2, 3$ the symmetry $\Gamma_{\varkappa\rho}^{\mu} = \Gamma_{\rho\varkappa}^{\mu}$. Thus altogether we have $40 = 4 \times 10$ independent components. Finally, $g^{\mu\nu}$ is the 4×4 matrix inverse to $g_{\mu\nu}$, i.e.

$$g^{\mu\nu} = \frac{g_{\mu\nu}^*}{\det(g_{\mu\nu})},$$

where the entries $g_{\mu\nu}^*$ form the matrix of algebraic adjoints (see [224]). Thus, the explicit expressing of the left-hand side of (19.8) would occupy several pages.

⁸In any space dimension N the Riemann tensor has $N^2(N^2 - 1)/12$ independent components due to the following conditions: $R_{\mu\rho\nu}^{\varkappa} + R_{\rho\nu\mu}^{\varkappa} + R_{\nu\mu\rho}^{\varkappa} = 0$ (the *First Bianchi identity*) and $R_{\lambda\mu\rho\nu} = -R_{\mu\lambda\rho\nu} = -R_{\lambda\mu\nu\rho}$, where $R_{\lambda\mu\rho\nu} = \sum_{\varkappa} g_{\lambda\varkappa} R_{\mu\rho\nu}^{\varkappa}$. From this it follows that $R_{\lambda\mu\rho\nu} = R_{\rho\nu\lambda\mu}$. For $N = 2$ the Riemann tensor can only be expressed by the scalar Gaussian curvature.

Since the system (19.8) of Einstein's equations is too complicated, its solution for two mutually orbiting bodies is not known. Therefore, many simplifications are made (see e.g. [184, p.1076]). The most popular is the parametrized post-Newtonian (PPN) formalism which is also an approximation of several other theories of gravity including General relativity. It assumes velocities v for which $v \ll c$. After many simplifications we get only a single algebraic expression and this is tested. But we should evaluate the modeling error between an actual solution of system (19.8) and measured quantities. Most of the tests are based only on Solar-system experiments (e.g. slowdown of electromagnetic waves [255] and bending of light in the gravitational field of the Sun [184], measuring the curvature of spacetime near the rotating Earth by means of the Lense-Thirring precession effect [161], and the perihelion advance of Mercury's orbit [234]). For instance, instead of solving system (19.8), the true perihelion advance of Mercury's orbit after one revolution⁹ is compared with the simple algebraic relation (see [70, s. 839])

$$\varepsilon = 24\pi^3 \frac{a^2}{T^2 c^2 (1 - e^2)} = 5.012 \cdot 10^{-7},$$

where $T = 7.6 \cdot 10^6$ s is the orbital period, $a = 57.909 \cdot 10^9$ m is the semimajor axis, and $e = 0.2056$ is the eccentricity of Mercury's orbit.

⁹According to Anatoli Vankov [284], calculating and measuring the perihelion advance of Mercury's orbit is an ill-conditioned problem. Einstein made a lot of simplifications. He assumes the metric tensor has a special form and takes $\Lambda = 0$. He does not consider the oblateness of the Sun, its mass losses, its rotation and the corresponding Lense-Thirring effect, gravitational aberration, tidal forces, and the influence of magnetic fields. Einstein also neglects many higher order terms in (19.8). Mercury is replaced by a mass point which does not curve the spacetime. The influence of other planets is not considered, too. Some elliptic integral, which has no analytical expression, is approximated numerically without any error estimate. Note that other planets shift the perihelion of Mercury's orbit about 530'' per century. Since their orbital periods cannot be expressed as fractions of small integer numbers, the perihelion is shifted very irregularly (see [190], [220]). A variable position of the Sun with respect to the center of gravity of the Solar system (see Fig. 13.1) has to be considered, too. Note that the eccentricity e is well defined only for a purely elliptic orbit. Moreover, we have to take into account the serious instability of the N -body problem (see Section 5.5) which cannot be solved analytically, the corresponding discretization errors, rounding errors, errors involved in physical constants and also in coordinates which are perturbed by nutations of the Earth's axis, refraction of the Earth's atmosphere, and so on. See also Remark 5.1.

From this we get $43'' = \varepsilon \cdot \tau \cdot 3600'' \cdot 180 / (\pi T)$ during $\tau = 3\,155\,814\,954$ s, i.e. just one century. Another simple formula tested is e.g. (8.15).

Now we will discuss verification of Einstein's equations in a distant stellar system. Concerning the reduction of orbital periods and the periastron shift of very distant binary pulsars, we should keep in mind that the minimal distance of their components is usually about 0.01 au (see e.g. [107], [275]), i.e., it is again on the scale of the Solar system. The distance 0.01 au is so short that strong magnetic fields of pulsars up to 10^{11} T and extremely fast rotation up to 1000 Hz have a nonnegligible influence on the periastron shift and on the reduction of orbital periods. Tidal forces are probably negligible.

The detected gravitational redshift $z = 0.4$ of neutron stars appears only in a very close vicinity of the star and is negligible several astronomical units away. On the other hand, distant galaxies represent very large gravitational lenses. However, they are so inhomogeneous that the relation (8.15) can give only a very rough approximation of the action of these lenses. Moreover, the distribution of mass along the trajectories of observed photons is not known. Therefore, the often-proclaimed statements that Einstein's equations describe reality with precision better than 99 % is questionable due to the above arguments.

Note that galaxies have a diameter on the order of 10^{10} astronomical units.¹⁰ The size of our universe is at least five orders of magnitude larger than a single galaxy. Hence, the Friedmann equation (19.1) was derived under a considerably unjustified extrapolation ignoring the modeling error [147]. So it probably does not describe reality well. This seems to be the main misconception of current cosmology.



19.4. Expansion function

In applying the standard cosmological model various “delicate” limits are performed: $a \rightarrow 0$, $a \rightarrow \infty$, $t \rightarrow 0$, $t \rightarrow \infty$, ... (see, for

¹⁰To see this note that the diameter of the Milky Way is about 150 000 ly (its visible part is 100 000 ly) and $1 \text{ ly} \approx 63\,242 \text{ au}$.

instance, [5], [184], [202], [293]). In this way, the amount of exotic dark energy, dark matter, and ordinary baryonic matter is derived up to three significant digits and the age of the universe is derived even up to four significant digits as $t_0 = 13.82$ Gyr (see [213]). On the other hand, the age of some small stars (e.g. HD 140283, SM 0313) in our Galaxy is estimated to be at least 13.6 Gyr independently of cosmological models [28], i.e., these stars should have been formed about $t_1 = 220$ million years after the Big Bang, which is too short a time period for star formation. According to current models, the temperature of clouds of molecular hydrogen should be about 10 K, which is necessary for star formation by Jeans' criteria when gravity dominates over pressure. The temperature of the cosmic microwave background radiation (see Fig. 18.4) was much higher, $2.73(z + 1) \approx 50$ K, where the cosmological redshift $z \approx 17.5$ corresponds by [211] to the time t_1 .

For the time being, only the two coefficients $H_0 = H(t_0)$ (see (19.4)) and $q_0 = q(t_0) \approx -0.6$ (see (10.11)) of the Taylor series (10.10) of the expansion function have been measured (with very low precision),

$$a(t) = a(t_0)(1 + H_0(t - t_0) - \frac{1}{2}q_0H_0^2(t - t_0)^2 + \dots),$$

where the deceleration parameter $q = -\ddot{a}/(\dot{a})^2$ depends on the second derivatives of $a = a(t)$ (see Fig. 8.7).¹¹ Let us emphasize that calculation of the second derivatives from biased supernova data (cf. e.g. Fig. 10.3) is a very ill-conditioned problem. Originally, cosmologists believed that the expansion of the universe slows down, and therefore, they did not introduce the acceleration parameter but a deceleration parameter. Notice that e.g. $q(t_2) = \infty$ for the cyclic universe from Fig. 19.2.

It is evident that the first three terms of the Taylor expansion at the point t_0 , which corresponds to present time, cannot describe the behavior of the expansion function well in the far past (e.g. Fig. 13.4 for 10 Gyr ago). Moreover, we cannot reliably estimate the remainder of the Taylor series on the whole domain of definition (see [224, p. 652]),

¹¹If $p = 0$ in equation (19.2), then by relations (19.7) we get $q = \frac{1}{2}\Omega_M - \Omega_\Lambda$.

since the first derivatives of the expansion function $a = a(t)$ were extremely large just after the Big Bang (cf. e.g. (18.3) and (19.9)). Therefore, the Hubble age of the universe $H_0^{-1} = 13.6$ Gyr derived from the linear part of the above Taylor expansion cannot well approximate the real age of the universe.

Further, we have to emphasize that the Friedmann equation (19.1) was derived only for the gravitational interaction. However, shortly after the Big Bang, electromagnetic forces that are many orders of magnitude higher played an important role. Before that even stronger nuclear forces also surely had an influence on the initial values of the true expansion function. Although nongravitational forces are studied on large accelerators, their behavior in extremely strong gravitational fields right after the Big Bang is not known. Up to now we do not know which ones of the four fundamental forces played a crucial role during the Big Bang. In other words, the Friedmann equation (19.1) can hardly describe the evolution of the universe for small $t > 0$.

Since the product $\rho(t)a^3(t)$ is constant during the time period when matter dominates over radiation, equation (19.1) takes the equivalent form

$$\boxed{\dot{a}^2 = Aa^2 + B + \frac{C}{a}} \quad (19.9)$$

with time independent constant coefficients $A = \Lambda c^2/3$, $B = -kc^2$, and $C > 0$. Note the the right-hand side of (19.9) has to be nonnegative. From such a simple ordinary differential equation far-reaching conclusions about the deep past and the far future are made in [5], [202], ... Since the terminal condition $\dot{a}(t_0)/a(t_0) = H_0$ corresponding to the present time is known, we may solve equation (19.9) backward and also forward in time for given parameters A, B, C . Nevertheless, from such a simple ordinary differential equation (19.9) we should not make any categorical conclusions about the deep past and the future of the universe as is often done. For the time period when radiation dominates over matter, the term D/a^2 is added to the right-hand side of equation (19.9).

At present it is believed that $a(t) \rightarrow \infty$ for $t \rightarrow \infty$. By (19.6) and (19.7) for $k \leq 0$ it follows that $1 > \Omega_\Lambda(t)$, and thus $\frac{1}{3}\Lambda c^2 < H^2(t)$ for arbitrary time. From this and (19.5) we also observe that the time derivative of the expansion function grows beyond all bounds if Λ is a positive constant. Hence, also $\dot{a}(t) \rightarrow \infty$ for $t \rightarrow \infty$ in an infinite universe (hyperbolic or Euclidean).



19.5. Dark matter versus baryonic matter

In this section we present other arguments showing why the unjustified extrapolations from Section 19.3 are incorrect. According to the interpretation of measurements of the Planck satellite (see [212], [213, p. 11]) the parameter of the mass density in the standard cosmological model is equal to

$$\Omega_M = \Omega_{DM} + \Omega_{BM} \approx 0.32, \quad \Omega_{DM} \approx 0.27, \quad \Omega_{BM} \approx 0.05, \quad (19.10)$$

i.e., 27 % consists of dark matter (DM) and 5 % consists of baryonic matter (BM), from which less than 1 % is made up of luminous matter. Although the model arose by excessive extrapolations, and thus is probably not correct, some dark matter may exist. Scientific results should be independently verified. We suspect that the proposed ratio 27 : 5 of the mass density of dark matter to baryonic matter is highly exaggerated (see Fig. 9.9). Therefore, we will now briefly recall other methods that are independent of the normalized Friedmann equation (19.6).

From Chapter 7 we already know that the existence of dark matter was postulated in 1933 by Fritz Zwicky [304] after discovering large velocities of galaxies in the Coma cluster A1656. With the help of classical Newtonian mechanics he derived a very simple relation for the virial mass of the cluster (see (7.14))

$$M = \frac{5R\bar{v}^2}{3G}, \quad (19.11)$$

where R is its radius and \bar{v} is the root-mean-square speed of all galaxies with respect to the center of mass of the cluster. However, can we claim on the basis of such a trivial algebraic relation as (19.11) that dark matter in the Coma cluster (see Fig. 7.3) really exists? In Sections 8.3 and 8.4 we presented a number of arguments showing that the real amount of mass can be more than 50 % lower than the virial mass (19.11) and that the total mass of the cluster is at most $5\times$ larger than the mass of luminous matter. This essentially reduces the amount of dark matter and on the other hand increases the amount of baryonic matter.

Astronomical measurements are continually improving and thus we should soon find whether the current estimates of dark matter are true or not. For instance, recently Tutukov and Fedorova [282] found that the intergalactic medium of galaxy clusters contains 30–50 % of the total number of stars in the cluster (cf. also (8.18)). Moreover, by [2], [27], and [290] clusters of galaxies contain five times more baryonic matter in the form of hot gas producing X-rays than baryonic matter contained in galaxies. Consequently, the large velocities of galaxies in the Coma cluster observed by Zwicky have an entirely natural explanation by means only of baryonic matter [138].

Vera Rubin postulated dark matter in spiral galaxies, since by her calculation stars close to the edge of galaxies orbit too fast about the galactic center. At the end of the last century the number of red dwarfs in our Galaxy was considerably underestimated. Astronomers somewhat surprisingly believed that only 3 % of all stars are red dwarfs (see [22, p. 93]). However, at present we know that about 70 % of all stars are red dwarfs of the spectral class M (keeping in mind that brown dwarfs are difficult to detect). To assume at present the existence of 5–6 times larger amount of dark matter than baryonic matter to keep galaxies together seems to be quite overestimated. In Chapter 9 we showed that the observed large velocities of stars of the edge of our Galaxy can be naturally explained without dark matter by means of only baryonic matter, see e.g. inequality (9.6) or Theorem 9.1.

The above arguments show that dark matter can possibly be only a modeling error. Notice that the density parameter Ω_M of dark and baryonic matter exists by definition (19.7). Furthermore, Pavel Kroupa in [150] states several other arguments that point to the absence of dark matter around our Galaxy. A number of other works (see e.g. [13], [77], [82], [111], [138], [151], [192], [257]) also confirm that it is not necessary to assume the existence of dark matter. Finally note that the famous MOND (Modified Newtonian Dynamics) assumes infinite speed of gravity, which surely also contributes to the modeling error.



19.6. Dark energy versus cosmological constant

Generally, the prevailing conviction is that dark energy is some mysterious substance which is responsible for the accelerated expansion of the universe. According to data measured by the Planck satellite [213] the present density of dark and baryonic matter is almost 32 % and the density of dark energy about 68 % (see Fig. 9.9), since

$$\Omega_M \approx 0.3175, \quad \Omega_\Lambda \approx 0.6825, \quad \Omega_K \approx 0. \quad (19.12)$$

However, it is not said how to define such a percentage for $\Lambda < 0$ or $\Omega_K < 0$. Can it be negative or does one have to consider absolute values?

The above values (19.12) were obtained by a combination of the methods of Baryonic Acoustic Oscillations (BAO), Cosmic Microwave Background (CMB), and Supernovae type Ia explosions (SNe), see Fig. 19.5. Similar pictures can be found e.g. in [4] and [205]. It is argued that these three methods are independent and that the corresponding sets of admissible cosmological parameters intersect in a small region whose coordinates are close to (19.12). However, we should keep in mind that all three methods originate from the same normalized Friedmann equation (19.6) that was derived by inordinate extrapolations, and thus these methods **are not independent**. In

other words, the amount of exotic dark matter and dark energy was obtained from a questionable model that was solved by three different methods up to four significant digits, see (19.12).

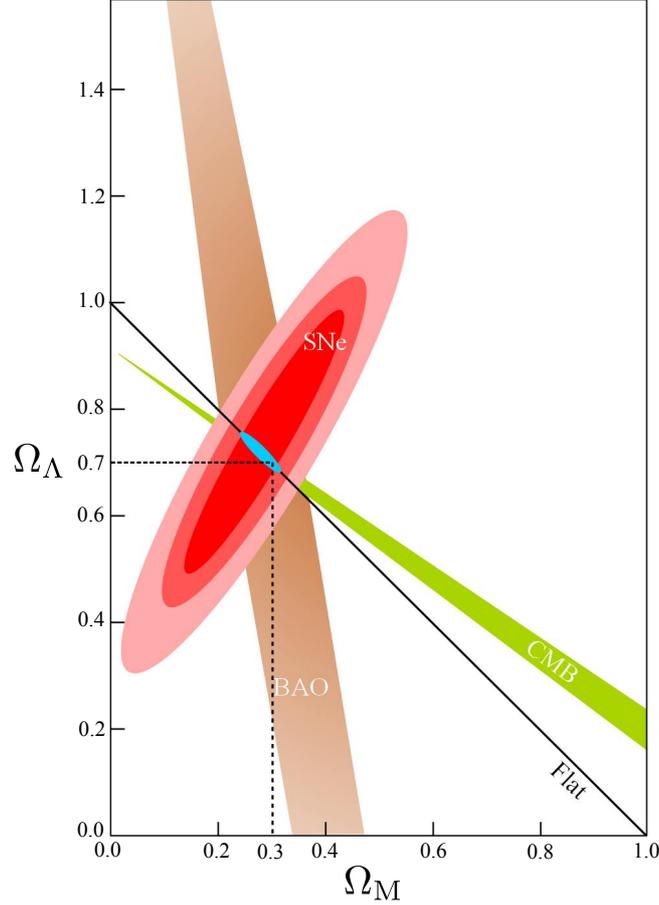


Fig. 19.5. Admissible values of cosmological parameters, obtained by three different methods: BAO, CMB, and SNe, intersect in a small region containing parameters (19.12).

The method SNe treats type Ia supernovae as standard candles. However, according to [12] and [236], they cannot be considered as standard candles due to the possible large extinction (reduction) of light from the supernova essentially depending on its location in the host galaxy, if it is in the middle or at the edge of a galaxy. It also depends on the direction of supernova rotational axis. In this way we may receive several orders of magnitude weaker light. Therefore, the vertical error bars in Fig. 10.4 should be extended downward.

For all of that, the following relation (see [47, p.511]), derived from the Friedmann equation, for the so-called *luminosity distance*¹² was used [226, p.1021] to establish distances of supernovae

$$d_L = \frac{c(1+z)}{H_0|\Omega_K|^{\frac{1}{2}}} \text{sinn}\left\{|\Omega_K|^{\frac{1}{2}} \int_0^z [(1+z)^2(1+\Omega_M z) - z(2+z)\Omega_\Lambda]^{-\frac{1}{2}} dz\right\},$$

where $H_0 \approx 70 \text{ km s}^{-1}\text{Mpc}^{-1}$ is the current value of the Hubble constant, $\Omega_K = 1 - \Omega_M - \Omega_\Lambda$ by (19.6),

$$\text{sinn } x = \begin{cases} \sin x & \text{for } k = 1, \\ x & \text{for } k = 0, \\ \sinh x & \text{for } k = -1, \end{cases}$$

and k is the normalized curvature.¹³ In [226, p.1021], the measured luminosity distance is then expressed by means of the measured redshift z and unknown cosmological parameters Ω_M , Ω_Λ , Ω_K , and H_0 . The likelihood for these cosmological parameters (19.12) is determined from a χ^2 statistic. Notice that for $\Omega_K = 0$ (cf. (19.12)) a division by zero appears in the relation for d_L . Thus, when $k = 0$ we should first reduce by $|\Omega_K|^{\frac{1}{2}}$, and then define d_L , cf. [226, p.1021]. In the case of an oscillating (or stationary) universe, the Hubble parameter vanishes at one point and we again divide by zero.¹⁴

The above values (19.12) were also derived by the method of baryonic acoustic oscillations [73] in the fluctuations of the cosmic microwave background radiation. The image of the Milky Way was carefully removed from Fig. 18.4. From Section 9.6 we know that the cosmic microwave background radiation was continually deformed by means of weak gravitational lensing of many galaxies and their clusters for more than 13 billion years (see Fig. 8.3). On the basis of such noisy data an extrapolation from $z = 1089$ to the present is made by

¹²Due to extinction we can get only a lower bound for the actual distance.

¹³Let us point out that the upper bound in the integral is the same as the integration variable, which should be denoted by a different symbol.

¹⁴Weinberg [293, p.421] presents a different relation for the “luminosity distance” which is independent of the Friedmann equation.

means of equation (19.1). In this way relations (19.12) and also (19.4) were obtained.

From Chapters 7–9 we know that the amount of dark matter (if it exists) is essentially not six times larger than the mass of baryonic matter as suggested in (19.10). Therefore, the value $\Omega_\Lambda \approx 0.6825$ in time $t_0 \approx 13.82$ Gyr is probably also far from reality. More precisely, we should say that the estimated age of the universe derived from the Λ CDM model for the parameters (19.12) is $t_0 \approx 13.82$ Gyr. The true age can be completely different.

From the relations (19.12) we observe that the sum of the measured values $\Omega_M(t_0)$ and $\Omega_\Lambda(t_0)$ is approximately equal to 1. This does not allow us to claim that from (19.6) and (19.7) it follows that $k = 0$ and that the true space is flat (i.e. infinite Euclidean) as it is often stated at present. Even if the sum were to be

$$\Omega_M(t_0) + \Omega_\Lambda(t_0) = 1.00000000000000000001,$$

we would still have a bounded universe that can be described by the sphere (18.1) with an incredibly large radius. Such a space is locally almost Euclidean, but finite. There is a big difference between a bounded and unbounded space (see also the very end of Chapter 18). Moreover, the sphere \mathbb{S}_r^3 has an entirely **different topology** than the flat space \mathbb{E}^3 which is promulgated by cosmologists at present.

In the Euclidean universe, very distant galaxies would recede by superluminal speeds — even arbitrarily large speeds,¹⁵ if they were to be sufficiently distant. This does not happen in a universe with a positive normalized curvature.

The physical dimension of the cosmological constant Λ is m^{-2} , since the left-hand side of (19.1) has dimension s^{-2} (the same as the sectional curvature from Section 18.5). Cosmologists describe it as the

¹⁵Here it is important to realize that these galaxies cannot be associated with inertial systems (cf. [294, p. 30])! For instance, for a constant value of the Hubble parameter (19.5) the expansion function is exponential.

density of energy which has another physical dimension in the SI units (International System of Units), namely

$$\text{kg m}^{-1}\text{s}^{-2}.$$

From definition (19.7) it is obvious that in quantities defining the density of dark energy $\Omega_\Lambda(t)$, the kilogram (kg) does not appear. Can we thus talk about density of energy?

We can easily verify that the physical dimension of the fraction $\frac{c^4}{G} \cdot \text{m}^{-2}$ is the same as the density of energy in the units $\text{kg m}^{-1}\text{s}^{-2}$. In the system $c = 1$ and $G = 1$ this is the same physical dimension as Λ has, since we may arbitrarily exchange kilograms, seconds, and meters using some appropriate multiplicative constants. In such a system, force, velocity, and power are dimensionless and we may evaluate energy and also time in kilograms. It is true that many relations will be much simpler in these restricted units, but the constants c and G in equation (19.1) are not equal to unity. Therefore, Λ cannot be interpreted as density of energy in the system SI.

Another possibility is to consider only the case $c = 1$. In this system we may define the density of energy by the relation $\rho_\Lambda = \Lambda/(8\pi G)$, since meters and seconds may be arbitrarily exchanged. This is again not the density of energy in the system SI.

Why should a single constant Λ truly model the accelerated expansion of the real universe. Is this not too big a simplification and too rough an approximation?

Dark energy was introduced into the standard cosmological model to explain the observed accelerated expansion of the universe and to eliminate the obvious violation of the energy conservation law. However, gravitational aberration (see Chapter 17) also has a repulsive character and thus it may generate the sought for energy necessary for the accelerated expansion. Antigravity (sometimes called a *dark force*) acts as a hidden repulsive force between planets, stars, galaxies, and their clusters and thus influences the expansion of the universe. It is only a side effect of gravity.

Nevertheless, the local expansion cannot be described by a single constant, since it depends on position and time. Its average values are not described by a fundamental constant. Therefore, we should rather consider a time dependent function $\Lambda = \Lambda(t)$ (like the Hubble parameter $H(t)$ which also depends on time).

The standard cosmological model assumes that the expansion of the universe is manifested only globally and not locally. However, according to [64], [65], [145], [175], and [302], the universe expands locally by a speed comparable with the Hubble constant H_0 , see also Chapters 11–16.



19.7. Conclusions

Current cosmological models are often identified with reality. Sometimes it is categorically claimed that *the universe is flat and that it consists of 68 % dark energy, 27 % of dark matter, and 5 % of baryonic matter.*

Rightly we should claim that *according to the standard cosmological model based on the Friedmann equation, the universe seems to be locally flat and it could consist of 68 % dark energy, 27 % of dark matter, and 5 % of baryonic matter.*

It is important to understand the difference between the above two assertions. We showed that the proclaimed dark matter and partly also dark energy can be explained as a modeling error of the Friedmann model. Alexander Friedmann applied Einstein's equations on the whole universe. When he published his famous paper [80] in 1922, astronomers had, of course, no idea about the real size of the universe, because other galaxies were only discovered later [104]. Hubble found the first Cepheid in the Andromeda nebulae M31 on October 6, 1923.

For the time being nobody knows how gravity behaves on cosmological scales. Nevertheless, cosmologists solve the normalized Friedmann equation (10.6) obtained by invalid extrapolations very exactly (up to four significant digits, see (19.12)).

If the Friedmann equation were to perfectly describe the rate of evolution of our universe, then the standard cosmological model would not possess so many problems and paradoxes, like e.g. the existence of some mysterious dark matter and dark energy, the horizon problem, the problem of homogeneity and isotropy, the flatness problem, the problem of exact setting of initial conditions, the problem of hierarchical structures, the problem of the existence of young stars orbiting the center of our Galaxy, the problem of the existence of giant black holes¹⁶ in early universe, and the problem of Big Bang itself.

At present it is very difficult to be familiar with the huge amount of information concerning cosmology. When reading literature on cosmology, it is often not clear what is a definition, what is an assumption, what is a statement, what is an experimentally verified fact, what are measured values and which values follow from some model, what is an attractive numerical simulation or artificially colored picture, and what is a serious estimate. A number of definitions are vague, confusing notation is used, measured data are wrongly interpreted, “serious” conclusions about the evolution of the universe are made from incorrectly derived equations, and so on. We often do not know in which way some statement was derived and then spread further without any verification. In this way a cosmological “folklore” arises.

For instance, in the current cosmology, we also often meet the following argumentation. Distances between galaxies increase and thus the entire universe was concentrated at one point in the past (see, for example, [202, p. 70], [294]). This implication is wrong from a mathematical point of view. As a counterexample it is enough to take the everywhere increasing non-Big-Bang expansion function

$$a(t) = C_1 + C_2 e^{C_3 t}, \quad t \in (-\infty, \infty), \quad (19.13)$$

which is not zero — nor arbitrarily close to 0 as t approaches $\pm\infty$ for

¹⁶It is not clear how antigravity acts inside black holes. Some black holes (e.g. the one in the center of the M87 galaxy) produce so giant jets that it seems to be impossible that these jets are supplied only from accretion disks.

positive constants C_1, C_2, C_3 . In this case the horizon problem does not arise.

We should not wonder why the difference between the measured and theoretically derived density of vacuum energy is 120 orders of magnitude (see [5, pp. 3, 109]). From this it is evident that the vacuum energy is not the main reason of the accelerated expansion of the universe.

No two different points in the universe (space) for a fixed time are causally connected. On the other hand, the observable universe, which is marked by a light cone in Fig. 18.5, is causally connected with our present time and position represented by the cone vertex. The present speed of expansion of the Universe at time t_0 should thus depend on the density of mass in the past, since gravitation has a finite speed of propagation. For instance, baryonic matter, which slows the expansion, should be influenced by the mass density at all previous time periods. Hence, the expansion function should be described by an equation whose solution depends on history,¹⁷ i.e. on all values $a(t)$ for $t \in (0, t_0)$. However, the Friedmann equation (19.1) does not have this property. It does not contain any delay given by the finite speed of gravitational interaction. It is a mere ordinary differential equation whose solution on the interval (t_0, ∞) depends only on the value of the expansion function at point t_0 and not on the history. This is another drawback of the standard cosmological Λ CDM model.



¹⁷The evolution of many dynamical systems (in biology, transport, robotics, theory of materials, telecommunication, . . .) essentially depends on the way in which the system got into its immediate state. A typical example is system (17.9)–(17.11) with delayed arguments.

20. Apparent superluminal speeds in the universe

*Anyone who claims that he understands cosmology,
just proves that he knows nothing.*

A paraphrase by ANTONÍN VRBA
of a quotation by Richard Feynmann

20.1. Observation of superluminal speeds

According to Einstein's General theory of relativity, no signal or mass can travel faster than the speed of light in the vacuum

$$c = 299\,792\,458 \text{ m/s.} \quad (20.1)$$

However, in October 1970 a group of radio-astronomers quite unexpectedly discovered quickly moving plasma jets from the quasar¹ 3C 279 while examining the fourth effect of the General theory of relativity [255]. The velocities of these jets calculated from angular measurements exceeded c , see [201, p. 3]. This phenomenon was independently confirmed (also for the neighboring quasar 3C 273) in February 1971 by two other teams of experts. Subsequently a series of hypotheses, seeking to explain this paradox of superluminal speeds, were created. Proponents of one theory argued that the distance to

¹Quasars are extremely distant objects with a very large redshift. They represent an early developmental stage of galaxies. They probably contain a supermassive black hole weighing 10^6 to 10^{10} Suns in their center.

the quasar by means of redshift measurements was estimated incorrectly. Others argued that the universe was much smaller at the time of plasma ejection, because the distance of the quasar is several billion light years from us, and therefore the measurement of the visual angle must be interpreted quite differently.

Since 1970, superluminal velocities were observed in many other quasars, see [183], [201], [219]. Some plasma jets achieved velocities far exceeding the speed of light up to ten times! The explanation based on incorrectly established distances from redshifts seemed to be very unlikely for such a high speed. Therefore, some scientists have begun to doubt the General theory of relativity, others the determination of the Hubble constant which characterizes the expansion of the universe, etc. But then a sensational discovery (cf. [181], [182]) of a microquasar² GRS1915+105 came which is only 40 000 light years ($\approx 3.78 \cdot 10^{17}$ km) from the Earth. Angular measurements showed that the plasma jets of this microquasar again apparently exceeded the speed of light. However, the most significant fact was that this object is in the Milky Way. Therefore, it was no longer necessary to take into account the expansion of our universe to explain the paradox of superluminal speeds, because the distance of the Earth from the microquasar is practically constant. The paper [183] demonstrates that the observed phenomenon is not just a flash of light passing through the plasma jets, but really the moving plasma.

In Fig. 20.1 we see that in merely 29 days, from 18 March 1994 to 16 April 1994, the angular distance between both plasma jets has increased to $0.816''$, which corresponds to approximately 10 000 au in the projection on the celestial sphere. The symbol + denotes the microquasar source location, which lies in the center of gravity of this system. Plasma jets had practically no impact on its position.

²A microquasar is a binary star, of which one component is a black hole weighing 6–10 solar masses, which sucks plasma from the second component and surrounds itself by the so-called accretion disk. A part of the plasma is ejected at tremendous speeds in two oppositely oriented narrow jets perpendicular to the plane of the disk due to extremely fast rotation of the black hole. The mechanism of this formation is still the subject of intense research.

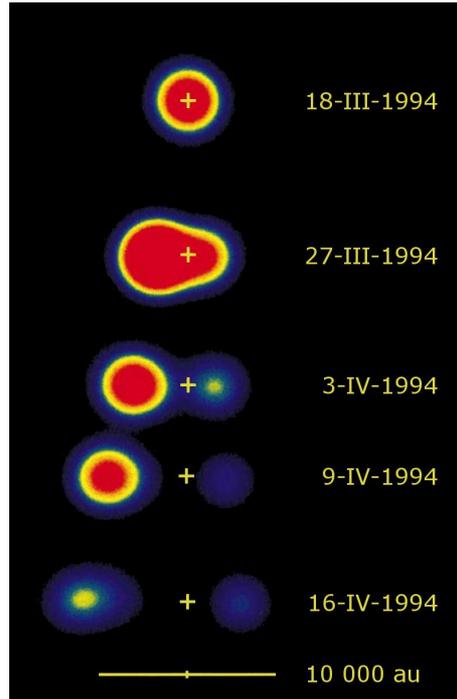


Fig. 20.1. Plasma jets from the microquasar GRS1915+105, which is denoted by the symbol +. Adapted from [183].

The brighter left jet is about 6235 au away from the microquasar and the right jet about 3360 au. As we shall see in the next section, this asymmetry can be explained by the fact that the axis of the left jet forms an acute angle with the viewing direction, i.e., the left plasma jet is facing us. From the perspective of the observer (see Fig. 20.1), it seems that this phenomenon corresponds to the speed

$$v^* = \frac{6235 \cdot 149.6 \cdot 10^6}{29 \cdot 24 \cdot 3600} = 372\,269 \text{ (km/s)} \quad (20.2)$$

on the celestial sphere, which obviously exceeds the speed of light (20.1).



20.2. Mathematical justification of the observed paradox

Suppose that a microquasar has a constant distance from us (otherwise we would have to use calculations involving the summation of relativistic speeds). Let $\alpha \leq 90^\circ$ denote the angle between the line microquasar–observer and the line along which the plasma jets move

(see Fig. 20.2). For simplicity, assume further that the actual speed v of the plasma jets is constant. Thus, the radial and tangential components relative to the observer are $v \cos \alpha$ and $v \sin \alpha$, respectively. The plasma reaches the distance vt from microquasar at time t .

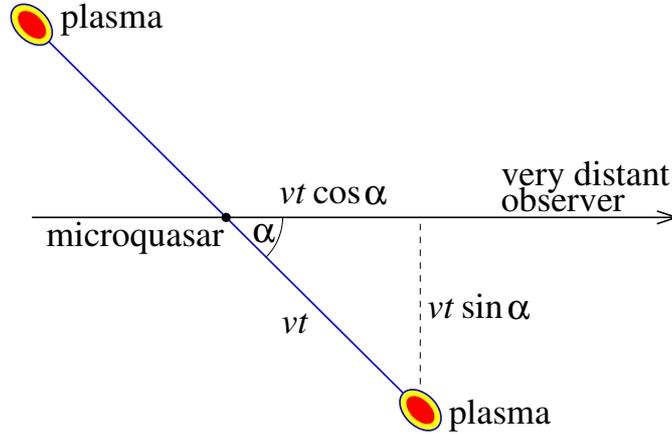


Fig. 20.2. A schematic illustration of the examined phenomenon

For $\alpha < 90^\circ$ the left plasma jet actually approaches the observer. After the time period t it is about $vt \cos \alpha$ closer to him/her than the microquasar. Hence, the time period t^* , at which the observer sees the plasma jet moving from the microquasar until its current position in Fig. 20.2, is smaller than the actual time period t . Therefore,

$$t^* = t - \frac{v}{c} t \cos \alpha, \quad (20.3)$$

where $(vt \cos \alpha)/c$ is the time period during which light covers the distance $vt \cos \alpha$ at the **finite** speed c . This is the main reason for the paradox, because the apparent speed of the left plasma jet equals by (20.3)

$$v^* = \frac{vt \sin \alpha}{t^*} = \frac{v \sin \alpha}{1 - (v/c) \cos \alpha} \quad (20.4)$$

and this fraction can easily be greater than c for appropriate data. For example, for the microquasar from Fig. 20.1 according to [181] and [183] (see also Remark 20.3), the angle is $\alpha = 71^\circ$ and the actual speed³

³Almost the same speed of jets was found at the microquasar GRO J1655-40, whose distance from us is just 10000 light years, see [183].

is $v = 0.92c$. Substituting these values into (20.4), the corresponding apparent superluminal speed is

$$\boxed{v^* = 1.24c}$$

and is fully in accordance with the measured speed (20.2). The time period for which the plasma jets actually traveled is

$$t = 42.3 \text{ days}$$

and the light covers the distance $vt \cos \alpha$ from Fig. 20.2 during 13.3 days. However, the time period of the phenomenon observed from the Earth is only

$$t^* = 42.3 - 13.3 = 29 \text{ days.}$$

Light from the final phase of the phenomenon flew to us for about 13.3 days less than from the initial phase, because it already does not need to overcome the distance $vt \cos \alpha$.

Remark 20.1. Notice that the function $v^* = v^*(\alpha, v)$ defined by the right-hand side of relation (20.4) on the set $[0, \frac{1}{2}\pi] \times [0, c)$ has an essential singularity at the point $(0, c)$. By an appropriate choice of the angle α and actual speed $v < c$, we can even reach an arbitrarily large value of the apparent velocity v^* . For example, taking $\alpha = 8^\circ$ and $v = 0.99c$, we get that $v^* = 7c$. On the other hand, for $\alpha = 90^\circ$ the relation (20.4) gives $v^* = v$, and therefore, no superluminal speed is obtained.

Remark 20.2. In Fig. 20.3, we see for which of the pairs α and v the apparent velocity v^* is larger or smaller than the speed of light. The corresponding interface between these two regions is given by the function

$$v_c(\alpha) = \frac{c}{\sin \alpha + \cos \alpha},$$

which can be directly derived from (20.4) in such a way that v is expressed as a function of α for a fixed $v^* = c$. The dashed lines show the graphs of the functions $v_{2c}(\alpha)$ and $v_{c/2}(\alpha)$ which correspond to the apparent velocities $v^* = 2c$ and $v^* = c/2$, respectively.

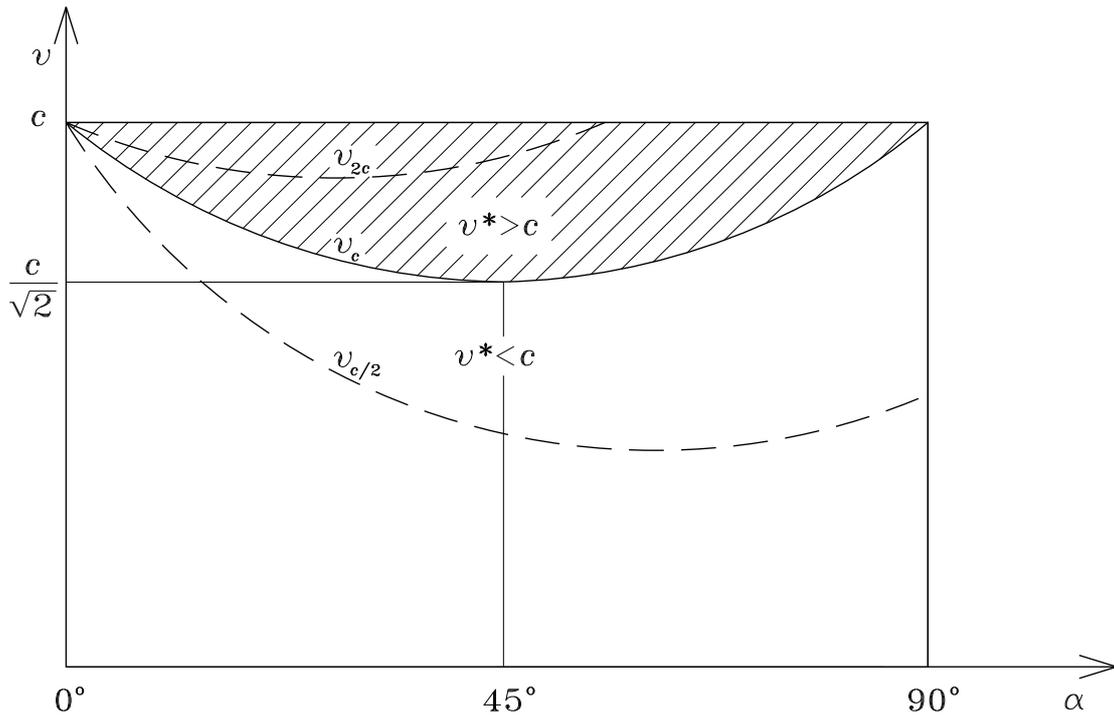


Fig. 20.3. If the pair α and v falls resides in the hatched area, then the observed speed is apparently superluminal.

Remark 20.3. The apparent speed of the left plasma jet satisfies equation (20.4) whose left-hand side is given by (20.2). Replacing the minus sign in (20.4) by the plus sign, we get another equation for the apparent velocity $\bar{v}^* = 200\,815$ km/s of the right plasma jet. In this way we obtain a system of two nonlinear algebraic equations for two unknowns v and α .



20.3. Superluminal velocities at cosmological distances

To explain the observed superluminal speeds of ejections of plasma at distant quasars (i.e. objects outside our Galaxy), other effects in addition to relations (20.3)–(20.4) must be taken into account. According to the General theory of relativity, apparently superluminal speeds can theoretically be caused by gravitational lensing, which may occur when an intermediate galaxy distorts the appearance of a more distant galaxy by bending the light from the more distant galaxy as

it approaches the neighborhood of the intermediate galaxy like a converging glass lens (see e.g. (8.15) and [101]).

In the next section we will show how the intrinsic expansion of the universe provides for a phenomenon which we will call a time lens and which also has an influence on the observed superluminal velocities at cosmological distances. To this end, let us briefly recall the generally accepted model of the universe with positive curvature from Chapter 18 (see also [184, p.724]). In this idealized model, the universe (space) is characterized by the three-dimensional surface (18.1) of an expanding four-dimensional sphere of radius $r = r(t)$. For any fixed time instant t the expansion at all points and in all directions is the same. Such a model is called homogeneous and isotropic, because all local irregularities are ignored. This is an important simplifying assumption based on Einstein's cosmological principle from Section 10.3.

The intrinsic expansion of the universe is often demonstrated by the surface of an inflating balloon provided that one space dimension is removed. Removing another spatial dimension, we get the situation of Fig. 20.4. Here the "universe" is a circle whose radius increases with time.

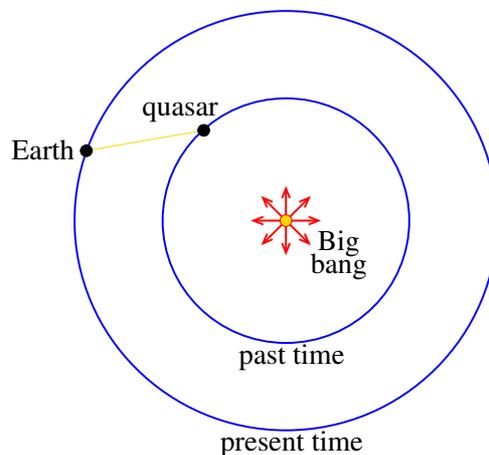


Fig. 20.4. A schematic illustration of an expanding universe. The projection of spacetime (cf. Fig. 18.5) in the direction of the time axis t gives a model of the universe similar to that of an inflating balloon. The trajectory of a photon from a distant quasar to the Earth is given in yellow.



20.4. The time-lens principle

In this section, we show that the intrinsic expansion of the universe seemingly magnifies very distant objects. Let us point out that the time lens has nothing to do with gravitational lenses. Its magnification depends nonlinearly on the distance to an observed object. It functions especially at very “large” distances — at least several billion light years. On the other hand, magnification of the time lens for “short” distances is almost negligible.

We see quasars several billion light years from us with a time delay given by the finite speed of light. Therefore, we must consistently distinguish between “at that time” and “actual dimensions”, i.e. the size of the universe when observed photons left the quasar and the universe was much smaller, and “today’s” dimensions when “ancient” photons (see the line Earth–quasar on Fig. 20.4) arrived at our ground-based telescopes. Roughly speaking, the younger the objects which are observed, the larger the magnification appears.

Therefore, by angular measurements we paradoxically see a very distant object as being larger. We will call this phenomenon the *time lens*. The magnification is given by the relation

$$Z = z + 1,$$

where z is the associated redshift. Let us illustrate supporting evidence of the functioning of the time lens by three real-world examples.

Example 20.1. Fig. 16.1 shows the universe as it appeared approximately 10–12 Gyr ago. This is the famous Hubble Deep Field, the center of which has coordinates RA = 12 h 37 min and DE = 62° 13′. The width of this image corresponds to the small angle 3′. By the graph on Fig. 8.7 it contains galaxies with redshifts between $z \approx 2$ and $z \approx 3$. Galaxies were being formed at that time and thus were smaller than today’s already developed galaxies (see Chapter 16). The corresponding magnification is up to $Z \approx 4$.

In 1998, photographs of the so-called Hubble Deep Field South (HDFS) and later also the Hubble Ultra Deep Field⁴ (HUDF) and Hubble Extreme Deep Field (XDF) were obtained, where some galaxies are even more distant. In this case, the magnification effect is slightly larger than that in Fig. 16.1.

Example 20.2. Another argument for the functioning of the time lens is given by the existence of the cosmic microwave background (CMB) radiation, which comes from the period 380 000 years after the Big Bang, when atoms were created and the universe started to be transparent for photons. This almost homogeneous and almost isotropic radiation comes from all directions of the sky at the distance of about 13.7 Gly — the so-called horizon of the observable universe. It came into existence in the period when the universe was more than 1000 times smaller than it is at present. According to [73], the corresponding redshift is $z = 1089$. From this we can estimate the behavior of the expansion function $r = r(t) = a(t)$ from relation (18.3) in a neighborhood of the origin, i.e. $1090 r(t_1) \approx r(t_0)$ for $t_1 = 380\,000$ yr, where $r(t_0)$ is the present value of the expansion function. The expansion speed given by $H = H(t)$ was very large at that time. When the cosmic microwave background radiation was created, the universe would be represented in Fig. 20.4 by an extremely small circle with radius ca. 1090 times smaller than the radius of the circle corresponding to the present time. The associated magnification is given by $Z = 1090 = r(t_0)/r(t_1)$.

Example 20.3. Finally, the last example of an enormous magnification by the time lens is the Big Bang itself, which appeared roughly 13.8 Gyr ago. Although it happened in a minimal volume, its present position is on the possibly greatest sphere (the so-called horizon) with an unimaginable large radius (much larger than in the previous exam-

⁴HUDF contains about 5500 galaxies on the area $1.5' \times 1.5'$. From Einstein's cosmological principle and (2.14) we find that the number of galaxies in the observable universe is approximately $5500 \cdot (60/1.5)^2 \cdot 41\,253 = 363 \cdot 10^9$. This represents a lower bound of all galaxies of the universe at present. For instance, galaxies that were formed on the manifold corresponding to $z \approx 1000$ (cf. Fig. 18.4) are not visible yet.

ple). Relict neutrino or relic gravitational waves would be needed for the detection of this horizon.

Thus, the farther we look, the corresponding sphere seems to be bigger and bigger, even though the universe was smaller and smaller. This is the main principle of the functioning of the time lens. However, it manifests itself at very large distances on the order of billions of light years. Monitored objects from the time t are seemingly magnified $r(t_0)/r(t)$ -times and values of the function $r = r(t)$ are “small” just close to the origin. Therefore, it may happen that a very distant galaxy for $z \approx 1$, which is ten times farther away from us than another galaxy of the same size at $z \approx 0.1$, will apparently be only about five times smaller than the nearer galaxy (see [293, p. 423] for angular distance).

Note that the observed superluminal speeds in no way contradict the General theory of relativity. Roughly speaking, they can be explained by relations (20.3)–(20.4). For distant objects outside our Galaxy it is also necessary to take into account the expansion of the universe itself, which causes the time lens effect. It is also necessary to give a proper interpretation of the size of the measured angle when looking into the depths of the universe. Shortly after the Big Bang, the universe could actually expand faster than light, because the corresponding expansion function has large derivatives near the origin (see Fig. 13.4) than is probably so at present. For example, if \mathbb{S}_r^3 were to be the correct model of our universe, then by (10.8) its radius $r = r(t)$ would increase on average by the highly superluminal speed $140 \text{ Gly}/14 \text{ Gyr} = 10c$.



20.5. The Big Bang singularity

The standard cosmological model assumes that the time flows completely uniformly from the Big Bang on. Therefore, we often hear the question (see e.g. [184, p. 769]): *What was before the Big Bang?*

It is important to realize that in the observable universe we actually look in any direction into the vast spacetime singularity. The more

distant the objects that are observed, the more it seems to us that time passes more slowly. If there were a huge clock placed at $z = 1$ from the Earth, we would see that it goes twice as slow. The largest currently observed distance corresponds to the CMB with redshift $z = 1089$ which appeared 380 000 years after the Big Bang. In this case, during one earthly hour we would observe that the CMB clock had moved only about $3.3 = \frac{3600}{1090}$ seconds. If we would ever see objects more distant than $z = 1089$ (using e.g. detectors of relic neutrinos), we would find that time there passes even more slowly, and so on. Thus the Big Bang singularity deforms not only space but also time.

In cosmological models we should therefore define exactly what is one second at least in the period 380 000 years after the Big Bang. One second is currently defined through the transition between the two levels of the basic state of the cesium atom at rest at 0 K. However, how do we define one second in the period when no cesium existed? It is mostly done by extrapolating half-lives of known particles. It is not clear whether we can use this kind of definition for the extremely strong gravitational fields that were present just after the Big Bang, and to talk about the time period 10^{-43} s, when no known particles existed.

According to Example 20.3, the Big Bang is located beyond the sphere of the cosmic microwave background radiation in every direction. It was caused by a yet unknown antigravitational force.



References

- [1] C. Adami et al., *The build-up of the Coma cluster by infalling substructures*. *Astron. Astrophys.* **443** (2005), 17–27.
- [2] S.W. Allen, A.E. Evrard, A.B. Mantz, *Cosmological parameters from observations of galaxy clusters*. *Annual Rev. Astron. Astrophys.* **49** (2011), 409–470.
- [3] R. A. Alpher, H. Bethe, G. Gamow, *The origin of chemical elements*. *Phys. Rev.* **73** (1948), 803–804.
- [4] R. Amanullah et al., *Spectra and HST light curves of six type Ia supernovae at $0.511 < z < 1.12$ and the Union2 compilation*. *Astrophys. J.* **716** (2010), 712–738.
- [5] L. Amendola, S. Tsujikawa, *Dark energy – Theory and observations*. Cambridge Univ. Press, Cambridge, 2010.
- [6] C. Amsler et al., *Review of particle physics*. *Phys. Lett. B* **667** (2008), Sec. 19: Big-Bang Cosmology, 217–227.
- [7] A.I. Arbab, *Evolution of angular momenta and energy of the Earth-Moon system*. *Acta Geod. Geoph. Hung.* **40** (2005), 33–42.
- [8] A. I. Arbab, *Flat rotation curve without dark matter: the generalized Newton’s law of gravitation*. *Astrophys. Space Sci.* **355** (2015), 343–346.
- [9] Aristarchus of Samos, *Peri megethón kai apostemáton helión kai selénes*. Translated from ancient Greek to Latin by F. Commandino in 1572.
- [10] Aristote, *Du ciel*. 350 BC, text établi et traduit par P. Moraux, Les Belles Lettres, Paris, 1965.
- [11] W. Baade, F. Zwicky, *Cosmic rays from super-novae*. *Proc. Nat. Acad. Sci.* **20** (1934), 254–263.
- [12] L.G. Balázs et al., *A possible interrelation between the estimated luminosity distances and the internal extinctions of type Ia supernovae*. *Astronom. Nachr.* **327** (2006), 917–924.
- [13] D. G. Banhatti, *Newtonian mechanics & gravity fully model disk galaxy rotation curves without dark matter*. ArXiv: 0806.1131, 2008, 1–6.

- [14] J. D. Barrow, F. J. Tipler, *The anthropic cosmological principle*. Oxford Univ. Press, 1986.
- [15] G. Battaglia et al., *The radial velocity dispersion profile of the Galactic halo: Constraining the density profile of the dark halo of the Milky Way*. *Astroph/0506102v2*, 2008, 1–11.
- [16] A. F. Beardon, *The geometry of discrete groups*. Springer, New York, 1983.
- [17] M. Beijersbergen et al., *Deprojection of luminosity functions of galaxies in the Coma cluster*. *Astron. Astrophys.* **390** (2002), 817–820.
- [18] J. Bekenstein, *Relativistic gravitation theory for the modified Newtonian dynamics paradigm*. *Phys. Rev. D* **70** (2004), 083509.
- [19] M. Belet, A. Belet, *Look at the stars and become a geometer!* In: *History of Mathematics: Histories of Problems, Ellipses*, Paris, 1997, 255–283.
- [20] B. Bertotti, P. Farinella, D. Vokrouhlický, *Physics of the Solar system*. Kluwer, Dordrecht, 2003.
- [21] B. G. Bills et al., *Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos*. *J. Geophys. Res.* **110** (2005), E07004, 15 pp.
- [22] J. Binney, M. Merrifield, *Galactic astronomy*. Princeton, 1998.
- [23] J. Binney, S. Tremaine, *Galactic dynamics*. Princeton Univ. Press, 1987.
- [24] A. Biviano et al., *A catalogue of velocities in the central region of the Coma cluster*. *Astron. Astrophys. Suppl. Ser.* **111** (1995), 265–274.
- [25] R. D. Blandford, *Strong gravitational lensing on the Hubble deep field*. In: *Galaxy Interactions at Low and High Redshift*, eds. J. E. Barnes, D. B. Sanders, *Internat. Astron. Union*, vol. 186, 1999, 439–445.
- [26] D. Blanuša, *Über die Einbettung hyperbolischer Räume in euklidische Räume*. *Monatsh. Math.* **59** (1955), 217–229.
- [27] H. Böhringer, N. Werner, *X-ray spectroscopy of galaxy clusters: studying astrophysical processes in the largest celestial laboratories*. *Astron. Astrophys. Rev.* **18** (2010), 127–196.
- [28] H. E. Bond et al., *A star in the Solar neighborhood that formed shortly after the Big Bang*. *Astrophys. J.* **765** (2013), L12.

- [29] A. Bosma, *Dark matter in galaxies: Observational overview*. In: Dark Matter in Galaxies, IAU Sympos. 220 (eds. S. Ryder, D. J. Pisano, M. Walker, K. C. Freeman), 2003, 1–12.
- [30] R. J. Bouwens et al., *A candidate redshift $z \approx 10$ galaxy and rapid changes in that population at an age of 500 Myr*. Nature **469** (2011), 504–507.
- [31] J. Bovy, S. Tremaine, *On the local dark matter density*. Astrophys. J. **756** (2012), 89, 6 pp.
- [32] D. Brander, *Isometric embeddings between space forms*. Master Thesis, Univ. of Pennsylvania, 2003, 1–48.
- [33] J. Brandts, M. Křížek, Z. Zhang, *Paradoxes in numerical calculations*. Neural Netw. World **26** (2016), 12 pp.
- [34] P. Bretagnon, G. Francou, *Planetary theories in rectangular and spherical variables. VSOP 87 solutions*. Astron. Astrophys. **202** (1988), 309–315.
- [35] T. Broadhurst, E. Scannapieco, *Detecting the gravitational redshift of cluster gas*. Astrophys. J. **533** (2000), L93–L97.
- [36] M. Brooks, *13 things that do not make sense*. New Scientist **2491** (2005), 30–37.
- [37] D. J. A. Brown et al., *Are falling planets spinning up their host stars?*. Mon. Not. R. Astron. Soc. **415** (2011), 605–618.
- [38] G. Bruno, *De l'Infinito, Universo e Mondi*. Venezia, 1584.
- [39] F. Buitrago et al., *Shaping massive galaxies: their morphology and kinematics at $z = 1 - 3$* . Highlights of Spanish Astrophysics VI, Proc. of the IX Sci. Meeting of the Spanish Astronom. Soc., Madrid (ed. M. R. Zapatero et al.), 2010, 154–160.
- [40] M. Burša, *Decrease in spin rate of Mars due to tidal torques exerted by Phobos and Sun*. Bull. Astron. Inst. Czechosl. **39** (1988), 168–171.
- [41] M. Burša, K. Peč, *Gravity field and dynamics of the Earth*. Springer, Berlin, 1993.
- [42] J. W. Cannon, W. J. Floyd, R. Kenyon, W. R. Parry, *Hyperbolic geometry*. In: Flavors of Geometry, Math. Sci. Res. Inst. Publ. **31**, Cambridge Univ. Press, 1997, 59–115.

- [43] A. Cappi, *Gravitational redshift in galaxy clusters*. *Astron. Astrophys.* **301** (1995), 6–10.
- [44] S. Carlip, *Aberration and the speed of gravity*. *Phys. Lett. A* **267** (2000), 81–87.
- [45] M. Carrera, D. Giulini, *Influence of global cosmological expansion on local dynamics and kinematics*. *Rev. Mod. Phys.* **82** (2010), 169–208.
- [46] B.W. Carroll, D.A. Ostlie, *Introduction to modern astrophysics*. Pearson Addison-Wesley, 2007.
- [47] S.M. Carroll, W.H. Press, E.L. Turner, *The cosmological constant*. *Annu. Rev. Astron. Astrophys.* **30** (1992), 499–542.
- [48] B. Carter, *Large number coincidences and the Anthropic Principle in cosmology*. In: *IAU Symposium 63, Confrontation of Cosmological Theories with Observational Data* (ed. M. S. Longair), Riedel, Dordrecht, 1974, 291–298.
- [49] C.V.L. Charlier, *How an infinite world may be built up?*. *Arkiv för Mat. Astronom. Fys.* **16** (1922), 1–34.
- [50] C.F. Chyba, D.G. Jankowski, P.D. Nicholson, *Tidal evolution in the Neptune-Triton system*. *Astron. Astrophys.* **219** (1989), L23–L26.
- [51] B.E. Clotfelter, *The Cavendish experiment as Cavendish knew it*. *Amer. J. Phys.* **55** (1987), 210–213.
- [52] D. Clowe et al., *A direct empirical proof of the existence of dark matter*. *Astrophys. J. Lett.* **648** (2006), L109–L113.
- [53] W. Cochran, F.H.C. Crick, V. Vand, *The structure of synthetic polypeptides. I. The transform of atoms on a helix*. *Acta Cryst.* **5** (1952), 581–586.
- [54] M. Colless, A.M. Dunn, *Structure and dynamics of the Coma cluster*. *Astrophys. J.* **458** (1996), 435–454.
- [55] F.I. Cooperstock, V. Faraoni, D.N. Vollick, *The influence of the cosmological expansion on local systems*. *Astrophys. J.* **503** (1998), 61–66.
- [56] N. Copernicus, *Complete works, vol. II, On the revolutions*. Polish Sci. Publishers, Warsaw–Krakw, 1978.
- [57] C.M. Cox, B.F. Chao, *Detection of large-scale mass redistribution in the terrestrial system since 1998*. *Science* **297** (2002), 831–833.

- [58] H. D. Curtis, *Novae in spiral nebulae and the island universe theory*. Publ. Astronom. Soc. Pacific **29** (1917), 206–207.
- [59] I. Damjanov et al., *Red nuggets at high redshift: structural evolution of quiescent galaxies over 10 Gyr of cosmic history*. Astrophys. J. Lett. **739** (2011), L44.
- [60] G. F. Davies, *Thermal evolution of the mantle*. Treatise on Geophysics, vol. 9, Evolution of the Earth (ed. D. J. Stevenson), Elsevier, 2007, 197–216.
- [61] T. M. Davis, C. H. Lineweaver, *Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe*. Publ. Astronom. Soc. Australia **21** (2004), 97–109.
- [62] W. De Sitter, *On the relativity of inertia. Remarks concerning Einstein latest hypothesis*. Proc. Kon. Ned. Acad. Wet. **19** (1917), 1217–1225.
- [63] J. O. Dickey et al., *Lunar laser ranging: A continuing legacy of the Apollo program*. Science **265** (1994), 482–490.
- [64] Y. V. Dumin, *A new application of the Lunar laser retroreflectors: Searching for the “local” Hubble expansion*. Adv. Space Res. **31** (2003), 2461–2466.
- [65] Y. V. Dumin, *Testing the dark-energy-dominated cosmology by the Solar-System experiments*, Proc. of the 11th Marcel Grossmann Meeting on General Relativity (eds. H. Kleinert, R. T. Jantzen, R. Ruffini), World Sci., Singapore, 2008, 1752–1754, arXiv: 0808.1302.
- [66] Y. V. Dumin, *The faint young Sun paradox in the context of modern cosmology*. Astron. Tsirkulyar **1623** (2015), 1–5, arXiv: 1505.03572v1.
- [67] G. Dvali, A. Gruzinov, M. Zaldarriga, *The accelerated universe and the Moon*. Phys. Rev. D **68** (2003), 024012.
- [68] A. Eddington, *Space, time and gravitation*. Cambridge Univ. Press, Cambridge, 1966.
- [69] A. Einstein, *Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes*. Ann. d. Phys. **35** (1911), 898–908.
- [70] A. Einstein, *Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie*. Königlich-Preuß. ische Akad. Wiss. Berlin (1915), 831–839. English translation “Explanation of the perihelion motion of Mercury from general relativity theory” by R. A. Rydin with comments by A. A. Vankov, 1–34.

- [71] A. Einstein, *The foundation of the general theory of relativity*, Ann. Phys. **49** (1916), 769–822.
- [72] A. Einstein, *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*. Königlich-Preuss. Akad. Wiss. Berlin (1917), 142–152.
- [73] D. J. Eisenstein, C. L. Bennett, *Cosmic sound waves rule*. Physics Today **61** (2008), 44–50.
- [74] A. E. Evrard et al., *Virial scaling of massive dark matter halos: Why clusters prefer a high normalization cosmology*. Astrophys. J. **672** (2008), 122–137.
- [75] F. J. M. Farley, *Does gravity operate between galaxies? Observational evidence re-examined*. Proc. Roy. Soc. A **466** (2010), 3089–3096.
- [76] J. Q. Feng, C. F. Gallo, *Mass distribution in rotating thin-disk galaxies according to Newtonian dynamics*. Galaxies **2** (2014), 199–222.
- [77] J. Q. Feng, C. F. Gallo, *Deficient reasoning for dark matter in galaxies*. Phys. Internat. **6** (2015), 1–12.
- [78] A. Ferré-Mateu, I. Trujillo, *Superdense massive galaxies in the nearby universe*. Proc. of the XXVII. General Assembly of IAU, S262 (eds. G. Bruzual, S. Charlot), Kluwer, 2010, 331–332.
- [79] A. V. Filippenko, *Einstein’s biggest blunder? High-redshift supernovae and the accelerating universe*. ArXiv: astro-ph/0109399v2, 2001.
- [80] A. Friedman, *Über die Krümmung des Raumes*. Z. Phys. **10** (1922), 377–386. English translation: *On the curvature of space*. General Relativity and Gravitation **31** (1999), 1991–2000.
- [81] A. Friedmann, *Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes*. Z. Phys. **21** (1924), 326–332. English translation: *On the possibility of a world with constant negative curvature of space*. General Relativity and Gravitation **31** (1999), 2001–2008.
- [82] C. F. Gallo, J. Q. Feng, *Galactic rotation described by a thin-disk gravitational model without dark matter*. J. Cosmology **6** (2010), 1373–1380.
- [83] G. Gamow, *Rotating universe?* Nature **158** (1946), 549.
- [84] R. Gavazzi, C. Adami et al., *A weak lensing study of the Coma cluster*. Astron. Astrophys. **498** (2009), L33–L36.

- [85] I. Gilmour, M. A. Sephton (eds.), *An introduction to astrobiology*. Cambridge Univ. Press, Cambridge, 2003.
- [86] J. Glanz, *Astronomers see a cosmic antigravity force at work*. *Science* **279** (1998), no. 5355, 1298–1299.
- [87] E. N. Glass, *Gravothermal catastrophe, an example*. *Phys. Rev. D* **82** (2010), 044039.
- [88] K. Gödel, *An example of a new type of cosmological solutions of Einstein's field equations of gravitation*. *Rev. Mod. Phys.* **21** (1949), 447–450.
- [89] B. R. Goldstein, *Eratosthenes on the “measurement” of the Earth*. *Historia Math.* **11** (1984), 411–416.
- [90] V. González et al., *Evolution of galaxy stellar mass functions, mass densities, and mass to light ratios from $z \sim 7$ to $z \sim 4$* . *Astrophys. J. Lett.* **735** (2011), L34.
- [91] M. D. Gregg, M. J. West, *Galaxy disruption as the origin of intracluster light in the Coma cluster of galaxies*. *Nature* **396** (1998), 549–552.
- [92] J. B. Griffiths, J. Podolský, *Exact space-times in Einstein's general relativity*. Cambridge Univ. Press, 2009.
- [93] P. Guillermier, S. Koutchmy, *Total eclipses*. Springer, 1999.
- [94] J. E. Gunn, B. Tinsley, *An accelerating Universe?* *Nature* **257** (1975), 454–457.
- [95] E. Halley, *Methodus singularis qua Solis parallaxis sive distantia a Terra, ope Veneris intra Solem conspicienda, tuto determinari poterit*. *Trans. Roy. Soc. London* (1716), 454–564.
- [96] P. S. Harrington, *Eclipse!* John Wiley, New York, 1997.
- [97] W. E. Harris, *Catalog of parameters for Milky Way globular clusters: The database*. Feb. 2003. See also *Astrophys. J.* **112** (1996), 1487.
- [98] W. K. Hartmann, *Mars*. Workman Publ., New York, 2003.
- [99] W. D. Heintz, *A study of multiple-star systems*. *Astronom. J.* **111** (1996), 408–411.

- [100] C. Hellier, *An orbital period of 0.94 days for the hot-Jupiter planet WASP-18b*. *Nature* **460** (2009), 1098–1100.
- [101] J.N. Hewitt, *Gravitational lenses*. *Ann. New York Acad. Sci.* **688** (1993), 250–259.
- [102] D. Hilbert, *Über Flächen von constanter gausscher Krümmung*. *Trans. Amer. Math. Soc.* **2** (1901), 87–99.
- [103] R.D. Holder, S. Mitton (eds.), *Georges Lemaître: Life, science and legacy*. Springer, Berlin, Heidelberg, 2012.
- [104] E. Hubble, *Cepheids in spiral nebulae*. *The Observatory* **48** (1925), 139–142.
- [105] E. Hubble, *A relation between distance and radial velocity among extragalactic nebulae*. *Proc. Nat. Acad. Sci. USA* **15** (1929), 168–173.
- [106] J.P. Hughes, *The mass of the Coma cluster: Combined X-ray and optical results*. *Astrophys. J.* **337** (1989), 21–33.
- [107] R. A. Hulse, J. H. Taylor, *Discovery of a pulsar in a binary system*. *Astrophys. J.* **105** (1975), L51–L53.
- [108] C. Impey, W.K. Hartmann, *The universe revealed*. Brooks-Cole, 2000.
- [109] A. Irrgang, B. Wilcox, E. Tucker, L. Schiefelbein, *Milky Way mass models for orbit calculations*. *Astronom. Astrophys.*, arXiv: 1211.4353v4, 2014, 1–13.
- [110] R. A. Jacobson, *The orbits and masses of the Martian satellites and the libration of Phobos*. *Astronom. J.* **139** (2010), 668–679.
- [111] J. Jaloča, Ł. Bratek, M. Kutschera, *Is dark matter present in NGC 4736? An iterative spectral method for finding mass distribution in spiral galaxies*. *Astrophys. J.* **679** (2008), 373–378.
- [112] G.C. Jordan IV et al., *Three-dimensional simulations of the deflagration phase of the gravitationally confined detonation model of type Ia supernovae*. *Astrophys. J.* **681** (2008), 1448–1457.
- [113] N. A. Kaib, R. Roškar, T. Quinn, *Sedna and the Oort cloud around a migrating Sun*. *Icarus* **251** (2011), 491–507.
- [114] S.M. Kent, J.E. Gunn, *The dynamics of rich clusters of galaxies, I. The Coma cluster*. *Astronom. J.* **87** (1982), 945–971.

- [115] J. Kepler, *Astronomia nova ...* Heidelberg, 1609.
- [116] A. W. Kerr, J. C. Hauck, B. Mashhoon, *Standard clocks, orbital precession and the cosmological constant*. *Classical Quant. Grav.* **20** (2003), 2727–2736.
- [117] H. Kienert, G. Feulner, V. Petoukhov, *Faint young Sun problem more severe due to ice-albedo feedback and higher rotation rate of the early Earth*. *Geophys. Res. Lett.* **39** (2012), L23710.
- [118] Y.-R. Kim, R. A. C. Croft, *Gravitational redshifts in simulated galaxy clusters*. *Astrophys. J.* **607** (2004), 164–174.
- [119] R. Kippenhahn, *Der Stern, von dem wir leben*. Deutschen Taschenbuch Verlag, München, 1993.
- [120] E. Kokubo, S. Ida, J. Makino, *Evolution of a circumterrestrial disk and formation of a single Moon*. *Icarus* **148** (2000), 419–436.
- [121] V. Kolář, J. Šístek, F. Cirak, P. Moses, *Average corotation of line segments near a point and vertex identification*. *AIAA J.* **51** (2013), 2678–2694.
- [122] C. T. Kowal, *Absolute magnitudes of supernovae*. *Astronom. J.* **73** (1968), 1021–1024.
- [123] C. T. Kowal, S. Drake, *Galileo's observations of Neptune*. *Nature* **287** (1980), 311–313.
- [124] G. V. Kraniotis, S. B. Whitehouse, *Compact calculation of the perihelion precession of Mercury in general relativity, the cosmological constant and Jacobi's inversion problem*. *Classical Quant. Grav.* **20** (2003), 4817–4835.
- [125] G. A. Krasinski, V. A. Brumberg, *Secular increase of astronomical unit from analysis of the major planet motions, and its interpretation*. *Celest. Mech. Dyn. Astr.* **90** (2004), 267–288.
- [126] F. Křížek, M. Křížek, *Vera Rubin and rotational curves of spiral galaxies* (in Czech). *Pokroky Mat. Fyz. Astronom.* **59** (2014), 223–236.
- [127] F. Křížek, M. Křížek, J. Šolc, *How massive is the black hole at the centre of our Galaxy?* *Obzory mat. fyz. inf.* **36** (2007), 43–51.
- [128] M. Křížek, *Numerical experience with the three-body problem*. *J. Comput. Appl. Math.* **63** (1995), 403–409.

- [129] M. Křížek, *Numerical experience with the finite speed of gravitational interaction*. Math. Comput. Simulation **50** (1999), 237–245.
- [130] M. Křížek, *The rôle of the protractor in understanding the universe*. Obzory mat. fyz. inf. **37** (2008), 36–47.
- [131] M. Křížek, *Does a gravitational aberration contribute to the accelerated expansion of the Universe?* Comm. Comput. Phys. **5** (2009), 1030–1044.
- [132] M. Křížek, *Numerical simulation and the origin of dark energy*. Proc. Conf. Computational Analysis and Optimization (eds. S. Repin, T. Tiihonen, T. Tuovinen), Univ. of Jyväskylä, 2011, 25–31.
- [133] M. Křížek, *Dark energy and the anthropic principle*. New Astronomy **17** (2012), 1–7.
- [134] M. Křížek, *Antigravity* (in Czech). Prague, 2014, 1–265.
- [135] M. Křížek, *Zdeněk Kopal – Numerical Analyst*. Astron. Soc. Pacific Conf. Ser. **496** (2015), 19–31.
- [136] M. Křížek, J. Brandts, *Manifestations of dark energy in the dynamics of the Solar system*. S264 – Solar and Stellar Variability: Impact on Earth and Planets (eds. A. G. Kosovichev, A. H. Andrei, and J.-P. Rozelot), Proc. of the IAU XXVIIth General Assembly in Rio de Janeiro, August 2009, Cambridge Univ. Press, Cambridge, 2010, 410–412.
- [137] M. Křížek, J. Brandts, L. Somer, *Is gravitational aberration responsible for the origin of dark energy?* Dark Energy: Theory, Implications and Roles in Cosmology (eds. C. A. Del Valle and D. F. Longoria), Nova Sci. Publishers, New York, 2012, 29–57.
- [138] M. Křížek, F. Křížek, L. Somer, *Which effects of galactic clusters can reduce the amount of dark matter*. Bulg. Astronom. J. **21** (2014), 1–23.
- [139] M. Křížek, F. Křížek, L. Somer, *Dark matter and rotational curves of galaxies*. Submitted, preprint Inst. of Math., Prague, 2014, 1–12.
- [140] M. Křížek, P. Křížek, *Why has nature invented three stop codons of DNA and only one start codon?* J. Theor. Biol. **304** (2012), 183–187.
- [141] M. Křížek, P. Neittaanmäki, *Finite element approximation of variational problems and applications*. Pitman Monographs and Surveys in Pure and Applied Mathematics vol. 50, Longman Scientific & Technical, Harlow; copublished with John Wiley, New York, 1990.

- [142] M. Křížek, P. Neittaanmäki *Mathematical and Numerical Modelling in Electrical Engineering: Theory and Applications*. Kluwer Academic Publishers, Dordrecht, 1996.
- [143] M. Křížek, J. Pradlová, *On the nonexistence of a Lobachevsky geometry model of an isotropic and homogeneous universe*. Math. Comput. Simulation **61** (2003), 525–535.
- [144] M. Křížek, H.-G. Roos, W. Chen, *Two-sided bounds of the discretization error for finite elements*. ESAIM Math. Model. Numer. Anal. **45** (2011), 915–924.
- [145] M. Křížek, L. Somer, *Antigravity – its manifestations and origin*. Internat. J. Astron. Astrophys. **3** (2013), 227–235.
- [146] M. Křížek, L. Somer, *Manifestations of dark energy in the Solar system*. Grav. Cosmol. **21** (2015), 58–71.
- [147] M. Křížek, L. Somer, *A critique of the standard cosmological model*. Neural Netw. World **24** (2014), 435–461.
- [148] M. Křížek, L. Somer, *Excessive extrapolations in cosmology*. Submitted, preprint Inst. of Math., Prague, 2014, 1–14.
- [149] P. Kroupa, *Star-cluster formation and evolution*. In: Proc. IAU S237, Cambridge Univ. Press, Cambridge, 2007, 230–237.
- [150] P. Kroupa, *Local-group tests of dark-matter concordance cosmology*. ArXiv: 1006.1647v3, 2010, 1–26.
- [151] P. Kroupa, *The dark matter crisis: Falsification of the current standard model of cosmology*. Publ. Astron. Soc. Australia **29** (2012), 395–433.
- [152] J.M. Kubo, A. Stebbins et al., *The mass of the Coma cluster from weak lensing in the Sloan Digital Sky Survey*. Astrophys. J. **671** (2007), 1466–1470.
- [153] L. R. Kump, J. F. Kastings, R. G. Crane, *The Earth system*. Prentice Hall, New Jersey, 1999.
- [154] L. Landau, *On the theory of stars*. Phys. Zeitschrift der Sowjetunion **1** (1932), 285–288.
- [155] K. K. Lang, *Cambridge encyclopedia of the Sun*. Cambridge Univ. Press, Cambridge, 2001.

- [156] K. R. Lang, *Astrophysical formulae, vol. II*. Springer, Berlin, 2006.
- [157] P. S. Laplace, *A treatise in celestial mechanics, vol. IV, book X*. Translated by N. Bowditch, Chelsea, New York, 1966.
- [158] C. Ledoux et al., *Identification of active galaxies behind the Coma cluster of galaxies*. *Astron. Astrophys. Suppl. Ser.* **138** (1999), 109–117.
- [159] G. E. Lemaître, *Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extragalactiques*. *Ann. Soc. Sci. de Bruxelles* (1927), April, 49–59.
- [160] G. E. Lemaître, *Evolution of the expanding universe*. *Proc. Nat. Acad. Sci.* **20** (1934), 12–17.
- [161] J. Lense, H. Thirring, *Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie*. *Phys. Zeitschrift* **19** (1918), 156–163.
- [162] A. P. Lightman, W. H. Press, R. H. Price, S. A. Teukolsky, *Problem book in relativity and gravitation*. Princeton Univ. Press, 1975.
- [163] C. H. Lineweaver, D. Schwartzman, *Cosmic thermobiology*. In: *Origins* (ed. J. Seckbach), Kluwer, Dordrecht, 2003, 233–248.
- [164] F. Link, *Sur les conséquences photométriques de la déviation d’Einstein*. *Comptes Rendus Acad. Sci.* **202** (1936), 917–919.
- [165] E. L. Lokas, G. A. Mamon, *Dark matter distribution in the Coma cluster from galaxy kinematics: Breaking the mass-anisotropy degeneracy*. *Mon. Not. R. Astron. Soc.* **343** (2003), 401–412.
- [166] O. L. Mangasarian, J.-S. Pang, *Computational optimization: a tribute to Olvi Mangasarian, vol. 1*. Springer, 1999.
- [167] E. Maor, *Venus in transit*. Princeton Univ. Press, Princeton, 2000.
- [168] C. Marchal, *The three-body problem*. Elsevier, Amsterdam, 1991.
- [169] V. Marra, L. Amendola, I. Sawicki, W. Valkenburg, *Cosmic variance and the measurement of the local Hubble parameter*. *Phys. Rev. Lett.* **110** (2013), 241305.
- [170] G. E. Marsh, C. Nissim-Sabat, *Comment on “The speed of gravity”*. *Phys. Lett. A* **262** (1999), 257–260.

- [171] B. Mashhoon et al., *Relativistic effects in the motion of the Moon*. Lect. Notes Phys. **562** (2001), 310–316.
- [172] B. Mashhoon, N. Mobed, D. Singh, *Tidal dynamics in cosmological space-times*. Classical Quant. Grav. **24** (2007), 5031–5046.
- [173] S.S. McGaugh, *Milky Way mass models and MOND*. Astrophys. J. **683** (2008), 137–148.
- [174] R. McLachlan, *A gallery of constant-negative-curvature surfaces*. Math. Intelligencer **16** (1994), 31–37.
- [175] C.G. McVittie, *The mass-particle in expanding universe*. Mon. Not. R. Astron. Soc. **93** (1933), 325–339.
- [176] P. Mészáros, A. Mészáros, *The brightness distribution of bursting sources in relativistic cosmologies*. Astrophys. J. **449** (1995), 9–17.
- [177] Z. Mikulášek, J. Krtička, *Foundations of physics of stars* (in Czech). Masaryk Univ., Brno, 2005.
- [178] M. Milgrom *A modification of Newtonian dynamics as a possible alternative to the hidden mass hypothesis; Implications for galaxies; Implications of galaxy systems*. Astrophys. J. **270** (1983), 365–370; 371–383; 384–389.
- [179] E. A. Milne, *Relativity, gravitation and world structure*. Clarendon Press, Oxford, 1935.
- [180] E. A. Milne, *A Newtonian expanding universe*. General Relativity and Gravitation. Vol. 32, no. 9, Springer, Berlin, 2000.
- [181] I. F. Mirabel, L. F. Rodríguez, *A superluminal source in the Galaxy*. Nature **371** (1994), 46–48.
- [182] I. F. Mirabel, L. F. Rodríguez, *Superluminal sources in the Galaxy*. Vistas in Astronomy **41** (1997), 15–16.
- [183] I. F. Mirabel, L. F. Rodríguez, *Microquasars in our Galaxy*. Nature **392** (1998), 673–676.
- [184] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation*. 20th edition, W. H. Freeman, New York, 1997.

- [185] B. Mobasher et al., *A photometric and spectroscopic study of dwarf and giant galaxies in the Coma cluster. IV. The luminosity function*. *Astrophys. J.* **587** (2003), 605–618.
- [186] S. Mollerach, E. Roulet, *Gravitational lensing and microlensing*. World Scientific, Singapore, 2002.
- [187] C. Moni Bidin, G. Carraro, R. A. Méndez, R. Smith, *Kinematical and chemical vertical structure of the Galactic thick disk, II. A lack of dark matter in the solar neighborhood*. *Astrophys. J.* 751 (2012), 30, 14 pp.
- [188] R. Montgomery, *A new solution to the three body problem*. *Notices Amer. Math. Soc.* **5** (2001), 471–481.
- [189] T. A. Morley, *A catalogue of ground-based astrometric observations of the Martian satellites 1877–1982*. *Astron. Astrophys. Suppl. Ser.* **77** (1989), 209–226.
- [190] J. V. Narlikar, N. C. Rana, *Newtonian N -body calculations of the advance of Mercury's perihelion*. *Mon. Not. R. Astron. Soc.* **213** (1985), 657–663.
- [191] D. M. Neumann, D. H. Lumb, G. W. Pratt, U. G. Briel, *The dynamical state of the Coma cluster with XMN-Newton*. *Astron. Astrophys.* **400** (2003), 811–821.
- [192] K. F. Nicholson, *Galactic mass distribution without dark matter or modified Newtonian mechanics*. ArXiv: astro-ph/0309762v2, 2007, 1–16.
- [193] J. T. Nielsen, A. Guffanti, S. Sarkar, *Marginal evidence for cosmic acceleration from type Ia supernovae*. Dostupn z arXiv: 1506.01354v2, 2015, 1–5.
- [194] P. D. Noerdlinger, *Solar mass loss, the astronomical unit, and the scale of the Solar system*. ArXiv: 0801.3807, 2008.
- [195] O. Novotný, *Motions, gravity field and figure of the Earth*. Lecture Notes, Univ. Federal da Bahia, Brazil, 1998.
- [196] H. Nussbaumer, L. Bieri, *Discovering the expanding universe*. Cambridge Univ. Press, Cambridge, 2009.
- [197] J. Oberst et al., *Astrometric observations of Phobos and Deimos with the SRC on Mars Express*. *Astron. Astrophys.* **447** (2006), 1145–1151.

- [198] J. Oort, *The force exerted by the stellar system in the direction perpendicular to the Galatic plane and some related problems*. Bull. Astronom. Inst. Neth. **6** (1932), 249–287.
- [199] R. Panek, *The 4 percent universe: dark matter, dark energy, and the race to discover the rest of reality*. Houghton Mifflin Harcourt, 2012.
- [200] G. Pannella, *Paleontological evidence on the Earth's rotation history since early precambrian*. Astrophys. Space Sci. **16** (1972), 212–237.
- [201] T. J. Pearson, J. A. Zensus (eds.), *Superluminal radio sources*. Cambridge Univ. Press, 1987.
- [202] P. J. E. Peebles, *Principles of physical cosmology*. Princeton Univ. Press, New Jersey, 1993.
- [203] W. R. Peltier, *History of Earth rotation*. In: Treatise on Geophysics (ed. G. Schubert), vol. 9, Evolution of the Earth (vol. ed. D. Stevenson), Elsevier, Amsterdam, 2007.
- [204] R. Penrose, *The road to reality*. Vintage Books, London, 2005.
- [205] S. Perlmutter, *Supernovae, dark energy, and the accelerating universe*. Physics Today **56** (2003), April, 53–60.
- [206] S. Perlmutter, G. Aldering et al., *Measurements of Ω and Λ from 42 high-redshift supernovae*. Astrophys. J. **517** (1999), 565–586.
- [207] S. Perlmutter, S. Gabi et al., *Measurements of the cosmological parameters Ω and Λ from the first seven supernovae at $z \geq 0.35$* . Astrophys. J. **483** (1997), 565–581.
- [208] J. T. Perron et al., *Evidence for an ancient martian ocean in the topography of deformed shorelines*. Nature **447** (2007), 840–843.
- [209] P. Petersen, *Riemannian geometry*, 2nd edition, Springer, 2006.
- [210] M. Pick, J. Pícha, V. Vyskočil, *Theory of the Earth's gravity field*. Academia, Praha, 1973.
- [211] S. V. Pilipenko, *Paper-and-pencil cosmological calculator*. Preprint, arXiv: 1303.5961v1, 2013, 1–4.
- [212] Planck Collaboration, *Planck 2013 results, I. Overview of products and scientific results*. Astron. Astrophys. **571** (2014), A1, 48 pp.

- [213] Planck Collaboration, *Planck 2013 results, XVI. Cosmological parameters*. *Astron. Astrophys.* **571** (2014), A16, 66 pp.
- [214] Planck Collaboration, *Planck 2013 results, XVIII. Gravitational lensing-infrared background correlation*. *Astron. Astrophys.* **571** (2014), A18, 24 pp.
- [215] Planck Collaboration, *Planck 2013 results, XX. Cosmology from Sunyaev–Zeldovich cluster counts*. *Astron. Astrophys.* **571** (2014), A20, 20 pp.
- [216] H. Poincaré, *Sur le problème des trois corps et les équations de la dynamique*. *Acta Math.* **13** (1890), 1–270.
- [217] H. Poincaré, *La théorie de Lorentz et le principe de réaction*. *Arch. Néerland. Sci. Exactes et Naturelles* **5** (1900), 252–278.
- [218] H. Poincaré, *Sur la dynamique de l'électron*. *C. R. Acad. Sci. Paris* **140** (1905), 1504–1508.
- [219] R. W. Porcas, *Summary of known superluminal sources*. In: *Superluminal radio sources*, T. J. Pearson, J. A. Zensus (eds.), Cambridge Univ. Press, 1987, 12–25.
- [220] N. C. Rana, *An investigation of the motions of the node and perihelion of Mercury*. *Astron. Astrophys.* **181** (1987), 195–202.
- [221] D. Rapetti et al., *The observed growth of massive galaxy clusters — III. Testing general relativity on cosmological scales*. *Mon. Not. R. Astron. Soc.* **406** (2010), 1796–1804.
- [222] B. Ratra, M. S. Vogeley, *Resource letter: BE-1: The beginning and evolution of the Universe*. ArXiv: 0706.1565v1, 2007, 1–95.
- [223] D. Rawlins, *The great unexplained residual in orbit of Neptune*. *Astronom. J.* **75** (1970), 856–857.
- [224] K. Rektorys, *Survey of applicable mathematics I*. Kluwer Acad. Publ., Dordrecht, 1994.
- [225] K. Rektorys, *Survey of applicable mathematics II*. Kluwer Acad. Publ., Dordrecht, 1994.
- [226] A. G. Riess, A. V. Filippenko, . . . , B. Schmidt et al., *Observational evidence from supernovae for an accelerating universe and a cosmological constant*. *Astronom. J.* **116** (1998), 1009–1038.

- [227] A. G. Riess, P. E. Nugent, . . . , B. Schmidt et al., *The farthest known supernova: Support for an accelerating universe and a glimpse of the epoch of deceleration*. *Astrophys. J.* **560** (2001), 49–71.
- [228] A. G. Riess, L.-G. Strolger et al., *Type Ia supernova discoveries at $z > 1$ from the Hubble space telescope: Evidence for past deceleration and constraints on dark energy evolution*. *Astrophys. J.* **607** (2004), 665–687.
- [229] A. G. Riess, L.-G. Strolger et al., *New Hubble space telescope discoveries of type Ia supernovae at $z \geq 1$: Narrowing constraints on the early behavior of dark energy*. *Astrophys. J.* **659** (2007), 98–121.
- [230] K. Rines et al., *Infrared mass-to-light profile throughout the infall region of the Coma cluster*. *Astrophys. J.* **561** (2001), L41–L44.
- [231] H. P. Robertson, *On the foundation of relativistic cosmology*. *Proc. Nat. Acad. Sci.* **15** (1929), 822–829.
- [232] C. Ron, J. Vondrák, *Expansion of annual aberration into trigonometric series*. *Bull. Astron. Inst. Czechosl.* **37** (1986), 96–103.
- [233] P. Rosenblatt, *The origin of the Martian moons revisited*. *Astronom. Astrophys. Rev.* **19** (2011), #44.
- [234] N. T. Roseveare, *Mercury’s perihelion from Le Verrier to Einstein*. Clarendon Press, Oxford, 1982.
- [235] R. Roškar et al., *Riding the spiral waves: implications of stellar migration for the properties of galactic disks*. *Astrophys. J.* **684** (2008), L79–L82.
- [236] M. Rowan-Robinson, *Do type Ia supernovae prove $\Lambda > 0$?* *Mon. Not. R. Astron. Soc.* **332** (2002), 352–360.
- [237] V. C. Rubin, *Dark matter in spiral galaxies*. *Scientific Amer.* **248** (1983), 88–101.
- [238] V. C. Rubin, *A brief history of dark matter*. *The Dark Universe: Matter, Energy, and Gravity* (ed. M. Livio), Cambridge Univ. Press, Cambridge, 2003, 1–13.
- [239] V. C. Rubin et al., *Kinematic studies of early-type stars, I. Photometric survey, space motions, and comparison with radio observations*. *Astrophys. J.* **67** (1962), 491–531.

- [240] V. C. Rubin, W. K. Ford, N. Thonnard, *Rotational properties of 21 Sc galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4$ kpc) to UGC 2885 ($R = 122$ kpc).* *Astrophys. J.* **238** (1980), 471–487.
- [241] G. Rudnick et al., *Measuring the average evolution of luminous galaxies at $z < 3$: The rest-frame optical luminosity density, spectral energy distribution, and stellar mass density.* *Astrophys. J.* **650** (2006), 624–643.
- [242] D. G. Saari, Z. Xia, *Off to infinity in finite time.* *Notices of Amer. Math. Soc.* **42** (1995), 538–546.
- [243] I. J. Sackmann, A. I. Boothroyd, K. E. Kraemer, *Our Sun. III. Present and future.* *Astrophys. J.* **418** (1993), 457–468.
- [244] C. Sagan, *Reducing greenhouses and the temperature history of Earth and Mars.* *Nature* **269**, 224–226 (1977).
- [245] C. Sagan, G. Mullen, *Earth and Mars: Evolution of atmospheres and surface temperatures.* *Science* **177**, 52–56 (1972).
- [246] C. Sagan, O. B. Toon, P. J. Gierasch, *Climatic change on Mars.* *Science* **181**, 1045–1049 (1973).
- [247] S. S. Said, F. R. Stephenson, *Solar and lunar eclipse measurements by medieval Muslim astronomers.* *J. Hist. Astronom.* **27, 28** (1996/97), 259–273, 29–48.
- [248] R. H. Sanders, *The dark energy problem – a historical perspective.* Cambridge Univ. Press, Cambridge, 2010.
- [249] R. H. Sanders, S. S. McGaugh, *Modified Newtonian dynamics as an alternative to dark matter.* *Ann. Rev. Astronom. Astrophys.* **40** (2002), 263–317.
- [250] R. Schödel, T. Ott, R. Genzel et al., *A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way.* *Nature* **419** (2002), 694–696.
- [251] B. Schwarzschild, *Discoverers of the Hubble expansion’s acceleration share Nobel physics prize.* *Physics Today* **64** (2011), Dec., 14–17.
- [252] K. Schwarzschild, *Über das zulässige Krümmungsmaaß des Raumes.* *Vierteljahrsschrift der Astronomischen Gesellschaft* **35** (1900), 337–347; English translation: *On the permissible numerical value of the curvature of space.* *Abraham Zelmanov J.* **1** (2008), 64–73.

- [253] D. Sellers, *The transit of Venus: the quest to find the true distance of the Sun*. Megavelda Press, 2001.
- [254] D.N. Sergel et al., *Three-years Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for cosmology*. *Astrophys. J.* **170** (2007), 377–408.
- [255] I.I. Shapiro, *Fourth test of general relativity*. *Phys. Rev. Lett.* **13** (1964), 789–791.
- [256] G. A. Shields, *A brief history of Active Galactic Nuclei*. *Publ. Astronom. Soc. Pacific* **111** (1999), 661–678.
- [257] S. Sikora, Ł. Bratek, J. Jałocha, M. Kutschera, *Gravitational microlensing as a test of a finite-width disk model of the Galaxy*. *Astron. Astrophys.* **546** (2012), A126, 9 pp.
- [258] A. J. Simoson *Periodicity domains and the transit of Venus*. *Amer. Math. Monthly* **121** (2014), 283–298.
- [259] P. Sisterna, H. Vucetich, *Time variation of fundamental constants: Bounds from geophysical and astronomical data*. *Phys. Rev. D* **41** (1990), 1034–1046.
- [260] V. M. Slipher, *The radial velocity of the Andromeda nebula*. *Lowell Observatory Bull.* **1** (1913), 56–57.
- [261] V. M. Slipher, *Spectrographic observations of nebulae*. *Amer. Astronom. Soc., Popular Astronomy* **23** (1915), 21–24.
- [262] S. Smith, *The mass of the Virgo cluster*. *Astrophys. J.* **83** (1936), 23–30.
- [263] J. Southworth et al., *Physical properties of the 0.94-day period transiting planetary system WASP-18, 2009*. *Astrophys. J.* **707** (2009), 167–172.
- [264] E. M. Standish, *Planet X: no dynamical evidence in the optical observations*. *Astronom. J.* **105** (1993), 2000–2006.
- [265] J. M. Steele, *Solar eclipse times predicted by the Babylonians*. *J. Hist. Astron.* **28** (1997), 133–139.
- [266] J. M. Steele, F. R. Stephenson, *The accuracy of eclipse times measured by the Babylonians*. *J. Hist. Astron.* **28** (1997), 337–345.
- [267] H. Stephani, *General relativity. An introduction to the theory of the gravitational field*. 2nd edition, Cambridge Univ. Press, Cambridge, 1990.

- [268] B. Stephenson, *Kepler's physical astronomy*. Springer-Verlag, New York, 1987.
- [269] F. R. Stephenson, *Historical eclipses and Earth's rotation*. *Astronomy & Geophysics* **44** (2003), 22–27.
- [270] F. R. Stephenson, L. J. Fatoohi, *Thales's prediction of a Solar eclipse*. *J. Hist. Astron.* **28** (1997), 279–282.
- [271] F. R. Stephenson, L. V. Morrison, G. J. Whitrow, *Long-term changes in the rotation of the Earth: 700 B.C. to A.D. 1980 [and Discussion]*. *Phil. Trans. Roy. Soc. Lond. A* **313** (1984), 47–70.
- [272] G. Strömberg, *Analysis of radial velocities of globular clusters and non-galactic nebulae*. *Astrophys. J.* **LXI** (1925), 353–362.
- [273] R. A. Sunyaev, Y. B. Zeldovich, *The spectrum of primordial radiation, its distortions and their significance*. *Comments. Astrophys. Space Phys.* **2** (1970), 66–74.
- [274] A. M. Swinbank et al., *Intense star formation within resolved compact regions in a galaxy at $z = 2.3$* . *Nature* **464** (2010), 733–736.
- [275] J. H. Taylor, J. M. Weisberg, *A new test of general relativity — Gravitational radiation and the binary pulsar PSR 1913+16*. *Astrophys. J.* **253** (1982), 908–920.
- [276] J. A. Thorpe, *Elementary topics in differential geometry*. Springer, New York, Berlin, 1979.
- [277] B. Tinsley, *Accelerating Universe revisited*. *Nature* **273** (1978), 208–211.
- [278] I. Trujillo, *Origin and fate of the most massive galaxy*. *Highlights of Spanish Astrophysics VI, Proc. of the IX Sci. Meeting of the Spanish Astronom. Soc., Madrid* (ed. M. R. Zapatero et al.), 2010, 120–130.
- [279] I. Trujillo et al., *Strong size evolution of the most massive galaxies since $z \sim 2$* . *Mon. Not. R. Astron. Soc.* **382** (2007), 109–120.
- [280] K. Tsiganis, R. Gomes, A. Morbidelli, H. F. Levison, *Origin of the orbital architecture of the giant planets of the Solar System*. *Nature* **435** (2005), 459–461.

- [281] S. G. Turyshev et al., *Advancing tests of relativistic gravity via laser ranging to Phobos*. *Exp. Astron.* **28** (2010), 209–249.
- [282] A. V. Tutukov, A. V. Fedorova, *The origin of intergalactic stars in galaxy clusters*. *Astron. Reports* **55** (2011), 383–391.
- [283] J. van de Sande et al., *The stellar velocity dispersion of a compact massive galaxy at $z = 1.80$ using X-shooter confirmation of the evolution in the mass-size and mass-dispersion relations*. *Astrophys. J. Lett.* **736** (2011), L9.
- [284] T. C. van Flandern, *A determination of the rate of change of g* . *Mon. Not. R. Astron. Soc.* **170** (1975), 333–342.
- [285] T. C. van Flandern, *The speed of gravity – what the experiments say*. *Phys. Lett. A* **250** (1998), 1–11.
- [286] A. A. Vankov, *General relativity problem of Mercury’s perihelion advance revisited*. *ArXiv: 1008.1811v1*, 2010, 1–46.
- [287] F. Verbund, *The Earth and Moon: from Halley to lunar ranking and shells*. Preprint, Utrecht Univ., 2002, 1–10.
- [288] L. Verde, R. Jimenez, S. Feeney, *The importance of local measurements for cosmology*. *ArXiv:1303.5341v2*, 2013.
- [289] E. B. Vinberg (ed.), *Geometry, I. Geometry of spaces of constant curvature, II. Discrete groups of motions of spaces of constant curvature*. Springer, Berlin, 1991.
- [290] G. M. Voit, *Tracing cosmic evolution with clusters of galaxies*. *Rev. Mod. Phys.* **77** (2005), 207–258.
- [291] R. M. Wald, *General relativity*. Univ. of Chicago Press, 1982.
- [292] Walker A. G.: *On Milne’s theory of world-structure*. *Proc. London Math. Soc.* 42 (1937), 90–127.
- [293] S. Weinberg, *Gravitation and cosmology: Principles and applications of the general theory of relativity*. John Wiley, New York, London, 1972.
- [294] S. Weinberg, *The first three minutes*. Basic Books, New York, 1977.
- [295] S. Weinberg, *Cosmology*. Oxford Univ. Press, 2008.
- [296] J. W. Wells, *Coral growth and geochronometry*. *Nature* **197** (1963), 948–950.

- [297] G. E. Williams, *Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit*. Rev. Geophys. **38** (2000), 37–60.
- [298] R. Wojtak, S. H. Hansen, J. Hjorth, *Gravitational redshift of galaxies in clusters as predicted by general relativity*. Nature **477** (2011), 567–569.
- [299] J. A. Wolf, *Spaces of constant curvature*. Amer. Math. Soc., Sixth Edition, 1984.
- [300] R. Wordsworth, R. Pierrehumbert, *Hydrogen-nitrogen greenhouse warming in Earth's early atmosphere*. Science **339** (2013), 64–67.
- [301] C. F. Yoder, J. G. Williams et al., *Secular variation of Earth's gravitational harmonic J_2 coefficient from Lageos and nontidal acceleration of Earth rotation*. Nature **303** (1983), 757–762.
- [302] W. J. Zhang, Z. B. Li, Y. Lei, *Experimental measurements of growth patterns on fossil corals: Secular variation in ancient Earth-Sun distance*. Chinese Sci. Bull. **55** (2010), 4010–4017.
- [303] F. Zwicky, *On the red shift of spectral lines through interstellar space*. Proc. Nat. Acad. Sci. **15** (1929), 773–779.
- [304] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*. Helv. Phys. Acta **6** (1933), 110–127.
- [305] F. Zwicky, *Remarks on the redshifts from nebulae*. Phys. Rev. **48** (1935), 802–806.
- [306] F. Zwicky, *On the masses of nebulae and of clusters of nebulae*. Astrophys. J. **86** (1937), 217–246.
- [307] http://en.wikipedia.org/wiki/NGC_4874



Name Index

- Adams J. C. 46
Adams W. 34
Anaximander 14
Archimedes 17, 41, 191
Aristarchus 16–19
Aristotle 17, 143
- Baade W. 91, 151
Barrow J. 211
Bessel F. W. 27, 34
Blanuša D. 278
Bolyai J. 263
Bouwers R. J. 240
Bradley J. 28
Brahe T. 7, 8, 15, 27
Brander D. 278
Brumberg V. A. 206
Bruno G. 145, 262
Bürgi J. 10
Burša M. 232
- Carlip S. 252
Carrera M. 207
Carter B. 211
Cassini G. D. 21
Cavendish H. 45, 46
Chandrasekhar S. 153
Charlier C. V. L. 150
Che 83
Clark A. G. 34, 46
Clarke A. C. 53
Clausius R. 95
Clowe D. 137
Cook J. 23
Cooper V. 122
Cooperstock F. I. 207
Copernicus N. 5, 20, 21, 211
Curtis H. 147
- d'Alembert J. 69
da Vinci L. 55
de Sitter W. 157
- Doppler C. 28
Drake F. 221
Dumin Y. V. 188, 199
- Eddington A. 145, 251
Einstein A. 32, 34, 81, 100, 150, 157, 239, 257, 284, 293
Eratosthenes 18, 19
Escher M. C. 274
Euler L. 41, 69
- Fedorova A. V. 298
Fermi E. 211, 220
Feynmann R. 307
Flandro G. A. 74
Friedmann A. A. 148, 157, 161, 282–283, 291, 304
- Galileo G. 5, 234
Galle J. G. 20, 46
Gamow G. 122
Gauss C. F. 263
Giulini D. 207
Gödel K. 122, 151
Green C. 23
- Hall A. 49
Halley E. 23, 56
Hamilton W. R. 69
Hamuy M. 156
Heinrich W. W. 69
Hell M. 23
Herschel W. F. 20, 57, 239
Hilbert D. 272, 273
Hipparchos 5
Hohmann W. 53
Hoyle F. 148
Hubble E. P. 92, 98, 147, 157, 161
Humason M. L. 92, 98, 148
Huygens C. 28, 280
- Jacobi C. G. J. 69
Jacobson R. A. 232

Kepler J. 5–11, 15, 21, 23, 36, 41, 47, 48, 69
 Kepler L. 7
 Klein F. 263
 Kowal C. T. 151
 Krasinsky G. A. 206
 Kroupa P. 136, 245, 299

 Lagrange J. L. 69
 Lalande M. L. 234
 Landau L. D. 91, 262
 Laplace P. 69, 257
 Leavitt H. S. 147
 Lemaître G. H. É. 148, 150, 161, 162, 207, 283
 Leško J. 82
 Le Verrier U. J. J. 46
 Lie S. 263
 Lobachevskii N. I. 3, 263
 Lomonosov M. V. 81
 Lowell P. 234
 Luhman K. 129

 MacLaurin C. 41
 Mashhoon R. 207
 Mather J. C. 143
 McVittie C. G. 118
 Messier C. 239
 Michelson A. A. 224
 Milne E. A. 150
 Minovitch M. A. 73

 Napier J. 10
 Nechvíle V. 69
 Newton I. 6, 42, 65, 69, 262
 Nobel A. 144
 Novotný O. 188

 Oort J. 91

 Panek R. 260
 Penrose R. 14
 Perelman G. J. 271
 Perlmutter S. 143, 144, 155, 159
 Petr K. 69

 Pick G. 157
 Poincaré H. 70, 75, 257, 271, 273
 Potočník H. 53
 Prusinger S. B. 222
 Ptolemy K. 5

 Richer J. 22
 Riemann B. 263
 Riess A. G. 143, 144, 158
 Rømer R. 28
 Röntgen W. C. 145
 Rubin R. 122
 Rubin V. 122–124, 128, 134, 135, 298
 Savary F. 57
 Schmidt B. 143, 144, 153, 155, 159
 Schopenhauer A. 249
 Schwarzschild K. 145, 161, 262, 265
 Si 83
 Slipher V. M. 95, 146, 148, 161
 Smith S. 90
 Smoot G. F. 143
 Strömberg G. 148, 150

 Thales 83
 Tinsley B. 160, 161
 Tipler F. 211
 Tombaugh C. 234
 Trujillo I. 240
 Tutukov A. V. 298

 van den Bergh G. 85
 van Flandern T. C. 181, 257
 Vankov A. A. 293
 Verbunt F. 181
 Vrba A. 307
 Vřčková F. 91

 Weinberg S. iii, 116, 150
 Wells J. W. 196
 Wheeler J. A. xii
 Williams G. E. 180

 Zhang W. 195, 196, 199
 Zwicky F. 47, 90–93, 95, 99–108, 112, 119, 151, 297

Subject Index

- aberration
 - gravitational 163, 165, 252, 260
 - of light 27, 81, 87, 89
- acceleration gravitational 63, 64
- age
 - Hubble 156, 160, 296
 - of the universe 285, 295, 302
- albedo 173, 174, 193
- Alpha Centauri 164
- Altair 30, 31
- angle
 - aberration 28, 29, 87, 88, 252
 - aberration of light 255
 - bending 117
 - of gravitational aberration 250, 253, 258
 - parallactic 23
- anomaly
 - eccentric 36–38
 - mean 40
 - true 37, 38
- antigravity 141, 165, 181, 182, 190, 192, 203, 209–211, 219, 220, 224, 226, 228–230, 239, 246, 247, 249, 261, 303, 305
- Apophis 77
- argument of perihelion 40, 41
- astrolabe 14
- axis
 - semimajor 9, 10, 20, 22, 26, 27, 37, 40, 47, 50, 54, 56, 58, 59, 63, 64, 196, 197, 293
 - semiminor 9, 11, 37, 54, 60
- barycenter 192, 201
- Big Bang 118, 136, 146, 148, 270, 271, 281, 288, 295, 296, 305, 313, 315–317
- blueshift 95, 150
- bound
 - lower 41, 199, 301, 315
 - two-sided 41
- catastrophe
 - gravithermal 246
 - tidal 181
- Cepheid 147, 149, 304
- cluster 97, 99
 - Coma 90, 95, 100–104, 297, 298
 - galaxy 90, 92, 96, 239, 298
 - globular 104, 145, 240, 246
 - Virgo 90
- CMB 33, 138, 139, 268, 270, 299, 317
- comet 72, 203
 - Halley's 56, 57
 - Shoemaker-Levy nine 72
- cone
 - light 269, 277, 306
- constant
 - cosmological 157, 159, 252, 253, 283, 284, 286, 299, 302
 - fundamental physical 211, 212, 253, 291
 - gravitational 43, 50, 63, 66, 93, 157, 189, 197, 212, 213, 253, 283
 - Hubble 95, 118, 150, 166, 175, 181, 187, 214, 227, 284, 301, 304, 308
 - Newton–Cavendish 46
 - of fine structure 213
 - Planck's 205, 213
 - solar 167, 176, 193, 194
- coordinates spherical 266
- cosmology 90, 116, 143, 268, 285, 294, 305, 307
- curvature 33, 154, 263, 279, 288
 - Gaussian 272, 280, 281, 292
 - global 33
 - normalized 157, 283, 284, 301
 - of spacetime 293
 - of the universe 158
 - sectional 302
 - spatial 157, 283, 288
- curve rotational 122, 123, 127, 135

declination 30, 107
 Deneb 30, 31
 density
 critical 285, 286
 mean 45, 47, 55
 of baryonic matter 157, 285, 286
 of dark energy 157, 285, 287
 of dark matter 157, 285, 286
 of spatial curvature 157
 diagram
 Hertzsprung–Russell 167, 191, 203
 O–C 46
 DNA 193, 221, 222
 dwarf
 black 129
 brown 129, 298
 infrared 129
 orange 128
 red 128, 298
 red-brown 129
 white 34, 46, 128, 152, 167
 eccentricity 9, 11, 37, 40, 56, 59, 67, 84,
 178, 236, 293
 linear 9, 54
 numerical 9, 54
 eclipse 17–19, 24, 25, 32, 81–87, 89,
 184, 190
 ecliptic 14, 23, 30, 40, 99, 202
 ecosphere 193, 194, 216
 effect
 Doppler 58, 61, 99, 151
 greenhouse 170, 174, 216
 inverse Compton scattering 140
 Lense–Thirring precession 34, 293
 Pioneer 238
 relativistic 105, 115, 116, 189
 selection 193
 selffensing 107, 117
 side 258, 303
 Sunyaev–Zeldovich 140
 Yarkovsky 189, 206
 YORP 206
 ellipse 11–13, 36, 37, 41, 67
 aberration 28
 parallactic 26
 energy
 dark 91, 104, 119, 138, 140, 141, 144,
 155, 158, 162, 163, 166, 208, 213,
 215, 220, 249, 285, 295, 299, 300,
 303, 304
 kinetic 52, 93, 94, 100, 210, 247,
 251, 261
 mechanical 94
 potential 52, 93, 94, 97, 98, 103, 113,
 225, 247, 261
 solar 219
 equation
 algebraic 312
 Drake 221
 differential 4, 68, 69, 76, 93, 157,
 204, 283, 284, 306
 Friedmann 267, 282, 284, 288, 291,
 294, 296, 301, 306
 Friedmann normalized 158, 285, 297,
 299, 304
 heat 289
 Kepler 36–39
 Kepler’s time 39
 Pogson’s 110
 Poisson 281
 quadratic 59, 61
 Schrödinger 290
 transcendental 39
 equations
 Einstein’s 157, 282–284, 288, 290,
 291, 293, 294, 304
 of mathematical physics 288, 290
 error
 discretization 76, 206, 293
 extrapolation 290
 iteration 76
 modeling 65, 76, 79, 131, 136, 206,
 235, 282, 289, 291–294, 299, 304
 relative 29

- rounding 76, 206, 293
 - systematic 284
- estimate
 - two-sided 213, 215–217
 - upper 159
- exoplanet 58, 220, 246, 247
- expansion
 - accelerated 156, 162, 303, 306
 - Hubble 241
 - local 213
 - local Hubble 175, 187, 195, 197, 199, 219
 - Taylor 133, 159, 295, 296
- extinction 300, 301
- formula
 - Einstein’s 257
- function
 - concave 160, 277
 - convex 277
 - decreasing 287
 - distribution 212
 - expansion 154, 155, 160, 207, 208, 218, 270, 283–288, 295, 297, 302, 305, 306
 - exponential 302
 - increasing 133, 305
 - linear 208, 218
 - monotonic 287
 - quadratic 208
 - rational 216–218
 - strictly concave 287
 - strictly convex 133, 160, 287, 288
- galaxy 307, 315
 - barred spiral 135, 240
 - elliptic 96, 240
 - irregular 240
 - lenticular 240
 - M31 101, 125, 134, 147, 148, 304
 - spiral 122–126, 129, 131, 135, 240, 257, 298
 - superdense 240
- Galaxy 58, 101, 102, 104, 111, 112, 124, 128–130, 134–136, 148–150, 164, 166, 220, 221, 241, 242, 246, 248, 260, 284, 298, 305, 312
- geodesic 30, 32, 264, 273, 274
- geometry
 - elliptic 264
 - Euclidean 29, 264, 269
 - hyperbolic 268, 269, 277, 278
 - Lobachevskian hyperbolic 32, 272
 - non-Euclidean 263
 - non-Euclidean spherical 32
 - Riemannian elliptic 32
 - Riemannian spherical 29
 - spherical 263, 269
- gnomon 14, 18, 19
- Hipparcos 27, 127
- hole
 - black 35, 46, 58, 59, 104, 124, 125, 129, 152, 240, 245, 305, 307, 308
- horizon 272, 315
- hypersphere 263, 283
- inclination 23, 40, 41, 99, 225, 230
- inequality
 - Bishop–Gromov 106
 - Cauchy–Schwarz 276
 - triangle 264, 273, 276, 277
 - triangle converse 277
- Io 28
- Jupiter 5, 10, 28, 46, 55, 64, 67, 72–74, 79, 201, 225, 227, 234, 238, 247, 259
- Larissa 226, 227
- law
 - gravitational 42, 47, 93, 136
 - harmonic 10
 - Hubble 101, 115, 149, 284
 - Kepler’s first 8, 9, 36, 38
 - Kepler’s second 8, 9, 11, 38
 - Kepler’s third 6, 9, 22, 42, 47–49, 51, 53, 54, 57, 61, 67, 186, 197, 199–201, 209, 225, 232, 233, 247

Kepler's third generalized 48, 198
 Newton's of gravitation 43, 47, 55,
 57, 62, 65, 68, 123, 136
 Newton's second 71, 93
 of action and reaction 47
 of conservation of energy 163, 224,
 249, 251, 252, 260
 of conservation of momentum 163,
 251, 252
 of cosines 31, 131
 of gravity 6, 42
 of sines 22
 Stefan–Boltzmann 173, 175
 lensing
 gravitational 91, 92, 137–139, 312
 weak 139, 301
 line of nodes 60
 longitude of the ascending node 40, 41
 manifold 264, 266, 274, 278
 maximally symmetric 283, 291
 n-dimensional 263
 Mars 8, 9, 22, 47, 49, 50, 53, 54, 63,
 64, 163, 166–177, 191, 220, 224, 225,
 227, 229–232, 236, 259
 mass
 of baryonic matter 127
 virial 98, 100, 115, 119, 297
 matter
 baryonic 121, 127, 129, 130, 138,
 285, 286, 295, 298, 306
 dark 1, 47, 90, 91, 93, 97, 99, 101,
 102, 111, 112, 120–122, 130, 134,
 136–138, 140, 155, 285, 286, 295,
 297–300, 304
 degenerate baryonic 102
 luminous 134
 nonluminous baryonic 102, 112
 ultrarelativistic 284
 Mercury 21, 34, 50, 58, 63, 64, 79, 125,
 166, 172, 237, 293
 method
 geometric 23
 Halley's 23
 Runge-Kutta 256
 symplectic 256
 Zwicky's 100
 metric 264, 272, 273, 275, 277
 Euclidean 274, 277
 microquasar 310
 Milky Way 35, 97, 108, 124, 127, 130,
 148, 150, 220, 240, 241, 245, 246,
 294, 301, 308
 model 140, 284, 297
 Copernican 27
 cosmological 140, 144, 267, 282–284,
 294, 295, 303, 304, 316, 317
 discrete 76
 Friedmann 304
 geocentric 140
 geometric 7
 heliocentric 7, 28, 140
 idealized 313
 Λ CDM 158, 282, 302, 306
 mathematical 62, 75, 76, 79, 267,
 268, 291
 Newtonian 232
 Nice 234
 non-Euclidean 262
 of an inflating balloon 288
 of the universe 271
 postnewtonian 250, 254, 255
 Ptolemaic 27
 three-dimensional 273
 MOND 136, 299
 NASA 74, 92, 96, 169, 171, 173, 176,
 177, 224, 236, 243
 Neptune 20, 46, 51, 62, 64, 74, 79, 125,
 166, 225–227, 229, 233–237
 Opportunity 171
 orbit
 circular 47, 123, 225, 236, 254
 Clarke 53
 Earth's 196, 206

- elliptic 9, 11, 22, 37, 40, 47, 50, 51, 54, 57, 59, 84, 104, 230, 293
- geostationary 53
- Hohmann transfer 54
- Keplerian 123, 124
- Mercury's 293
- planetary 203
- stable 51
- stationary 226–230
- paradox 166, 189, 203, 305, 309, 310
 - faint young Sun 166, 191, 192, 213
 - Fermi 220, 221
 - of the large angular momentum 166
 - of superluminal speeds 307, 308
 - of tidal forces of the Moon 166, 180, 181
 - twin 277
 - Zwicky's 121
- parallax 28, 145
 - annual 26, 27, 58
- parameter
 - cosmological 285, 287, 299, 301
 - cosmological scale 154
 - curvature 285, 287, 288
 - deceleration 159, 161, 295
 - dimensionless 157, 285
 - Hubble 117, 118, 150, 154, 156, 161, 165, 206, 242, 285, 302
 - Keplerian 206
- parsec 58
- perihelion 38, 57, 293
- period
 - orbital 8, 10, 13, 49, 53, 182, 201, 225, 250, 293
 - sidereal rotation 53
- Phobos 49, 50, 166, 224, 225, 229–232
- ping-pong gravitational 73, 74
- plasma quark-gluon 91
- Pluto 46, 55, 234, 255
- point
 - autumn 30
 - freezing 175
- Lagrangian 152, 219
 - spring 30
 - triple 171, 172
 - vernal 30
- Polaris 32, 110
- potential gravitational 45
- principle
 - anthropic 211, 213, 215
 - Einstein's cosmological 33, 34, 150, 267, 281, 313, 315
 - equivalence 260
 - Heisenberg uncertainty 212
 - time-lens 314
- problem
 - elliptic 289
 - flatness 305
 - heat conduction 289
 - horizon 281, 305, 306
 - ill-conditioned 293
 - N*-body 65, 73, 80, 93, 235, 290, 293
 - of hierarchical structures 305
 - of homogeneity and isotropy 305
 - restricted three-body 69
 - three-body 70, 240
 - two-body 68, 70, 105, 254
- Proxima Centauri 27
- pseudometric 277
 - Minkowski 265, 275, 277, 281
- pseudosphere 265, 272, 277–280
- pulsar 294
 - binary 182, 294
- quadrant 7, 14, 15
- quasar 307, 308, 313
- radius
 - Schwarzschild 61
- radius-vector 70
- recombination 143
- redshift 34, 93, 95, 109, 117, 146, 270, 284, 301, 307, 314, 315
 - cosmological 92, 295
 - gravitational 106, 117, 294

- relation
 - Pogson's 110
 - relativistic 116
- resonance 185
- right ascension 30, 31
- RNA 221, 222
- saros 85
- Saturn 5, 20, 46, 64, 74
- sequence
 - main 167
- Sirius 34, 46
- space 267–270
 - Euclidean 145, 158, 253, 262, 264, 269, 270, 278, 281
- spacetime 146, 268, 269, 313
 - curved 80, 182, 252
- speed
 - angular 125
 - constant 124, 125, 213
 - of expansion 306
 - of gravitational interaction 249–251, 254, 258, 306
 - of light 27, 28, 157, 205, 257, 283, 309, 314
 - orbital 127
 - recession 146, 185, 199, 216, 218, 284
 - relativistic 309
 - root-mean-square 97, 298
 - superluminal 257, 302, 307, 311, 316
- sphere 135, 151, 263, 264, 270–272, 277, 278, 302, 313, 315
 - armillary 14
 - celestial 30, 33, 57, 268, 308, 309
- spiral 181
- star
 - binary 104, 308
 - neutron 46, 91, 129, 152, 205, 294
 - solitary 121
- supernova 91, 151–153, 156, 158, 160, 300
- system
 - binary 152, 182, 250, 251, 255
 - Cartesian 179
 - geocentric 5
 - gravitationally stabilized 94
 - heliocentric 5
 - heliocentric coordinate 40, 192
 - International of Units 48
 - inertial 192
 - metric 48
 - of differential equations 69, 71, 73, 74, 251, 253
 - of units SI 49, 303
- temperature
 - central 213
 - effective 168
 - equilibrium 173, 174
 - surface 168, 174
- tensor 283
 - metric 292, 293
 - Ricci 292
 - Riemann curvature 292
- theorem
 - Banach fixed-point 254
 - binomial 148
 - Newton's first 43, 44, 98, 113, 126, 134
 - Newton's second 45, 98, 113, 120, 121, 134
 - Poynting's 257
 - Pythagorean 59, 60, 99
 - Pythagorean generalized 265
 - Shell 45, 130
 - Thales' 250
 - Virial 47, 90, 93, 95–98, 100, 101, 103, 105, 108, 115, 119
- theory
 - Darwin's evolutionary 223
 - General of relativity 32, 34, 182, 252, 257, 258, 275, 281, 284, 307, 308, 316
 - Newtonian of gravity 65, 69, 136, 209, 210, 249, 284
 - of differential equations 71

Titan 170
 tractrix 279, 280
 triangle 283
 curved 32
 equilateral 179, 256
 right 16, 60, 250
 right-angled 86
 spherical 29, 31, 32, 264
 Summer 31
 Triton 166, 226, 235–237

 unit astronomical 22, 48, 165, 194, 294
 universe 267–271
 bounded 302
 cyclic 286, 287, 295
 Einstein’s stationary 159, 283, 286
 Euclidean 302
 expanding 244, 247, 282, 313
 infinite 297
 maximally symmetric 282
 non-Euclidean 262, 263
 observable 146, 268, 269, 271, 285,
 306, 315
 oscillating 287, 301
 real 288
 stationary 283, 284, 301
 Uranus 46, 63, 64, 74, 225–229, 234,
 235

 Vega 30, 31
 velocity
 angular 231
 apparent 311, 312
 areal 11
 circular 66
 constant 217
 escape 51, 52, 64, 66
 mean orbital 50, 64
 orbital 51, 189
 radial 99, 104, 120
 superluminal 69
 Venus 5, 20, 21, 23, 50, 64, 81, 82, 164,
 201, 220

 year
 calendar 28
 Gregorian calendar 202
 sidereal 28, 49, 55, 185, 187, 196,
 204
 tropical 202

 zone
 habitable 192, 216, 221, 248