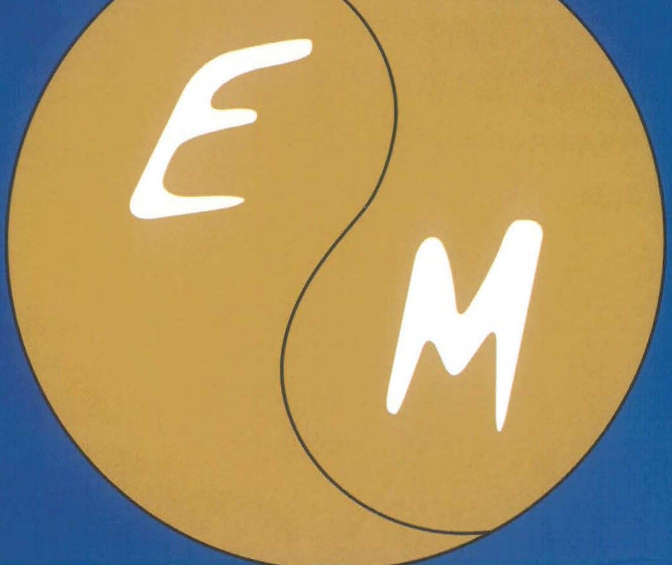


Energy and Mass in Relativity Theory

Lev B Okun

World Scientific

Energy and Mass in Relativity Theory

This page intentionally left blank

Energy and Mass in Relativity Theory

Lev B Okun

A I Alikhanov Institute of Theoretical and Experimental Physics, Moscow, Russia

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

The author and publisher would like to thank the following publishers of the various journals and books for their assistance and permissions to include the selected reprints found in this volume:

Uspekhi Fizicheskikh Nauk (Articles 1, 6, 11, 12, 20, 26, 29 & 30)
Harwood Academic Publishers (Article 2)
American Institute of Physics (Articles 3, 4, 5 & 13)
Springer Science and Business Media (Articles 7, 8, 16 & 21)
IOP Publishing Limited (Articles 9 & 17)
Elsevier (Article 18)
Acta Physica Polonica (Articles 24 & 25)
Progress of Theoretical Physics (Article 27)

ENERGY AND MASS IN RELATIVITY THEORY

ISBN-13 978-981-281-411-1
ISBN-10 981-281-411-6

Printed in Singapore by World Scientific Printers

Preface

This book contains reprints of articles on various facets of teaching relativity theory, with its symmetry principles, conservation laws of energy E, momentum $\mathbf{p}$ and mass m, and the fundamental kinematic equation, $(E/c^2)^2 - (\mathbf{p}/c)^2 = m^2$, where c is the velocity of light.

This fundamental equation has been tested again and again in various experiments with radioactive substances, cosmic rays and particle accelerators. It is equally applicable to massive particles, such as protons and electrons, to very light neutrinos and to massless particles of light — photons. From this equation, when $\mathbf{p} = 0$, Einstein's great formula follows: $E_0 = mc^2$, according to which even a small mass contains a huge energy in the form of rest energy E_0.

Doing physics research is like climbing a mountain; to see the path to the truth, one has to look down from the standpoint of higher, advanced knowledge. In the case of relativity this standpoint — the vantage ground of truth — was found by Minkowski, who united time t and space $\mathbf{r}$ into a single four-dimensional world. Similarly, E/c^2 and $\mathbf{p}/c$ are united in a four-dimensional vector, for which m is its pseudo-Euclidean length.

If one attempts to pour the new wine of relativistic four-dimensional physics into an old wineskin of three-dimensional Newtonian formulas, one ends up with such confusing terms as "relativistic mass" and "rest mass" and the famous but misleading equation $E = mc^2$ which differs from Einstein's true equation by the absence of subscript zero.

This famous equation and the concept of mass increasing with velocity indoctrinate teenagers through the popular science literature, and through college text-books. According to Einstein, "common sense is a collection of prejudices acquired by age eighteen. "It is very difficult to get rid of this "common sense" later: "better untaught than ill taught." As a result one can find the term "rest mass" even in serious professional physics journals. One of the aims of this book is to help the reader to get out of the habit of using this term.

I am grateful to Professor K. K. Phua, who invited me to publish this book, and to Mister Chee-Hok Lim, who carefully edited it.

The book consists of 30 articles published over the course of 40 years (1968–2008). The three most recent papers resulted from attempts to write this preface. They elaborate on the historical evolution of the concepts of energy, momentum and mass from the standpoint of relativity principle, and suggest a simple method of deriving the main results of both special and general relativity. They acknowledge those who helped me to write these papers and hence this book.

Lev Okun
Moscow
August 2008

Contents

SOVIET PHYSICS USPEKHI VOLUME 11, NUMBER 3 NOVEMBER-DECEMBER, 1968

534.14

ON THE PHOTON MASS

I. Yu. KOBZAREV and L. B. OKUN'

Institute of Theoretical and Experimental Physics, Moscow

Usp. Fiz. Nauk 95, 131–137 (May, 1968)

1. INTRODUCTION

THIS article is devoted to the experimentally known data concerning the mass of the photon. This question has not been discussed recently in the physics literature, and it is usually implied that the photon mass is exactly equal to zero.

The arguments frequently advanced in favor of the idea that the photon mass is strictly equal to zero are as follows:

1. The existence of electromagnetic action as a distance implies that the photon mass is very small compared with the masses of other particles, and there should be no small parameters in the theory.

2. The theory (relativity theory, quantum electrodynamics) requires that the photon mass be equal to zero.

It is easy to see, however, that both arguments are wrong. To verify that the first argument is incorrect, it is sufficient to recall that the ratio of the constants of the gravitational and weak interaction is approximately 10^{-34}. Thus, small parameters are encountered in physics. We note that if the photon mass m_γ were smaller than the electron mass by 34 orders of magnitude, then its Compton wave length $\lambda_\gamma = \hbar/m_\gamma c$ would equal* approximately 10^{23} cm $\approx 10^5$ light years.

As regards the assertions that the vanishing of the photon mass follows from the theory, we can state the following.

If the photon mass were not equal to zero, then no harm would come to the special theory of relativity; the velocity that enters in the Lorentz transformation would simply be not the velocity of light, but the limiting velocity c to which velocities of all the bodies tend when their energy becomes much larger than their mass.

Within the framework of quantum electrodynamics, the vanishing of the photon mass is the consequence of the so-called gauge invariance of the second kind. However, the absence of gauge invariance of the second kind does not lead to any difficulties, unlike the gauge invariance of the first kind, violation of which denotes nonconservation of the charge.

Actually, gauge invariance of the second kind is not the cause but the mathematical expression of the vanishing of the photon mass. Quantum electrodynamics with a nonzero photon mass has no theoretical flaws: the charge in it is conserved and is renormalizable. In some respects it is even simpler than ordinary electrodynamics, since its quantization does not require an indefinite metric.

Thus, the question of the photon mass is not theoretical but experimental. This circumstance was apparently first formulated distinctly by deBroglie[1,2], who obtained the first estimates of the upper limit of the photon mass. More rigorous estimates were obtained in 1943 by Schrödinger[3,4]. These two authors apparently exhausted[1-7] all the presently known methods of determining the upper limit of the photon mass. The question of limits for m_γ was discussed also by Ginzburg[8]. In the present brief review we shall refine somewhat the estimates on the basis of the latest experimental data, and compare the estimates obtained by different methods. We shall see how the photon mass will affect its free motion in vacuum (Sec. 2), the interaction between charges and currents realized as the result of exchange of virtual photons (Sec. 3), and finally, the properties of blackbody radiation (Sec. 4). The main results of this analysis are listed in a table at the end of the article.

We shall not discuss here the possible cosmological manifestation of a nonzero photon mass.*

2. PHOTON MASS AND VELOCITY OF LIGHT AND OF RADIO WAVES

As already mentioned, the presence of a photon mass would bring about a situation wherein the velocity of the photon in vacuum would no longer be a universal constant, and would depend on the photon energy, just as in the case of other particles with nonzero mass. For the group velocity of the electromagnetic waves we would have

$$\frac{v}{c}=\frac{k}{\sqrt{k^2+m_\gamma^2c^2}}=\frac{1}{\sqrt{1+\left(\frac{\lambda}{\lambda_\gamma}\right)^2}}.$$

As a result the velocity of, say, blue light would be larger than that of red light. It was precisely this circumstance which was used by de Broglie to estimate the upper limit of the photon mass. He noted that the dispersion of the speed of light in vacuum would lead to the occurrence of color phenomena in the case of eclipses of double stars: the blue light would arrive earlier than

*We shall use frequently also the system of units with $\hbar = c = 1$, and then $\lambda\!\!\!^{-}_\gamma = 1/m_\gamma$.

*De Broglie [6] gave an estimate $m_\gamma < 10^{-66}$ g, meaning $\lambda\!\!\!^{-}_\gamma > 10^{10}$ light years ($\lambda\!\!\!^{-} = \lambda/2\pi$). This estimate was obtained from the assumption that the masses of the photon and of the graviton, if they differ from zero, should be commensurate. The graviton mass in Einstein's equations for the gravitation field can be connected with the so-called cosmological constant. From a comparison of the Friedmann theory of the expanding universe with experiment it is seen that the Compton wavelength of the graviton cannot greatly exceed the radius of the visible part of the universe ($\sim 10^{10}$ light years). It is obvious, however, that such an estimate cannot be regarded as an experimental lower limit for the Compton wavelength of the photon.

the red light. If the minimum delay time of the red light compared with the blue light, which can be noticed, is denoted by δt, and the time required for the light to travel from the sun to the earth is denoted by t, then we readily obtain

$$\frac{\delta t}{t} \approx \frac{\lambda_1^2-\lambda_2^2}{2\lambda_\gamma^2} \approx \frac{1}{2}\left(\frac{\lambda_1}{\lambda_\gamma}\right)^2,$$

where λ_1 is the wavelength of the red light and λ_2 that of the blue light. Assuming, as did de Broglie, that $\lambda_1 = 10^{-5}$ cm, $\delta t = 10^{-3}$ sec, and $t = 10^{10}$ sec (on the order of 10^2 light years), we get

$$\lambda_\gamma \gg \lambda_1 \sqrt{\frac{t}{\delta t}} \sim 10 \text{ cm},$$

corresponding to*

$$m_\gamma \lesssim \frac{10^{-37}\,\text{g-cm}}{10\text{ cm}} = 10^{-38}\text{ g}.$$

Color phenomena lasting several minutes were observed in eclipses of variable stars (the Tikhov-Nordman effect). However, as noted by P. N. Lebedev[9], this effect is due to the fact that different sections of the stellar atmospheres have different spectral characteristics. Therefore the limit of λ_γ should be even lowered, and it is therefore too low to be of interest.

A somewhat better lower limit for λ_γ, but still too low, can be obtained by using data on radio-wave propagation. It follows from measurements made by a group headed by L. I. Mandel'shtam[10] that the phase velocity of propagation of radio waves with $\lambda = 300$ m coincides with the velocity of visible light accurate to 5×10^{-4}.

From this we get

$$\lambda_\gamma \gg 1.5 \text{ km}.$$

A further increase of the limit for λ_γ by this method encounters the difficulty that the velocity change, connected with the influence of the earth's surface already becomes significant at the attained accuracy, as was already noted in[10].

3. PHOTON MASS AND STATIC FIELDS

The lowest limit for the photon mass follows from data on the static magnetic field (the earth's field). As will be seen below, data on the static electric field give a much worse limitation. In order to understand how these limitations arise, let us consider in greater detail the properties of electrodynamics with $m_\gamma \neq 0$.

In such a theory, the expression for the electromagnetic field has the usual form, and consequently the photon-emission vertex is also usual. The only deviation from the usual electrodynamics is that the photon propagator $D_{\mu\nu}$ does not have the form

$$D_{\mu\nu} = \frac{\delta_{\mu\nu}}{k^2},$$

but becomes equal to the propagator of the neutral vector meson:

$$D_{\mu\nu} = \frac{\delta_{\mu\nu} - \dfrac{k_\mu k_\nu}{m_\gamma^2}}{k^2 - m_\gamma^2}.$$

By virtue of the conservation of the current, the vertices Γ_μ are transverse:

$$k_\mu \Gamma_\mu = 0$$

(we assume here and throughout that the current with which the photon interacts is conserved, since nonconservation of the current, and in particular nonconservation of the charge, would of necessity lead to the appearance of a photon mass), and therefore in the simplest cases, which we shall consider, the second term $k_\mu k_\nu / m_\gamma^2$ in the numerator of the propagator makes no contribution and we can write

$$D_{\mu\nu} = \frac{\delta_{\mu\nu}}{k^2 - m_\gamma^2}.$$

It is well-known that the static potential corresponding to the propagator $\delta_{\mu\nu}/k^2$ is the Coulomb potential

$$V(r) = \frac{Q}{r}.$$

The potential corresponding to the propagator $\delta_{\mu\nu}/(k^2 - m^2)$ is the Yukawa potential

Thus, the presence of a photon mass should lead to an exponential decrease of the interaction at distances exceeding the photon Compton wavelength $\lambda_\gamma = 1/m_\gamma$.

A similar change should take place also in the well-known expressions for the magnetic field. Thus, for example, in place of the ordinary expression for the field of a magnetic dipole*

$$\mathbf{A} = \frac{[\mathbf{mr}]}{r^3}$$

we get the expression

$$\mathbf{A} = \left(\frac{[\mathbf{mr}]}{r^3}\right)(1 + m_\gamma r)\, e^{-m_\gamma r}.$$

using this expression to describe the earth's magnetic field, and making use of the fact that the earth's magnetic field extends to distances on the order of 10^4 km, Schrödinger concluded that $\lambda_\gamma > 10^4$ km[3,4]. Data on the earth's magnetic field, obtained up to the time of Schrödinger work from cosmic-ray studies and studies of the northern lights, have by now been greatly supplemented by satellite measurements (see, for example,[11]). These measurements show that the magnetic field of the earth has the form of a magnetic dipole up to distances on the order of $6R_+ \approx 30\,000$ km. This indeed gives the best presently available limit of m_γ. Further refinement of this limit by measuring the earth's magnetic field is impossible, since at large distances the earth's magnetic field becomes comparable with the magnetic field of the solar plasma flowing around the earth—the so-called solar wind.

As noted by Gintsburg[8], the limit of λ_γ can be refined further by measuring the magnetic field of Jupiter. Results of radio observations of Jupiter in the decimeter band (200—300 MHz) and in the dekameter band (5—43 MHz) point to the existence on Jupiter of radiation belts at distances on the order of $2R_{\text{Ц}}$, and conse-

*The estimate $m < 10^{-44}$ g given in [2,5] is apparently due to a misprint.

*$[\mathbf{m\,r}] \equiv \mathbf{m} \times \mathbf{r}$.

quently of a strong magnetic field (from 1 to 10 G in accordance with various estimates). The radius of Jupiter $R_♃$ is larger than the radius of the earth by approximately one order of magnitude ($R_♃ = 11.2R_+$ = 71 400 km); therefore the measurement of the magnetic field of Jupiter at distances of several times $R_♃$ would make it possible to raise the limit of λ_γ by approximately one order of magnitude, possibly up to a million kilometers.

From the point of view of the question considered by us, it is of interest to increase further the accuracy of interference radio observation of Jupiter, which give a high spatial resolution, and particularly to perform an accurate measurement of Jupiter's magnetic field with the aid of rockets.

Let us consider now the limitations imposed on the photon mass by experiments aimed at verifying the Coulomb law. From among the experiments of this type known to us, the most accurate was performed in 1936 by Plimpton and Lawton[15]. In this experiment they measured the potential difference between two concentric spheres, the outer one being charged to a certain potential V_1. At an outer-sphere radius $R_1 \approx 75$ cm, inner-sphere radius $R_2 \approx 60$ cm, and potential V_1 = 3000 V they obtained $V_1 - V_2 \lesssim 10^{-6}$ V. (In the case of $m_\gamma = 0$, as is well-known, we should get $V_1 - V_2 = 0$.) The deviations from the Coulomb law were sought by Plimpton and Lawton in the form

$$V(r) = \frac{Q}{\gamma^{1+\varepsilon}}.$$

In this case we should get

$$\frac{V_2 - V_1}{V_1} \sim \varepsilon,$$

and it follows from the experimental result that $\epsilon \lesssim 10^{-9}$. However, this experiment is less sensitive to the type of modification of Coulomb's law which is of interest to us. Indeed, in the case when

$$V(r) = \frac{Qe^{-m_\gamma r}}{r}$$

the field intensity at the outer surface of the charged sphere is

$$E = \frac{Qm_\gamma^2}{3},$$

and we get for the potential difference in the Plimpton-Lawton experiment

$$\frac{V_1 - V_2}{V_1} \sim \frac{1}{3} m_\gamma^2 (R_1 - R_2) R_1 \lesssim 3 \cdot 10^{-10}.$$

It follows therefore that the lower limit for the Compton wavelength of the photon is

$$\lambda_\gamma \approx 3 \cdot 10^4 \sqrt{(R_1 - R_2) R_1} \sim 10 \text{ km}.$$

This limitation is weaker by three orders of magnitude than that obtained from data on the earth's magnetic field.

4. PHOTON MASS AND BLACKBODY RADIATION

A photon with nonzero mass, unlike a photon with zero mass, has not two but three polarization states. It may therefore turn out that if the photon has a mass then an additional factor 3/2 appears in Planck's formulas for the density of blackbody radiation, and thus the Boltzmann black-body constant should differ by a factor 1.5 from the observed one. It is easy to see, however, that this will not take place. The reason why the statistical description of black-body radiation remains practically unchanged[7] lies in the fact that the probabilities of transitions in which "longitudinal" photons take part contain a small factor m_γ^2/ω^2, where ω is the photon frequency.*

Let us consider, following[7] a photon gas situated in a cavity with linear dimensions L. For the time of transformation of the transverse photons in the cavity into longitudinal ones we get

$$t \sim \frac{L}{c}\left(\frac{\omega}{m_\gamma}\right)^2 = \frac{L}{c}\left(\frac{\lambda_\gamma}{\lambda}\right)^2.$$

Assuming that ω is in the optical region ($\lambda \sim 10^{-5}$ cm) and putting L = 10 cm and $\lambda_\gamma = 10^9$ cm, we get $t \sim 10^{18}$ sec $\sim 10^{11}$ years. Thus, no real effect arises. We note that at such small values of w_z the walls of practically any cavity will be transparent to the "longitudinal" photons.

CONCLUSION

Thus, the Compton wavelength of the photon is certainly larger than 30 000 km. It is possible to increase this limit by measuring the magnetic field of Jupiter at distance from the planet where this field is small. At present we see no other experiment capable of yielding a comparable accuracy.

We can raise the question whether it is necessary to strive to lower the limit of the photon mass if the theory with a zero photon mass seems more attractive from the esthetic point of view than the theory with nonzero mass, and if there are no theoretical grounds whatever for introducing the photon mass. After all, we do not see how the solution of any of the numerous problems of

---

*In this note we present a simple derivation of this statement. Let the photon-emission amplitude A be given by

$$A = M_\alpha e_\alpha,$$

where e_α is the 4-vector-potential of the photon normalized to a single photon, and M_α is a 4-vector that depends on the remaining 4-vectors of the problem (4-momenta of the particles and their spins). Let the photon momentum be directed along the z axis. In the system in which the photon is at rest, the probabilities of emission of transverse photons (w_x and w_y) and of a "longitudinal" photon (w_z) are respectively

$$w_x = |M_x^c|^2, \quad w_y = |M_y^c|^2, \quad w_z = |M_z^c|^2,$$

Where the index c denotes the rest system.

The same probabilities, expressed in terms of the components of the 4-vector M in the laboratory system, are respectively

$$w_x = |M_x|^2, \quad w_y = |M_y|^2, \quad w_z = |M_z|^2 - |M_0|^2.$$

We now use the transversality of the 4-vector M:

$$k_0 M_0 - k_z M_z = 0,$$

and obtain

$$w_z = |M_z|^2 \frac{k_0^2 - k_z^2}{k_0^2} = |M_z|^2 \frac{m_\gamma^2}{\omega^2}.$$

since our currents are the same as in ordinary electrodynamics, M_z is finite when $m_\gamma^2 \to 0$, and we can assume that M_x, M_y, and M_z are of the same order of magnitude. It follows therefore that

$$w_x \sim w_y, \quad w_z \sim \frac{m_\gamma^2}{\omega^2} w_x.$$

Upper limit of the photon mass m_γ (lower limits of the Compton wavelength of the photon λ_γ), which follow from different experimental data

Physical phenomenon	λ_γ	m_γ/m_e*	Literature**
Dispersion of the velocity of light from double stars	0.1 km	10^{-9}	de Broglie [2]
Velocity of radio waves	1 km	10^{-15}	Mandel'shtam [9,10]
Coulomb's law (measurement of the field inside a charged sphere)	10 km	10^{-16}	Plimpton-Lawton [15]
Extent of the earth's magnetic field	30 000 km	10^{-20}	Schrodinger [3,4]
Extent of the magnetic field of Jupiter***	10^6 km	10^{-21}	Gintsburg [8]

*m_e – electron mass.
**The numerical values of m_γ/m_e and χ_γ listed in the table differ in a number of cases from those presented in the cited papers.
***The measurements were not made.

modern theory of elementary particles can be facilitated by the presence of a photon mass.

However, esthetic arguments are frequently in error, and it is necessary to understand clearly that all we know at present is that $\lambda_\gamma > 30\,000$ km. We cannot guarantee that attempts to increase this limit will not lead to unexpected results.

We take the opportunity to thank N. L. Grigorov, Ya. B. Zel'dovich, S. B. Pikel'ner, A. E. Salomonovich, I. B. Khriplovich, and A. E. Chudakov for interesting discussions.

The question of an experimental limit of the photon mass was suggested to us in the Fall of 1966 by I. Ya. Pomeranchuk.

[1] L. de Broglie, Phil. Mag. 47, 446 (1924).
[2] L. de Broglie, La Mecanique ondulatoire du photon. Une nouvelle theorie de la Lumiere, t. 1, Paris 1940, p. 39, 40.
[3] E. Schrödinger, Proc. Roy. Irish. Acad. A49, 43 (1943).
[4] E. Schrödinger, Proc. Roy. Irish. Acad. A49, 135 (1943).
[5] L. de Broglie, Mechanique ondulatoire du photon et theorie quantique de champs, Paris, 1949.
[6] L. de Broglie, La Theorie generale des particules á spin, Paris, 1943, p. 191.
[7] L. Bass, E. Schrödinger, Proc. Roy. Soc. **A232**, 1 (1955).
[8] M. A. Gintsburg, Astron. zh. 40, 703 (1963) [Sov. Astron. AJ 7, 536 (1963)].
[9] P. N. Lebedev, Sobranie sochineniĭ (Collected Works), AN SSSR, 1963, p. 282.
[10] L. I. Mandel'shtam, Polnoe sobr. trudov (Complete Collect. Works) AN SSSR, 1947, V. 2, pp. 277–305 and V. 3, p. 238.
[11] L. Biermann, Sitzber. Bayer. Akad. Wissenschaft. 37 (1965).
[12] V. V. Zheleznyakov, Radioizluchenie Solntsa i planet (Radio Emission from the Sun and the Planets), Nauka, 1964, p. 509.
[13] G. R. A. Ellis, Radio Science, J. Research NBS/USNC-URSI, 1513 (1965).
[14] J. A. Roberts, Radio Science, J. Research NBS/USNC-URSI, 1543 (1965).
[15] S. J. Plimpton and W. E. Lawton, Phys. Rev. 50, 1066 (1936).

Translated by J. G. Adashko

Mass, energy, and momentum in Einstein's mechanics

Newtonian mechanics provides a perfect description of the motion of a body when its velocity v is much less than the velocity of light: $v \ll c$. But this theory is manifestly incorrect when the velocity is comparable with c, especially when $v = c$. To describe motion with arbitrary velocity, up to that of light, we have to turn to Einstein's special theory of relativity, i.e., relativistic mechanics. Newton's non-relativistic mechanics is but a particular (although very important) limiting case of Einstein's relativistic mechanics.

The word "relativity" originates in Galileo's relativity principle. In one of his books, Galileo explained in very graphic terms that no mechanical experiment performed inside a ship could establish whether the ship is at rest or moving relative to the shore. Of course, this would be easy to do by simply looking out. But if one were in a cabin without portholes, then rectilinear motion of the ship could not be detected.

Galileo's relativity can be shown mathematically to demand that the equations of motion of bodies, i.e., the equations of mechanics, must be identical in all the so-called inertial reference frames. These are coordinate systems attached to bodies moving uniformly along straight lines relative to very distant stars. (Obviously, in the case of Galileo's ship, we disregard the diurnal rotation of the Earth, its rotation around the Sun, and the rotation of the Sun around the center of our Galaxy.)

Einstein's extremely important achievement was that he extended Galileo's relativity principle to all physical phenomena, including electrical and optical processes involving photons. This generalization of Galileo's principle has resulted in drastic changes in fundamental ideas, such as those of space, time, mass, momentum, and energy. For example, in Einstein's relativity the concepts of total energy and rest energy are introduced. The kinetic energy T is related to total energy E by

$$E = E_0 + T$$

where E_0 is the rest energy, related to the mass m of a body by the famous formula

$$E_0 = mc^2.$$

The mass of photon being zero, its rest energy E_0 is also zero. For a photon, "a rest is but a dream": it always moves with the velocity of light

Reprinted from: *α, β, γ ... Z. A Primer in Particle Physics* (Harwood Academic Publishers, 1987), pp. 9–13.

c. Other particles, e.g., electrons and nucleons, have nonzero masses and therefore nonzero rest energy.

The relations between energy, velocity, and momentum for particles with $m \neq 0$ take the following forms in Einstein's mechanics:

$$E = \frac{mc^2}{\sqrt{(1-v^2/c^2)}}, \; p = Ev/c^2.$$

Consequently,

$$m^2c^4 = E^2 - p^2c^2.$$

Each of the two terms on the right-hand side of the latter equation increases as the body travels faster, but the difference remains constant (physicists prefer to say that it is invariant). The mass of a body is a relativistic invariant, since it is independent of the reference frame in which the motion of the body is described.

Einstein's relativistic formulas for momentum and energy become identical with the non-relativistic Newtonian expressions when $v/c << 1$. Indeed, if we expand the right-hand side of the relation $E = mc^2/\sqrt{1-v^2/c^2}$ in a series in powers of the small parameter v^2/c^2, we obtain the expression

$$E = mc^2 \left[1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\left(\frac{v^2}{c^2}\right)^2 + \ldots\right],$$

where the dots represent higher powers of v^2/c^2.

■ When $x << 1$, a function $f(x)$ can be expanded in a series in powers of the small parameter x. By differentiating the left- and right-hand sides of the expression

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \ldots,$$

and each time evaluating the result for $x = 0$, you will readily verify its validity (when $x << 1$, the discarded terms are negligible). Thus,

$$f(x) = (1 - x)^{-1/2},\; f(0) = 1,$$
$$f'(x) = \tfrac{1}{2}(1 - x)^{-3/2},\; f'(0) = \tfrac{1}{2},$$
$$f''(x) = \tfrac{3}{4}(1 - x)^{-5/2},\; f''(0) = \tfrac{3}{4}.$$ ■

Note that for the Earth moving in its orbit with a velocity of 30 km/s, the ratio v^2/c^2 is 10^{-8}. For an air-liner flying at 1000 km/h, this parameter is much smaller: $v^2/c^2 \approx 10^{-12}$. Consequently, the nonrelativistic relations

$$T = \tfrac{1}{2} mv^2, \mathbf{p} = m\mathbf{v}$$

are valid for the airliner to an accuracy of about 10^{-12}, and the relativistic corrections are definitely negligible.

Let us now return to the formula that relates mass, energy, and momentum, rewriting it in the form

$$m^2c^2 = (E/c)^2 - p_x^2 - p_y^2 - p_z^2.$$

The fact that the left-hand side of this equality is unaffected by going from one inertial reference frame to another is similar to the invariance of the square of momentum

$$\mathbf{p}^2 = p_x^2 + p_y^2 + p_z^2,$$

or of any squared three-dimensional vector, under rotations of the coordinate system (see above, Figure 1) in ordinary Euclidean space. In terms of this analogy, m^2c^2 is said to be the square of a four-dimensional vector, namely the four-dimensional momentum p_μ, where the subscript μ takes on the four values $\mu = 0, 1, 2, 3$ and $p_0 = E/c$, $p_1 = p_x$, $p_2 = p_y$, $p_3 = p_z$. The space in which the vector $p_\mu = (p_0, \mathbf{p})$ has been defined is said to be pseudo-Euclidean. The prefix "pseudo" signifies in this case that the invariant is not the sum of the squares of all four components but is the expression

$$p_0^2 - p_1^2 - p_2^2 - p_3^2.$$

The transformation that relates the time and space coordinates in two different inertial systems is called the Lorentz transformation. Without writing out this transformation in full, we note that if two events are separated by t in time and $\mathbf{r}$ in space, the quantity

$$s = (ct)^2 - \mathbf{r}^2,$$

called the *interval*, remains constant under the Lorentz transformation, i.e., it is a Lorentz invariant. Note that neither t nor $\mathbf{r}$ are invariants individually. When $s < 0$, the interval is said to be timelike; when $s > 0$, it is spacelike; when $s = 0$, it is lightlike. When $s < 0$, two spatially distinct events can be simultaneous in one reference frame but nonsimultaneous in another.

Let us now consider a system of n noninteracting free particles. Let E_i be the energy of the i-th particle, $\mathbf{p}_i$ be its momentum, and m_i its mass.

The total energy and momentum of the system are, respectively,

$$E = \sum_{i=1}^{n} E_i \text{ and } \mathbf{p} = \sum_{i=1}^{n} \mathbf{p}_i.$$

Since, by definition, the mass is given by

$$M^2 = E^2/c^4 - \mathbf{p}^2/c^2,$$

the mass of the system is not, in general, equal to the sum of the masses of its constituents.

In our daily life, we are used to the equality $M = \sum_{i=1}^{n} m_i$, but this fails for fast particles. For example, the combined mass of two electrons colliding head-on with equal energies is $2E/c^2$, where E is the energy of each electron. In accelerator experiments, this combined mass is greater by many orders of magnitude than the quantity $2m_e$, where m_e is the mass of the electron.

We conclude this section with a few remarks about terminology.

Some books and popular-science periodicals use the phrases "rest mass" and "motional mass" (or the equivalent phrase "relativistic mass"). The latter mass increases with increasing velocity of the body. The "rest mass" m_0 is then taken to mean the physical quantity that we have simply called *mass* and denoted m, and the *relativistic mass* m is taken to mean the energy of the body divided by the square of the velocity of light, $m = E/c^2$, which definitely increases with increasing velocity. This obsolete and essentially inadequate terminology was widespread at the beginning of this century when it seemed desirable, for purely psychological reasons, to retain the Newtonian relation between momentum, mass, and velocity: $\mathbf{p} = m\mathbf{v}$. However, as we approach the end of the century, this terminology has become archaic and only obscures the meaning of relativistic mechanics for those who have not mastered its foundations sufficiently well.

In relativistic mechanics, the "rest mass" is neither the inertial mass (i.e., the proportionality coefficient between force and acceleration) nor the gravitational mass (i.e., the proportionality coefficient between the gravitational field and the gravitational force acting on a body). It must be emphasized that the infelicitous "motional mass" cannot be interpreted in this way, either.

The correct relation between the force **F** and the acceleration $d\mathbf{v}/dt$ can be found if we use the relativistic expression for the momentum

$$\mathbf{p} = m\mathbf{v}/\sqrt{(1-v^2/c^2)},$$

and recall that $\mathbf{F} = d\mathbf{p}/dt$. The formula $\mathbf{F} = m\mathbf{a}$, which is familiar from school textbooks, is valid only in the non-relativistic limit in which $v/c \ll 1$.

That the gravitational attraction is not determined by the "rest mass" is evident, for example, from the fact that the photon is deflected by a gravitational field despite its zero "rest mass". The gravitational attraction exerted on the photon by, say, the Sun, increases with increasing photon energy. We are therefore tempted to say that the phrase "motional mass" is meaningful at least in this case. In fact, this is not so. A consistent theory of the motion of a photon (or any object moving with velocity comparable with the velocity of light) in a gravitational field will show that the energy of a body is *not* equivalent to its gravitational mass.

To conclude this discussion of "mass", I must ask the reader never to use the phrases "rest mass" or "motional mass" and always to mean by "mass" the relativistically invariant mass of Einstein's mechanics.

Another example of unfortunate terminology is the false claim that high-energy physics and nuclear physics are somehow able to transform energy into matter and matter into energy. We have already pointed out that energy is *conserved*. Energy does not transform into anything, and it is only different particles that transform into one another. We shall discuss many examples of such transformation in the following pages. The point is well illustrated by the chemical reaction between carbon and oxygen that we observe in a bonfire. This reaction is

$$C + O_2 \rightarrow CO_2 + \text{photons}.$$

The kinetic energy of photons and CO_2 molecules is produced in this reaction because the combined mass of the C atom and the O_2 molecule is slightly greater than the mass of the CO_2 molecule. This means that while all the energy of the initial ingredients of the reaction is in the form of rest energy, that of the final products is the sum of rest energy and kinetic energy.

Energy is thus conserved but the carriers of it, and the form in which it appears, do change.

This page intentionally left blank

THE CONCEPT OF MASS

In the modern language of relativity theory there is only one mass, the Newtonian mass m, which does not vary with velocity; hence the famous formula $E = mc^2$ has to be taken with a large grain of salt.

Lev B. Okun

Mass is one of the most fundamental concepts of physics. Understanding and calculating the masses of the elementary particles is the central problem of modern physics, and is intimately connected with other fundamental problems such as the origin of *CP* violation, the mystery of the energy scales that determine the properties of the weak and gravitational interactions, the compositeness of particles, supersymmetry theory and the properties of the not-yet-discovered Higgs bosons.

But instead of discussing all these subtle and deep connections, I feel obliged to raise and discuss an elementary question—that of the connection between mass and energy. I agree with those readers who think this topic would be more appropriate for high school pupils than for physicists, but just to find out how far off I am, I would like to propose a simple test and tell you about an opinion poll related to it.

The famous Einstein relation between mass and energy is a symbol of our century. Here you have four equations:

$$E_0 = mc^2 \qquad (1)$$

$$E = mc^2 \qquad (2)$$

$$E_0 = m_0c^2 \qquad (3)$$

$$E = m_0c^2 \qquad (4)$$

In these equations c is the velocity of light, E the total energy of a free body, E_0 its rest energy, m_0 its rest mass and m its mass.

Lev Okun is head of the laboratory of elementary-particle theory at the Institute of Theoretical and Experimental Physics, in Moscow, USSR.

Now I ask two simple questions:

▷ Which of these equations most rationally follows from special relativity and expresses one of its main consequences and predictions?

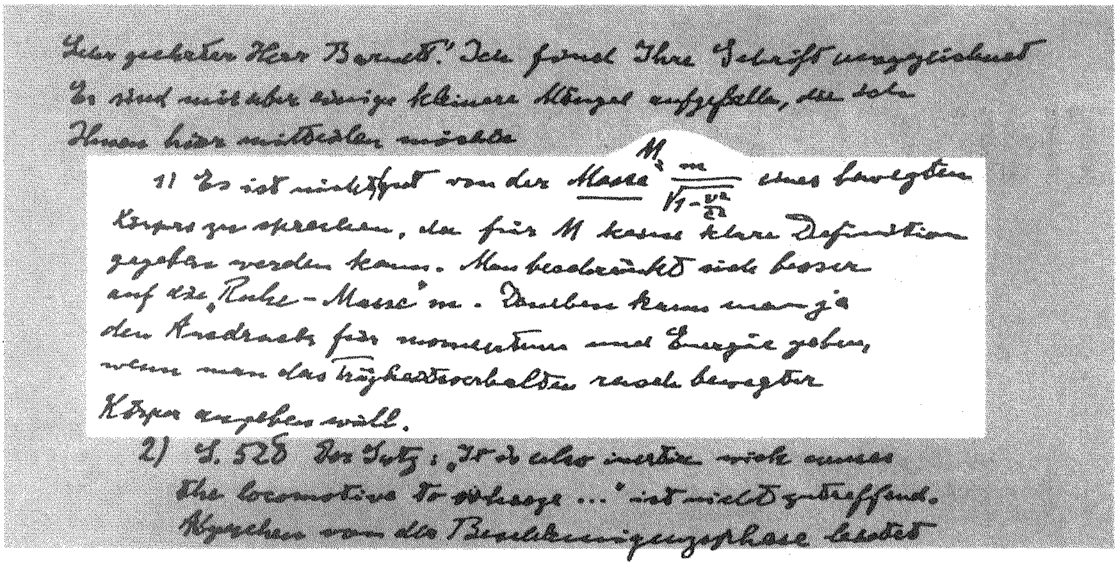
▷ Which of these equations was first written by Einstein and was considered by him a consequence of special relativity?

The correct answer to these two questions is equation 1, while opinion polls that I have carried out among professional physicists have shown that the majority prefers equation 2 or 3 as the answer to both questions. This choice is caused by the confusing terminology widely used in the popular science literature and in many textbooks. According to this terminology the body at rest has a "proper mass" or "rest mass" m_0, whereas a body moving with velocity v has "relativistic mass" or "mass" m, given by

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

As I will show, this terminology had some historical justification at the beginning of our century, but it has no rational justification today. When doing relativistic physics (and often when teaching relativistic physics), particle physicists use only the term "mass." According to this rational terminology the terms "rest mass" and "relativistic mass" are redundant and misleading. There is only one mass in physics, m, which does not depend on the reference frame. As soon as you reject the "relativistic mass" there is no need to call the other mass the "rest mass" and to mark it with the index 0.

The purpose of this article is to promote the rational terminology. You may wonder whether this subject is really so important. I'm deeply convinced, and I will try to


Sehr geehrter Herr Barnett! Ich fand Ihre Schrift ausgezeichnet. Es sind nicht aber einige kleinere Mängel aufzustellen, die ich Ihnen hier mitteilen möchte.

1) Es ist nicht gut von der Masse $M = \frac{m}{\sqrt{1-\frac{v^2}{c^2}}}$ eines bewegten Körpers zu sprechen, da für M keine klare Definition gegeben werden kann. Man beschränkt sich besser auf die „Ruhe-Masse" m. Daneben kann man ja den Ausdruck für momentum und Energie geben, wenn man das Trägheitsverhalten rasch bewegter Körper angeben will.

2) S. 52 Der Satz: „It is also inertia which causes the locomotive to ..." ist nicht zutreffend.

Letter from Albert Einstein to Lincoln Barnett, 19 June 1948. Einstein wrote in German; the letter was typed and sent in English. The highlighted passage in this excerpt says: "It is not good to introduce the concept of the *mass* $M = m/(1 - v^2/c^2)^{1/2}$ of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the 'rest mass' m. Instead of introducing M it is better to mention the expression for the momentum and energy of a body in motion." (Reprinted by permission of the Hebrew University of Jerusalem, Israel.)

persuade you, that the use of the proper terminology is extremely important in explaining our science to other scientists, to the taxpayers and especially to students in high schools and colleges. Nonrational, confusing language prevents many students from grasping the essence of special relativity and from enjoying its beauty.

Two fundamental equations

Let us return to equation 1. Its validity is apparent when one recalls two fundamental equations of special relativity for a free body:

$$E^2 - \mathbf{p}^2c^2 = m^2c^4 \qquad (5)$$

$$\mathbf{p} = \mathbf{v}\frac{E}{c^2} \qquad (6)$$

Here E is the total energy, $\mathbf{p}$ the momentum, $\mathbf{v}$ the velocity and m the ordinary mass, the same as in Newtonian mechanics.

When $\mathbf{v} = 0$, we get $\mathbf{p} = 0$ and $E = E_0$, the energy of the body at rest. Then, from equation 5,

$$E_0 = mc^2$$

This is equation 1. Rest energy was one of Einstein's great discoveries.

Now why have I written m but not m_0 in equation 5? To see the answer, let's consider the case $v \ll c$. In this case

$$\mathbf{p} \simeq \mathbf{v}\frac{E_0}{c^2} = \mathbf{v}m \qquad (7)$$

$$E = E_0 + E_{\text{kin}} = \sqrt{\mathbf{p}^2c^2 + m^2c^4} = mc^2 + \frac{\mathbf{p}^2}{2m} + \ldots$$

and

$$E_{\text{kin}} = \frac{\mathbf{p}^2}{2m}$$

Thus we obtain in the nonrelativistic limit the well-known Newtonian equations for momentum and kinetic energy. This means that m in equation 5 is the ordinary Newtonian mass. Hence, if I were to use m_0 instead of m, the relativistic and nonrelativistic notations would not match.

If the notation m_0 and the term "rest mass" are bad, why then are the notation E_0 and the term "rest energy" good? The answer is, because mass is a relativistic invariant and is the same in different reference systems, while energy is the fourth component of a four-vector $(E,\mathbf{p})$ and is different in different reference systems. The index 0 in E_0 indicates the rest system of the body.

Let us look again at equations 5 and 6, and consider them in the case when $m = 0$—the extreme "anti-Newtonian" case. We see that in this case the velocity of the body is equal to that of light: $v = c$ in any reference system. There is no rest frame for such bodies. They have no rest energy; their total energy is purely kinetic.

Thus, equations 5 and 6 describe the kinematics of a free body for all velocities from 0 to c, and equation 1 follows from them directly. Every physicist who knows special relativity will agree on this.

On the other hand, every physicist and many nonphysicists are familiar with "the famous Einstein formula $E = mc^2$." But it is evident that equations 1 and 2, $E_0 = mc^2$ and $E = mc^2$, are absolutely different. According to equation 1, m is constant and the photon is massless. According to equation 2, m depends on energy (on velocity) and the photon has mass $m = E/c^2$.

$E = mc^2$ as historical artifact

We have seen the origin of equation 1. Now let us look at the origin of equation 2. It was first written by Henri Poincaré[1] in 1900, five years before Einstein formulated special relativity.[2] Poincaré considered a pulse of light, or a wave train, with energy E and momentum $\mathbf{p}$. (I am using modern terminology.) Recalling that, according to the Poynting theorem, $p = E/c$, and applying to the pulse of light the nonrelativistic Newtonian relation of equation 7, $\mathbf{p} = m\mathbf{v}$, Poincaré concluded that a pulse of light with energy E has mass $m = E/c^2$.

The idea that mass increases with velocity is usually ascribed, following Hendrik Lorentz,[3] to J. J. Thomson. But Thomson, who considered in 1881 the kinetic energy of a freely moving charged body, calculated only the correction proportional to v^2 and therefore derived only the velocity-independent contribution to the mass.[4] In subsequent papers by Oliver Heaviside, George Searle and others, the energy was calculated for various kinds of charged ellipsoids in the whole interval $0 \leqslant v \leqslant c$, but I have not found in the papers I have read any suggestion that mass depends on velocity.[5]

The notion of the dependence of mass on velocity was introduced by Lorentz in 1899 and then developed by him[6] and others in the years preceding Einstein's formulation of special relativity in 1905, as well as in later years. The basis of this notion is again the application of the nonrelativistic formula $\mathbf{p} = m\mathbf{v}$ in the relativistic region, where (as we know now) this formula is not valid.

Consider a body accelerated by some force $\mathbf{F}$. One can show that in the framework of special relativity the formula

$$\frac{d\mathbf{p}}{dt} = \mathbf{F} \tag{8}$$

is valid. If we start from equations 5 and 6, for the case in which the body is massive (as opposed to massless) we can easily obtain

$$\mathbf{p} = m\mathbf{v}\gamma \tag{9}$$

$$E = mc^2\gamma \tag{10}$$

where

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \tag{11}$$

$$\boldsymbol{\beta} = \frac{\mathbf{v}}{c} \tag{12}$$

Substituting equation 9 into equation 8, it is again easy to get the following relation between acceleration $\mathbf{a}$, given by $\mathbf{a} = d\mathbf{v}/dt$, and force $\mathbf{F}$:

$$\mathbf{a} = \frac{\mathbf{F} - (\mathbf{F}\cdot\boldsymbol{\beta})\boldsymbol{\beta}}{m\gamma} \tag{13}$$

We see that in the general case the acceleration is not parallel to the force, unlike the Newtonian situation to which we are accustomed. Hence one cannot cling to the Newtonian relation of proportionality between $\mathbf{a}$ and $\mathbf{F}$,

$$\mathbf{a} = \frac{\mathbf{F}}{m}$$

with mass defined as a scalar, because $\mathbf{a}$ has a nonvanishing component along $\mathbf{v}$. However, when $\mathbf{F}$ is perpendicular to $\mathbf{v}$, one can consider a "transverse mass"

$$m_t = m\gamma$$

and when $\mathbf{F}$ is parallel to $\mathbf{v}$, one can consider a "longitudinal mass"

$$m_l = m\gamma^3$$

These are the very expressions with which Lorentz introduced the two masses. Together with the "relativistic mass" in the relation $\mathbf{p} = m_r\mathbf{v}$, where $m_r = E/c^2$ (which is equal to m_t when $m \neq 0$, but which had a more general meaning applicable also in the case of photons), these masses formed the basis of the language physicists used at the beginning of the century.

Making the trouble even more lasting, however, it was decided to call the "relativistic mass" m_r simply "mass" and to denote it by m, while the normal mass m was nicknamed "rest mass" and denoted m_0.

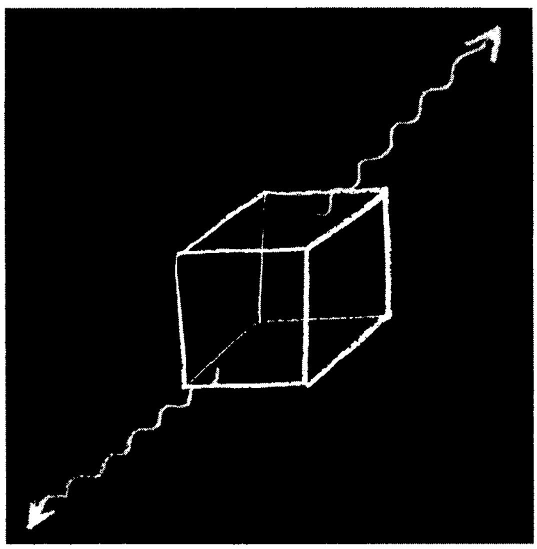
Einstein's papers of 1905 and 1906

In his first paper on relativity Einstein didn't use the term "rest mass," but he did mention the transverse and longitudinal masses.[2] He formulated the famous mass–energy relation in the second of his 1905 papers on relativity[7] in the form

$$\Delta E_0 = \Delta mc^2 \tag{14}$$

Einstein had considered a free body at rest with rest energy E_0 that emits two light waves in opposite directions, as indicated in the figure below. By looking at the same process from a slowly moving frame and applying energy conservation he arrived at equation 14, and, in fact, conjectured equation 1 as universal by writing that "the mass of a body is a measure of its energy content."

From our present point of view we can say that the proof was facilitated by the fact that the two-photon system is at rest with respect to the body and therefore it was easy to see that its mass, which is equal to the sum of the energies of the two photons, is Δm.



***Gedanken* experiment** that Einstein described[7] in 1905. A body at rest with rest energy E_0 emits two equal pulses of light in opposite directions. Applying conservation of energy to the process in stationary and slowly moving reference frames leads to the equation $\Delta E_0 = \Delta mc^2$.

You can also find equation 1 as equation 44 in the famous book *The Meaning of Relativity*,[8] which is based on four lectures Einstein gave at Princeton in 1921. (The figure at the right reproduces the relevant page.)

But in between, Einstein was not absolutely consistent in preferring equation 1 to equation 2. In 1906, for example, he rederived[9] Poincaré's formula (equation 2) by considering a photon (to use modern language) that is emitted at one end of a hollow cylinder and absorbed at the other end of the cavity, as indicated in the figure at the top of page 35. Requiring that the center of mass not move, he essentially equated the product of the large mass M of the cylinder and its small displacement l with the product of the small mass m of the photon and its large displacement L, the length of the cylinder:

$$lM = Lm \tag{15}$$

The small displacement l, however, is a product of the photon's time of flight L/c and the cylinder velocity $v = E/(cM)$, where E is the photon's energy and E/c is both its momentum and the momentum of the cylinder. From equation 15 one immediately obtains equation 2. The conclusion of the paper was that light with energy E transfers mass $m = E/c^2$ (which is the correct expression in this thought experiment), and that to any energy E there corresponds a mass equal to E/c^2 (which we now know is not so correct because the photon is massless).

As we understand it today, the subtle point, which Einstein did not discuss in the 1906 paper, was that in special relativity the absorption of a massless particle changes the mass of the absorbing body. Thus a massless photon may "transfer" nonvanishing mass. In absorbing a massless photon, the end of the cylinder becomes heavier, but its mass increase will be E/c^2 only if it is heavy enough that its recoil kinetic energy is negligible. (For the sake of "physical purity," it is better to consider the cylinder as being cut into two "cups.")

The above inconsistent conclusion was extremely fruitful for Einstein's further thinking, which led him finally to general relativity. It implied that a photon possessing inertial mass $m = E/c^2$ has to possess the same gravitational mass and hence has to be attracted by a gravitational force. This idea served as a sort of springboard, as Einstein explained in his "Autobiographical Notes."[10] However, when general relativity was ready, Einstein no longer needed this inconsistent conclusion. This is evidenced by equation 44 in *The Meaning of Relativity*, written 15 years after the 1906 paper.

A few years ago I came across a cartoon that showed Einstein contemplating two equations he had written on a blackboard and then crossed out: $E = ma^2$ and $E = mb^2$. This humorous image of how science is done (reproduced at the bottom of page 35) may be closer to reality than is the usual description in books on the history of relativity,[11] which neglects the striking difference between Einstein's papers of 1905 (with $E_0 = mc^2$) and of 1906 (with $E = mc^2$) and presents a "*coup d'etat*" as a quiet evolution.

'Gravitational mass'

Many physicists still believe that gravitational mass is equal to E/c^2 and quite often use this as an argument in favor of equation 2. Contrary to this belief, the gravitational attraction between two relativistic bodies is determined by their energy–momentum tensors, not just by their ener-

SPECIAL RELATIVITY 51

$$\left.\begin{aligned} I_x &= \frac{m\mathrm{q}_x}{\sqrt{1-q^2}}\\ &\cdot\quad\cdot\quad\cdot\\ &\cdot\quad\cdot\quad\cdot\\ E &= \frac{m}{\sqrt{1-q^2}}\end{aligned}\right\}\quad .\ .\ .\ .\ (43)$$

We recognize, in fact, that these components of momentum agree with those of classical mechanics for velocities which are small compared to that of light. For large velocities the momentum increases more rapidly than linearly with the velocity, so as to become infinite on approaching the velocity of light.

If we apply the last of equations (43) to a material particle at rest ($q = 0$), we see that the energy, E_0, of a body at rest is equal to its mass. Had we chosen the second as our unit of time, we would have obtained

$$E_0 = mc^2 \quad .\ .\ .\ .\ (44)$$

Mass and energy are therefore essentially alike; they are only different expressions for the same thing. The mass of a body is not a constant; it varies with changes in its energy.* We see from the last of equations (43) that E becomes infinite when q approaches 1, the velocity of light. If we develop E in powers of q^2, we obtain,

$$E = m + \frac{m}{2}q^2 + \frac{3}{8}mq^4 + \ldots \quad .\ (45)$$

* The emission of energy in radioactive processes is evidently connected with the fact that the atomic weights are not integers. Attempts have been made to draw conclusions from this concerning the structure and stability of the atomic nuclei.

Page from Einstein's book *The Meaning of Relativity*,[8] which is based on lectures he gave at Princeton University in May 1921. (Copyright by the estate of Albert Einstein; published by Princeton University Press.)

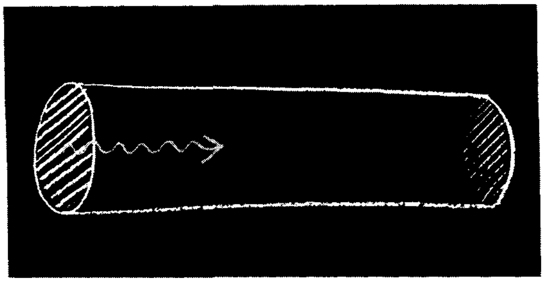
gies. In the simplest case of a light relativistic body such as a photon or an electron of mass m traveling with energy E and velocity $v = \beta c$ in the gravitational field of a very heavy body of mass M such as the Earth or the Sun, the force acting on the light body has the form

$$\mathbf{F}_g = \frac{-G_N M(E/c^2)[\mathbf{r}(1+\beta^2) - \boldsymbol{\beta}(\boldsymbol{\beta}\cdot\mathbf{r})]}{r^3} \tag{16}$$

Here G_N is Newton's constant, $6.7\times10^{-11}\ m^3\,\mathrm{kg}^{-1}\,\mathrm{sec}^{-2}$. When $\beta \ll 1$, equation 16 coincides with the classical expression

$$\mathbf{F}_g = \frac{-G_N M m\mathbf{r}}{r^3}$$

When $\beta \simeq 1$, however, the force is not directed along the radius $\mathbf{r}$: It has a component along $\boldsymbol{\beta}$ as well. So there is no such notion as "relativistic gravitational mass" entering the coefficient of proportionality between $\mathbf{F}_g$ and $\mathbf{r}$. The so-called gravitational mass of a photon falling vertically toward Earth is, incidentally, given by E/c^2. As you can see from equation 16, however, a horizontally moving photon ($\boldsymbol{\beta} \perp \mathbf{r}$) is twice as heavy. (See the figure on page 36). It is this extra factor of 2 that gives the correct angle



A light pulse is emitted at one end of a hollow cylinder and absorbed at the other end in a thought experiment described[9] by Einstein in 1906. Taking E/c as the momentum of the photon and requiring that the center of mass of the system not move leads to the conclusion that light with energy E transfers mass $m = E/c^2$.

of deflection of starlight by the Sun: $\theta = 4G_N M_\odot / R_\odot c^2$. With $M_\odot = 2\times10^{30}$ kg and $R_\odot = 7\times10^8$ m, we get $\theta \simeq 10^{-5}$, in agreement with observations.

I have sketched the changes in Einstein's views during the first two decades of our century. But there were many other important protagonists on the stage.[12] Since the beginning of the century experimenters had tried hard to test equations 8–13 for electrons (beta rays and cathode rays) in various combinations of electric and magnetic fields. According to the standard cliché these experiments were done "to test the velocity dependence of longitudinal and transverse masses," but actually they tested the velocity dependence of momentum. The first results "disproved" relativity theory. Gradually technique improved and agreement started to appear. The confirming results were not terribly convincing, however, as you can see from a letter of 10 November 1922 sent to Einstein by the secretary of the Swedish Academy of Sciences[13]:

> . . . the Royal Academy of Sciences decided to award you last year's Nobel Prize for physics, in consideration of your work on theoretical physics and in particular for your discovery of the law of the photoelectric effect, but without taking into account the value which will be accorded your relativity and gravitation theories after these are confirmed in the future.

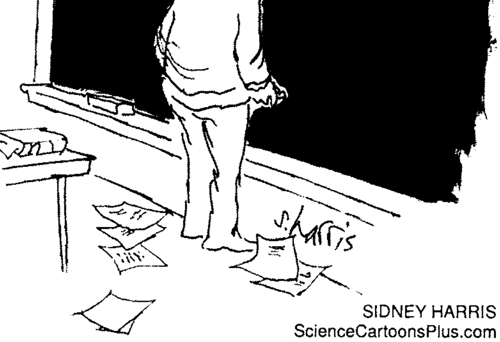
Nor were theorists unanimous in accepting relativity theory or in interpreting its equations. (This article is itself a remote echo of their disputes.) It is well known that the views of Poincaré and Lorentz were different from Einstein's. Important contributions revealing the four-dimensional symmetry of the theory came from Max Planck and especially Hermann Minkowski.[14] But in forming public opinion Gilbert Lewis and Richard Tolman played a particularly important role.[15] It was Tolman who in 1912, starting as before from $\mathbf{p} = m\mathbf{v}$, insisted that m, given by $m_0\gamma$, is *the* mass.[16]

When the 21-year-old student Wolfgang Pauli published in 1921 his encyclopedic article "Relativitätstheorie," which all of us know as the book *The Theory of Relativity*,[17] he discarded the longitudinal and transverse masses as obsolete, but retained the "rest mass" m_0 and the "mass" m, given by $m_0\gamma$, along with the Newtonian relation $\mathbf{p} = m\mathbf{v}$. Pauli's book served as the introduction to relativity for many generations of physicists. It is a great book, but with all its virtues it gave an undesirably long life to the notorious notion that mass depends on velocity, to the term "rest mass" and to the so-called Einstein formula $E = mc^2$.

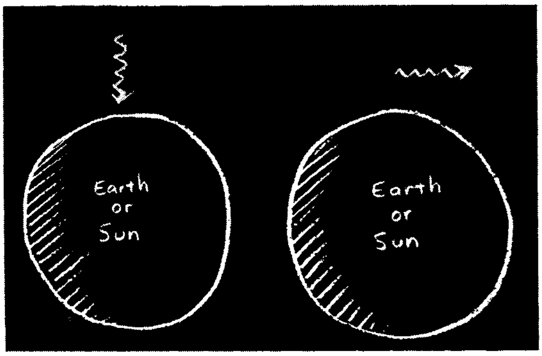
$E = mc^2$ as an element of mass culture

Not only has this terminology flooded the popular science literature and textbooks, but for a long time it dominated most serious monographs on relativistic physics. To my knowledge, the first authors to ignore this archaic terminology consistently were Lev Landau and Evgenii Lifshitz. In their classic 1940 book *The Classical Theory of Fields*, they called the invariant mass by its correct name, mass.[18] They didn't use the term "relativistic mass" or "rest mass." Their language was consistently relativistic.

In 1949 the introduction of Feynman diagrams generalized this relativistic terminology to include antiparticles.[19] Since then all monographs and scientific papers on elementary particles have used consistently relativistic language. Nevertheless, the popular science literature and high school and college textbooks are still full of archaic notions, terms and notation. (One of the rare exceptions is the 1963 book *Spacetime Physics* by Edwin F. Taylor and John A. Wheeler.) As a result we have a kind of pyramid: At the top are books and articles that use consistently relativistic language and are published in thousands of copies; at the bottom are books and articles that use inconsistently relativistic language and are published in the millions. At the top we have $E_0 = mc^2$; at the bottom $E = mc^2$. In between, all four of the equations listed at the beginning of this article peacefully coexist. I have seen many books in which all the notions, consistent and inconsistent, are so mixed up that one is reminded of nightmare cities in which right- and left-side traffic rules apply simultaneously. The situation is aggravated by the fact that even great scientists such as Landau and Feynman, when addressing nonscientists, have sometimes—though not always—used the equation $E = mc^2$. (Compare, for instance, *The Feynman Lectures on Physics*[20] and Feynman's last published



SIDNEY HARRIS
ScienceCartoonsPlus.com



Gravitational force attracting a horizontally moving photon to the Earth or Sun is twice as large as that attracting a vertically moving photon.

lecture, "The Reason for Antiparticles."[21])

The latest example comes from Stephen Hawking's 1988 book *A Brief History of Time.*[23] On the very first page Hawking says: "Someone told me that each equation I included in the book would halve the sales. I therefore resolved not to have any equations at all. In the end, however, I did put in one equation, Einstein's famous equation $E = mc^2$. I hope that this will not scare off half of my potential readers."

I think that in such cases the equation $E = mc^2$, because it is an element of mass culture, is successfully exploited as a kind of "attractor." But the global result of its use is confusion. Readers begin to believe that E/c^2 is a genuine relativistic generalization of inertial and gravitational mass; that whenever you have energy, you have mass (a photon is a counterexample); and that $E = mc^2$ is an inevitable consequence of special relativity (actually it follows from the special and non-natural assumption that $\mathbf{p} = m\mathbf{v}$). The scaffolding used many years ago in the construction of the beautiful building of special relativity has been and is now presented as the central part of the building. The important difference between a Lorentzian scalar and a Lorentzian vector is lost, and with it the four-dimensional symmetry of the theory. The confusion in terminology cannot but lead to confusion in many minds.

"Does Mass Depend on Velocity, Dad?" This is the title of a 1987 *American Journal of Physics* article by Carl Adler.[23] The answers Adler gave to his son were "No!" "Well, yes..." and "Actually, no, but don't tell your teacher." The next day the boy dropped physics. Adler gives several examples of how relativistic mass slowly disappears from university textbooks. There is an interesting quotation in the article from a letter Einstein wrote to Lincoln Barnett in 1948 (the original letter, written in German, is reproduced on page 32):

> It is not good to introduce the concept of the *mass* $M = m/(1 - v^2/c^2)^{1/2}$ of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the "rest mass" m. Instead of introducing M it is better to mention the expression for the momentum and energy of a body in motion.

In the autumn of 1987 I was asked to be a member of a committee set up by what was then the Ministry of Secondary Education to judge a competition for the best physics textbook for secondary schools. I looked over more than a dozen competing books and was shocked to learn that all were promoting the idea that mass increases with velocity and that $E = mc^2$. I was shocked even more when I discovered that my colleagues on the committee—teachers and specialists in teaching physics—had never heard about the equation $E_0 = mc^2$, where E_0 is rest energy and m is mass. I explained this equation to them, and one of them suggested that I write about the topic in *Physics in the School,* a journal for physics teachers. The next day I asked the assistant editor whether the journal would like to publish such an article, and after three months I got a phone call: The editorial board decided it did not want an article that explained special relativity without using $E = mc^2$.

Every year millions of boys and girls throughout the world are taught special relativity in such a way that they miss the essence of the subject. Archaic and confusing notions are hammered into their heads. It is our duty—the duty of professional physicists—to stop this process.

* * *

I am grateful to the members of the physics textbook committee for suggesting that I write this article. I am also grateful to B. M. Bolotovsky, I. Yu. Kobzarev, P. A. Krupchitsky, A. K. Mann, I. S. Tsukerman and M. B. Voloshin for helpful discussions and remarks.

References

1. H. Poincaré, Arch. Neerland. **5**, 252 (1900).
2. A. Einstein, Ann. Phys. (Leipzig) **17**, 891 (1905).
3. H. Lorentz, *The Theory of Electrons*, Brill, Leiden (1909; revised ed. 1916); Dover, New York (1952).
4. J. J. Thomson, Philos. Mag. **11**, 229 (1881).
5. O. Heaviside, Philos. Mag **27**, 324 (1889). G. F. C. Searle, Philos. Mag. **44**, 329 (1897).
6. H. Lorentz, Proc. R. Acad. Sci. Amsterdam **1**, 427 (1899); **6**, 809 (1904).
7. A. Einstein, Ann. Phys. (Leipzig) **18**, 639 (1905).
8. A. Einstein, *The Meaning of Relativity: Four Lectures Delivered at Princeton University, May 1921*, fifth ed., E. P. Adams, translator, Princeton U. P., Princeton, N. J. (1970).
9. A. Einstein, Ann. Phys. (Leipzig) **20**, 371 (1906).
10. A. Einstein, in *Albert Einstein: Philosopher–Scientist*, P. A. Schilpp, ed., Tudor, New York (1949).
11. See, for example, M. Jammer, *Concepts of Mass in Classical and Modern Physics*, Harvard U. P., Cambridge, Mass. (1961); E. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 2, Nelson, London (1953).
12. A. I. Miller, *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911)*, Addison-Wesley, Reading, Mass. (1981).
13. A. Pais, *'Subtle is the Lord...,'* Oxford U. P., New York (1982).
14. M. Planck, Verh. Dtsch. Phys. Ges. **4**, 136 (1906); Sitzungsber. Akad. Wiss. Berlin **13**, 542 (1907). H. Minkowski, Phys. Z. **20**, 104 (1909).
15. G. Lewis, R. Tolman, Philos. Mag. **18**, 510 (1909).
16. R. Tolman, Philos. Mag. **23**, 375 (1912).
17. W. Pauli, *The Theory of Relativity*, G. Field, translator, Pergamon, New York (1958).
18. L. D. Landau, E. M. Lifshitz, *The Classical Theory of Fields*, translation from the second Russian ed., Pergamon, New York (1951).
19. R. Feynman, Phys. Rev. **76**, 749, 769 (1949).
20. R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman Lectures on Physics*, three volumes, Addison-Wesley, Reading, Mass. (1963, 1964, 1965); vol. 1, chaps. 15, 16; vol. 2, chap. 28.
21. R. P. Feynman, in *Elementary Particles and the Laws of Physics*, Cambridge U. P., New York (1987).
22. S. W. Hawking, *A Brief History of Time*, Bantam, New York (1988).
23. C. Adler, Am. J. Phys. **55**, 739 (1987). ■

PUTTING TO REST MASS MISCONCEPTIONS

I am disturbed by the harm that Lev Okun's earnest tirade (June 1989, page 31) against the use of the concept of relativistic mass ("It is our duty ... to stop this process") might do to the teaching of relativity. It might suggest to some who have not thought these matters through that there are unresolved logical difficulties in elementary relativity or that if they use the quantity $m = \gamma m_0$ they commit some physical blunder, whereas in fact this entire ado is about terminology. There are perhaps 40% of us who find it useful ocassionally to write m for E/c^2 and 60% who don't. But why the latter should try to coerce the former beats me.

One can perhaps understand a desire that everyone should use the term "mass" in the same sense. I myself have never found this a stumbling block, since the context tells me which mass the author means. Nevertheless, if I am told by the particle physicists — and they *are* the largest user group of special relativity these days — that henceforth I must use the symbol m for rest mass and call it mass, so be it. But I refuse to stop using the concept of relativistic mass, which I would then denote by m_r.

I know a man who can drive a shift car without ever using the clutch — it's a question of timing. The ingenious Ernest Vincent Wright in 1939 wrote a 50 000-word novel, called *Gadsby*, in all of which the letter *e* never occurs. Its sentences look like this: "A busy day's traffic had had its noisy run." Sure, such feats can be performed. But to what end? I like using my clutch, the letter *e* and the relativistic mass.

To me, $m = \gamma m_0$ is a useful heuristic concept. It gives me a feeling for the magnitude of the momentum $\mathbf{p} = m\mathbf{v}$ at various speeds. The formula $E = mc^2$ reminds me that energy has mass-like properties such as inertia and gravity, and it tells me how energy varies with speed. I will confess to even occasionally using the heuristic concepts of longitudinal mass $\gamma^3 m_0$ and

Reprinted with permission from: *Physics Today*, May 1990, pp. 13, 14, 115 and 117.

transverse mass γm_0 to predict how a particle will move in a given field of force.

Wolfgang Rindler
University of Texas at Dallas

9/89

I read with great interest the article "The Concept of Mass" by Lev B. Okun. As the author points out, the notion of "relativistic mass" is both unnecessary and confusing, and should therefore be avoided.

In this note I would like to go even one step further and suggest that one should also avoid equations such as $d\mathbf{p}/dt = \mathbf{f}$ and $\mathbf{p} = m\gamma\mathbf{v}$, which obscure the Lorentz covariance of the theory because of the appearance of the operator d/dt. Their use is, however, common practice, even in the best textbooks.[1]

The well-known manifestly Lorentz-covariant treatment employs the equation $d\mathbf{P}/d\tau = \mathbf{F}$, where $\mathbf{P} \equiv m\, d\mathbf{X}/d\tau$, τ denoting the proper time and the capitals representing four vectors. Not only is the Lorentz-covariant formulation conceptually clearer; it is also simpler to use in practice due to its geometrical significance.

These points are made clearly in a book by John L. Synge,[2] to which the reader is referred for more details, in particular for an enlightening discussion of the concept of "rest mass."

REFERENCES

1. See, for example, R. Feynman, R. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. I. Addison-Wesley, Reading, Mass. (1966), ch. 15.8.
2. J. Synge, *Relativity: The Special Theory*, North-Holland, Amsterdam (1965), ch. VI.

Michael A. Vandyck
Cork Regional Technical College
Bishopstown, County Cork, Ireland
and Dublin Institute
for Advanced Studies
Dublin, Ireland

7/89

Lev Okun's article on the concept of mass, despite the apologetic tone, is an important one. Many introductory physics and physical science texts include the erroneous concept implied in the equation

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

Students are taught that as the velocity of an object of "rest mass" m_0 gets close to c, the velocity of light, its mass becomes extremely large. The argument is extended to explain why an object of "rest mass" m_0 cannot reach the velocity of light: If $v = c$, then the mass becomes infinite! The same students are also told that particles in cosmic rays could have velocities greater than $0.9c$ without any change in their masses. It is about time we eliminate such contradictory and confusing statements from our textbooks.

Okun's definition, $E_0 = mc^2$, may not eliminate the confusion, as m could still be interpreted as "rest mass" since E_0 is rest energy. To avoid that confusion, I recommend that the mass-energy relation be stated as $E = \gamma mc^2$, where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

For the special case of a reference frame at rest, $v = 0, \gamma = 1$ and $E_{\text{rest}} = mc^2$. Let us eliminate the subscript "0" altogether.

This definition may not gain popularity among the public. On the other hand, physics students will learn the correct concepts of special relativity.

Poovan Murugesan
Stetson University
DeLand, Florida

6/89

The article "The Concept of Mass" contains, in my opinion, a curious mistranslation of Einstein's letter of 19 June 1948 to Lincoln Barnett.

The pertinent part in the caption on page 32 reads:

> It is not good to introduce the concept of the *mass* $M = m/(1-v^2/c^2)^{1/2}$ of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the "rest mass" m. Instead of introducing M it is better to mention the expression for the momentum and energy of a body in motion.

The German word *daneben* does not mean "instead of," but rather "besides," "in addition to" or "moreover." I would therefore translate the passage:

> It is not proper to speak of the *mass* $M = m/(1 - v^2/c^2)^{1/2}$ of a moving body, because no clear definition can be given for M. It is preferable to restrict oneself to the "rest mass" m. Besides, one may well use the expression for momentum and energy when referring to the inertial behavior of rapidly moving bodies.

It should be noted that according to Einstein's letter, the expression for momentum and energy may be used to describe the inertial behavior of rapidly moving bodies, not the motion as such.

I am unqualified to evaluate the article. But it seems to me that the inaccurate translation may not be irrelevant to the argument.

Siegfried Ruschin
Linda Hall Library
Kansas City, Missouri

8/89

I found many of the remarks in the article "The Concept of Mass" thought provoking and interesting. This past semester I took a course where we discussed some of Einstein's theory of special relativity. My professor wished for us to use m_0 as the rest mass and $m = \gamma m_0$ as the relativistic mass in our equations. Our text,[1] however, did not show this notation. This conflict between my professor and the text was a source of great confusion for me. One of my main concerns was that my professor had probably been taught about relativistic mass and then passed the idea along to his students like so many other teachers. Why is it that the academic system has been able to pass along information like this for so long?

Everyone should now fully realize what terminology one should use in explaining mass, and the reasons behind it. The article was a source of greater understanding for me and I hope, for other readers. There are still concepts in physics that confuse me: Does mass truly depend on velocity?

REFERENCE

1. R. Serway, *Physics for Scientists and Engineers*, 2nd ed., Saunders, Philadelphia (1987).

Catherine Sauter
Baltimore, Maryland

7/89

OKUN REPLIES: I fully agree with Siegfried Ruschin that the English text of Einstein's letter does not correspond exactly to the German original. Moreover, I was aware of this circumstance when I published my article. However, both the German manuscript and a copy of the English typescript that Einstein sent to Lincoln Barnett are in the same file of the Einstein archives, and I did not and do not think that an inaccurate translation that did not bother Einstein was worth remarking on in the article.

I agree with Michael A. Vandyck that the most adequate way of presenting special relativity is by writing manifestly Lorentz-covariant equations. However, the connection with Newtonian mechanics is also essential. Nonrelativistic physics is an important (and correct within its realm) part of physics. And it would be a kind of "relativistic extremism" always to start from four-dimensional coordinates and momenta when considering everyday phenomena.

It is not quite clear to me why Poovan Murugesan believes that the equation $E_0 = mc^2$, where E_0 is the rest energy and m is the mass, would provoke a student to call m the rest mass. But I don't think we have a real disagreement here.

In connection with Catherine Sauter's letter I would like to point out that the textbook by Raymond Serway to which she refers contains a modern and concise essay on relativity written by George O. Abel. Curiously, the chapter on relativity in a more recent book cowritten by Serway[1] is based entirely on $m = \gamma m_0$.

Sauter wonders why "the academic system has been able to pass along information like [the idea of relativistic mass] for so long." A partial answer is given by Wolfgang Rindler, who has written an eloquent letter defending the notion of velocity-dependent mass.

My article cannot suggest that there are unresolved logical difficulties in elementary relativity, as Rindler worries, because I stressed that the matter is absolutely clear to all specialists. What is unresolved is the way the subject is treated in many textbooks. I do insist that by using $m = \gamma m_0$ the authors of these textbooks commit a "physical blunder" because they use misleading and confusing terminology.

Rindler writes, "One can perhaps understand a desire that everyone should use the term 'mass' in the same sense." I appreciate his understanding that terms in physics should mean the same thing to everyone and his readiness to "use the symbol m for rest mass and call it mass." However, the equations $m = \gamma m_0$, $\mathbf{p} = m\mathbf{v}$ and $E = mc^2$ — the heuristic value of which Rindler praises — preclude uniform usage of the word "mass." (By the way, it seems to me that $E = \gamma mc^2$ tells how energy varies with speed much better than the "heuristic" $E = mc^2$.)

As for Rindler's statement that understanding the term "mass" in various senses never was a stumbling block for him, it would be extremely strange if it were otherwise. Rindler is not a student, but an expert in relativity. I have seen three editions of his book *Essential Relativity: Special, General and Cosmological* (Springer-Verlag, New York, 1969, 1977, 1979). It is a nice introduction to the subject, except for the sections where the so-called relativistic mass and the whole bunch of other masses are introduced at length. I believe that these sections do present a "stumbling block" to students. If Rindler uses the term "mass" in only one sense in the fourth edition of his book, it will not contain such misleading statements as one that accompanies the equation $m = \gamma(v)m_0$ in section 5.3 of the last edition — "This conclusion is inevitable if momentum is conserved" — or the whole paragraph about "mass-energy equivalence according to the formula $E = mc^2$" in the same section.

The relativistic mass is dear to Rindler's heart. For him not to use it is like driving a car without using the clutch, or writing a novel without using the letter *e*. But the fact is that most leading physics journals, such as *Physical Review, Physical Review Letters* and *Physics Letters*, don't use relativistic mass.

You will not find it in professional books on particle physics. Are all these cars without clutches? To me, texts using the letter m in many senses look like a novel where m stands for $e, m, n \dots$.

It is obvious that Rindler is simply accustomed to the archaic language. But is that enough to justify the promotion of this language to new generations of students?

I would like to use this opportunity to make some remarks concerning equation 16 of my article for gravitational force:

$$\mathbf{F}_g = -\frac{G_N M(E/c^2)[\mathbf{r}\ (1+\boldsymbol{\beta}^2) - \boldsymbol{\beta}(\boldsymbol{\beta}\cdot\mathbf{r})]}{r^3}$$

According to this equation the force acting on a horizontally moving photon is twice as large as that on a vertically moving one. This factor of 2 explains the famous extra factor of 2 in Einstein's expression for the angle of light bending by the Sun. Equation 16 needs some comments, which unfortunately were missing from my article.

It is common knowledge that in a locally inertial frame the gravitational force is equal to zero. That means that equation 16 is valid only for locally non-inertial frames, such as the usual laboratory frame. Moreover, it is important to stress that we are considering the static, spherically symmetric gravitational field of the Sun or the Earth (neglecting rotation) and that the metric chosen to obtain equation 16 is isotropic: $g_{\alpha\beta} = (g_{00}, g_{ik})$, where $g_{ik} = \eta_{ik} f$, so that $\mathrm{d}s^2 = g_{00}x^0x^0 - f(x^1x^1 + x^2x^2 + x^3x^3)$. Using this metric, equation 16 is easily derived as a first-order approximation in the gravitational potential $G_N M/r$ from equation 87.3 of *Field Theory* by Lev D. Landau and Evgenii M. Lifshitz. As the final step of this derivation one has to change to the frame with the usual local rulers and clocks. The choice of an isotropic metric does not permit us to get rid of our force by using equivalence-principle elevators, and therefore one can say that the result for the light bending angle does arise from the global geometry of the central field — the point that is usually stressed in textbooks on gravity.

Engelbert Schucking from New York University has informed me that he has obtained a generalized exact formula for what he calls "the relativistic apple," valid in all approximations with respect to the parameter $G_N M/rc^2$:

$$\mathbf{F}_g = -\frac{G_N M(E/c^2)}{r^3\left(1+\frac{G_N M}{2rc^2}\right)^3}\times$$
$$\times\left[\mathbf{r}\left(1+\beta^2+\frac{G_N M}{2rc^2-G_N M}\right)-\boldsymbol{\beta}(\boldsymbol{\beta}\cdot\mathbf{r})\right]$$

(I'm grateful to Schucking for this communication. I am also grateful to him and to Mikhail Voloshin and Alexander Dolgov for very enlightening discussions.) The first-order approximation in gravitational coupling implicit in equation 16 is very good for the cases of the Sun and the Earth.

I have found equation 16 in only one book.[2] Unfortunately the formula is constructed there semiempirically, and the book itself is full of $E = mc^2$ and all that.

The lack of space and time didn't allow me to discuss in my article such important questions as the mass of a system of particles. I consider this and some other problems in more detail in an extended version of the article.[3]

I don't think we should try to banish $E = mc^2$ from T-shirts, badges and stamps. But in the textbooks it should appear only as an example of a historical artifact, with an explanation of its archaic origin.

REFERENCES

1. R. S. Serway, J. S. Faughn, *Physics*, Saunders, Philadelphia (1989).
2. M. G. Bowler, *Gravitation and Relativity*, Pergamon, Elmsford, N. Y. (1976).
3. L. B. Okun, Usp. Fiz. Nauk. **158**, 511 (1989) [Sov. Phys. Usp. **32**, 629 (1989)].

Lev B. Okun
Institute of Theoretical and Experimental Physics
Moscow, USSR

1/90

This page intentionally left blank

METHODOLOGICAL NOTES

The concept of mass (mass, energy, relativity)

L. B. Okun'

Institute of Theoretical and Experimental Physics, Moscow
Usp. Fiz. Nauk **158**, 511–530 (July 1989)

Present-day ideas concerning the relationship between mass and energy are presented. The history of the origin of archaic terms and concepts that are widely used in the literature in discussing the problem of mass and energy is related, and arguments are presented for the necessity of abandoning these archaic terms and concepts.

1. A small test instead of an introduction

Einstein's relation between the mass of a body and the energy contained in it is undoubtedly the most famous formula of the theory of relativity. It permitted a new and deeper understanding of the world that surrounds us. Its practical consequences are vast and, to a large degree, tragic. In a certain sense this formula has become the symbol of science in the 20th century.

What is the need for yet another paper on this famous relation, on which thousands of papers and hundreds of books have already been written?

Before I answer this question, let us think about the form in which, in your opinion, the physical meaning of the relation between mass and energy is most adequately expressed. Here are four relations:

$$E_0 = mc^2, \qquad (1.1)$$

$$E = mc^2, \qquad (1.2)$$

$$E_0 = m_0c^2, \qquad (1.3)$$

$$E = m_0c^2; \qquad (1.4)$$

where c is the speed of light, E is the total energy of the body, m is its mass, E_0 is the rest energy, and m_0 is the rest mass of the same body. Please write the numbers of these relations in the order in which you regard them as most "correct." Now continue reading.

In the popular scientific literature, school textbooks, and the overwhelming proportion of university textbooks the dominant relation is (1.2) (and its consequence (1.3)), which is usually read from the right to the left and interpreted as follows: The mass of a body increases with its energy—both internal and kinetic.

The overwhelming majority of serious monographs and scientific papers on theoretical physics, particularly on the theoretical physics of elementary particles, for which the special theory of relativity is a working tool, does not contain relations (1.2) and (1.3) at all. According to these books, the mass of a body m does not change when it is in motion and, apart from from the factor c, is equal to the energy contained in the body at rest, i.e., the relation (1.1) is valid. The implication is that both the term "rest mass" and the symbol m_0 are superfluous and therefore are not used. Thus, a kind of pyramid exists where base is formed by the popular science books and school textbooks with press runs of millions, and at whose apex are monographs and papers on the theory of elementary particles with press runs in the thousands.

Between the apex and base of this theoretical pyramid we find a significant number of books and papers in which all three (and even four!) relations coexist peacefully in a mysterious manner. Responsible in the first place for this situation are the theoretical physicists who have not yet explained to large circles of educated people this absolutely simple matter.

The aim of this paper is to explain as simply as possible why the relation (1.1) adequately reflects the essence of the theory of relativity while (1.2) and (1.3) do not, and thus to foster the adoption in the textbook and popular scientific literature of a clear terminology that does not introduce confusion and misunderstandings. In what follows I shall call such terminology the correct terminology. I hope that I shall succeed to convince the reader that the term "rest mass" m_0 is superfluous, that instead of speaking of the "rest mass" m_0 one should speak of the mass m of a body which for ordinary bodies is the same, in the theory of relativity and in Newtonian mechanics, that in both theories the mass m does not depend on the reference frame, that the concept of mass dependent on velocity arose at the beginning of the twentieth century as a result of an unjustified extension of the Newtonian relation between momentum and velocity to the range of velocities comparable to the velocity of light in which it is invalid, and that at the end of the twentieth century one should bid a final farewell to the concept of mass dependent on velocity.

The paper consists of two parts. In part I (Secs. 2–12) the role played by mass in Newtonian mechanics is discussed. We then consider the basic relations of the theory of relativity that connect the energy and momentum of a particle to its mass and velocity, establish the connection between acceleration and force, and give the relativistic expression

0038-5670/89/070629-10$01.80

for the gravitational force. We show how the mass of a system consisting of several particles is defined and consider examples of physical processes that result in a change in the mass of a body or system of bodies, this change being accompanied by the absorption or emission of particles carrying kinetic energy. The first part of the paper ends with a brief account of modern attempts to calculate theoretically the masses of the elementary particles.

In Part II (Secs: 13–20) we discuss the history of the development of the notion of the mass of a body that increases with its energy, the so-called relativistic mass. We show that the use of this archaic concept does not correspond to the four-dimensionally symmetric form of the theory of relativity and leads to numerous confusions in the textbook and popular scientific literature.

I. FACTS

2. Mass in Newtonian mechanics

In Newtonian mechanics, mass possesses a number of important properties and presents, as it were, several faces:

1. Mass is a measure of the amount of matter.
2. The mass of a composite body is equal to the mass of the bodies that constitute it.
3. The mass of an isolated system of bodies is conserved—it does not change with time.
4. The mass of a body does not change on the transition from one system of coordinates to another; in particular, the mass is the same in different inertial systems of coordinates.
5. The mass of a body is a measure of its inertness (inertia).
6. The masses of bodies are the sources of their gravitational attraction to each other.

We discuss in more detail the last two properties of mass.

As a measure of the inertia of a body, the mass m appears in the formula that connects the momentum $\mathbf{p}$ of the body to its velocity $\mathbf{v}$:

$$\mathbf{p}=m\mathbf{v}. \tag{2.1}$$

The mass also occurs in the expression for the kinetic energy E_{kin} of the body:

$$E_{\text{kin}}=\frac{\mathbf{p}^2}{2m}=\frac{m\mathbf{v}^2}{2}. \tag{2.2}$$

By virtue of the homogeneity of space and time the momentum and energy of a free body are conserved in an inertial system of coordinates. The momentum of a given body changes with the time only under the influence of other bodies:

$$\frac{d\mathbf{p}}{dt}=\mathbf{F}, \tag{2.3}$$

where $\mathbf{F}$ is the force that acts on the body. If it is borne in mind that in accordance with the definition of the acceleration $\mathbf{a}$

$$\mathbf{a}=\frac{d\mathbf{v}}{dt}, \tag{2.4}$$

and allowance is made for the relations (2.1) and (2.3), then

$$\mathbf{F}=m\mathbf{a}. \tag{2.5}$$

In this relation the mass again appears as a measure of inertia. Thus, in Newtonian mechanics mass as a measure of inertia is determined by two relations: (2.1) and (2.5). Some authors prefer to define the measure of inertia by the relations (2.1), others by the relation (2.5). For the subject of our paper it is only important that these two definitions are compatible in Newtonian mechanics.

We now turn to gravitation. The potential energy of the attraction between two bodies with masses M and m (for example, the earth and a stone) is

$$U_g=-\frac{GMm}{r}, \tag{2.6}$$

where $G=6.7\cdot10^{-11}\mathrm{N\cdot m^2\cdot kg^{-2}}$ (we recall that $1N=1$ $\mathrm{kg\cdot m\cdot sec^{-2}}$). The force with which the earth attracts the stone is

$$\mathbf{F}_g=-\frac{GMm\mathbf{r}}{r^3}, \tag{2.7}$$

where the radius vector $\mathbf{r}$, which joins the centers of mass of the bodies, is directed from the earth to the stone. (The stone attracts the earth with a force of the same magnitude but opposite direction.) It follows from (2.7) and (2.5) that the acceleration of a body falling freely in a gravitational field does not depend on its mass. The acceleration in the field of the earth is usually denoted by g:

$$\mathbf{g}=\frac{\mathbf{F}_g}{m}=-\frac{GM\mathbf{r}}{r^3}. \tag{2.8}$$

Substituting the mass and radius of the earth ($M_\oplus\simeq6\cdot10^{24}$ kg, $R_\oplus\simeq6.4.10^6$ m) in(2.9), we readily obtain the estimate $g\approx9.8$ m/sec.

The universality of g was first established by Galileo, who concluded that the acceleration of a falling sphere depended neither on its mass nor the material of which it was made. This independence was verified to a very high degree of accuracy at the beginning of the 20th century by Eötvös and in a number of recent experiments. The fact that the gravitational acceleration does not depend on the mass of the accelerated body is usually characterized in school physics courses as equality of the inertial and gravitational masses, by which it is meant that one and the same quantity m occurs in the relation (2.5), on the one hand, and (2.6) and (2.7), on the other.

We shall not discuss here the other properties of mass listed at the beginning of this section, since they appear obvious from the point of view of common sense. In particular, no one doubts that the mass of a vase is equal to the sum of its fragments:

$$m=\sum_i m_i. \tag{2.9}$$

Nor does anyone doubt that the mass of two automobiles is equal to the sum of their masses irrespective of whether they are parked or race toward each other at maximal speed.

3. Galileo's principle of relativity

If one does not go into the actual formulas, one can say that the quintessence of Newtonian mechanics is the principle of relativity.

In one of Galileo's books there is a brilliant argument about the fact that in the cabin of a ship with portholes closed it is not possible to detect by any mechanical experiments a uniform and rectilinear motion of the ship relative to the shore. In giving this example, Galileo emphasized that no mechanical experiments could distinguish one inertial frame of reference from another. This assertion became known as Galileo's principle of relativity. Mathematically, this principle is expressed by the fact that the equations of Newtonian mechanics do not change on the transition to new coordinates: $\mathbf{r}\to\mathbf{r}'=\mathbf{r}-\mathbf{V}t$ $t\to t'=t$, where $\mathbf{V}$ is the velocity of the new inertial system with respect to the original one.

4. Einstein's principle of relativity

At the beginning of the 20th century a more general principle, which became known as Einstein's principle of relativity, was formulated. According to this principle not only mechanical but also all other experiments (optical, electrical, magnetic, etc.) are incapable of distinguishing one inertial system from another. The theory constructed on this principle has become known as the theory of relativity, or relativistic theory.

Relativistic theory, in contrast to nonrelativistic (Newtonian) mechanics, takes account of the fact that there exists in nature a limiting speed c of the propagation of physical signals: $c=3\cdot 10^8$ m/sec.

Usually, c is called the speed of light in vacuum. Relativistic theory makes it possible to calculate the motion of bodies (particles) with all speeds v up to $v=c$. Nonrelativistic Newtonian mechanics is the limiting case of Einstein's relativistic mechanics as $v/c\to 0$. Formally, in Newtonian mechanics there is no limit to the speed of propagation of signals, i.e., $c=\infty$.

The introduction of Einstein's principle of relativity required a modification in our view of fundamental concepts such as space, time, and simultaneity. It was found that, considered separately, the distances between two events in space, $\mathbf{r}$, and in time, t, do not remain unchanged on the transition from one inertial frame of reference to another but behave like the components of a four-dimensional vector in a four-dimensional Minkowski spacetime. All that remains unchanged, invariant, is the interval s: $s^2=c^2t^2-\mathbf{r}^2$.

5. Energy, momentum, and mass in the theory of relativity

The fundamental relations of the theory of relativity for a freely moving particle (system of particles, body) are

$$E^2-\mathbf{p}^2c^2=m^2c^4, \tag{5.1}$$

$$\mathbf{p}=\frac{\mathbf{v}E}{c^2}, \tag{5.2}$$

where E is the energy, $\mathbf{p}$ is the momentum, m is the mass, and $\mathbf{v}$ is the velocity of the particle (or system of particles, or body). It should be emphasized that the mass m and the velocity $\mathbf{v}$ for a particle or a body are the same quantities with which we deal in Newtonian mechanics. Like the four-dimensional coordinates t and $\mathbf{r}$, the energy E and the momentum $\mathbf{p}$ are the components of a four-dimensional vector. They change on the transition from one inertial system to another in accordance with the Lorentz transformations. The mass, however, is not changed—it is a Lorentz invariant.

It should be emphasized that, as in Newtonian mechanics, in the theory of relativity there are laws of conservation of the energy and momentum of an isolated particle or an isolated system of particles.

In addition, as in Newtonian mechanics, the energy and momentum are additive—the total energy and total momentum of n free particles are, respectively,

$$E=\sum_{i=1}^{n}E_i, \qquad \mathbf{p}=\sum_{i=1}^{n}\mathbf{p}_i. \tag{5.3}$$

With regard to the mass, in theory of relativity the mass of an isolated system is conserved (does not change with the time), but does not possess the property of additivity (see below).

The most important difference of the theory of relativity from nonrelativistic mechanics is that the energy of a massive body does not vanish even when the body is at rest, i.e., for $\mathbf{v}=0$, $\mathbf{p}=0$. As can be seen from (5.1), the rest energy of a body (it is usually denoted by E_0) is proportional to its mass:

$$E_0=mc^2. \tag{5.4}$$

Indeed, the assertion that inert matter at rest hides within it a huge (by virtue of the square of the limiting velocity c) store of energy, made by Einstein in 1905, is the main practical consequence of the theory of relativity. All nuclear energy and all military nuclear technology is based on the relation (5.4). It may not be quite so well known that all ordinary energy production is based on the same relation.

6. Limiting cases of the relativistic equations

It is a remarkable property of Eqs. (5.1) and (5.2) that they describe the motion of particles in the complete interval of speeds: $0\leqslant v\leqslant c$. In particular, for $v=c$ it follows from (5.2) that

$$pc=E. \tag{6.1}$$

Substituting this relation in (5.1), we conclude that if a particle moves with speed c, then its mass is equal to zero, and vice versa. For a massless particle there is no coordinate system in which it is at rest—it "can only dream" of rest.

For massive particles (as we shall call all particles with nonzero mass, even if they are very light) the relations for the energy and momentum can be conveniently expressed in terms of the mass and velocity. For this we substitute (5.2) in (5.1):

$$E^2\left(1-\frac{v^2}{c^2}\right)=m^2c^4, \tag{6.2}$$

and, taking the square root, we obtain

$$E=mc^2\left(1-\frac{v^2}{c^2}\right)^{-1/2}. \tag{6.3}$$

Substituting (6.3) in (5.2), we obtain

$$\mathbf{p}=m\mathbf{v}\left(1-\frac{v^2}{c^2}\right)^{-1/2}. \tag{6.4}$$

It is obvious from (6.3) and (6.4) that a massive body (with $m\neq 0$) cannot move with the speed of light, since then the energy and momentum of the body would have to be infinite.

In the literature on the theory of relativity it is customary to use the notation

$$\beta=\frac{v}{c}, \tag{6.5}$$

and

$$\gamma=(1-\beta^2)^{-1/2}. \tag{6.6}$$

Using γ, we can express E and $\mathbf{p}$ in the form

$$E=mc^2\gamma, \tag{6.7}$$

$$\mathbf{p} = m\mathbf{v}\gamma. \tag{6.8}$$

We define the kinetic energy E_{kin} as the difference between the total energy E and the rest energy E_0:

$$E_{\text{kin}} = E - E_0 = mc^2(\gamma - 1). \tag{6.9}$$

In the limit when $v/c \ll 1$, only the first terms in the series in β need be retained in the expressions (6.8) and (6.9). Then we return in a natural manner to the equations of Newtonian mechanics:

$$\mathbf{p} = m\mathbf{v}, \tag{6.10}$$

$$E_{\text{kin}} = \frac{\mathbf{p}^2}{2m} = \frac{m\mathbf{v}^2}{2}, \tag{6.11}$$

from which it can be seen that the mass of the body in Newtonian mechanics and the mass of the same body in relativistic mechanics are the same quantity.

7. Connection between the force and acceleration in the theory of relativity

One can show that in the theory of relativity the Newtonian relation between the force F and the change in the momentum remains the same:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}. \tag{7.1}$$

Using the relation (7.1) and the definition of acceleration,

$$\mathbf{a} = \frac{d\mathbf{v}}{dt}, \tag{7.2}$$

we readily obtain

$$\mathbf{F} = m\gamma\mathbf{a} + m\gamma^3\boldsymbol{\beta}(\boldsymbol{\beta}\mathbf{a}). \tag{7.3}$$

We see that, in contrast to the nonrelativistic case, the acceleration in the relativistic case is not directed along the force but also has a component along the velocity. Multiplying (7.3) by $\mathbf{v}$, we find

$$\mathbf{a}\mathbf{v} = \frac{\mathbf{F}\mathbf{v}}{m\gamma(1+\gamma^2\beta^2)} = \frac{\mathbf{F}\mathbf{v}}{m\gamma^3}. \tag{7.4}$$

Substituting this in (7.3), we obtain

$$F - (F\boldsymbol{\beta})\boldsymbol{\beta} = m\gamma\mathbf{a}. \tag{7.5}$$

Despite the unusual appearance of Eq. (7.3) from the point of view of Newtonian mechanics (we should say, rather, precisely because of this unusual appearance), this equation correctly describes the motion of relativistic particles. From the beginning of the century it was frequently submitted to experimental verification for different configurations of the electric and magnetic fields. This equation is the foundation of the engineering calculations for relativistic accelerators.

Thus, if $\mathbf{F}\perp\mathbf{v}$, then

$$\mathbf{F} = m\gamma\mathbf{a}, \tag{7.6}$$

but if $\mathbf{F}\parallel\mathbf{v}$, then

$$\mathbf{F} = m\gamma^3\mathbf{a}. \tag{7.7}$$

Thus, if one attempts to define an "inertial mass" as the ratio of the force to the acceleration, then in the theory of relativity this quantity depends on the direction of the force relative to the velocity, and therefore cannot be unambiguously defined. Consideration of the gravitational interaction leads to the same conclusion with regard to the "gravitational mass."

8. Gravitational attraction in the theory of relativity

Whereas in Newtonian theory the force of the gravitational interaction is determined by the masses of the interacting bodies, in the relativistic case the situation is much more complicated. The point is that the source of the gravitational field is a complicated quantity possessing ten different components—the so-called energy-momentum tensor of the body. (For comparison we point out that the source of the electromagnetic field is the electromagnetic current, which is a four-dimensional vector and has four components.

We consider a very simple example, when one of the bodies has a very large mass M and is at rest (for example, the sun or the earth), and the other has a very small or even zero mass, for example, an electron or photon with energy E. On the basis of the general theory of relativity, one can show that in this case the force acting on a light particle is

$$\mathbf{F} = -GM\frac{E}{c^2}[(1+\boldsymbol{\beta}^2)\mathbf{r} - (\mathbf{r}\boldsymbol{\beta})\boldsymbol{\beta}]r^{-3}. \tag{8.1}$$

It is easy to see that for a slow electron, with $\beta \ll 1$, the expression in the square bracket reduces to $\mathbf{r}$, and, bearing in mind that $E_0/c^2 = m$, we return to Newton's nonrelativistic formula. However, for $v/c \sim 1$ or $v/c = 1$ we encounter a fundamentally new phenomenon, namely, the quantity that plays the role of the "gravitational mass" of the relativistic particle depends not only on its energy but also on the mutual direction of the vectors $\mathbf{r}$ and $\mathbf{v}$. If $\mathbf{v}\parallel\mathbf{r}$, then the "gravitational mass" is E/c^2, but is $\mathbf{v}\perp\mathbf{r}$, it is $(E/c^2)(1+\beta^2)$, and for a photon $2E/c^2$.

We use the quotation marks to emphasize that for a relativistic body the concept of gravitational mass is invalid. It is meaningless to speak of the gravitational mass of a photon if for a vertically falling photon this quantity is half that for one traveling horizontally.

Having discussed different aspects of the dynamics of a single relativistic particle, we now turn to the question of the mass of a system of particles.

9. Mass of a system of particles

We have already noted above that in the theory of relativity the mass of a system is not equal to the mass of the bodies that make up the system. This assertion can be illustrated by several examples.

1. Consider two photons moving in opposite directions with equal energies E. The total momentum of such a system is zero, and the total energy (it is the rest energy of the system of the two photons) is $2E$. Therefore, the mass of this system is $2E/c^2$. It is easy to show that a system of two photons will have zero mass only when they move in the same direction.

2. We consider a system consisting of n bodies. The mass of this system is determined by

$$m = \left[\left(\sum_{i=1}^{n}\frac{E_i}{c^2}\right)^2 - \left(\sum_{i=1}^{n}\frac{\mathbf{p}_i}{c}\right)^2\right]^{1/2}, \tag{9.1}$$

where $\Sigma_i E_i$ is the sum of the energies of these bodies, and $\Sigma\,\mathbf{p}_i$ is the vector sum of their momenta.

The first two examples are characterized by being systems of free particles; the sizes of these systems increase without limit with the time as the particles that constitute them move away from each other. We now consider systems whose sizes remain unchanged.

3. We consider a hydrogen atom consisting of a proton and an electron. To a good accuracy, the rest energy E_0 of the atom can be represented as a sum of four terms:

$$E_0 = m_p c^2 + m_e c^2 + E_{kin} + U, \tag{9.2}$$

where m_p is the mass of the proton, m_e is the mass of the electron, and E_{kin} and U are the kinetic and potential energies of the electron.

The potential energy U is due to the mutual attraction of the electric charges of the proton and electron, which prevents the electron from leaving the proton. From a theory exhaustively tested by experiment it follows that

$$E_{kin} + U = -E_{kin} = -\frac{1}{2} m_e v_e^2, \tag{9.3}$$

where $v_e \approx c/137$ is the speed of the electron in the hydrogen atom. Hence

$$m_H = \frac{E_0}{c^2} = m_p + m_e - \frac{m_e v_e^2}{2c^2}. \tag{9.4}$$

Thus, the mass of the hydrogen atom is less than $m_p + m_e$ by a few millionths of the electron mass.

4. Let us consider the deuteron, the nucleus of the heavy isotope of hydrogen, consisting of a proton and a neutron. The proton and neutron attract each other more strongly and move more rapidly than the electron in the hydrogen atom. As a result, the mass of the deuteron is about 0.1% less than the sum of the masses of the proton and neutron.

Essentially, we treated the last two examples on the basis of nonrelativistic mechanics, since the considered mass differences, or, as they are called, the mass defects, are, although important, fairly small compared with the masses themselves.

Now is the time to recall the broken vase mentioned in Sec. 2. The sum of the masses of the fragments is equal to the mass of the vase to the accuracy with which the binding energy of these fragments is small compared with their rest energy.

10. Examples of transformations into each other of rest energy and kinetic energy

In nuclear or chemical reactions the rest energy must, by virtue of the law of conservation of energy, be transformed into the kinetic energy of the reaction of products if the total mass of the particles that interact is greater than the total mass of the reaction products. We consider four examples:

1. When an electron and positron annihilate into two photons, the entire rest energy of the electron and positron is transformed into the kinetic energy of the photons.

2. As a result of thermonuclear reactions taking place in the sun, there are transformations of two electrons and four protons into a helium nucleus and two neutrinos:

$$2e + 4p \rightarrow {}^4He + 2\nu + E_{kin}. \tag{10.1}$$

The energy released is $E_{kin} = 29.3$ MeV. Remembering that the mass of the proton is 938 MeV and the mass of the electron 0.5 MeV, the relative decrease of the mass is of the order of a percent ($\Delta m/m = 0.8 \cdot 10^{-2}$).

3. If a slow neutron collides with a ^{235}U nucleus, the nucleus breaks up into two fragments and also emits two or three neutrons capable of striking other uranium nuclei, and an energy $E_{kin} \approx 200$ MeV is released. In this case, as is readily seen, $\Delta m/m = 0.9 \cdot 10^{-3}$.

4. In the combustion reaction of methane in the gas burner of a kitchen stove,

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \tag{10.2}$$

an energy equal to 35.6 MJ per cubic meter of methane is released. Since the density of methane is 0.89 kg/m^3, we can readily see that in this case $\Delta m/m = 10^{-10}$. In chemical reactions $\Delta m/m$ is 7–8 orders of magnitude less than in nuclear reactions, but the essence of the mechanism of energy release is the same—rest energy is transformed into kinetic energy.

To emphasize that the mass of a body changes whenever its internal energy changes, we consider two common examples:

1) if a flat iron is heated to 200°, its mass increases by $\Delta m/m = 10^{-12}$ (this is readily estimated using the specific heat 450 J·kg^{-1}·deg^{-1} of iron);

2) if a certain amount of ice is transformed completely into water, $\Delta m/m = 3.7 \cdot 10^{-12}$.

11. Comparison of the role played by mass in the theories of Einstein and Newton

Summarizing what was said above, it is expedient to compare the role played by mass in the mechanics of Einstein and of Newton.

1. In the theory of relativity, in contrast to Newtonian mechanics, the mass of a system is not a measure of the amount of matter. In relativistic theory the very concept of matter is much richer than in nonrelativistic theory. In relativistic theory there is no fundamental difference between matter (protons, neutrons, electrons) and radiation (photons).

Protons, neutrons, electrons, and photons are the most commonly encountered representatives in nature of the large family of so-called elementary particles. It is possible that the photons are not the only particles having zero mass. For example, certain types of neutrinos could also have zero mass. It is also possible that there exist other massless particles that have not yet been detected because of the great difficulty of detecting them by means of existing instruments.

2. In nonrelativistic theory, the more individual particles (atoms) a system (a scale weight) contains, the greater its mass. In relativistic theory, when the energies of particles are very large compared with their masses, the mass of a system of particles is determined not so much by their number as by their energies and mutual orientations of their momenta. The mass of a composite body is not equal to the sum of the masses of the bodies that constitute it.

3. As in Newtonian mechanics, the mass of an isolated system of bodies is conserved, i.e., does not change with time. However, it is now necessary to include among the bodies not only "matter," say atoms, but also "radiation" (photons).

4. As in Newtonian mechanics, in the theory of relativity the mass of a body does not change on the transition from one inertial frame of reference to another.

5. The mass of a relativistically moving body is not a

measure of its inertia. Indeed, a single measure of inertia for relativistically moving bodies does not exist at all, since the resistance of a body to the force accelerating it depends on the angle between the force and the velocity.

6. The mass of a relativistically moving body does not determine its interaction with the gravitational field. This interaction is determined by an expression that depends on the energy and momentum of the body.

Despite these four "noes" the mass of a body is also an extremely important property in the theory of relativity. A vanishing mass means that the "body" must always move with the speed of light. A nonvanishing mass characterizes the mechanics of a body in a frame of reference in which it moves slowly or is at rest. This frame of reference is distinguished compared with other inertial systems.

7. According to the theory of relativity, the mass of a particle is a measure of the energy "sleeping" in the particle at rest; it is a measure of the rest energy: $E_0 = mc^2$. This property of mass was unknown in nonrelativistic mechanics.

The mass of an elementary particle is one its most important characteristics. Attempts are made to measure it as accurately as possible. For stable or long-lived particles the mass is determined by independent measurement of the energy and momentum of the particle and application of the formula $m^2 = (E^2/c^4) - (\mathbf{p}^2/c^2)$. The masses of short-lived particles are determined by measuring the energies and momenta of the particles produced by their decay or of particles that are "present" when they are produced.

Information about the masses of all elementary particles together with their other properties (lifetime, spin, decay modes) is contained in reference collections that are regularly updated.

12. The nature of mass: Question No. 1 of modern physics

During recent decades great progress has been made in understanding the properties of elementary particles. We have seen the construction of quantum electrodynamics—the theory of the interaction of electrons with photons, and the foundations have been laid of quantum chromodynamics—the theory of the interaction of quarks with gluons and of the theory of the electroweak interaction. In all these theories the particles that transmit the interactions are the so-called vector bosons—particles that have spin equal to unity: the photon, gluons, and the W and Z bosons.

As regards the masses of the particles, the achievements here are much more modest. At the turn of the 19th and 20th centuries it was believed that mass could have a purely electromagnetic origin, at least for the electron. We know today that the electromagnetic fraction of the mass of the electron is of the order of a percent of its total mass. We know that the main contribution to the masses of the protons and neutrons are made by the strong interactions mediated by gluons and not the masses of the quarks that are present in protons and neutrons.

But we know absolutely nothing of what produces the masses of the six leptons (electron, neutrino, and four further particles analogous to them) and six quarks (of which the first three are significantly lighter than the proton, the fourth is somewhat lighter than the proton, the fifth is five times heavier, while the sixth is so massive that the attempts to produce and detect it have hitherto failed).

There are theoretical guesses that hypothetical particles with spin zero play a decisive role in creating the masses of the leptons and quarks, and also of the W and Z bosons. The searches for these particles represent one of the fundamental problems of high-energy physics.

II. ARTIFACTS

13. At the turn of the century: Four masses

Everything that has been said in the first part of this paper is well known to any theoretical physicist who has ever dealt with the special theory of relativity. On the other hand, any physicist (and not only a physicist) has heard of Einstein's "famous" relation $E = mc^2$. It is therefore natural to ask how it comes about that there is a peaceful coexistence of mutually exclusive formulas in the literature and in the minds of readers:

$$E_0 = mc^2,$$

$$E = mc^2.$$

Before we seek to answer this question, we recall once more that in accordance with the first formula the rest energy E_0 corresponds to the mass of a body at rest, while according to the second any body with energy E has mass E/c^2. According to the first, the mass of a body does not change when it is in motion. According to the second, the mass of the body increases with increasing velocity of the body. According to the first, the photon is massless, but according to the second it has a mass equal to E/c^2.

To answer the question we have posed about the coexistence of the formulas, we must examine the history of the creation, interpretation, and recognition of the special theory of relativity.

In discussions of the connection between mass and energy, the starting point is usually taken to be the paper of J. J. Thomson[1] published in 1881. In this paper, the first attempt was made to estimate the contribution to the inertial mass of an electrically charged-body made by the energy of the electromagnetic field of this body.

The creation of the theory of relativity is usually associated with Einstein's 1905 paper[2] in which the relativity of simultaneity was clearly formulated. But, of course, the work on the creation and interpretation of the theory began long before 1905 and continued long after that date.

If one speaks of interpretation, the process must still be regarded as continuing today. Otherwise it would not be necessary to write this paper. As regards recognition, one can say that even at the end of 1922, when Einstein was awarded the Nobel prize, the theory of relativity was not generally accepted.

The secretary of the Swedish Academy of Sciences wrote to Einstein that the Academy had awarded him the Nobel prize for the discovery of the law of the photoelectric effect "but without taking into account the value which will be accorded your relativity and gravitation theories after these are confirmed in the future" (quoted from Pais's book[3]).

The formula $E = mc^2$ appeared in 1900 before the creation of the theory of relativity. It was written down by Poincaré whose point of departure was that a plane light wave carrying energy E has a momentum $\mathbf{p}$ of absolute magnitude that, in accordance with Poynting's theorem, is E/c. Using Newton's nonrelativistic formula for the momentum,

$\mathbf{p} = m\mathbf{v}$, and the fact that for light $v = c\mathbf{c}$, Poincaré[4] concluded that a photon must possess an inertial mass $m = E/c^2$.

Already a year before this, in 1899, Lorentz[5] had first introduced the concept of longitudinal and transverse masses of ions, the first of which increases with the velocity as γ^3, the other as γ. He arrived at this conclusion by using the Newtonian relation between the force and the acceleration, $\mathbf{F} = m\mathbf{a}$. A detailed consideration of these masses for electrons is contained in his paper[6] published in 1904.

Thus, at the turn of the century, there arose, through, as we now understand, the incorrect use of nonrelativistic equations to describe relativistic objects, a family of "masses" that increase with the energy of the body:

"relativistic mass" $m = E/c^2$,

"transverse mass" $m_t = m\gamma$,

"longitudinal mass" $m_l = m\gamma^3$.

Note that for $m \neq 0$ the relativistic mass is equal to the transverse mass, but, in contrast to the latter, it also exists for massless bodies, for which $m = 0$. We here use the letter m in the usual sense, since we used it in the first part of this paper. But all physicists during the first five years of this century, i.e., before the creation of the theory of relativity, and many after the creation of that theory called the relativistic mass the mass and denoted it by the letter m, as did Poincaré in his 1900 paper. And then there must necessarily arise, and did arise, a further, fourth term: the "rest mass," which was denoted by m_0. The term "rest mass" came to be used for the ordinary mass, which, in a consistent exposition of the theory of relativity, is denoted m.

This is the origin of the "gang of four,' which successfully established itself in the incipient theory of relativity. Thus were created the prerequisites for the confusion that continues to the present day.

From 1900 special experiments were made with β rays and cathode rays, i.e., energetic electrons, beams of which were deflected by magnetic and electric fields (see Miller's book[7]).

These experiments were called experiments to measure the velocity dependence of the mass, and during almost the whole of the first decade of our century their results did not agree with the expressions for m_t and m_l obtained by Lorentz and, essentially, refuted the theory of relativity and were in good agreement with the incorrect theory of Abraham. Subsequently, agreement with the formulas of Lorentz was established, but it can be seen from the letter of the secretary of the Swedish Academy of Sciences quoted earlier that it did not appear absolutely convincing.

14. Mass and Energy in Einstein's papers in 1905

In Einstein's first paper on the theory of relativity,[2] he, like everyone at that time, used the concepts of longitudinal and transverse mass, but did not denote them by special symbols, while for the kinetic energy W he obtained the relation

$$W = \mu V^2 \left\{ \frac{1}{[1-(v^2/V^2)]^{1/2}} - 1 \right\},$$

where μ is the mass, and v is the speed of light. Thus, he did not use the concept of "rest mass."

In the same 1905 Einstein published a short note[8] in which he concluded that "the mass of a body is a measure of the energy contained in it.' If we use modern notation, this conclusion is expressed by the formula

$$E_0 = mc^2.$$

Actually, the symbol E_0 occurs already in the first phrase with which the proof begins: "Suppose that in the system (x,y,z) there is a body at rest whose energy, referred to the system (x,y,z) is E_0." This body radiates two plane light waves with equal energies $L/2$ in opposite directions. Considering this process in a system moving with velocity v using the circumstance that in this system the total energy of the photons is $L(\gamma - 1)$, and equating it to the difference of the kinetic energies of the body before and after the emission, Einstein concluded that "if the body gives up energy L in the form of radiation, its mass is reduced by L/V^2," i.e., $\Delta m = \Delta E_0/c^2$. Thus, in this paper he introduced the concept of rest energy of the body and established equivalence between the mass of the body and the rest energy.

15. "Generalized Poincaré formula"

If in the 1905 paper Einstein was completely clear, in his subsequent paper[9] of 1906 the clarity is somewhat lost. Referring to the 1900 paper of Poincaré that we mentioned earlier, Einstein proposed a more transparent proof of Poincaré's conclusion and asserted that to every energy E there corresponds an inertia E/V^2 (inertial mass E/V^2, where V is the speed of light), and he ascribed "to the electromagnetic field a mass density (ρ_e) that differs from the energy density by a factor $1/V^2$." Moreover, it can be seen from the text of the paper[9] that he regards these assertions as a development of his 1905 paper. And although in the paper[10] that appeared in 1907 Einstein again clearly speaks of the equivalence of mass and the rest energy of a body (§11), he does not draw a clear distinction between the relativistic formula $E_0 = mc^2$ and the prerelativistic formula $E = mc^2$ and in his paper[11] "On the influence of gravitation on the propagation of light" he wrote: "... If the increment of the energy is E, then the increment of the inertial mass is E/c^2."

At the end of the first decade of this century, Planck[12,13] and Minkowski[14] played an important part in creating the modern unified four-dimensional spacetime formalism of the theory of relativity. At about the same time, in the papers of Lewis and Tolman[15,16] the "prerelativistic mass," equal to E/c^2, was finally elevated to the throne of the theory of relativity. It received the title "relativistic mass" and, most unfortunate of all, usurped the simple name of "mass." Meanwhile, the true mass suffered Cinderella's fate and was given the nickname "rest mass." Lewis and Tolman based their papers on the Newtonian definition $\mathbf{p} = m\mathbf{v}$ of momentum and the law of conservation of "mass," which in essence was the law of conservation of energy, divided by c^2.

It is remarkable that in the literature on the theory of relativity this "palace revolution" has remained unnoted, and the development of the theory of relativity has been represented as a logically consistent process. In particular, the historians of physics (for example, in the books of Refs. 3, 7, 17, and 18) do not note the fundamental difference between Einstein's paper of Ref. 8, on the one hand, and the papers of Poincaré[4] and Einstein[9] on the other.

I once saw a cartoon representing the process of scientific creativity. A scientist, with a back like Einstein, is writing at the blackboard. He has written $E = ma^2$, crossed it out, below that $E = mb^2$, and again crossed it out, and, finally, still lower; $E = mc^2$. A humorous trifle, but this cartoon may be nearer the truth than the received description of the process of scientific creation as a continuous logical development.

It is not by chance that I mentioned Cinderella. A mass that increased with speed—that was truly incomprehensible and symbolized the depth and grandeur of science, bewitching the imagination. Compared with it, what was ordinary mass, so simple, so comprehensible!

16. A thousand and two books

The title of this section is not to be taken literally as giving the total number of books which discuss the theory of relativity, which I do not know. The number is certainly greater than several hundred, and may be a thousand. However, two books that appeared at the beginning of the twenties need to be considered especially. They are both very famous and have been admired by more than one generation of physicists. The first is the encyclopaedic monograph of the 20-year-old student Wolfgang Pauli "Relativitätstheorie",[19] which appeared in 1921. The second is the *The Meaning of Relativity*,[20] published in 1922 by the creator of the special and general theories himself—Albert Einstein. In these two books the question of the connection between energy and mass is treated in radically different ways.

Pauli decisively rejects, as obsolete, the longitudinal and transverse masses (and with it the formula $\mathbf{F} = ,m\mathbf{a}$) but regards it as "expedient" to use the formula $\mathbf{p} = m\mathbf{v}$ and, therefore, the concept of a velocity-dependent mass, to which he devotes several sections. He gives much space to the "law of equivalence of mass and energy" or, as he calls it, "the law of inertia of energy of any form," according to which "to every energy there corresponds a mass $m = E/c^2$."

In contrast to Pauli, Einstein denotes the ordinary mass by the letter m. Expressing the four-dimensional energy-momentum vector in terms of m and the velocity of the body, Einstein then considers a body at rest and concludes that "the energy, E_0, of a body at rest is equal to its mass." It should be noted that he had earlier adopted c as the unit of velocity. Further, he wrote: "Had we chosen the second as our unit of time, we would have obtained

$$E_0 = mc^2. \qquad (44)$$

Mass and energy are therefore essentially alike; they are only different expressions for the same thing. The mass of a body is not a constant; it varies with changes in its energy." The two last phrases acquire an unambiguous meaning through the introductory word "therefore" and the circumstance that they follow directly after the equation $E_0 = mc^2$. Thus, a mass that depends on the velocity is not found in the book *The Meaning of Relativity*.

It is possible that if Einstein had commented in more detail and systematically on his equation $E_0 = mc^2$, the equation $E = mc^2$ would have disappeared from the literature already in the twenties. But that he did not do, and the majority of subsequent authors followed Pauli, and a velocity-dependent mass captured the majority of popular scientific books and brochures, encyclopaedias, school and university textbooks on general physics, and also monographs, including books of eminent physicists specially devoted to the theory of relativity.

One of the first monograph textbooks in which the theory of relativity was given a systematically relativistic exposition was the *The Classical Theory of Fields* of Landau and Lifshitz.[21] It was followed by a number of other books.

The diagram method of Feynman, which he created in the middle of this century,[22] occupied a central position in the systematically relativistic four-dimensional formalism of quantum field theory. But the tradition of using a velocity-dependent mass was so ingrained that in his famous lectures published at the beginning of the sixties[23] Feynman made it the basis of the chapters devoted to the theory of relativity. It is true that the discussion of the velocity-dependent mass ends in Chap. 16 with the two following sentences:

"That the mass in motion at speed v is the mass m_0 at rest divided by $\sqrt{1 - v^2/c^2}$, surprisingly enough, is rarely used. Instead, the following relations are easily proved, and turn out to be very useful:

$$E^2 - p^2c^2 = M_0^2c^4 \qquad (16.13)$$

and

$$pc = \frac{Ev}{c} ». \qquad (16.14'')$$

In the last lecture published in his life (it was read in 1986, is dedicated to Dirac, and is called "The reason for antiparticles,"[24]) Feynman mentions neither a velocity-dependent mass nor a rest mass and speaks merely of mass and denotes it by m.

17. Imprinting and mass culture

Why is the formula $m = E/c^2$ so tenacious? I cannot give a complete explanation. But it seems to me that here the popular scientific literature has played a fatal role. For from it we draw our first impressions about the theory of relativity.

In ethology there is the concept of imprinting. An example of imprinting is the way chicks learn to follow a hen in a short period after their hatching. If at this period the chick is palmed off with a moving child's toy, it will subsequently follow the toy and not the hen. It is known from numerous observations that the result of imprinting cannot be subsequently changed.

Of course, children and, still less, young people, are not chicks. And, having become students, they can learn the theory of relativity in covariant form, "according to Landau and Lifshitz," so to speak, and without a mass that depends on the velocity and all the nonsense that goes with it. But when, having grown up, they start to write booklets and textbooks for young people, the imprinting takes over again.

The formula $E = mc$ long ago became an element of mass culture. This gives it a particular tenacity. Sitting down to write about the theory of relativity, many authors start with the assumption that the reader is already familiar with this formula, and they attempt to exploit this knowledge. Thus a self-sustaining process arises.

18. Why it is bad to call E/c^2 the mass

Sometimes one of my physicist friends says to me: "Now why do you bother about this relativistic mass and rest mass? At the end of the day nothing terrible can happen if a certain combination of letters is denoted by some one letter and described by one or two words. After all, even although they use these concepts, which are indeed archaic, engineers correctly design relativistic accelerators. The important thing is that the formulas should not contain mathematical errors."

Of course, one can use formulas without fully understanding their physical meaning, and one can make correct calculations despite having a distorted idea of the essence of the science that these formulas represent. But, first, distorted concepts can sooner or later lead to an incorrect result in some unfamiliar situation. Second, a clear understanding of the simple and beautiful foundations of science is more important than the unthinking substitution of numbers in equations.

The theory of relativity is simple and beautiful, but its exposition in the language of two masses is confused and ugly. The formulas $E^2 - \mathbf{p}^2 = m^2$ and $\mathbf{p} = E\mathbf{v}$ (I now use units in which $c = 1$) are among the most transparent, elegant, and powerful formulas of physics. Quite generally, the concepts of a Lorentz vector and a Lorentz scalar are very important, since they reflect a remarkable symmetry of nature.

On the other hand, the formula $E = m$ (I again set $c = 1$) is ugly, since it represents an extremely unfortunate designation of the energy E by a further letter and term, these being, moreover, the letter and term associated in physics with another important concept. The only justification of the formula is the historical justification—at the beginning of the century it helped the creators of the theory of relativity to create this theory. From the historical point of view, this formula and everything associated with it can be regarded as the remnants of the scaffolding used in the construction of the beautiful edifice of modern science. But to judge from the literature, it is today regarded as almost the principal portal of this edifice.

If the first argument against $E = mc^2$ can be called aesthetic—beautiful as against ugly—the second can be called ethical. Teaching the reader this formula usually entails deceiving him, hiding from him at least part of the truth and provoking in his mind unjustified illusions.

First, it is hidden from the inexperienced reader that this formula is based on the arbitrary assumption that the Newtonian definition $\mathbf{p} = m\mathbf{v}$ of the momentum is natural in the relativistic domain.

Second, it creates implicitly in the reader the illusion that E/c^2 is a universal measure of inertia and that, in particular, proportionality of the inertial mass to γ is sufficient to ensure that a massive body cannot be accelerated to the speed of light, even if its acceleration is defined bv the formula $\mathbf{a} = \mathbf{F}/m$. But from

$$\frac{dv}{dt} = \frac{F}{m_0}\left(1 - \frac{v^2}{c^2}\right)^{1/2} \tag{18.1}$$

it follows that

$$\int_0^c dv\left(1 - \frac{v^2}{c^2}\right)^{-1/2} = \frac{1}{m_0}\int_0^T F\,dt. \tag{18.2}$$

Assuming that the force F is constant, we readily find that the time T required for the body to reach the speed c is

$$T = \frac{\pi m_0}{2Fc}. \tag{18.3}$$

This incorrect result is due to the fact that it is necessary to substitute in the formula $a = F/m$, not the "relativistic mass," but the "longitudinal mass," which is proportional to γ^3, a fact that, as a rule, modern authors do not remember.

Third, an illusion is created for the reader that E/c^2 is the universal gravitational mass. In reality, as we have seen, in the relativistic case, in contrast to the nonrelativistic case, there is no universal gravitational mass—for the force that acts on a photon traveling horizontally is twice the force that acts on a photon falling vertically.

Fourth, in calling this Einstein's formula, one hides Einstein's true formula, $E_0 = mc^2$, from the reader.

A third argument can be called philosophical. After all, the definition $E = mc^2$ provides the basis for tens of pages of profound philosophical discussions about the complete equivalence of mass and energy, the existence of a single essence "mass-energy," etc., whereas, according to the theory of relativity, an energy does indeed correspond to any mass but the opposite is by no means true—a mass does not correspond to every energy. Thus, there is not a complete equivalence between mass and energy.

A fourth argument is terminological. The literature on the theory of relativity contains such a confusion in the notation and terminology that is resembles a city in which the transport must simultaneously observe the rule of driving on the right and driving on the left. For example, in the Great Soviet Encyclopaedia, in various physics encyclopedias, and in handbooks the letter m is used to denote the mass and the relativistic mass; the ordinary mass is sometimes called the mass, but more often the rest mass, while the relativistic mass is also called the kinetic mass, but frequently simply the mass. In some papers the authors adhere to a mainly consistent relativistic terminology, in others to a consistently archaic one. It is difficult for an inexperienced reader who wishes to compare, say, a "mass" paper with a "relativity" paper.

The same mixing of notation and terms can also be found in many textbooks and monographs. And all this confusion flourishes at a time in which there is in the theory of relativity essentially just one term, mass, and all the others come "from the devil."

A fifth argument is pedagogical. Neither a school pupil, nor a school teacher, nor a student of first courses who has learnt dogmatically that the mass of a body increases with its velocity can truly understand the essence of the theory of relativity without then spending considerable efforts on reeducation.

As a rule, someone who has not subsequently become a professional relativist has the most false ideas about mass and energy. At times the formula $m = m_0[1 - (v^2/c^2)]^{-1/2}$ is all that remains in his memory, together, of course, with the formula $E = mc^2$.

It is clear that any independently thinking student must feel an intellectual discomfort when studying the theory of relativity using a standard school textbook.

19. "Does mass really depend on velocity, dad?"

Such is the title of a paper published by C. Adler[24] published in the American Journal of Physics in 1987. The question posed in the title was put to the author by his son. The answer was: "No!" "Well, yes...", "Actually, no, but don't tell your teacher." The next day his son dropped physics.

Adler writes that with every year the concept of a relativistic mass plays an ever decreasing role in the teaching of the special theory of relativity. He illustrates this assertion with quotations from four successive editions of a textbook widely used in the United States, *University Physics*, from 1963 through 1982.

Speaking of the opinions of Einstein, Adler gives extracts from an unpublished letter of Einstein to Lincoln Barnett, written in 1948:

"It is not good to introduce the concept of the mass $M = m[1-(v^2/c^2)]^{-1/2}$ of a body, for which no clear definition can be given. It is better to introduce no other mass than the rest mass' m. Instead of introducing M, it is better to mention the expression for the momentum and energy of a body in motion."[2)]

Viewed historically, Adler regards the relativistic mass as an inheritance from the prerelativistic theories of Lorentz and Poincaré. He criticizes this concept and expresses optimism with regard to the decrease of its use.

20. Fizika v shkole (Physics at School)

It so happened that in the same year 1987 in which Adler's paper appeared I had to work in a commission created by the former Ministry of Education of the USSR to determine the winners of the All-Union competition for the best textbooks on physics. Having become acquainted with about 20 submitted textbooks, I was struck by the fact that they all treated a velocity-dependent mass as one of the central points of the theory of relativity.

My surprise increased still more when I found that the majority of the members of the commission—pedagogues and specialists on methods of teaching—had not even heard that a different point of view existed. In a brief improvised talk I told them about the two basic formulas $E^2 - \mathbf{p}^2c^2 = m^2c^4$ and $\mathbf{p} = (E/c^2)\mathbf{v}$. One of them then said to me: "Now you know about this and we know, but nobody else knows. You must write an article about mass for the journal *Fizika v shkole*. Then 100,000 physics school teachers will know about it."

Somewhat flippantly, as it subsequently turned out, I assured them that everything I had said was known not only to all professional physicists but also to students of nonpedagogical universities. But I promised to write the paper.

A few days later, encountering at the next meeting of the commission the deputy of the editor-in-chief of the journal *Fizika v Shkole*, I told her about the proposal that had been made and asked if the journal would commission from me a paper on the concept of mass in the theory of relativity. For about two months there was no answer, and then the person to whom I had spoken rang me up and said that the editorial board had decided not to commission such a paper. It appears that the imprinting that I wrote about earlier had been at work.

This refusal only strengthened my conviction of the need for such a paper. Working on it, I studied more than 100 books and about 50 papers. I saw that the school textbooks were not much worse than the university textbooks, and I became interested in the history of the question. The material expanded, and the work began to absorb me. And there appeared no end to it.

I then decided to sit down and write this short text, putting away in a separate file the detailed bibliography and pages with an analysis of different papers and books.

Time does not wait. Every year books are published in millions of copies that hammer into the heads of the young generations false ideas about the theory of relativity. This process must be stopped.

I am grateful to the members of that competition commission for initiating the writing of this paper. I am also grateful for helpful discussions and comments to B. M. Bolotovskiĭ, M. B. Voloshin, P. A. Krupchitskiĭ, and I. S. Tsukerman.

1) (Editor's Note. Essentially the same material by the same author appears in Phys. Today **40**(6), 31 (1989).

2) In connection with the extract from the letter to Barnett, it is appropriate to give extracts from the "Autobiographical notes" by Einstein[26] published in 1949 (in Vol. 4 of the Soviet edition of Einstein's scientific works these come directly after the foreword to the book of L. Barnett: *The Universe and Dr Einstein* (New York, 1949)). In these notes, recalling the initial stage of work on the creation of the relativistic theory of gravitation, Einstein writes: "... the theory had to combine the following things:

1) From general considerations of special relativity theory it was clear that the *inert* mass of a physical system increases with the total energy (therefore, e.g., with the kinetic energy).

2) From very accurate experiments ... it was empirically known with very high accuracy that the gravitational mass of a body is exactly equal to its *inert* mass."

This extract confirms that in the work on the creation of the general theory of relativity the concept of a mass that increases with increasing kinetic energy was the point of departure for Einstein. The extract may also indicate that, recalling in 1949 this concept without any reservations, Einstein was not completely consistent. It is possible that he assumed that in this way he would be understood by a larger number of readers.

1. J. J. Thomson, Philos. Mag. **11**, 229 (1881).
2. A. Einstein, Ann. Phys. (Leipzig) **17**, 891 (1905).
3. A. Pais, *Subtle is the Lord. The Science and the Life of Albert Einstein* (Clarendon Press, Oxford, 1982).
4. H. Poincaré, *Lorentz Festschrift*, Arch. Neerland **5**, 252 (1900).
5. H. Lorentz, Proc. R. Acad. Sci. Amsterdam **1**, 427 (1899).
6. H. A. Lorentz, Proc. R. Acad. Sci. Amsterdam **6**, 809 (1904).
7. A. I. Miller, *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911)*, Addison-Wesley, Reading, 1981.
8. A. Einstein, Ann. Phys. (Leipzig) **18**, 639 (1905).
9. A. Einstein, Ann. Phys. (Leipzig) **20**, 627 (1906).
10. A. Einstein, Ann. Phys. (Leipzig) **23**, 371 (1907).
11. A. Einstein, Ann. Phys. (Leipzig) **35**, 898 (1911).
12. M. Planck, Verh. Dtsch. Phys. Ges. **4**, 136 (1906).
13. M. Planck, Sitzungsber. Akad. Wiss. Berlin **13**, 542 (1907).
14. H. Minkowski, Phys. Z. **10**, 104 (1909).
15. G. Lewis and R. Tolman, Philos. Mag. **18**, 510 (1909).
16. R. Tolman, Philos. Mag., **23**, 375 (1912).
17. M. Jammer, *Concepts of Mass in Classical and Modern Physics*, Harvard University Press, Cambridge, Mass., 1961 [Russ. Transl. Progress, M., 1967].
18. E. A. Whittaker, *History of the Theories of Aether and Electricity*, Vol. 2, Nelson, London, 1953.
19. W. Pauli, "Relativistätstheorie," in: *Enzykl. Math. Wiss.*, Bd 19 Teubner, Leipzig, 1921. [Russ. transl., Gostekhizdat, M., 1947, Nauka, M. 1973; Engl. transl., Pergamon Press, Oxford, 1958].
20. *Einstein The Meaning of Relativity: Four Lectures Delivered at Princeton University May 1921*, Princeton Univ. Press, Princeton, 1970.
21. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (2nd. ed. Pergamon Press, Oxford, 1962. [Russ. original, Gostekhizdat, M., 1955].
22. R. Feynman, Phys. Rev. **76**, 749, 769 (1949).
23. R. P. Feynman, R. B. Leighton, and M. Sands, *Feynman Lectures on Physics*, Vols. 1 and 2, Addison-Wesley, 1963, 1964, Chaps. 15, 16, and 28. [Russ. transl., Mir, M., 1966].
24. R. P. Feynman, "The reason for antiparticles," in: *Elementary Particles and the Laws of Physics. The 1986 Dirac Memorial Lectures*, Cambridge University Press, Cambridge, 1987. [Russ. transl., Usp. Fiz. Nauk **157**, 163 (1989)].
25. C. Adler, Am. J. Phys. **55**, 739 (1987).
26. A. Einstein, "Autobiographical notes," in: *Albert Einstein: Philosopher-Scientist*, (Ed.) P. A. Schilpp, Tudor, N.Y., 1949, reprinted, Harper, N.Y., 1959.

Translated by Julian B. Barbour

The fundamental constants of physics

L. B. Okun'

Institute of Theoretical and Experimental Physics

(The paper is based on a report presented at the First Sakharov Conference, Moscow, May 1991)
(Submitted May 20, 1991)
Usp. Fiz. Nauk **161**, 177–194 (September 1991)

A natural system of fundamental physical units c, $\hbar$, and m_P is discussed, where c is the velocity of light, $\hbar$ is Planck's constant, and m_P is the Planck mass, which is related to the Newtonian gravitational constant by the equation $m_P^2 = \hbar c/G$. In a natural system of units, such questions as: "How does the anthropic nature of the physical universe arise? Is it unique, or does an infinite set of universes exist?" and "Are the fundamentals of the physical universe knowable and what is the strategy for knowing them?" become particularly urgent.

1. INTRODUCTION

This paper is devoted to the question concerning the fundamental constants of physics. It is well known that an adequate choice of physical units is one of the most important prerequisites for solving any specific physical problem. This is especially true in relation to the fundamental questions of physics. A discussion of the question of the choice of fundamental physical units makes it possible to judge with greater understanding not only the history of fundamental physics, but also predictions of its development. Such a discussion binds together the physics of elementary particles and cosmology, and inevitably touches upon the most systematically different questions: from ones of the politics of science (Is it necessary to construct gigantic colliders, or can one see the entire plan for the structure of the physical universe by an effort of pure reasoning?) to philosophical ones (Why is the physical universe so well suited for the existence of life, and is it unique?).

2. LET US PLAY WITH THE VELOCITY OF LIGHT

It makes sense to start a discussion of fundamental physical constants from a somewhat unexpected question: what would be changed in the universe around us if the velocity of light were different than what it actually is, let us say, faster by ten orders of magnitude, i.e., $3\cdot 10^{20}$ cm/sec? In order that this question make sense, one must stipulate what happens at the same time to other physical constants. So then, let Planck's constant $\hbar$, the electron charge e, and the masses of the electron m_e and of the proton m_p remain constant, while the velocity of light become different.

The three quantities $\hbar$, e, and m_e enable one to obtain the dimensionalities of all physical quantities. Thus, it is convenient to choose the Bohr radius, the radius of the hydrogen atom

$$r_B = \hbar^2/m_e e^2,$$

as the unit of length, the Bohr energy

$$E_B = e^2/r_B = e^4 m_e/\hbar^2,$$

as the unit of energy,

$$t_B = \hbar/E_B = \hbar^3/e^4 m_e$$

as the unit of time, and, as the unit of velocity,

$$v_B = r_B/t_B = e^2/\hbar.$$

Thus, an atom is both a clock (t_B) and a ruler (r_B). Therefore, the question of changing the velocity of light is not an empty one, is not a question of redesignating and choosing units.

Since chemical reactions are basically determined by electron exchange, then neither chemistry nor biochemistry would be seriously changed. And nevertheless, the universe would be radically changed. The point is that the properties of a photon would be changed radically. For the same energy E that is determined by the energies of the atomic levels, the photon emitted by the atom would have a momentum k that is ten orders of magnitude smaller:

$$k = E/c,$$

and its wavelength

$$\lambda\!\!\!^- = \hbar/k = \hbar c/E$$

would be ten orders of magnitude longer. Let us note that its frequency ω would remain unchanged: $\omega = E/\hbar$.

The probability of the emission of a photon by an excited atom is proportional to its phase space and consequently, to $k^2 dk$. But

$$k^2 dk = E^2 dE/c^3,$$

and the time for an excited atom to radiate it away optically would exceed the age of the universe. (Atoms would go over into the ground state due to collisions with each other.) The Thomson cross section for photon scattering by free electrons would be reduced by 40 orders of magnitude:[1]

$$\sigma_T = (8\pi/3)r_0^2,$$

where $r_0 = e^2/m_e c^2$, and the Rayleigh non-resonant scattering of light by atoms is[2]

$$\sigma_R \sim (e^2 r_B^2/E_b)^2(\omega/c)^4 \sim r_B^2(v_B/c)^4.$$

It is such that photons would be practically uncoupled from matter. There would be neither the Sun nor a light bulb to shine, nor eyes to see.[1)] All the remaining changes in the universe would possibly be less dramatic. Thus, for example, by Maxwell's equation

 0038-5670/91/090818-09$01.00

$\text{curl } \mathbf{H} = \mathbf{j}/c$

the magnetic field **H** and electric current **j** would be uncoupled. So there would be neither dynamos nor electric motors. But chemical sources of current would remain, although the question as to whether one could buy a battery in a store does not have a definite answer.

At first glance, the circumstance that photons with $c \to \infty$ are uncoupled from charges, while the Coulomb interaction between charges remains unchanged is found to contradict the widely known theoretical statement that the interaction between charges is determined by the exchange of virtual photons. However, one must nevertheless take this statement, which is absolutely correct within a four-dimensional formalism, "with a grain of salt." This is most simply seen if one recalls how the effect which characterizes the interaction of a four-dimensional potential A_i with an electric charge will look; $-(e/c)\int A_i dx^i$, and if one allows for the fact that $A_i = (A,\varphi)$ and $x_i = (x,ct)$ (see Ref. 1).

The example with the velocity of light shows how very relativistic our universe is: in the Galilean limit, it becomes unrecognizably different.

A photon is a relativistic particle. This becomes apparent not only in the kinematics but, as we see, also in its dynamical properties and in its interactions. For some reason, this circumstance is not emphasized in the popular scientific literature. And it's a pity. If only the readers of popular scientific books on the theory of relativity recognized this, then possibly fewer refuters and improvers of this theory would come forth from their ranks.

After considering the case of a very high velocity of light, it is instructive to turn ourselves to the case of a low velocity of light. Effects of the type of slowing of the time measurements of moving clocks during a trip in an automobile were considered by O. A. Vol'berg, who wrote the appendix "An entertaining trip into Einstein's country" to Ya. I. Perel'man's book[3] "*Entertaining Mechanics*" in 1935. In order to make the discussion understandable, Vol'berg assumed that the velocity of light was twice the velocity of the automobile. In 1939, in his book "*Mr. Tompkins in Wonderland*," George Gamow[4] placed the hero on a bicycle and reduced the imagined velocity of light even more. Here all everyday life, except for the slowing of clocks and the contraction of scales, remained unchanged both for Vol'berg and for Gamow. However, it is easy to verify that this cannot be obtained without any accompanying changes of the constants $\hbar$, e, m_e, and m_p. If one holds these constants fixed and reduces c, then the universe will be changed long before the velocity of light will approach the velocity of an automobile.

Indeed, the value of the so-called fine structure constant α,

$$\alpha = e^2\hbar c.$$

is a parameter which characterizes the role of relativistic effects in a hydrogen atom. From experiment, $\alpha \approx 1/137$. It is easy to verify that $\alpha = v_B/c$. Thus, by reducing the velocity of light by at most two orders of magnitude, it would approach the velocity of an electron in a hydrogen atom.

Here one must emphasize that the binding energy of an electron in a hydrogen atom would be, as before, of the order of $E_B = m_e e^4/\hbar^2$, but the rest mass energy of the electron $m_e c^2$ would become comparable with this value. In this sense, the ordinary Coulomb interaction in an atom would become strong.

With a further reduction of c, the quantity $2m_e c^2$ would become less than the binding energy of an electron in a hydrogen atom, and a hydrogen atom H would become so much lighter than a bare proton p that it would become energetically favorable for decay of the proton to a hydrogen atom and a positron e^+:

$$p \to H + e^+.$$

One may say that, in this case, the electron is superbound in hydrogen.

Until now, we have discussed only the lightest atom. But superbound electrons would have shown up considerably sooner in heavy atoms, in their inner shells. The appropriate parameter here is $0.8Ze^2/\hbar c$, where Z is the nuclear charge, and the numerical coefficient 0.8 allows for the fact that one may not neglect the radius of a heavy nucleus in comparison with the radius of an inner electron orbit. As a result, let us say, at $c = 3 \cdot 10^9$ cm/sec, all atoms heavier than silicon would contain superbound electrons, and at $c = 3 \cdot 10^8$ cm/sec, as has been said above, superbound electrons would have shown up also in hydrogen.

A further increase of the value of $\alpha = e^2/\hbar c$ compared to unity would have to lead to a very strong interaction of electrons with positrons. It is not clear as to whether free electrons can exist in general under these conditions, or will confinement ensue for them; life imprisonment in electrically neutral positronium atoms similar to the manner of the confinement of colored quarks and gluons occurs in white hadrons. Perhaps free electric charges would become just as impossible as are free color charges. I write "not clear" since electrodynamics with $\alpha \gg 1$ has not yet been subjected to a systematic theoretical analysis.

3. THREE FUNDAMENTAL CONSTANTS

The example considered in the previous section was based on the assumption that $\hbar$, e, and m_e are the most fundamental dimensional constants, while the velocity of light appears to be less fundamental. But actually we know that, besides the electromagnetic interaction, there are at least two more gauge interactions, the weak and the strong, which are characterized by the charges g_w and g_s that have the same dimensionality as e. Besides the electron, there exist 15 more particles (5 leptons, 6 quarks, 1 photon, 1 gluon, and 2 weak bosons, not counting the anti-particles and the varieties of color). In viewing this diversity of charges and masses, e and m_e already do not appear to be the chosen constants for the role of the most fundamental constants.

Everyone who is even slightly acquainted with elementary particle physics does not doubt that $\hbar$ and c are certainly such constants, for each of them is unique and universal in its role. The velocity c is the limiting velocity for the propagation of physical signals. The constant $\hbar$ is the quantum of angular momentum and, what is no less important, it is the fundamental quantum unit of action. As far as the third fundamental constant is concerned, the opinion has gradually been building up among specialists (a consensus, as it is conventional to say now) that the Newtonian constant G of universal gravitational interaction or some kind of combination of the quantities G, $\hbar$, and c is the best candidate for this missing constant. The most popular of these combinations is called the Planck mass and is denoted by m_P.

As is well known, V_g is the potential energy of the gravitational interaction of two bodies with masses m which are located at a distance r from each other, and it equals

$$V_g = -Gm^2/r.$$

If one recalls that the potential energy of the Coulomb interaction of two charges equals

$$V_e = -e^2/r\,,$$

and allows for the fact that $\alpha = e^2\hbar c$ is a dimensionless quantity, then it is easy to understand that it is natural to represent G in the form

$$G = \hbar c/m_P^2,$$

where m_P is the so-called Planck mass introduced by Planck at the very end of the last century[5] by combining the constants $G, \hbar$, and c. The Planck length $l_P = \hbar/m_P c$ and Planck time $t_P = \hbar/m_P c^2$ were introduced in the same paper. Starting from the known value of G, it is easy to find that

$$m_P \approx 1{,}2\cdot 10^{19}\ \mathrm{GeV}/c^2,$$

$$l_P \approx 10^{-33}\ \mathrm{cm},$$

$$t_P \approx 10^{-43}\ \mathrm{sec}.$$

One often speaks of the Planck energy $E_P = m_P c^2$ and of the Planck momentum $k_P = m_P c$. The physical meaning of the Planck scale started to become clear considerably later.

The first detailed paper devoted to the c, G, and $\hbar$ system, "Universal constants and the ultimate transition" was published at the start of 1928 by G. Gamow, D. Ivanenko, and L. Landau.[6] (The first two authors were then 24 years old, and the last one was 20 years old. According to the testimony of one of the authors, the paper was written in the form of a joking birthday present to a female student acquaintance. Not one of them referred to this paper later on[2]), although there are a number of profound ideas in it.)

At the start of the 1930s, M. P. Bronshteĭn gave a detailed classification of physical theories based on c, G, and $\hbar$ units, and used them to quantize gravity. He introduced the term $cG\hbar$-physics (M. P. Bronshteĭn was executed by shooting in 1938 in his 32nd year). Then L. D. Landau[7] and J. A. Wheeler[8,9] in the mid-1950s started to address the role of the Planck mass. This time the discussion was about whether, as one approaches Planck distances l_P or momenta k_P, the gravitational interaction must become comparable in strength with the other interactions, and significant quantum fluctuations must arise for it (G. Gorelik[10,11,12] discussed the history of this question in more detail).

At present, the Planck mass m_P, along with the constants $\hbar$ and c, is considered as a fundamental physical quantity, which characterizes the energy scale for theories of super-unification for all interactions, including gravitation. As is well known, superstring theory is considered to be the most promising direction for creating a super-unification theory (see the books by A. M. Polyakov[13] and M. B. Green, J. Schwartz, and E. Witten[14]). Instead of point particles, extended one-dimensional objects, "strings," which have characteristic Planck dimensions l_P, are the fundamental objects of this theory.

At the same time, the Planck scale with its characteristic time t_P forms the basis of quantum cosmology, of which the wave function of the universe is the fundamental object (see, for example, the books by C. W. Misner, K. S. Thorne, and J. A. Wheeler[15] and S. Hawking,[16] the papers by S. Coleman,[17] S. Weinberg,[18] M. Gell-Mann and J. Hartle,[19,20] the book and review by A. Linde,[21,22] and also the popular review by J. J. Halliwell[23]). One of the objectives of quantum cosmology is to understand how, in the process of the evolution of the early universe, the properties of particles and the vacuum were fixed.

One can express the dimensions of any physical quantity by the dimensions of length L, time T, and mass M. Starting from our worldly experience, one might expect that nature chooses fundamental length, time, and mass as three natural independent units. But nature decided otherwise: the limiting velocity of signal propagation c ($[c] = [L/T]$) and the quantum of action $\hbar$ ($[\hbar] = [ET] = [ML^2/T]$) acquired fundamental meaning.

In order to represent M. P. Bronshteĭn's ideas in graphic form, A. L. Zel'manov drew[24,25] in the 1960s a "physical theories cube" constructed on the three $1/c$, G, and $\hbar$ orthogonal axes (see Fig. 1). When one of the units goes to zero, a limiting transition into a plane occurs. For fixed values of $\hbar$ and c, it is already unimportant what the third fundamental unit is; mass, length, or the Newtonian constant G itself.

Today one usually replaces G by $1/m_P$ and speaks of physics on the Planck scales.

The physical theories cube, which is an integral part of modern physics folklore, is depicted in Fig. 2. Newtonian mechanics (NM), or more accurately, that part of it which does not take gravity into account, is located at the origin of coordinates. Above it is located non-relativistic (Newtonian) gravity (NG), to its right is quantum mechanics (QM), and in front of it is the Special Theory of Relativity (STR). A synthesis of the Special Theory of Relativity and quantum mechanics gives quantum field theory (QFT). A synthesis of non-relativistic gravity and the Special Theory of Relativity gives the General Theory of Relativity (GTR). A synthesis of quantum mechanics and non-relativistic gravity gives non-relativistic quantum gravity (NQG), a theory with respect to which it is not clear whether there exist objects which it describes (see below about this). And finally, the synthesis of all theories in the future may lead to a universal Theory of Everything (TOE, the English acronym for this theory).

The main achievements of twentieth century physics, which led to radical changes in the whole way of life of humanity, lie in the $\hbar c$ plane. And although we have traversed a

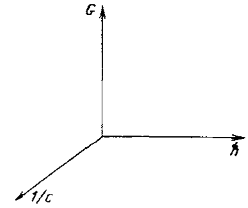


FIG. 1. The $1/c$, G, and $\hbar$ axes.

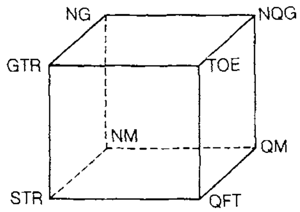


FIG. 2. The physical theories cube.

path from 10^{-8} GeV to 100 GeV along the energy axis, we are still 17 orders of magnitude away from the Planck mass. So that the energy layer investigated appears to be vanishingly thin on the Planck scale.

It is obvious that the well known charges e, g_w, and g_s are dimensionless quantities in $c\hbar$ units. One usually characterizes them by the dimensionless quadratic values:

$$\alpha = e^2/\hbar c, \quad \alpha_w = g_w^2/\hbar c, \quad \alpha_s = g_s^2/\hbar c.$$

According to quantum field theory, the charges of interacting particles change as a function of the distance between the particles (of the transferred momentum or energy). We know that, in the momentum interval from 0 to 100 GeV, α increases slightly, from 1/137 to 1/128, and α_s decreases sharply, from a value of the order of one (at confinement scales of $E \lessdot 1$ GeV) to ~ 0.1 for $E \sim 100$ GeV. As far as α_w is concerned, it is approximately constant in this interval, $\sim 1/30$, but must, in principle, change at higher energies. Extrapolation of the curves for all three "running" constants indicates that they all will "converge" at common values, that are appoximately 1/40 at energies from 10^{13} to 10^{16} GeV. The closeness of the energy for this grand unification to the Planck mass serves as one more argument for the latter being a natural fundamental energy unit in physics.

Let us now turn to masses. The mass scale for hadrons which consist of light u, d, and s quarks is mainly (but not only) determined by the characteristic confinement radius. The masses of hadrons which consist of heavy (c and b) quarks are mainly determined by the masses of these quarks. In turn, according to the hypothesis of Higgs bosons, the masses of quarks, just like the masses of leptons, are determined by the value of η, the vacuum condensate of a Higgs field, which is approximately 250 GeV. In its simplest variant, each of the masses is a product of the η values and of one of the constants which characterizes the interaction of one or another lepton or quark with a Higgs field. Such constants are called Yukawa constants; they have the same dimensions as do the e, g_w, and g_s charges and, consequently, they are dimensionless in units of $\hbar$ and c. (Let us note in passing that the mass of the W-boson is determined by the product of η and $\frac{1}{2}g_w$, and g_w is replaced by $g_w^2/\sqrt{g_w^2 - e^2}$ for the Z-boson).

The scheme outlined here for the origin of the masses is the very simplest one. Other, more refined schemes also exist. In some of them there are no Yukawa interactions at all, and they are replaced by complicated non-Abelian gauge structures (technicolor).

But what is an undoubted fact is that, at present, elementary particle theory contains over twenty dimensionless parameters which appear as arbitrary today. (The three angles and the phase of the Kobayashi–Maskawa matrix, which determines the interaction of weak quark currents and the violation of CP-invariance in weak interactions, are also among the arbitrary parameters. It is possible that an analogous but still more complicated matrix describes charged lepton currents with neutrino participation. The more complicated nature of weak neutrino interactions may be determined by the fact that not only Dirac, but also Majorana masses are possible for them.) At the modern level of science, it appears that we shall solve nothing in the structure of the theory by changing one or another of these parameters, although, of course, here the appearance of the universe that is described by it is radically changed.

Purely kinematic exercises with the velocity of light a la Mr. Tompkins are an extremely frivolous activity from the point of view of fundamental physics. (No wonder, according to the testimony of Z. Kazimir,[26] this activity "seemed to Bohr stupid rather than funny".) Also a discussion of the dynamical properties of a universe in which it is imagined that $c \to \infty$, but $\hbar$, e, and m_e remain unchanged is "rather stupid" from the point of view of the $cG\hbar$ system. But from this same point of view, the efforts of many years to construct a single theory of the gravitational and electromagnetic fields without taking the constant $\hbar$ into account certainly appear to be doomed to failure. In fact, not the dimensional quantity e, but the dimensionless α is the fundamental parameter of electrodynamics. (As is well known, classical electrodynamics becomes internally inconsistent at the distances which characterize the classical electron radius $r_0 = e^2/m_e c^2 \sim 10^{-13}$ cm, which are 20 orders of magnitude longer than the Planck distances $l_P \sim \hbar/m_P c \sim 10^{-33}$ cm.) The problem of constructing a single fundamental theory in one plane taken separately, be that $1/c \to 0$, $\hbar \to 0$, or $G \to 0$, is Utopian.

4. THE ANTHROPIC NATURE OF THE PHYSICAL UNIVERSE

The question as to whether all the dimensionless parameters in the final physical theory will be fixed by a condition of consistency, or if certain of them will remain arbitrary, is today a question of belief. It does not have a scientific answer at present. The word "arbitrary" means in this context that a given dimensionless parameter assumed its value in the process of the cosmological evolution of the universe at an early stage of it. Here, with a greater or lesser probability, it could have also assumed other values.

Even a fleeting glance at the arbitrary, "free" parameters, suffices to astonish one how favorable their values are for our existence. The elementary particle masses appear especially surprising on the Planck scale.

For example, the difference of the masses of the neutron and proton, $m_n - m_p$, is 1.33 MeV $\sim 10^{-22} m_P$. But if this difference were, let us say, 1 MeV less, the neutron would become stable, and the hydrogen atom, as I. L. Rozental' emphasized,[27-30] would be unstable:

$$e^- + p \to n + \nu_e.$$

The set of reactions $\nu_e p \leftrightarrow n e^+$ and $e^- p \leftrightarrow n \nu_e$, which determine the ratio between the numbers of neutrons and protons that are left to us as the legacy of primeval nucleosynthesis, would have, under these conditions, shifted the balance towards an equal abundance of protons and neutrons. As a result, not hydrogen, but helium atoms would have been the most abundant matter in the universe. And the entire evolution of the formation and burning of stars would be changed radically. They would have exploded rapidly. Life would have been impossible for many reasons. A small (~ 1 MeV) increase of the electron mass leads to analogously radical consequences.

Thus, the entire structure of the universe is extremely sensitive to small "disturbance" of the value of the electron mass and/or the difference of the masses of the proton and neutron, and essentially, to the difference of the masses of the u and d quarks. In fact, the neutron is heavier than the proton because of the fact that the d quark is heavier than the u quark ($m_d \sim 7$ MeV, $m_u \sim 5$ MeV). Let us note that, in the two other generations of quarks, unlike the first generation, the lower quarks (s,b) are significantly lighter than their upper partners (c,t).

Until now, we have discussed the sensitivity of our universe to values of the masses of the fundamental fermions, m_e, m_u, and m_d. The sensitivity to less fundamental quantities, such as the binding energy of the nucleons in a deuteron,[28] is even more remarkable. From experiment, $\varepsilon_d \sim 2.2$ MeV. A reduction of this binding energy by at most 0.4 MeV would lead to the situation that the main reaction for burning hydrogen in the Sun $pp \to de^+\nu_e$ would turn out to be forbidden, and only the considerably less probable reaction $pe^-p \to d\nu_e$ could go on. Essentially, everything depends on the details of the nuclear forces between the nucleons which, from the point of view of quantum chromodynamics, are something like the "chemistry of strong interactions."

Another, even subtler example; this is the details in the situation of the energy levels of the ^{12}C and ^{16}O nuclei. The famous level of the carbon nucleus with an excitation energy of 7.65 MeV lies at most 0.3 MeV above the sum of the masses of the ^{4}He + ^{8}Be nuclei. The ^{8}Be nucleus is unstable, and therefore, without this level, which by resonance amplifies the cross-section of the $3(^4\mathrm{He}) \to {}^{12}\mathrm{C}$ reaction, carbon would be formed considerably less efficiently than it would be burned up in the reaction $^{12}\mathrm{C} + {}^4\mathrm{He} \to {}^{16}\mathrm{O} + \gamma$, and the universe would be so poor in carbon that life would hardly have arisen. Arguments of just this kind led Fred Hoyle to his prediction at the start of 1953 of the existence of the 7.65 MeV level and to his discovery of it in cooperation with experimenters at the California Institute of Technology about a week later. (The history of this discovery has been very clearly described by Hoyle in the symposium "*Essays in Nuclear Astrophysics: Dedicated to W. Fowler on His Seventieth Birthday,*"[31] and the astrophysical role of the energy levels of ^{12}C and ^{16}O has been described in Hoyle's book "*Galaxies, Nuclei, and Quasars.*"[32]) When you look at the energy level diagram for the ^{12}C nucleus (there are about 30 of them in an interval of the order of 30 MeV; see Ref. 33) and you see the first three levels at 4.43 MeV, 7.65 MeV, and 9.64 MeV, then a feeling of profound gratitude captures the spirit for the 7.65 MeV level, and for the fact that it was not lowered by 0.5 MeV. What a small safety factor for all which is dear to us!

One more example was indicated by F. Dyson,[34] who noted that the presence of even a weakly bound state for two protons, i.e., for the ^{2}He nucleus also would have had a decisive effect on the entire development of the universe. One could multiply examples such as these many times. But what do small changes of the nuclear forces mean from the Planck point of view?

If a Grand Unification Theory is valid, then the characteristic confinement momentum Λ_{QCD} (the inverse confinement radius) is related to the Grand Unification characteristic momentum Λ_{GUT} by the relation

$$\frac{1}{\alpha_s(\Lambda_{GUT})} - \frac{1}{\alpha_s(\Lambda_{QCD})} \approx \ln\frac{\Lambda_{GUT}}{\Lambda_{QCD}}.$$

(The missing coefficient, which is of the order of one, in front of the logarithm is determined by the contribution of particles with color. If one does not take into account superparticles and other hypothetical particles, but allows for only the ordinary gluons and quarks, then it equals $7/2\pi$). If we now neglect $1/\alpha_s(\Lambda_{QCD})$, which is of the order of one, in comparison with $1/\alpha_s(\Lambda_{GUT})$ which, as was already stated above, is of the order of 40, then a simple relation is found

$$\Lambda_{QCD} \sim \Lambda_{GUT} e^{-1/\alpha_s(\Lambda_{GUT})}.$$

We see that a decrease of $\alpha_s(\Lambda_{GUT})$ by half, from 0.02 to 0.01, decreases Λ_{QCD} by 17 orders of magnitude. It is sufficient to decrease $\alpha_s(\Lambda_{GUT})$ by at most 10% in order that Λ_{QCD} be decreased by a factor of 10, and that the masses of the nucleons would already be determined not by Λ_{QCD}, but by the current masses of the light quarks. Allowing for the fact that $m_u \sim 5$ MeV and $m_d \sim 7$ MeV, we would have $m_p \sim 17$ MeV and $m_n \sim 19$ MeV.

Here, perhaps, the time has arrived to speak of the comment of J. Sal'vini, who pointed out that, if the masses of nuclei and electrons were comparable, there could be neither crystals nor solid bodies, and consequently, also, no classical instruments of quantum mechanics. The question of its probabilistic interpretation does not even arise here, although all the dynamic consequences of quantum mechanics remain. A small disturbance of the "free" parameters would lead to a situation that not only physicists, not only all live creatures such as we know them, but also individual chapters of physics textbooks would fall through into Tartarus.[3)]

5. THE ANTHROPIC PRINCIPLES

Along with the already mentioned "lucky accidents," there are many others, which are associated with the evolution of the Milky Way and of stars.[36-39] Thus, crude dimensional estimates show that the lifetime of an ordinary star, in whose interior the carbon, nitrogen, oxygen, and heavier elements that are necessary for life are created, must be of the order of $(\hbar/m_p c^2)(\hbar c/G m_p^2)^2$. This time is sensitive to the ratio $Gm_p^2/\hbar c = m_p^2/m_P^2$, which one often denotes in the literature as α_G by analogy with α_e, α_w, and α_s. (Let us note that the designation α_G does not seem fortunate, since in this case the discussion is not about the magnitude of a gravitational charge, but about the magnitude of the proton mass. In fact, the magnitude of G is a fundamental dimensional unit in the $cG\hbar$ units. But the use of the designation α_G supposes that the proton mass m_p, and not the Planck mass m_P, is the fundamental unit of the dimensionality of mass.)

An enormous literature is devoted to the anthropic nature of the universe. P. C. W. Davies' book "*The Accidental*

Universe"[40] may serve as a good popular introduction. The most complete book on this theme is the one by J. D. Barrow and F. J. Tipler, "*The Anthropic Cosmological Principles*"[41]; it contains over 700 pages, over 1,500 references, and it covers diverse aspects from physics, astrophysics, cosmology, biochemistry, biology, and computer science to history, philosophy, problems of extraterrestrial civilizations, and religion.

The anthropic properties of the universe led to the formulation of a number of hypothetical (speculative) principles.

The weak anthropic principle starts from the idea of an ensemble containing an infinite number of universes. The *a priori* probability of creating an anthropic universe is vanishingly small. But this small value has no bearing on the subject, since the *a posteriori* probability is significant. It follows from the fact of our existence that we are not able not to live in one of the "very best of universes."

An infinite network of universes, each of which generates innumerable daughter universes in its early inflationary stage, is the theoretical realization of this statistical ensemble. Not only different symmetry breaking schemes, but also different numbers of space-time dimensions can be realized in each of them. However, an infinite variety of values for the dimensionless free parameters turns out to be possible even for a given number of space-time dimensions and for a given symmetry breaking scheme. Ya. B. Zel'dovich[42] made a contribution to the anthropic principle in its weak form. A. D. Sakharov returned to it many times in his publications. Thus, he wrote in Ref. 43: "Some authors consider the anthropological principle to be unfruitful and even not in accord with the scientific method. I do not agree with this. In particular, I note that the requirement that the fundamental laws of nature apply under conditions that are significantly different from those in our universe can have heuristic importance for finding these laws."

A discussion of the anthropic principle within the framework of the inflationary universe is contained in papers by A. Linde.[21,22]

In the context of an infinite abundance of variants, it no longer seems so surprising that at least one turned up in which intelligent life capable of knowing the universe is possible.

It is important to emphasize that, notwithstanding the fundamental impossibility of transmitting information from one universe into another one, they all (on paper) are "children" of one "primeval Lagrangian" with the same dimensional fundamental units: c, G, and $\hbar$.[4)]

In contrast to the weak principle, the strong anthropic principle states that the universe must necessarily be constructed so as to provide for the possibility to know itself. A number of different formulations of the strong principle exists, which are discussed in detail in the book by Barrow and Tipler. It is possible that all its broken symmetries and all values of the free dimensionless parameters are fixed by a self-consistency condition for this unimaginably complicated nonlinear system. (L. Maiani expressed this point of view.)

6. IN WHAT IS THERE HOPE?

The questions which arise in connection with the anthropic principles are immeasurably more complicated than the questions that are solved by modern elementary particle physics. The strength of physics in general, and of fundamental physics in particular, be that high energy physics on colliders or the physics of subterranean, low-background laboratories, lies in the capability to find and solve questions which lend themselves to solutions. As I. Pomeranchuk once said in the studio of the sculptor Vadim Sidur: it is important, both in creating sculptures and in solving a physical problem, to understand and to feel what one can neglect.

Can one, by making finite, local steps, arrive at a comprehension of very profound global (or, more accurately) universal truths, in which nothing may be neglected? A comparison of the part of the energy scale accessible to experimental investigation (with a realistic prospect for $\leqslant 10^5$ GeV) with the Planck energy scale $\sim 10^{19}$ GeV which is the natural one for a Theory of Everything and seemingly symbolizes the tragic gap between the ideal and reality in twentieth century physics, can cause an especially strong attack of pessimism. The confusion is also brought about by the fact, that the number of particles which remains to be discovered in the TeV range is no less than the number of fundamental particles already known. So that it seems that, at least in the immediate future, we shall move towards an ever larger variety of the fundamental building blocks of matter.

However, progress in creating a unified picture in modern fundamental physics is characterized not so much by a decrease in the number of fundamental particles as by a decrease in the number of "free" parameters. The establishment of a theoretical quantitative relation between dimensionless parameters, which previously were independent, raises physics to a new, higher level of unity. The scope of phenomena that are described from a unified point of view is greatly expanded here. The creation of the modern standard model for the electroweak and the strong interactions was the most recent such stage.

It is interesting to compare theoretical physics with mathematics. In mathematics, along with the seeming unbounded process of growth and branching, there also occurs a process of synthesis, when by the effort of pure reason, profound relations are established between fields and concepts that seem distant at first glance.

In theoretical physics there are also processes of differentiation and integration. But here experiment and the observation of nature also play a very important role. They throw the seeds of new theoretical sprouts into the ground, they stimulate the growth of some theoretical phantasies and mercilessly weed out others.

We have great expectations from experiments on future TeV colliders. Of course, much may be discovered on machines that are already operating: the t quark, light Higgs bosons, and the very lightest of the supersymmetry particles. But only TeV colliders can reveal the entire richness of scalar and supersymmetry particles, can give a hint as to how particle masses are constructed, and can sharply reduce the number of free parameters of fundamental physics.

Data on the current constants α_s, α_w, and α_1 (the last one is a combination of α and α_w) are a clear example of how TeV physics could serve as a launching pad for very far-reaching extrapolations in energy. As Ugo Amaldi recently emphasized,[44] data obtained on the LEP accelerator indicate that, with the known set of particles, these current constants do not meet at a single "triple" point, but at three "double" points at 10^{13} GeV, 10^{14} GeV, and 10^{16} GeV. At the same time, as J. Ellis, S. Kelley, and D. V. Nanopoulos noted,[45] allowance for the supersymmetry partners of the ordinary particles focuses all three curves at one point at 10^{16} GeV. Immediately there are two corollaries of this result. The first one refers to the fact that the supersymmetry particles mentioned must be light with masses not greatly exceeding 1 TeV. Otherwise, the curves of the current constants will acquire break points, and the focusing will be disrupted. And the second one is that new detectors for searches for proton decay that are orders of magnitude more massive than previous detectors are needed.

Experiments with neutrinos, especially with solar neutrinos, may turn out to be another invaluable source of fundamental data, since specifically, they are especially sensitive to small differences of mass and to small angles of mixing of the different types of neutrinos.

The solution of the problem of dark matter in the universe may turn out to be an unexpected gift.

Of course no imaginable accelerators will enable either us or our descendants to reach Planck energies. But it is not impossible (and keeping in mind the just considered example of extrapolating the three current constants it is even plausible!) that the exponential inaccessibility for experimenters is one side of a coin, the other side of which is the logarithmic accessibility for theoreticians. In fact, perhaps it will turn out to be sufficient for theoreticians to analyze and understand only the exponents that are expressed by only two-digit numbers.

If in the next few decades we succeed in sorting out the physics of phenomena up to TeV energies, then this may prove to be a sufficiently broad base platform for examining by a mental effort many details of physics at Planck energies. In fact, the astronomical distances to celestial bodies have not hindered studying them in more detail than the interior of our own planet. Yes, and helium was first discovered on the Sun and not on the Earth.

In speaking of theoretical "insight," one must not forget the intensive process of synthesizing theoretical physics and mathematics, as a result of which new fields of mathematics are created. Ideas of beauty or simply beautiful ideas also can perform an important role (see the remarks above on mathematics).

It is appropriate to conclude this section with the following remark. For many years, the early universe was considered by theoretical physicists as a natural Planck laboratory. If, however, the temperature of the universe after inflation never exceeded values of the order of 1 TeV, as many cosmologists are inclined to think,[5)] then, as a result of inflation, the "Planck laboratory" receded exponentially far beyond the horizon. And the next generation of supercolliders may tell us more about the early universe than do astrophysical observational data. And the detection of proton decay by the next generation of subterranean detectors would give information about energies that are absolutely inaccessible to observational cosmology.

7. MENTAL EXPERIMENTS. IS THE PLANCK SCALE NOT A MIRAGE?

From the definition of the Planck mass, it follows that the gravitational interaction between two particles, each of which has a Planck mass and a unit electric charge e is 137 times stronger than the electromagnetic interaction between them. But just by itself, this still does not mean that the gravitational interaction for such particles actually becomes strong: in fact, they may be located fairly far from each other.

The gravitational attraction will become actually strong when the distance between the particles is decreased so much that the interaction potential will become comparable with their rest mass energy:

$$Gm_P^2/l_P = m_P c^2.$$

Keeping in mind that $G = \hbar c/m_P^2$, we find $l_P = m_P c$ for the Planck length. In this case, the quantum fluctuations of the gravitational field are of the order of magnitude of the field itself. We come into the quantum gravity region. In order that such a situation be realized, it is necessary that the dimensions of the particles themselves be no larger than l_P.

The differential cross-section for elastic gravitational scattering $d\sigma/dq^2$ for two energetic point particles is of the order of

$$d\sigma/dq^2 \sim G^2E^4/q^4,$$

where E is their energy in the center of mass system, and q is the transferred momentum (units of $c = 1$ and $\hbar = 1$ are used). For $E \sim q \sim m_P$, the cross section $d\sigma/dq^2 \sim 1/m_P^4$, and the unitarity of the amplitude is saturated in the s-wave. However, the formation of a black hole becomes important at such energies and transferred momenta. Already John Michell[51] (in 1784) and Pierre S. Laplace[52] (in 1796) spoke of black holes.[6)] At $r < r_g = 2Gm/c^2$, the potential energy of the gravitational attraction for a photon "attempting" to escape from a body with mass m and a radius smaller than r_g is larger than its kinetic energy.

The smaller is the mass of the body, the smaller is its gravitational radius r_g. For such astronomical objects as ordinary stars, $r_g \ll R$, where R is the star's radius (for example, for our Sun, $r_g \approx 3$ km). If, as a result of evolution, a mass of the order of the solar mass is concentrated in a region with a radius smaller than r_g, then a black hole arises. Such black holes arise during the explosions of supernovae.

If one imagines a system consisting of two identical black holes which are located at a distance from each other of the order of r_g, then the potential energy of their attraction would equal

$$\frac{Gm^2}{r_g} \sim \frac{Gm^2c^2}{Gm} \sim mc^2,$$

and the gravitational field strength (force) is

$$\frac{Gm^2}{r_g^2} \sim \frac{Gm^2c^4}{G^2m^2} \sim \frac{c^4}{G} \sim \frac{c^3m_P^2}{\hbar},$$

i.e., it would be characteristically Planckian. Thus, one can obtain the Planck strength of a gravitational field in a mental experiment with two macroscopic black holes. However this field would be classical; its fluctuations on the Planck scale would still be small.

If there existed particles with $m \sim 0.01m_P$ to $m \sim 0.1m_P$, then the gravitational attraction between them would lead to the formation of atom-like quantum systems, in which their motions would be non-relativistic. These systems would be described by non-relativistic quantum gravity (see Fig. 2). The present concentration of such particles must not significantly exceed $\sim 10^{-16}$ of the concentration of ordinary hydrogen. Otherwise they would already make an unacceptably large contribution to the mass of the dark matter. The detection of such particles is a very difficult task.

In particle physics, the Planck scale could turn into a mirage if the particles (including the graviton?) had intrinsic sizes l such that $l \gg l_P$. In this case the Planck sizes would, even in principle, be unattainable in particle interactions. A large number of papers (see the books of Ref. 53, the popular science paper by A. D. Sakharov,[54] and several recent theoretical papers[55]) are devoted to attempts to construct a non-local field theory. However, no one has succeeded in constructing a consistent fundamental non-local field theory (truly, and not phenomenologically, non-local): casuality is violated. In this sense, the quest for internal consistency of the theory drives us to Planck distances.

A good mental experiment for thinking out how non-locality would affect gravitational interaction is the collision of two particles with a specified impact parameter $l \gg l_P$ and with such high (super-Planck $E \gg m_P c^2$) energies in the center of mass system that $G(E/c^2)^2/l \sim E$. After approaching to the distance l, these particles would form a black hole.

A. D. Sakharov wrote in Ref. 54: "At present, more and more physicists are leaning towards the position that specifically the boundary L_0 will determine the most significant changes in our ideas.

It is nevertheless very important to become convinced that no intermediate characteristic length between $r = 2.8 \cdot 10^{-13}$ cm and $L_0 = 1.61 \cdot 10^{-33}$ cm plays such a fundamental role.

8. CONCLUSION

In the main stair landing of the Physics Department of the University of Padua, the university at which Galileo Galilei lectured in his time, hangs a marble slate with his words which, translated into English, sound approximately so:

"In my opinion, it is better to find the truth if only in a small matter than to argue at length about very great questions without having obtained any truth at all."

These words of Galileo[56] express the credo of every professional physicist. But what, besides purely subjective reasons, impelled me to violate the precept of Galileo? I think that, first of all, it is that the experimentally inaccessible Planck mass occupies a larger and larger place in physics with every year. This forces one to look for in some sense a metaphysical frame for a physical picture of the universe.

In 1918, A. S. Eddington,[57] following Planck in emphasizing that, of all the physical systems of units, the $cG\hbar$ system is absolutely the preferred one, noted that the Planck length "must serve as the key to some very significant structure." Several years later, in his excellent in all remaining respects book "*Dimensional Analysis*," P. W. Bridgman,[58] after ridiculing this statement, declared that these units have no relation to real physics. By the way, the first and last Russian edition of this book appeared almost 60 years ago, and it would be very useful to reprint it. But as far as disputes with Eddington and Planck go, the pan of the balance moves steadily in their direction. Even within ordinary astrophysics the Planck units are accepted in our time as the natural ones (see the book by É. Dibai and S. Kaplan[59]).

Over the time of working on this paper (August 1990 to May 1991), I discovered that a number of ideas, which I had at the start or which occurred to me in the course of the work, had been published earlier by other authors. So that only certain details, accentuations, comments, and comparisons are properly mine.

The idea of the plurality of universes has been discussed in the scientific literature since the times that scientific literature itself appeared.[60] Many striking words have been said about it. I should like to quote the words with which the Nobel Lecture by Andreĭ Dmitrievich Sakharov[61] is concluded:

"I also defend the cosmological hypothesis, according to which the cosmological development of the universe is repeated in its main features an infinite number of times. Here other civilizations, including more "successful" ones, must exist an infinite number of times on the pages of the Book of the Universe "preceding" and "following" our universe. But all this must not belittle our holy quest specifically in this universe, where we, like a flare in the darkness, arose for one instant from the black void of the unconscious existence of matter, to achieve the requirements of Reason and to create a life worthy of ourselves and of The Purpose that is vaguely guessed by us."

I thank M. Baldo-Cheolin, R. Barbieri, V. I. Gol'danskiĭ, G. E. Gorelik, Yu. A. Danilov, D. D. Ivanenko, G. M. Idlis, V. I. Kisin, V. I. Kogan, L. Maiani, A. Yu. Morozov, D. K. Nadezhin, N. A. Nekrasov, Yu. F. Orlov, I. L. Rozental', J. Sal'vini, S. G. Tikhodeev, and É. Shuging for useful comments and/or for indicating reference sources previously unknown to me.[7)]

[1)] At first glance, the statement that the interaction of photons with matter vanishes seems to contradict the circumstance that the cross section for the resonance scattering of light increases as λ^2. However, with the enormous lifetime and vanishingly small width of a resonance line, it is practically impossible to get into resonance.

[2)] If one leaves out of consideration the indirect reference of Mr. Tompkins' initials: C. G. H.

[3)] Let us note that the binding energy of an electron in a positronium ion is ~ 0.2 eV, and the binding energy of a molecule consisting of two positronium atoms is ~ 0.1 eV (for example, see Ref. 35). So that a condensed state of them could exist at very low temperatures in the case when $m_e = m_p$.

[4)] However, in the case of universes absolutely isolated from each other, the statement about the generality of the dimensional units c, G, and $\hbar$ does not have operational meaning. If one imagines a universe in which the values of c, G, and $\hbar$ expressed in our grams, centimeters, and seconds are different than for us, but all the dimensionless parameters which we discussed above are the same as for us, then all of physics in that and in our universes will be indistinguishable. Of course, their grams, centimeters, and seconds will be different from ours, but the numerical values of c, G, and $\hbar$ expressed in their grams, centimeters, and seconds will be the same as for us.

[5)] At a temperature of the order of several TeV a phase transition occurred in which the baryon asymmetry of the universe[46-50] was apparently formed in final form.

[6)] In the appendix to the Russian translation of Pierre Simon Laplace's book "*Exposition du Systeme du Monde*,"[52] the translator quotes the following statement by Laplace: "A shining star with a density equal to that of the Earth and with a diameter 250 times larger than the Sun's diameter will not allow a single light beam to reach us because of its gravity; therefore, it is possible that the most luminous celestial bodies in the universe turn out to be invisible for this reason." It is only strange that the quotation is given with a somewhat unusual reference: the newspaper "*Leninskaya Pravda*," the edition of December 21, 1980.

7) Of course, the reference list lays no claim to be either a complete or even a well thought out selection. These are simply those books and papers which I used, read or scanned during my work on this paper. Those pages or paragraphs which have a direct bearing on the questions that I have discussed are marked for some references.

[1] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Pergamon Press, N.Y., 1975 [Russ. original, Nauka, M. 1973 and 1988, §§ 16, 27, 26, 30, and 78].

[2] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Pergamon Press, Oxford, 1982 [Russ. original, Nauka, M., 1973 and 1989, §§ 45, 56, and 59].

[3] O. A. Vol'berg, "An entertaining trip into Einstein's country," in Ya. I. Perel'man, Entertaining Mechanics (In Russian) GRN-PYul, L., M., 1935, Ch. 11, p. 188.

[4] G. Gamow, Mr. Tompkins in Paperback. Containing Mr. Tompkins in Wonderland and Mr. Tompkins Explores the Atom, Cambridge University Press, Cambridge, England, 1988.

[5] M. Planck, Selected Works [Russ. transl., Nauka, M., 1975, p. 232].

[6] G. Gamov, D. Ivanenko, and L. D. Landau, Zh. Russ. Fiz. Khim. Ova. Chast' Fiz. **60**, No. 1, 13 (1928).

[7] L. D. Landau, in *Niels Bohr and the Development of Physics*, (W. Pauli, ed.), McGraw-Hill, N.Y., 1955 (Russ. Transl., IL, M., 1958, p. 85).

[8] J. A. Wheeler, Phys. Rev. **97**, 511 (1955).

[9] J. A. Wheeler, *Gravitation, the Neutrino, and the Universe* [Russ. transl., IL, M., 1962].

[10] G. E. Gorelik, Einstein Volume 1978-1979, (In Russian), Nauka, M., 1983, p. 334.

[11] G. E. Gorelik, *The Dimensions of Space. A Historical and Methodological Analysis*, (In Russian) Izdatel'stvo Moskovskogo Universiteta, M., 1983.

[12] G. E. Gorelik and V. Ya. Frenkel', *Matveĭ Petrovich Bronshteĭn*, (In Russian) Nauka, M. (1990).

[13] A. M. Polyakov, *Gauge Fields and Strings*, Harwood Acad. Publ., 1987.

[14] M. B. Green, J. Schwartz, and E. Witten, *Superstring Theory*, Cambridge Univ. Press, Cambridge, England, 1987.

[15] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*, W. H. Freeman, San Francisco, 1973 [Russ. Transl., Mir, M., 1977, Vol. 3, § 44].

[16] S. W. Hawking, *A Brief History of Time*, Bantam Books, N.Y., 1988.

[17] S. Coleman, Nucl. Phys. B **310**, 643 (1988).

[18] S. Weinberg, Usp. Fiz. Nauk **158**, 679 (1989) [Sov. Phys. Usp. **32**, 697 (1989)].

[19] M. Gell-Mann and J. Hartle, *Quantum Mechanics in the Light of Quantum Cosmology* Preprint CALT-68-1647, UCSB TH 89-36. *Proceedings of the Santa Fe Institute Workshop on Complexity, Entropy, and the Physics of Information*, May 1989, *Proceedings of the Third International Symposium on the Foundations of Quantum Mechanics in the Light of New Technology*, Tokyo, August 1989.

[20] M. Gell-Mann, Invited Talk, in *Proceedings of the 25th International Conference on High Energy Physics*, August 1990, World Scientific, Singapore, 1991.

[21] A. D. Linde, *The Physics of Elementary Particles and Inflationary Cosmology*, (In Russian), Nauka, M., 1990.

[22] A. D. Linde, *Inflation and Quantum Cosmology*, Stanford University Preprint ITP-878, September 1990.

[23] J. J. Halliwell, *Quantum Cosmology in the Creation of the Universe*, CTP Preprint No. 1928 (Massachusetts Inst. of Technology), December, 1990; Submitted to Scientific American.

[24] A. L. Zel'manov, *Infinity and the Universe*, (In Russian), Mysl', M., 1960, p. 274

[25] A. L. Zel'manov, *Development of Astronomy in the USSR* (In Russian), Nauka, M., 1967, p. 320.

[26] Z. Kazimir, *Niels Bohr. His Life and Work* (In Russian) M. Mysl', 1967, p. 190.

[27] I. L. Rozental', Usp. Fiz. Nauk **131**, 239 (1980) [Sov. Phys. Usp. **23**, 296 (1980)].

[28] I. L. Rozental', Elementary *Particles and the Structure of the Universe*, (In Russian) Nauka, M., 1984.

[29] I. L. Rozental', *Geometry, Dynamics, and the Universe*, (In Russian) Nauka, M., 1987.

[30] I. L. Rozental', *Interpreting the Problems of Elementary Particle Physics*, (In Russian), Space Research Institute Preprint No. P-1671, M., 1990.

[31] C. A. Barnes, D. D. Clayton, and D. N. Schramm (editors), *Essays in Nuclear Astrophysics: Dedicated to W. Fowler on His Seventieth Birthday*, Cambridge University Press, Cambridge, England, 1982. (Russ. Transl., Mir, M., 1986).

[32] F. Hoyle, *Galaxies, Nuclei, and Quasars*, Harper and Row, N.Y., 1965 (Russ. Transl., Mir, M., 1968).

[33] F. Ajzenberg-Selove, Nucl. Phys. A **33**, 55 (1985).

[34] F. Dyson, Sci. Am. **22**, No. 3, 50 (1971).

[35] V. I. Gol'danskiĭ, *The Physical Chemistry of the Positron and Positronium* (In Russian), Nauka, M., 1968, p. 20.

[36] G. M. Idlis, Izv. Astrofiz. Inst. Akad. Nauk KazakhSSR **7**, 39 (1958).

[37] R. H. Dicke, Nature **192**, 440 (1961).

[38] B. Carter, *Cosmology. Theories and Observations* (Russ. Transl., Mir, M., 1978, p. 369).

[39] B. J. Carr and M. J. Rees, Nature **278**, 605 (1979).

[40] P. C. W. Davies, *The Accidental Universe*, Cambridge Univ. Press, Cambridge, N.Y., 1982, [Russ. Transl., Mir, M., 1985].

[41] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principles*, Clarendon Press, Oxford, 1986.

[42] Ya. B. Zel'dovich, Pis'ma Astron. Zh. 7, 579 (1981) [Sov. Astron. Lett. 7, 322 (1981)].

[43] A. D. Sakharov, Zh. Eksp. Teor. Fiz. **87**, 375 (1984) [Sov. Phys. JETP **60**, 214 (1984)].

[44] CERN Courier **21**, No. 2, 1 (1991); **21**, No. 3, 3 (1991).

[45] J. Ellis, S. Kelley, and D. V. Nanopoulos, Phys. Lett. B **249**, 441 (1990); M. Einhorn and D. R. T. Jones, Nucl. Phys. B **196**, 475 (1982); J. Ellis, D. Nanopoulos, and S. Rudaz, Nucl. Phys. B **202**, 43 (1982); W. Marciano and G. Senjanovich, Phys. Rev. D **25**, 3092 (1982).

[46] N. S. Manton, Phys. Rev. D **28**, 2019 (1983).

[47] F. R. Klinkhamer and N. S. Manton, Phys. Rev. D **30**, 2212 (1984).

[48] V. A. Kuzmin, V. A. Rudakov, and M. E. Shaposhnikov, Phys. Lett. B **155**, 36 (1985).

[49] V. A. Rudakov, Nucl. Phys. B **256**, 509 (1985).

[50] P. Arnold and L. McLerran, Phys. Rev. D **36**, 581 (1987); D **37**, 1020 (1988).

[51] J. Michell, Philos. Trans. R. Soc. London **74**, 35 (1784); W. Israel, *Three Hundred Years of Gravitation* (S. W. Hawking and W. Israel, (Eds)), Cambridge Univ. Press, Cambridge, 1987, p. 199.

[52] P. S. Laplace, *Exposition du Systeme du Monde* (J. B. M. Duprat (Ed.)), Vol 2, Paris (1976), (Russ. Transl., Discussion of Systems of the Universe, Nauka, M., 1982, see V. M. Vasil'ev, Appendix, p. 368). S. W. Hawking and G. F. R. Ellis, *The Large-Scale Structure of Space-Time*, Cambridge Univ. Press, Cambridge, England (1973), p. 365.

[53] *Problems of Theoretical Physics. Symposium in Memory of Igor Evgen'evich Tamm*, (In Russian), Nauka, M., 1972; G. V. Efimov, *Non-local Interactions of Quantum Fields*, (In Russian), Nauka, M., 1977; G. V. Efimov, *Problems of the Quantum Theory of Non-local Interactions*, (In Russian), Nauka, M., 1985.

[54] A. D. Sakharov, "Does an elementary length exist?" (In Russian), Fizika v Shkole, No. 2, 6 (1968); *A. D. Sakharov: Studies for a Scientific Portrait. With the Eyes of Friends and Colleagues. Free Thinking*, (In Russian), Fizicheskoe Obshchestvo SSSR, Mir, M., 1991.

[55] J. M. Moffat, Phys. Rev. D **41**, 1117 (1990). D. Evens, J. W. Moffat, G. Kleppe, and R. P. Woodard, Phys. Rev. D**43**, 499 (1991); J. W. Moffat, Mod. Phys. Lett. A **6**, 1011 (1991).

[56] A. Favaro (Ed.), *Le Opere di Galileo Galilei, Vol 4*, p. 738, Note 2.

[57] A. S. Eddington, Proc. Phys. Soc. London **30**, 91 (1918).

[58] P. W. Bridgman, *Dimensional Analysis* (Russ. Transl. from Second English Edition (S. I. Vavilov, (Ed.)) (1934), Ch. 8, pp. 105–110).

[59] É. A. Dibaĭ and S. A. Kaplan, *Dimensionality and Similarity of Astrophysical Quantities*, (In Russian), Nauka, M., 1976.

[60] V. P. Vizgin, *The Idea of a Multiplicity of Universes* (In Russian), Nauka, M., 1988.

[61] A. D. Sakharov, *Peace, Progress, and The Rights of Man, Nobel Lecture*, Read by E. Bonner December 11, 1975, at the Swedish Academy of Sciences; *Appendix to Literaturnaya Gazeta*, January 1990.

Translated by Frederick R. West

This page intentionally left blank

THE PROBLEM OF MASS: FROM GALILEI TO HIGGS

L.Okun

ITEP, 117259, Moscow, USSR

The subject of my talk was chosen by Professor Zichichi and the title was formulated by him. I have never given such historical talks before: about four hundred years of historical development in some field. Therefore I will stick to physics. My talk will consist of ten parts:

1. Introduction
2. The mass of a particle
3. The mass in XVI-XVII centuries
4. The mass in XVIII-XIX centuries
5. The mass of a system of free particles
6. Interactions and binding energy
7. Quarks and gluons confined in hadrons
8. The pattern of masses of elementary particles
9. Scalars and the nature of mass
10. Conclusions.

As you can see, there will be only two historical parts.

1. INTRODUCTION

The great Galilei started our science, physics, by fighting the prejudices and by bringing to this science its main constituents: observation, experiment, mathematics — and last but not least — common sense.

All these elements could be traced back in the works of his predecessors, but I will not do it. Books about Galilei start usually by castigating Aristotle as a silly person. I will break with this tradition. On the other hand, I will not praise Lucretius Carus whose ideas were often rather close to those ideas of Galilei.

What I would like to do first of all is to define what we are speaking about: to make clear what we mean when we say the word "mass", to make clear the relations between mass and energy, between mass and gravity. It seems to me that only by defining notations, terminology, language would it be possible to go in fifty minutes from Galilei to Higgs.

Physics at the Highest Energy and Luminosity, Edited by A. Zichichi
Plenum Press, New York, 1992

In the booklet about Erice, which all of us were given by hospitable organizers of the school, you can read that "Erice is a singular synthesis of myth and history ... of fantasy and reality". This applies also to the subject of my talk.

The main message of my talk is that there is only one mass. This message is connected with the poster of our school, according to which $m_i = m_g$. I will insist that there is no m_i, no m_g, no other "indexed" masses, there is only one mass m.

2. THE MASS OF A PARTICLE

In special relativity the mass is defined as a Lorentz invariant quantity defined by the square of the energy-momentum four-vector of a body:

$$m^2 = E^2 - \vec{p}^{\,2}$$

(the velocity of light is used as the unit of velocity).

The velocity of a body is defined by its momentum and energy:

$$\vec{v} = \vec{p}/E.$$

If $m = 0$ (the massless particle) then $|\vec{p}| = E$ and hence $v \equiv |\vec{v}| = 1$.

If $m \neq 0$, then for a body at rest we have

$$v = 0, \quad \vec{p} = 0, \quad m = E_0,$$

where E_0 is the rest energy. By using our starting equations it is easy to express energy E and momentum $\vec{p}$ of a massive particle as functions of its velocity $\vec{v}$:.

$$E = m\gamma,$$

$$\vec{p} = m\vec{v}\gamma,$$

where

$$\gamma = 1/\sqrt{1 - v^2}.$$

It is important to stress that the difference between energy E and the rest energy E_0 is the kinetic energy T:

$$T = E - E_0.$$

When $v \ll 1$, then from

$$(E - m)(E + m) = \vec{p}^{\,2}$$

it immediately follows that

$$T = E - m \simeq \vec{p}^{\,2}/2m$$

and

$$\vec{p} \simeq m\vec{v}.$$

Now, there exists relation between force and the time derivative of momentum:

$$\vec{F} = \frac{d\vec{p}}{dt}$$

which is valid at arbitrary values of v.

In the case of $v \ll 1$,

$$d\vec{p}/dt = md\vec{v}/dt = m\vec{a}.$$

In the non-relativistic case, therefore,

$$\vec{F} = m\vec{a},$$

and, hence, the mass is the measure of inertia: the smaller the mass of a body the easier to accelerate it with a given force.

In the general case, which has to be considered when v is of the order of unity, the differentiation of $\vec{p} = m\vec{v}\gamma$ gives a more complex equation:

$$\vec{F} = \frac{d\vec{p}}{dt} = m\gamma[\vec{a} + \vec{v}(\vec{v}\vec{a})\gamma^2],$$

or, taking into account that $m\gamma = E$,

$$\vec{a} = [\vec{F} - (\vec{F}\vec{v})\vec{v}]/E.$$

You see that in the general case $\vec{a}$ is not directed along $\vec{F}$, and neither m nor E play the role of the measure of inertia.

Of course, one can define velocity and acceleration of a particle in a four-dimensionally covariant way by using instead of t the proper time $\tau(\tau = t/\gamma)$:

$$u^i = dx^i/d\tau \ , \ \dot{u}^i = du^i/d\tau \ .$$

Then the energy-momentum four-vector p^i is proportional to the velocity:

$$p^i = mu^i \ ,$$

and the four-vector of fource f^i is proportional to the acceleration:

$$f^i = dp^i/d\tau = m\dot{u}^i \ .$$

However the clocks usually do not accompany particles in their flights, they are at rest in the laboratory. Therefore we have to express $u^i, \dot{u}^i$ and f^i through $\vec{v}, \vec{a}$ and $\vec{F}$ — the quantities that one actually measures:

$$\begin{aligned} \vec{u} &= \vec{v}\gamma \ , \ u^0 = \gamma \ ; \\ \dot{\vec{u}} &= \gamma^2[\vec{a} + \vec{v}(\vec{v}\vec{a})\gamma^2] \ , \dot{u}^0 = (\vec{v}\vec{a})\gamma^4 \ ; \\ \vec{f} &= \gamma\vec{F} \ , \ f^0 = m(\vec{v}\vec{a})\gamma^4 \ . \end{aligned}$$

This brings us back to the equations we started with.

In many books you may read about relativistic mass $m = E/c^2$ which increases with velocity, about rest mass $m_0 = E_0/c^2$, about equivalence between mass and energy. Many authors even use the term mass-energy. This terminology is extremely wide-spread and extremely misleading. It stems from the tradition of keeping the non-relativistic relation for relativistic particles: $\vec{p} = m\vec{v}$ instead of $\vec{p} = m\gamma\vec{v}$. Please, don't use this archaic jargon. Use the professional language, which is adequate and precise.

Now let us consider the gravitational interaction of a particle. The key idea of general relativity — the classical theory of gravitation — is to postulate the universality (mass independence) of gravitational acceleration, as proved with increasing accuracy by the experiments of Galileo, Newton and Eötvös. For a non-relativistic particle the coupling to gravitational field is proportional to its mass. However, for a relativistic particle this coupling is proportional to energy-momentum tensor T^{ik}. This tensor is symmetric and therefore has ten components. Gravitational properties of a particle are thus described in a general case not by one quantity — mass, but by ten quantities — components of energy-momentum tensor.

I will not go into full discussion of general relativity, instead I will consider a simple case of a relativistic particle in a weak gravitational field. In this case the energy-momentum tensor T^{ik} may be presented in the form

$$T^{ik} = mu^i u^k \frac{d\tau}{dt} = \frac{p^i p^k}{E},$$

where, as before,

$$p^i = mu^i, \quad u^i = \frac{dx^i}{d\tau},$$

τ - the proper time of a particle, $dt/d\tau = \gamma$. I have omitted a factor which is non-essential to us, namely, $\delta(\vec{r} - \vec{R})$, (see "Classical Field Theory" by L.Landau and E.Lifshitz, sections 87 and 7).

When $v = 0$, the only non-vanishing component of T^{ik} is $T^{00} = m$, where m is just the same mass we have considered before. $T^{00} = m$ remains the dominant component also for $v \ll 1$. In this nonrelativistic limit the gravitational interaction between two bodies of masses M and m is described by the Newtonian potential

$$V = -G_N \frac{Mm}{r}$$

where G_N is the Newtonian gravitational constant:

$$G_N \simeq 6.7 \cdot 10^{-39} GeV^{-2}$$

(we use units in which $\hbar, c = 1$).

In modern physics, there is only one mass m which at low velocities determines both the inertial and gravitational properties of a body ($m_i \equiv m_g \equiv m$). As for high velocities, the mass does not determine either inertia or gravity. Therefore, the notions of m_i and m_g are redundant for slow particles and are irrelevant and even false for fast particle.

3. THE MASS IN XVI-XVII CENTURIES

Now, after giving a short review of modern "mass terminology", I am going to state that Galileo Galilei, in full agreement with modern science, has never written the equation $m_i = m_g$. We have no $m_i = m_g$ today and there was no $m_i = m_g$ four hundred years ago. (It was in between).

The present Galilei festivities are connected in time with the period (1589-1592) when he was a young professor at Pisa. Before that, in 1586, he wrote "La Bilancetta" ("The Small Balance"), a manuscript in which he considered a scale with weights immersed in various liquids. Due to this work a position of professor was given to him in Pisa (with a very modest salary).

In 1590 he prepared a collection of manuscripts dealing with mechanical motion under the title "De motu antiquiora". As I understand, the word "antiquiora" referred to his own "old" papers (at that time he was 26 years old), not to antique papers of ancient authors.

There was no "publish or perish" rule at that time: many manuscripts were not printed at all. Fragments of "De motu" were incorporated later by Galilei into his books: "Le meccaniche" (written in 1593, published in French in 1634 and in Italian in 1649, seven years after Galileo had passed away), "Dialogo" (1632) and "Discorsi" (1638).

There is no definition or discussion of the notion of mass in any of Galilei's writings. As you can easily guess, I have not studied all of them (they were first published in twenty volumes of "Opere" a century ago). But I did read some of them. And I trust historians of physics, who contend that the term "mass" was unknown to Galilei and to other scientists of his time. No m_i, no m_g, no mass at all. Instead of the word "mass" he used the word "weight".

Historians assert also that dropping balls from the Pisa tower was only a gedanken experiment (such experiment was actually performed much later at St.Paul Cathedral in London and was described by Newton in "Principia" (1687)). What Galilei really did were experiments with balls rolling on an inclined plane. These were very ingenious experiments, with bells ringing when balls were passing certain distances.

What was Galilei's main result? His main result was that he established that g – the gravity acceleration – is independent of weight, size, material of a body. It was only approximately true in the air, but he postulated that this is an exact law for the free fall in vacuum. In modern language this means that there is no "fifth force", which is presumed to depend on the material of a body (on its baryon or lepton numbers, or proton-neutron ratio).

It was Newton who introduced the notion of mass. He defined mass as a quantity of matter. (This definition has repercussions even today. In many popular science books and articles, you may read about transformation of matter into energy, say, in electron-positron annihilation. But the fact is that energy is conserved and what is transformed are particles: electron and positron into photons.)

According to Newton, the mass of a body is proportional to the body's density and volume. This definition was considered by many physicists and philosophers as a kind of vicious circle, because it was not defined what density was. Some of them have proposed other definitions of mass. Others argued that identification of mass and matter by Newton implied in fact the atomic structure of matter. Without following this debate, let us note that this identification definitely implied conservation and additivity of mass.

Newton defined mass in "Principia" (1687) – in the main work of his life. Galilei was mentioned in the book ten times and with great respect. Newton first referred to Galilei in connection with the action at distance. (At that point he was arguing against Descartes). But he referred to Galilei mainly when describing parabolic trajectories in gravitational field. This parabola impressed Newton greatly. You can see it from the text of "Principia".

Looking from a distance of four centuries, one of the greatest achievements of Galilei is certainly the formulation of the relativity principle. This formulation is presented in "Dialogo" as a picture of a gedanken experiment in a cabin of a ship.

There are a lot of processes going on in the cabin: the water is dripping, fishes are swimming in an aquarium etc. Galilei insists that by observing all possible processes in a cabin with the windows covered one cannot tell whether the ship is at rest or in uniform rectilinear motion.

This picture is in fact much broader than what is called Galilean relativity or "non-relativistic relativity". In fact it has in it all of relativity. Indeed, there is light in the cabin, and the velocity of light is finite.

The velocity of light was first measured by Romer in 1676 by observing the motion of the satellites of Jupiter discovered by Galilei. As soon as you include the finiteness of velocity of signals in the picture of the cabin, you are immediately forced to generalize the "non-relativistic relativity" to full relativity in Einstein's sense.

The two great gedankenexperiments of Galilei were like two seeds. The campanile of Pisa was the seed of general relativity, the cabin of a ship was the seed of special relativity. The "Dialogo" and "Discorsi" were the cradle of the modern physics in general, and of the modern concept of mass, in particular.

4. THE MASS IN XVIII AND XIX CENTURIES

In order to trace the further evolution of the concept of mass I will mention here only three milestones.

First, it is the discovery by A.Lavoisier (1789) that mass is conserved and additive in chemical reactions.

Second, it is the discovery of the conservation of energy in the process of its transformation from its mechanical form into heat and vice versa (R.Mayer, 1842; H.Helmholtz, 1847).

Third, it is Maxwell's equations (1872) which have opened the direct road to special relativity.

These three discoveries are beyond mechanics and at first sight have little connection to the problem of mass. But I think that they are much more important for the real understanding of mass (remember: mass is equal to the rest energy) than many philosophical treaties on the definition of mass. Quite a number of them were written during these two centuries, discussing whether to define mass as a ratio of F over a, or consider it as a purely mathematical symbol in equations of mechanics, etc. But let us return from the history of physics to physics.

5. THE MASS OF A SYSTEM OF FREE PARTICLES

Let us consider n free particles. Both energy and momentum are conserved and additivie, therefore the total energy E and the total momentum $\vec{p}$ are

$$E = \sum_{i=1}^{n} E_i, \quad \vec{p} = \sum_{i=1}^{n} \vec{p}_i.$$

The word "additive" is extremely important when you speak about the difference between energy and mass. I am stressing this, because usually people somehow forget about this word which should be kept in mind.

The mass of a system of particles is expressed through E and $\vec{p}$ in the same way as the mass of a single particle:

$$m^2 = E^2 - \vec{p}^{\,2}.$$

As an example, let us consider the mass of a system of two photons:

$$m^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2 = 2(E_1E_2 - \vec{p}_1\vec{p}_2) = 2E_1E_2(1 - cos\vartheta)$$

You see that when photons fly in the same direction, $\vartheta = 0, m = 0$. When photons fly in opposite directions, $\vartheta = \pi, m = 2\sqrt{E_1E_2}$. In the rest frame of the system ($\vec{p} = 0$), the two photons fly in opposite directions with equal energies: $E_1 = E_2 = E_0/2$, so in that frame $m = E_0$: the mass of a system is equal to its rest energy. Please note that here E_0 is the rest energy of the system and that the rest energy of a photon is equal to zero.

Mass being equal to the rest energy, and energy being conserved, it is absolutely clear that mass is also conserved. That means that the mass of any isolated system before any reactions take place in it and afterwards is the same. But there is no additivity of mass. This is demonstrated by the two photons: each of them is massless, but the mass of the two photons may be zero or may be very large; it depends on the energy of the photons and on the angle between their momenta.

You see that the often claimed "equivalence" between mass and energy does not hold! Saying this, I fully realize that I contradict occasional statements of many a great physicist. But what I'm saying is not my invention. It is known to every professional theorist in particle physics. As for the great physicists, I believe that sometimes they did not care too much about dotting all the "i's" and crossing all the "t's", especially when addressing non-physicists.

It is true: whenever you have mass, you have energy. But you cannot reverse this statement. One may have energy and have no mass: an example is a photon.

For energy: conservation — yes! additivity — yes!

For mass: conservation — yes! additivity — no!

The energy is the time-like component of a four-vector in four-dimensional space-time. The mass is a scalar in the four-dimensional space-time. This is our professional language.

6. INTERACTIONS AND BINDING ENERGY

Let us consider now bound systems of particles. We have two long-range static potentials in Nature — Newton's and Coulomb's. We have Quantum Field Theory which describes all interactions including the Newton and the Coulomb potentials as the exchange of virtual particles. But the last statement should be taken with a grain of salt, because of an essential difference between these two potentials and all other potentials and forces. When you go to the limit $c \to \infty$, the former survive, while the latter (magnetic forces, Yukawa potential, exchanges of W-bosons and of gluons) die out.

Nature displays an impressive spectrum of bound systems: nuclei, atoms, molecules... galaxies... . Each of these systems has a certain mass defect which is equal to the difference between the sum of the masses of its constituents and the mass of a system as a whole. In the case of a nucleus, the constituents are nucleons. In the case of a galaxy, the constituents are stars and interstellar matter. With c as the unit of velocity, the mass defect is equal to the binding energy. The binding energy is defined as energy needed to pull the constituents of a body (at rest) apart.

As an extreme example of the mass defect, I shall consider the closed universe. For the universe as a whole the gravitational attraction of galaxies may compensate ("eat up") the sum of their masses (rest energies) and their kinetic energies. In this case the universe is closed and its total mass is equal to zero. For the observed velocities of Hubble expansion, the critical mean density above which the universe becomes closed is

$$\rho_c = 3H^2/8\pi G_N \simeq 10^{29}\,\mathrm{g/cm^3}.$$

Here G_N - is Newtonian constant and H is the Hubble constant: $H \simeq 100\,\mathrm{km/s}\cdot\mathrm{Mps}$.

The observed mean density of luminous matter is about $10^{-2}\rho_c$. There is strong evidence, however, that the actual density ρ is much larger. The evidence comes from the observation of motion of stars on the periphery of galaxies. These stars are moving too fast, thus indicating that the actual masses of galaxies are substantially larger than their visible masses. The same is true for clusters of galaxies. According to the inflationary cosmological scenario, $\rho \simeq \rho_c$ and the universe is flat.

When speaking about the Universe, one cannot avoid mentioning the problem of the energy density of vacuum, the so called cosmological term. According to all evidence, this energy density is vanishingly small. On the other hand, according to quantum field theory it must be tremendously large because of vacuum quantum fluctuations. This is one of the deepest mysteries of physics!

7. QUARKS AND GLUONS CONFINED IN HADRONS

One of the great pinnacles of physics is the discovery (or creation?) of Quantum Chromodynamics. It turned out that Quantum Field Theory can describe particles which have no mass in the normal sense, i.e. in the sense in which I used it before: mass as rest energy of a free particle. Quarks and gluons are never at rest and never free: they dash around inside hadrons, being jailed for life. Therefore, their masses cannot be defined in the usual sense. Although masses of quarks and gluons cannot be defined at large distances, they can be defined at short distances, deep inside their jail cells where they become asymptotically free. At short distances (large momentum transfers q), the role of the strong interaction becomes less and less important, thus we can see quarks almost naked inside their gluonic "fur coats". As for the mass of the "coat", it is about 300 MeV, being determined by the running of the strong coupling constant, $\alpha_s \sim 1/ln(\Lambda_{QCD}/q)$, and, thus, by the radius of confinement: $r_{conf} \sim 1/\Lambda_{QCD}$.

The Lagrangian of QCD is known, but the calculation of hadron masses is a very difficult problem. I will mention here two approaches to the problem: QCD sum rules and computer calculations on lattices.

The QCD sum rules are based on dispersion relations and asymptotic freedom. In this approach, dispersion relations serve as a bridge between the region of asymptotic freedom and the region of confinement. (I remind you that dispersion relations are based on two general principles: analyticity and unitarity of the S-matrix). With QCD sum rules, masses were calculated for a number of particles in the lower part of the hadron spectrum.

The space-time lattices allow calculations in the strong coupling limit and therefore are very promising.

8. THE PATTERN OF MASSES OF ELEMENTARY PARTICLES

The notion of elementarity is time-dependent. In 1991 the particles we consider as elementary are gauge bosons, quarks and leptons.

The "oldest" of the gauge bosons — photon — is "experimentally" massless. What do I mean by this? We usually consider it to be massless, but what we really know is that it is very light. The upper limits on its mass (in fact, lower limits for the Compton wave length) are given by astronomical observations of the magnetic fields of Jupiter and of the galactic magnetic field. These limits are very impressive, but we have no compelling theoretical reason, no theoretical principle that would require the photon to be massless.

Sometimes people say it is gauge invariance. But gauge invariance in the case of abelian symmetry is not sacred. For electrodynamics it means only that the photon is massless. It is tautology. An intrinsic anomalous magnetic moment of the electron is gauge invariant, but it would destroy the renormalizability of QED and the theory would become a mess. A tiny photon mass is gauge non-invariant, but it does not destroy renormalizability and its presence would not spoil the beautiful agreement between QED and experiment. The abelian gauge invariance is neither sufficient nor necessary.

Contrary to photons, gluons (and also gravitons) have to be massless on theoretical grounds: for the consistency of the theory. A tiny mass-term in QCD lagrangian makes that theory unrenormalizable because of ultra-violet divergences.

The masses of W- and Z-bosons essentially give us the Fermi scale, the scale that determines the masses of all known elementary particles. Another parameter which plays an important role in determining this scale is $\sqrt{\alpha_W}$, where α_W is the weak gauge coupling constant ($\alpha_W \sim 1/30$). As a result, the Fermi scale is in the ballpark of several hundred GeV.

Among all elementary fermions, the top quark seems to have the most natural value of the mass: of the order of Fermi scale. More precisely, on the basis of the data on electroweak radiative corrections and on the transitions $K^0 - \bar{K}^0$ and $B^0 - \bar{B}^0$ we expect, that m_t is somewhere between 100 GeV and 200 GeV. We have not yet discovered the top quark just because its mass is so natural.

The masses of lighter quarks are strongly suppressed with respect to Fermi scale:

$$m_b, m_c \sim \alpha m_t,$$

$$m_s \sim \alpha^{3/2} m_t,$$

$$m_d, m_u \sim \alpha^2 m_t.$$

We have no theory explaining this empirical pattern of suppression.

Charged leptons are lighter than their quark partners (down-components of quark doublets):

$$m_\tau < m_b, \quad m_\mu < m_s, \quad m_e < m_d.$$

Neutral leptons — neutrinos — are the most enigmatic particles: we know only upper limits on their masses. According to the 1990 Particle Data Booklet,

$$m_{\nu_e} < 17eV \text{ at } 95\% \text{ CL},$$

$$m_{\nu_\mu} < 0.27MeV \text{ at } 90\% \text{ CL},$$

$$m_{\nu_\tau} < 35MeV \text{ at } 95\% \text{ CL}.$$

Many years ago it was a common opinion that neutrinos are massless because of γ_5–symmetry. This option is preserved by the Standard Model which deals only with massless left-handed neutrinos and right-handed antineutrinos. But at present we know that all discrete symmetries are violated in nature: C, P, CP, T. (Only gauge symmetries stand). Why then respect the discrete γ_5–symmetry? Also, with the advent of grand- and superunification, nobody is afraid anymore of very big or very small numbers in physics. Therefore, the vogue now is to believe that neutrino masses are small but non-vanishing. Unfortunately, I have no time to discuss these extremely interesting particles. I will mention only that they may have Majorana masses in addition to the usual (Dirac) masses.

In connection with CP and all that, let me remark that CPT is okay. The equality of particle and antiparticle masses was tested in the case of K^0 and $\bar{K}^0$ with fantastic accuracy: $\Delta m/m \simeq 10^{-18}$. For all other particles the accuracy is more than a billion times worse, and special experiments to improve it are underway, in particular, for the equality of masses of proton and antiproton and of their gravitational properties. They will get improved limits, but I personally do not expect that any mass–difference will be discovered.

9. SCALARS AND THE NATURE OF MASS

The nature of mass is the most important unsolved problem of modern particle physics. It is really fascinating that this problem may be solved (at least partially) by experiments on future colliders such as SSC or LHC.

Theoretically, the nature and the origin of mass are connected with the expected existence of scalar particles, among which higgses are the most popular ones. Thus, we finally arrived from Galilei to Higgs.

Professor Higgs, undoubtedly, will discuss scalars in great detail in his closing lecture. I therefore consider that my task is accomplished and I shall confine myself only to a few remarks.

The existence of elementary scalars — higgses — is one of several possibilities. Another possibility are composite scalars, for instance, scalar which is a bound state of $t\bar{t}$. They have been discussed very extensively during the last two years. There is also technicolor, in which mass appears due to a confinement mechanism similar to that in QCD. The technicolor model predicts heavy technihadrons (with masses above 1 TeV) and comparatively light pseudogoldstone bosons (technianalogs of pions). Unfortunately, we have no selfconsistent technicolor theory.

If scalars are elementary, then supersymmetric particles must exist. These superparticles cancel quadratic divergences connected with elementary scalars. Otherwise the Fermi scale would blow up to the Planck scale. The minimal standard model Higgs scalar cannot be very heavy, definitely not heavier than 1 TeV. Otherwise the theory would become strongly inconsistent at the level of logarithmic divergences.

If there are no scalars at all, there must exist a strong WW–interaction at energies between 1 TeV and 2 TeV.

The lightest scalars (elementary or not) and superparticles, if they exist, would be discovered at SSC and LHC. But the full exploration of Scalarland and Superland is the task for a bigger machine.

I would like to make here a comparison with the study of ordinary hadrons. To study production and decay of pions, several hundred MeV was enough. To produce strange particles and antiprotons, energies above 1 GeV were needed. To answer that need, the Cosmotron and the Bevatron were built in the 1950's. But even today our understanding of the strong and weak interactions of quarks is far from being complete, and we need machines of much, much higher energies. Thus, it turned out that Quarkland stretches from subGev– to TeV–energies.

Similarly, it is quite natural to expect that the Scalarland and Superland would stretch from subTeV– to PeV–energies. It is fine that such machine as Eloisatron, with energy in subPeV–range, is under active discussion. But it would be better if its energy were raised to 1 PeV.

The detailed study of physics in the range up to 1 PeV may allow extrapolation to physics at the deepest level, the level of the Planck scale. As you know, this scale is determined by the three most fundamental constants: $\hbar, c$ and G_N. At present we believe that it is the Planck scale, where superunification of all interactions, including gravity, takes place. Clues to macroscopic physics are hidden at atomic and nuclear scales. Clues to atomic and nuclear physics are hidden at TeV and PeV scales. Clues to TeV and PeV physics are hidden at Planck scale.

The graphs of the three running gauge coupling constants published recently in CERN Courier may be considered as a crude example of the extrapolation I'm speaking about. These graphs, which are based on the data of DELPHI Collaboration at LEP bring additional support to the observation made a decade ago, namely, that the existence of superparticles drastically improves the convergence of the three couplings to a common grand unification point. (A critical analysis of the experimental uncertainties is given by F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi in a paper submitted to "Il Nuovo Cimento"). In a similar way, the extrapolation from the PeV scale to the Planck scale may become possible.

10. CONCLUSIONS

Galileo Galilei was the first who limelighted acceleration. Acceleration is the key word, if we think about his fundamental contribution to physics. His accelerators were gravitational. Our accelerators are electromagnetic. We have to build them with the highest possible energies and luminosities in order to unravel the origin of mass and the most profound properties of matter, including gravity.

Chairman: L. Okun

Scientific Secretaries: P. Hernandez, Z. Pluciennik, K. Qureshi, U. Vikas

DISCUSSION

– *U. Vikas:*

What according to you is the significance of the equivalence principle if you do not consider as any kind of a principle the equality of inertial mass and gravitational mass?

– *Okun:*

The local equivalence of gravitational field and accelerated reference frame does not become less significant if instead of using two letters with indices, m_i and m_g, you say that $m_i \equiv m_g \equiv m$ and use only one mass m.

– *U. Vikas:*

Now you are stating this almost like principle of equivalence. But how do we know that the mass involved in acceleration under a force different from gravitation is the same as that under the force of gravitation?

– *Zichichi:*

How do we know that the identity $m_i \equiv m_g \equiv m$ exists? The first man who discovered experimentally this equation was Galilei.

– *Okun:*

We know it from both experiment and theory. Experiments started by Galilei and continued by Newton have achieved very high accuracy during the XX century. As for theory, I refer here to special and to general relativity and to quantum field theory.

In the nonrelativistic limit, the acceleration of a body in a given gravitational field does not depend on its mass. You may call this mass inertial mass and denote it by m_i, but better call it simply mass and denote it by m. You may introduce gravitational mass m_g. Then you conclude from the above independence that m_g is proportional to m: $m_g = km$, where the coefficient k is arbitrary. You may choose, for instance, $k = 1/2$. Then you have to redefine the Newtonian gravitational constant G. However, it is better not to introduce m_g, not to introduce k, but directly conclude that the gravitational force is proportional to m. This immediately fixes the value of G. The indices i and g are thus redundant.

– *Sivaram:*

What will be the implications of a particle with a negative mass?

– *Okun:*

At first sight, the sign of the mass is irrelevant in relativistic quantum theory. In the case of the fermions, the unitary transformation:

$$\psi \rightarrow \gamma_5 \psi$$

changes the sign of mass in the Dirac equation. In the bosonic case, it is m^2 that enters the Klein-Gordon equation and hence the sign of m is unimportant. However, we should not forget about the relation between mass and energy:

$$m^2 = E^2 - \vec{p}\,^2.$$

The energy of a free particle is positive by definition: $E = +\sqrt{m^2 + \vec{p}\,^2}$. (The sign is the same for all particles, otherwise the vacuum would be unstable.) Therefore in order to preserve the equation $E_0 = m$ for the rest energy (in order not to write $E_0 = \mid m \mid$), we have to define mass as positive.

Forget for a moment about relativity and consider the non-relativistic relation $E_{kin} = mv^2/2$. With negative mass you have to change this equation into $E_{kin} = \mid m \mid v^2/2$. Those who don't want to change would obtain a terrible variant of perpetuum mobile: a particle which would accelerate itself by giving its energy to other, normal particles. But now remember relativity and the instability of vacuum in the case of negative rest energy.

– *Sivaram:*

Could you comment on the antiparticles behaving differently under gravitational force?

– *Okun:*

Experimentally, the equality of masses of particles and antiparticles have been checked with a very poor accuracy. Galilei taught us that we have to measure everything which can be measured and to make unmeasurable, measurable. If experimentalists can do the appropriate experiments, they should do them.

The inequality of masses of particles and antiparticles would involve the breaking of CPT-symmetry which in a consistent theory is difficult to imagine. Therefore I don't expect that such inequality would be discovered.

– *Ellis:*

I would agree with you if you restrict your attention to only a graviton exchange of spin 2, but these experiments would be also sensitive to the exchange of vector particles, which would change the sign of the coupling.

– *Okun:*

Yes, that would imply the fifth force, which is definitely very interesting to observe. What I meant is difference of masses, not attraction for particles and repulsion for antiparticles.

– *Sivaram:*

Would you comment on a massive graviton? I believe there is a result which states that, in the case of a massive graviton, the deflection of light in a gravitational field would be 3/4 of the Einstein value.

– *Okun:*

I do not know this specific result. Anyway, a massive graviton because of its longitudinal components would make the theory badly divergent. There is no continuous limit $m \to 0$ for graviton.

– *Junk:*

How would the gravitational force change when a particle is embedded in some medium, e.g., electron in a piece of silicon or neutrino in the sun?

– *Okun:*

I would say that there is nothing fundamental that happens to the gravitational force in a medium. There is no fundamental difference between an elementary particle embedded in silicon and a macroscopic body immersed in water. Both are not free, owing to the electromagnetic interaction with the medium.

– *Junk:*

It seems it might be more fundamental because of the quantum mechanical wave length.

– *Okun:*

A solar neutrino has different wave lengths inside and outside the sun. This is described by the refraction index in the same way as for a photon inside a piece of glass. But I don't see any fundamental problem here as far as gravity is concerned.

– *Rothstein:*

Why is the abelian QED with a massive photon renormalizible, while the non-abelian QCD with a massive gluon is not?

– *Okun:*

This is connected with the emission of longitudinal components of the gauge bosons. In the abelian case, the corresponding amplitudes are suppressed by a factor m/E, where m is the mass of the boson and E its energy. In the non-abelian case, they are enhanced by a factor E/m.

– *Hasan:*

If the Universe is closed and massless, why do we need massive particles at all? What is then the meaning of mass at all?

– *Okun:*

It is not up to us to choose the content of the Universe. Whether Universe is closed or not, nothing would change locally, here in Erice. The mass of a proton does not vanish when the mass of the Universe vanishes. The mass of the Universe is a global property, while the mass of a particle or a star is a local one. Even if you believe in Mach's principle, you cannot say that the mass of a proton is proportional to the mass of the Universe. The closed Universe has vanishing mass, open Universe has infinite mass.

– *Weselka:*

What does it mean to say that the Universe is massless?

– *Okun:*

It means that the absolute value of the potential energy of gravitational attraction between galaxies is exactly equal to the sum of their rest energies and kinetic energies.

– *Zichichi:*

The notion of the total energy of the Universe is not so obvious. There is no way to measure the total energy of a Universe externally.

– *Okun:*

No, but you can do it internally.

– *Ellis:*

In my opinion mass is just an index to label representations of the Lorentz group, which acquires its physical significance from the existence of an asymptotic Minkowski space. The Universe is not sitting in such a space and the comment you made that a closed Universe is massless seems to be independent of the possible existence of such an asymptotic Minkowski space. So how can we talk about the Universe having a mass?

– *Okun:*

There may be an asymptotic Minkowski space, there may be no asymptotic Minkowski space. We don't know. The result I'm quoting is derived in "The

Classical Theory of Fields" by L.Landau and E.Lifshitz, section 111. The closed Universe has vanishing total electric charge and vanishing total four-momentum, obtained by integrating over all space.

As for the external Minkowski space, there is a drastic difference between a massless Universe and a massless particle, e.g. the photon. The energy of the photon depends on the reference frame and there is no frame in which it is equal to zero. The energy of the massless Universe is equal to zero regardless of the choice of reference frame. A bubble of true vacuum created in false vacuum has a similar property. It is an object which is invariant with respect to Lorentz transformations.

– *Qureshi:*

If there is only one mass m, then is it a conceptual mistake on the part of those who use $m = m_0/\sqrt{1-\beta^2}$?

– *Okun:*

It is not an algebraic mistake, but it is a conceptual mistake. If you use units where $c = 1$, then you have two different notations, E and m, for the same quantity. Instead of one conservation law you have now two, for energy and mass, with only one symmetry behind them. You have to speak about mass of a massless photon. All this is very misleading.

– *Czyzewski:*

I would like to ask a question about two kinds of masses: m_i and m_g. If any other feature of matter, for example, baryon number, would give correction to the energy-momentum tensor and hence to the space-time curvature, would it mean that we have two different masses or would it be included in both of them?

– *Okun:*

It would be included in both of them if you use notations m_i and m_g. If there is a force connected with the baryon number, the acceleration of a body in the field of the Earth would depend on the material of the body, because now the Earth would have two different fields: gravitational and baryonic. But still there will be only one mass and the gravitational attraction will be proportional to the mass. There is no difference in this respect between new barionic force and the old Coulomb force.

– *Pluciennik:*

You mentioned in your lecture three ideas: closed Universe, fifth force and confinement. I would like to combine them in my question. Can it be that the Universe is closed due to a new confining force with confinement radius of astronomic scale?

– *Okun:*

A new confining force between particles of ordinary matter (electrons and nucleons) is impossible. The point is that such a force would imply a new degree of freedom like colour. You need at least two degenerate electrons, or protons, or neutrons. This would drastically change the Periodic Mendeleev Table. A new confining force may exist between heavy not-yet-discovered particles. But even in this case the radius of confinement cannot be larger than centimetres, otherwise the gas of relic "gluons" would be too hot ($\sim 3K$) to form "gluonic" strings.

– *Sanchez:*

You have presented mass in the context of gravity, whether relativistic or non-relativistic, at the classical level. When discussing the elementary particle mass spectrum without gravity, you talked, of course, at the level of Q.F.T.

I would like to comment about the combination of the two problems, namely the mass in the context of gravity at the quantum level. When the gravitational radius of a particle (which is proportional to the mass) becomes of the order of the Compton length (which is inversely proportional to the mass), we reach the Planck mass (or Planck energy) scale. This is the highest mass for point-like particles in Q.F.T., compatible with locality. The existence of this highest scale energy has several conceptual implications: a quantum theory of gravity must be finite and it must be a Theory of Everything (that is, all dimensionless numbers must be compatible in it and there is no hope to include gravity in a consistent way in the context of a renormalizable Q.F.T.). These implications are based on renormalization group arguments.

– *Okun:*

I agree with your comment. I would like also to add that the future theory beyond the Planck mass would not use our conventional idea of space-time. It would not involve a metric and would need new concepts for causality.

– *Gourdin:*

Why do you not define mass as an invariant of the Poincare group, P^2, using space-time translation operator P_μ?

– *Okun:*

I have defined mass as a Lorentz scalar: $m^2 = E^2 - \vec{p}^{\,2}$. I agree with you: it is better to define it as a Poincare scalar. This emphasises that conservation of energy and momentum is connected with the 4-dimensional translational symmetry.

– *Gourdin:*

My second question: why do you make a definite difference between an exact SU(3) in QCD with massless gluons and exact U(1) in QED with massless photon? Is it more important for theory or for experiment, this type of difference?

– *Okun:*

As I said this morning, the photon is massless, so to say, experimentally. We know from astronomical observations that the upper limit on its mass is very small. But a non-vanishing mass within this limit will not spoil the agreement between QED and experiment. As for the gluons, they are massless theoretically. Even a tiny gluon mass would make QCD non-renormalizable.

– *Khoze:*

Could you, please, comment on a scenario of a Higgs boson connected with the $t\bar{t}$ condensate.

– *Okun:*

Electroweak radiative corrections tell us that the top quark is not very heavy: $m_t \leq 2m_W$. For the Yukawa coupling of the top, this gives: $g_t^2 \leq 4\alpha_w \sim 1/7 \ll 1$. This coupling is not large enough to have fast running and to produce $t\bar{t}$ condensate. So special four-fermion interaction at much higher energies has to be introduced. The $t\bar{t}$-scenario would be more attractive in the case the mass of the top is of the order of 0.3 TeV or higher.

– *Neubert:*

I would like to come back to the question of the photon mass. Since gauge invariance is a fundamental concept of field theory would not you agree that the discovery of the photon mass would also be a disaster for theory?

– *Okun:*

My point is that gauge invariance in the case of abelian theory is not so fundamental. The non-abelian gauge invariance is sacred, but the abelian is not.

– *Zichichi:*

We had a special session on Gauge Invariance here in 1982, (P.A.M. Dirac, S. Ferrara, H. Kleinert, A. Martin, E.P. Wigner, C.N. Yang and A. Zichichi, "Special Session on Symmetries and Gauge Invariance", in *"Gauge Interactions"*, Plenum Press, New York, p. 125-140, 1984). It would be interesting, as this is a school, if you illustrate the link between the origin and the present meaning of gauge invariance.

– *Okun:*

It seems to me I have to make a digression into history. The term gauge invariance was coined by Hermann Weil in 1919, when he was (unsuccessfully) trying to create a unified theory of electromagnetism and gravity. In German it was Eichinvarianz and it was used by Weyl in its direct sense of scale invariance (Maßstabinvarianz).

With the advent of quantum mechanics, Vladimir Fock in 1926 invented the Klein-Fock-Gordon equation (after Klein but before Gordon) and discovered that the equation is invariant under the transformation

$$A_\mu \to A_\mu + df/dx_\mu, \quad \varphi \to \varphi e^{if}.$$

Fock called it gradient transformation. Note that if i is dropped, the phase factor becomes a scale factor. This observation was made in 1927 by F.London, who thus related the new gradient transformation to Weyl's old gauge transformation. In 1929 Weyl proclaimed Eichinvarianz (in the sense of Fock's gradient invariance) as a general principle. The old name was given thus to a new concept.

– *Neubert:*

But gauge invariance is a fruitful concept to build theories. And I think it would really change the understanding of field theory, of forces if you would not have this concept.

– *Okun:*

Historically, of course, the idea of electromagnetic gauge invariance played a very important role. When the notion of isospin appeared, the abelian gauge invariance was generalized to non-abelian in the late 1930's by Oscar Klein. Then in the 1950's this generalization was independently introduced by Yang and Mills. Today the non-abelian gauge invariance is the basis of the Standard Model. But the abelian gauge invariance, in spite of its historical importance, is not important physically. You may violate it without any penalty.

– *Zichichi:*

Why is there a catastrophe in non-abelian case? What do you lose when you give a tiny mass to gluons?

– *Okun:*

Photons are truly neutral: they do not emit photons. Gluons are charged in the sense of colour. Because of their colour charges, gluons emit gluons. Here is the drastic difference between photons and gluons. If you give mass to gluons via Higgs mechanism, you lose confinement. If you give mass to gluons by hand, you lose renormalizability, because of the catastrophic emission of longitudinally polarized gluons.

– *Zichichi:*

It would be great to break confinement.

– *Hernandez:*

Is it possible to consider the Standard Model, with gauge symmetry broken by explicit mass terms, as phenomenological model? Is it possible to describe in this way, by effective chiral Lagrangian, the non-unitarity in WW-scattering?

– *Okun:*

It is possible. But just because of this non-unitarity the phenomenological model would become non-self-consistent at high energies and of course loop diagrams would be divergent.

– *Danielsson:*

You mentioned dark matter. It explains the discrepancy between observations of stars rotation and our naive expectations. Could it not be that something is fundamentally wrong with our theories and with the notion of mass on a very large scale? Can you give an example of a really good test of gravity at a very large distance?

– *Okun:*

I do not know any other observations except those I have mentioned. If you do not believe in dark matter then these observations could be interpreted as evidence for anomalously strong gravity at large distances. But I don't know any reasonable theoretical framework for discussion of such anomalous behaviour of gravity.

– *Zichichi:*

Several months ago I have seen a paper in which it was claimed that the amount of luminous matter was grossly underestimated because the interstellar space is in fact less transparent than it is usually assumed.

– *Ellis:*

I think that this claim was not generally accepted and certainly it would be very difficult to explain constant rotational velocity in that way.

– *Okun:*

I have not seen this paper.

– *Duff:*

With three parameters G, $\hbar$, c we can form a fundamental mass, a fundamental length and a fundamental time. This is in accord with what we were taught in high school that there are the three basic dimensions of mass, length taught in high school that there are the three basic dimensions of mass, length and time. Does this mean that the Theory of Everything must have at least three fundamental parameters?

- *Okun:*

Yes, there must be three dimensional fundamental constants. The arguments I am going to present were first formulated, slightly differently, by Matvei Petrovich Bronshtein in the paper "On the possible theory of the world as a whole" which was published (in Russian) in 1933 (Uspekhi Astronom. Nauk. Sbornik 3, p.p. 3–30, Moscow ONTI). Professor Bronshtein was one of the most brilliant theorists. During Stalin purges he was arrested and executed in 1938 when he was 32 years old.

Place G, $\hbar$ and $1/c$ on three orthogonal axes (Fig. 1).

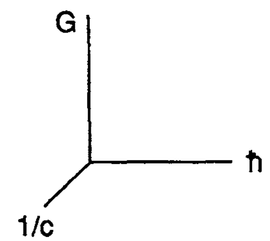


Fig. 1

Now we can construct the cube of theories (Fig. 2).

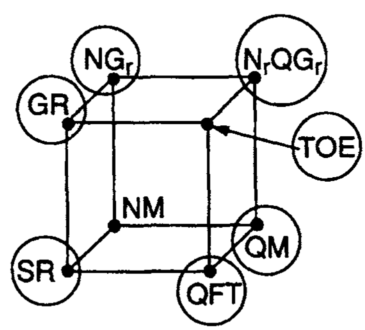


Fig. 2

At the origin (where $G = 0, \hbar = 0, 1/c = 0$), we have Newtonian Mechanics (without gravity, without relativity, without quantum effects). If we take into account G we have Newtonian Gravity. If we take into account finiteness of c, we have Special Relativity. If we take into account $\hbar$, we have Quantum Mechanics.

When we combine NG and SR we obtain General Relativity. When we combine SR and QM we obtain Quantum Field Theory. NG and QM give us Nonrelativistic Quantum Gravity. (This is a science which we don't need at present, because we have not observed objects which are described by this theory. But these objects may exist. These are particles, say, ten times lighter than Planck mass. Gravitationally attracting each other they would form "atoms". If these particles are neutral, it is very difficult to discover them.)

If we combine QFT and GR we obtain Theory of Everything. It is impossible to construct the Theory of Everything staying in one of the planes of the cube of theories. This was stated even earlier, in 1928, by George Gamow, Dimitri Ivanenko and Lev Landau (ZhRFCHO (Journal of Russian Physico-Chemical Society) 60,13(1928)). Let me remark that it was the great Einstein who for many years tried to build a unified theory in Gc-plane, without taking into account $\hbar$.

– *Zichichi:*

We will have a Bronshtein scholarship in Erice.

– *Duff:*

Now that we have set up the straw man, may we knock him down? Namely, now that we all agree that the Theory of Everything must have at least three fundamental parameters, where does that leave the string theory that has less than three?

– *Okun:*

This is a very good question which, I suppose, you will answer in your lecture.

– *Ellis:*

I don't agree with what Lev and Mike have said. I don't think you would need in the Theory of Everything three parameters. I think there will be no parameters at all. I have already told that to Mike at the coffee break this morning.

– *Zichichi:*

A Theory of Everything starting from nothing: fantastic!

– *Okun:*

I fully agree with John. There will be no parameters. G, $\hbar$, c are in fact not what we usually call parameters, they are units in which we will calculate everything. These units will drop out of physics. Physics depends only on dimensionless parameters. The Theory of Everything will give us relations between different dimensionless quantities like α or m_e/m_p, so there will be no free parameters.

– *Ellis:*

Exactly. All we measure is numbers.

- *Okun:*

Yes, all we measure is numbers. Some of them are obviously dimensionless like electron-proton mass ratio. But others like $\alpha = e^2/\hbar c$ involve fundamental units. You have the maximal velocity of signals c, you have the quantum of action $\hbar$ and you have the Planck mass m_P or what you prefer instead. You cannot get rid of them. In a natural theory you measure everything in these natural units. Therefore you must know that such three fundamental units exist.

- *Ellis:*

Yes, historically somehow we grew up with units of mass, length and time, which have no fundamental significance. We re-express them through $G,\ \hbar,\ c$.

- *Okun:*

The values of the three fundamental units are of historical origin. When we say that $c = 3 \cdot 10^{10} cm/s$, it depends on our definitions of centimetre and of second. The only fact that we really use is that our unit of velocity is the maximal velocity of signals.

The existence of $G,\ \hbar,\ c$ is a fundamental property of Nature.

- *Ellis:*

The statement that there is an ultimate velocity is just a statement that Minkowski space has an indefinite metric. It is just group theory.

- *Okun:*

There would be group theory even in the case of infinite velocity — Galilean group.

- *Ellis:*

That would be a different group.

- *Okun:*

Of course, you may speak about Minkowski space, but it is simpler to say that velocity is limited.

- *Levi:*

Is gravitational field the only field that deforms space-time? If yes, why is gravitation so special in comparison to other fields?

- *Okun:*

The space is deformed only by gravitation. This is connected with the fact that the graviton is the only fundamental particle with spin equal to two.

– *Ungkitchanukit:*

Galilei discovered that the motion in gravitational field does not depend on the mass (it depends only on g). But in quantum mechanics mass appears in the phase of wave function. So it seems that in classical mechanics mass drops out while in quantum mechanics it does not.

– *Okun:*

Of course, quantum mechanical phase depends on energy and for massive particles, on mass. But there are lots of effects in classical mechanics as well in which energy and mass are essential. The acceleration of free fall is not the only measurable quantity of classical mechanics.

– *Shabelski:*

Is it possible for different free electrons to have slightly different mass values?

– *Okun:*

This is a very interesting question. There exist about 10^{80} electrons in our Universe. All of them are identical. But maybe some of them are more identical than others? Is it possible that there is a distribution of their masses which is characterized by a quantity δm?

Such a question was asked by Enrico Fermi in 1933. He noticed that a non-vanishing δm would spoil the property of atoms during the long time T of their existence. The non-observation of such effects means that $\delta m \ll 1/T$.

– *Pluciennik:*

What are your sources on Galileo Galilei in addition to the references given in the lecture?

– *Okun:*

– A.Koyré. Etudes Galiléennes. Paris 1939.
– M.Fierz. Vorlesungen zur Entwicklungsgeschichte der Mechanik (Lecture Notes in Physics, Vol. 15) Springer-Verlag. Berlin, 1972.
– Galilée: L'Expérience sensible. Textes de P.Caluzzi, G.Micheli, A.Porta, L.Rosino, G.Taboreli. Editions VILO. Paris, 1990 (French translation of Italian edition: Amilcare Pizzi S.P.A. 1988)
– Antologia Galileana, con traduzioni e note bibliografiche a cura di Carlo Maccagni. Firenze. Barbera Editore. 1964.

6

VACUA, VACUUM: THE PHYSICS OF NOTHING

L. B. Okun

Theoretical Physics Division, CERN
CH - 1211 Geneva 23
Switzerland
and ITEP, Moscow 117259, Russia*

ABSTRACT

The origin and the early history of several ideas concerning possible properties of the vacuum are briefly sketched. The topics include the CP properties of the vacuum, the possible existence of a mirror world, phase transitions in the early universe, vacuum domain walls, and the decay of the false vacuum.

1. INTRODUCTION

To think is difficult. This is my personal experience. To think about nothing is more difficult than about something. But this is not the only reason why there exists an old and fruitful tradition of attributing complex properties to the vacuum. In this century alone we had such outstanding examples as the ether, the Dirac sea and the non-vanishing vacuum expectation values of fields.

Sometimes, when using the word "vacuum," physicists do not really mean the vacuum itself, but certain virtual states of particles. For instance, the polarization of vacuum by a photon is not different from any other one-loop (or many-loop) Feynman diagram with two (or more than two) external lines. We are dealing here with particles, not with vacuum. Graphs with only one external line (tadpoles), and especially without any external lines at all, are really the ones that describe the evolution of the vacuum. (An important role of the isolated vacuum graphs without external lines has been discussed by Feynman [1].)

2. POMERANCHUK ON VACUUM

Many years ago, being a student of Isaak Pomeranchuk, I first heard from him the sentence: "The vacuum is filled with the most profound physical content!" Later I learned

*Permanent Address.

that young Pomeranchuk came early in the morning to one of his friends, an experimentalist, and told him: "The most important physical object is the vacuum. You have to stop all other activities, to buy vacuum pumps and to explore the vacuum." This was a humorous way to stress the utmost importance of the subject and to express the frustration of being unable to suggest an experiment to study it. Cosmology, at that time, was not so intimately connected with particle physics.

3. LANDAU ON VACUUM

In November 1956, Boris Ioffe, Alexei Rudik and myself, studying the preprint of the famous paper by T. D. Lee and C. N. Yang [2] on parity non-conservation, and the famous article by W. Pauli on the CPT theorem [3], came to the conclusion that the spin–momentum correlations discussed in that paper were possible only if charge conjugation symmetry C was violated, together with parity P [4]. The two neutral kaons, the short-lived and long-lived one, are in this case respectively even and odd under time reversal T (or, better, under CP), but not under C, as had been believed until then [5]. (T. D. Lee, R. Oehme and C. N. Yang [6] came independently to similar conclusions, see references [7] and [8].)

A few weeks later, Lev Landau, with whom we had discussed our paper, became absolutely convinced overnight that the vacuum could not have a CP-odd admixture, and put forward the law of exact CP invariance [9]. The idea looked very attractive and was widely accepted. I considered it to be beautiful, but, on the other hand, I did not understand why the Lagrangian could not have complex coefficients. Therefore in lectures given at ITEP (1959) and Dubna (1961), I stressed the importance of experimental tests of CP invariance, and in the book based on these lectures the search for the decay of the long-lived kaon into two pions is mentioned as one of the most decisive experiments [10]. The upper limit on the branching ratio of this decay attained in Dubna in 1962 [11] was at approximately the same level at which this process was discovered in 1964 by J. Christenson, J. Cronin, V. Fitch and R. Turlay [12].

4. "MIRROR WORLD"

Still, the appeal of Landau's idea was so strong that in 1965 Igor Kobzarev, Isaak Pomeranchuk and myself suggested the hypothesis of a "mirror world" [13]. We assumed "CPA invariance," where the A — which stands for "Alice Through the Looking Glass" — transforms our part of the Lagrangian into its mirror part. In this way the "sin" of CP violation was committed by our particles, and not by the vacuum. Each of our particles has its mirror counterpart. The mirror particles have, between them, the same interactions as ours. In principle there may exist mirror nuclei, atoms, molecules, stars, planets, galaxies, even mirror life. Whether they actually exist depends on cosmological evolution. The simplest assumption is that the only interaction between ours and mirror matter is gravitational.

I vividly remember how stunned I was when skiing in a forest not far from Moscow I had a vision of a mirror train crossing the clearing in front of me. It was of course clear to me that this mirror train had to move on the surface of a mirror planet, the existence of which is excluded by astronomical and gravimetric observations. But the image of the train was very graphic.

A number of papers have been written dealing with mirror particles as a substantial component of dark matter in the universe. In 1983 I came back to the idea, and considered a stronger coupling between two worlds, which could be due to the exchange of some new neutral particle, and suggested a number of experimental tests [14].

Note that the modern usage of the term "mirror symmetry" refers, in the context of superstrings, to a duality between large and short distances and has nothing to do with "mirror particles."

5. SAKHAROV AND BAU

In 1967 I had the privilege to witness the creation of Sakharov's seminal paper [15] on the Baryon Asymmetry of the Universe (BAU). He started out from the CP-violating charge asymmetry of the semileptonic decays of long-lived neutral kaon and non-leptonic decays of sigma and antisigma hyperons. Here a paper by S. Okubo was very important [16]. Using it as a springboard, he jumped from strange particles to the universe and from strangeness violation to baryon-number violation, which he postulated.

With the advent of the Standard Model based on the broken SU(3)×SU(2)×U(1) gauge symmetry, and of grand unification theories (GUTs), baryon number violation appeared as a byproduct. In GUTs it occurs at the tree level. In electroweak theory it comes from a non-perturbative anomaly discovered by G. 't Hooft [17]. At ordinary temperatures, the anomalous term that triggers baryon number violation is negligibly small, but it becomes very efficient at the temperature of the electroweak phase transition, $T \simeq 1$ TeV [18]. This phase transition may wipe out the BAU left from the GUT era and create it from scratch.

Here I have to return to the early 1970's, when the idea of the vacuum phase transitions was formulated.

6. KIRZHNITZ ON VACUUM

It is well known that the vacuum expectation value (VEV) of the Higgs field plays a crucial role in the reduction of SU(2)×U(1) to U(1). An important step was made in 1972 by David Kirzhnitz — a theorist at the Lebedev Institute, who realized that the VEV of a Higgs is a function of temperature and has to vanish in the early universe when we go backward in time to temperatures much higher than the present value of the VEV [19]. This idea was further developed by Kirzhnitz in collaboration with his student Andrei Linde [20] and later by Linde [21] and by many others. I first heard about the phase transition of the vacuum from David Kirzhnitz on a street during a conference in Tashkent in 1972.

7. VACUUM DOMAIN WALLS

Two years later Igor Kobzarev, Yakov Zeldovich and myself [22] merged the idea of the vacuum phase transition with a model of spontaneous CP violation (again CP!) proposed by T. D. Lee [23]. According to this model a neutral pseudoscalar CP-odd field has two degenerate vacua which differ only by sign and transform into each other under CP reflection. If we assume the validity of this model, then the vacuum around us has a sign that was chosen by chance during the cooling of the universe and the formation of VEVs. But at any distant enough place the other sign may have been chosen. The domains of different signs are separated from each other by "domain walls," the thickness and density of which are determined by VEVs of the pseudoscalar field and its self-coupling, with the characteristic scale of, say, several hundred GeV. We discussed in [18] the cosmological evolution of a universe filled with domain walls and concluded that at present the nearest wall should have gone beyond the horizon, leaving as a farewell an anisotropy in the black-body radiation. The advent of inflational cosmology greatly weakened this argument against spontaneous CP violation.

The vacuum domain wall was the first megascopic elementary object considered in the framework of quantum field theory. Soon cosmic strings [24, 25, 26] and magnetic monopoles [27, 28] were considered. These objects emerge as a result of spontaneous breaking of gauge symmetries (U(1) for the former and SU(2) for the latter). An exotic example of a cosmic string was invented by Albert Schwarz [29, 30] after I told him about the "mirror world." A particle, after making a circle around his "Alice string," transforms into a mirror particle and becomes invisible. If an "Alice string" passes between the Earth and the galaxy, the galaxy becomes a mirror one from the point of view of a terrestrial observer, so it becomes invisible. Conversely, as a result of the passage of a string, a mirror galaxy becomes visible.

8. FALSE VACUUM

Up to now I have spoken about the breaking of symmetry between degenerate vacua. The next subject is the transition between two vacua, one of which lies a little bit higher than the other. In this case the upper vacuum is metastable. Such a model was proposed by T. D. Lee and G. C. Wick [31] and elaborated by Igor Kobzarev, Mikhail Voloshin and myself [32]. The decay of the metastable vacuum (later C. Callan and S. Coleman dubbed it "false vacuum" [33, 34, 35]) starts by the formation, through quantum tunneling, of the smallest bubble of the lower vacuum surrounded by a wall which separates the two vacua. The size of the smallest bubble is such that the gain of energy proportional to its volume becomes large enough to compensate the mass of the wall which is proportional to the surface. After that, the bubble expands classically with velocity close to that of the light, destroying the universe.

When I first thought that the creation of a bubble could be catalyzed at collider, my back shivered. Then I reassured myself: all possible collisions have already occurred in the early universe. A few months later I told Andrei Sakharov about the bubble. His reaction was: "Such theoretical work should be forbidden!" My argument about collisions in the early universe was rejected by him: "Nobody has collided two nuclei of lead." I still believe that, although we live in an unstable world, our vacuum is stable. Vacuum bubbles became a toy for cosmologists.

9. THE SUBJECTS I HAVE TO OMIT

Lack of time (and partly of knowledge) prevents me from discussing such important topics as

- the structure of the QCD vacuum with its quark and gluon condensates, which play a crucial role in the mechanism of confinement and in the ITEP Sum Rules;
- the structure of the vacuum in grand unified theories and their supersymmetric variants, which determines the hierarchy of scales;
- the gravitational vacuum, and especially the cosmological constant;

- the analogy between the complex structure of the vacuum and topological defects in condensed matter physics.

To discuss all this, there would have to be a special conference, fully devoted to nothing.

Dedication

I dedicate this lecture to the memory of those who are dear to my heart and with whom I had the privilege and fun to coauthor some of the papers briefly described in this lecture:

Igor Yuryevich Kobzarev (1932–1991),
Isaak Yakovlevich Pomeranchuk (1913–1966),
Alexey Petrovich Rudik (1922–1993),
Yakov Borisovich Zeldovich (1914–1987).

REFERENCES

1. R. Feynman, *The Reason for Antiparticles* in *Elementary Particles and the Laws of Physics; The 1986 Dirac Memorial Lectures* (Cambridge Univ. Press, 1987).
2. T. D. Lee and C. N. Yang, *Phys. Rev.* **104** (1956) 254.
3. W. Pauli, in *Niels Bohr and the Development of Physics,* edited by W. Pauli, London, Pergamon Press, 1955.
4. B. L. Ioffe, L. B. Okun and A. P. Rudik, *Zh. ETF* **32** (1957) 396 [*Sov. Phys. JETP* **5** (1957) 328].
5. M. Gell-Mann and A. Pais, *Phys. Rev.* **97** (1955) 1387.
6. T. D. Lee, R. Oehme and C. N. Yang, *Phys. Rev.* **106** (1957) 340.
7. C. N. Yang, in *Nobel lectures. Physics 1942 - 1962* (Elsevier, 1964), p. 393.
8. T. D. Lee, in *Nobel lectures. Physics 1942 - 1962* (Elsevier, 1964), p. 406.
9. L. D. Landau, *ZhETF* **32** (1957) 405 [*Sov. Phys. JETP* **5** (1957) 336]; *Nucl. Phys.* **3** (1957) 127.
10. L. B. Okun, *Weak Interactions of Elementary Particles* (Pergamon, 1965) (a translation of *Slaboe vzaimodeystvie elementarnych chastiz* (Fizmatgiz, Moskva, 1963).
11. M. Anikina et al., *ZhETF* **42** (1962) 130, [*Sov. Phys. JETP* **15** (1962) 93].
12. J. H. Christenson, J. W. Cronin, V. C. Fitch and R. Turlay, *Phys. Rev. Lett.* **13** (1964) 138.
13. I. Yu. Kobzarev, L. B. Okun and I. Ya. Pomeranchuk, *Yad. Fiz.* **3** (1966) 1154 [*Sov. J. Nucl Phys.* **3** (1966), 837].
14. L. B. Okun, *On Search for mirror particles.* Preprint ITEP-149 (1983).
15. A. D. Sakharov, *ZhETF Pis'ma* **5** (1967) 32, [*JETP Lett.* **5** (1967) 24].
16. S. Okubo, *Phys. Rev.* **109** (1958) 984.
17. G. 't Hooft, *Phys. Rev. Lett.* **37** (1976) 8.
18. V. A. Kuzmin, V. A. Rubakov and M. E. Shaposhnikov, *Phys. Lett.* **B155** (1985) 36.
19. D. A. Kirzhnitz, *Pisma Zh. ETF* **15** (1972) 745, [*JETP Lett.* **15** (1972) 529].
20. D. A. Kirzhnitz and A. D. Linde, *Phys. Lett.* **42B** (1972) 471.
21. A. D. Linde, *Particle Physics and Inflationary Cosmology (1990), Fizika Elementarnykh Chastits Kosmologiya.* (Nauka, Moscow, 1990).
22. I. Yu. Kobzarev, L. B. Okun and Ya. B. Zeldovich, *Zh. ETF* **67** (1974) 3 [*JETP* **40** (1975) 1]; *Phys. Lett.* **B50** (1974) 340.
23. T. D. Lee, *Phys. Reports* **9C** (1974) No. 2.
24. T. W. B. Kibble, *J. Phys.* **A9** (1976) 1387.
25. Ya. B. Zeldovich, *Mon. Not. Roy. Astron. Soc.* **192** (1980) 663.
26. A. Vilenkin, *Phys. Lett.* **46** (1981) 1169.
27. G. 't Hooft, *Nucl. Phys.* **B79** (1974) 276.
28. A. M. Polyakov, *ZhETF Pis'ma* **20** (1974) 276, [*JETP Lett.* **20** (1974) 194].
29. A. S. Schwarz, *Nucl. Phys.* **B208** (1982) 141.

30. A. S. Schwarz and Yu. S. Tyupkin, *Nucl. Phys.* **209** (1982) 427.
31. T. D. Lee and G. C. Wick, *Phys. Rev.* **D9** (1974) 2291.
32. M. B. Voloshin, I. Yu. Kobzarev and L. B. Okun, *Yad. Fiz.* **20** (1974) 1229; [*Sov. J. Nucl. Phys.* **20** (1975) 644].
33. S. Coleman, *Phys. Rev.* **D15** (1977) 2929; Err. **D16** (1977) 1248.
34. C. Callan, S. Coleman, *Phys. Rev.* **D16** (1977) 1762.
35. P. H. Frampton, *Phys. Rev.* **D15** (1977) 2922.

Eur. J. Phys. **15** (1998) 403–406. Printed in the UK PII: S0143-0807(98)93440-7

LETTERS AND COMMENTS

Note on the meaning and terminology of Special Relativity

L B Okun†
Institute for Theoretical and Experimental Physics, 117218 Moscow, Russia

Received 20 April 1998

Abstract. Special Relativity is a classical and simple theory. In spite of this, the pedagogical literature on it is still not free from confusion. This confusion is caused by the frequent usage of archaic notions, notations and equations, which have nothing to do with the essence of the theory, only with the history of its development. The aim of this short comment is to explain the basic notions of Special Relativity and to help the beginner to recognize and avoid the confusing notations and equations.

1. The relativity of Galileo Galilei

According to the principle of relativity, the relative motion of inertial reference frames cannot be detected by any experiments within these frames. The essence of relativity is expressed by the 'gedanken ship' described by Galileo Galilei in his famous book the *Dialogo*, published in Florence in 1632 [1]:

'Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal: jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that, you will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. In jumping, you will pass on the floor the same spaces as before, nor will you make larger jumps toward the stern than toward the prow even though the ship is moving quite rapidly, despite the fact that during the time you are in the air the floor under you will be going in a direction opposite to your jump. In throwing something to your companion, you will need no more force to get it to him whether he is in the direction of the bow or the stern, with yourself situated opposite. The droplets will fall as before into the vessel beneath without dropping toward the stern, although while the drops are in the air the ship runs many spans. The fish in their water will swim toward the front of their bowl with no more effort than toward the back, and will go with equal ease to bait placed anywhere around the edges of the bowl. Finally the butterflies and flies will continue their flights indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired

† E-mail: okun@heron.itep.ru

out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air. And if smoke is made by burning some incense, it will be seen going up in the form of a little cloud, remaining still and moving no more toward one side than the other. The cause of all these correspondences of effects is the fact that the ship's motion is common to all the things contained in it, and to the air also. That is why I said you should be below decks; for if this took place above in the open air, which would not follow the course of the ship, more or less noticeable differences would be seen in some of the effects noted.'

All laws of nature and hence equations of motion and observable effects look the same in reference frames moving rectilinearly with constant velocity relative to the stars. In order to preserve the form of the Newtonian equations the time and space coordinates of two inertial frames should obey the relations

$$t' = t \tag{1a}$$

$$x' = x + vt \tag{1b}$$

$$y' = y \qquad z' = z \tag{1c}$$

where v is the velocity of the ship, which moves along the x-axis, primed coordinates refer to the frame of the shore, and unprimed to that of the ship.

2. The relativity of Albert Einstein

The discovery of Maxwell's equations and the necessity of combining them with the equations of mechanics has led to modification of Newton's equations and to the Lorentz transformations:

$$t' = \gamma\left(t + \frac{vx}{c^2}\right) \tag{2a}$$

$$x' = \gamma(x + vt) \tag{2b}$$

$$y' = y \qquad z' = z \tag{2c}$$

where

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} \tag{3}$$

and c is the maximum velocity a particle (body) may have in a vacuum, i.e. the velocity of light. Michelson's experiment showed that relative motion of inertial frames cannot be detected even if one uses an optical interferometer. The interference pattern of two perpendicular rays of light did not depend on the orientation of the interferometer.

One may say that Einstein brought Michelson's interferometer onto Galileo's ship. In the course of the 20th century it was proved that not only mechanical, optical and electromagnetic phenomena, but also all atomic and nuclear interactions are invariant under the Lorentz transformations (2). At $v/c \ll 1$, $\gamma \to 1$ and equations (2) coincide with equations (1). According to equations (2) not only the velocities of bodies on the ship, but also the velocity of the ship itself may be of the order of the velocity of light.

The new feature of equation (2) was that time was no longer universal: it became relative, i.e. dependent on the relative velocity of the reference frames. The word relativity acquired several new meanings. In particular particles with velocities close to c are called relativistic. According to relativity theory the momentum $\boldsymbol{p}(p_x, p_y, p_z)$ of a particle and E/c^2, where E is its energy, transform similarly to $\boldsymbol{r}(x, y, z)$ and t:

$$E' = \gamma(E + vp_x) \tag{4a}$$

$$p'_x = \gamma\left(p_x + \frac{vE}{c^2}\right) \tag{4b}$$

$$p'_y = p_y \qquad p'_z = p_z. \tag{4c}$$

It is easy to see from equations (2) that the quantity

$$\tau^2 = t^2 - \frac{\boldsymbol{r}^2}{c^2} \tag{5}$$

is invariant under Lorentz transformation:

$$\tau'^2 = \tau^2. \tag{6}$$

The physical meaning of τ is that of time in the rest frame of the particle, the proper time of the particle. Another important invariant is the mass m of the particle, its 'proper mass':

$$m^2 = \left(\frac{E}{c^2}\right)^2 - \frac{\boldsymbol{p}^2}{c^2}. \tag{7}$$

A great discovery, the one that Einstein made in 1905 [2], was that a massive body at rest still had energy, which he denoted by

E_0 and called rest energy. As follows from equation (7)

$$E_0 = mc^2. \tag{8}$$

The rest energy plays a unique role in all types of energy transformations: from nuclear to chemical reactions, as explained in many books and articles. If we apply equations (4) to a particle in its rest frame, in the laboratory frame (where the particle is seen to move with velocity v) we obtain

$$E = mc^2\gamma = \gamma E_0 \tag{9a}$$

$$\boldsymbol{p} = m\gamma\boldsymbol{v}. \tag{9b}$$

Here we may forget about the ship and consider v as the velocity of the particle. Equation (9*a*) may be considered as another definition of γ:

$$\gamma = \frac{E}{mc^2}. \tag{10}$$

Note that for $v \ll c$ we obtain from (9)

$$T = E - mc^2 \simeq \tfrac{1}{2}mv^2 \tag{11a}$$

$$\boldsymbol{p} \simeq m\boldsymbol{v} \tag{11b}$$

where T is the non-relativistic kinetic energy, while $\boldsymbol{p}$ is the non-relativistic momentum. Equations (11) stress that in all equations m is the usual Newtonian mass. A consistent presentation of Special Relativity can be found in [3–5].

3. Archaic notions and notations

Many years ago, when Special Relativity was in the process of development and experimental verification, many notions and kinds of notation were introduced which were later abandoned in favour of those which are described in the first two sections of this comment. Unfortunately some of these outdated notions are still being promoted by authors of many textbooks, popular books and articles.

One of the main sources of confusion is the use of the non-relativistic relation between momentum and velocity in place of the relativistic one (equation (11*b*) instead of equation (9*b*)).

As a result m in (11*b*) plays the role of $m\gamma$ in (9*b*). This leads to the false conclusion that the mass of a body increases with its velocity. This 'mass' (which is actually E/c^2) is called relativistic mass. This leads immediately to the 'most famous formula'

$$E = mc^2 \tag{12}$$

which according to many authors is the central formula of Special Relativity. Many famous physicists, when writing popular texts, used this formula (probably because it is so famous and hence self-promoting) instead of the correct formula (8*c*): $E_0 = mc^2$. (Einstein himself alternated between (12) and (8*b*), although he preferred (8*b*). Sometimes he used the relation $E = mc^2$ in the following sense: if E is the energy of radiation emitted or absorbed by a massive body at rest, then m is the change (negative, or positive, respectively) of its mass.)

If one uses m to denote E/c^2 one has to introduce the so called 'rest mass' m_0 to denote the Newtonian mass:

$$m = m_0\gamma. \tag{13}$$

For experts in relativistic physics the misuse of (11*b*) and use of equations (12) and (13) do not present any difficulty in understanding texts which contain these pseudo-relativistic equations. However, in the minds of beginners and outsiders they create confusion, and misunderstanding, and produce wrong intuition.

A decade ago an effort was made [5, 6] to explain the origin and the danger of archaic terminology. As a result recent editions of many university and college textbooks have dropped it, and use only rational modern terminology. However, some are still promoting confusion [7]. Some papers published recently by *European Journal of Physics* are also based on confusing terminology. The Editorial Board of the journal decided to publish this Comment in order to rectify this deplorable situation.

References

[1] Galilei G 1632 *Dialogue Concerning the Chief World Systems—Ptolemaic and Copernican* Engl. transl. S Drake, foreword A Einstein, 1967 (Berkeley, CA: University of California Press) 2nd edn

[2] Einstein A 1905 *Ann. Phys., Lpz* **18** 639

[3] Einstein A 1921 *The Meaning of Relativity: Four Lectures Delivered at Princeton University, May 1921* Engl. transl. E P Adams 1970 (Princeton, NJ: Princeton University Press) 5th edn

[4] Landau L D and , Lifshitz E M 1951 *The Classical Theory of Fields* (Oxford: Pergamon) transl. from 2nd Russian edn

[5] Taylor E F and Wheeler J A 1992 *Spacetime Physics. Introduction to Special Relativity* 2nd edn (New York: Freeman) (see especially pp 246–52: 'Dialog: use and abuse of the concept of mass')

[6] Okun L B 1989 The concept of mass *Phys. Today* June pp 31–6

For discussion of this paper see: Putting to rest mass misconceptions *Phys. Today* May 1990 pp 13, 15, 115–7

Okun L B 1989 The concept of mass (mass, energy, relativity) *Sov. Phys.–Usp.* **32** 629–38

[7] Tipler P A 1996 *Physics for Scientists and Engineers* 3rd edn, extended version (Worth) pp 1125–32

On Relativistic Apple: Exchange of Letters with M.R. Kleemans in August 1998

Letter No. 1

Date: 6-AUG-1998 15:29:09.88
From: "M.R.Kleemans@phys.uu.nl" "Machiel Kleemans"
To: lev.okun@cern.ch
CC: M.R.Kleemans@phys.uu.nl (Machiel Kleemans)
Subj: gravitational force

Dear Prof. Okun,

I am an undergraduate student of theoretical physics at Utrecht University, the Netherlands. For some time I have been trying to derive the equation for the general relativistic force working on a light relativistic body in the field of the earth or the sun. This is eq. (16) in your article in Phys. Today, June 1989 and it also appears in Phys. Today, may 1990. Although it seems a simple problem I do not get the right expression.

I have been trying to derive this equation from Weinberg (Grav. & Cosmology), from an equation for the acceleration of the particle as it appears on page 221, between eqs. 9.2.4 and 9.2.5, but have failed so far. Does there exist an explicit derivation somewhere (I have not been able to find one, the book by Serway and Faughn as referred to in the may 1990 issue, does not seem to exist in the Netherlands)? Otherwise maybe you could tell me explicitly how to obtain the right expression. If you could help me, I would be much obliged and would like to thank you in advance,

Sincerely,

M.R. Kleemans.

Letter No. 2

Date: Tue, 18 Aug 1998 14:47:59 +0400 (GMT)
From: "L. B. Okun" <okun@heron.itep.ru>
To: Machiel Kleemans <M.R.Kleemans@phys.uu.nl>
Subject: gravity

Dear Mr. Kleemans,

The book by Serway and Faughn , which you looked for, is cited in my reply to Catherine Sauter's letter and obviously has nothing to do with eq.(16). The book, in which this eq. is written, is that by Bowler. The two references were interchanged in Phys. Today. As I wrote in 1990, I am not satisfied with the derivation given in the book by Bowler; therefore I would not recommend it to you.

The correct derivation is sketched on page 117. By using eq. (87.3) from LL (Landau and Lifshitz), which is identical with the eq. on page 221 of SW (Steven Weinberg), and substituting in it the Christoffel symbols (9.1.67) and (9.1.70) from SW , you derived , I guess, an equation, which differs from eq. (16) by an extra factor of 2 in front of the second term. Now if you change to local rulers and clocks and use isotropic metric (see problem 4 in section 100 of LL), you get for the local momentum an extra factor (1 + phi), which after differentiation reduces 2 to 1 and gives you eq. (16).

When will you graduate? Who is your Professor on Gravity? Who is your adviser?

With best wishes,

Lev Okun

Letter No. 3

Date: Fri, 28 Aug 1998 15:28:09 +0200 (MET DST)
From: Machiel Kleemans <M.R.Kleemans@phys.uu.nl>
To: L. B. Okun <okun@heron.itep.ru>
Cc: Machiel Kleemans <M.R.Kleemans@phys.uu.nl>
Subject: Re: gravity

Dear Prof. Okun,

Thank you very much for your most kind reply to my letter on gravitation. I was away for a few days and read your letter only very recently, hence my delayed response. I found the derivation in the book by Bowler (1976, Pergamon Press) on pages 63/64 (not on page 117), which, if this is the one you mean, I agree, is not a very good derivation.

Calculating the force from the local momentum with an extra factor of (1+phi) indeed gives me the right expression. I am not sure, however, that I completely understand how this factor comes about, but that is a small matter which I think I can find out. I did have a factor of two wrong, as you guessed correctly, and I'm glad to see the proper result coming out now.

I hope to be graduating within a few months. While having finished all my exams, I am presently writing my graduation thesis, my supervisor is Prof. B. de Wit (Utrecht). The subject of my thesis will be interactions of BPS-states, coupled to different fields (tensor, vector, scalar), such as they appear in Kaluza-Klein theory. In an introductory chapter on general relativity, I wanted to mention the equation for the general relativistic force and found out that I could not properly derive it. This was the reason for mailing you and I am grateful for your response to my problem.

Sincerely yours,

Machiel Kleemans.

Letter No. 4

Date: Mon, 31 Aug 1998 12:31:36 +0200 (MET DST)
From: Machiel Kleemans <M.R.Kleemans@phys.uu.nl>
To: L. B. Okun <okun@heron.itep.ru>
Cc: Machiel Kleemans <M.R.Kleemans@phys.uu.nl>
Subject: Re: gravity

> Dear Mr. Kleemans,
> Thank you for fast reply.
> The page 117 is not from Bowler, but from Phys.Today ,1990.
> I still do not know who was your lecturer on gravity.
> Best wishes,
> Lev Okun
>

I had in fact two lecturers on gravity, one of them was Prof. 't Hooft, who both gave a similar series of lectures on the introduction to general relativity. This started my interest for the subject.
I am, by the way, familiar with the 1990 issue of Phys. Today, where the proof is sketched. At the time of reading it, it wasn't clear to me what exactly was meant with local coordinates in this case. I thought it just meant the transformation of proper time to time in a laboratory frame, as in special relativity. Thank you again for helping me with this issue.

Sincerely yours,

M.R. Kleemans.

This page intentionally left blank

Physics – Uspekhi **41** (6) 553 – 557 (1998)

ON THE 90TH ANNIVERSARY OF L D LANDAU'S BIRTH PACS numbers: **10, 11, 13.90.+i, 14**

Current status of elementary particle physics

L B Okun

Contents

Abstract. A brief review is given of the state-of-the-art in elementary particle physics based on the talk of the same title given on January 22, 1998, at the seminar marking the 90th anniversary of the birth of L D Landau. (The seminar was hosted by the P L Kapitza Institute for Physical Problems in cooperation with the L D Landau Institute of Theoretical Physics).

1. Sixteen fundamental particles

Elementary (or fundamental) particles are defined as particles that are not composed, at the current level of knowledge, of other elementary particles.

Experiments have revealed twelve elementary fermions (with spin $s = 1/2$) and four bosons (with spin $s = 1$) (not counting the respective antiparticles).

2. Fundamental fermions

Table 1 shows that the twelve fermions form three generations, each comprising two leptons and two quarks.

The last column of Table 1 gives the electric charge of the particle in the given row. The names and notation of quarks stand for the English words: u for up, d for down, c for charmed, s for strange, t for top (or truth), b for bottom (or beauty).

Table 1.

Generations		1st	2nd	3rd	Q
Quarks	upper	u	c	t	+2/3
	lower	d	s	b	−1/3
Leptons	neutrino	ν_e	ν_μ	ν_τ	0
	charged	e	μ	τ	−1

L B Okun State Scientific Center of Russian Federation "Institute of Theoretical and Experimental Physics"
ul. B. Cheremushkinskaya 25, 117218 Moscow, Russia
E-mail: okun@heron.itep.ru

Received 26 March 1998
Uspekhi Fizicheskikh Nauk **168** (6) 625 – 629 (1998)
Translated by V Kisin; edited by L V Semenova

Each charged fermion has an antiparticle counterpart: $\bar{u}$, $\bar{d}$, $\bar{c}$, $\bar{s}$, $\bar{t}$, $\bar{b}$, e^+, μ^+, τ^+. It is not yet known if the neutrinos have antiparticles. They (at least some of them) may even be truly neutral, that is, they constitute their own antiparticles. These truly neutral neutrinos are known as Majorana neutrinos, as against the ordinary ones, known as Dirac neutrinos. A search for neutrinoless double beta-decay is being actively pursued now. The highest accuracy was achieved in the search for the neutrinoless decay of ^{76}Ge ($T_{1/2} \gtrsim 10^{25}$ years). If neutrinoless double beta-decay is discovered at an improved level of instrument sensitivity, it will indicate the Majorana nature of the neutrino and also that its mass is non-zero.

Looking at Table 1, we cannot help marveling at how far elementary particle physics has gone since fate snatched L D Landau first from physics (January 1962) and then from life (1 April 1968). Landau knew only three leptons (e, ν_e and μ). The discovery of the muon neutrino was first announced in July 1962 at the Rochester conference in Geneva; all the participants of the conference sent Landau a letter, wishing him the speediest recovery.

As for the hadrons, in 1962 physicists knew, in addition to nucleons, π and K mesons, plus a number of strange hyperons and resonances. However, the composite hadron model based on three sakatons failed to provide an adequate description of baryons, even though it gave an adequate description of weak hadron currents and mesons. That only came in 1964 when M Gell-Mann introduced three quarks: u, d and s.

It would of course be possible to say that both the world around us and ourselves are built of nucleons and electrons; the fact that nucleons consist of three quarks (p = uud, n = ddu) cannot change anything in the 'everyday' life of a human being. For our world to function, it also needs, in addition to p, n and e, the electron neutrinos ν_e, without which hydrogen could not burn in the Sun and the stars:

$$2e + 4p \to {}^4\mathrm{He} + 2\nu_e + 27\ \mathrm{MeV}\,.$$

This is the source of all the energy we can count on. Neutrinos are born in the Sun in enormous numbers and immediately fly away into the surrounding space. Seventy billion neutrinos cross each square centimeter of the surface of the Earth each second. We do not notice this, though: we are transparent to neutrinos. Huge detector systems had to be erected to record several hundred solar neutrinos. This number was roughly what was expected, which was a triumphant success of science. But the triumph was not complete: the number of neutrinos was less than that calculated by a factor of 2 – 3. We will return to this puzzle later. At the moment we can emphasize that out of the three generations of fermions in Table 1, a 'pedestrian' would be quite happy if the first was the only one. We can only guess what the second and third generations are needed for. According to some very simple models, for example, CP invariance cannot be violated without them. And without this violation of CP invariance the Universe could not have generated the baryonic asymmetry at the early stages of its evolution: it would contain equal numbers of protons and antiprotons, electrons and positrons; they would all be converted by annihilation into photons and neutrinos and we would never have appeared!

3. Fundamental vector bosons

The tradition of science-popularizing literature is to treat fundamental fermions as 'building blocks' of the Universe (even though very few of them justify this term). As for the four vector bosons, they are usually described as interaction carriers, as a glue that fixes the 'building blocks' to one another. This comparison describes some of them extremely well. Experiments have thus established four vector bosons (we again ignore their antiparticles): the photon γ, the gluon g, the neutral weak boson Z^0, and the charged weak bosons $W^{\pm}$, which are antiparticles to each other.

The photon has been studied better than the rest. Free photons create free electromagnetic fields: radio waves, light, x-rays, and gamma-quanta. The photon mass is zero. Consequently, the photon energy E in vacuum equals the absolute value of $\mathbf{k}$, of its momentum (times c): $E = c|\mathbf{k}|$. Photons with $E^2 \neq c^2\mathbf{k}^2$ are said to be virtual. The Coulomb field in the hydrogen atom is created by virtual photons with $-c^2\mathbf{k}^2 \gg E^2$. The electric charge e is the source of photons. We know from optics and atomic physics that $e^2/\hbar c = \alpha \simeq 1/137$. (Here $\hbar = h/2\pi$ and h is the Planck quantum of action). All electromagnetic interactions are due to photon exchange. The theory describing electromagnetic interactions is known as quantum electrodynamics (QED).

The carriers of the strong interactions are the gluons. Gluons' sources are the so-called 'color' charges. These charges have no relation to ordinary color and were given this name only for convenience of talking about them. Each of the six species of quarks (also known as quark flavors, u, d, c, s, t, b) exists in three color forms: yellow, blue or red (y, b, r). Antiquarks carry the corresponding color anticharges: $\bar{y}$, $\bar{b}$, $\bar{r}$. It is important to emphasize that the three charges (y, b, r) and the three anticharges ($\bar{y}$, $\bar{b}$, $\bar{r}$) are completely independent of quark flavors. We have hidden part of the truth when stating that there exist six different quarks: in fact, they are twelve if antiquarks are taken into account and 36 if color charges are. However, this 'sextupling' can be interpreted as six different states of one particle.

The color states of gluons are even more complicated. A gluon has two color indices, not one. It is easy to show that the number of different gluons is eight: $3 \times \bar{3} = 8 + 1$. The combination $y\bar{y} + b\bar{b} + r\bar{r}$ is 'snow-white' and thus carries no color charge. In contrast to photons in QED, which are electrically neutral, gluons in quantum chromodynamics (QCD) are carriers of color charges and thus have to emit and absorb gluons. The result is a completely unfamiliar behavior for the strong interactions of gluons and quarks: their interaction energy increases in response to attempts to separate them.

As a result, free gluons and quarks cannot exist: they are 'self-confined' within colorless hadrons. The complete quantitative theory of this confinement has not been constructed yet but qualitative confirmations have been obtained by computer simulations on four-dimensional lattices and by the so-called QCD sum rules.

4. Running coupling constants

The dimensionless quantities $\alpha = e^2/\hbar c$, $\alpha_s = g^2/\hbar c$, $\alpha_W = g_W^2/\hbar c$, $\alpha_Z = g_Z^2/\hbar c$ are often referred to as coupling constants. It was established as early as in the 1950s that $\alpha \approx 1/137$ is a constant only at zero (or very low) momentum transfer q^2. Owing to vacuum polarization, α grows with increasing q^2 and tends to infinity at high but still finite q^2. This phenomenon is known as the Landau – Pomeranchuk pole and was interpreted as a manifestation of the non-self-consistency of QED. Once QCD was born, it became clear that $\alpha_s(q^2)$ [in contrast to $\alpha(q^2)$] tends to zero as $q^2 \to \infty$. This phenomenon is known as asymptotic freedom. As a result of the asymptotic freedom, the perturbation theory provides an adequate description of gluon and quark collisions at small distances (at high q^2), generating the so-called jets of hadrons. The reverse face of asymptotic freedom is confinement: color attraction becomes irresistibly strong at large distances. Of considerable experimental significance for studying the confinement is the search for the so-called 'exotic hadrons' that contain, in addition to three valent quarks (in the case of a baryon) and a quark and an antiquark (in the case of a meson), additional quark – antiquark pairs or valent gluons. Of special interest are the so-called glueballs that consist only of gluons.

In comparison with the fast-running α_s, the weak interaction constants α_Z and α_W crawl rather than run: they grow by about one per cent from $q^2 = 0$ to $q^2 \simeq 100$ GeV^2. However, if the progress of all constants is extrapolated towards higher q^2, they show a tendency to coming together at a single point somewhere around $q^2 = 10^{15} - 10^{16}$ GeV, where $\alpha_s \sim \alpha_W \sim (8/3)\alpha \sim 1/40$. We can hope, therefore, that a unified theory of the electroweak and strong interactions exists at such high q^2. Such theories are described by symmetry groups that incorporate leptons and quarks in a single multiplet and predict proton decay. The search for proton instability was especially intense 10 – 15 years ago but unfortunately failed to discover the expected effect.

To complete the survey of the sixteen fundamental particles, we give in Table 2 the particle masses (or the upper

Table 2.

$\nu_e < 10$ eV 1956	$\nu_\mu < 170$ KeV 1962	$\nu_\tau < 24$ MeV 1975, 1998	$\gamma < 10^{-15}$ eV 1926
e 0.51 MeV 1897	μ 105.7 MeV 2×10^{-6} s 1937, 1947	τ 1777 MeV 3×10^{-13} s 1975	g 0 1973
u 5 MeV 1964	c 1300 MeV 10^{-12} s 1974	t 176 GeV $\Gamma = 2$ GeV 1994	Z 91.2 GeV $\Gamma = 2.5$ GeV 1983
d 10 MeV 1964	s 150 MeV 1964	b 4.3 GeV 10^{-12} s 1977	W 80.4 GeV $\Gamma = 2.1$ GeV 1983

limits for the masses of the neutrinos and the photon), their lifetimes (or decay widths) and the years of their experimental discovery.

The quark masses in Table 2 should not be taken too literally; indeed, quarks do not exist as free isolated particles. These values characterize quark masses deep inside hadrons. Quark lifetimes should not be interpreted literally either. The only quark that is 'born free and dies free' is the heaviest of them, the t quark. Its lifetime is so short that it simply has no time to form hadrons with the quarks that accompany its birth.

5. Colliders

The heavy quarks c, b, t, the τ leptons, the gluons and the W and Z bosons were mostly discovered and studied in experiments with colliding particle beams. Storage ring accelerators with colliding beams are known as colliders. Table 3 lists twelve electron – positron colliders that have been functioning (or functioned) from the beginning of the 1970s, and one electron – proton collider. We list in Tables 3 to 5 the name of the collider, the institute, the center or town where it is located, the duration of operation, and the energy of each beam in GeV.

Table 3.

SPEAR	SLAC	1972–1990	4 × 4
DORIS	DESY	1973–1993	5.6 × 5.6
CESR	Cornell U.	1979–	6 × 6
PETRA	DESY	1978–1986	23.4 × 23.4
PEP	SLAC	1980–1990	15 × 15
BEPC	Beijing	1989–	2.2 × 2.2
VEPP-4M	Novosibirsk	1994–	6 × 6
TRISTAN	KEK	1987–1995	32 × 32
SLC	SLAC	1989–	50 × 50
LEP I	CERN	1989–1995	50 × 50
VEP-2M	Novosibirsk	1992–	0.7 × 0.7 (0.55 × 0.55)
LEP II	CERN	1996–2000	up to 100 × 100
HERA	DESY	1992–	$e^{\pm}$30 × p820

The ψ meson, the first of the numerous energy levels of charmonium (the system consisting of c and $\bar{c}$), was discovered at the SPEAR collider in 1974. The same particle, christened J, was discovered simultaneously at the Brookhaven accelerator in an experiment with a fixed target.

The Υ meson, the first of the numerous levels of upsilonium (or bottomonium, the system consisting of b and $\bar{b}$), was also discovered in a collider. To study the properties of b quarks, e^+e^- colliders were built with beam energies around 6 GeV.

LEP I and SLC were constructed for precision measurements of the properties of Z bosons. The four detectors of LEP I recorded 20 million events of the resonance creation and decay of Z bosons. The number of Z bosons at the SLC collider is fewer by two orders of magnitude but they are produced by a polarized electron beam, which makes it possible to conduct unique measurements of Z-boson properties.

The LEP II collider, which operates in the same 27 km-long ring tunnel in which LEP I worked before it, was built to study the creation of W^+W-boson pairs and also to search for new particles: the scalar Higgs boson and the so-called supersymmetric particles that we will discuss later. To conclude the comments to Table 3, we note that the highest precision in measuring the τ lepton mass was achieved with the Beijing accelerator, and that the experiments at the lowest energies in Novosibirsk are required for the high-precision testing of QED.

Three high-current e^+e^- colliders listed in Table 4 will start operation in 1999 specially to study the mechanisms of CP-invariance violation.

Table 4.

DAΦNE	Frascati	1999	0.51 × 0.51 (0.75 × 0.75)
KEKB	KEK	1999	8 × 3.5
PEP II	SLAC	1999	3.1 × 9

The collider in Italy will study K mesons and those in Japan and USA, B-meson decays. Note that to be able to observe B-meson decays in flight (not at rest), the colliding electrons and positrons must have unequal energies.

The proton – antiproton colliders are shown in Table 5.

Table 5.

$Sp\bar{p}S$	CERN	1981–1990	315 × 315
Tevatron	FNAL	1987–	1000 × 1000

The European collider was used to discover, in 1983, the W and Z bosons (the first publications contained less than ten events). This collider was designed on the basis of the already existing proton accelerator SpS, specially for the discovery of the W and Z bosons whose masses were predicted by the theory of the electroweak interaction. The t quark was discovered in 1994 at the American collider.

The fate of the proton – proton colliders proved to be rather sad. The largest of them, with a ring tunnel more than 80 km long and an energy for each proton beam of 20000 GeV = 20 TeV, was called the SSC: the superconducting supercollider. In 1993 the US Congress canceled the decision to build the SSC after three billion dollars had been spent on its construction near Dallas in Texas.

The superconducting acceleration-storage complex UNK of the Institute of High Energy Physics (in Protvino, near Serpukhov, Russia) was expected to reach colliding beam energies of 3 TeV. The ring tunnel longer than 20 km was completed; however, further building work was discontinued, owing to lack of funding.

There are grounds for hoping that the LHC — the Large Hadron Collider at CERN with a proton energy of 7 TeV, whose construction is to begin in the tunnel in which LEP II is now operating, will be completed in 2005. Not only the CERN member countries but also many others are participating in the construction of LHC, including USA, Japan and Russia.

The design plans for the construction of linear electron – positron colliders are still at a much earlier stage; the colliders may start operation after 2010. The energy range involved is 250 to 1000 GeV in a beam. The beam energy is only limited by the collider's length. In contrast to electron – positron ring colliders, the synchrotron radiation in linear machines is insignificant.

The $\mu^+\mu^-$ colliders, discussed since the early 1960s, would be truly tiny in comparison to the electron – positron colliders. The main difficulty here is the short lifetime of muons (two microseconds); however, a muon with energy 1 TeV lives for 20 ms and can travel up to 6000 km.

6. Why do we need still larger colliders?

The first goal is to search for new particles that are required for theorists to achieve self-consistency in the description of nature. The so-called Standard Model is based on the local (gauge) symmetry described by the $SU(3)_c \times SU(2)_W \times U(1)_Y$ group. Here $SU(3)_c$ is the symmetry of the strong color interaction of quarks and gluons. $SU(2)_W \times U(1)_Y$ describes the electroweak interaction. All fermions and vector gauge bosons in non-broken symmetry are massless. As a result of a spontaneous breaking of the $SU(2)_W \times U(1)_Y$ symmetry the neutral bosons W^0 and B^0 are mixed. One of their superpositions describes a massless photon that corresponds to the non-broken gauge invariance of electrodynamics, as discovered by V A Fock in 1926. The other superposition describes the massive Z^0 boson. Charged $W^{\pm}$ bosons also gain masses as a result of spontaneous symmetry breaking.

In each generation massless fermions form isotopic doublets if they are left-handed (the spin points against the momentum) and singlets if they are right-handed (the spin points along the momentum). (Left-handed neutrinos and right-handed antineutrinos were first discussed by L D Landau in 1956.) The 100% breaking of the mirror symmetry P and the charge symmetry C is thus built into the theory 'by hand'.

In the Standard Model fermions gain masses, just as the Z and W bosons, via the spontaneous breaking of the $SU(2)_W \times U(1)_Y$ symmetry realized by the Higgs mechanism. The Higgs mechanism in quantum field theory is a relativistic version of the Ginzburg–Landau mechanism in superconductors. The electrically neutral Higgs field forms a non-zero vacuum mean field — the 'vacuum condensate' — filling the entire Universe. The quanta of this field are particles with spin $s = 0$: Higgs bosons, or simply higgses. The masses of all particles in the model are $m = f\eta$, where m is the mass of a given particle, f is the coupling constant for the particle and the Higgs field, and η is the value of the condensate. We have no theoretical predictions for the values of f but we know the value of η with high accuracy:

$$\eta = (\sqrt{2}G_F)^{-1/2} = 246\ \text{GeV}\,.$$

(Here G_F is the familiar constant of the week four-fermion interaction.)

On the η scale, the most natural is the mass of the t quark: it corresponds to $f \simeq 1$. However, the hierarchy of fermion masses (of coupling constants f) remains unclear. The smallness of the neutrino masses looks especially baffling.

The W and Z boson masses were predicted using the Standard Model. The t quark mass was predicted on the basis of a theoretical analysis of high-precision data obtained by measuring electroweak radiative corrections in Z boson decays (20 million events!). Predictions of the higgs mass are much less reliable because it affects the radiative corrections much less than the t quark mass. The simplest variants of supersymmetric models predict 130 GeV for the upper bound on the higgs mass. The schemes of Grand Unification of the strong and electroweak interactions predict this limit to be 200 GeV (the prediction is based on renormalization group equations). Most theorists expect that $m_Z < m_h < 2m_Z$. If $m_h < 2m_W$, the principal decay signal of higgs creation must be the decay to two gamma-quanta. Although this is an extremely rare decay, other decays are very hard to extract from the background. The hunt for the higgs at LEP uses the reaction $e^+e^- \to Z^0h$. If LEP II fails to find higgs, LHC will. The scenarios with a very heavy higgs appear to be too ugly to a theorist. If $m_h > 800$ GeV, the Landau–Pomeranchuk pole in the higgs sector shifts to a physically observable region.

The search for the higgs is Problem No. 1 of high-energy physics. Once the higgs is discovered, it will be necessary to find out whether there are additional types of higgs. If there are several of them, the properties of the Higgs sector of the theory will become an area of detailed analysis.

Problem No. 2 of high-energy physics is the search for supersymmetry (SUSY). There are three arguments in favor of believing that SUSY exists:

1. The cancellation of quadratic divergences in the higgs sector: they threaten to transform the Fermi scale ($\sim 10^2$ GeV) into the Planck scale ($\sim 10^{19}$ GeV, see clarification below).
2. The unification of all interactions, including gravitation (see below).
3. The mathematical elegance of SUSY.

A host of supersymmetric theories is available. The simplest of them is the $N = 1$ SUSY in which each of 'our' particles has a supersymmetric partner differing only in the value of spin (in a non-broken SUSY) and mass (in a broken SUSY; we know that the SUSY in nature is indeed broken!). Here is a brief list of supersymmetric pairs:

1. Photon; $s = 1$ — photino; $s = 1/2$.
2. Gluon; $s = 1$ — gluino; $s = 1/2$.
3. Z boson; $s = 1$ — zino; $s = 1/2$.
4. Higgs; $s = 0$ — higgsino; $s = 1/2$.
5. Neutrino; $s = 1/2$ — sneutrino; $s = 0$.
6. W boson; $s = 1$ — wino; $s = 1/2$.
7. Leptons; $s = 1/2$ — sleptons; $s = 0$.
8. Quarks; $s = 1/2$ — squarks; $s = 0$.

The 3 through 5 superparticles are also known as neutralinos, and the 6 through 8 superparticles, as chargino. The lightest of the superparticles must be stable, provided the so-called R-parity is conserved. Such stable particles may constitute the main component of the so-called 'dark matter' which comprises from 90% to 99% of the total mass of the Universe. The search for dark matter is conducted both on the Earth, in low-background laboratories, and by astronomical methods (in the search for gravitational lenses). Superparticles may be discovered at LEP II, Tevatron, and LHC.

Another particle, in whose existence theorists firmly believe but which will never be detected by experimenters, is the graviton. It is the quantum of the gravitational field, a neutral massless particle with spin $s = 2$. Gravitational antennae will soon be able to detect gravitational waves but the individual particles of these waves interact sufficiently strongly only at fantastically high (Planckian) energies in the center-of-mass reference frame:

$$E \sim m_P c^2 = \left(\frac{\hbar c}{G_N}\right)^{1/2} c^2 = 1.22 \times 10^{19}\ \text{GeV}\,,$$

where G_N is Newton's constant.

The quantum theory of gravitation has not yet been constructed. In the supergravity theory, the graviton is accompanied by the gravitino ($s = 3/2$).

The studies devoted to the particles already discovered and those likely to be discovered in the foreseeable future are known as 'phenomenological'. Those devoted to particles that will never be observed in experiments are known as 'theoretical'. This group of studies assumes that the quantum field theory in general and the Standard Model in particular are phenomenological theories. The fundamental theory is then not a theory of pointlike particles but one dealing with strings whose size is of the order of $\hbar/m_{P}c \sim 10^{-33}$ cm. These strings move in multidimensional spaces and possess boson – fermion symmetry, that is, they are superstrings. Superstrings are coupled in ten-dimensional space to multidimensional supermembranes (branes). The latest achievement in this field is the idea that there exists an all-encompassing M-theory. Various versions of superstring models suggested previously have been generated by perturbative expansions at different points in the parameter space of the future M-theory.

Can theorists succeed in devising, by a shear mental effort, a unified theory in which all the experimentally observed masses, mixing angles and coupling constants will be fixed by the imposed self-consistency of the theory, instead of being free parameters? In my opinion, theorists cannot do it without experimenters' help. The foremost problem is to study the properties of higgses and superparticles in the mass range up to and of the order of 1 TeV. It is necessary to understand the mechanism of violation of CP invariance. We need to clarify how the neutrino sector works (masses, mixing angles, phases). Taken together, this may offer a launching pad for extrapolation from 10^3 to 10^{19} GeV. But we will also need luck.

This page intentionally left blank

Physics – Uspekhi **42** (10) 1045 – 1050 (1999)

METHODOLOGICAL NOTES

PACS numbers: **01.40.-d, 01.55.+b, 04.20.-q**

Gravitation, photons, clocks

L B Okun', K G Selivanov, V L Telegdi

Contents

Abstract. This paper is concerned with the classical phenomenon of gravitational red shift, the decrease in the measured frequency of a photon moving away from a gravitating body (e.g., the Earth). Of the two current interpretations, one is that at higher altitudes the frequency-measuring clocks (atoms or atomic nuclei) run faster, i.e. their characteristic frequencies are higher, while the photon frequency in a static gravitational field is independent of the altitude and so the photon only reddens relative to the clocks. The other approach is that the photon reddens because it loses the energy when overcoming the attraction of the gravitational field. This view, which is especially widespread in popular science literature, ascribes such notions as a 'gravitational mass' and 'potential energy' to the photon. Unfortunately, also scientific papers and serious books on the general theory of relativity often employ the second interpretation as a 'graphic' illustration of mathematically immaculate results. We show here that this approach is misleading and only serves to create confusion in a simple subject.

1. Introduction

In the literature, two types of red shift are known: gravitational and cosmological. As a rule, they are considered independently of each other. Gravitational red shift occurs when a photon moves away from a massive body (for instance, the Earth or the Sun) that can be considered as a static object. Observable values of the red shift are usually very small. This paper is devoted to the gravitational red shift.

The cosmological red shift is the red shift of light (photons) from remote galaxies caused by their recession. This shift is often called the Hubble shift. It is large in magnitude: for the most-distant observed galaxies, $\Delta\lambda/\lambda \approx 5$, where λ is the wavelength of the radiated light. In what follows, we will not discuss cosmological red shift.

The phenomenon of gravitational red shift was predicted by A Einstein in 1907 [1], and he discussed it in 1911 [2] before the creation of general relativity (GR). When GR was constructed by Einstein [3], gravitational red shift became one of the three classical effects of this theory (see Refs [4, 5]; as to the history of creation of GR, see Refs [6 – 8]).

Phenomenologically, without theoretical interpretation of the phenomenon, it can be described as follows: the frequency of light emitted by two identical atoms is lower for the atom 'sitting' deeper in the gravitational potential. Starting from 1960, unique experiments were carried out aimed at measuring various manifestations of the phenomenon [9 – 14]. These experiments have been discussed in excellent review papers [15 – 25], whose main purpose is to compare the experimental data with theoretical predictions, not only of GR, but of different nonstandard theories of gravitation as well. Interpretations of the gravitational red shift in the framework of the standard theory are not discussed in those reviews.

The authors of the majority of monographs on GR (see, for instance, Refs [26 – 35]) follow the interpretation given by Einstein in 1916 – 1920 (see Refs [3 – 5]), according to which the gravitational red shift is caused by the universal property of the standard clocks (atoms, atomic nuclei). In the general case of an arbitrary gravitational field and arbitrary velocities of emitting and absorbing atoms, the proper time interval between events of emission of two photons measured by the standard clock at the point of emission differs from the proper time interval between events of absorption of these photons measured by an identical standard clock at the point of absorption. The ratio of these time intervals gives the invariant description of the red shift. This formulation of the red shift was first given by H Weyl in 1923 [36].

In the case of a static gravitational potential and fixed atoms, the picture gets simplified, since there exists a distinct time (the time coordinate) upon which the metric does not depend. This time can be taken as the universal (world) time. With this choice, the energy difference of two atomic levels

L B Okun', K G Selivanov State Scientific Center of Russian Federation 'Institute for Theoretical and Experimental Physics',
ul. B. Cheremushkinskaya 25, 117259 Moscow, Russian Federation
E-mail: okun@heron.itep.ru, selivanov@heron.itep.ru
V L Telegdi EP Division, CERN, CH-1211 Geneva 23, Switzerland
E-mail: valentine.telegdi@cern.ch

Received 19 May 1999
Uspekhi Fizicheskikh Nauk **169** (10) 1141 – 1147 (1999)
Translated by T Dumbrajs; edited by A Radzig

increases with increasing distance between an atom and the Earth, whereas the energy of a photon remains unchanged. (In what follows, we will speak about the Earth, but it could be any massive body.) Thus, the phenomenon called the red shift of a photon is actually the blue shift of an atom. As to the values of proper time at different points, they are expressed in terms of the universal time with the help of a factor that depends on the gravitational potential and, therefore, has different values at different points.

In textbooks and monographs published in recent years [37–40], red shift is described with the use of mathematical constructions such as orthonormalized bases (a sequence of proper frames of reference), with respect to which both the energy and the parallel transport of the 4-momentum of a photon along its world line are defined. Frequently, such a rigorous mathematical description is accompanied by nonstrict verbal representations of a photon that loses its energy when it 'gets out' of the gravitational potential well. Even some classical textbooks and monographs [41–43] make use of a similar 'visual phraseology'. The experts on GR do not pay attention to it — for them, this is merely a tribute to the tradition of scientific popularization. However, nonexperts should be warned that the content of mathematical formulae underlying the description of the gravitational red shift drastically differs from the 'heuristic' (and incorrect) arguments, discussed above and widespread in many elementary textbooks (see, for instance, Refs [44–50]).

Their authors proceed from the implicit supposition that a massless photon is similar to a conventional massive nonrelativistic particle, call the photon energy E divided by the speed of light squared c^2 the photon mass, and consider the 'photon potential energy' in the gravitational field. Only exceptional popular-science texts (see, e.g., Ref. [51]) do not contain this incorrect picture and emphasize that the energy and frequency of a photon do not change as it moves higher and higher.

2. Experiments

The first laboratory measurements of the gravitational red shift were performed at Harvard in 1960 by R Pound and G Rebka [9, 10] (with an accuracy of 10%). Within an accuracy of 1%, an experiment of that sort was carried out later by R Pound and J Snider [11]. Photons were moving in a tower 22.5 m high. The source and an absorber of photons (γ-rays with energy 14.4 keV) were nuclei of the isotope ^{57}Fe. To diminish possible systematic errors, observations were made both for reddening and blue-shifting of a photon. In the first case, the source was placed in a basement and the absorber, in an attic. In the second case, they changed places.

The measured frequency shift was very small, $\Delta\omega/\omega \approx 10^{-15}$. This accuracy could be achieved owing to the Mössbauer effect discovered in 1958, due to which photon lines in a crystal are extremely monochromatic. The gravitational red shift was compensated by the Doppler effect: the absorber was slowly moving in the vertical direction, thus restoring the resonance absorption of photons.

As to the interpretation of the results obtained, there is some ambiguity in the papers by Pound and colleagues. Though they mention the interpretation in terms of the clock, with reference to Einstein's papers, their paper [9] is entitled "Apparent weight of photons"; and the report made by Pound in Moscow [10] was entitled "On photon weight". From the title of the paper by Pound and Snider [11] "Effect of gravity on nuclear resonance" it can be concluded that they did not want to choose between alternative interpretations.

Unlike the original papers by Pound and colleagues, most of the reviews covering gravitational experiments [18–24] consider their result as a test of clock behavior in a gravitational field. Actually, the experiments themselves do not provide the choice between the two interpretations, unless GR is taken as the basis in making that choice. The reason is that they measure the relative shift of photon and nuclear frequencies, and each of the frequencies is not measured separately. The same remark also concerns the shift of the photon (radio wave) frequency with respect to the frequency of the atomic standard (a hydrogen maser) measured with a rocket that was launched to altitude 10,000 km and then fell into the ocean [12]. In this experiment, the theory of gravitational red shift was verified to an accuracy of the order 10^{-4}.

For a direct test of the dependence of the atomic-clock rate on height (without photons), experiments were carried out, in which the clock was in the air for a long time on aircraft [13, 14] (see also reviews [18–24]). In these experiments, the clock was brought back to the laboratory, where its reading was compared to that of an identical clock staying put on the Earth. (Besides, in experiment [14], the shift of the aircraft clock was observed from the Earth telemetrically.) It turned out that, in accordance with general relativity, the clock onboard went ahead by $\Delta T = (gh/c^2)T$, where T is the flight duration at height h, g is the gravitational acceleration, and c is the speed of light. (The accuracy of the experiment [13] that used a beam of cesium atoms was of the order of 20%. The accuracy of experiment [14] was 1.5%.)

This result was, of course, obtained when numerous background effects were taken into account. One of them was the famous 'paradox of twins': according to special theory of relativity, the moving clock after coming back to the starting point would lag behind the clock at rest. It is not difficult to derive the general formula describing the influence of both the gravitational potential ϕ and velocity u (see, for instance, Ref. [27]):

$$d\tau = dt\left[1 + \frac{2\phi}{c^2} - \frac{u^2}{c^2}\right]^{1/2}, \qquad (1)$$

where τ is the proper 'physical' time of the clock, and t is the above-mentioned so-called universal world time that can be introduced in the case of a static gravitational potential and that is sometimes called laboratory time, because it is this time which is shown by a clock at rest in the laboratory, where the value of ϕ is taken zero.

In his lectures on gravitation [35], R Feynman gave a detailed explanation for the change in the clock rate because of ϕ and u. He concluded that the center of the Earth should be "a day or two younger than its surface".

Apart from the tower, rocket, and aircraft experiments, 'desk' experiments were also carried out [52, 53] with the use of the Josephson effect that compensated the gravitational shift.

As to satellite experiments, they have been repeatedly discussed in the literature (see, for instance, Refs [54, 55]), however, we do not know their results.

Besides the shift of the photon frequency, the shift of the wavelength was also measured [56] (see also Fig. 38.2 in monograph [37] that illustrates experiment [56] and reviews [15–25]). In this experiment, the shift of a sodium line in the solar spectrum was measured (with an accuracy of 5%) with the help of a diffraction grating. The theory of such grating experiments will be discussed below (see Section 6). The shift of the solar absorption line of potassium was measured by means of the resonance scattering of sunlight by an atomic beam [57] (with an accuracy of 6%).

3. Theory up to 1916: Einstein elevator

Since in this paper we mainly discuss the gravitational red shift in the field of the Earth, we choose the frame of reference, in which the Earth is at rest (its rotation is neglected).

As is well known, the potential is determined to within an additive constant. When a gravitational potential $\phi(r)$ is considered at a certain distance r from the center of the Earth, it is convenient to choose $\phi(\infty) = 0$. Then ϕ at any finite r is negative.

At an altitude h near the Earth surface ($h = r - R \ll R$, where R is the Earth radius), use can be made of the linear approximation:

$$\delta\phi(h) = \phi(R+h) - \phi(R) = gh\,, \tag{2}$$

where g is the standard gravitational acceleration. Note that $\delta\phi(h) > 0$ for $h > 0$. We will discuss the red shift only in the first order of the parameter gh/c^2.

The linear approximation, Eqn (2), is valid for laboratory and aircraft experiments. It is, however, obvious that it does not work for a rocket at a high altitude ($h \simeq 10^4$ km). In this case, the potential $\delta\phi(h)$ is to be replaced by the Newton potential $\phi(r)$, which is, however, unessential for the dilemma 'clocks or photons' that is the subject of the present paper.

The first papers by Einstein [1, 2] on gravitational red shift contained a lot of basic ideas that entered into numerous texts of various authors (sometimes, without a proper critical analysis). He considered the Doppler effect in a freely falling system and found the frequency of an atom (clock) to increase with increasing height (potential). A corner-stone of his reasoning was the principle of equivalence formulated by him: local equivalence between the behavior of physical systems in the gravitational potential (2) and in a properly accelerated frame of reference (an elevator). In an elevator like that, the observer cannot detect any manifestations of the gravitational field, whatever local experiments he might carry out. (Notice that experiments with unshielded electric charges are not local since the Coulomb field of such charges extends to infinity.)

Consider now, from an elevator freely falling with acceleration g, the emission and absorption of a photon. A photon with frequency ω is emitted upward by an atom at rest on the Earth surface. An identical atom that should absorb the photon is at rest at the height h. In a freely falling elevator, the gravitational field does not affect the photon, and therefore it conserves its initial frequency. Let us assume that at the moment of emission of a photon ($t = 0$) the velocity of the elevator equals zero. Then at the moment $t = h/c$, when the photon reaches the upper atom, the velocity of the latter with respect to the elevator is equal to $v = gh/c$ and directed upward: the atom 'runs away' from the photon. As a result, the photon frequency seen by the absorbing atom is diminished by the linear Doppler effect, and the photon reddens:

$$\frac{\Delta\omega}{\omega} = -\frac{v}{c} = -\frac{gh}{c^2}\,. \tag{3}$$

Consider now a different design of the experiment. Let the upper atom (an absorber) be moving in the laboratory system downward with a constant velocity $v = gh/c$. Then its velocity in the reference frame of the elevator is zero at the moment of absorption, and it can absorb the photon emitted with frequency ω resonantly, in complete agreement with experiments [9–11]. Obviously, in the reference frame connected with the freely falling elevator, it is impossible to interpret the red shift as energy loss by a photon in the process of overcoming the gravitational attraction, since there is no gravitational attraction in that elevator.

No less visual is the interpretation in the laboratory reference frame. In a static field, the frequency of a photon is conserved, whereas in the reference frame of an atom moving towards the photon, it increases due to the Doppler effect and compensates the 'blue shift' of the atom in the gravitational field.

4. General relativity: metric

Till now, we used only the special theory of relativity (constancy of the speed of light and the Doppler effect) and Newtonian gravitation in the approximation of a linear potential. As is known, a consistent relativistic description of classical gravitation is provided by general relativity with its curved space-time metric. The theory is based on the metric tensor $g_{ik}(x)$, $i,k = 0,1,2,3$ that is transformed under changes of coordinates so that the interval $\mathrm{d}s$ between two events with coordinates x^i and $x^i + \mathrm{d}x^i$

$$\mathrm{d}s^2 = g_{ik}(x)\,\mathrm{d}x^i\,\mathrm{d}x^k \tag{4}$$

remains unchanged. Setting $\mathrm{d}x^1 = \mathrm{d}x^2 = \mathrm{d}x^3 = 0$, we arrive at the relation between the interval of proper time $\mathrm{d}\tau = \mathrm{d}s/c$ and the interval of world [1] time $\mathrm{d}t = \mathrm{d}x^0/c$ for an observer at rest:

$$\mathrm{d}\tau = \sqrt{g_{00}}\,\mathrm{d}t\,. \tag{5}$$

In the static case, the integration of equation (5) gives

$$\tau = \sqrt{g_{00}}\,t\,, \tag{6}$$

where in the general case g_{00} is a function of $\mathbf{x}$, whereas g_{00} in Eqn (2) depends only on $x^3 = z = h$.

The proper time τ is measured with any standard clock. It can also be considered as the coordinate time in the so-called comoving locally inertial reference frame, i.e. in the locally inertial system that at a given moment at a given point has a zero velocity with respect to the laboratory system. (Imagine a stone thrown up from the earth at the top point of its trajectory.) If there is a set of standard clocks at different points, their proper times τ are related to the world (laboratory) time t in different ways, because g_{00} depends on $\mathbf{x}$ [see formula (6)]. This explains the aircraft experiments [13, 14].

A weak gravitational field can be described in terms of the gravitational potential ϕ, and in this case g_{00} is expressed through ϕ as follows:

$$g_{00} = 1 + \frac{2\phi}{c^2}\,. \tag{7}$$

[1] Recall that world time is sometimes called laboratory time. The first name reflects the fact that this time is the same for the whole world; the second, that it can be set using a standard clock in the laboratory [see the text after Eqn (1)]. Many authors call t coordinate time.

The physical meaning of this expression will be explained somewhat later [see formulae (8)–(10)]. According to equations (5) and (7), the clock runs slower in the laboratory which is deeper in the gravitational potential.

In analogy with Eqn (5), the rest energy of a body in the laboratory system, $E_0^{\rm lab}$, and that in the comoving locally inertial system, $E_0^{\rm loc}$, are related by the formula

$$E_0^{\rm lab} = E_0^{\rm loc}\sqrt{g_{00}} \qquad (8)$$

(note that $E_0^{\rm lab}\,dt = E_0^{\rm loc}\,d\tau$; this relationship holds true, since the energy E is a zero component of a covariant 4-vector, whereas dt is a zero component of a contravariant 4-vector).

The rest energy of a body in the locally inertial system is the same as in the special theory of relativity (see, for instance, Refs [58, 59] and [48], p. 246, 3rd English edition):

$$E_0^{\rm loc} = mc^2\,, \qquad (9)$$

whereas the rest energy in the laboratory system $E_0^{\rm lab}$ also contains the potential energy of a body in the gravitational field. Notice that in Eqn (8) this potential energy is 'hidden' in g_{00}, in conformity with the fundamental principle of general relativity: gravitation enters only through the metric. The relationship between the metric and potential, Eqn (7), can be considered as a consequence of equations (8), (9), and the relation

$$E_0^{\rm lab} = mc^2 + m\phi \qquad (10)$$

that generalizes the notion of rest energy of a free particle to such in a weak gravitational field.

Now we are able to explain the red shift in the laboratory frame of reference. According to Eqn (8) or (10), the difference between energies of atomic or nuclear levels in this system, $\varepsilon_{\rm lab}$, depends on the position of an atom. The deeper the atom sits in the gravitational potential, the smaller is $\varepsilon_{\rm lab}$. For an atom-absorber that is situated at height h relative to an identical atom emitting a photon, the relative energy difference of levels equals

$$\frac{\Delta\varepsilon^{\rm lab}}{\varepsilon^{\rm lab}} = \frac{gh}{c^2}\,. \qquad (11)$$

[In formula (11), like in Eqn (2), we made use of the linear approximation.] It can be said that the energy levels of an absorbing atom are slightly 'bluer' than those of an emitting atom. Equation (11) is certainly nothing else than a way of describing the difference in the rate of two atomic clocks one above the other located at height h. On the other hand, the energy (frequency) of a photon is conserved as it moves in the static gravitational potential. This can be seen, for instance, from the wave equation for the electromagnetic field in the presence of a static gravitational potential or from the equation of motion of a particle (massless or massive) in the static metric. From all the aforesaid it is clear that in the laboratory reference frame, there is no place for the interpretation according to which 'a photon loses its energy for overcoming the action of the gravitational field'.

And finally, we can discuss the experiment on the red shift by using a sequence of locally inertial reference frames that are comoving the laboratory clocks (atoms) at the moment when a photon passes through them. As we have explained above, the standard clocks in such reference frames run with the same rate, the rest energy of an atom is equal to its mass times c^2 [see Eqn (9)], and the energies of atomic levels are the same as at infinity. On the other hand, the energy of a photon in the laboratory system, $E_\gamma^{\rm lab} = \hbar\omega^{\rm lab}$, and that in a comoving locally inertial system, $E_\gamma^{\rm loc}$, are related as follows

$$E_\gamma^{\rm lab} = E_\gamma^{\rm loc}\sqrt{g_{00}}\,. \qquad (12)$$

Equation (12) can be derived from equation (8) if we notice that a photon can be absorbed by a massive body and consider the increase in energy of this body. Thus, since $E_\gamma^{\rm lab}$ is conserved and $E_\gamma^{\rm loc}$ decreases with increasing height, we arrive at the following expression

$$\frac{\omega^{\rm loc}(h) - \omega^{\rm loc}(0)}{\omega^{\rm loc}(0)} = \frac{E_\gamma^{\rm loc}(h) - E_\gamma^{\rm loc}(0)}{E_\gamma^{\rm loc}(0)} = -\frac{gh}{c^2}\,, \qquad (13)$$

which is just the observed red shift of a photon. However, note is to be made that the decrease in $E^{\rm loc}$ is not at all caused by the work done by the photon against the gravitational field. (The gravitational field is absent in a locally inertial system.) The energy $E_\gamma^{\rm loc}$ changes since in the given description one should pass from one locally inertial reference frame to another (from the one comoving the laboratory at the moment of emission to that at the moment of absorption).

5. Pseudoderivation and pseudointerpretation of the gravitational red shift

The simplest (and incorrect) explanation of the red shift is based on assigning the inertial gravitational mass $m_\gamma = E_\gamma/c^2$ to a photon. Owing to this mass, a photon is attracted to the Earth with the force gm_γ, as a result of which the relative change in its energy (frequency) at height h equals

$$\frac{\Delta E_\gamma}{E_\gamma} = \frac{\Delta\omega}{\omega} = -\frac{gm_\gamma h}{m_\gamma c^2} = -\frac{gh}{c^2}\,. \qquad (14)$$

Notice that (up to a sign) this is exactly the formula for the blue shift of an atomic level, which is not surprising. An atom and a photon are here considered in the same way: both of them are treated nonrelativistically! This is certainly wrong for the photon. If the explanation in terms of the gravitational attraction of a photon to the Earth were correct, one should expect red-shift doubling (summation of the effects of the clock and photon) in an experiment of Pound–Rebka type.

Some readers may attempt to use Einstein's authority in order to defend the above pseudoderivation. In a paper of 1911 [2], Einstein put forward the idea that the energy is not only the source of inertia, but also the source of gravitation. He used the heuristic argument: "If there is a mass, there is an energy, and vice versa". As he realized later, this "vice versa" was not so correct as the direct statement was (a photon possesses energy, whereas its mass equals zero). Identifying the energy and mass, he calculated the energy loss of a photon moving in the vertical direction in the gravitational field of the Earth, as discussed above. Taking advantage of the same heuristic principle, he also determined the deviation of a ray of light by the Sun that was half the correct deviation. Subsequently, in the framework of GR, Einstein found this missing factor of two [3–5]. The correct formula was verified experimentally.

6. Measurement of the wavelength

In previous sections, we discussed the gravitational red shift in terms of the frequency of a photon and that of a clock.

Now, we will discuss the same effect in terms of the photon wavelength and the diffraction-grating period. Consider two gratings at different heights. The lower grating serves as a monochromator, i.e. as a source of monochromatic light. The wavelength of a photon $\lambda^{\rm lab}(z)$ corresponds to its frequency, whereas the grating period in the vertical (z) direction, $l^{\rm lab}(z)$, corresponds to the frequency of a clock.

Though the energy of a photon $E^{\rm lab}$ is conserved in a static gravitational field, its momentum $p^{\rm lab}$ is not conserved. The relation between these quantities is provided by the condition for a photon being massless. This condition in the gravitational field reads as follows

$$g^{ij}p_i p_j = 0\,, \tag{15}$$

where g^{ij}, $i, j = 0, \ldots, 3$ are contravariant components of the metric tensor, and p_j are components of the 4-momentum, $p_0 = E^{\rm lab}$, $p_3 = p^{\rm lab} = 2\pi\hbar c/\lambda^{\rm lab}(z)$ (for a photon moving along the z-axis). In our case, the metric g^{ij} can be taken in diagonal form, then, in particular, $g^{zz} = 1/g_{zz}$.

From Eqn (15) one can easily determine the change of $\lambda^{\rm lab}(z)$ with height:

$$\lambda_{\rm lab}(z) = \sqrt{\frac{g^{zz}(z)}{g^{zz}(0)}}\sqrt{\frac{g^{00}(0)}{g^{00}(z)}}\,\lambda_{\rm lab}(0)\,. \tag{16}$$

On the other hand, the grating period in the z direction, $l^{\rm lab}(z)$, changes with height, too. This is merely a familiar change of the scale in the gravitational field explained, e.g., in the book by Landau and Lifshitz (see Sect. 84 in Ref. [28]):

$$l^{\rm lab}(z) = \sqrt{-g^{zz}(z)}\,l^0\,, \tag{17}$$

where l^0 is the 'intrinsic period' of the grating in the direction z, an analog of the proper frequency of the standard clock. Thus, the grating period $l^{\rm lab}(z)$ depends on z in the following way:

$$l^{\rm lab}(z) = \sqrt{\frac{g^{zz}(z)}{g^{zz}(0)}}\,l^{\rm lab}(0)\,. \tag{18}$$

Let us now take into account that the 'grating' version of an experiment of the Pound et al. type, at the difference of heights h, would measure the double ratio $[\lambda(h)/l(h)]/[\lambda(0)/l(0)]$. The result can be represented in the form

$$\frac{\Delta\lambda^{\rm lab}}{\lambda^{\rm lab}} - \frac{\Delta l^{\rm lab}}{l^{\rm lab}} = \sqrt{\frac{g^{00}(0)}{g^{00}(h)}} = \frac{gh}{c^2}\,, \tag{19}$$

where $\Delta\lambda^{\rm lab}/\lambda^{\rm lab} = [\lambda^{\rm lab}(h) - \lambda^{\rm lab}(0)]/\lambda^{\rm lab}(0)$, $\Delta l^{\rm lab}/l^{\rm lab}$ is defined analogously. Note that this result does not depend on g^{zz}, as might be expected, since there is freedom in the choice of the scale along the z-axis, and the observables should be independent of that choice. Equation (19) is analogous to the equation that describes the experiments performed by Pound et al.:

$$\frac{\Delta\omega}{\omega} - \frac{\Delta\epsilon}{\epsilon} = \sqrt{\frac{g_{00}(0)}{g_{00}(h)}} = -\frac{gh}{c^2}\,, \tag{20}$$

where ω is the photon frequency, and $\epsilon/\hbar$ is the clock frequency [see formula (11)]. Equation (20) requires a word of explanation. In the laboratory reference frame, the first term in the left-hand side equals zero, viz.

$$\frac{\Delta\omega^{\rm lab}}{\omega^{\rm lab}} = 0\,, \tag{21}$$

as discussed in Section 3; therefore, the whole contribution comes from the second term defined by equation (11).

However, we would like to emphasize a significant distinction from the case when the photon frequency is measured. In that case, one can independently measure the difference in the rate of upper and lower clocks [$\Delta\epsilon^{\rm lab}/\epsilon^{\rm lab}$ in equation (20)], which was done in aircraft experiments; whereas in the grating case, the change of the scale [$\Delta l^{\rm lab}/l^{\rm lab}$ in Eqn (19)] cannot be measured independently. This important distinction results from the fact that the metric is time-independent but essentially depends on z.

It is to be realized that such a purely laboratory experiment cannot be performed at the present level of development of experimental physics (recall the importance of the Mössbauer effect in experiments carried out by Pound et al.). However, an experiment of that sort is feasible in measuring a sufficiently large red shift, like the shift of the sodium line from the Sun [56]. Needless to say, the initial wavelength of light in this experiment was fixed by an atom on the Sun's surface rather than by the grating.

7. Conclusions

The present paper contains little original material: for the most part, it is pedagogic. Since gravitational red shift is one of the corner-stones of general relativity, both from theoretical and experimental points of view, it is highly important that its explanation should be maximally simple but, at the same time, correct, as the explanation based on the change of the clock rate in the gravitational field. An alternative explanation in terms of the mass ascribed to the photon — and the corresponding potential energy — is wrong and produces confusion. We have demonstrated that it is incorrect and schematically discussed experiments on the red shift in the framework of the correct approach. We would like to emphasize the importance of those experiments, in which an atomic clock was lifted to a high altitude, kept there for a sufficiently long time, and then compared with a twin that never left the Earth. The clock at height was fast compared to its twin. Thus, the 'blue' shift of the clock was established as an absolute effect. Hence it follows immediately that a naive explanation of the gravitational red shift in terms of the attraction of a photon by the Earth is incorrect.

Acknowledgments

We would like to thank V V Okorokov for his question concerning the consistency of experiments performed by Pound et al. with aircraft experiments. We are also grateful to S I Blinnikov, G Veneziano, A D Dolgov, A Yu Morozov, N Straumann, and K Thorne for interesting discussions. Special thanks are to E L Schucking for his help in an essential revision of our bibliography and for his insistence on a unified invariant description of the gravitational and cosmological red shift based on the Killing vectors. However, we could not follow his advice, since we wished to focus the attention of readers, not specialists on GR, on the erroneousness of the widespread naive interpretation of the gravitational red shift. One of us (L B Okun') is sincerely grateful to the theoretical division of CERN, where part of this work was accomplished, for hospitality. Our special gratitude is to J A Wheeler for his encouragement. This work was partially supported by grants of the Russian Foundation for Basic Research 96-15-96578 and 98-02-17372.

References

1. Einstein A *Jahrb. Radioaktivität Elektronik* **4** 411 (1907) [Translated into Russian in *Sobranie Nauchnykh Trudov* Vol. 1 (Moscow: Nauka, 1965) p.110]
2. Einstein A *Ann. Phys.* (Leipzig) **35** 898 (1911) [Translated into Russian in *Sobranie Nauchnykh Trudov* Vol. 1 (Moscow: Nauka, 1965) p. 167]
3. Einstein A *Ann. Phys.* (Leipzig) **49** 769 (1916) § 22 [Translated into Russian in *Sobranie Nauchnykh Trudov* Vol. 1 (Moscow: Nauka, 1965) p. 502]
4. Einstein A *Über die Specielle und Allgemeine Relativitäts Theorie* (*Gemeinverständlich*) (Braunschweig: Vieveg-Verlag, 1920) [Translated into English as *Relativity: The Special and General Theory* (New York: Crown Trade Paperbacks, 1961) p. 130] [Translated into Russian in *Sobranie Nauchnykh Trudov* Vol. 1 (Moscow: Nauka, 1965) p. 597]
5. Einstein A *The Meaning of Relativity* (New York: Princeton University Press, 1921) Eqn (106)] [Translated into Russian in *Sobranie Nauchnykh Trudov* Vol. 2 (Moscow: Nauka, 1965) p. 5]
6. Pais A "Subtle is the Lord ..." *The Science and the Life of Albert Einstein* (Oxford: Oxford University Press, 1982) Ch. 9 [Translated into Russian as *Nauchnaya Deyatel'nost' i Zhizn' Al'berta Eĭnshteĭna* (Moscow: Nauka, 1989) Ch. 9, p. 5]
7. Vizgin V P *Relyativistskaya Teoriya Tyagoteniya* (*Istoki i Formirovanie*) [Relativistic Theory of Gravitation (Sources and Formation)] (Moscow: Nauka, 1981)
8. Corry L, Renn J, Stachel J *Science* **278** 1270 (1997)
9. Pound R, Rebka G *Phys. Rev. Lett.* **4** 337 (1960); **4** 275 (1960); **3** 439 (1959)
10. Pound R *Usp. Fiz. Nauk* **72** 673 (1960) [*Sov. Phys. Usp.* **3** 875 (1961)]
11. Pound R, Snider J *Phys. Rev. B* **140** 788 (1965); *Phys. Lett.* **13** 539 (1964)
12. Vessot R, Levine M *Gen. Rel. Grav.* **10** 181 (1979)
13. Haefele J, Keating R *Science* **117** 166, 168 (1972)
14. Alley C et al., in *Experimental Gravitation: Proc. of the Conf. at Pavia* (*September 1976*) (Ed. B Bertotti) (New York: Academic Press, 1977)
15. Konopleva N P *Usp. Fiz. Nauk* **123** 537 (1977) [*Sov. Phys. Usp.* **20** 973 (1977)]
16. Rudenko V N *Usp. Fiz. Nauk* **126** 361 (1978) [*Sov. Phys. Usp.* **21** 893 (1978)]
17. Ginzburg V L *O Teorii Otnositel'nosti* (On the Theory of Relativity) (Moscow: Nauka, 1979) p. 157
18. Will C, in *General Relativity* (Eds S Hawking, W Israel) (Cambridge: Cambridge University Press, 1979) Table 2 [Translated into Russian (Moscow: Mir, 1983)]
19. Will C *Theory and Experiment in Gravitational Physics* (Cambridge: Cambridge University Press, 1981) Sect. 2.4 [Translated into Russian (Moscow: Energoatomizdat, 1986) p. 34]
20. Will C *Int. J. Mod. Phys. D* **1** 13 (1992)
21. Will C, Preprint WUGRAV-95-5; gr-qc/9504017 (1995)
22. Shapiro I, in *Astrofisica e Cosmologia, Gravitazione Quanti e Relatività* (Firenze: Giunti Barbera, 1979) Sect. 3 [Translated into Russian (Moscow: Mir, 1982) p. 218]
23. Ashby N, Spilker J, in *The Global Positioning System: Theory and Applications* Vol. 1 (American Institute of Aeronautics and Astronautics Inc., 1995) Ch. 18
24. Taylor J "Astronomical and space experiments to test relativity", in *General Relativity and Gravitation* (Cambridge: Cambridge University Press, 1987) p. 214
25. Tonnelat M A "Les frequences en relative generale" ("Frequencies in general relativity") *Ann. Inst. Henri Poincare* **1** (1) 79 (1964) [Translated into Russian in *Eĭnshteĭnovskiĭ Sbornik* (Moscow: Nauka, 1967) p. 175]
26. Pauli W *Theory of Relativity* (Oxford: Pergamon Press, 1967) Sect. 53 [Translated into Russian (Moscow: Nauka, 1983) § 53]
27. Möller C *The Theory of Relativity* (Oxford: Clarendon Press, 1960) §§ 92, 93 [Translated into Russian (Moscow: Atomizdat, 1975) § 8.12]
28. Landau L, Lifshitz E *Teoriya Polya* (Field Theory) (Moscow: Nauka, 1988) § 88 [Translated into English as *The Classical Theory of Fields* (Oxford: Pergamon Press, 1962) § 89]
29. Weinberg S *Gravitation and Cosmology* (New York: Wiley, 1972) p. 79 [Translated into Russian (Moscow: Mir, 1975) p. 95]
30. Schneider P, Ehlers J, Falco E E *Gravitational Lenses* (Berlin: Springer-Verlag, 1996) p. 93
31. Straumann N *General Relativity and Relativistic Astrophysics* (Berlin: Springer-Verlag, 1984) p. 96
32. Dirac P A M *General Theory of Relativity* (New York: Wiley, 1975) [Translated into Russian (Moscow: Atomizdat, 1978)]
33. Berkov A V, Kobzarev I Yu *Teoriya Tyagoteniya Eĭnshteĭna. Obshchie Printsipy i Eksperimental'nye Sledstviya* (Einstein's Theory of Gravitation) (Moscow: Izd. MIFI, 1989) p. 88
34. Ciufolini I, Wheeler J A *Gravitation and Inertia* (Princeton: Princeton University Press, 1995) p. 97
35. Feynman R *Lectures on Gravitation* (Ed. B Hatfield) (Redwood City, Calif.: Addison-Wesley Publ., 1995) Sect. 5.2
36. Weyl H *Raum, Zeit, Materie* 5te Auflage (Berlin: Verlag von Julius Springer, 1923) S. 322; see also 7th edition published by J Ehlers (Springer-Verlag, 1988)
37. Misner C, Thorne K, Wheeler J A *Gravitation* (San Francisco: Freeman and Company, 1973) §§ 7.2–7.5, 25.4, 38.5 [Translated into Russian (Moscow: Mir, 1977)]
38. Wald R W *General Relativity* (Chicago, London: University of Chicago Press, 1984) p. 137
39. Stephani H *General Relativity* 2nd ed. (Cambridge: Cambridge University Press, 1996) p. 117
40. Kenyon I R *General Relativity* (Oxford: Oxford University Press, 1996) p. 15
41. Born M *Einstein's Theory of Relativity* (New York: Dover Publications, 1962) Ch. VII, § 11 [Translated into Russian (Moscow: Mir, 1972) pp. 60, 342]
42. Sciama D W *The Physical Foundations of General Relativity* (New York: Doubleday and Co., 1969) Ch. 5 [Translated into Russian (Moscow: Mir, 1971)]
43. Zel'dovich Ya B, Novikov I D *Teoriya Tyagoteniya i Evolyutsiya Zvezd* (The Theory of Gravitation and Evolution of Stars) (Moscow: Nauka, 1971) Ch. 3, § 4 [Translated into English as *Relativistic Astrophysics* Vol. 1 *Stars and Relativity* (Chicago, London: University of Chicago Press, 1971) p. 89]
44. Braginskiĭ V B, Polnarev A G *Udivitel'naya Gravitatsiya* (*ili Kak Izmeryayut Kriviznu Mira*) (Wonderful Gravitation and How to Measure the World Curvature) (Moscow: Nauka, 1985) p. 29
45. Syunyaev R A (Ed.) *Fizika Kosmosa. Malen'kaya Entsiklopediya* (Physics of Cosmos. Small Encyclopedia) (Moscow: Bol'shaya Sovetskaya Entsiklopediya, 1986) p. 331
46. *The New Encyclopedia Britannica* Vol. 20, 15th ed. (1994) p. 174
47. Wheeler J A *A Journey into Gravity and Spacetime* (New York: Scientific American Library, 1990) pp. 48, 166
48. Taylor E, Wheeler J A *Spacetime Physics* 3rd ed. (New York: Freeman and Co., 1992) pp. 258, 272 [Translated into Russian 2nd ed. (Moscow: Mir, 1971) pp. 200, 201, 295, 296]
49. Layzer D *Constructing the Universe* Ch. 6 (New York: Scientific American Library, 1984) p. 210
50. Bowler M G *Gravitation and Relativity* (New York: Pergamon Press, 1976) [Translated into Russian (Moscow: Mir, 1979)]
51. Schwinger J *Einstein's Legacy* (New York: Scientific American Library, 1986) p. 141
52. Jain A K et al. *Phys. Rev. Lett.* **58** 1165 (1987)
53. Anandan J *Phys. Lett. A* **105** 280 (1984)
54. Ginzburg V L *Usp. Fiz. Nauk* **59** 11 (1956); **63** 119 (1957)
55. Ginzburg V L "Cosmic studies and the theory of relativity" *Astronautica Acta* **12** 136 (1966) [Translated into Russian in *Eĭnshteĭnovskiĭ Sbornik* (Moscow: Nauka, 1967) p. 80]
56. Brault J W *Bull. Am. Phys. Soc.* **8** 28 (1963)
57. Snider J L *Phys. Rev. Lett.* **28** 853 (1972)
58. Okun' L B *Usp. Fiz. Nauk* **158** 511 (1989) [*Sov. Phys. Usp.* **32** 629 (1989)]
59. Okun L B *Phys. Today* **42** (6) 31 (1989); **43** (5) 115 (1990)

On the interpretation of the redshift in a static gravitational field

L. B. Okun[a] and K. G. Selivanov[b]
ITEP, Moscow, 117218, Russia
V. L. Telegdi[c]
EP Division, CERN, CH-1211 Geneva 23, Switzerland

(Received 12 April 1999; accepted 27 July 1999)

The classical phenomenon of the redshift of light in a static gravitational potential, usually called the gravitational redshift, is described in the literature essentially in two ways: On the one hand, the phenomenon is explained through the behavior of clocks which run faster the higher they are located in the potential, whereas the energy and frequency of the propagating photon do not change with height. The light thus appears to be redshifted relative to the frequency of the clock. On the other hand, the phenomenon is alternatively discussed (even in some authoritative texts) in terms of an energy loss of a photon as it overcomes the gravitational attraction of the massive body. This second approach operates with notions such as the "gravitational mass" or the "potential energy" of a photon and we assert that it is misleading. We do not claim to present any original ideas or to give a comprehensive review of the subject, our goal being essentially a pedagogical one.

I. INTRODUCTION

There are two kinds of photon redshift known in the literature, a gravitational and a cosmological one. Though in the General Relativity framework the two shifts can be described very similarly, in the literature they are usually discussed separately. The cosmological redshift is that of light from distant galaxies which recede. It is generally referred to as the Hubble redshift. For the farthest observed galaxies it is quite large, $\Delta\lambda/\lambda \approx 5$. The gravitational redshift arises when light moves away from a static massive object, e.g., the earth or the sun. Its observed magnitudes are generally small. This paper is devoted exclusively to this type of redshift.

The gravitational redshift is a classical effect of Einstein's General Relativity (GR), one predicted by him[1] well before that theory was created[2] (for the historical background, see e.g., Ref. 3). Phenomenologically one can simply affirm that the frequency of light emitted by two identical atoms is smaller for the atom which sits deeper in the gravitational potential. A number of ingenious experiments have been performed[4–9] to measure various manifestations of this effect. They are discussed in a number of excellent reviews[10] whose main goal is to contrast the predictions of GR with those of various nonstandard theories of gravity. Explanations of the gravitational redshift *per se* within the standard framework are, however, not critically discussed in these reviews.

Most treatises on GR[11,12] follow the definitive reasoning of Einstein,[2] according to which the gravitational redshift is explained in terms of the universal property of standard clocks (atoms, nuclei). The proper time interval between events of emission of two photons as measured by the standard clock at the point of emission is different from the proper time interval between events of absorption of those photons as measured by the identical standard clock at the point of absorption (in this way it was first formulated in Ref. 13).

In the static gravitational potential the picture simplifies because there is a distinguished time—the one on which metric is independent. This time can be chosen as a universal (world) one. Under this choice the energy difference between two atomic levels increases with the distance of the atom from the earth while the energy of the propagating photon does not change. (In what follows we speak of the earth, but it could be any other massive body.) Thus what is called the redshift of the photon is actually a blueshift of the atom. As for the proper times at different points, they are related to the universal time via a multiplier which depends on gravitational potential and hence has different values at different points (see Sec. IV).

Actually, most modern textbooks and monographs[14] derive the redshift by using sophisticated general relativity calculations, e.g., using orthonormal bases (a sequence of proper reference frames) to define photon energy and parallel transporting the photon's 4-momentum along its world line. Sometimes this description is loosely phrased as a degradation of the photon's energy as it climbs out of the gravitational potential well. (Some other classical textbooks also use this loose phrasing.[15]) However, the nonexperts should be warned that the mathematics underlying this description is radically different from the heuristic (and wrong) arguments presented in many elementary texts, e.g., Ref. 16. These authors claim to deduce the "work against gravity" viewpoint by pretending that the photon is like a normal, low-velocity, massive particle and thus has a "photon mass" and "photon potential energy." Such derivations are incorrect and should be avoided. They are in fact avoided in exceptional popular texts, e.g., Ref. 17.

II. EXPERIMENTS

The first laboratory measurement of the gravitational redshift was performed at Harvard in 1960 by Robert Pound and Glenn Rebka[4,5] (with 10% accuracy) and in 1964 by Pound and Snider[6] (with 1% accuracy). The photons moved in a 22.5-m tall tower. The source and the absorber of the photons (γ rays of 14.4-keV energy) were ^{57}Fe nuclei. The experiment exploited the Mössbauer effect, which makes the photon lines in a crystal extremely monochromatic. The redshift was compensated through the Doppler effect, i.e., by slowly moving the absorber and thereby restoring the resonant absorption. The shift measured in this way[5] was $\Delta\omega/\omega \approx 10^{-15}$.

As to the interpretation of the result, there is some ambiguity in the papers by Pound *et al.* Although they mention the clock interpretation by referring to Einstein's original papers, the absolute reddening of the photon is also implied, as can be seen from the title of Pound's talk in Moscow,[5] "On the weight of photons." From the title of the paper by Pound and Snider,[6] "Effect of gravity on nuclear resonance," one might infer that they did not want to commit themselves to any interpretation.

By contrast, the majority of the reviews of gravitational experiments[10] quote the Harvard result as a test of the behavior of clocks. In fact the result must be interpreted as a relative shift of the photon frequency with respect to the nuclear one since the experiment does not measure these frequencies independently.

An experiment measuring the relative shift of a photon (radio-wave) frequency with respect to an atomic one was also performed with a rocket flying up to 10 000 km above the earth and landing in the ocean.[7]

Alongside the experiments described in Refs. 3–7, special measurements of the dependence of the atomic clock rate on the altitude were done directly by using airplanes[8,9] (see also the reviews in Ref. 10). In these experiments a clock which had spent many hours at high altitude was brought back to the laboratory and compared with its "earthly twin." The latter, once corrected for various background effects, lagged behind by $\Delta T = (gh/c^2)T$, where T is the duration of the flight at height h, g the gravitational acceleration, and c the speed of light.

One of these background effects is the famous "twin paradox" of Special Relativity, which stems from the fact that moving clocks run slower than clocks at rest. It is easy to derive a general formula which includes both the gravitational potential ϕ and the velocity u (see, e.g., the book by C. Møller[11]):

$$d\tau = dt[1 + 2\phi/c^2 - u^2/c^2]^{1/2}, \tag{1}$$

where τ is the proper "physical" time of the clock, while t is the so-called world time, which can be introduced in the case of a static potential and which is sometimes called laboratory time.

In his lectures on gravity,[12] Richard Feynman stresses the differences between the effects due to u and to ϕ. He concludes that the center of the earth must be by a day or two younger than its surface.

III. THEORY BEFORE GENERAL RELATIVITY: "ELEVATORS"

Since most of the conclusive experiments on the gravitational redshift were earthbound, we shall throughout use that frame in which the earth is at rest (neglecting its rotation).

As is well known, a potential is defined up to a constant. When considering the gravitational potential $\phi(r)$ at an arbitrary distance r from the earth's center, it is convenient to set $\phi(\infty) = 0$; then ϕ is negative everywhere.

Near the earth's surface (at $h = r - R \ll R$) it is legitimate to approximate ϕ linearly:

$$\delta\phi(h) = \phi(R+h) - \phi(R) = gh, \tag{2}$$

where g is the usual gravitational acceleration. Note that $\delta\phi(h)$ is positive for $h \neq 0$. We shall discuss the redshift only to the first order in the parameter gh/c^2.

[The linear approximation Eq. (2) is valid for the description of experiments 8 and 9. It is obvious, however, that for the high-flying rockets[7] ($h \approx 10^4$ km) it is not adequate and the Newtonian potential should be used, but this is not essential for the dilemma "clocks versus photons" which is the subject of this paper.]

Einstein's first papers[1] on the gravitational redshift contain many of the basic ideas on the subject, which were incorporated (sometimes without proper critical analysis) into numerous subsequent texts. He considered the Doppler effect in the freely falling frame and found the increase of the frequency of an atom (clock) with increasing height (potential). The cornerstone of his considerations was the local equivalence between the behavior of a physical system in a gravitational field and in a properly accelerated reference frame.

For the potential (2) it is particularly convenient to appeal to Einstein's freely falling reference frame ("elevator"). In such an elevator an observer cannot detect any manifestation of gravity by any *strictly local* experiment (equivalence principle). Operationally, "strictly local" means that the device used is sufficiently small so as not to be sensitive to curvature effects.

Consider from such an elevator falling with the acceleration g a photon of frequency ω which is emitted upwards by an atom at rest on the surface of the earth and which is expected to be absorbed by an identical atom fixed at height h. The frequency of light is not affected by any gravitational field in a freely falling elevator: it keeps the frequency with which it was emitted. Assume that at the moment of emission ($t=0$) the elevator had zero velocity. At the time $t = h/c$, when the photon reaches the absorbing atom, the latter will have velocity $v = gh/c$ directed upwards in the elevator frame. As a result the frequency of the photon, as seen by the absorbing atom, will be shifted by the linear Doppler effect by $v/c = gh/c^2$ toward the red, that is

$$\frac{\Delta\omega}{\omega} = -\frac{gh}{c^2}. \tag{3}$$

(Minute corrections of higher order in gh/c^2 to the "elevator formulas" are lucidly discussed on the basis of a metric approach in Ref. 18.) Consider now another situation, when the upper atom (absorber) moves in the laboratory frame with a velocity $v = gh/c$ downwards. Then in the elevator frame it will have zero velocity at the moment of absorption and hence it will be able to absorb the photon resonantly in complete agreement with experiments.[4,5] Obviously, in the elevator frame there is no room for the interpretation of the redshift in terms of a photon losing its energy as it climbs out of the gravitational well.

IV. THEORY IN THE FRAMEWORK OF GENERAL RELATIVITY: METRIC

Up to now we have used only Special Relativity and Newtonian gravity. As is well known, a consistent relativistic description of classical gravity is given in the framework of GR with its curved space–time metric. One introduces a metric tensor, $g_{ik}(x)$, $i,k = 0,1,2,3$, which is, in general, coordinate dependent and transforms by definition under a change of coordinates in such a way that the interval ds between two events with coordinates x^i and $x^i + dx^i$,

$$ds^2 = g_{ik}(x) dx^i dx^k, \tag{4}$$

is invariant. Setting $dx^1 = dx^2 = dx^3 = 0$, one obtains the relation between the proper time interval $d\tau = ds/c$ and the world time interval $dt = dx^0/c$ for an observer at rest,

$$d\tau = \sqrt{g_{00}}\, dt. \tag{5}$$

For a static case, Eq. (5) integrates to

$$\tau = \sqrt{g_{00}}\, t, \tag{6}$$

where g_{00} is a function of x in the general case while in the case of Eq. (2) it is a function of $x^3 = z = h$.

The time τ is displayed by a standard clock and can also be viewed as a time coordinate in the so-called comoving locally inertial frame, i.e., the locally inertial frame which at a given instant has zero velocity with respect to the laboratory frame. If one has a set of standard clocks at different points, then their proper times τ are differently related to the world (laboratory) time t, due to x dependence of g_{00} [see Eq. (6)]. This explains the airplane experiments.[8,9] Let us recall that sometimes the world time is called laboratory time. The former term reflects the fact that it is the same for the whole world; the latter signifies that it can be set with standard clocks in the laboratory. Many authors refer to t as the coordinate time.

A weak gravitational field can be described by a gravitational potential ϕ, and g_{00} is related to the gravitational potential:

$$g_{00} = 1 + 2\phi/c^2. \tag{7}$$

We shall explain the meaning of this relation a bit later [see Eqs. (8)–(10)]. According to Eqs. (6) and (7) a clock runs slower in the laboratory the deeper it sits in the gravitational potential.

Analogous to Eq. (5) is a relation between the rest energy E_0 of a body in the laboratory frame and in the comoving locally inertial frame,

$$E_0^{\text{lab}} = E_0^{\text{loc}} \sqrt{g_{00}} \tag{8}$$

(notice that $E_0^{\text{lab}} dt = E_0^{\text{loc}} d\tau$; this is because the energy E is the zeroth component of a covariant 4-vector, while dt is the zeroth component of a contravariant 4-vector).

The rest energy in the locally inertial frame is the same as in special relativity (see, e.g., Ref. 19 and the book by E. Taylor and J. A. Wheeler,[16] p. 246),

$$E_0^{\text{loc}} = mc^2, \tag{9}$$

while the rest energy in the laboratory system E_0^{lab} also includes the potential energy of the body in the gravitational field. This is in accordance with the main principle of general relativity: All effects of gravity arise only via the metric tensor. Equation (7) for the g_{00} component of the metric tensor in a weak gravitational field can be considered as a consequence of Eqs. (8) and (9) and of the relation

$$E_0^{\text{lab}} = mc^2 + m\phi, \tag{10}$$

which generalizes the notion of the rest energy of a free particle to that of a particle of mass m in a gravitational weak potential.

Now we are in a position to explain the redshift in the laboratory frame. According to Eq. (8) or Eq. (10), the energy difference ϵ_{lab} of atomic or nuclear levels in that frame depends on the location of the atom. The deeper the atom sits in the gravitational potential, the smaller ϵ_{lab} is. For an absorber atom which is located at height h above an identical atom which emits the photon, the relative change in the energy difference is gh/c^2,

$$\frac{\Delta \epsilon^{\text{lab}}}{\epsilon^{\text{lab}}} = \frac{gh}{c^2}. \tag{11}$$

[We use in Eq. (11) the linear approximation of Eq. (2).] One can say that the energy levels of the absorber atoms are shifted toward the blue in the laboratory frame. Equation (11) is, of course, nothing but a way to describe the difference in the rates of atomic clocks located at a height h one above the other. On the other hand, the energy (frequency) of the photon is conserved as it propagates in a static gravitational field. This can, for example, be seen from the wave equation of the electromagnetic field in the presence of a static gravitational potential or from the equations of motion of a massless (or massive) particle in a static metric. Clearly, in the laboratory system there is no room for the interpretation in which the photon loses its energy when working against the gravitational field.

Finally, one can discuss the experiment using a sequence of locally inertial frames which are comoving with the laboratory clocks (atoms) at the instants when the photon passes them. As we explained above, in such systems the standard clocks run with the same rates, the rest energy of the atom is equal to its mass, times c^2, Eq. (9), and the energy levels of the atom are the same as at infinity. On the other hand, the energy of the photon in the laboratory frame $E_\gamma^{\text{lab}} = \hbar\omega^{\text{lab}}$ and in the comoving locally inertial frame E_γ^{loc} is related as

$$E_\gamma^{\text{lab}} = E_\gamma^{\text{loc}} \sqrt{g_{00}}\,. \tag{12}$$

Equation (12) follows from Eq. (8) by noticing that the photon can be absorbed by a massive body and by considering the increase of the rest energy of that body. Thus, since E_γ^{lab} is conserved, E_γ^{loc} decreases with height:

$$\frac{\omega^{\text{loc}}(h) - \omega^{\text{loc}}(0)}{\omega^{\text{loc}}(0)} = \frac{E_\gamma^{\text{loc}}(h) - E_\gamma^{\text{loc}}(0)}{E_\gamma^{\text{loc}}(0)} = -\frac{gh}{c^2} \tag{13}$$

and this is the observed redshift of the photon. But the decrease in E^{loc} is not because the photon works against the gravitational field. The gravitational field is absent in any locally inertial frame. E_γ^{loc} changes because one passes from one of the locally inertial frames to the other: one comoving with the laboratory at the moment of emission, the other—at the moment of absorption.

V. PSEUDODERIVATION AND MISINTERPRETATION OF GRAVITATIONAL REDSHIFT

The simplest (albeit wrong) explanation of the redshift is based on ascribing to the photon both an inertial and a gravitational "mass," $m_\gamma = E_\gamma/c^2$. Thereby a photon is attracted to the earth with a force gm_γ, while the fractional decrease of its energy (frequency) at height h is

$$\frac{\Delta E_\gamma}{E_\gamma}=\frac{\Delta\omega}{\omega}=-gm_\gamma h/m_\gamma c^2=-gh/c^2. \quad (14)$$

Note that (up to a sign) this is exactly the formula for the blueshift of an atomic level. That should not surprise. The atom and the photon are treated here on the same footing, i.e., both nonrelativistically! This is of course inappropriate for the photon. If the explanation in terms of gravitational attraction of the photon to the earth were also correct, then one would be forced to expect a doubling of the redshift (the sum of the effects on the clock and on the photon) in the Pound-type experiments.

Some readers may invoke Einstein's authority to contradict what was said above. In his 1911 paper,[1] he advanced the idea that energy is not only a source of inertia, but also a source of gravity. Loosely speaking, he used the heuristic argument "whenever there is mass, there is also energy and vice versa." As he realized later, this "vice versa" was not as correct as the direct statement (a photon has energy while its mass is zero). By applying this energy–mass argument, he calculated the energy loss of a photon moving upwards in the potential of the earth as previously discussed. With the same heuristic principle he also derived an expression for the deflection of light by the sun which, however, underestimated the deflection angle by a factor of 2. Subsequently, in the framework of GR, Einstein recovered this factor.[2] The correct formula was confirmed by observation.

VI. THE WAVELENGTH MEASUREMENT

Up to now we have discussed the gravitational redshift in terms of the photon frequency and clocks. Let us now consider the same phenomenon in terms of the photon wavelength and gratings. To do this we consider two identical gratings at different heights, inclined with respect to the z axis along which the light propagates between them. We do not go into the details of this gedanken experiment. The z projections of grating spacing are used only as a standard of length. The lower grating serves as monochromator, i.e., as the light source. The wavelength of the photon $\lambda^{\text{lab}}(z)$ corresponds to its frequency, while the spacings of the grating in the vertical (z) direction, $l^{\text{lab}}(z)$, correspond to the rates of the clocks.

For the sake of simplicity, one may consider a very small angle of incidence on the gratings, i.e., the grazing incidence of the light. In that case, the vertical projection of the spacing is practically the spacing itself. (Recall that, for grazing incidence, the spacing l^{lab} must be of the same order as the wavelength λ^{lab}.)

While the photon energy E^{lab} is conserved in a static gravitational field, the photon momentum p^{lab} is not. The relation between these quantities is given by the condition that the photon remains massless, which in a gravitational field reads

$$g^{ij}p_ip_j=0, \quad (15)$$

where the g^{ij}, $i,j=0,...,3$ are the contravariant components of the metric tensor, p_j are the components of the four momentum, $p_0=E^{\text{lab}}$, $p_3=p^{\text{lab}}=2\pi\hbar/\lambda^{\text{lab}}(z)$ (for the photon moving along the z axis). For the cases we are discussing the metric g^{ij} can be taken in diagonal form, in particular $g^{33}=1/g_{33}$.

From Eq. (15) one readily finds how $\lambda^{\text{lab}}(z)$ changes with height:

$$\lambda_{\text{lab}}(z)=\sqrt{g^{33}(z)/g^{33}(0)}\sqrt{g^{00}(0)/g^{00}(z)}\;\lambda_{\text{lab}}(0). \quad (16)$$

On the other hand, the grating spacing in the z direction, $l^{\text{lab}}(z)$, also changes with height. This is just the standard change of scale in the gravitational field, explained, e.g., in the book by L. Landau and E. Lifshitz, Sec. 84:[11]

$$l^{\text{lab}}(z)=\sqrt{-g^{33}(z)}\;l^0, \quad (17)$$

where l^0 is the "proper spacing" in the z direction, the counterpart of the proper period of the standard clock. Thus the spacing $l^{\text{lab}}(z)$ depends on z as follows:

$$l^{\text{lab}}(z)=\sqrt{g^{33}(z)/g^{33}(0)}\;l^{\text{lab}}(0). \quad (18)$$

Finally, in the wavelength analogue of the Pound *et al.* experiment with $z=h$ one would measure the double ratio $(\lambda(h)/l(h))/(\lambda(0)/l(0))$. The result can be presented in the form

$$\Delta\lambda^{\text{lab}}/\lambda^{\text{lab}}-\Delta l^{\text{lab}}/l^{\text{lab}}=\sqrt{g^{00}(0)/g^{00}(h)}=gh/c^2, \quad (19)$$

where $\Delta\lambda^{\text{lab}}/\lambda^{\text{lab}}=[\lambda^{\text{lab}}(h)-\lambda^{\text{lab}}(0)]/\lambda^{\text{lab}}(0)$ and analogously for $\Delta l^{\text{lab}}/l^{\text{lab}}$. Notice that g^{33} drops out from the result. This should be so because there is a freedom in the choice of z scale, and the observed quantities cannot depend on this choice. Equation (19) is analogous to the one describing the Pound *et al.* experiments:

$$\Delta\omega/\omega-\Delta\epsilon/\epsilon=\sqrt{g_{00}(0)/g_{00}(h)}=-gh/c^2, \quad (20)$$

where ω is the frequency of the photon and $\epsilon/\hbar$ is the frequency of the clock [see Eq. (11)]. A word of explanation should be added about Eq. (20). In the laboratory frame the first term on the left-hand side is equal to zero,

$$\Delta\omega^{\text{lab}}/\omega^{\text{lab}}=0, \quad (21)$$

as discussed in Sec. III; thus only the second term given by Eq. (11) contributes.

We would, however, like to stress an important difference as compared to the case of frequency. There one can independently measure the difference in the rates of the upper and lower clocks, $\Delta\epsilon^{\text{lab}}/\epsilon^{\text{lab}}$ [analogue of the second term on the left-hand side of Eq. (19)], and that was done in the airplane experiments. Here the change of the scale, $\Delta l^{\text{lab}}/l^{\text{lab}}$, cannot be measured independently. This important difference comes from the fact that the metric is static, though it is z dependent.

One has to realize that such a laboratory experiment with gratings cannot be performed at the present state of the art in experimental physics (recall the importance of the Mössbauer effect in the experiments of Pound *et al.*). However, for the measurement of a large value of the redshift, e.g., that of the sodium spectral line from the sun, it is feasible. Such a grating experiment was performed by J. W. Brault in 1962 and was described in Sec. 38.5 in the monograph by C. Misner, K. Thorne, and J. A. Wheeler.[14] In this experiment the wavelength of the emitted light was fixed not by the lower grating, but by the atom on the sun surface.

VII. CONCLUSIONS

The present article contains little original material; it is primarily pedagogical. The gravitational redshift being, both theoretically and experimentally, one of the cornerstones of

General Relativity, it is very important that it always be taught in a simple but nevertheless correct way. That way centers on the universal modification of the rate of a clock exposed to a gravitational potential. An alternative explanation in terms of a (presumed) gravitational mass of a light pulse—and its (presumed) potential energy—is incorrect and misleading. We exhibit its fallacy, and schematically discuss redshift experiments in the framework of the correct approach. We want to stress those experiments in which an atomic clock was flown to, and kept at, high altitude and subsequently compared with its twin that never left the ground. The traveller clock was found to run ahead of its earthbound twin. The blueshift of clocks with height has thus been exhibited as an absolute phenomenon. One sees once again that the explanation of the gravitational redshift in terms of a naive "attraction of the photon by the earth" is wrong.

ACKNOWLEDGMENTS

We would like to thank V. V. Okorokov, who asked the question on compatibility in the framework of General Relativity of experiments by Pound *et al.*, and the airplane experiments. We also thank S. I. Blinnikov, A. D. Dolgov, A. Yu. Morozov, N. Straumann, K. Thorne, and G. Veneziano for very interesting discussions. We would like to express our gratitude especially to E. L. Schucking for his help in substantially improving our bibliography and for his insistence on a unified invariant approach to both gravitational and cosmological redshifts based on Killing vectors. We did not, however, follow his advice, wanting to focus the general reader's attention on the fallacy of the widespread naive interpretation of the gravitational redshift. Last but not least we want to thank J. A. Wheeler for his encouragement. One of us (LO) would like to thank the Theoretical Physics Division of CERN, where part of this work was done, for their hospitality.

[a]Electronic mail: okun@heron.itep.ru
[b]Electronic mail: selivano@heron.itep.ru
[c]Electronic mail: valentine.telegdi@cern.ch

[1]A. Einstein, "Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen," Jahrb. f. Radioaktivität u. Elektronik **4**, 411–462 (1907); "Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes," Ann. Phys. (Leipzig) **35**, 898–908 (1911).
[2]A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," Ann. Phys. (Leipzig) **49**, 769–822, (1916), Sec. 22; *The Meaning of Relativity* (Princeton U.P., New York, 1921), Eq. (106).
[3]A. Pais, *'Subtle is the Lord...', The Science and the Life of Albert Einstein* (Oxford U.P., Oxford, 1982), Chap. 9.
[4]R. V. Pound and G. A. Rebka, "Apparent weight of photons," Phys. Rev. Lett. **4**, 337–341 (1960); "Variation with temperature of the energy of recoil-free gamma rays from solids," **4**, 274–275 (1960); "Gravitational redshift in nuclear resonance," **3**, 439–441 (1959).
[5]R. Pound, Usp. Fiz. Nauk **72**, 673–683 (1960); "On the weight of photons," Sov. Phys. Usp. **3**, 875–883 (1961) (English translation).
[6]R. V. Pound and J. L. Snider, "Effect of gravity on gamma radiation," Phys. Rev. B **140**, 788–803 (1965); "Effect of gravity on nuclear resonance," Phys. Rev. Lett. **13**, 539–540 (1964).
[7]R. Vessot and M. Levine, "A test of the equivalence principle using a space-borne clock," Gen. Relativ. Gravit. **10**, 181–204 (1979).
[8]J. Haefele and R. Keating, "Around the world atomic clocks: Predicted relativistic gains," Science **177**, 166–167 (1972); "Around the world relativistic clocks: Observed relativistic time gains," **177**, 168–170 (1972).
[9]C. Alley *et al.*, in *Experimental Gravitation*, Proceedings of the Conference at Pavia (September 1976), edited by B. Bertotti (Academic, New York, 1977).
[10]C. Will, *Theory and Experiment in Gravitational Physics* (Cambridge U.P., Cambridge, 1981), Sec. 2.4; "The confrontation between general relativity and experiment: A 1992 upgrade," Int. J. Mod. Phys. D **1**, 13–68 (1992); preprint WUGRAV-95-5, gr-qc/9504017 1995; I. Shapiro, *Gravitazione Quanti e Relatività. Astrofisica e Cosmologia* (Giunti Barbera, Florence, 1979), Sec. 3; N. Ashby and J. Spilker, *The Global Positioning System, Theory and Applications*, Vol. 1 (American Institute of Aeronautics and Astronautics, Washington, DC, 1995), Chap. 18; J. Taylor, *Astronomical and Space Experiments to Test Relativity*, in General Relativity and Gravitation (Cambridge U.P., Cambridge, 1987), p. 214.
[11]W. Pauli, *Theory of Relativity* (Pergamon, Oxford, 1967), Sec. 53; C. Møller, *The Theory of Relativity* (Clarendon, Oxford, 1960), Secs. 92, 93; L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1962), Sec. 89; S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), pp. 79–85; P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (Springer, Berlin, 1996), pp. 93–97; N. Straumann, *General Relativity and Relativistic Astrophysics* (Springer-Verlag, Berlin, 1984), pp. 96–99; I. Ciufolini and J. A. Wheeler, *Gravitation and Inertia* (Princeton U.P., Princeton, 1995), pp. 97–108.
[12]R. P. Feynman, *Lectures on Gravitation*, edited by Brian Hatfield (Addison–Wesley, Reading, MA, 1995), Sec. 5.2.
[13]H. Weyl, *Raum, Zeit, Materie* (Verlag von Julius Springer, Berlin, 1923), 5te Auflage, p. 322; see also the seventh edition republished by J. Ehlers (Springer-Verlag, Berlin, 1988).
[14]C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Secs. 7.2–7.5, 25.3, 25.4, 38.5; R. W. Wald, *General Relativity* (University of Chicago Press, Chicago and London, 1984), p. 137; H. Stephani, *General Relativity* (Cambridge U.P., Cambridge, 1996), 2nd ed.; I. R. Kenyon, *General Relativity* (Oxford U.P., Oxford, 1996), pp. 15–21.
[15]Max Born, *Einstein's Theory of Relativity* (Dover, New York, 1962), Chap. VII, Sec. 11; D. W. Sciama, *The Physical Foundations of General Relativity* (Doubleday, Garden City, NY, 1969), Chap. 5; Ya. B. Zeldovich and I. D. Novikov, *Stars and Relativity*, Relativistic Astrophysics Vol. I (University of Chicago Press, Chicago and London, 1971), pp. 89–93.
[16]*The New Encyclopaedia Britannica* (Encyclopaedia Britannica, Inc., Chicago), 15th Ed., Vol. 20, p. 174; J. A. Wheeler, *A Journey into Gravity and Spacetime* (Scientific American Library, New York, 1990), pp. 48–53, 166–168; E. Taylor and J. A. Wheeler, *Spacetime Physics* (Freeman, New York, 1992), 3rd ed., pp. 258, 272; D. Layzer, *Constructing the Universe* (Scientific American Library, New York, 1984), Chap. 6, pp. 210–212.
[17]J. Schwinger, *Einstein's Legacy* (Scientific American Library, New York, 1986), pp. 141–153.
[18]E. A. Desloge, "Nonequivalence of a uniformly accelerating reference frame and a frame at rest in a uniform gravitational field," Am. J. Phys. **57**, 1121–1125 (1989); "The gravitational redshift in a uniform field," **58**, 856–858 (1990).
[19]L. B. Okun, "The concept of mass," Phys. Today **42**, 31–36 (June 1989); "Putting to rest mass misconceptions," **43**, 13, 15, 115, 117 (May 1990); Usp. Fiz. Nauk **158**, 511–530, (1989); "The concept of mass (mass, energy, relativity)," Sov. Phys. Usp. **32**, 629–638 (1989); "Note on the meaning and terminology of special relativity," Eur. J. Phys. **15**, 403–406 (1998).

This page intentionally left blank

Modern Physics Letters A, Vol. 15, No. 31 (2000) 1941–1947

PHOTONS AND STATIC GRAVITY

L. B. OKUN*
ITEP, Moscow, 117218, Russia

Received 12 October 2000

The influence of static gravitational field on frequency, wavelength and velocity of photons and on the energy levels of atoms and nuclei is considered in the most elementary way. The interconnection between these phenomena is stressed.

1. Introduction

The behavior of light in a static gravitational field has a long history. It has been discussed in many monographs on General Relativity and in vast popular literature. (An extensive list of references may be found in recent publications.[1,2])

In this letter we discuss from a single point of view four effects which are usually discussed separately in the literature:

(a) the gravitational redshift of photons flying away from a gravitating body,
(b) the increase of energy difference between levels in atoms with the increase of their distance from gravitating body,
(c) the retardation of radar echo from planets,
(d) the deflection of star light observed during a solar eclipse.

The relation between effects (a) and (b) was thoroughly discussed in Refs. 1 and 2. This letter may be considered as a continuation of those articles.

2. Redshift and Clocks

The phenomenon of gravitational redshift is widely explained by using (explicitly or implicitly) a presumed analogy between a photon and a stone: when moving away from a gravitating body (sun or earth) both a photon and a stone are supposed to lose energy overcoming the gravitational attraction.

This explanation, as was stressed recently,[1,2] is wrong. The energy of a photon E and hence its frequency $\omega = E/\hbar$ do not depend on the distance from the gravitating body, because in the static case the gravitational potential does not depend on the

*E-mail: okun@heron.itep.ru

time coordinate t. The reader who is not satisfied with this argument may look at Maxwell's equations as given e.g. in Sec. 5.2 of Ref. 3. These equations with time-independent metric have solutions with frequencies equal to those of the emitter.

In what follows we will often use both the quantum notion of photon and the classical notions of electromagnetic wave or even ray of light. To a certain extent this is caused by the physical nature of experiments in which the effects we are discussing were observed. For instance, in the famous experiments by Pound *et al.* the γ-quanta were flying vertically in a tower. The language of quantum physics is necessary to discuss the behavior of nuclear and atomic levels, and hence of atomic clocks. On the other hand, in the discussion of the radar echo and deflection of light quantum effects are absolutely not essential.

The proper explanation of gravitational redshift lies in the behavior of clocks (atoms, nuclei). The rest energy E_0 of any massive object increases with increase of the distance from the gravitating body because of the increase of the potential ϕ:

$$E_0 = mc^2(1 + \phi/c^2)\,. \tag{1}$$

The gravitational potential is

$$\phi = -GM/r\,, \tag{2}$$

where $G = 6.67 \times 10^{-8}\,\mathrm{cm}^3/\mathrm{g}\cdot\mathrm{s}^2$ is the Newton constant, M is the mass of the body, and r the distance from its center; $M = 2 \times 10^{33}\,\mathrm{g}$ for the sun, while for the earth $M = 6 \times 10^{27}\,\mathrm{g}$.

The universal increase of rest energy of an atom in an excited and the ground state means that the energy difference between levels also universally increases by the factor $(1 + \phi/c^2)$. As a result of this increase the energy of a photon emitted in a transition of an atom downstairs is not enough to excite a reverse transition upstairs. For the observer upstairs this looks like a redshift of the photon.

A "semi-competent" observer who knows nothing about the dependence of the rate of the standard clocks (atoms, nuclei) on the value of gravitational potential could conclude that photon is redshifted not relatively, but absolutely. A competent observer knows about special experiments on airplanes which proved that, in accord with General Relativity, (atomic) clocks run faster high above the earth (see e.g. Refs. 1 and 2). Therefore for a competent observer the apparent redshift of the photon is a result of the blueshift of the clock.

A naive (but obviously wrong!) way to derive the formula for the redshift is to ascribe to the photon with energy E a mass $m_\gamma = E/c^2$ and to apply to the photon a nonrelativistic formula $\Delta E = -m_\gamma \Delta\phi$ treating it like a stone. Then the relative shift of photon energy is $\Delta E/E = -\Delta\phi/c^2$, which coincides with the correct result. But this coincidence cannot justify the absolutely thoughtless application of a nonrelativistic formula to an ultrarelativistic object.

It is interesting that one can still speak about the redshift of the photon if one considers not its frequency, but its wavelength or momentum.[a] To see this let us consider the condition that photon is massless:

$$g^{ij}p_i p_j = 0\,, \tag{3}$$

where g^{ij}, $i,j = 0,1,2,3$ are contravariant components of the metric tensor. In a spherically symmetric potential when photon moves along a radius we have

$$g^{00}p_0p_0 - g^{rr}p_rp_r = 0\,. \tag{4}$$

In the case of standard Schwarzschild metric[3–5]:

$$g^{00} = \left(1 - \frac{r_g}{r}\right)^{-1}, \tag{5}$$

$$g^{rr} = \left(1 - \frac{r_g}{r}\right), \tag{6}$$

where

$$r_g = 2MG/c^2\,. \tag{7}$$

For the sun $r_g = 3$ km.

We identify the covariant vector p_i (and not contravariant vector p^i) with energy–momentum four-vector because

$$p_i = \partial S/\partial x^i\,, \tag{8}$$

where S is the action. This guarantees that if time is uniform (g^{ij} is time-independent), then energy of a particle is conserved.

As the energy of the photon $E = cp_0$ does not depend on r, we immediately see from Eq. (4) that its momentum p_r does depend:

$$p_r = \frac{E}{c}\left(1 - \frac{r_g}{r}\right)^{-1}. \tag{9}$$

The closer the photon to the sun, the larger its momentum. Hence its wavelength is blueshifted:

$$\lambda = \frac{2\pi\hbar}{p_r} = \frac{2\pi\hbar c}{E}\left(1 - \frac{r_g}{r}\right). \tag{10}$$

In that sense (but only in that!) the widely used words about redshift and blueshift of the photon are correct.

[a]It is instructive to compare the behavior of λ in the spherically symmetric case with that in a linear potential near the earth's surface considered in detail in Refs. 1 and 2.

3. Photon Velocity

With ω being constant and λ decreasing with radius, the velocity of the photon

$$v = \frac{\lambda\omega}{2\pi} = c\left(1 - \frac{r_g}{r}\right) \tag{11}$$

also decreases with decreasing radius r. In that respect photon drastically differs from a nonrelativistic stone, falling into the sun. (The velocity of the stone obviously increases with the decrease of r.)

The same conclusion could be reached by an explicit calculation of velocity from the expression for the interval:

$$ds^2 = g_{00}c^2\, dt^2 - g_{rr}\, dr^2 = 0\,, \tag{12}$$

where

$$g_{00} = 1/g^{00} = (1 - r_g/r)\,, \tag{13}$$

$$g_{rr} = 1/g^{rr} = (1 - r_g/r)^{-1}\,, \tag{14}$$

$$v = dr/dt = c\left(\frac{g_{00}}{g_{rr}}\right)^{1/2} = c(1 - r_g/r)\,. \tag{15}$$

[Note that if we defined (incorrectly!) v as in Special Relativity, $v = pc^2/E = cp_r/p_0$, we would get a different behavior of velocity: $v = c(1-r_g/r)^{-1}$, because momentum increases both for the photon and for the stone.]

The time t is often called the coordinate time, or world time, as it is set by clocks infinitely far from the gravitating body. Therefore it is proper to call v the coordinate or world velocity. It is worth noting that while the velocity v changes with the potential, the velocity of light in a locally inertial reference frame is always equal to c.

The decrease of coordinate velocity v discussed above leads to two types of effects both of which are well known. One of them is the famous deflection of light by the sun, predicted by Einstein and first observed during 1919 solar eclipse. This observation was reported by newspapers worldwide and started the cult of Einstein. The other effect is not so widely known. It is the delay of radar echo from inner planets (Venus, Mercury), the measurement of which was proposed[6] and realized by I. I. Shapiro. (For the description of the measurements see Sec. 8.7 of Ref. 3.)

It is appropriate to discuss both effects (the echo and the deflection) in the framework of geometrical optics (see Ref. 5), as the wavelength λ is negligible compared to any other characteristic length. The refraction index n is given by:

$$n = \frac{c}{v} = (1 - r_g/r)^{-1} = (1 + r_g/r)\,. \tag{16}$$

Here and in what follows we consistently neglect terms of the order of $(r_g/r)^2$: the gravitational field in the solar system is weak, and therefore

$$n - 1 = r_g/r \ll 1\,. \tag{17}$$

4. Radar Echo

Let us start with radar echo and consider for simplicity the echo from the sun. (This is of course only a gedanken experiment.) The duration of a two-way travel of a signal in this case is

$$t = 2\int_{R_s}^{r_e} \frac{dr}{v(r)} = \frac{2}{c}\int_{R_s}^{r_e} n(r)dr = \frac{2}{c}\int_{R_s}^{r_e}(1 + r_g/r)dr \tag{18}$$

and the delay compared to the case $v = c$ is

$$\Delta t = 2\frac{r_g}{c}\ln\frac{r_e}{R_s}, \tag{19}$$

where $r_e = 150 \times 10^6$ km is the distance between the sun and the earth, while $R_s = 0.7 \times 10^6$ km is the radius of the sun.

Consider now the echo from a planet close to its superior conjunction (when the planet is at the largest distance from the earth, so that the radar ray almost grazes the sun). In that case, unlike the previous ones, we have a non-radial trajectory with a nonvanishing impact parameter ρ. Let us denote by z the coordinate along the straight line connecting the earth at (z_e, ρ) and the planet at (z_p, ρ). (The axis ρ is obviously orthogonal to z.) Note that we can use straight lines even for non-radial trajectories because the gravitational field in solar system is weak: $(n-1) = r_g/r \ll 1$. The deflection of light (which is considered in the next section) gives a negligible contribution of order $(n-1)^2$ to the echo delay time. To calculate this delay time we use isotropic metric:

$$\Delta t = \frac{2}{c}\int_{z_e}^{z_p} dz\left(\frac{c}{v} - 1\right) = \frac{2}{c}\int_{z_e}^{z_p} dz(r_g/r) = 2\frac{r_g}{c}\int_{z_e}^{z_p}\frac{dz}{\sqrt{z^2+\rho^2}}$$

$$= 2\frac{r_g}{c}\left(\ln\frac{|z_p| + r_p}{\rho} + \ln\frac{|z_e| + r_e}{\rho}\right) = 2\frac{r_g}{c}\ln\frac{4r_p r_e}{\rho^2}. \tag{20}$$

The effect is maximal when $\rho \simeq R_s$. The logarithmic dependence for the echo from the sun and from a planet is the same. The non-logarithmic terms ($\sim r_g/r_e$) are neglected here: we do not take into account the retardation of clocks on the earth with respect to the world time. Their inclusion would increase Δt by $\sim 9\%$. For Mercury $r_p = 58 \times 10^6$ km, hence $\Delta t \simeq 240$ μs.

It is interesting to note that retardation must take place not only for photons but also for any extreme relativistic particle if its mass m and energy E are such that

$$(mc^2)^2/2E^2 \ll r_g/r_e. \tag{21}$$

For the sun–earth system $r_g/r_e \sim 2 \times 10^{-8}$, while for electrons from LEP I $(mc^2/E)^2 \sim 10^{-10}$. It is easy to see that even for low energy solar and reactor neutrinos, the inequality is also fulfilled. Unfortunately echo experiments with electrons and neutrinos are not realistic for many reasons.

5. Deflection of Light

Consider a light ray connecting a star and the earth. In this case the trajectory of light is bent, but we will continue to denote the variable along this trajectory as z (its end points z_* and z_e, respectively) and the impact parameter as ρ. Assume that we deal with a "thick pencil" ray of diameter $d\rho$. It is easy to see from Eq. (15) that the inward side of the ray moves slower than the outward one. (Here inward means closer to the sun.) As a result the cross-section of the ray is tilted: When its inward edge propagates by dz the outer edge propagates by $dz[1 + (\partial v / c\partial\rho)d\rho]$. Thus the angle of the tilt is

$$d\alpha = -dz(\partial v / c\partial\rho) = \frac{dz\,\partial(r_g/r)}{\partial\rho}\,. \tag{22}$$

Hence the direction of the ray would be inward deflected, the differential deflection angle of the ray being given by Eq. (22), while the total deflection angle

$$\alpha = \int_{z_*}^{z_e} dz \frac{\partial(r_g/r)}{\partial\rho} = \frac{\partial}{\partial\rho}\int_{z_*}^{z_e} dz(r_g/r)\,. \tag{23}$$

If we now substitute z_* by z_p which is not essential, because both of them are much larger than ρ and will not enter the result, we get from Eq. (20):

$$\alpha = \frac{\partial}{\partial\rho}\left(\frac{c\Delta t}{2}\right) = -\frac{2r_g}{\rho}\,. \tag{24}$$

As was stressed in Ref. 5, the first part of Eq. (24) is a consequence not of General Relativity, but of geometrical optics.

For the case of solar eclipse $\rho = R_s$ and

$$\alpha = -2r_g/R_s = -1.75''\,. \tag{25}$$

This is exactly the angle predicted by Einstein[7] in 1915 in the framework of General Relativity. He used the coordinate velocity v given by Eq. (15). Note that in 1911 by implicitly assuming that $g_{rr} = 1$ and hence, using $v = c(1 - r_g/2r)$ Einstein derived an expression for the angle in which coefficient 2 was missing.[8]

6. Conclusions

Thus in a static gravitational field the frequency ω of the photon is a constant; it is equal to the frequency at emission. The phenomenon of gravitational redshift is explained by the increase of the energy difference between levels in atoms or nuclei (in general, by the increase of the rate of clocks) with the increase of their distance from the gravitating body. The constant energy of photon appears as redshifted with respect to the blueshifted energy difference of atomic levels.

The momentum of a photon p and hence its wavelength λ change with the decrease of the distance from gravitating body: p increases, λ decreases. Hence the coordinate (world) velocity v of the photon decreases. This explains the two well established effects: The delay of radar echo from planets and the deflection of star light by the sun.

Acknowledgments

I am grateful to A. D. Dolgov, V. L. Ginzburg and K. G. Selivanov for insightful discussions. It is a pity that the latter refused to co-author this article. I would also like to thank S. D. Drell, Ya. I. Granowski, J. D. Jackson and especially V. L. Telegdi who read the draft of the article for their constructive criticism.

This work has been partly supported by A. von Humboldt award and by RFBR grant No. 00-15-96562.

References

1. L. B. Okun, K. G. Selivanov and V. L. Telegdi, *Usp. Fiz. Nauk* **169**, 1141 (1999) (in Russian).
2. L. B. Okun, K. G. Selivanov and V. L. Telegdi, *Amer. J. Phys.* **68**, 115 (2000).
3. S. Weinberg, *Gravitation and Cosmology* (Wiley, 1972).
4. L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Pergamon Press, 1962).
5. A. V. Berkov and I. Yu. Kobzarev, *Teoriya Tyagoteniya Einshteina. Obshchie Principy i Eksperimentalnye Sledstviya* (Moskva, MIFI, 1989) (in Russian).
6. I. I. Shapiro, *Phys. Rev. Lett.* **13**, 789 (1964).
7. A. Einstein, *Sitzungsber. Preuss. Acad. Wiss.* **47**, 831 (1915); *Ann. Phys.* **49**, 769 (1916); *The Meaning of Relativity* (Princeton Univ. Press, 1921).
8. A. Einstein, *Ann. Phys.* **35**, 898 (1911).

This page intentionally left blank

Modern Physics Letters A, Vol. 15, No. 32 (2000) 2007–2009

A THOUGHT EXPERIMENT WITH CLOCKS IN STATIC GRAVITY

L. B. OKUN*

ITEP, Moscow 117218, Russia

Received 23 October 2000

In order to directly demonstrate that in static gravitational field the rate of clocks increases with their distance from the source, a simple thought experiment is proposed.

1. Introduction

The redshift of the frequency of photons as they travel upwards in a static gravitational field and the corresponding increase of the frequency of clocks (e.g. of the spacings between atomic levels) were theoretically discovered by Einstein long before he constructed general relativity. These phenomena are discussed in a vast literature going from monographs on general relativity to popular books. In the recent articles[1–3] it has been argued that in a static field for a static observer the frequency of the photon does not change and that it is redshifted only with respect to the blueshifted atomic (nuclear) levels. Strangely enough, this statement met with objections from some of the experts in general relativity. These gave us a thought experiment involving only clocks but no photons, which could prove that standard clocks are running faster when they are raised above the ground. This letter proposes such a thought experiment. In order to make the *pro* and *contra* arguments clearer, the following text is written as a dialogue between a professional relativist (R) and the author (A).

2. Dialogue on the Thought Experiment

A: In Refs. 1–3 two kinds of arguments were presented: theoretical and experimental. According to the theoretical argument, the wave solution of Maxwell's equations in a static gravitational field must have a fixed frequency equal to that of the e.m. source since the coefficients of Maxwell's equations in that case do not depend on time.

R: The conservation of energy (or frequency) of the photon depends on the definition, depends on what you choose to call the energy or frequency, on the

*E-mail: okun@heron.itep.ru

definition of time and the system of reference. To be operationally meaningful you have to specify the way of measuring these entities.

A: As we consider a static gravitational field we should choose observers which are at rest in that field, say in laboratories situated at ground floor and the upper floor of the same building. That was how the first measurements of gravitational redshift were performed in the famous experiments of Pound and Rebka four decades ago. We should not consider observers in moving elevators. To them the gravitational field is not static. Moreover in a falling elevator the gravitational field vanishes locally.

R: OK. Let choose the static observers and consider the experiments by Pound *et al.* In these experiments photons were emitted by excited iron nuclei at the ground floor and their absorption by iron nuclei in their ground state was measured at the top of Harvard tower. It was observed that the energy of photons was not sufficient to excite in the iron nuclei upstairs the same level from which they were emitted downstairs. The relative shift of the frequency of photons and nuclei was $-gh/c^2$, where h is the height of the tower, $g = 10$ m/s^2 and c – the velocity of light. One can interpret the experiment by saying (as the authors of Refs. 1–3 did) that the energy of the photon has not changed, but that the energy of excitation of iron nuclei (the rate of the clock) has fractionally increased by $+gh/c^2$.

Alternatively, there is an equally valid viewpoint according to which the clock rates are the same while the photon frequency is different. For that it is enough to consider local inertial frames momentarily at rest at the worldpoints of the emission and absorption of the photon.

A: But neither Pound, nor his coworkers used in their experiments falling elevators.

R: One can introduce local time by simply assuming that standard clocks do not change their rate with their height. Anyway, you are attempting improperly to subdivide the gravitational redshift phenomenon into two separate processes, that separately have no operational meaning. The only operationally meaningful statement is that when a photon is emitted at the bottom of the tower, its frequency as measured by the identical clock at the top of the tower is lower. You cannot determine whether it was the photon or the clocks that were responsible for the shift.

My view would be different if you could carry out the following thought experiment. Two identical clocks are placed one at the top of a tower, the other at the bottom. Demonstrate without connecting the clocks by photons and otherwise disturbing the clocks that the top clock runs faster than the bottom clock by the standard gravitational shift. This would be a demonstration of your assertion that the photon frequency does not change. I believe such a demonstration to be impossible.

A: There were special experiments described in Refs. 1 and 2 in which one of the two clocks traveled in an aircraft and then was brought back to another one which was all this time at rest in the laboratory. These experiments showed that in agreement with general relativity the clock which traveled (after taking account of special relativity twin effect) was ahead of that which was at rest.

R: These results do not change my point of view, because the two clocks, in the course of the experiment, were treated differently: one made a loop in space, while the other was not moved at all. Thus, those experiments demonstrate actually only that the total elapsed times differ and say nothing (except by inference) about the instantaneous difference in their rates.

A: Inference is an inalienable part of physics. But I shall not dwell on this here. I shall just propose that thought experiment about which you asserted "I believe such a demonstration to be impossible".

Consider two rooms: one at the ground floor, the other at the top floor of a building. One starts with two identical standard clocks at the bottom room. At a certain moment one raises the first clock in the room upstairs. After some time T one raises the second clock in the same way and compares the time it shows with that of the first clock. If general relativity is correct, their time difference should be $\Delta T = (gh/c^2)T$. In this thought experiment both clocks are treated absolutely identical. Nevertheless the first clock will be ahead of the second. As time is uniform and both clocks are not moved during the time T, this thought experiment yields the instantaneous difference in the rates of the clocks, which is of course equal to $\Delta T/T = gh/c^2$.

3. Conclusion

By a thought experiment (which with improvements in the precision of atomic clocks could one day become a real one) it is shown that when measuring the "gravitational redshift of photons" one actually measures the gravitational blueshift of clocks.

Acknowledgments

I am grateful to Prof. R. H. Romer for stimulating me to invent the thought experiment described in this note. I am grateful to Val Telegdi for helpful remarks. This work was supported by RFBR (grant # 00-15-96562), A. von Humboldt Foundation and CERN Theory Division.

References

1. L. B. Okun, K. G. Selivanov and V. L. Telegdi *Am. J. Phys.* **68**, 115 (2000).
2. L. B. Okun, K. G. Selivanov and V. L. Telegdi, *Usp. Fiz. Nauk* **169**, 1141 (1999); *Phys. Usp.* **42**, 1045 (1999).
3. L. B. Okun, *Mod. Phys. Lett.* **A15**, 1941 (2000).

This page intentionally left blank

Physics of Atomic Nuclei, Vol. 64, No. 3, 2001, pp. 536–539. Translated from Yadernaya Fizika, Vol. 64, No. 3, 2001, pp. 590–593.

90th ANNIVERSARY OF A.B. MIGDAL'S BIRTHDAY
ELEMENTARY PARTICLES AND FIELDS

Relation between Energy and Mass in Bohr's Essay on His Debate with Einstein

L. B. Okun*

Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia

Received July 12, 2000

Abstract—The famous debate between Einstein and Bohr on the (in)consistency of quantum mechanics was described in detail by Bohr in his essay of 1949. The present article comments not on the main subject of the debate but only on the terminology that is relevant to the notions of the theory of relativity and which was used by the participants. In particular, their statement on the equivalence of mass and energy should not be taken literally. In fact, the rest energy is meant here. The authority of the two great physicists should not be misused to preserve the confusing terminology. © *2001 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The debate between Einstein and Bohr on quantum mechanics achieved a culmination point at the Sixth Solvay Congress in October 1930. In this debate, the former tried to prove the intrinsic inconsistency of quantum mechanics, while the latter aimed at showing that the gedanken experiments proposed by Einstein do not lead to such contradictions. This controversy was skillfully described in Bohr's article [1] published in 1949 and dedicated to Einstein's 70th birthday.

Bohr's article stimulated a wealth of contradictory responses from physicists, historians of physics, and philosophers. These include (in chronological order) M. Born, V. Fock, K. Popper, Y. Aharonov, D. Bohm, L. Rosenfeld, J. Holton, M. Klein, K. von Weizsäcker, M. Jammer, and A. Pais. A separate chapter of Jammer's monograph [2] is devoted to these responses. Articles on this theme have appeared in succeeding years as well (see, for example, [3, 4]). The latest of these was published in February 2000.

Various aspects of the debate between Bohr and Einstein and further development of the ideas stated in it are described in the miscellanea *Quantum Theory and Measurement* (over 800 pages) [5], which contains original articles and comments by the main participants and other eminent physicists.

2. EINSTEIN'S GEDANKEN EXPERIMENT

In the gedanken experiment proposed by Einstein in October 1930, there was a box filled with photons and hung by a spring balance, so that its vertical translation could be measured with a special scale. Thus, variations in the box mass could be measured by weighing.

In a lateral wall of the box, there was a small hole closed by a shutter that was controlled by a clock placed within the box. The clock opened the hole for a very short time interval, so that the instant at which a single photon escaped from the box could be determined to a precision as high as was desired.

On the other hand—as Bohr wrote [1]—the general energy–mass relation expressed by Einstein's famous formula

$$E = mc^2$$

made it possible to measure the photon energy, to any desired precision, by weighing the whole box prior to and after this event. According to Einstein's idea, this resulted in a contradiction with the known uncertainty relation

$$\Delta E \Delta T > h,$$

where ΔE is the uncertainty in the photon energy, ΔT is the uncertainty in the instant of its escape, and h is the Planck constant. Since the aforesaid suggested that both ΔE and ΔT could be arbitrarily small, their product could also be arbitrarily small in the gedanken experiment being discussed.

3. FAMOUS FORMULA $E = mc^2$

Prior to considering the way in which Bohr refuted Einstein's arguments, it is worth emphasizing that, in referring to Einstein's famous formula $E = mc^2$, Bohr did not specify the definitions of energy and mass. In [1], Bohr wrote that the discovery of induced radioactivity provided the most direct test of Einstein's fundamental law of mass–energy equivalence.

From these statements of Bohr, a reader that is insufficiently conversant with these matters could deduce that a photon with energy E has the mass $m = E/c^2$, and so does any particle or body.

This false interpretation of Einstein's famous formula is widespread in the popular-science literature

* e-mail: okun@heron.itep.ru

despite attempts at stamping it out (see, for example, [6–9]). Within the theory of relativity, it inevitably entails the completely superfluous notion of the rest mass and the use of nonrelativistic notions like the inertia mass and the gravitational mass (which were mentioned by Bohr in connection with the formula $E = mc^2$).

The phrase "equivalence of mass and energy" is repeated in the monograph [2] and in the articles [3, 4], where it is stated that energy has a weight. In [4], m is defined either as a "mass" or as a "rest mass."

4. E OR E_0? 1905–1921

Possibly, there would have been no reason for that inexact interpretation if Bohr had explained in more detail that E in Einstein's formula was the energy inherent in a body of mass m at rest, or, in other words, the rest energy of the body. It was just this meaning that was assigned to it by Einstein when he was deducing his formula in 1905 [10]. This follows from the fact that, in this article, he considered a body at rest that had emitted two oppositely directed light waves carrying away energy but not momentum, so that the body remained at rest after the emission event.

Einstein's discovery was as follows: if a body emits energy L in the form of radiation, then the mass of the body decreases by L/V^2. Obviously, it is of no importance in this case that the energy emitted by the body is converted directly into radiant energy. Owing to this, we arrive at the more general conclusion that the mass of a body is a measure of its energy content: if the energy changes by L, the mass changes accordingly by $L/(9 \times 10^{20})$, the energy and the mass being measured in ergs and in grams, respectively.

Obviously, V is the speed of light. It is interesting to note that there was no special symbol for mass in [10]. In present-day notation, L/V^2 is the right-hand side of the equation

$$\Delta m = \Delta E_0/c^2,$$

where E_0 is the rest energy. This notation dates back to Einstein's monograph *The Meaning of Relativity* [11], where Eq. (44) has the form

$$E_0 = mc^2$$

and where E_0 is referred to as the energy of a body at rest.

It should be emphasized that Einstein was not consistent in the period between the publication of the article [10] and the publication of the monograph [11]; occasionally, he interpreted E in his famous formula as the total energy of a body and wrote on mass–energy equivalence. It is possible that this inconsistency helped him to predict the effect of the gravitational red shift (1907), the deviation of a light ray by the Sun (1911),[1] and eventually to create the general theory of relativity (1916). However, this inconsistency gave rise to the terminological confusion that has existed to date.

5. E OR E_0? 1921–1955

An interchange of the correct and incorrect formulations of the mass–energy relation in Einstein's articles published in the 1920s–1940s contributed greatly to this confusion. This can easily be seen in the second of the four volumes of his collected works.[2] He mentions the equivalence of energy and inertial mass in his report [12]. The title of the lecture [13] is "Elementary Derivation of the Equivalence of Mass and Energy." At the same time, it is the rest energy, and not energy in general, that is mentioned in the main body of this lecture, where he writes that, here, it is natural to assign the rest energy m (mc^2 in standard units) to a particle at rest and that the supposed mass–energy equivalence exists as well; a further statement is that the relation $E_0 = m$ proves the equivalence of the inertia mass and the rest energy.

In the minor note [14], however, which has the same title "Elementary Derivation of the Equivalence of Mass and Energy," the term rest energy is absent, and the last paragraph says that this relation expresses the law of energy–mass equivalence. The increase of E in energy is associated with the increase of E/c^2 in mass. Since the definition of energy usually admits arbitrariness in an additive constant, one can choose this constant in such a way that $E = Mc^2$.

The first phrase of the next article, "$E = mc^2$: An Urgent Problem of Our Time" [15], comments on the principle of mass–energy equivalence. Two pages later, we can find the statement that the equivalence of mass and energy is usually expressed (although this is not quite correct) by the formula $E = mc^2$, where c is the speed of light equal to about 186000 miles per second, E is the energy content of a body at rest, and m is its mass.

In the article [16], which was written for an encyclopedia, Einstein omits the term "rest energy" once again. Summarizing the basic results of the special theory of relativity, he states that this demonstrates the equivalence of mass and energy.

These "terminological oscillations" are a real psychological puzzle. It seems impossible that Einstein was annoyed to pinpoint each time that he meant the rest energy E_0. He rather tried to proceed on equal terms with the reader, who had got used to "the famous formula $E = mc^2$" from his school days. Possibly, Einstein tried to emphasize the philosophical aspect of the problem: it is important that mass is related to energy, but it is not so important to which energy specifically. Only one thing is doubtless: from 1921 to the end of his life (1955), Einstein had spoken and written either explic-

[1] The angle of deviation as calculated in 1911 was one-half as large as the correct value predicted by Einstein on the basis of the general theory of relativity.

[2] Translated and published in Russian.

itly or implicitly about the equivalence of the mass of a body and its rest energy.

6. BOHR'S COUNTERARGUMENT

Bohr did not sleep all night long, thinking over Einstein's arguments contra the relation $\Delta T\Delta E > h$ (see, for example, [2]). In the morning, he disproved these arguments using Einstein's theory of general relativity. Bohr described it himself as follows [1]. In order to discuss the problem, it appeared to be necessary to consider the consequences of identifying the inertia mass with the gravitational mass that are caused by the use of the relation $E = mc^2$. It was especially necessary to take into account the relation between the rate of the clock and its position in the gravitational field. This relation, which is well known from the red shift of the lines of the Sun spectrum, follows from Einstein's principle of equivalence of effects of gravity and phenomena observed in accelerated frames of reference.

Suppose that measurement of the box weight by a spring balance to a precision Δm requires estimating its vertical translation to a precision Δq. According to the relation $\Delta q\Delta p \approx h$, this generates an uncertainty Δp in the box momentum. This uncertainty in turn must be smaller than the total momentum that body of mass Δm can acquire over the weighing period T owing to the effect of the gravitational field; that is,

$$\frac{h}{\Delta q} \approx \Delta p < Tg\Delta m,$$

where g is the acceleration due to gravity.

The position of the clock that is fixed rigidly within the box and which controls the shutter must also have the uncertainty Δq. But according to the general theory of relativity, a vertical translation of the clock changes its rate in such a way that

$$\frac{\Delta T}{T} = \frac{g\Delta q}{c^2}.$$

From the last two relations, it follows that

$$\Delta T > \frac{h}{c^2\Delta m}.$$

By using the "famous formula" in the form $\Delta E = \Delta mc^2$, Bohr obtained the required relation

$$\Delta T\Delta E > h.$$

Jammer writes that, after the Sixth Solvay Congress, Einstein concentrated on attempts at proving the incompleteness of quantum mechanics rather than on attempts at proving its inconsistency (see [2]).

7. ADDITIONAL COMMENTS ON TERMINOLOGY

Many authors who critically analyzed Bohr's proof mentioned above discussed various details of time measurement (so-called internal and external measurement) and the interpretation of the relation $\Delta T\Delta E > h$. However, I am unaware of any critical comment on energy–mass equivalence or on the identity of the inertia mass and the gravitational mass. Moreover, some authors wrote about weighing energy or stated that energy has a weight.

It is absolutely obvious that, in order to describe Einstein–Bohr's gedanken experiment unequivocally, it is necessary and sufficient to substitute the clause "rest energy" for the word "energy." The same concerns the weight of energy. As to identifying the inertia mass with the gravitational mass, these terms, as well as their identification, belong to the notions of Newtonian nonrelativistic theory. After all, it is widely known that the inertial properties of relativistic particles are controlled by their energy and that the gravitational properties are determined by the energy–momentum tensor, and not by the mass.

It should be emphasized that Bohr was absolutely right to explain the red shift of the spectral lines in a static gravitational field by the variation of the clock rate—that is, by the shift of atomic levels—and not by the photon weight. (A comparison of these two alternative explanations is drawn in [17, 18], where it is explained in detail that the second is wrong).

8. CONCLUSION

In summary, Bohr's analysis in [1] does not lead to the conclusion, despite some terminological faults, that the photon with energy E has the mass $m = E/c^2$.

ACKNOWLEDGMENTS

I am grateful to V.L. Telegdi and V.B. Braginsky, who called my attention to the articles [1] and [4], respectively, in connection with the problem of mass–energy equivalence and to A.Yu. Morozov and K.G. Selivanov for stimulating discussions.

This work was supported by the Russian Foundation for Basic Research (project no. 00-15-96562) and in part by A. von Humboldt Foundation.

POSTSCRIPT

When I got acquainted with A.B. Migdal fifty years ago, problems of history and philosophy of physics were of small interest to him, since he was fascinated by solving specific scientific problems and taught his pupils, me among them, to do that. In the late 1970s, the focus of his interests shifted drastically. He kept studying the articles by Bohr, Einstein, and other classical scientists. He lectured on the history of physics and on the psychology of creative activity in science. His popular-science articles appeared one after another [19–23], and many of them were translated into foreign languages. He became the editor-in-chief of the *Encyclopedic Dictionary for the Young Physicist* [24]. I often recollect A.B. Migdal with deep gratitude, and I thought about him very frequently when I was writing this note about the debate between Bohr and Einstein. I believe that he would have taken interest in this note.

REFERENCES

1. N. Bohr, in *Albert Einstein: Philosopher-Scientist*, Ed. by P. A. Schilpp (Library of Living Philosophers, Evanston, 1949), Vol. VII, p. 200.
2. M. Jammer, *The Philosophy of Quantum Mechanics: The Interpretations of Quantum Mechanics in Historical Perspective* (Wiley, New York, 1974), Chap. 5.
3. W. G. Unruh and G. I. Opat, Am. J. Phys. **47**, 743 (1979).
4. Y. Aharonov and B. Reznik, Phys. Rev. Lett. **84**, 1368 (2000).
5. *Quantum Theory and Measurement*, Ed. by J. A. Wheeler and W. H. Zurek (Princeton Univ. Press, Princeton, 1983).
6. E. Taylor and J. A. Wheeler, *Spacetime Physics* (Freeman, New York, 1992, 3rd ed.), pp. 258, 272.
7. L. B. Okun', Usp. Fiz. Nauk **158**, 511 (1989) [Sov. Phys. Usp. **32**, 629 (1989)].
8. L. B. Okun, Phys. Today **42**, 31 (1989); **43**, 13 (1990); **43**, 15 (1990); **43**, 115 (1990); **43**, 117 (1990).
9. L. B. Okun, Eur. J. Phys. **15**, 403 (1998).
10. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 1, p. 36.
11. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 5.
12. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 109.
13. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 416.
14. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 650.
15. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 653.
16. A. Einstein, *Collection of Scientific Works in Four Volumes* (Nauka, Moscow, 1966), Vol. 2, p. 657.
17. L. B. Okun', K. G. Selivanov, and V. L. Telegdi, Usp. Fiz. Nauk **169**, 1141 (1999).
18. L. B. Okun, K. G. Selivanov, and V. L. Telegdi, Am. J. Phys. **68**, 115 (2000).
19. A. B. Migdal, *Subscription Popular Scientific Series "Physics,"* No. 7: *Search for Truth* (Znanie, Moscow, 1978).
20. A. B. Migdal, *Search for Truth* (Molodaya Gvardiya, Moscow, 1983).
21. A. B. Migdal, *How Physical Theories Occur* (Pedagogika, Moscow, 1984).
22. A. B. Migdal, *Subscription Popular Scientific Series "Physics,"* No. 3: *Quantum Physics and Niels Bohr* (Znanie, Moscow, 1987).
23. A. B. Migdal, *Library "Quantum,"* Vol. 75: *Quantum Physics for Grown-up and Youngs* (Nauka, Moscow, 1989).
24. *Encyclopedic Dictionary for the Young Physicist. For Older and Middle Age*, Ed. by A. B. Migdal (Pedagogika, Moscow, 1984).

Translated by E. Kozlovskiĭ

JHEP Published by Institute of Physics Publishing for SISSA/ISAS
Received: *March 4, 2002*
Accepted: *March 9, 2002*

Trialogue on the number of fundamental constants

Michael J. Duff
Michigan Center for Theoretical Physics, Randall Laboratory
Ann Arbor, MI, USA
E-mail: mduff@umich.edu

Lev B. Okun
ITEP
Moscow, 117218, Russia
E-mail: okun@heron.itep.ru

Gabriele Veneziano
Theory Division, CERN
CH-1211 Geneva 23, Switzerland, and
Laboratoire de Physique Thèorique, Universitè Paris Sud
91405, Orsay, France
E-mail: Gabriele.Veneziano@cern.ch

Abstract: This paper consists of three separate articles on the number of fundamental dimensionful constants in physics. We started our debate in summer 1992 on the terrace of the famous CERN cafeteria. In the summer of 2001 we returned to the subject to find that our views still diverged and decided to explain our current positions. LBO develops the traditional approach with three constants, GV argues in favor of at most two (within superstring theory), while MJD advocates zero.

Keywords: Models of Quantum Gravity, Standard Model.

http://jhep.sissa.it/archive/papers/jhep032002023/jhep032002023.pdf

Contents

Part I
Fundamental constants: parameters and units

Lev B. Okun

Abstract. There are two kinds of fundamental constants of Nature: dimensionless (like $\alpha \simeq 1/137$) and dimensionful (c — velocity of light, $\hbar$ — quantum of action and angular momentum, and G — Newton's gravitational constant). To clarify the discussion I suggest to refer to the former as fundamental parameters and the latter as fundamental (or basic) units. It is necessary and sufficient to have three basic units in order to reproduce in an experimentally meaningful way the dimensions of all physical quantities. Theoretical equations describing the physical world deal with dimensionless quantities and their solutions depend on dimensionless fundamental parameters. But experiments, from which these theories are extracted and by which they could be tested, involve measurements, i.e. comparisons with standard dimensionful scales. Without standard dimensionful units and hence without certain conventions physics is unthinkable.

1. Introduction: parameters and units

There is no well established terminology for the fundamental constants of Nature. It seems reasonable to consider as fundamental the dimensionless ratios, such as the famous $\alpha = e^2/\hbar c \simeq 1/137$ and similar gauge and Yukawa couplings in the framework of standard model of elementary particles or its extensions.

It is clear that the number of such constants depends on the theoretical model at hand and hence depends on personal preferences and it changes of course with the evolution of physics. At each stage of this evolution it includes those constants which cannot be expressed in terms of more fundamental ones, because of the absence of the latter [1]. At present this number is a few dozens, if one includes neutrino mixing angles. It blows up with the inclusion of hypothetical new particles.

On the other hand the term "fundamental constant" is often used for such dimensionful constants as the velocity of light c, the quantum of action (and of angular momentum) $\hbar$, and the Newton gravitational coupling constant G . This article is concerned with these dimensionful constants which I propose to call fundamental (or basic) units.

Physics consists of measurements, formulas and "words". This article contains no new formulas, it deals mainly with "words" because, unlike many colleagues of mine, I believe that an adequate language is crucial in physics. The absence of accurately defined terms or the uses (i.e. actually misuses) of ill-defined terms lead to confusion and proliferation of wrong statements.

2. Stoney's and Planck's units of L, T, M

The three basic physical dimensions: length L, time T, and mass M with corresponding metric units: cm, sec, gram, are usually associated with the name of C.F. Gauss. In spite of tremendous changes in physics, three basic dimensions are still necessary and sufficient to express the dimension of any physical quantity. The number three corresponds to the three basic entities (notions): space, time and matter. It does not depend on the dimensionality of space, being the same in spaces of any dimension. It does not depend on the number and nature of fundamental interactions. For instance, in a world without gravity it still would be three.

In the 1870's G.J. Stoney [2], the physicist who coined the term "electron" and measured the value of elementary charge e, introduced as universal units of Nature for L, T, M: $l_S = e\sqrt{G}/c^2$, $t_S = e\sqrt{G}/c^3$, $m_S = e/\sqrt{G}$. The expression for m_S has been derived by equating the Coulomb and Newton forces. The expressions for l_S and t_S has been derived from m_S, c and e on dimensional grounds: $m_S c^2 = e^2/l_S$, $t_S = l_S/c$.

When M. Planck discovered in 1899 $\hbar$, he introduced [3] as universal units of Nature for L, T, M: $l_P = \hbar/m_P c$, $t_P = \hbar/m_P c^2$, $m_P = \sqrt{\hbar c/G}$.

One can easily see that Stoney's and Planck's units are numerically close to each other, their ratios being $\sqrt{\alpha}$.

3. The physical meaning of units

The Gauss units were "earth-bound" and "hand-crafted". The cm and sec are connected with the size and rotation of the earth.[1] The gram is the mass of one cubic cm of water.

An important step forward was made in the middle of XX century, when the standards of cm and sec were defined in terms of of wave-length and frequency of a certain atomic line.

Enormously more universal and fundamental are c and $\hbar$ given to us by Nature herself as units of velocity $[v] = [L/T]$ and angular momentum $[J] = [MvL] = [ML^2/T]$ or action $[S] = [ET] = [Mv^2T] = [ML^2/T]$. (Here [ ] denotes dimension.)

3.1 The meaning of c

It is important that c is not only the speed of light in vacuum. What is much more significant is the fact that it is the maximal velocity of any object in Nature, the photon being only one of such objects. The fundamental character of c would not be diminished in a world without photons. The fact that c is the maximal v leads to new phenomena, unknown in newtonian physics and described by relativity. Therefore Nature herself suggests c as fundamental unit of velocity.

In the Introduction we defined as fundamental those constants which cannot be calculated at our present level of fundamental knowledge (or rather ignorance). This "negative" definition applies equally to parameters and to units (to α and to c). At first sight α looks superior to c because the value of α does not depend on the choice of units, whereas the numerical value of c depends explicitly on the units of length and time and hence on conventions. However c is more fundamental than α because its fundamental character has not only a "negative" definition, but also a "positive" one: it is the basis of relativity theory which unifies space and time, as well as energy, momentum and mass.

[1]metre was defined in 1791 as a 1/40,000,000 part of Paris meridian.

By expressing v in units of c (usually it is defined as $\beta = v/c$) one simplifies relativistic kinematics. On the other hand the role of c as a conversion factor between time and distance or between mass and rest-energy is often overstated in the literature. Note that in spite of the possibility of measuring, say, distance in light-seconds, the length does not become identical to time, just as momentum is not identical to energy. This comes from the pseudoeuclidian nature of four-dimensional space-time.

3.2 The meaning of $\hbar$

Analogously to c, the quantity $\hbar$ is is also fundamental in the "positive" sense: it is the quantum of the angular momentum J and a natural unit of the action S. When J or S are close to $\hbar$, the whole realm of quantum mechanical phenomena appears.

Particles with integer J (bosons) tend to be in the same state (i.e. photons in a laser, or Rubidium atoms in a drop of Bose-Einstein condensate). Particles with half-integer J (fermions) obey the Pauli exclusion principle which is so basic for the structure of atoms, atomic nuclei and neutron stars.

Symmetry between fermions and bosons, dubbed supersymmetry or SUSY, is badly broken at low energies, but many theorists believe that it is restored near the Planck mass (in particular in superstrings and M-theories).

The role of $\hbar$ as a conversion factor between frequency and energy or between wavelength and momentum is often overstated.

It is natural when dealing with quantum mechanical problems to use $\hbar$ as the unit of J and S.

3.3 The status of G

The status of G and its derivatives, m_P, l_P, t_P, is at present different from that of c and $\hbar$, because the quantum theory of gravity is still under construction. The majority of experts connect their hopes with extra spatial dimensions and superstrings.[2] But the bridge between superstrings and experimental physics exists at present only as wishful thinking. Recent surge of interest to possible modifications of Newton's potential at sub-millimetre distances demonstrates that the position of G is not as firm as that of c and $\hbar$.

4. The cube of theories

The epistemological role of c, $\hbar$, G units in classifying theories was first demonstrated in a jocular article by G. Gamov, D. Ivanenko and L. Landau [4], then quite seriously by M. Bronshtein [5, 6], A. Zelmanov [8, 7] and others (see e.g. [9, 10]); and it is known now as the cube of theories.

The cube is located along three orthogonal axes marked by c (actually by $1/c$), $\hbar$, G. The vertex (000) corresponds to non-relativistic mechanics, (c00) — to special relativity, (0$\hbar$0) — to non-relativistic quantum mechanics, ($c\hbar$0) — to quantum field theory, ($c0G$)

[2]The characteristic length of a superstring $\lambda_s = l_P/\sqrt{\alpha_{\mathrm{GUT}}}$, where $\alpha_{\mathrm{GUT}} = \alpha(q^2 = M^2{}_{\mathrm{GUT}})$. (As is well known, the fundamental parameters are "running": their values depend on q^2.)

— to general relativity, $(c\hbar G)$ — to futuristic quantum gravity and the Theory of Everything, TOE. There is a hope that in the framework of TOE the values of dimensionless fundamental parameters will be ultimately calculated.

5. The art of putting $c = 1$, $\hbar = 1$, $G = 1$

The universal character of $c, \hbar, G$ and hence of m_P, l_P, t_P makes natural their use in dealing with futuristic TOE. (In the case of strings the role of l_P is played by the string length λ_s.) In such natural units all physical quantities and variables become dimensionless. In practice the use of these units is realized by putting $c = 1$, $\hbar = 1$, G (or λ_s) $= 1$ in all formulas. However one should not take these equalities too literally, because their left-hand sides are dimensionful, while the right-hand sides are dimensionless. It would be more proper to use arrows "$\rightarrow$" (which mean "substituted by") instead of equality signs "=".

The absence of $c, \hbar, G$ (or any of them) in the so obtained dimensionless equations does not diminish the fundamental character of these units. Moreover it stresses their universality and importance.

It is necessary to keep in mind that when comparing the theoretical predictions with experimental results one has anyway to restore ("$\leftarrow$") the three basic units $c, \hbar, G$ in equations because all measurements involve standard scales.

The above arguments imply what is often dubbed as a "moderate reductionism", which in this case means that all physical phenomena can be explained in terms of a few fundamental interactions of fundamental particles and thus expressed in terms of three basic units and a certain number of fundamental dimensionless parameters.

6. International system of units

An approach different from the above underlies the International System of Units (Systéme Internationale d'Unitées — SI) [11, 12]. This System includes 7 basic units (metre, second, kilogram, ampere, kelvin, mole, candela) and 17 derivative ones. The SI might be useful from the point of view of technology and metrology, but from the point of view of pure physics four out of its seven basic units are evidently derivative ones. Electric current is number of moving electrons per second. Temperature is up to a conversion factor (Boltzman constant $k = 1.38 \times 10^{-23}$ joules/kelvin) is the average energy of an ensemble of particles. Mole is trivially connected with the number of molecules in one gram-molecule, called Avogadro's number $N_A = 6.02 \times 10^{23}$/mole. As for unit of optical brightness or illumination (candela), it is obviously expressed in terms of the flux of photons.

It is interesting to compare the character of k with that of $c, \hbar, m_P$. The Boltzman constant is an important conversion factor which signals the transition from a few (or one) particle systems to many particle systems. However it radically differs from $c, \hbar, m_P$, as there is no physical quantity with the dimension of k, for which k is a critical value. The role of conversion factor is only a secondary one for $c, \hbar, m_P$, whereas for k it is the only one.

In the framework of SI vacuum is endowed with electric permittivity $\varepsilon_0 = 8.85 \times 10^{-12}$ farad/m and magnetic permeability $\mu_0 = 12.57 \times 10^{-17}$ newton/(ampere)^2, whereas $\varepsilon_0 \mu_0 = 1/c^2$. This is caused by electrodynamic definition of charge, which in SI is secondary with respect to the current. In electrostatic units $\varepsilon_0 = \mu_0 = 1$. According to the SI standard this definition is allowed to use in scientific literature, but not in text-books (see critical exposition of SI in ref. [13]).

7. Remarks on Gabriele's part II

I note with satisfaction that some of the original arguments and statements do not appear in his part of this Trialogue II. Among them there are the following statements: 1. that in string theory there is room only for two and not three dimensionful constants [14, 15]; 2. that units of action are arbitrary [which means that $\hbar$ is not a fundamental unit (LO)]; 3. that masses unlike length and time intervals are not measured directly [16]. Gabriele admits in section 6 that his two units can be "pedagogically confusing" and the set $c, \hbar, \lambda_s$ is "most practical", but he considers the latter "not economical" and in other parts of the part II he insists on using ${\lambda_s}^2$ instead of $\hbar$.

Of course, if you forget about the pedagogical and practical sides of physics, the most economical way is not to have fundamental units at all, like Mike, but that is a purely theoretical approach (`hep-th`), and not physical one (`physics`, `hep-ph`).

It seems to me inconsistent to keep two units (c, λ_s) explicitly in the equations, while substituting by unity the third one $(\hbar)$, as Gabriele is doing in part II and refs. [14, 15, 16]. According to my section 5 above, this corresponds to using $\hbar$ as a unit of J and S, while not using c and λ_s as units of velocity and length.

I also cannot agree that the electron mass, or G_F are as good for the role of fundamental unit as the Planck mass or G.

8. Remarks on Mike's part III

In section 4 of Mike's part III he introduces a definition of fundamental constants with the help of an alien with whom it is possible to exchange only dimensionless numbers. According to Mike, only those constants are fundamental the values of which can be communicated to the alien. Thus Mike concludes that there exist no fundamental units. According to my section 5 above, this actually corresponds to the use of $c, \hbar, G$ as fundamental units.

In fact, at the end of section 2 Mike writes "that the most economical choice is to use natural units where there are no conversion factors at all." Mike explained to me that his natural units are $c = \hbar = G = 1$. As these equalities cannot be considered literally, I believe that Mike uses the same three units as I do. However he concludes section 2 with a statement: "Consequently, none of these units or conversion factors is fundamental."

(In response to the above paragraph Mike added a new paragraph to his section 2, in which he ascribed to me the view that one cannot put $c = 1$. According to my section 5, one can (and should!) put $c = 1$ in relativistic eqations, but must understand that this means that c is chosen as the unit of velocity.)

The "alien definition" of fundamental constants is misleading. We, theorists, communicate not with aliens, but with our experimental colleagues, students, and non-physicists. Such communication is impossible and physics is unthinkable without standardized dimensionful units, without conventions..

Concerning Mike's criticism of my article [10], I would like to make the following remark. The statement that only dimensionless variables, functions and constants have physical meaning in a theory does not mean that every problem should be explicitly presented in dimensionless form. Sometimes one can use dimensionful units and compare their ratios with ratios of other dimensionful units. This approach was used in ref. [10], where entertaining stories by O. Volberg [17] and G. Gamov [18] were critically analyzed. In these stories, in order to demonstrate the peculiarities of relativistic kinematics, the velocity of light was assumed to be of the order of that of a car, or even bicycle, while the everyday life remained the same as ours. In ref. [10] I have shown that if c is changed, while dimensions of atoms are not changed (mass and charge of electron as well as $\hbar$, are the same as in our world), then electromagnetic and optical properties of atoms (and hence the everyday life) would change drastically because of change of α, which is the ratio of electron velocity in hydrogen atom to that of light. It is not clear to me why in section 5 of his paper Mike disagrees with these considerations.

9. Conclusions

It is obvious that using proper language (terms and semantics) three fundamental units are the only possible basis for a selfconsistent description of fundamental physics. Other conclusions are viable only through the improper usage of terms.

Acknowledgments

I am very grateful to Gabriele and Mike for their patience in discussing their points of view. I am indebted to Val Telegdi who critically read the draft of this article and helped to improve it. The hospitality of the CERN Theory Division and Michigan Center of Theoretical Physics during the completion of this article is appreciated, as well as Humboldt Award and the grant of RFBR 00-15-96562.

Part II
Fundamental units in physics: how many, if any?

Gabriele Veneziano

Abstract. I summarize my previous work on the question of how many fundamental dimensionful constants (fundamental units) are needed in various theoretical frameworks such as renormalizable QFT + GR, old-fashioned string theory, and modern string/M-theory. I will also try to underline where past and present disagreement on these issues between Lev Okun, Mike Duff, and myself appears to be originating from.

1. Introductory remarks

Some fifteen years ago I wrote a short letter [14] on the number of (dimensionful) fundamental constants in string theory, where I came to the somewhat surprising conclusion that *two* constants, with dimensions of space and time, were both necessary and sufficient. Somewhat later, I became aware of S. Weinberg's 1983 paper [1], whose way of looking at the question of defining fundamental constants in physics I tried to incorporate in my subsequent work on the subject [15, 16].

After reading those papers of mine once more, I still subscribe to their content, even if I might have expressed some specific points differently these days. Here, rather than repeating the details of my arguments, I will try to organize and summarize them stressing where, in my opinion, the disagreement between Lev, Mike and myself arises from. I have the impression that, in the end, the disagreement is more in the words than in the physics, but this is what we should try to find out.

The rest of this note is organized as follows: In section 2 I make some trivial introductory statements that are hopefully uncontroversial. In sections 3, 4 and 5 I describe how I see the emergence of fundamental units (the name I will adopt for fundamental dimensionful constants following Lev's suggestion) in QFT+GR, in the old Nambu-Goto formulation of quantum string theory (QST), and in the so-called Polyakov formulation, respectively. In sections 6 I will try to point at the origin of disagreement between myself and Lev while, in section 7, the same will be done w.r.t. Mike. Section 8 briefly discusses the issue of time-varying fundamental units.

2. Three questions and one answer

Let me start with two statements on which we all seem to agree:

- Physics is always dealing, in the end, with dimensionless quantities, typically representing ratios of quantities having the same dimensions, e.g.

$$\alpha = \frac{e^2}{\hbar c}, \qquad \frac{m_e}{m_p}, \qquad \dots \tag{2.1}$$

- It is customary to introduce "units", i.e. to consider the ratio of any physical quantity q to a fixed quantity u_q *of the same kind* so that

$$q = (q/u_q)u_q \,, \tag{2.2}$$

where u_q is a *name* (e.g. *centimetre* or *second*) and (q/u_q) is a *number*. Obviously, $q_1/q_2 = (q_1/u_q)/(q_2/u_q)$.

- Let us now ask the following three questions

  Q_1: are units arbitrary?

  Q_2: are there units that are more fundamental than others according to S. Weinberg's definition [1]?

  Q_3: How many units (fundamental or not) are necessary?

  and try to answer them in the context of different theories of elementary particles and interactions.

I hope we agree that the answer to the first question is yes, since only q_i/q_j matter and these ratios do not depend on the choice of units.

I think that the answer to the other two questions depends on the framework we are considering (Cf. Weinberg, ref. [1]). The next three sections therefore analyze Q_2 and Q_3 within three distinct frameworks, and provide, for each case, answers A_2 and A_3, respectively.

3. Fundamental units in QFT+GR

Quantum Field Theory (QFT) (or more specifically the Standard Model (SM)) plus General Relativity (GR) represent the state of the art in HEP before the string revolution of 1984. Weinberg's 1983 paper [1] reflects therefore the attitude about FC's at the dawn of the string revolution. I would summarize it briefly as follows:

- A_2: a qualified yes.

  At the QFT level of understanding c and $\hbar$ appear to be more fundamental units of speed and action than any other. In newtonian mechanics only the ratios of various velocities in a given problem matter. By contrast, in (special) relativity the ratio of each velocity appearing in the problem to the (universal) speed of light, c, also matters. Likewise, in classical mechanics only the ratios of various terms in the action matter, the overall normalization being irrelevant while, in QM, the ratio of the total action to the (universal) quantum of action $\hbar$ does matter (large ratios, for instance, correspond to a semiclassical situation). It appears therefore that both c and $\hbar$ have a special status as the most basic units of speed and action.

  Indeed, let's apply S. Weinberg's criterion [1] and ask: can we compute c and $\hbar$ in terms of more fundamental units? Within QFT the answer appears to be an obvious no. Had we chosen instead some other arbitrary units of speed and action, then, within a given theory, we would be able to compute them, in principle at least, in terms of c and $\hbar$, i.e. in terms of something more fundamental (and of some specified dimensionless constants such as α).

- A_3: most probably three

It is quite clear, I think, that in QFT+GR we cannot compute everything that is observable in terms of c, $\hbar$, and of dimensionless constants, without also introducing some mass or length scale. Hence it looks that the answer to the third question is indeed three. Unlike in the case of c and $\hbar$, it is much less obvious, however, which mass or length scale, if any, is more fundamental in the sense of SW. The Planck mass, M_P, does not look like a particularly good choice since it is very hard, even conceptually, to compute, say, m_e or m_p in terms of M_P in the SM + GR framework. This is a bit strange: we seem to need three units, but we can only identify two fundamental ones. So why three? Why not more? Why not less?

Why not more? This is because it looks unnecessary (and even"silly" according to present understanding of physical phenomena) to introduce a separate unit for temperature, for electric current and resistance, etc., or separate units for distances in the x, y and z directions. I refer to Lev for a discussion about how to go from the seven units of the International System of Units (SI) down to three [13], and for how three fundamental units define the so-called "'cube" of physical theories [10].

And why not less, say just two? Well because mass or energy appear as concepts that are qualitatively different from, say, distances or time intervals. Let us recall how mass emerges in classical mechanics (CM). We can base CM on the action principle and get $F = ma$ by varying the action

$$S = \int \left(\frac{1}{2} m\dot{x}^2 - V(x) \right) dt \Rightarrow ma = F \equiv -\frac{dV}{dx} , \tag{3.1}$$

but, as it's well known, classically the action can be rescaled by an arbitrary factor. If we had only one species of particles in Nature we could use, instead of S,

$$\tilde{S} = \int \left(\frac{1}{2} \dot{x}^2 - \frac{V(x)}{m} \right) dt \equiv \int \left(\frac{1}{2} \dot{x}^2 - \tilde{V}(x) \right) dt \Rightarrow a = \tilde{F} \equiv -\frac{1}{m} \frac{dV}{dx} . \tag{3.2}$$

No physical prediction would change by using units in which masses are pure numbers provided we redefine forces accordingly! In this system of units $\hbar$ would be replaced by $\hbar/m$ and would have dimensions of $v^2 \times t$. If we have already decided for c as unit of velocity, $\hbar$ would define therefore a fundamental unit of time (the Compton wavelength of the chosen particle divided by c). However, in the presence of many particles of different mass, we cannot decide which mass to divide the action by, which choice is most fundamental.

I think there is even a deeper reason why QFT+GR needs a separate unit for mass. QFT is affected by UV divergences that need to be renormalized. This forces us to introduce a cut-off which, in principle, has nothing to do with c, $\hbar$ or M_P, and has to be "removed" in the end. However, remnants of the cut-off remain in the renormalized theory. In QCD, for instance, the hadronic mass scale (say m_p) originates from a mechanism known as dimensional transmutation, and is arbitrary. Perhaps one day, through string theory or some other unified theory of all interactions, we will understand how m_p is related to M_P, but in QFT+GR it is not. We do not know therefore which of the two, M_P or m_p, is more fundamental and the same is true for the electron mass m_e, for G_F etc. etc.

The best we can do, in QFT+GR, is to take any one of these mass scales (be it a particle mass or a mass extracted from the strength of a force) as unit of mass and consider the ratio of any other physical mass to the chosen unit as a pure number that, in general, we have no way to compute, even in principle.

4. Fundamental units in old-fashioned quantum string theory (QST)

- A_2: yes, c and λ_s!

  With string theory the situation changes because it is as if there were a single particle, hence a single mass. Indeed, a single classical parameter, the string tension T, appears in the Nambu-Goto (NG) action:

$$S = T \int d(\text{Area})\,, \qquad \frac{S}{\hbar} = \lambda_s^{-2} \int d(\text{Area})\,, \tag{4.1}$$

  where the speed of light c has already been implicitly used in order to talk about the area of a surface embedded in space-time. This fact allows us to replace $\hbar$ by a well defined length, λ_s, which turns out to be fundamental both in an intuitive sense and in the sense of S. Weinberg. Indeed, we should be able, in principle, to compute any observable in terms of c and λ_s (see below for an example). Of course, I could instead compute c and λ_s in terms of two other physical quantities defining more down-to-earth units of space and time, but this would not satisfy SW's criterion of having computed c and λ_s in terms of something more fundamental!

- A_3: the above two constants are also sufficient!

  This was the conclusion of my 1986 paper: string theory only needs two fundamental dimensionful constants c and λ_s, i.e. one fundamental unit of speed and one of length.

  The apparent puzzle is clear: where has our loved $\hbar$ disappeared? My answer was then (and still is): it changed its dress! Having adopted new units of energy (energy being replaced by energy divided by tension, i.e. by length), the units of action (hence of $\hbar$) have also changed. And what about my reasoning in QFT+GR? Obviously it does not hold water any more: For one, QFT and GR get unified in string theory. Furthermore, the absence of UV divergences makes it unnecessary to introduce by hand a cut off.

  And indeed the most amazing outcome of this reasoning is that the new Planck constant, λ_s^2, *is* the UV cutoff. We can express this by saying that, in string theory, first quantization provides the UV cutoff needed in order to make second quantization well defined. Furthermore, in quantum string theory (QST), there are definite hopes to be able to compute both M_P and m_p (in the above string units, i.e. as lengths) in terms of λ_s, c and of a dimensionless parameter, the string coupling (see below).

The situation here reminds me of that of pure quantum gravity. As noticed by Novikov and Zeldovich [19, part V, ch. 23, par. 19], such a theory would only contain *two* fundamental units, c, and the Planck length $l_P = \sqrt{G_N \hbar c^{-3}}$, but not $\hbar$ and G_N separately. We may view string theory as an extension of GR that allows the introduction of all elementary particles and all fundamental forces in a geometrical way. No wonder then to find that only geometrical units are necessary.

Let us consider for instance, within the string theory framework, the gravitational acceleration a_2 induced by a string of length L_1 on a string of length L_2 sitting at a distance r from it. A simple calculation gives (for $r \gg L_1, L_2$):

$$a_2 = g_s^2 c^2 \left(\frac{L_1}{r^2} \right), \tag{4.2}$$

where g_s is the (dimensionless!) string coupling discussed in the next section. Clearly, the answer does not contain anything else but geometrical quantities and a pure number.

Another more familiar example is the computation of the energy levels of atoms in terms of the electron mass, its charge, and $\hbar$. These are given, to lowest order in α, by

$$E_n = -\frac{1}{2n^2} m_e \left(\frac{e^2}{\hbar} \right)^2 = -\frac{1}{2n^2} (m_e c^2) \alpha^2 \tag{4.3}$$

Weinberg argues, convincingly I think, that the quantities E_n are less fundamental than the electron charge, mass and $\hbar$. However, if we argue that what we are really measuring are not energies by themselves, but the transition frequencies

$$\omega_{mn} = \frac{1}{\hbar}(E_m - E_n) = \frac{1}{2} \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \frac{\alpha^2 c}{\lambda_s} \epsilon_e \, , \tag{4.4}$$

we see that, once more, only c and λ_s, and some in principle calculable dimensionless ratios (such as the electron mass in string units, $\epsilon_e = m_e/M_s$), appear in the answer [14]. Obviously, if we follow Weinberg's definition, λ_s and λ_s/c, and *not* for instance c/ω_{12} and $1/\omega_{12}$ (which are like the "modern" units of length, and time), play the role of fundamental units of length and time.

5. Fundamental units in modern QST/M-theory

We now turn to the same set of questions within the context of first-quantized string theory in the presence of background fields. Here I will attempt to give A_2 and A_3 together. The beautiful feature of this formulation is that all possible parameters of string theory, dimensionful and dimensionless alike, are replaced by background fields whose vacuum expectation values (VEV) we hope to be able to determine dynamically. As a prototype, consider the bosonic string in a gravi-dilaton background. The dimensionless action (i.e. the action divided by $\hbar$ in more conventional notation) reads:

$$S = \frac{1}{2}\int \sqrt{-\gamma}\left(\gamma^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X^\nu G_{\mu\nu}(X) + R(\gamma)\phi(X)\right) d^2z \tag{5.1}$$

where $X^\mu = X^\mu(\sigma,\tau)$, $\mu = 0, 1, \ldots, D-1$, are the string coordinates as functions of the world-sheet coordinates $z = (\sigma, \tau)$, with respect to which the the partial derivatives are defined. Furthermore, $G_{\mu\nu}$ is the so-called string metric and ϕ is the so-called dilaton. Finally, $\gamma_{\alpha\beta}$ and $R(\gamma)$ are, respectively, the metric and scalar curvature of the two-dimensional Riemann surface having coordinates σ and τ. ϕ is clearly dimensionless, while the dimensions of the metric components $G_{\mu\nu}$ are such that $G_{\mu\nu}X^\mu X^\nu$ is also dimensionless.

The exponential of the expectation value of ϕ gives the dimensionless parameter — known as the string coupling g_s — that controls the strength of all interactions (e.g. α) and thus also the string-loop expansion. Instead, the expectation value of $G_{\mu\nu}$ converts lengths and time intervals into pure numbers. Thus, through its non trivial dimension, the *metric* $G_{\mu\nu}$ actually provides the *metre/clock*, i.e. the fundamental units of space and time that we are after.

If the VEV of $G_{\mu\nu}$ is proportional to $\eta_{\mu\nu}$, the flat minkowskian metric, then it will automatically introduce the constants c and λ_s of the previous section via:

$$\langle G_{\mu\nu}(X)\rangle = \mathrm{diag}\left(-c^2\lambda_s^{-2}, \lambda_s^{-2}, \ldots\right) \tag{5.2}$$

The mere finiteness of c and λ_s is clearly of fundamental importance. However, in our context, the real question is: do the actual values of c and λ_s mean something (in the same way in which the actual value of $\langle\phi\rangle$ does)? What is, in other words, the difference between dimensionful and dimensionless constants? The answer is a bit subtle. String theory should allow to compute α in terms of the VEV of ϕ. Similarly, it should allow to compute $(\Delta X)^2 \equiv G_{\mu\nu}\Delta X^\mu \Delta X^\nu$ for some physical length ΔX^μ (say for the Hydrogen atom). Calling that pure number so many centimetres would fix the string length parameter in cm but, of course, this would be just a convention: the truly convention-independent (physical) quantity is just $(\Delta X)^2$. Both $\langle\phi\rangle$ and $(\Delta X)^2$ are pure numbers whose possible values distinguish one theory (or one vacuum) from another.

The difference between the two kinds of constants, if any, simply stems from the fact that, while different values of $\langle\phi\rangle$ (or α) define genuinely different theories, values of $\langle G_{\mu\nu}\rangle$ that are related by a General Coordinate Transformation (GCT) can be compensated by a GCT on X and thus define the same theory as long as $(\Delta X)^2$ remains the same. In particular, if $\langle G_{\mu\nu}\rangle \sim \eta_{\mu\nu}$ as in the example discussed above, the actual proportionality constants c and λ_s appearing in (5.2) can be reabsorbed by a GCT. This is why it does not make sense to talk about the *absolute* values of c and λ_s or to compare them to those of an alien: only the dimensionless numbers $(\Delta X)^2$, i.e. the values of some physical length or speed in those units are physically relevant and can be compared (see section 7).

The situation would be very different if $\langle G_{\mu\nu}\rangle$ would not be reducible to $\eta_{\mu\nu}$ via a GCT. That would mean a really different world, like one with a different value of α. In ref. [20] I gave the example of $\langle G_{\mu\nu} \rangle$ proportional to the de-Sitter metric, stressing the fact that, in such a vacuum, even λ_s disappears in favour of a dimensionless parameter similar to $\langle \phi \rangle$. Thus, as stressed in [15, 16], my early statement in [14] about having just two constants should be considered valid if the vacuum of QST is minkowskian, in particular in the NG formulation of QST.

To summarize, QM provides, through the string metric $G_{\mu\nu}$, a truly fundamental metre/clock allowing us to reduce space-time distances to pure numbers whose absolute value *is* physically meaningful. Note, incidentally, that in Classical GR only $g_{\mu\nu}\Delta X^{\mu}\Delta X^{\nu}$ is an invariant. However, in the classical case (and even for classical strings), only ratios of quantities of this type matter while in QST, $(\Delta X)^2$ is, for each single ΔX, a meaningful pure number.

In conclusion, I still stand by my remark in [15] that the fundamental constants of Nature are, in QST, the constants of the vacuum. How many (physically distinct) choices of its VEV's does QST allow? We now believe that all known consistent string theories correspond to perturbations around different vacua of a single, yet unknown, "M-theory". We still do not know, however, how many physically inequivalent non-perturbative vacua M-theory has. Until then, I do not think we can really answer the question of fundamental units in QST, but I would be very surprised if, in any consistent string vacuum, we would find that we need more than one unit of length and one of time.

6. The disagreement with Lev

Lev cannot accept (part I) that $\hbar$ has disappeared from the list. He claims that, without $\hbar$, there is no unit of momentum, of energy, and, especially, of angular momentum. But, as I said in the previous two sections, $\hbar$ has not really disappeared: it has actually been promoted, in string theory, to a grander role, that of providing also, through QM, an UV cutoff that hopefully removes both the infinities of QFT and ordinary Quantum Gravity and the ubiquitous singularities of Classical GR.

I would concede, however, that, given the fact that momentum and energy are logically distinct from lengths and times for ordinary objects, insisting on the use of the same (or of reciprocal) units for both sets can be pedagogically confusing. Therefore I do agree that the set c, $\hbar$, and λ_s define at present, within QST, the most practical (though not the most economical) set of fundamental units.

To rephrase myself: within the NG action there seems to be no reason to introduce a tension T or $\hbar$. The action is naturally *the* area and the Planck constant is the unit of area needed to convert the action into a number. However, by the standard definition of canonically conjugate variables, this would lead to identical dimensions for momenta and lengths (or for times and energies). For strings that's fine, since we can identify the energy of a string with its length, but when it comes to ordinary objects, i.e. to complicated bound states of fundamental strings or branes, it looks less confusing to give momentum a unit other than length. In order to do that we introduce, somewhat artificially, a conversion factor, the string tension T, so that energies are now measured in ergs, in GeV, or whatever we wish, different choices being related by irrelevant redefinitions of T.

7. The disagreement with Mike

Two issues appear to separate Mike's position from my own:

- The alien story

  Mike quotes an example, due to Feynman, on how we could possibly tell an alien to distinguish left from right. Then he asks: can we similarly communicate to an alien our values for c and λ_s and check whether they agree with ours? I claim the answer to be: yes, we can, and, to the same extent that the alien will be able to tell us whether her[3] α agrees with ours, she will also be able to tell us whether her c and λ_s agree with ours.

  In order to do that, we "simply" have to give the alien *our* definitions of cm. and s. in terms of a physical system she can possibly identify (say the H atom) and ask: which are your values of c and λ_s in these units? If the alien cannot even identify the system then she lives in a different world/string-vacuum; if she does, then she should come up with the same numbers (e.g. $c = 3 \times 10^{10}$ cm/s) or else, again, her world is not like ours. It thus looks to me that the alien story supports the idea that we do have, in our own world, some fundamental units of length and time. Mike seems to agree with me on the alien's reply, but then concludes that c is *not* a fundamental unit because a completely rescaled world, in which both c and the velocity of the electron in the H atom are twice as large, is indistinguishable from ours. I conclude, instead, that c *is* a fundamental unit because the velocity of our electron in units of c *is* a relevant number to be compared with the alien's.

  Incidentally, the same argument can be applied either to some ancestors (or descendants) of ours, or to inequivalent string vacua. A value of c in cm/s for any of those which differs from ours would really mean different worlds, e.g. worlds with different ratios of the velocity of the electron in the Hydrogen atom and the speed of light. We may *either* express this by saying that, in the two different worlds, c is different in atomic units, *or* by saying that c is the same but atomic properties differ. No experimental result will be able to distinguish about these two physically equivalent statements since a rescaling of all velocities is inessential.

- Reducing fundamental units to conversion factors

  Mike's second point is that these units can be used as conversion factors, like k_B, in order to convert any quantity into any other and, eventually, everything into pure numbers. However, I do insist that the point is *not* to convert degrees Kelvin into MeV, centimetres into seconds, or everything into numbers. The important point is that there are units that are arbitrary and units that are fundamental in the sense that, when a quantity becomes $O(1)$ in the latter units, dramatic new phenomena occur. It makes a huge difference, for instance, having or not having a fundamental length. Without a fundamental length, properties of physical systems would

[3]To stress that my alien's reaction is different from that of Mike's alien I have also changed the alien's gender.

be invariant under an overall rescaling of their size, atoms would not have a characteristic size, and we would be unable to tell the alien which atom to use as a metre. By contrast, with a fundamental quantum unit of length, we can meaningfully talk about short or large distances (as compared to the fundamental length, of course).

Going back to the discussion at the end of section 5, the pure number $(\Delta X)^2$ has a meaning in itself. In the absence of any fundamental units of length and time I would be able to rescale this number arbitrarily (e.g. by rescaling $G_{\mu\nu}$) without changing physics. Only ratios of two lengths in the problem, like $(\Delta X_1)^2/(\Delta X_2)^2$ would matter. Because of QM, however, there is a fundamental rod (and clock) that gives, out of any *single* physical length or time interval, *a relevant* pure number.

On this particular point, therefore, I tend to agree with Lev. There is, in relativity, a fundamental unit of speed (its maximal value); there is, in QM, a fundamental unit of action (a minimal uncertainty); there is, in string theory, a fundamental unit of length (the characteristic size of strings). QST appears to provide the missing third fundamental unit of the three-constants system. These three units form a very convenient system except that, classically, the units of action are completely arbitrary (and the same is true therefore of mass, energy etc.), while, quantum mechanically, only $S/\hbar$ matters. In string theory this allows us to identify the Planck constant with the string length eliminating the necessity, but perhaps not the convenience, of a third unit besides those needed to measure lengths and time intervals.

I also agree with Mike that all that matters are pure numbers. As I stressed in section 2, it is easy to convert any quantity into a pure number by choosing arbitrarily some unit. I only add to this the observation that relativity and quantum mechanics provide, in string theory, units of length and time which look, at present, more fundamental than any other. The number of distinct physical quantities (and of corresponding units) is a matter of choice and convenience, and also depends on our understanding of the underlying physical laws. Within QFT + GR it looks most useful to reduce this number to three, but there is no obvious candidate for the third unit after c and $\hbar$. With QST, the third unit naturally emerges as being the string length λ_s. However there appears an interesting option to do away with $\hbar$. Going further down, say from two to one or to zero, means considering space as being *equivalent* to time or as both being *equivalent* to pure numbers, while, keeping the two units c and λ_s, allows to *express* space and time intervals *in terms* of pure numbers.

This is what distinguishes, in my opinion, fundamental units from conversion factors. While I see no reason to distinguish the units of temperature from those of energy, and thus to introduce Boltzmann's constant, I see every reason to distinguish space from time and to introduce c as a fundamental unit of speed and *not* as a trivial conversion factor. Another clear difference is that, while the ratio $E(T)/T$ is always the same, we do observe, in Nature, a variety of speeds (all less than c, so far), of lengths, and of frequencies.

8. Time variation of fundamental units?

I think that the above discussion clearly indicates that the "time variation of a fundamental unit", like c, has no meaning, unless we specify what else, having the same units, is kept fixed. Only the time variation of dimensionless constants, such as α or $(\Delta X)^2$ for an atom have an intrinsic physical meaning.

We do believe, for instance, that in a cosmological background the variation in time of $G_{\mu\nu}$ is accompanied by a corresponding variation of the ΔX^μ of an atom so that $(\Delta X)^2$ remains constant. The same is usually assumed to be true for α. However, this is not at all an absolute theoretical necessity (e.g. α can depend on time, in QST, if ϕ does), and should be (and indeed is being) tested. For instance, the same $(\Delta X)^2$ is believed to grow with the expansion of the Universe if ΔX^μ represents the wavelength of light coming to us from a distant galaxy. The observed red shift only checks the *relative* time-dependence of $(\Delta X)^2$ for an atom and for the light coming from the galaxy.

However, I claim that, in principle, the time variation of $(\Delta X)^2$ has a physical meaning for each one of the two systems separately because it represents the time variation of some physical length w.r.t. the fundamental unit provided by string theory. For instance, in the early Universe, this quantity for the CMBR photons was much smaller than it is today ($O(10^{30})$). If it ever approached values $O(1)$, this may have left an imprint of short-distance physics on the CMBR spectrum.

Acknowledgments

I am grateful to Lev and Mike for having given their serious thoughts to this issue and for pushing me to clarify my point of view to them, and to myself. I also wish to acknowledge the support of a "Chaire Internationale Blaise Pascal", administered by the "Fondation de L'Ecole Normale Supérieure", during the final stages of this work.

Part III
A party political broadcast on behalf of the Zero Constants Party

Michael J. Duff

Abstract. According to the manifesto of Okun's Three Constants Party, there are three fundamental dimensionful constants in Nature: Planck's constant, $\hbar$, the velocity of light, c, and Newton's constant, G. According to Veneziano's Two Constants Party, there are only two: the string length λ_2 and c. Here we present the platform of the Zero Constants Party.

1. The false propaganda of the Three Constants Party

As a young student of physics in high school, I was taught that there were three basic quantities in Nature: Length, Mass and Time [21]. All other quantities, such as electric charge or temperature, occupied a lesser status since they could all be re-expressed in terms of these basic three. As a result, there were three basic units: centimetres, grams and seconds, reflected in the three-letter name "CGS" system (or perhaps metres, kilograms and seconds in the alternative, but still three-letter, "MKS" system).

Later, as an undergraduate student, I learned quantum mechanics, special relativity and newtonian gravity. In quantum mechanics, there was a minimum quantum of action given by Planck's constant $\hbar$; in special relativity there was a maximum velocity given by the velocity of light c; in classical gravity the strength of the force between two objects was determined by Newton's constant of gravitation G. In terms of length, mass, and time their dimensions are

$$\begin{aligned} [c] &= LT^{-1} \\ [\hbar] &= L^2MT^{-1} \\ [G] &= L^3M^{-1}T^{-2} . \end{aligned} \tag{1.1}$$

Once again, the number three seemed important and other dimensionful constants, such as the charge of the electron e or Boltzmann's constant k, were somehow accorded a less fundamental role. This fitted in perfectly with my high school prejudices and it seemed entirely natural, therefore, to be told that these three dimensionful constants determined three basic units, first identified a century ago by Max Planck, namely the Planck length L_P, the Planck mass M_P and the Planck time T_P:

$$\begin{aligned} L_P &= \sqrt{G\hbar/c^3} = 1.616 \times 10^{-35}\,\mathrm{m} \\ M_P &= \sqrt{\hbar c/G} = 2.177 \times 10^{-8}\,\mathrm{kg} \\ T_P &= \sqrt{G\hbar/c^5} = 5.390 \times 10^{-44}\,\mathrm{s} \end{aligned} \tag{1.2}$$

Yet later, researching into quantum gravity which attempts to combine quantum mechanics, relativity and gravitation into a coherent unified framework, I learned about the

Bronshtein-Zelmanov-Okun (BZO) cube [10], with axes $\hbar$, c^{-1} and G, which neatly summarizes how classical mechanics, non-relativistic quantum mechanics, newtonian gravity and relativistic quantum field theory can be regarded respectively as the $(\hbar, c^{-1}, G) \to 0$, $(c^{-1}, G) \to 0$, $(\hbar, c^{-1}) \to 0$, and $(G) \to 0$ limits of the full quantum gravity. Note, once again that we are dealing with a three-dimensional cube rather than a square or some figure of a different dimension.

What about Kaluza-Klein theories which allow for $D > 4$ spacetime dimensions? Unlike $\hbar$ and c, the dimensions of G depend on D:

$$[G_D] = M^{-1}L^{D-1}T^{-2} \tag{1.3}$$

and hence (dropping the P subscript), the D-dimensional Planck length L_D, mass M_D and time T_D are given by

$$\begin{aligned} {L_D}^{D-2} &= G_D \hbar c^{-3} \\ {M_D}^{D-2} &= {G_D}^{-1} \hbar^{D-3} c^{5-D} \\ {T_D}^{D-2} &= G_D \hbar c^{-1-D} . \end{aligned} \tag{1.4}$$

After compactification to four dimensions, $G \equiv G_4$ then appears as

$$\frac{1}{G_4} = \frac{1}{G_D} V , \tag{1.5}$$

where V is the volume of the compactifying manifold. Since V has the four-dimensional interpretation as the vacuum expectation value of scalar modulus fields coming from the internal components of the metric tensor, it depends on the choice of vacuum but does not introduce any more fundamental constants into the lagrangian.

Adherents of this conventional view of the fundamental constants of Nature have been dubbed the "Three Constants Party" by Gabriele Veneziano [16]. Lev Okun is their leader. Until recently I was myself, I must confess, a card-carrying member.[4]

2. The false propaganda of the Two Constants Party

My faith in the dogma was shaken, however, by papers by Gabriele [14, 15, 16], self-styled leader of the rebel Two Constants Party. As a string theorist, Gabriele begins with the two-dimensional Nambu-Goto action of a string. He notes that, apart from the velocity of light still needed to convert the time coordinate t to a length coordinate $x^0 = ct$, the action divided by $\hbar$ requires only one dimensionful parameter, the string length λ_2 (denoted λ_s by Gabriele).

$${\lambda_2}^2 = \frac{\hbar}{cT_2} , \tag{2.1}$$

[4] It seems that the choice of length, mass and time as the three basic units is due to Gauss [27], so we could declare him to be the founder of the Three Constants Party, although this was long before the significance of c and $\hbar$ was appreciated.

where $T_2 = 1/2\pi c\alpha'$ is the tension of the string and α' is the Regge slope. This is because the Nambu-Goto action takes the form

$$\frac{S_2}{\hbar} = \frac{1}{{\lambda_2}^2}\text{Area} \tag{2.2}$$

So if this were to describe the theory of everything (TOE), then the TOE would require only two fundamental dimensionful constants c and λ_2. In superstring theory, the ten-dimensional Planck length is given in terms of the string length λ_2 and the vacuum expectation value of the dilaton field ϕ

$${L_{10}}^2 = {\lambda_2}^2\langle e^{\phi}\rangle \tag{2.3}$$

Once again, the vev of ϕ will be different in different vacua but does not introduce any new constants into the lagrangian.

A similar argument for reducing the three constants h, c, G to just two was made previously by Zeldovich and Novikov [19] with regard to quantum gravity. The Einstein-Hilbert action divided by $\hbar$ involves G and $\hbar$ only in the combination $G\hbar$ appearing in the square of the Planck length, and so we need only L_P and c. Of course quantum gravity does not pretend to be the TOE and so this argument still leaves open the number of dimensionful constants required for a TOE.

In the light of the 1995 M-theory [22] revolution, we might wish to update Gabriele's argument by starting with the corresponding three-dimensional action for the M2-brane,

$$\frac{S_3}{\hbar} = \frac{1}{{\lambda_3}^3}(\text{3d-volume})\,, \tag{2.4}$$

where the corresponding parameter is the membrane length λ_3.

$${\lambda_3}^3 = \hbar/cT_3 \tag{2.5}$$

and where T_3 is the membrane tension. Alternatively, we could start with the six-dimensional action of the dual M5-brane,

$$\frac{S_6}{\hbar} = \frac{1}{{\lambda_6}^6}(\text{6d-volume}) \tag{2.6}$$

where the corresponding parameter is the fivebrane length λ_6

$${\lambda_6}^6 = \frac{\hbar}{cT_6} \tag{2.7}$$

and where T_6 is the fivebrane tension. Eleven-dimensional M-theory is, in fact, simpler than ten-dimensional superstring theory in this respect, since there is no dilaton. Consequently, the three lengths: membrane length, fivebrane length and eleven-dimensional Planck length are all equal [23] up to calculable numerical factors $\lambda_3 \sim \lambda_6 \sim L_{11}$. So the fundamental length in M-theory is λ_3 rather than λ_2 and will be shorter for string coupling less than unity [26].

However, even if we substitute λ_3 for λ_2, Gabriele would say that we are still left with the number two. This also reduces the number of basic units to just two: length and time.

Gabriele's claim led to many heated discussions in the CERN cafeteria between Lev, Gabriele and myself. We went round and round in circles. Back at Texas A&M, I continued these arguments at lunchtime conversations with Chris Pope and others. There at the College Station Hilton, we eventually reached a consensus and joined what Gabriele would call the Zero Constants Party [16].

Our attitude was basically that $\hbar$, c and G are nothing but conversion factors e.g. mass to length, in the formula for the Schwarzschild radius R_S

$$R_S = \frac{2Gm}{c^2},$$

or energy to frequency

$$E = \hbar\omega$$

energy to mass

$$E = mc^2$$

no different from Boltzmann's constant, say, which relates energy to temperature

$$E = kT\,.$$

As such, you may have as many so-called "fundamental" constants as you like; the more different units you employ, the more different constants you need.[5] Indeed, no less an authority than the *Conférence Générale des Poids et Mesures*, the international body that administers the SI system of units, adheres to what might be called the Seven Constants Party, decreeing that seven units are "basic": metre(length), kilogram (mass), second (time), ampere (electric current), kelvin (thermodynamic temperature), mole (amount of substance), candela (luminous intensity), while the rest are "derived" [27, 28]. The attitude of the Zero Constants Party is that the most economical choice is to use natural units where there are no conversion factors at all. Consequently, none of these units or conversion factors is fundamental.

Incidentally, Lev (part I) objects in his section 5 that equations such as $c = 1$ cannot be taken literally because c has dimensions. In my view, this apparent contradiction arises from trying to use two different sets of units at the same time, and really goes to the heart of my disagreement with Lev about what is real physics and what is mere convention. In the units favored by members of the Three Constants Party, length and time have different dimensions and you cannot, therefore, put $c = 1$ (just as you cannot put $k = 1$, if you want to follow the conventions of the Seven Constants Party). If you want to put $c = 1$, you must trade in your membership card for that of (or at least adopt the habits of) the Two Constants Party, whose favorite units do not distinguish length from time.[6] In these units, c *is* dimensionless and you may quite literally set it equal to one. In the natural units favored by the Zero Constants Party, there are no dimensions at all and $\hbar = c = G = \cdots = 1$ may

[5] In this respect, I take the the number of dimensionful fundamental constants to be synonymous with the number of fundamental (or basic) units.

[6] This $(\hbar, G)$ wing of the Two Constants Party is different from Gabriele's (c, λ_2) wing, which prefers not to introduce a separate unit for mass.

be imposed literally and without contradiction. With this understanding, I will still refer to constants which have dimensions in some units, such as $\hbar, c, G, k \ldots$, as "dimensionful constants" so as to distinguish them from constants such as α, which are dimensionless in any units.

3. Three fundamental theories?

Lev and Gabriele remain unshaken in their beliefs, however. Lev (part I) makes the, at first sight reasonable, point (echoed by Gabriele in part II) that $\hbar$ is more than just a conversion factor. It embodies a fundamental physical principle of quantum mechanics that there is a minimum non-zero angular momentum. Similarly, c embodies a fundamental physical principle of special relativity that there is a maximum velocity c. If I could paraphrase Lev's point of view it might be to say that there are three "fundamental" units because there are three fundamental physical theories: quantum mechanics, special relativity and gravity. According to this point of view, temperature, for example, should not be included as a basic unit (or, equivalently, Boltzmann's constant should not be included as a fundamental constant.)

However, I think this elevation of $\hbar$, c and G to a special status is misleading. For example, the appearance of c in $x^0 = ct$ is for the benefit of people for whom treating time as a fourth dimension is unfamiliar. But once you have accepted $O(3,1)$ as a symmetry the conversion factor becomes irrelevant. We have become so used to accepting $O(3)$ as a symmetry that we would not dream of using different units for the three space coordinates,[7] but to be perverse we could do so.

To drive this point home, and inspired by the *Conférence Générale des Poids et Mesures*, let us introduce three new superfluous units: xylophones, yachts and zebras to measure intervals along the x, y and z axes. This requires the introduction of three superfluous "fundamental" constants, c_x, c_y and c_z with dimensions length/xylophone, length/yacht and length/zebra, respectively, so that the line element becomes:

$$ds^2 = -c^2dt^2 + {c_x}^2dx^2 + {c_y}^2dy^2 + {c_z}^2dz^2 \,. \tag{3.1}$$

Lev's point is that the finiteness of c ensures that we have $O(3,1)$ symmetry rather than merely $O(3)$. This is certainly true. But it is equally true that the finiteness of c_x, say, ensures that we have $O(3,1)$ rather than merely $O(2,1)$. In this respect, the conversion factors c and c_x are on an equal footing.[8] Both are, in Gabriele's terminology (part II), equally "silly". Both can be set equal to unity and forgotten about.

Similarly, the "fundamental" lengths λ_d appearing in brane actions (2.2), (2.4) and (2.6) can be removed from the equations by defining new dimensionless worldvolume coordinates, ξ', related to the old ones, ξ, by $\xi = \lambda_d\xi'$.

[7] I am grateful to Chris Pope for this example.

[8] To put this more rigorously, the Poincaré group admits a Wigner-Inönü contraction to the Galileo group, obtained by taking the $c \to \infty$ limit. However, this is by no means unique. There are other contractions to other subgroups. For example, one is obtained by taking the $c_x \to \infty$ limit. Although of less historical importance, these other subgroups are mathematically on the same footing as the Galileo group. So, in my opinion, the singling out of c for special treatment has more to do with psychology than physics.

So I would agree with Lev that the finiteness of the conversion factors is important (minimum angular momentum, maximum velocity) but, in my view, no significance should be attached to their value and you can have as many or as few of them as you like.

The reason why we have so many different units, and hence conversion factors, in the first place is that, historically, physicists used different kinds of measuring apparatus: rods, scales, clocks, thermometres, electroscopes etc. Another way to ask what is the mimimum number of basic units, therefore, is to ask what is, in principle, the minimum number of basic pieces of apparatus.[9] Probably Lev, Gabriele and I would agree that $E = kT$ means that we can dispense with thermometers, that temperature is not a basic unit and that Boltzmann's constant is not fundamental. Let us agree with Lev that we can whittle things down to length, mass and time or rods, scales and clocks. Can we go further? Another way to argue that the conversion factor c should not be treated as fundamental, for example, is to point out that once the finiteness of c has been accepted, we do not need both clocks and rulers. Clocks alone are sufficient since distances can be measured by the time it takes light to travel that distance, $x = ct$. We are, in effect, doing just that when we measure interstellar distances in light-years. Conversely, we may do away with clocks in favor of rulers. It is thus superfluous to have both length and time as basic units. Similarly, we can do away with rulers as basic apparatus and length as a basic unit by trading distances with masses using the formula for the Compton wavelength $R_C = h/mc$. Indeed, particle theorists typically express length, mass and time units as inverse mass, mass and inverse mass, respectively. Finally, we can do away with scales by expressing particle masses as dimensionless numbers, namely the ratio of a particle mass to that of a black hole whose Compton wavelength equals its Schwarzschild radius. So in this sense, the black hole acts as our rod, scale, clock, thermometer etc. all at the same time. In practice, the net result is as though we set $\hbar = c = G = \cdots = 1$ but we need not use that language.

J-M. Levy-LeBlond [29] puts it like this: "This, then, is the ordinary fate of universal constants: to see their nature as concept synthesizers be progressively incorporated into the implicit common background of physical ideas, then to play a role of mere unit conversion factors and often to be finally forgotten altogether by a suitable redefinition of physical units."

4. An operational definition

"If, however, we imagine other worlds, with the same physical laws as those of our own world, but with different numerical values for the physical constants determining the limits of applicability of the old concepts, the new and correct concepts of space, time and motion, at which modern science arrives only after very long and elaborate investigations, would become a matter of common knowledge."

George Gamow, *Mr. Tompkins in paperback* [18]

It seems to me that this issue of what is fundamental will continue to go round and around until we can all agree on an operational definition of "fundamental constants".

[9] I am grateful to Chris Isham for this suggestion.

Weinberg [1] defines constants to be fundamental if we cannot calculate their values in terms of more fundamental constants, not just because the calculation is too hard, but because we do not know of anything more fundamental. This definition is fine, but does not resolve the dispute between Gabriele, Lev and me. It is the purpose of this section to propose one that does. I will conclude that, according to this definition, the dimensionless parameters, such as the fine structure constant, are fundamental, whereas all dimensionful constants, including $\hbar$, c and G, are not.[10]

In physics, we frequently encounter ambiguities such as "left or right" and "matter or antimatter". Let us begin by recalling Feynman's way of discriminating between what are genuine differences and what are mere conventions. Feynman imagines that we can communicate with some alien being [30]. If it were not for the violation of parity in the weak interactions we would have no way of deciding whether what he[11] calls right and left are the same as what we call right and left. However, we can ask him to perform a cobalt 60 experiment and tell him that the spinning electrons determine a left handed thread. In this way we can agree on what is left and right. When we eventually meet the alien, of course, we should beware shaking hands with him if he holds out his left hand (or tentacle). He would be made of antimatter and annihilate with us! Fortunately, after the discovery of CP violation we could also eliminate this ambiguity.

In a similar vein, let us ask whether there are any experiments that can be performed which would tell us whether the alien's universe has the same or different constants of nature as ours. If the answer is yes, we shall define these constants to be fundamental, otherwise not. In particular, and inspired by Gamow's Mr. Tompkins [18], we will ask whether there is in principle any experimental difference that would allow us to conclude unambiguously that his velocity of light, his Planck's constant or his Newton's constant are different from ours. By "unambiguously" I mean that no perceived difference could be explained away by a difference in conventions. (Of course, even Feynman's criterion is not devoid of theoretical assumptions. We have to assume that the cobalt behaves the same way for the alien as for us etc. To be concrete, we might imagine that we are both described by a TOE (perhaps M-theory) in which the fundamental constants are given by vacuum expectation values of scalar fields. The alien and we thus share the same lagrangian but live in possibly different vacua. Let us further assume that both vacua respect $O(3,1)$ symmetry.)

5. The operationally indistinguishable world of Mr. Tompkins

The idea of imagining a universe with different constants is not new, but, in my opinion, the early literature is very confusing. For example, Vol'berg [17] and Gamow [18] imagine a universe in which the velocity of light is different from ours, say by ten orders of magnitude, and describe all sorts of weird effects that would result:

[10] My apologies to those readers to whom this was already blindingly obvious. A similar point of view may be found in [32]. On the other hand, I once read a letter in Physics World from a respectable physicist who believed that a legitimate ambition of a TOE would be to calculate the numerical value of $\hbar$.

[11] I will follow Feynman and assume that the alien is a "he", without resolving the "he or she" ambiguity.

"The initials of Mr. Tompkins originated from three fundamental physical constants: the velocity of light c; the gravitational constant G; and the quantum constant h, which have to be changed by immensely large factors in order to make their effect easily noticeable by the man on the street."

George Gamow, *Mr. Tompkins in paperback* [18]

In this one sentence, Gamow manages to encapsulate everything I am objecting to! First, he takes it as axiomatic that there are three fundamental constants. Second, he assumes a change in these constants can be operationally defined. I for one am mystified by such comparisons. After all, an inhabitant of such a universe (let us identify him with Feynman's alien) is perfectly free to choose units in which $c = 1$, just as we are. To use the equation

$$k = \frac{E}{c}$$

to argue that in his universe, for the same energy E, the photon emitted by an atom would have a momentum k that is ten orders of magnitude smaller than ours is, to my mind, meaningless. There is no experimental information that we and the alien could exchange that would allow us to draw any conclusion.

By contrast, in his critique of Vol'berg and Gamow, Lev [10] imagines a universe in which the binding energy of an electron in a hydrogen atom $E = me^4/\hbar^2$ exceeds twice the electron rest energy $2mc^2$, where m and e are the electron mass and charge respectively. In such a universe it would be energetically favorable for the decay of the proton to a hydrogen atom and a positron $p \to H + e^+$. This universe is demonstrably different from ours. But, in my opinion, the correct conclusion has nothing to do with the speed of light, but simply that in this universe the dimensionless fine structure constant $\alpha = e^2/\hbar c$ exceeds $\sqrt{2}$.

I believe that these two examples illustrate a general truth: no experimental information that we and the alien could exchange can unambiguously determine a difference in dimensionful quantities. No matter whether they are the $\hbar$, c and G sacred to the Three Constants Party, the λ_2 and c of the Two Constants Party or the seven constants of the *Conférence Générale des Poids et Mesures*. Any perceived difference are all merely differences in convention rather than substance. By contrast, differences in dimensionless parameters like the fine structure constants are physically significant and meaningful.[12] Of course, our current knowledge of the TOE is insufficient to tell us how many such dimensionless constants Nature requires. There are 19 in the Standard model, but the aim of M-theory is to reduce this number. Whether they are all calculable or whether some are the result of cosmological accidents (like the ratios of distances of planets to the sun) remains one of the top unanswered questions in fundamental physics.[13]

[12] In his section 7, Gabriele (part II) claims to disagree with me on this point, but I think the first two sentences of his section 8 indicate that we are actually in agreement. If, for example, the alien tells us that he observes the decay $p \to H + e^+$, then we can be sure that his α is different from ours. Choosing to attribute this effect (or any other effect) to a difference in c rather than $\hbar$ or e, however, is entirely a matter of convention, just as the difference between left and right would be a matter of convention in a world with no CP violation. So c fails the Feynman test.

[13] Indeed, participants of the Strings 2000 conference placed it in the top ten [24].

6. What about theories with time-varying constants?

Suppose that our "alien" came not from a different universe but from a different epoch in our own universe and we stumbled across his historical records. In this way of thinking, the issue of whether $\hbar$, c and G are fundamental devolves upon the issue of whether the results of any experiments could require the unambiguous conclusion that $\hbar$, c and G are changing in time. According to our criterion above, any such time-dependence would be merely convention, without physical significance.

On the other hand, many notable physicists, starting with Dirac [25], have nevertheless entertained the notion that G or c are changing in time. (For some reason, time-varying $\hbar$ is not as popular.) Indeed, papers on time-varying c are currently in vogue as as an alternative to inflation. I believe that these ideas, while not necessarily wrong, are frequently presented in a misleading way and that the time-variation in the physical laws is best described in terms of time-varying dimensionless ratios, rather than dimensionful constants.[14] So, in my opinion, one should talk about time variations in the dimensionless parameters of the standard model but not about time variations in $\hbar$, c and G. For example, any observed change in the strength of the gravitational force over cosmological times should be attributed to changing mass ratios rather than changing G. For example, the proton is approximately 10^{19} times lighter than the black hole discussed in section 3, whose Compton wavelength equals its Schwarzschild radius. It is then sensible to ask whether this dimensionless ratio could change over time.[15]

Unfortunately, this point was made insufficiently clear in the recent paper presenting astrophysical data suggesting a time-varying fine structure constant [34]. As a result, a front page article in the New York Times [35] announced that the speed of light might be changing over cosmic history.[16]

In the context of M-theory which starts out with no parameters at all, these standard model parameters would appear as vacuum expectation values of scalar fields.[17] Indeed, replacing parameters by scalar fields is the only sensible way I know to implement time varying constants of Nature. The role of scalar fields in determining the fundamental constants in a TOE was also emphasized by Gabriele [14, 15, 16].

7. Conclusions

The number and values of fundamental dimensionless constants appearing in a Theory of Everything is a legitimate subject of physical enquiry. By contrast, the number and values

[14]This point of view is also taken in [33].

[15]One could then sensibly discuss a change in the number of protons required before a star reaches its Chandrasekar limit for gravitational collapse. I am grateful to Fred Adams for this example.

[16]I am reminded of the old lady who, when questioned by the TV interviewer on whether she believed in global warming, responded: "If you ask me, it's all this changing from Fahrenheit to Centigrade that causing it!".

[17]The only other possibility compatible with maximal four-dimensional spacetime symmetry is the vacuum expectation value of a 4-index field strength. For example, the cosmological constant can receive a contribution from the vev of the M-theory 4-form [31].

of dimensionful constants, such as h, c, G,... is a quite arbitrary human construct, differing from one choice of units to the next. There is nothing magic about the choice of two, three or seven. The most economical choice is zero. Consequently, none of these dimensionful constants is fundamental.

Acknowledgments

My interest in this subject was fired by conversations with Lev Okun and Gabriele Veneziano. I am also grateful for discussions on fundamental constants with colleagues at Texas A&M, especially Chris Pope, at the University of Michigan, especially Gordy Kane, and at Imperial College, especially Chris Isham. I have enjoyed correspondence with Thomas Dent, Harald Fritsch, James Glanz, Robert C. Helling, Saibal Mitra, Jose Sande Lemos, Warren Siegel and Steve Weinberg. Alejandro Rivero points out that the Greek root of *dia* in *dialogue* means *through* and not *two* so *Trialogue* is misnamed. The hospitality of the High Energy Theory Group, Imperial College and the Mathematical Institute, Oxford, during the preparation of this article is gratefully acknowledged. This research was supported in part by DOE Grant DE-FG02-95ER40899.

Note added. Warren Siegel (private communication) makes the following interesting points:

1. Planck was actually a member of the Four Constants Party, since his original paper introduced not only a basic length, mass and time but also a temperature.[18]

2. In 1983, the *Conférence Générale des Poids et Mesures* declared c to have the value 299,792,458 metres/second exactly, *by definition*, thus emphasizing its role as a nothing but a conversion factor.[19]

3. Sailors use the perverse units of section 3, when they measure intervals along the x and y axes in nautical miles and intervals along the z axis in fathoms. The same observation was made independently by Steve Weinberg (private communication).

[18] By analyzing Planck's papers [3] Lev came to the conclusion that by adding k to c, h and G, Planck contradicts his definition of natural units [36].

[19] So asking whether the value of c has changed over cosmic history is like asking whether the number of litres to the gallon has changed.

References

[1] S. Weinberg, *Overview of theoretical prospects for understanding the values of fundamental constants*, in *The constants of physics*, W.H. McCrea and M.J. Rees eds., *Phil. Trans. R. Soc. London* **A310** (1983) 249.

[2] G.J. Stoney, *The philosophical magazine and journal of science*, **11** (1881) 381.

[3] M. Planck, *Über irrevesible Strahlungsvorgänge*, *S.-B. Preuss Akad. Wiss.* (1899) 440-480; *Ann. d. Phys.* **1** (1900) 69 reprinted in *Max Planck, Physikalische Abhandlungen und Vorträge, Band I. Friedr. Vieweg.* 1958, pp. 560–600, pp. 614–667.

[4] G. Gamov, D. Ivanenko and L. Landau, *Zh. Russ. Fiz. Khim. Obstva. Chast' Fiz.* **60** (1928) 13 (in Russian).

[5] M. Bronshtein, *K voprosu o vozmozhnoy teorii mira kak tselogo, (On a possible theory of the world as a whole)*, in *Osnovnye problemy kosmicheskoy fiziki (Basic problems of cosmic physics).* Kiev, ONTI (1934), pp. 186–218, in particular p. 210 (in Russian).

[6] M. Bronshtein, *Physikalische Zeitscrift der Sowjetunion* **9** (1936) 140.

[7] A. Zelmanov, *Kosmologia (Cosmology)*, in *Razvitie astronomii v SSSR (Development of astronomy in USSR)*, Nauka, Moscow, 1967, pp. 320–390, in particular p. 323 (in Russian).

[8] A. Zelmanov, *Mnogoobrazie materialnogo mira i problema beskonechnosti Vselennoi (Diversity of the material world and the problem of infinity of the Universe)*, in *Beskonechnost' i Vselennaya (Infinity and Universe)*, Mysl', Moscow, 1969, pp. 274–324, in particular p. 302 (in Russsian).

[9] G. Gorelik, *Razmernost' prostranstva (Dimension of space)*, MGU, Moscow, 1983, chapter 5 (in Russian).

[10] L.B. Okun, *The fundamental constants of physics*, *Sov. Phys. Usp.* **34** (1991) 818.

[11] *Symbols, units and nomenclature in physics: document U.I.P.* **20** (1978), International Union of Pure and Applied Physics, S.U.N. Commission.

[12] *Review of Particle Properties*, *Phys. Rev.* **D 45** (1992) Part II; page III.4.

[13] L.B. Okun, *Particle physics: the quest for the substance of substance*, Harwood academic publishers, Chur, New York 1985, Appendix 1. On systems of physical units.

[14] G. Veneziano, *A stringy nature needs just two constants*, *Europhys. Lett.* **2** (1986) 199.

[15] G. Veneziano, *Quantum strings and the constants of nature*, in *The challenging questions*, proceedings of the 27th Course of International School held July 26 - August 3, 1989, Erice, Italy, A. Zichichi ed., Plenum Press, New York 1990, p. 199.

[16] G. Veneziano, *Fundamental constants in field and string theory*, CERN-TH.6725/92, talk given at the 6th session of the International Workshop of Theoretical Physics, in *String quantum gravity and physics at the Planck scale*, Erice, Sicily, 21-28 June 1992, N. Sanchez and A. Zichichi eds., World Scientific 1993, p. 552.

[17] O.A. Volberg, *Zanimatel'naya progulka v strane Einshteina (An entertaining trip into Einstein's country)*, in *Zanimatel'naya Mekhanika (Entertaining Mechanics)*, Ya.I. Perel'man ed., GRN-PYuL, Leningrad - Moscow 1935, ch. 11, p. 188 (in Russian).

[18] G. Gamow, *Mr. Tompkins in paperback*, containing *Mr. Tompkins in wonderland* and *Mr. Tompkins explores the atom*, Cambridge University Press, Cambridge, England 1988.

[19] Ya.B. Zeldovich and I.D. Novikov, *The structure and evolution of the universe*, University of Chicago press 1982.

[20] G. Veneziano, *Physics with a fundamental length*, in *Physics and mathematics of strings, Vadim Knizhnik Memorial Volume*, L. Brink, D. Friedan and A.M. Polyakov eds., WSPC, 1990, p. 509.

[21] N. Feather, *Mass, length and time*, Edinburgh University Press, 1959.

[22] M.J. Duff, *The world in eleven dimensions: supergravity, supermembranes and M-theory*, I.O.P. Publishing 1999, `http://bookmark.iop.org/bookpge.htm/book=815p`.

[23] M.J. Duff, J.T. Liu and R. Minasian, *Eleven-dimensional origin of string/string duality: a one loop test*, *Nucl. Phys.* **B 452** (1995) 261.

[24] M.J. Duff, *The top ten questions in fundamental physics*, in *Strings 2000*, proceedings of the International superstrings conference, University of Michigan, Ann Arbor, M.J. Duff, J.T. Liu and Lu eds., World Scientific, 2001, `http://feynman.physics.lsa.umich.edu/strings2000/millennium.html`.

[25] P.A.M. Dirac, *The cosmological constants*, *Nature* **139** (1937).

[26] M.R. Douglas, D. Kabat, P. Pouliot and S.H. Shenker, *D-branes and short distances in string theory*, *Nucl. Phys.* **B 485** (1997) 85 [`hep-th/9608024`].

[27] The NIST reference on constants, units and uncertainty, `http://physics.nist.gov/cuu/Units/international.html`.

[28] Particle Data Group, *Review of particle properties*, *Phys. Rev.* **D 45** (1992) S1, part II; page III.4., erratum *ibid.* **D46** (1992) 5210.

[29] J.-M. Levy-Leblond, *The importance of being (a) constant*, in *Problems in the foundations of physics*, Enrico Fermi School LXXII, G. Torraldi ed., (North Holland 1979), p. 237.

[30] R.P. Feynman, *The character of physical law*, Penguin, 1992, p. 101.

[31] M.J. Duff and P. van Nieuwenhuizen, *Quantum inequivalence of different field representations*, *Phys. Lett.* **B 94** (1980) 179.

[32] J. Gribbin and P. Wesson, *Fickle constants of physics*, *New Scientist*, 4 July 1992, p. 30.

[33] J.D. Barrow and J. Magueijo, *Varying-α theories and solutions to the cosmological problems*, `astro-ph/9811072`.

[34] J.K. Webb et al., *Further evidence for cosmological evolution of the fine structure constant*, `astro-ph/0012539`.

[35] J. Glanz and D. Overbye, *Anything can change, it seems, even an immutable law of nature*, *New York Times*, August 15, 2001.

[36] L. Okun, *Cube or hypercube of natural units?*, `hep-ph/0112339`.

This page intentionally left blank

ELSEVIER

Nuclear Physics B (Proc. Suppl.) 110 (2002) 151–155

NUCLEAR PHYSICS B
PROCEEDINGS
SUPPLEMENTS

www.elsevier.com/locate/npe

Photons, Clocks, Gravity and the Concept of Mass

L.B.Okun[a] *

[a]ITEP, Moscow, 117218, Russia

Various aspects of the concept of mass in modern physics are discussed: 1. Mass in Special Relativity, 2. Atoms in Static Gravity, 3. Photons in Static Gravity, 4. Misleading Terminology, 5. Unsolved Problems of mass.

1. MASS IN SPECIAL RELATIVITY

1. The term mass was introduced by Newton in *Principia*, 1687: "Definition I: The quantity of matter is the measure of the same, arising from its density and bulk conjointly".

2. The principle of relativity was formulated by Galileo in *Dialogue*, 1632 (Galileo's ship).

3. His predecessor: Nicolaus Cusanus in "De docta ignorata" ("On the scientific ignorance"), 1440.

4. Velocity of light c was first measured by O.Roemer, 1676.

5. A. Michelson and E. Morley, 1887:
   c is the maximal velocity of signals. $c = 1$.

6. Special relativity: Interferometer of Michelson on the ship of Galileo.

7. Lorentz transformations of t, $\mathbf{r}$ and E, $\mathbf{p}$, where $\mathbf{r}$ and $\mathbf{p}$ are 3-vectors.

8. $m^2 = E^2 - \mathbf{p}^2$; E, $\mathbf{p}$ - 4-vector; m - scalar.

9. Einstein 1905: rest energy $E_0 = m$.

10. $E = m\gamma\ , \quad \mathbf{p} = m\gamma\mathbf{v} = E\mathbf{v}\ , \quad \mathbf{v} = \frac{\mathbf{p}}{E}$

$$v = |\mathbf{v}|\ , \quad \gamma = 1/\sqrt{1-v^2}\ , \quad v \le 1\ .$$

11. Photon: $m = 0$, $E = p$, $v = 1$, no rest frame.

12. System of two bodies:

$$E = E_1 + E_2\ , \quad \mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2\ ,$$

$$m^2 = E^2 - \mathbf{p}^2 = (E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2 =$$

$$= m_1^2 + m_2^2 + 2E_1E_2(1 - \mathbf{v}_1\mathbf{v}_2)\ .$$

   Two photons: $m_{2\gamma}^2 = 2E_1E_2(1-\cos\theta)$.

13. Mass of a system is conserved.
   Example $e^+e^- \to 2\gamma$ at rest.
   $m_{e^+e^-} = m_{2\gamma} = 2m_e$, though $m_\gamma = 0$.

14. Mass is not a measure of inertia.
   This measure is derived from: $d\mathbf{p}/dt = \mathbf{F}$, $\mathbf{p} = E\mathbf{v}$.

15. Mass is not additive at $v \ne 0$.
   $m = m_1 + m_2$ only at $v_1 = v_2 = 0$.

16. Mass is not a source and receptor of gravity at $v \ne 0$ (see item 20).

17. Mass is not a measure of an amount of matter. The concept of matter is much broader in relativity than in classical physics (examples: gas or flux of photons and/or neutrinos).

18. From mathematical point of view mass occupies a higher rank position in special relativity than in non-relativistic physics: it is a relativistic invariant.

19. From practical point of view the role of mass of a particle is decreasing with increase of its energy.

*email: okun@heron.itep.ru

PII S0920-5632(02)01472-X

2. ATOMS IN STATIC GRAVITY

20. According to General Relativity, the source and receptor of gravitational field is the density of energy-momentum tensor T^{ik} which is coupled to gravitational field like density of electric current is coupled to electromagnetic field.

    For a point-like particle

$$T^{ik} = m\delta(\mathbf{r})\frac{dx^i}{d\tau}\frac{dx^k}{dt} \quad , \quad \tau = \sqrt{t^2 - \mathbf{r}^2} \ .$$

    For a particle at rest ($v = 0$) mass m is the source and receptor of gravity: $T^{00} = m\delta(\mathbf{r})$.

    Thus, according to GR, there is no such notion as gravitational mass m_g.

21. ϕ – gravitational potential

    a) the potential of sun:

$$\phi = -\frac{GM_\odot}{rc^2} = -\frac{1}{2}\frac{r_g}{r} \ , \quad \phi(\infty) = 0$$

    G – Newton constant, $M_\odot$ – mass of the sun,
    gravitational radius $r_g = 2GM_\odot/c^2 = 3$ km.

    b) the potential near the surface of earth at height h:

$$\phi = \frac{gh}{c^2} \ , \quad \phi(0) = 0 \ , g = 9.8 \text{ m s}^{-2}$$

22. $E_{0g} = m + m\phi = m(1+\phi)$, $E_{0g} \neq m$.
    E_{0g} – rest energy in a static gravitational field. It increases with increasing distance from the source.

23. For an excited level with mass m^*

$$E^*_{0g} = m^*(1+\phi) \ .$$

    The splitting between levels increases with h:

$$\Delta E_{0g} = E^*_{0g} - E_{0g} = (m^* - m)(1+\phi) \ .$$

24. Note that for electric potential the splitting is constant:

$$E_{0e} = m^* + e\phi_e \ .$$

    Consider an excited ion He^+ in a capasitor:

$$\Delta E_{0e} = E^*_{0e} - E_{0e} = m^* - m \ .$$

25. The frequency ω of a photon emitted in the process of deexcitation

$$\omega = \Delta E_{0g}/h \ ,$$

    where $h = 2\pi\hbar$ is the Planck constant.
    Hence atomic clocks in gravitational field increase their rate with height, as predicted by Einstein in 1907.

26. This was confirmed by airplane experiments in 1970s.
    Feynman: "atoms at the surface of the earth are a couple of days older than at its center".

27. Thought experiment: Carry clock A from the first floor to the second. In a year carry clock B the same way. A will be ahead of B. $\Delta T/T = gh/c^2$. Absolute effect!
    This thought experiment is used to get rid of disparity between a clock on an airplane and that on a desk. (Think of twin paradox in special relativity.)

28. Pendulum clocks are not standard clocks similar to wristwatches, or atomic clocks.
    $T \sim \sqrt{l/g}$.
    They are gravimeters, measuring the strength of gravitational field.

3. PHOTONS IN STATIC GRAVITY

29. In 1911 Einstein predicted redshift of photon frequency in static gravitational field: the gravitational redshift.

30. In 1960 Pound and Rebka discovered the shift $\Delta\omega/\omega = gh/c^2$ in ^{57}Fe. Interpretation in terms of "the weight of the photon".

31. For static field and static observers the frequency of photon does not depend on height. (Maxwell equations in static field). Hence the shift observed by Pound et al. was due to the larger splitting of ^{57}Fe levels at the top of the Harvard tower. Redshift of photons relative to clocks.

32. Unlike frequency, the momentum of photon decreases with height.
Schwarzshild metric in general relativity g^{ij}.
The masslessness of photon:
$$g^{ij}p_ip_j = 0,\ i,j = 0,1,2,3,\ g^{00}p_0^2 - g^{rr}p_r^2 = 0$$
$$g^{00} = 1/(1+2\phi)\ ,\quad g^{rr} = (1+2\phi)\ ,$$
$p_0 \equiv E$, hence,
$$p_r = \frac{E}{1+2\phi} = \frac{E}{1-r_g/r}\ .$$

33. Photon's wave length λ increases with height. In that sense photon is redshifted
$$\lambda = \frac{2\pi\hbar}{p_r} = \frac{2\pi\hbar}{E}(1+2\phi)$$

34. Coordinate velocity is smaller near the sun:
$$v = \frac{\lambda\omega}{2\pi} = (1+2\phi)$$

35. Another derivation: Consider interval
$$ds^2 = g_{00}dt^2 - g_{rr}dr^2 = 0$$
$$g_{00} = \frac{1}{g^{00}} = 1+2\phi\ ,\quad g_{rr} = \frac{1}{g^{rr}} = \frac{1}{1+2\phi}$$
$$v = \frac{dr}{dt} = \sqrt{g_{00}/g_{rr}} = 1+2\phi = 1-\frac{r_g}{r} < 1$$

36. Before general relativity (1906, 1911) Einstein had $v = 1+\phi$ (there was no g_{rr} at that time). In 1915: $(1+2\phi)$.

37. Geometrical optics. Refraction index
$$n = \frac{1}{v} = \frac{1}{1+2\phi} \approx 1+\frac{r_g}{r}\ ;\quad n-1 = \frac{r_g}{r}\ .$$

38. Deflection of star light by sun (1911, 1915, 1919)
$$\alpha = \frac{2r_g}{R_\odot} \sim 10^{-5}$$

39. Gravitational lensing of galaxies.

40. Delay of radar echo from a planet in upper conjunction (Shapiro, 1964)
$$\Delta t = \frac{2}{c}\int_{z_e}^{z_p} dz(\frac{1}{v}-1) = \frac{2}{c}\int_{z_e}^{z_p} dz\frac{r_g}{r} =$$
$$= 2\frac{r_g}{c}\ln\frac{4r_pr_e}{R_\odot^2} \approx 240\mu s$$
Earth: $r_e = 150\cdot 10^6$ km.
Mercury: $r_p = 58\cdot 10^6$ km,
$R_\odot = 0.7\cdot 10^6$ km.

41. Deceleration of a particle approaching the sun is a relativistic effect, characteristic not only to photons, but to any fast enough particle. M.I. Vysotsky (private communication) has shown that "enough" means $v_\infty > c/\sqrt{3}$, where v_∞ is velocity of the particle at $r=\infty$.

To prove this consider the expression for energy of a particle of mass m falling on the sun along radius r with coordinate velocity $v = dr/dt$ (use eq. (88.9) from "Field Theory" by L. Landau and L. Lifshitz):
$$E = \frac{mc^2\sqrt{g_{00}}}{\sqrt{1-\frac{g_{rr}}{g_{00}}v^2}}\ .$$
Note that, according to Schwarzschild, $g_{00} = \frac{1}{g^{rr}} = 1 - r_g/r$ and that energy does not depend on r. Hence
$$1 - v_\infty^2 = g_{00}^{-1} - g_{00}^{-3}v^2\ .$$
For $r_g/r \ll 1$ one gets
$$v^2 = v_\infty^2 + r_g/r(1-3v_\infty^2)\ .$$
Thus, the particle accelerates like a nonrelativistic stone only if $v_\infty < c/\sqrt{3}$.

4. MISLEADING TERMINOLOGY

42. If, instead of $\mathbf{p} = E\mathbf{v}/c^2$, the non-relativistic formula $\mathbf{p} = m\mathbf{v}$ is kept, then $m = E/c^2$. This "relativistic mass", often denoted as m_r, is another notation for energy.

43. When applied to photon "$m_\gamma = E_\gamma/c^2$". Poincare, Lorentz, Born, Pauli, partly Einstein.

44. Einstein coined the expression "energy - mass equivalence", which sometimes he used in the sense "whenever there is energy, there is mass". (Recall massless photon.)

45. A redundant term "rest mass" m_0 is abundantly used.

46. Gravitational redshift is interpreted as "loss of photon's energy as it climbs out of gravitational potential".

47. From the point of view of relativity no m_r, m_0, m_g, m_i, m_l, m_t. Only m!

5. UNSOLVED PROBLEMS OF m

48. Is the Higgs mechanism correct? If yes, we have to discover higgses and to study the pattern of their couplings.

49. Can the hierarchy problem ($m_{Pl} \simeq 10^{19}$ GeV *vs* $m_Z \simeq 100$ GeV) be solved by SUSY? If yes, we have to discover numerous superparticles and to ascertain the pattern of their masses and couplings.

50. SUSY, if confirmed, would lead us into the quantum dimensions of space and time, into superspace.

51. Supergravity might lead to superunification of all forces.

52. The scale and mechanism of SUSY breaking is of paramount importance.

53. What makes the critical energy density of vacuum $\varepsilon_c \simeq 10^{-47}$ GeV^4 so small compared with η^4, the density of vacuum expectation value (VEV) of the Higgs field: $\eta = (246$ GeV$)$?

54. If the recent data on the universe expansion are correct, how to explain the pattern of cosmological energy densities of baryons of dark matter, and of vacuum:

$$\Omega_B \equiv \varepsilon_B/\varepsilon_c \simeq 0.03,\ \Omega_{DM} \equiv \varepsilon_{DM}/\varepsilon_c \simeq 0.3\ ,$$

$$\Omega_\Lambda \equiv \varepsilon_\Lambda/\varepsilon_c \simeq 0.7\ ?$$

55. Compared with the above enigmas, the problem of neutrino masses and mixings may seem to be not so lofty. But it could reveal important clues to the concept of mass.

56. A serious twist to the concept of mass was given by quarks and gluons. The basic definition $m^2 = E^2 - \mathbf{p}^2$ is not applicable to them, as they never exist as free particles with definite value of momentum. This colored particles are confined within colorless hadrons.

57. That reminds us that when speaking about mass of a particle we have to specify the distance (momentum tranfer q) at which it is measured. Masses as well as charges are running functions of q^2. It is not enough to write down the fundamental Lagrangian. One has to indicate the scale to which its couplings and masses refer.

58. At short distances the masses of light u-quarks and d-quarks, the building blocks of nucleons, are smaller than 10 MeV, while masses of nucleons are hundred times larger. These short-distance quarks are called "current quarks".

59. The same quarks with their "gluon coats" are called "constituent quarks", their masses being about 300 MeV. There should exist "constituent gluons". Hadrons without "valent" quarks are called "glueballs", hadrons containing both "valent" quarks and gluons are referred to as "hybrids".

60. The quark-antiquark and gluonic vacuum condensates: $< q\bar{q} >$, $< GG >$, etc. play an important role. The masses of hadrons appear due to "burning out" of these (negative energy density) vacuum condensates inside hadrons by the intense color fields of the valent constituents. Though this picture has not been rigorously proved, it seems evident that the origin of mass is in vacuum.

61. The existing terminology is obsolete and non-adequate. Fermions (quarks and leptons) are usually referred to as "matter", though neutrinos kinematically behave like photons. Even very heavy bosons are often referred to as "radiation". Actually all fields and their quantum excitations (including photons) are matter.

    Massive vacuum; is it also matter?

62. The literature on the subjects discussed in this lecture is very vast and controversial. The interested reader might look for further references in article [1–12]

ACKNOWLEDGEMENTS

I am grateful to Alessandro Bettini and Venya Berezinsky for the kind invitation to give this lecture at TAUP 2001. The work was partially supported by A. van Humbold Award and by RFBR grant No. 00-15-96562.

REFERENCES

1. On the photon mass (with I.Yu.Kobzarev). UFN **95** (1968) 131-137;
   Sov. Phys. Uspekhi **11** (1968) 338-344.
2. The concept of mass. Physics Today, June 1989, pp. 31-36
3. Putting to rest mass misconseptions. Physics Today, May 1990, pp.13,15,115-117.
4. The Problem of Mass: From Galilei to Higgs. Opening lecture at the 29th Intern. School of Subnuclear Physics, in "Physics at the Highest Energy and Luminosity to Understand the Origin of Mass", v.29, Ed. A.Zichichi (Plenum Press, N.Y., 1992) pp.1-24.
5. Note on the meaning and terminology of Special Relativity. Eur. J. Phys. **15** (1998) 403-406.
6. Comment on the physical meaning of 'c' in relativistic equations (with B. Webber). Eur. J. Phys. **20** (1999) L47.
7. Gravitation, Photons, Clocks (with K. Selivanov, V.Telegdi). UFN **169** (1999) 1141-1147; Phys. Uspekhi **42** (1999) 1045-1051.
8. On the interpretation of the redshift in a static gravitational field (with K. Selivanov, V. Telegdi). Am. J. Phys. **68** (2000) 115-119.
9. Photons and static gravity. Mod. Phys. Lett. **A15** (2000) 1941-1947; preprint hep-ph/0010120.
10. A thought experiment with clocks in static gravity. Mod. Phys. Lett. **A15** (2000) 2007-2009; preprint hep-ph/0010256.
11. Reply to the letter "What is mass?" by R.I.Khrapko. UFN **170** (2000) 1366-1371; Phys. Uspekhi **43** (2000) 1270-1275.
12. Relation between energy and mass in Bohr's essay on his debate with Einstein. Yad. Fiz. **64** (2001) 590-593; Phys. Atomic Nuclei **64** (2001) 536-539.

This page intentionally left blank

SPACETIME AND VACUUM AS SEEN FROM MOSCOW

L.B. OKUN
ITEP, Moscow, 117218, Russia
E-mail: okun@heron.itep.ru

Contents

1 Introduction

This talk presents: a few episodes from the history of ideas on the structure and evolution of vacuum during last 50 years as seen from ITEP; the impact on the theory of vacuum by I. Pomeranchuk, L. Landau, Ya. Zeldovich, A. Sakharov, D. Kirzhnits and their disciples; the importance of international exchange of ideas; the crucial role of experiments on P and CP violation; interconnection of particle physics and cosmology.

2 Pomeranchuk on vacuum

Being a student of Isaak Pomeranchuk I first heard from him in 1950: "The vacuum is filled with the most profound physical content" and "The Book of Physics consists of two volumes: v.1. Pumps and Manometers, v.2. Quantum Field Theory." In the middle of 1950's Pomeranchuk proved his famous theorem about equality of particle and antiparticle cross-sections at the same target. It is not an accident that the Regge pole with quantum numbers of vacuum, which is responsible for this equality, was called later Pomeranchuk pole or pomeron.

In 1956 Pomeranchuk encouraged Boris Ioffe, Alexei Rudik and myself to work on P-non-conservation. By applying CPT theorem[1] to the famous paper of T.D. Lee and C.N. Yang[2] we came to the conclusions[3] that: 1) $\vec{\sigma}\vec{p}$ correlations suggested by them would signal not only P-, but also C-violation. 2) K_S and K_L are respectively even and odd under T (or, better, CP), but not under C[4]. Hence $K_S \not\to 3\pi^0$, $K_L^0 \to 3\pi^0$ (November 1956). The same conclusions were reached by T.D. Lee, R. Oehme and C.N. Yang[5] (January 1957) (see Nobel lectures by T.D. Lee and C.N. Yang[6,7] (December 1957)).

3 Landau on parity, P, and combined parity, CP

Till the end of 1956 Lev Landau did not believe in the possibility of P-violation. In the summer of 1956 he did not encourage Iosif Shapiro to publish a Moscow University report,[8] in which an experiment of the type carried out later by C.S. Wu *et al.* was suggested (according to Shapiro's report energy of a system changed its value under mirror reflection). Landau was convinced that P must be conserved because vacuum is mirror symmetric. (One would not expect this argument from an author of 1950 article with Ginzburg[9] in which spontaneous symmetry breaking was introduced into theory of superconductivity.)

We (IOR) discussed our paper with Landau. A few weeks after it was submitted to JETP, Landau[10] put forward (December 1956) the idea of absolute CP-invariance and conservation of "combined parity". According to this idea the "sin" of P-violation was committed by particles and antiparticles (with opposite signs), but the mirror symmetry of vacuum was preserved.

The idea was widely accepted. I considered it to be beautiful, but on the other hand, I did not understand why the Lagrangian could not have complex coefficients. Therefore in lectures given at ITEP (1960-61) and Dubna (1961)[11] and in the book[12] based on these lectures (1963) the importance of experimental tests of CP was stressed, in particular search for $K_L \to 2\pi$.

4 Search and discovery of $K_L^0 \to \pi^+\pi^-$

A special search at Dubna was carried out by E. Okonov and his group. They have not found a single $K_L^0 \to \pi^+\pi^-$ event among 600 decays of K_L^0 into charged particles[13] (Anikina *et al.*, JETP, 1962). At that stage the search was terminated by administration of the Lab. The group was unlucky.

Approximately at the level 1/350 the effect was discovered by J. Christensen, J. Cronin, V. Fitch and R. Turlay[14] at Brookhaven in 1964 in an experiment the main goal of which was $K_L \to K_S$ regeneration in matter.

Thus absolute CP-invariance was falsified.

5 "Mirror world"

Still the appeal of Landau's idea of absolutely symmetric vacuum was so strong that in 1965 Igor Kobzarev, Isaak Pomeranchuk and myself suggested the hypothesis of a "mirror world".[15] We assumed CPA invariance, where A [from "Alice through the Looking Glass"] transforms our part of the Lagrangian into its mirror part.

Each of our particles has its mirror counterpart. The mirror particles have between them the same electromagnetic, weak and strong interactions as ours. In principle there might exist mirror nuclei, atoms, molecules, stars, planets, galaxies, even mirror life. Whether they actually exist depends on cosmological evolution.

The possibility of the existence of both "left protons", p_L, and "right protons", p_R, had been discussed by T.D. Lee and C.N. Yang in two last paragraphs of their famous article.[2] But they believed that p_L and p_R can interact with the same pion and the same photon. We have proved that this is impossible.

According to our original assumption, the only particles which belong to both our and mirror worlds are gravitons. If there were two gravitons, nothing would connect the two worlds and the idea of a mirror world would have no physical consequences.

Why the graviton but not, say, a photon? As soon as you assume that the photon is common to both worlds, you immediately come to contradiction with experiments. Colliding e^+e^- through a virtual photon would annihilate not only into our particles but also into mirror ones. Besides, the electromagnetic radiative corrections would be doubled destroying the beautiful agreement of QED with experiment.

In 1982 I studied the oscillations between our photons and sterile ones, "paraphotons";[16] in 1983 — considered a stronger coupling between two worlds which could be caused by exchange of some hypothetical neutral bosons.[17] A number of authors considered mirror particles as a substantial component of dark matter. Many enthusiastic papers about mirror matter were written recently by R. Foot (see Ref. [18] and references therein) without proper citation of Ref. [15].

6 Zeldovich and cosmological term

Yakov Zeldovich was very excited when he realized that a non-vanishing Λ-term is strongly required by quantum field theory. In 1967 he estimated ε_Λ on the basis of pairs of virtual elementary particles in vacuum. First he started with $\varepsilon_\Lambda \sim m_{Pl}^4$ and found enormous (124 orders) discrepancy with observations (private communication). Then[19] he assumed that quartically divergent loops ($\sim m_{Pl}^4$ and even m^4) vanish and estimated

$$\varepsilon_\Lambda \sim \frac{m^6}{m_{Pl}^2} \ ,$$

where m is the mass of a typical elementary particle of the order of a proton mass (1 GeV). That was the first attempt to put m_{Pl}^2 in denominator. (The contribution of the loop is proportional to gravitational interaction.) In order to bridge the gap of 10^8, he tried

$$\varepsilon_\Lambda \sim G G_F m^8 \ ,$$

where G is Newton constant ($m_{Pl} = G^{-1/2}$), while G_F is Fermi constant.

In 1968 Zeldovich noticed the negative sign of the contribution of fermions and indicated that the contributions of bosons and fermions might cancel each other![20] This was before Yuri Golfand and his student Evgeni Likhtman published[21] the first paper on supersymmetry.

Recently Johannes Blümlein and Paul Söding kindly informed me about Refs. [22,23], in which early unpublished estimates made by young W. Pauli are described. Pauli discussed vacuum fluctuations and their influence on cosmology. Note that he overlooked that cosmological term would lead to inflation and concluded that the universe would be smaller than the distance from the earth to the moon. Even earlier, in 1916, the "zero point energy" had been considered by W. Nernst (Ref. [24], see also Ref. [25]).

7 QCD vacuum condensates

Here I will make a digression from cosmology to Quantum Chromodynamics (QCD) with its quarks and gluons.

In 1965 V. Vanyashin and M. Terentyev[26] discovered the negative sign of the vacuum polarization of gauge bosons in the case of SU(2) gauge symmetry. Their numerical result was slightly incorrect, as the "ghost" was unknown at that time. A correct number (with the same negative sign!) was derived by I. Khriplovich in 1969 by using Coulomb gauge.[27] The negative sign of vacuum polarization was a novelty: for photons in QED it is positive. But at that time the nonabelian theories with gauge vector bosons were not considered seriously at ITEP. Even after SLAC experiments on deep inelastic scattering and their parton interpretation by J. Bjorken[28] and R. Feynman[29] we were not clever enough to realize the importance of negative vacuum polarization.

The situation has drastically changed in 1973 when H. Fritzsch, M. Gell-Mann, and H. Leutwyler[30] suggested QCD with its SU(3) color gauge symmetry. In papers by H.D. Politzer[31] and D.J. Gross and F. Wilczek[32] the negative sign was rediscovered, and its importance for QCD was stressed: it lead to decreasing color charge at short distances (and hence to asymptotic freedom at high momentum transfers).

It should be noted that Vanyashin and Terentyev[26] considered as "absurd" not the ultraviolet, but the infrared behaviour of the charge where its square first rises and then becomes negative. At present we know that this region is never reached: in Electroweak Model because of the Higgs mechanism, in QCD because of confinement. The increase of charge at large distances makes perturbation theory invalid at such distances. Non-perturbative effects manifest themselves in non-vanishing vacuum expectation values (VEVs) of colorless products of quark and gluon fields, the so-called QCD condensates. These condensates were used at ITEP by Michael Shifman, Arkady Vainshtein and Valentin Zakharov in order to successfully describe the properties of various hadrons.[33] This work was followed by many hundreds of theoretical papers (for a review see Ref. [34]).

8 Sakharov and baryonic asymmetry of the universe, BAU

In 1967 I had the privilege to witness the creation of Sakharov's seminal paper[35] on the Baryonic Asymmetry of the Universe (BAU). He started out from the CP-violating charge asymmetry of the semileptonic decays of K_L^0 and from the nonequality of branching ratios in the decays of Σ^+ and $\widetilde{\Sigma}^+$ discussed by S. Okubo.[36] Using it as a springboard, he jumped from strange particles to the universe and from strangeness violation to baryon number violation. A small excess of baryons over antibaryons in the hot big bang soup survived, while the bulk annihilated into photons, neutrinos and antineutrinos. Photons are observed today as CMBR. The cosmic background neutrinos are tantalizing. When devising the baryon nonconserving ad hoc interaction Sakharov was helped by ITEP team. In the framework of SM and its extensions there are several mechanisms of baryon number violation; among them grand unification[37] and triangle anomalies.[38] The latter mechanism raises serious problems[39] at electroweak scale: it might wipe out baryon asymmetry, unless the original lepton asymmetry is involved.[40] The problem of BAU can not be solved in the framework of Standard Model. (V. Rubakov and M. Shaposhnikov.[41])

9 Kirzhnits and phase transitions

With the advent of the Standard Model based on $SU(3) \times SU(2) \times U(1)$ gauge symmetry new concepts came to particle physics; one of them is the vacuum expectation value (VEV) of the Higgs field. It reduces $SU(2)_L \times U(1)_Y$ to $U(1)_{em}$ and provides with masses leptons, quarks, W and Z bosons. An important step was made[42] in 1972 by David Kirzhnits, a theorist at the Lebedev Institute, who realized that the Higgs VEV is a function of temperature and has to vanish in the early universe, when we go backwards in time to temperatures much higher than the present value of VEV. Hence the masses of particles have also to vanish. This idea was further developed by Kirzhnits and his student Andrei Linde[43] and later by many others.[44] (Both A. Sakharov and D. Kirzhnits were disciples of Igor Tamm.) I first heard about the vacuum phase transition from David on a street during a conference in Tashkent in 1972.

10 Vacuum domain walls

Two years later Igor Kobzarev, Yakov Zeldovich and myself[45] merged the idea of vacuum phase transition with a model of spontaneous CP violation proposed by T.D. Lee.[46] According to this model a neutral pseudoscalar field has two degenerate vacua (see Fig. 1).

The vacuum around us has a sign that was fixed by chance during the cooling of the universe and formation of VEVs. But at any distant enough place the other sign might have been chosen. The domains of different sign are separated from each other by "domain walls" the thickness and density of which are determined by VEVs of the field φ and by its self-coupling. Typical electroweak scale is several hundred GeV. We discussed the evolution of the universe filled with domain walls and concluded that at present the nearest wall should have gone beyond the horizon, leaving as a farewell an anisotropy of CMBR. The advent of inflational cosmology (see below) greatly weakened this argument against spontaneous CP-violation.

11 Monopoles, strings, instantons, and sphalerons

Vacuum domain walls were the first megascopic cosmological elementary objects considered in the framework of quantum field theory. They were caused

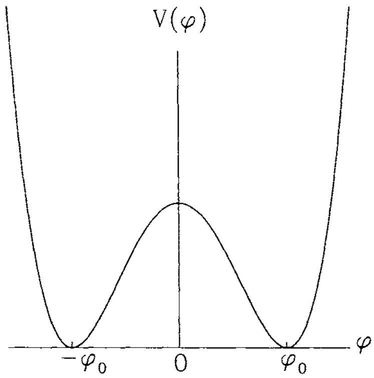


Figure 1.

by spontaneous breaking of discrete symmetry. Later in 1974 G. 't Hooft[47] and A. Polyakov[48] discovered magnetic monopoles as solutions of spontaneously broken SU(2) gauge symmetry. Cosmological creation of monopoles and of strings (the latter appear as solutions of a broken U(1) gauge symmetry) was considered in 1976 by T.W.B. Kibble.[49] (Microscopic U(1) strings of Nambu type underlying dual resonance model of hadrons were considered in 1973, without any reference to cosmology, by H.B. Nielsen and P. Olesen,[50] who were inspired by Abrikosov's vertices in a superconductor.[51])

Since the early 1980's a whole "industry of theoretical papers" on cosmological role of domain walls, strings and monopoles appeared. I will mention here only articles by Ya.B. Zeldovich,[52] A. Vilenkin[53] and an exotic example of a cosmic string invented by Albert Schwartz[54] in 1982 after he learned from me about the "mirror world". A particle, after making a circle around his "Alice string", transforms into a mirror particle and becomes invisible for a terrestrial observer.[55]

A special type of non-perturbative transitions between topologically different vacua of non-abelian gauge fields were constructed by A. Belavin, A. Polyakov, A. Schwartz, V. Tyupkin[56] in 1975 and by 'Hooft[57] in 1976. The latter called them instantons. In the four-dimensional Euclidean space (with imaginary time) instantons represent solutions with finite action. Many attempts to understand the mechanism of confinement in QCD have exploited instantons.

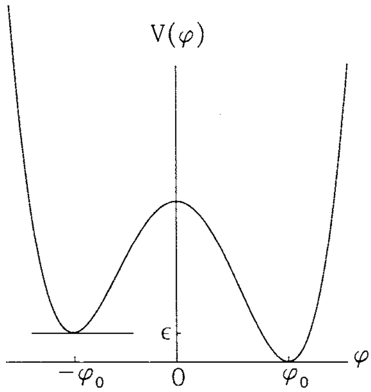


Figure 2.

Amplitudes of high temperature transitions between vacua in electroweak theory, called sphalerons, were first analyzed by N. Manton[58] in 1983 and by F. Klinkhamer and N. Manton[59] in 1984 in connection with baryon nonconserving phase transition at electroweak scale.

12 False vacuum

In 1974 T.D. Lee and G.C. Wick[60] suggested a Higgs potential with two vacua, one of which lies a little bit higher than the other (see Fig. 2).

The cosmological consequences for universe that lives in an upper metastable vacuum were elaborated[61] by Igor Kobzarev, Mikhail Voloshin and myself in 1974 and later, in 1977, by Curt Callan and Sidney Coleman[62,63] (they dubbed it "false vacuum") and also by Paul Frampton.[64]

The spontaneous decay of a false vacuum starts by formation through quantum tunneling of the smallest bubble of the true vacuum surrounded by a wall which separates the two vacua. The critical radius of this bubble is such that the gain of energy proportional to its volume becomes large enough to compensate the mass of the wall which is proportional to its surface:

$$\frac{4\pi}{3}R_c^3 \cdot \varepsilon = 4\pi R_c^2 \sigma \ ,$$

where ε is the energy density of the false vacuum (see Fig. 2), while σ is the surface density of the wall. After that the bubble expands classically, destroying the universe.

When I first thought that the creation of a bubble could be catalyzed at a collider, my back shivered. Then I reassured myself: all possible collisions have already occurred in the early universe. A few months later I told Andrei Sakharov about the bubble. His reaction was: "Such theoretical work should be forbidden". My argument about collisions in the early universe was rejected by him: "Nobody had collided two nuclei of lead".

In 1984 P. Hut[65] published an estimate of uranium-uranium collisions in cosmic rays during the time of the existence of universe. It turned out that the number of such collisions, N, was many orders of magnitude larger than at any future heavy ion collider: $N \sim 10^{47}$ for Fe with $E \geq 100$ GeV/nucleon. Experiments at RHIC (BNL) study Au - Au collisions at $\sqrt{s} \sim 200A$ GeV ~ 40 TeV ($A \simeq 197$ for gold). There will be $2 \cdot 10^{11}$ collisions in 10 years, first at 30 GeV, then at 70 GeV, and finally at 100 GeV/nucleon. A fixed target NA 50 experiment at CERN gathered $2 \cdot 10^{12}$ Pb - Pb collisions at $\sqrt{s} \sim 17A$ GeV~ 3.5 TeV ($A \simeq 207$ for lead). The experiment ALICE at LHC (CERN) will study Pb - Pb collisions at $\sqrt{s} \sim 1200$ TeV. The abundances of heavy nuclei may be $10^{-5} - 10^{-10}$ of that of Fe, which still leaves a huge safety factor in cosmic rays. For a recent update of these arguments see Refs. [66,67].

13 Inflation

The notion of a false vacuum with large energy density (VEV) naturally led Alan Guth[68] to the idea of inflationary universe (1980). The exponential expansion solves the problems of flatness and the horizon as well as homogeneity and isotropy problems. It also explains the vanishing abundance of magnetic monopoles. (For earlier models of exponential expansion of the universe see papers by E.B. Gliner[69] and A.A. Starobinsky.[70])

In 1982 Andrei Linde[71] and independently Andreas Albrecht and Paul Steinhardt[72] suggested how to stop this expansion by producing many vacuum bubbles and high entropy. Also in 1982 Alexander Dolgov[73] suggested a mechanism with a rolling scalar field which led to what later was dubbed "quintessence".

Recent measurements with high angular resolution of CMBR fluctuations confirm the inflation scenario (experiments BOOMERANG,[74] MAXIMA,[75] DAST[76]). The flatness of the universe is corroborated by observations of high-z supernova of type Ia.[77,78] On the other hand, they give a puzzling result: the fractional energy density of the universe $\Omega = \varepsilon/\varepsilon_c$, where ε_c is the critical density, corresponding to the flat universe, is dominated by cosmological term — *i.e.* energy density of vacuum ($\sim$ 70%), then follows the so-called dark matter ($\sim$ 30%), while ordinary baryonic matter constitutes only about 3%. To understand the cause and mechanism of such fine tuning is a great challenge.

During the 1990s the development of inflation cosmology has lead to a new view on the place of our universe in the world. According to this view, our universe is a part of eternal metauniverse, which consists of innumerable universes. They create their offsprings by forming Big-Bang bubbles in false vacua. The properties of vacua, fields, particles and even number of space-time dimensions differ from one universe to the other.[79] This enormous variety may explain the anthropic properties of our universe which is so nicely tuned for our existence (see, *e.g.* Ref. [80]).

14 Brane and bulk

The idea of extra dimensions[81,82] has been actively discussed for 60 years before it was first combined with the idea of "domain walls" in these extra dimensions.[83,84] According to this idea, our four-dimensional world is a "brane" in a multidimensional bulk.[86,87,88] All known particles and their fields are confined (completely, or almost completely) within the brane, while real and virtual gravitons are allowed to leave the brane and to propagate in the bulk. This drastically changes gravity. The Planck scale may become as low as 1 TeV. There could be deviations from the standard Newton's force at distances of the order of millimeters and many other strange phenomena. But at present all this is a kind of a daydreaming.

References

1. W. Pauli, in *Niels Bohr and the Development of Physics*, edited by W. Pauli, (Pergamon Press, London, 1955).
2. T.D. Lee and C.N. Yang, *Phys. Rev.* **104**, 254 (1956).
3. B.L. Ioffe, L.B. Okun and A.P. Rudik, *Zh.ETF* **32**, 396 (1957) [*Sov. Phys. JETP* **5**, 328 (1957)].
4. M. Gell-Mann and A. Pais, *Phys. Rev.* **97**, 1387 (1955).
5. T.D. Lee, R. Oehme and C.N. Yang, *Phys. Rev.* **106**, 340 (1957).
6. C.N. Yang, in *Nobel lectures. Physics 1942–1962*, p. 393 (Elsevier, 1964).
7. T.D. Lee, in *Nobel lectures. Physics 1942–1962*, p. 406 (Elsevier, 1964).
8. I.S. Shapiro, *Moscow State University report*, Summer 1956, unpublished.
9. V.L. Ginzburg and L.D. Landau, *ZhETF* **20**, 1064 (1950) (in Russian).
10. L.D. Landau, *ZhETF* **32**, 405 (1957) [*Sov. Phys. JETP* **5**, 336 (1957)]; *Nucl. Phys.* **3**, 127 (1957).
11. L.B. Okun, *Lectures on theory of weak interactions of elementary particles*, 16 ITEP preprints, 1960–1961 (in Russian).
    English translations: *Theory of weak interactions: Thirteen lectures*, AEC-tr-5226. US Atomic Energy Commission. Oak Ridge, Tenn. Another English translation of 13 lectures: NP-10254; of lectures 14–16: NP-10254, 10842, 10845, 10840.
    Revised edition: Preprint JINR P-833, Dubna, 1961 (in Russian).
12. L.B. Okun, *Weak Interactions of Elementary Particles* (Pergamon, 1965) a translation of Slaboe vzaimodeystvie elementarnych chastits (Fizmatgiz, Moskva, 1963); another translation: Israel Program Sci. Translations, 1965.
13. M. Anikina *et al.*, *ZhETF* **42**, 130 (1962) [*Sov. Phys. JETP* **15**, 93 (1962)].
14. J.H. Christenson, J.W. Cronin, V.C. Fitch and R. Turlay, *Phys. Rev. Lett.* **13**, 138 (1964).
15. I.Yu. Kobzarev, L.B. Okun and I.Ya. Pomeranchuk, *Yad. Fiz.* **3**, 1154 (1966) [*Sov. J. Nucl. Phys.* **3**, 837 (1966)].
16. L.B. Okun, *ZhETF* **83**, 892 (1982); *JETP* **56**, 502 (1982).
17. L.B. Okun, *On Search for mirror particles*, Preprint ITEP-149 (1983).
18. R. Foot, *Seven (and a half) reasons to believe in mirror matter*, astro-ph/0102294.
19. Ya.B. Zeldovich, *Pisma ZhETF* **6**, 883 (1967).
20. Ya.B. Zeldovich, *Uspekhi Fiz. Nauk* **95**, 209 (1968).
21. Yu.A. Golfand, E.P.Likhtman, *Pis'ma ZhETF* **13**, 323 (1971).
22. C.P. Enz, A. Thellung, *Helv. Phys. Acta* **33**, 839 (1960).

23. N. Straumann, *The mystery of the cosmic vacuum energy density and the accelerated expansion of the Universe*, astro-ph/9908342.
24. W. Nernst, *Verhandlungen der deut. physik. Gesellschaft* **13**, 83 (1916).
25. S.E. Rugh, H. Zinkernagel, *The quantum vacuum and the cosmological constant problem*, hep-th/0012253.
26. V.S. Vanyashin, M.V. Terentyev, *ZhETF* **48**, 565 (1965); *JETP* **21**, 375 (1965).
27. I.B. Khriplovich, *Yad. Fiz.* **10**, 409 (1969); *Sov. J. Nucl. Phys.* **10**, 235 (1970).
28. J.D. Bjorken, *A theorist view on e^+e^--annihilation*, in: *Proc. Intern. Symposium on Electron and Photon Interactions at High Energy*, August 1973. Preprint SLAC-PUB-1318.
29. R.P. Feynman, *Photon-Hadron interactions* (W.A. Benjamin Inc., 1972).
30. H. Fritzsch, M. Gell-Mann, H. Leutwyler, *Phys. Lett.* B **47**, 365 (1973).
31. H.D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973).
32. D. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973).
33. M. Shifman, A. Vainshtein and V. Zakharov, *Nucl. Phys.* B **147**, 385 (1979); 448 (1979).
34. M.A. Shifman, *ITEP Lectures on Particle Physics and Field Theory* (World Scientific, 1999).
35. A.D.Sakharov, *ZhETF Pis'ma* **5**, 32 (1967) [*JETP Lett.* **5**, 24 (1967)].
36. S. Okubo, *Phys. Rev.* **109**, 984 (1958).
37. H. Georgi, *Grand unified theories*, in: *History of Original Ideas and Basic Discoveries in Particle Physics*, Ed. Harvey B. Newman and Thomas Ypsilantis (Plenum Press, New York and London, 1996).
38. G. 't Hooft, *Phys. Rev. Lett.* **37**, 8 (1976).
39. V.A. Kuzmin, V.A. Rubakov and M.E. Shaposhnikov, *Phys. Lett.* B **155**, 36 (1985).
40. M. Fukugita and T. Yanagida, *Phys. Lett.* B **174**, 45 (1986).
41. V. Rubakov and M. Shaposhnikov, *UFN* **166**, 493 (1996); *Phys. Usp.* **39**, 461 (1996).
42. D.A.Kirzhnitz, *Pisma Zh.ETF* **15**, 745 (1972) [*JETP Lett.* **15**, 529 (1972)].
43. D.A. Kirzhnitz and A.D. Linde, *Phys. Lett.* B **42**, 471 (1972).
44. A.D. Linde, *Particle Physics and Inflationary Cosmology (1990), Fizika Elementarnykh Chastits i Inflatsionnaya Kosmologiya* (Nauka, Moscow, 1990).
45. I.Yu. Kobzarev, L.B. Okun and Ya.B. Zeldovich, *Zh.ETF* **67**, 3 (1974) [*JETP* **40**, 1 (1975)]; *Phys. Lett.* B **50**, 340 (1974).
46. T.D. Lee, *Phys. Reports* **9C**, No.2 (1974).

47. G. 't Hooft, *Nucl. Phys.* B **79**, 276 (1974).
48. A.M. Polyakov, *ZhETF Pis'ma* **20**, 276 (1974) [*JETP Lett.* **20**, 194 (1974)].
49. T.W.B. Kibble, *J. Phys.* A **9**, 1387 (1976).
50. H.B. Nielsen, P. Olesen, *Nucl. Phys.* B **61**, 45 (1973).
51. A.A. Abrikosov, *ZhETF* **32**, 1442 (1957).
52. Ya. B. Zeldovich, *Mon. Not. Roy. Astron. Soc.* **192**, 663 (1980).
53. A.Vilenkin, *Phys. Rev. Lett.* **46**, 1169 (1981).
54. A.S. Schwarz, *Nucl. Phys.* B **208**, 141 (1982).
55. A.S. Schwarz and Yu.S. Tyupkin, *Nucl. Phys.* **209**, 427 (1982).
56. A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu.S. Tyupkin, *Phys. Lett.* B **59**, 85 (1975).
57. G. 't Hooft, *Phys. Rev. Lett.* **37**, 8 (1976); 3432 (1976); *Phys. Rev.* D **14**, 3432 (1976).
58. N.S. Manton, *Phys. Rev.* D **28**, 2019 (1983).
59. F.R. Klinkhamer, N.S. Manton, *Phys. Rev.* D **30**, 2212 (1984).
60. T.D. Lee and G.C. Wick, *Phys. Rev.* D **9**, 2291 (1974).
61. M.B. Voloshin, I.Yu. Kobzarev and L.B. Okun, *Yad. Fiz.* **20**, 1229 (1974) [*Sov. J. Nucl. Phys.* **20**, 644 (1975)].
62. S. Coleman, *Phys. Rev.* D **15**, 2929 (1977); Err. **16**, 1248 (1977).
63. C. Callan, S. Coleman, *Phys. Rev.* D **16**, 1762 (1977).
64. P.H. Frampton, *Phys. Rev.* D **15**, 2922 (1977).
65. P. Hut, *Nucl. Phys.*A **418**, 301C (1984).
66. R.L. Jaffe, W. Busza, J. Sandweiss, and F. Wilczek, *Review of Speculative "Disaster Scenarios" at RHIC*, *Rev. Mod. Phys.* **72**, 1125 (2000) [hep-ph/9910333].
67. A. Dar, A. De Rujula and U. Heinz, *Phys. Lett.* B **470**, 142 (1999).
68. A.H. Guth, *Phys. Rev.* D **23**, 347 (1981).
69. E.B. Gliner, *ZhETF* **49**, 542 (1965).
70. A.A. Starobinssky, *Pisma ZhETF* **30**, 719 (1979); *Phys. Lett.* B **91**, 99 (1980).
71. A.D. Linde, *Phys. Lett.* B **108**, 389 (1982).
72. A. Albrecht, P.J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
73. A.D. Dolgov, *An attempt to get rid of the cosmological constant*, in: *Proc. Very early Universe*, Nuffield Workshop, Cambridge, 1982, pp. 449–458 (Cambridge Univ. Press, 1983).
74. C.B. Netterfield *et al.*, astro-ph/0104460.
75. A.T. Lee *et al.*, astro-ph/0104459.
76. N.W. Halverson, astro-ph/0104489.
77. B. Schmidt *et al.*, *Astrophys. J.* **507**, 46 (1998).

78. S. Perlmutter *et al.*, *Astrophys. J.* **517**, 563 (1999).
79. A. Linde, *Lectures in inflationary cosmology*, hep-th/9410082.
80. L.B. Okun, *Uspekhi Fiz. Nauk* **161**, 177 (1991); *Sov. Phys. Uspekhi* **34**, 818 (1991).
81. T. Kaluza, Sitzungsberichte Preussische Akad. Wissenschaften, p. 966 (1921).
82. O. Klein, *Z. Phys.* **37**, 895 (1926).
83. V.A. Rubakov and M.E. Shaposhnikov, *Phys. Lett.* B **125**, 136 (1983).
84. K. Akama, in *Gauge Theory and Gravitation, Proceedings of the International Symposium, Nara, Japan, 1982*, eds. K. Kikkawa, N. Nakanishi and H. Nariai (Springer-Verlag, 1983).
85. P. Horava and E. Witten, *Nucl. Phys.* B **460**, 506 (1996).
86. G. Dvali and M. Shifman, *Phys. Lett.* B **396**, 64 (1997).
87. L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 4690 (1999).
88. N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *Phys. Lett.* B **429**, 263 (1998).

Physics – Uspekhi **46** (10) 1099 – 1103 (2003)

METHODOLOGICAL NOTES

PACS numbers: **03.30.** + p, **45.50.** – j

Critical velocities $c/\sqrt{3}$ and $c/\sqrt{2}$ in the general theory of relativity

S I Blinnikov, L B Okun', M I Vysotskiĭ

DOI: 10.1070/PU2003v046n10ABEH001661

Contents

Abstract. **We consider a few thought experiments of radial motion of massive particles in the gravitational fields outside and inside various celestial bodies: Earth, Sun, black hole. All other interactions except gravity are disregarded. For the outside motion there exists a critical value of coordinate velocity $v_c = c/\sqrt{3}$: particles with $v < v_c$ are accelerated by the field like Newtonian apples, and particles with $v > v_c$ are decelerated like photons. Particles moving inside a body with constant density have no critical velocity; they are always accelerated. We consider also the motion of a ball inside a tower, when it is thrown from the top (bottom) of the tower and after elastically bouncing at the bottom (top) comes back to the original point. The total time of flight is the same in these two cases if the initial proper velocity v_0 is equal to $c/\sqrt{2}$.**

1. Introduction. Proper and coordinate velocities

According to the theory of relativity, a time interval depends both on the velocity of clocks and on the gravitational potential. There exist infinitely many different coordinates. In the majority of physical experiments the gravitational potential remains constant in the laboratory and it is convenient to use the so-called proper time τ which is determined by the clocks at rest in the laboratory frame. Maybe this is the reason why sometimes in the literature the proper time is called 'genuine', or 'physical'. A lightminded person would think that any other time (the coordinate time) is not physical, and thus should not be considered. As a result, some people working on the General Theory of Relativity (GTR) consider coordinate-dependent quantities as nonphysical, so to say 'second-quality' quantities.

S I Blinnikov, L B Okun', M I Vysotskiĭ Institute of Theoretical and Experimental Physics,
ul. B. Cheremushkinskaya 25, 117218 Moscow, Russian Federation
Tel. (7-095) 123 83 93
Fax (7-095) 127 08 33
E-mail: blinn@sai.msu.su; vysotsky@heron.itep.ru; okun@heron.itep.ru

Received 11 April 2003, revised 25 July 2003
Uspekhi Fizicheskikh Nauk **173** (10) 1131 – 1136 (2003)
Translated by M I Vysotskiĭ; edited by M S Aksent'eva

However the coordinate time t is even more important for some problems than the proper time τ. A good example is a particle moving in a static gravitational field, which does not depend on the time coordinate t. The proper time should be used when discussing experiments performed in one and the same gravitational potential (say, on a given floor of a building). However, to discuss events happening on different floors observers should introduce corrections which take into account the difference of the gravitational potentials. The coordinate time plays an important role in the global positioning system (GPS), where atomic clocks are running on several dozen of the Earth's satellites, and in metrology in general [1, 2]. That is why the coordinate quantities are also 'genuine', or 'physical'.

The solution of the Einstein equations outside a spherically symmetric massive body was given by Schwarzschild [3] and can be found in all textbooks on GTR (see, for example, [4 – 7]). Simultaneously, Schwarzschild found the metric inside a spherically symmetric body of constant density [8]. This paper is less known. Below we will compare the behavior of coordinate velocities in the Schwarzschild metrics [3, 8]. Some other coordinate systems will be considered at the end of the paper.

As found in papers [9 – 11], in the gravitational field of a spherically symmetric object (the Earth, the Sun, other stars) there exists a critical value of the coordinate speed $v_c = c/\sqrt{3}$. Particles moving radially with this speed do not accelerate or decelerate in the first order of the Newton constant G — thus, they 'ignore' gravity.

For $v < v_c$, a falling object accelerates (a famous example is Newton's apple). For $v > v_c$, a falling object decelerates (a well known example is the electromagnetic wave, or photon). This deceleration is a reason for the radar echo delay ([4], chapter 8, § 7) and the deviation of light by the Sun ([4], chapter 8, § 5).

That is why an intermediate critical coordinate velocity v_c should exist.

We will show that outside the Sun $v_c = c/\sqrt{3}$, while for a particle moving inside the Sun such a critical velocity does not exist: any particle (even a photon) accelerates when moving toward the center of the Sun.

To study gravitational effects for particles moving inside celestial bodies, one should select such particles, whose non-gravitational interaction with normal matter can be neglected. Neutrinos, neutralinos, and mirror particles [12] moving inside the Sun or the Earth satisfy this criterion. Another example is photons, electrons, and nucleons moving inside a mirror star. Finally, particles moving inside a vertical mine on the Earth also satisfy the above condition.

In Section 2 we introduce the necessary notations and obtain one of the main results: $\mathrm{v}_c = 1/\sqrt{3}$ for a particle moving in a weak gravitational field $r > R \gg r_g$, where r is the particle coordinate, R is the star radius, and r_g is the gravitational radius of the star [4–7]. Here and in what follows we put $c = 1$, that is why

$$r_g = 2GM\,, \tag{1}$$

where M is the mass of the star and G is the gravitational constant.

In Section 3 we consider motion outside a star with strong gravity (neutron star, black hole).

In Section 4 we consider motion inside a star, $r < R$, and demonstrate that both apples and photons are accelerated when freely falling inside a star.

In Section 5 the notion of a critical proper (not coordinate!) velocity is illustrated by the following thought experiment: a massive ball is thrown from the top (bottom) of a tower and after elastically bouncing at the bottom (top) comes back after time τ_+ (τ_-). The initial proper velocities in both cases are equal v_0. For $v_0 \ll 1$ we have $\tau_+ < \tau_-$, while for $v_0 = 1$ we have $\tau_+ > \tau_-$. We found that $\tau_+ = \tau_-$ for $v_0 = v_c \equiv 1/\sqrt{2}$.

In Section 6 other coordinate systems are considered (harmonic and isotropic) which differ from the Schwarzschild one for $r \sim r_g$, but describe flat space for $r \gg r_g$. In weak gravitational fields ($r/r_g \gg 1$) the critical coordinate velocity in these coordinates is also $1/\sqrt{3}$ (plus corrections $\sim r_g^2/r^2$). Concerning $v_c = 1/\sqrt{2}$, it is invariant under a change of coordinates.

2. Derivation of $\mathrm{v}_c = 1/\sqrt{3}$ for $r > R$

For a radial motion ($\dot{\theta} = \dot{\varphi} = 0$) the interval looks like [3]

$$ds^2 = g_{00}\, dt^2 - g_{rr}\, dr^2 \equiv d\tau^2 - dl^2\,, \tag{2}$$

where

$$g_{00} = 1 - \frac{r_g}{r}\,, \tag{3}$$

$$g_{rr} = \left(1 - \frac{r_g}{r}\right)^{-1} = \frac{1}{g_{00}}\,. \tag{4}$$

The so-called local (or proper) velocity v and coordinate velocity v are the main subjects of this paper. The local velocity is measured by a local observer with the help of a ruler and a clock which he has in his hands. Making this measurement, the local observer ignores that time is contracted and the radius is stretched by gravity:

$$v = \frac{dl}{d\tau}\,, \tag{5}$$

where the local (proper) coordinates l and τ are defined in Eqn (2).

The same observer will measure the coordinate velocity if he takes into account that his rulers and clocks are influenced by gravity and uses the coordinates r and t:

$$\mathrm{v} = \frac{dr}{dt}\,, \tag{6}$$

$$\mathrm{v} = v\left(\frac{g_{00}}{g_{rr}}\right)^{1/2}\,, \tag{7}$$

$$\mathrm{v} = v g_{00}\,. \tag{8}$$

Here the last equation follows from (7) and (4). From (3) and (4) it follows that at $r = \infty$ $g_{00}(\infty) = g_{rr}(\infty) = 1$, which is why the coordinate velocity coincides with the velocity measured by an observer who resides infinitely far from the source of gravity (star).

When a body moves radially from $r = a$ to $r = b$ with a coordinate velocity $\mathrm{v}(r)$, the following coordinate time elapses:

$$t = \int_a^b \frac{dr}{\mathrm{v}}\,. \tag{9}$$

That is why to calculate the echo delay for a radar located far from the star the coordinate velocity v is relevant.

For a particle freely moving in a static gravitational field, the conserved energy can be introduced:

$$E = \frac{m\sqrt{g_{00}}}{\sqrt{1 - v^2}} \tag{10}$$

[see [5], Eqn (88.9)]. The law of energy conservation,

$$E(r = \infty) = E(r)\,, \tag{11}$$

allows us to determine $v(r)$:

$$1 - v^2 = (1 - \mathrm{v}_\infty^2)\, g_{00}\,. \tag{12}$$

Thus for a falling massive particle the proper velocity v increases, while it does not change for a photon: $v_\gamma = 1$. $\mathrm{v}(r)$ has a more complicated behavior:

$$1 - \frac{\mathrm{v}^2}{g_{00}^2} = (1 - \mathrm{v}_\infty^2)\, g_{00}\,. \tag{13}$$

For $r > R \gg r_g$, when the gravitational field is weak, from (13) and (4) we get:

$$\mathrm{v}^2 = \mathrm{v}_\infty^2 + \frac{r_g}{r}\,(1 - 3\mathrm{v}_\infty^2)\,. \tag{14}$$

For $\mathrm{v}_\infty \ll 1$, a well-known law for the motion of a non-relativistic particle in a gravitational field follows:

$$\mathrm{v}^2 = \mathrm{v}_\infty^2 + \frac{2MG}{r}\,. \tag{15}$$

For $\mathrm{v}_\infty = \mathrm{v}_c = 1/\sqrt{3}$, the coordinate velocity does not depend on r: $\mathrm{v} = \mathrm{v}_\infty$.

For $\mathrm{v}_\infty > \mathrm{v}_c$, a freely falling particle slows down!

Concerning nonradial motion, one easily finds that a massive body is deflected by the Sun at any speed, and the deflection angle θ is larger than that of light, θ_γ:

$$\theta = \frac{1}{2}\,\theta_\gamma (1 + \mathrm{v}_\infty^{-2})\,, \tag{16}$$

$$\theta_\gamma = \frac{2r_g}{R}\,. \tag{17}$$

Here R is the minimum distance between the photon, radiated by the distant star, and the Sun [see [7], problem 15.9, Eqn (13)].

3. Variation of v outside a black hole

For a strong gravitational field $r_g/r \sim 1$ we get

$$1 - \frac{\mathrm{v}^2}{(1 - r_g/r)^2} = (1 - \mathrm{v}_\infty^2)\left(1 - \frac{r_g}{r}\right), \tag{18}$$

$$\mathrm{v}^2 = \left[1 - (1 - \mathrm{v}_\infty^2)\left(1 - \frac{r_g}{r}\right)\right]\left(1 - \frac{r_g}{r}\right)^2. \tag{19}$$

For $\mathrm{v}_\infty = 1/\sqrt{3}$,

$$\mathrm{v}^2 = \frac{1}{3}\left(1 + \frac{2r_g}{r}\right)\left(1 - \frac{r_g}{r}\right)^2. \tag{20}$$

For large values of r ($r_g/r \ll 1$), one can neglect the $(r_g/r)^2$ and $(r_g/r)^3$ terms, and from (18) and (20) it follows that $\mathrm{v}^2 = 1/3$ and does not depend on r. For $r_g/r \sim 1$, the behavior of $\mathrm{v}(r)$ changes; in particular, it follows from (19) that $\mathrm{v}(r_g) = 0$ for any initial value v_∞.

For $\mathrm{v}_\infty > 1/\sqrt{3}$, the coordinate velocity monotonically decreases when r decreases. Thus even moderately relativistic particles behave like a photon.

For $\mathrm{v}_\infty < 1/\sqrt{3}$, a particle starts to accelerate and reaches a maximum coordinate velocity

$$\mathrm{v}_{\max} = \frac{2}{3\sqrt{3}}\frac{1}{1 - \mathrm{v}_\infty^2} \tag{21}$$

at

$$r_{\max} \equiv r(\mathrm{v}_{\max}) = \frac{3(1 - \mathrm{v}_\infty^2)}{1 - 3\mathrm{v}_\infty^2}\, r_g. \tag{22}$$

After this it starts to slow down, like a photon. Let us note that

$$v(\mathrm{v}_{\max}) = \frac{1}{\sqrt{3}}. \tag{23}$$

Thus the proper velocity equals $1/\sqrt{3}$ at zero coordinate acceleration.

4. Variation of v inside a star

As already stated in the introduction, we ignore all interactions except gravitation.

The metric inside a celestial body of constant density (oversimplified model of a star) was found by Schwarzschild [8] (see [4], section 11, § 6).

For a radial motion,

$$g_{00} = \frac{1}{4}\left[3\left(1 - \frac{r_g}{R}\right)^{1/2} - \left(1 - \frac{r_g r^2}{R^3}\right)^{1/2}\right]^2, \tag{24}$$

$$g_{rr} = \left(1 - \frac{r_g r^2}{R^3}\right)^{-1}, \tag{25}$$

where R is the radius of the star. Now $g_{rr} \neq 1/g_{00}$ and while Eqn (7) is still valid, Eqn (8) is not satisfied. For a weak field, when $r_g/R \ll 1$, (let us note that the condition $r_g/R < 1$ is necessary for star stability),

$$g_{00} = 1 - \frac{3}{2}\frac{r_g}{R} + \frac{1}{2}\frac{r_g r^2}{R^3}, \tag{26}$$

$$\mathrm{v} = v\left(\frac{g_{00}}{g_{rr}}\right)^{1/2} = v\left(1 - \frac{3}{2}\frac{r_g}{R} - \frac{1}{2}\frac{r_g r^2}{R^3}\right)^{1/2}$$

$$= v\left(1 - \frac{3}{4}\frac{r_g}{R} - \frac{1}{4}\frac{r_g r^2}{R^3}\right). \tag{27}$$

We observe that even photons start to accelerate when they cross the surface of a mirror star. (At the boundary $r = R$ the coordinate acceleration of photons changes sign.)

5. A thought experiment in a tower and the critical proper velocity

It is clear that the existence of the critical velocity v_c is not connected with the infinite distance between the observer and the gravitating body. Let us imagine a tower on the Earth's surface and an experimentalist with a clock having a metal ball and a metal mirror which reflects the ball's elastically. A series of experiments is performed. Each experiment consists of two parts. In the first part, the experimentalist throws the ball from the top of the tower with initial proper speed $v_0 = \mathrm{d}l^-/\mathrm{d}\tau^-$. The ball is reflected by the steel mirror situated at the bottom of the tower and bounces back to the top. The local time of this flight τ^+ is measured at the top of the tower.

In the second part, the experimentalist throws a ball from the bottom of the tower upwards with initial velocity $v_0 = \mathrm{d}l^+/\mathrm{d}\tau^+$. The ball is reflected by the mirror situated at the top of the tower and the time interval of the flight τ^- is measured. For nonrelativistic velocities $v_0 \ll c$, we evidently get $\tau^+ < \tau^-$. For photons $v_0 = 1$ and $\tau^+ > \tau^-$. The times τ^+ and τ^- are equal for $v_0 = v_c = 1/\sqrt{2}$. Thus, for $1/\sqrt{2} < v_0 \leqslant 1$, the ball travels more quickly when thrown from the bottom. This puzzling result originates from the time delay of the clock in a gravitational field; the same reason according to which photons look redshifted when they move from the bottom of the tower to the top (the famous Pound-Rebka experiment).

Deriving the value of the critical velocity $v_c = 1/\sqrt{2}$, we will see that its value is universal: it does not depend on the value of the gravitational potential and is valid for strong fields as well. The only condition is $h/r \ll 1 - r_g/r$, where h is the height of the tower. Thus the tower should not be too high (and tidal forces should not destroy the tower, the ball, and the experimentalist).

Let us derive formulas which illustrate our statements.

The local time interval for the observer at the top of the tower is

$$\mathrm{d}(\tau^+)^2 = g_{00}^+\,\mathrm{d}t^2 \equiv \left(1 - \frac{r_g}{r^+}\right)\mathrm{d}t^2, \tag{28}$$

where r^+ is the radial coordinate of the top of the tower in the Schwarzschild coordinates. The time of flight of the ball according to the clock of this observer is

$$\tau^+ = 2\int\frac{\mathrm{d}r}{\mathrm{d}r/\mathrm{d}\tau^+} = \frac{2}{(1 - r_g/r^+)^{1/2}}\int\frac{\mathrm{d}r}{\dfrac{1}{1 - r_g/r^+}\dfrac{\mathrm{d}r}{\mathrm{d}t}}. \tag{29}$$

This result immediately follows from Eqns (9) and (28).

To find the expression for dr/dt, let us write the expression for the local speed (which is coordinate invariant):

$$v = \left(\frac{g_{rr}}{g_{00}}\right)^{1/2} \frac{dr}{dt} = \frac{1}{1 - r_g/r} \frac{dr}{dt}, \tag{30}$$

and apply the energy conservation equation

$$\frac{1 - v_0^2}{1 - r_g/r^+} = \frac{1 - v^2}{1 - r_g/r}, \tag{31}$$

where v_0 is the initial speed of the ball. From Eqns (30) and (31), for the square of the denominator of the integral in (29) we get:

$$\frac{1}{(1 - r_g/r^+)^2}\left(\frac{dr}{dt}\right)^2 = \left(\frac{1 - r_g/r}{1 - r_g/r^+}\right)^2 \times \left[1 - \frac{1 - r_g/r}{1 - r_g/r^+}(1 - v_0^2)\right]. \tag{32}$$

Expressing $r = r^+ - \varepsilon$ (where ε varies from zero to h, h is the height of the tower in Schwarzschild coordinates) and expanding (32) with respect to ε we obtain

$$\frac{1}{(1 - r_g/r^+)^2}\left(\frac{dr}{dt}\right)^2 = v_0^2 + \frac{r_g}{(r^+)^2(1 - r_g/r^+)}\varepsilon(1 - 3v_0^2), \tag{33}$$

where the weakness of the gravitational field is unnecessary and the only requirement is: $h \ll r^+ - r_g$.

Substituting Eqn (33) into Eqn (29), we obtain

$$\tau^+ = \frac{2}{(1 - r_g/r^+)^{1/2}} \times \int_0^h \frac{d\varepsilon}{v_0\left[1 + \dfrac{r_g(1 - 3v_0^2)}{2(r^+)^2(1 - r_g/r^+)\,v_0^2}\,\varepsilon\right]}, \tag{34}$$

where we suppose that $v_0^2 \gg h r_g/(r^+)^2 = 2gh$.

The same derivation applied to an observer at the bottom of the tower leads to the following expression for the time of flight of the ball thrown by him:

$$\tau^- = \frac{2}{(1 - r_g/r^-)^{1/2}} \times \int_0^h \frac{d\varepsilon}{v_0\left[1 - \dfrac{r_g(1 - 3v_0^2)}{2(r^-)^2(1 - r_g/r^-)\,v_0^2}\,\varepsilon\right]}, \tag{35}$$

where $r^- = r^+ - h$. For $v_0 = v_c = 1/\sqrt{3}$, the integrals in Eqns (34) and (35) coincide, though τ^+ will still be smaller than τ^- because of the difference in the factors which multiply integrals. To find the critical velocity, let us expand the expressions in Eqns (34) and (35) with respect to ε and calculate the integrals:

$$\tau^+ = \frac{2h}{v_0(1 - r_g/r^+)^{1/2}} - \frac{h^2}{v_0(1 - r_g/r^+)^{1/2}} \frac{r_g(1 - 3v_0^2)}{2(r^+)^2(1 - r_g/r^+)\,v_0^2}, \tag{36}$$

$$\tau^- = \frac{2h}{v_0(1 - r_g/r^-)^{1/2}} + \frac{h^2}{v_0(1 - r_g/r^-)^{1/2}} \frac{r_g(1 - 3v_0^2)}{2(r^-)^2(1 - r_g/r^-)\,v_0^2}.$$

Now we easily observe that these time intervals coincide for $v_0 \equiv v_c = 1/\sqrt{2}$:

$$\tau^+ = \tau^- = \frac{2h}{v_c\left[1 - 2r_g/(r^+ + r^-)\right]^{1/2}}. \tag{37}$$

Performing an analogous experiment in a mine, an experimentalist will find that v_c is absent. See Section 4.

6. Isotropic and harmonic coordinates

In this section we will demonstrate that the critical velocity v_c equals $1/\sqrt{3}$ in isotropic, harmonic, and all other asymptotically flat coordinates, related with Schwarzschild coordinates by the following relations:

$$t' = t\,, \quad r' = r + \text{const} + O\left(\frac{r_g^2}{r^2}\right).$$

According to well-known formulas for the interval in isotropic coordinates, we have ([4], chapter 8, §§ 1–3)

$$ds^2 = \frac{(1 - GM/2\rho)^2}{(1 + GM/2\rho)^2}\,dt^2 - \left(1 + \frac{GM}{2\rho}\right)^4 (d\rho^2 + \rho^2\,d\theta^2 + \rho^2 \sin^2\theta\,d\phi^2).$$

Thus in a weak field at $\rho \to \infty$ and for a radial motion ($d\theta = 0$, $d\phi = 0$) we have

$$ds^2 = \left(1 - \frac{2GM}{\rho}\right)dt^2 - \left(1 + \frac{2GM}{\rho}\right)d\rho^2.$$

In harmonic coordinates:

$$ds^2 = \frac{1 - GM/R}{1 + GM/R}\,dt^2 - \left(1 + \frac{GM}{R}\right)^2 d\mathbf{X}^2 - \frac{1 + GM/R}{1 - GM/R}\frac{G^2M^2}{R^4}(\mathbf{X}\cdot d\mathbf{X})^2,$$

where

$$X_1 = R\sin\theta\cos\phi\,, \quad X_2 = R\sin\theta\sin\phi\,,$$
$$X_3 = R\cos\theta\,, \quad R^2 \equiv \mathbf{X}^2.$$

Thus at $d\theta = 0$, $d\phi = 0$:

$$ds^2 = \frac{1 - GM/R}{1 + GM/R}\,dt^2 - \left[\left(1 + \frac{GM}{R}\right)^2 + \frac{1 + GM/R}{1 - GM/R}\frac{G^2M^2}{R^2}\right]dR^2.$$

In the limit $R \to \infty$,

$$ds^2 = \left(1 - \frac{2GM}{R}\right)dt^2 - \left(1 + \frac{2GM}{R}\right)dR^2$$

with an accuracy $O(r_g^2/R^2)$. Thus with this accuracy the metric is the same in the Schwarzschild, isotropic, and harmonic coordinates.

Performing a derivation analogous to that of Eqns (18) and (19) for arbitrary coordinates (r, t) in a spherically symmetric static metric, where for a radial motion we can write

$$ds^2 = g_{00}(r)\,dt^2 - g_{rr}(r)\,dr^2\,,$$

and generalizing this derivation to the case where the initial proper velocity v_0 is determined not at infinity, but at an arbitrary $r = r_0$, we get:

$$1 - \mathrm{v}^2 \frac{g_{rr}(r)}{g_{00}(r)} = (1 - v_0^2)\frac{g_{00}(r)}{g_{00}(r_0)}\,, \tag{38}$$

$$\mathrm{v}^2 = \left[1 - (1 - v_0^2)\frac{g_{00}(r)}{g_{00}(r_0)}\right]\frac{g_{00}(r)}{g_{rr}(r)}\,. \tag{39}$$

In weak fields:

$$g_{rr}(r) = \frac{1}{g_{00}(r)} + O\left(\frac{r_{\mathrm{g}}^2}{r^2}\right),$$

the coordinate velocity is:

$$\mathrm{v}^2 = \left[1 - (1 - v_0^2)\frac{g_{00}(r)}{g_{00}(r_0)}\right]g_{00}^2(r)\,. \tag{40}$$

One easily observes that the *coordinate* acceleration is zero when the local velocity $v_0 = 1/\sqrt{3}$ in all coordinates considered by us. For these coordinates at infinity $\mathrm{v} = v$, and $\mathrm{v}_{\mathrm{c}} = 1/\sqrt{3}$ at $r_0 = \infty$, precisely as stated at the beginning of this section.

7. Conclusion and acknowledgements

We demonstrated that in a spherically symmetric gravitational field, two critical velocities exist: proper $v_{\mathrm{c}} = 1/\sqrt{2}$ and coordinate $\mathrm{v}_{\mathrm{c}} = 1/\sqrt{3}$. The first is by definition invariant under change of coordinates; the second is with high accuracy the same in the Schwarzschild, isotropic, and harmonic coordinates, which are widely used to synchronize clocks on satellites, to analyze the motion of double neutron stars, and for other relativistic astrophysical objects. We criticized the terminology according to which the proper time is called 'physical' or 'genuine'.

The existence of critical velocities $c/\sqrt{3}$ and $c/\sqrt{2}$ does not contradict the unique role played by the speed of light c in the relativistic domain. However, the acquaintance with the critical velocities in Schwarzschild coordinates will help the reader to study better the general theory of relativity since these coordinates are 'sensible, useful and frequently used' (see [6], Vol. 2, p. 526).

We are grateful to S Deser and B Tekin, who pointed out references [9, 10] to us (after we published paper [11]).

We are grateful to the UFN anonymous referee of this paper. Due to his efforts, the paper was substantially expanded.

We are partially supported by FS NTP FYaF 40.052.1.1.1112. SB is partly supported by RFBR grant 02-02-16500 and ILE of the Osaka University. LO is partly supported by the A von Humboldt Award.

References

1. Ashby N, Allen D W *Radio Science* **14** 649 (1979)
2. Guinot B *Metrologia* **34** 261 (1997)
3. Schwarzschild K *Sitzungsber. Deutsch. Akad. Wissenschaft Berlin Kl. Math. Phys. Tech.* 189 (1916)
4. Weinberg S *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley, 1972) [Translated into Russian (Moscow: Mir, 1975)]
5. Landau L D, Lifshitz E M *Teoriya Polya* (Classical Theory of Fields) (Moscow: Fizmatlit, 2001) [Translated into English (Oxford: Pegramon Press, 1983)]
6. Misner C W, Thorne K S, Wheeler J A *Gravitation* (San Francisco: W.H. Freeman, 1973) [Translated into Russian (Moscow: Mir, 1977)]
7. Lightman A, Press W, Price R, Teukolsky S *Problems Book in Relativity and Gravitation* (Princeton, NJ: Princeton Univ. Press, 1975) [Translated into Russian (Moscow: Mir, 1979)]
8. Schwarzschild K *Sitzungsber. Deutsch. Akad. Wissenschaft Berlin Kl. Math. Phys. Tech.* 424 (1916)
9. Carmeli M *Lett. Nuovo Cimento* **3** 379 (1972)
10. Carmeli M *Classical Fields: General Relativity and Gauge Theory* (New York: J. Wiley, 1982)
11. Blinnikov S I, Okun L B, Vysotsky M I, gr-qc/0111103; in *Proc. of the Workshops "What Comes Beyond the Standard Model 2000, 2001"* Vol. 1 *Festschrift Dedicated to the 60th Birthday of Holger Bech Nielsen* (Bled Workshops in Physics, Vol. 2, No 2, Eds N M Borštnik, C D Froggatt, D Lukman) (Ljubljana: DMFA–Založništvo, 2001) p. 115; hep-ph/0212221
12. Kobzarev I Yu, Okun L B, Pomeranchuk I Ya *Yad. Fiz.* **3** 1154 (1966) [*Sov. J. Nucl. Phys.* **3** 837 (1966)]

This page intentionally left blank

Fundamental Units: Physics and Metrology

Lev B. Okun

ITEP, Moscow, 117218, Russia

Abstract. The problem of fundamental units is discussed in the context of achievements of both theoretical physics and modern metrology. On one hand, due to the fascinating accuracy of atomic clocks, the traditional macroscopic standards of metrology (second, metre, kilogram) are giving way to standards based on fundamental units of nature: velocity of light c and quantum of action h. On the other hand, the poor precision of the gravitational constant G, which is widely believed to define the "cube of theories" and the units of the future "theory of everything", does not allow to use G as a fundamental dimensional constant in metrology. The electromagnetic units in SI are actually based on concepts of prerelativistic classical electrodynamics such as ether, electric permittivity and magnetic permeability of vacuum. Concluding remarks are devoted to terminological confusion which accompanies the progress in basic physics and metrology.

1 Introduction

The problem of fundamental units has many facets, three of which seem to be most important: theoretical, experimental and technological. At present they are inseparable. Theory, the so called Standard Model, formulates basic physical laws and mathematical methods of their application. Theoretical laws were established and continue to be established and tested on the basis of ingenious experiments and astronomical observations of higher and higher accuracy for an expanding space of parameters.

Precision experiments and observations, in their turn, are unthinkable without modern high technologies, including lasers and computers. These technologies are indissolubly connected with metrology – creation, perfection and unification of standards of physical units, while metrology is widely using the results of such theories as quantum mechanics, relativity theory, electrodynamics, condensed matter theory etc. Thus the circle is closed.

The situation is additionally complicated by the fact that the Standard Model is not a complete theory. It has many unsolved problems. Perhaps the most burning is the problem of the existence of fundamental scalar particles (higgses), responsible for the masses of all fundamental particles (from the heaviest one – the t-quark – to the lightest of the three neutrinos). We still do not understand the role of the three families of leptons and quarks. It would be too naive to think that their only justification is CP-violation. We still lack a successful theory unifying electroweak and strong interactions. The hypothesis of the existence of moderately violated supersymmetry – symmetry between fundamental bosons and fermions – is still not confirmed by experiments. Mathematical constructs of the type of superstrings and M-theory which in the beginning were considered as attempts to unify quantum gravity with electroweak and strong interaction, as time goes by, withdraw into a separate field of mathematics, whose practioners do not promise any applications to the real physical world. The situation might become drastically different if manifestations of extra space dimensions are discovered; in particular if laws of gravity turn out to change at TeV scale.

Sections 2–6 are devoted to the history of units based on c, h, G. Sections 7–9 deal with the units based on c, h, e and precision frequency measurements. Section 10 compares the Gaussian units and SI units. It is argued that while the latter are more convenient for practical purposes, the former allow to better understand the basic notions of modern physics. Therefore the use of both systems of units should be allowed in physics textbooks. Section 11 contains concluding remarks.

2 Fundamental Parameters and Units

The essence of theoretical physics is expressed by dimensionless equations for dimensionless quantities. However one cannot do experimental physics (and teach physics) without dimensional quantities and dimensional units.

In what follows we shall refer to dimensionless fundamental constants such as $e^2/\hbar c$, or m_e/m_p as fundamental parameters (here e is the electron charge, $\hbar$ is the reduced quantum of action ($\hbar = h/2\pi$) and of angular momentum, c is the velocity of light in vacuum, m_e and m_p are the electron mass and the proton mass, respectively).

In the absence of established terminology we shall refer to dimensional fundamental constants as fundamental units. Examples of units: c (for velocity), $\hbar$ (for action and angular momentum). According to our definition G is also a fundamental unit (though indirectly).

3 Planck Units

When in 1899-1900 Planck discovered h [1], he used this discovery to introduce universal units, which at present are written in the form

$$l_P = \hbar/m_P c \ , \quad t_P = \hbar/m_P c^2 \ , \quad m_P = (\hbar c/G)^{1/2} \ , \tag{1}$$

where G is Newton's gravitational constant.

Planck derived his units by using dimensional order of magnitude relations:

$$c = \frac{l_P}{t_P} \ , \quad \frac{G m_P^2}{l_P} = m_P c^2 \ , \quad \frac{G m_P^2}{l_P} t_P = \hbar \ . \tag{2}$$

He was inspired by the idea that his units are universal (contrary to "handcrafted" earthbound ordinary units – meter, second, gram): they are the same at any far away corner of the universe.

Planck also considered as universal the Planck temperature $T_P = m_P c^2/k$. But Boltzmann's k is not a universal unit, it is a conversion factor: $k = 8.6 \cdot 10^{-5}$ eV/K (hint: $\hbar\omega/kT$).

4 c, h, G – Units

From the point of view of the future "theory of everything" it is natural to use c, h, G as fundamental dimensional constants.

In 1928 Gamov, Ivanenko and Landau [2] considered the theory "of the world as a whole" in terms of dimensional fundamental constant c, h, G. In 1928 Landau was 20 years old, Gamow and Ivanenko twentyfour (see Figs. 1,2). They had written the paper "World Constants and Limiting Transitions" (see Fig. 3) as a humorous birthday present to their friend, a young lady. None of them ever referred to this paper in their subsequent publications. But the ideas of the paper were fundamental. In 1936 Bronstein [3] worked on a theory in which all three constants are finite. It was one of the first papers on relativistic quantum gravity. In 1967 ideas of [2,3] were presented in the form of a cube (see Figs. 4,5) by Zelmanov [4]. Later on it was further developed by others [5,6].

The vertices of this cube represent nonrelativistic mechanics (NM), nonrelativistic gravity (NG), nonrelativistic quantum mechanics (QM), special relativity (SR), quantum field theory (QFT), general relativity (GR) and finally relativistic quantum gravity (QGR) or theory of everything (TOE).

The cube, made of units, is "endowed" with dimensionless parameters like α, α_s, mixing angles, mass ratios, etc. Their values are expected to follow from TOE. Similar to the cube is "dimensional pyramid" (Kuchar [7], Sanchez [8]) with 4 vertices and 4 planes, Fig. 6.

Note that Einstein tried to build a unified theory of electricity and gravity ("TOE") in the left-hand vertical plane of the cube, without Quantum Mechanics, without $\hbar$.

Planck units allow one to deal in the equations of TOE only with dimensionless functions of dimensionless variables and dimensionless parameters of the type $\alpha = e^2/\hbar c$ or m_e/m_p. Conceptually Planck units are excellent, but practically they have serious shortcomings, caused by G, the same G which allows to bring gravitation and cosmology into the realm of quantum phenomena. Thus the source of strength at the same time turns out to be a source of weakness.

5 Planck Units Are Impractical

The obvious shortcoming of Planck units is that they differ by many orders of magnitude from atomic units commonly used in physics. Their values are natural for the early universe and TOE, but not for mundane physics:

$$l_P = 10^{-35}\ \mathrm{m}\ ,\ \ t_P = 10^{-43}\ \mathrm{s}\ ,\ \ E_P = m_P c^2 = 10^{19}\ \mathrm{GeV}\ \ .$$

The energy which corresponds to the Planck mass is unattainable by accelerators even of the remotest future. (Note, however, that it is only a few orders

Fig. 1. Meeting in Kharkov, 1928, attended by Gamow, Ivanenko, and Landau

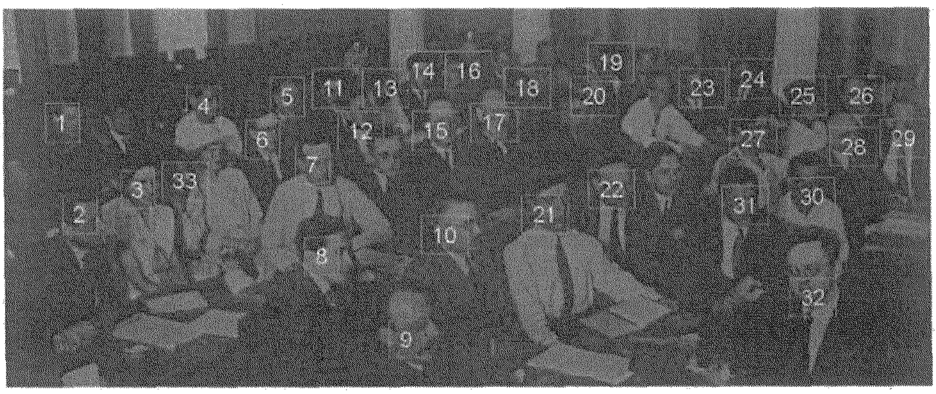


Fig. 2. Who is who in Fig.1.

1. Dunaev	8. Ivanenko	15. Efimovich	22. Mandel	29. Frenkel
2. Heitler	9. Obreimov	16. Ogievetsky	23. Vereschagin	30. Rosenkevich
3. Arsenyeva	10. Fock	17. Grommer	24. Slutsky	31. Finkelshtein
4. Davydov	11. Leipunsky	18. Muskelischvili	25. Gamov	32. Jordan
5. Todorovich (?)	12. Kopp	19. Korsunsky	26. Shubnikov	33. Timoreva
6. Frish	13. Kotsarova	20. Gorwitz	27. Landau	
7. Bursian	14. Khalfin	21. Ambartsumian	28. Shtrumm	

ЖУРНАЛ
РУССКОГО
ФИЗИКО-ХИМИЧЕСКОГО
ОБЩЕСТВА

ПРИ ЛЕНИНГРАДСКОМ УНИВЕРСИТЕТЕ

ЧАСТЬ ФИЗИЧЕСКАЯ

Ответственный редактор А. Ф. Иоффе
Помощник редактора П. С. Тартаковский

ТОМ LX—1928

ГЛАВНОЕ УПРАВЛЕНИЕ НАУЧНЫМИ УЧРЕЖДЕНИЯМИ (ГЛАВНАУКА)
ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
МОСКВА 1928 ЛЕНИНГРАД

ОГЛАВЛЕНИЕ IX ТОМА.—INHALT DES LX BANDES.

Выпуск 1. Стр.

Fig. 3. The title page and a part of the contents of the journal where the paper by G. Gamow, D. Ivanenko, and L. Landau was published.

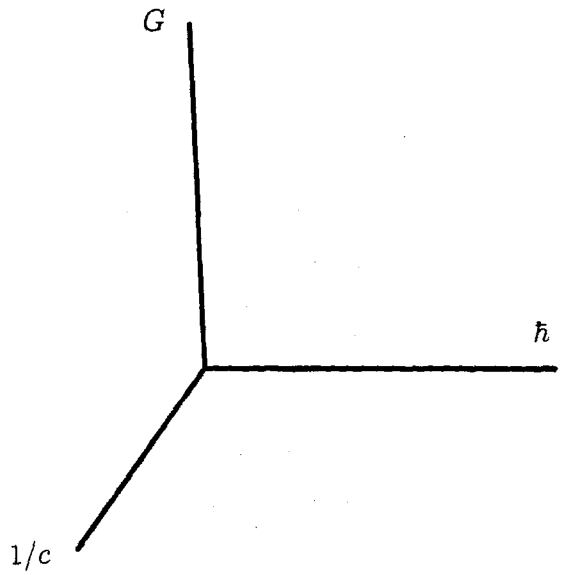


Fig. 4. Orthogonal axes. At the origin there is no gravity, no maximal velocity, no quantum effects

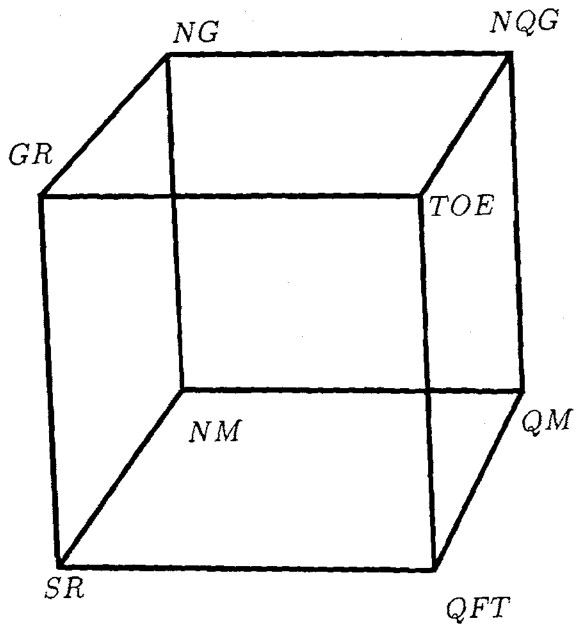


Fig. 5. Cube of theories

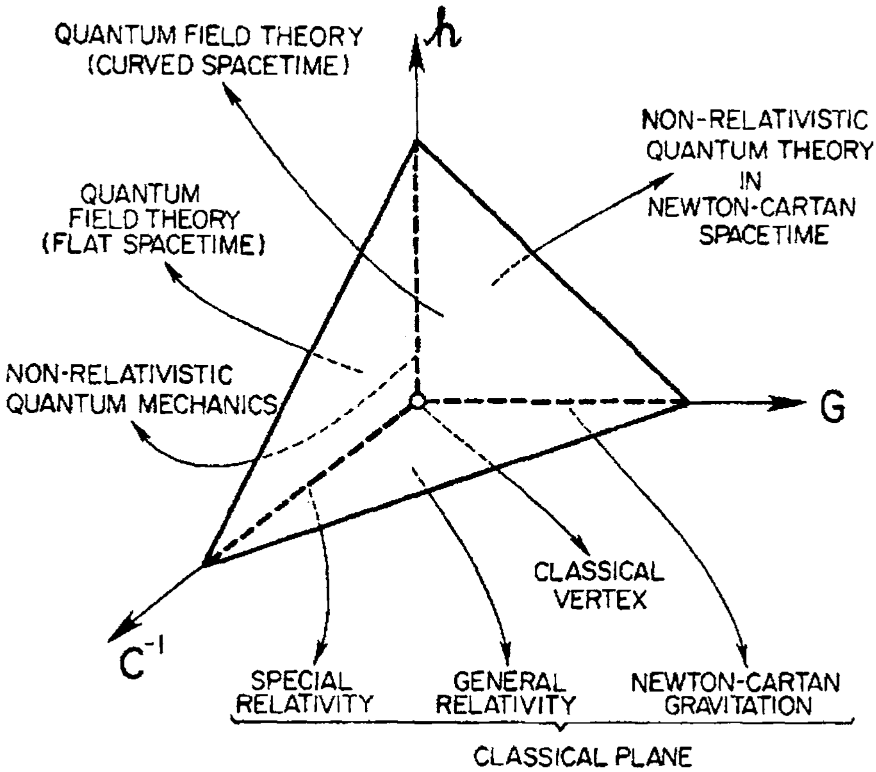


Fig. 6. Dimensional pyramid

of magnitude larger than the grand unification scale of electroweak and strong interactions.) The Planck units of length and time are vanishingly small compared with atomic units. Of course the huge powers of ten are not frightful by themselves. As is well known, atomic units also differ by many orders of magnitude from SI units, which does not prevent atomic standards to be the base of modern metrology.

Much more essential is another shortcoming of Planck units, which stems from the fact that G is known with rather poor accuracy [9] (by several orders worse than those of c and h and by approximately ten orders worse than the precision of atomic clocks). Thus it is impossible to use the Planck units as standards in modern precision physics and technology.

6 Units of Stoney

Planck's use of G as a basis for defining the unit of mass was caused by the absence at the beginning of the 20th century of another natural, not "handcrafted", candidate for the unit of mass. In that respect Planck's universal units resemble the universal units suggested 30 years earlier by the Irish physicist

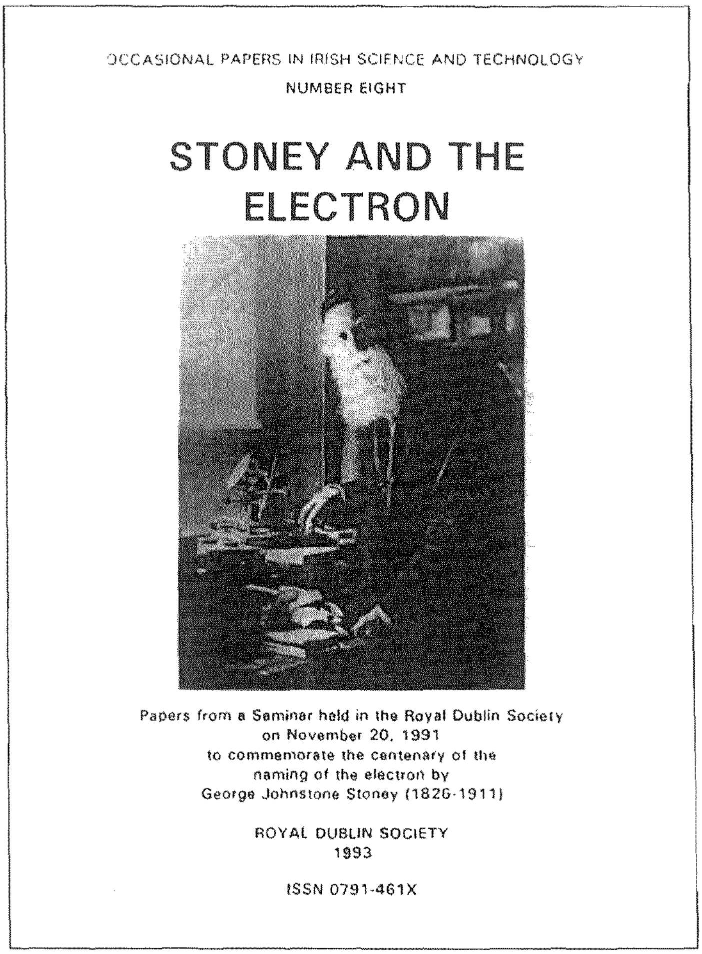


Fig. 7. Cover of the book dedicated to G.J. Stoney (with permission of the Royal Dublin Society)

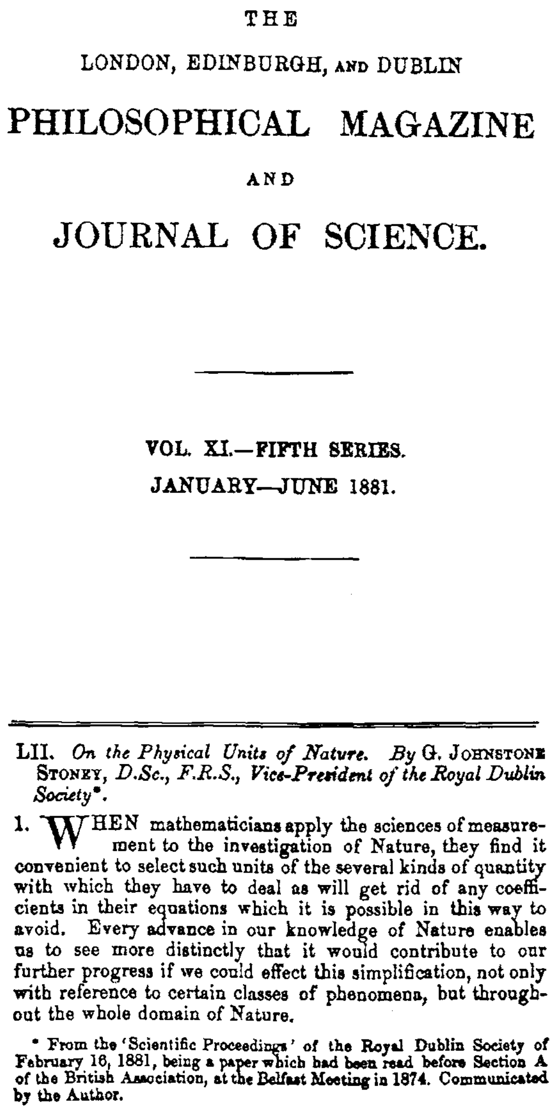
Stoney (1826 - 1910), secretary of the Royal Dublin Society (see Figs. 7,8). By studying electrolysis, he was the first who measured the value of elementary charge e and introduced into physics the term "electron" for the carrier of this charge (in modern terminology he was referring to ions). From e, c, G Stoney [10] constructed in 1870 - 1880 universal units with dimensions of length, time and mass:

$$l_S = e\sqrt{G}/c^2 \ , \quad t_S = e\sqrt{G}/c^3 \ , \quad m_S = e/\sqrt{G} \ , \tag{3}$$

which he derived from dimensional equations:

$$c = l_S/t_S \ , \quad e^2 = Gm_S^2 \ , \quad e^2/l_S = m_S c^2 \ .$$

Let us note that the units of Stoney are only by a factor $\sqrt{\alpha}$ smaller than those of Planck.



THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

VOL. XI.—FIFTH SERIES.

JANUARY—JUNE 1881.

LII. *On the Physical Units of Nature.* *By* G. JOHNSTONE STONEY, *D.Sc., F.R.S., Vice-President of the Royal Dublin Society**.

1. WHEN mathematicians apply the sciences of measurement to the investigation of Nature, they find it convenient to select such units of the several kinds of quantity with which they have to deal as will get rid of any coefficients in their equations which it is possible in this way to avoid. Every advance in our knowledge of Nature enables us to see more distinctly that it would contribute to our further progress if we could effect this simplification, not only with reference to certain classes of phenomena, but throughout the whole domain of Nature.

* From the 'Scientific Proceedings' of the Royal Dublin Society of February 16, 1881, being a paper which had been read before Section A of the British Association, at the Belfast Meeting in 1874. Communicated by the Author.

Fig. 8. The title page of "Philosophical Magazine" and the beginning of article [10] by G.J. Stoney

Stoney's units look "tailored" for Einstein's unified theory. Constants e, c, G contain the gist of classical electrodynamics and gravity. There is no $\hbar$ in them. Comparison with c, $\hbar$, G shows that $\hbar$ is brought into Stoney's set of constants "through the backdoor of α". Therefore e, c, G do not form a cube of theories with its limiting transitions considered by Gamov, Ivanenko and Landau [2].

7 Atomic Clocks and c

During the 20th century the situation with standards of mass (time, length) has changed drastically. The fundamental identity of elementary particles and hence of atoms produced many candidates for a standard of mass, known with much, much better precision than G. Thus, from the point of view of dimensions, the necessity to use G disappeared. However from the point of view of unifying physics the Planck units became even more attractive.

Let us now look at two other fundamental constants: c and h.

Let us start from c and the frequencies of light and radio waves. In the second half of the 20th century physicists learned how to measure them in a digital way by counting the number of crests. This raised the accuracy of atomic (Cesium-133) clocks (first suggested by I. Rabi in 1945) to the level of 1 second in 300 years (NBS, 1955). (Now this has become 1 second in $20 \cdot 10^6$ years: LPTF, NIST, PTB.) But even the first figure was sufficient for the introduction into SI of an atomic unit of a second (in 1967):

"1 s = 9 192 631 770 periods of radiation in the transition between levels of hyperfine splitting of the atomic ground state of Cs-133".

This, together with the independence of the velocity of light on its frequency, impelled Bay et al. [11] to suggest, instead of the unit of length (meter), to use as the basic unit the unit of velocity, namely the velocity of light c. In 1983 the definition

$$c = 299\,792\,458 \text{ m/s} \tag{4}$$

was introduced into the SI. The traditional standard of length gave way to the new standard based on the value of the velocity c. This velocity is defined as a number without uncertainty. Further improvements of experiments which measured c would mean further improvement of the realization of the meter. An international report "Practical realization of the definition of the metre, including recommended radiations of other optical frequency standards" (2001) was published by T. Quinn in 2003 [12]. (Note that both spellings "metre" and "meter" are used in the literature, the former in metrology, while the latter one in physics.)

Further progress in the accuracy of atomic clocks is connected with passing from microwave to optical frequencies [13,14].

8 Towards a Kilogram Based on h

Thus metrology made two momentous steps in the direction of fundamental physics: the place of macroscopic clocks and ruler (the famous rod at BIPM, in Sèvres, near Paris) became occupied by the velocity of light and by atoms of Cs-133. There remains now only one macroscopic standard - the kilogram at Sèvres. The prospect of expressing it through the quantum of action h is connected with precision measurements in atomic and condensed matter physics. There are many promising quantities which are good candidates for such measurements. I shall touch upon only one project which is connected with two outstanding discoveries in condensed matter physics: the Josephson effect [15] (Nobel Prize 1973) and the von Klitzing effect [16] (Nobel Prize 1985).

Josephson theoretically predicted the existence of a supercurrent and its remarkable properties. A supercurrent is a current of Cooper pairs tunneling through an insulator separating two superconductors. A supercurrent can exist without external voltage. An external voltage V creates an alternating supercurrent of frequency ν. The steps in V are given by the relation:

$$V(n) = \nu n / K_J \ , \tag{5}$$

where n is an integer, while the coefficient K_J is universal and is called the Josephson constant. It is reproduced in various experiments with unprecedented accuracy and is determined only by the ratio of fundamental constants:

$$K_J = 2e/h \ . \tag{6}$$

The effect, discovered by von Klitzing, is called the quantum Hall effect. This effect shows that there exists in Nature a universal electric resistance, one which can be expressed in terms of fundamental constants.

As is well known, the ordinary Hall effect occurs in a solid conductor (or semiconductor) with density of current $\mathbf{j}$ in a magnetic field $\mathbf{H}$ which produce an electric field $\mathbf{E}$ (with voltage V_H) orthogonal both to $\mathbf{j}$ and $\mathbf{H}$.

The quantum Hall effect was discovered in a two-dimensional electron system separating two parts of a silicon field transistor at very low temperature (< 4 K) and very strong magnetic field (~ 14 Tesla). It was established that the Hall resistance

$$R_H = V_H / I \ , \tag{7}$$

where I is the total current, has quantum jumps:

$$R_H(n) = \frac{R_K}{n} \ , \tag{8}$$

where n is an integer, while R_K is the von Klitzing constant:

$$R_K = h/e^2 \ . \tag{9}$$

It is obvious that

$$h = 4K_J^{-2} R_K^{-1} \ . \tag{10}$$

This permits the measurement of h using a macroscopic apparatus. A special two-story-high watt balance compared electrical and mechanical forces:

$$VI/v = mg \ , \tag{11}$$

where m is the measured mass of a body, g the local gravitational acceleration, V the voltage in a coil moving with a vertical velocity v in a magnetic field, while I is the current in the same coil, this time fixed in the same magnetic field. By calibrating V and V/I through the Josephson and von Klitzing effects Williams et al. [17] succeeded in connecting h and the kilogram within an uncertainty of $8.7 \cdot 10^{-8}$.

It is hoped that in the not too distant future this accuracy might be improved by an order of magnitude, which would allow to use the watt balance for gauging the standards of mass and thus get rid of the Sèvres kilogram and to define the value of h. As a result the value of h would have no uncertainties in the same way as it occurred with c. Thus fundamental units of nature c and h would become fundamental SI units of metrology.

9 Kilogram as Frequency ν_K

Another definition of the kilogram has been suggested [18] on the basis of equations

$$E = h\nu \ , \tag{12}$$

$$E = mc^2 \ : \tag{13}$$

"The kilogram is the mass of a body at rest whose equivalent energy equals the energy of collection of photons whose frequencies sum to $13.5639274 \times 10^{49}$ hertz".

This definition should be taken with a grain of salt. The combined use of (12) and (13) implies that a photon of frequency ν has mass $h\nu/c^2$. This implication persists in spite of the words "equivalent energy". The words "the mass of the body at rest" imply that mass is not Lorentz invariant, but depends on the velocity of a reference frame. It would be proper to replace (13) by

$$E_0 = mc^2 \ , \tag{14}$$

where E_0 is the rest energy (see e.g. [19]). But then it would take some additional considerations in order to define the frequency ν_K corresponding to one kilogram. In particular, massive atoms emitting and absorbing photons should be taken into account. From a practical point of view the measurement of "frequencies sum" of order 10^{50} hertz is by eight orders of magnitude more difficult than that of the Planck frequency $\nu_P = 1/t_P$.

10 Electromagnetism and Relativity

Electromagnetism – the kinship of electricity and magnetism – discovered in 1820 by Oersted, rather soon became the foundation of Ampére's electrodynamics. The development of the latter by Faraday and other outstanding physicists culminated in 1873 in the Treatise of Maxwell [20] who linked electric currents with electric and magnetic fields and with the properties of light. None of these great physicists knew the genuine nature of the phenomena. Maxwell considered a vacuum filled with ether; the carriers of charges were unknown to him. The electromagnetic field was described by four vector quantities: electric field $\mathbf{E}$, electric induction (or displacement) $\mathbf{D}$, magnetic field $\mathbf{H}$, and magnetic induction (or flux density) $\mathbf{B}$.

On the basis of these notions practical units (such as volt, ampere, coulomb, joule) were introduced by International Electrical Congresses in the 1880s. The electric permittivity ε_0 and magnetic permeability μ_0 ascribed by Maxwell to the ether were accepted by the community of engineers and physicists: $\mathbf{D} = \varepsilon_0 \mathbf{E}$, $\mathbf{B} = \mu_0 \mathbf{H}$. In the middle of the 20th century these practical units became the basis of the Système International d'Unités (SI).

The end of the 19th and beginning of the 20th century were marked by great successes in understanding and applying classical electrodynamics. On the practical side it was the use of electric currents in industry, transport and radio communications. On the theoretical side it was the unification of electrodynamics, optics and mechanics in the framework of special relativity [21].

According to special relativity, the position four-vector is $x^i = (ct, \mathbf{r})$ ($i = 0, 1, 2, 3$), the momentum four-vector is $p^i = (E/c, \mathbf{p})$, the four-potential of electromagnetic field $A^i = (\varphi, \mathbf{A})$, the density of the four-current $j^i = (c\rho, \mathbf{j})$, where $\mathbf{j} = \rho \mathbf{v}$, and $\rho = e\delta(\mathbf{r} - \mathbf{r}_a)$, e is the electric charge. (The current j^i is consistent with the definitions of p^i and A^i, due to an appropriate coefficient c in front of ρ. The source of the field, the charge, is pointlike. Otherwise there appears a problem of the field inside the finite-size cloud of charge.) The upper index i of a four-vector indicates a contravariant four-vector; a lower index i indicates a covariant four-vector, its space components have a minus sign. Raising or lowering of indices is done with the diagonal metric tensors g^{ik} or g_{ik} respectively.

The three-vectors $\mathbf{E}$ and $\mathbf{H}$ are components of the four-tensor of the electromagnetic field

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \ . \tag{15}$$

The tensors F_{ik} and F^{ik} can be represented by matrices:

$$F_{ik} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -H_3 & H_2 \\ -E_2 & H_3 & 0 & -H_1 \\ -E_3 & -H_2 & H_1 & 0 \end{pmatrix} , \tag{16}$$

and

$$F^{ik} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -H_3 & H_2 \\ E_2 & H_3 & 0 & -H_1 \\ E_3 & -H_2 & H_1 & 0 \end{pmatrix} , \tag{17}$$

respectively, or in a condensed form:

$$F_{ik} = (\mathbf{E}, \mathbf{H}) \ , \tag{18}$$

$$F^{ik} = (-\mathbf{E}, \mathbf{H}) \ . \tag{19}$$

This four-tensor is obviously antisymmetric. From the definition of F_{ik} it follows that the dimensions of $\mathbf{E}$ and $\mathbf{H}$ are the same: $[\mathbf{E}] = [\mathbf{H}]$.

The field equations have the form in Gaussian units:

$$\frac{\tilde{F}^{ik}}{\partial x^k} = 0 \ , \tag{20}$$

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c}j^i \ . \tag{21}$$

Here

$$\tilde{F}^{ik} = \frac{1}{2}\varepsilon^{iklm}F_{lm} \ , \tag{22}$$

where ε^{iklm} is fully antisymmetric tensor ($\varepsilon^{0123} = +1$).

The equation describing the motion of a charge in the electromagnetic field is given by

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + \frac{e}{c}[\mathbf{vH}] \ , \tag{23}$$

where

$$\mathbf{v} = \frac{\mathbf{p}c^2}{E} \ . \tag{24}$$

Note that according to special relativity there is no ether, $\varepsilon_0 \equiv \mu_0 \equiv 1$, and the strength of magnetic field in vacuum **H** has the same dimension as that of **E**; the identity of $\varepsilon_0 \equiv \mu_0 \equiv 1$ immediately follows from the fact that the same e determines the action of the charge on the field and of the field on the charge. (See expression for the action in [22], (27.6).) Thus, there is no need to consider **B** and **D** in the case of vacuum. In classical electrodynamics they appear only in the continuous media due to polarization of the latter [23].

In a number of classical monographs and textbooks on classical electrodynamics **E** and **H** are consistently used for the description of electric and magnetic fields in vacuum with $\varepsilon_0 \equiv \mu_0 \equiv 1$ ([21,22,24–26]). Their authors use Gaussian or Heaviside-Lorentz (with $1/4\pi$ in the Coulomb law) units.

Many other authors use **B** instead of **H**, sometimes calling **B** magnetic field and sometimes magnetic induction in vacuum [27]. Most of them use the SI units, according to which ε_0 and μ_0 are dimensional: $\mu_0 = 4\pi \cdot 10^{-7}\mathrm{HA}^{-2}$, $\varepsilon_0\mu_0 = c^2$, where H is henry, while A is the ampere. The classical electromagnetic fields in vacuum are described by four physical quantities **D**, **H** and **E**, **B**, all four of them having different dimensions at variance with the spirit of special relativity.[1] In that respect, the vacuum is similar to a material body. The SI units are very convenient for engineers, but not for theorists in particle physics.

In fact, theorists are not less responsible than metrologists for the gap between the deductive basis of modern physics and the mainly prerelativistic inductive basis of modern metrology. A good example is the 1935 article [28] by A. Sommerfeld and his book "Electrodynamics" based on lectures given in 1933-34 [29].

His argument against an absolute system (that is based on units of time, length and mass) was the presence of fractional exponents (for instance from the Coulomb law the dimension of charge is $\mathrm{g}^{1/2}\ \mathrm{cm}^{3/2}\mathrm{s}^{-1}$). This argument was not very compelling in the 1930s and is even less so today. His argument against the

[1] Sometimes one can hear that the identity $\varepsilon_0 \equiv \mu_0 \equiv 1$ is similar to putting $c = 1$, when using c as a unit of velocity. However this similarity is superficial. In the framework of special relativity one can use any unit for velocity (for instance, m/s). But the dimensions and values of ε_0 and μ_0 are fixed in SI.

Gaussian or Heaviside-Lorentz system was based on an inductive, prerelativistic view on electromagnetism. Though he was not quite happy[2] with the new clumsy expression for the fine structure constant α introduced by him during World War I, he kept insisting on MKSA units and against Gaussian units. His authority was not the least in the decision to legally enforce after World War II the SI as the obligatory system of units for all textbooks in physics.

Coming back to classical electrodynamics, let us note that it is not a perfect theory: it has serious problems at short distances. To a large part these problems are solved by quantum electrodynamics (QED). Therefore the latter should be used as a foundation of a system of electromagnetic units. By the way, QED is used to extract the most accurate value of α from the precision measurements of the magnetic moment of the electron.

In the framework of QED, α is not a constant but a function of momentum transfer due to the polarization of vacuum. Let us stress that this polarization has nothing to do with purely classical non-unit values of ε_0 and μ_0.

11 Concluding Remarks

The mutually fruitful "crossing" of fundamental physics and metrology gives numerous practical applications. One of them should be specially mentioned: the use of general relativity in global positioning systems [30,31].

Remarkable achievements of metrology are not always accompanied by elaboration of adequate terminology. Here we will mention only a few of wide-spread delusions.

The choice of c as a unit of velocity leads many authors to the false conclusion that c should be excluded from the set of fundamental units. They insist that $c = 1$, because c in units of c is equal to 1. (The same refers to h in units of h.) But the number 1 is not a unit of measurement, because such units are always dimensional. Equations $c, h = 1$ are simply wide spread jargon. Some authors go even further by identifying space and time. (A detailed discussion can be found in [32].)

The number of physical units is not limited. When solving a given problem the choice of units is determined by considerations of convenience. However, from the point of view of "the world as a whole" c, h and G (or instead of G some other quantity representing gravity) are definitely singled out as fundamental dimensional constants. Of course they must be accompanied by a number of dimensionless parameters. But the number of fundamental units could not be less than three [32].

The inclusion of the candela into the set of base units (see Fig. 9) seems to be unconvincing from the point of view of physics. Of course, practically it is convenient to use it when discussing the brightness of light. But it does not look logical to put it on the same footing as units of length, time and mass.

[2] "What is especially painful for me is that the fine structure constant is no more $e^2/\hbar c$, but $e^2/4\pi\varepsilon_0\hbar c$". Z. Phys. **36** (1935) 818.

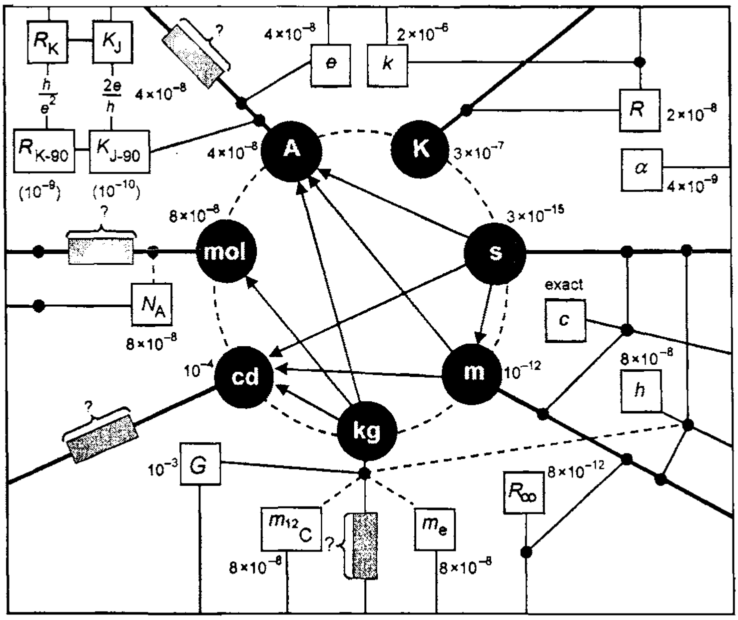


Fig. 9. The base units of the SI, with their present uncertainties of realization, and some of their links to atomic and fundamental constants with their present uncertainties in terms of the SI. The absence of a useful quantitative estimate of the long-term stability of the kilogram, indicated by "?", is reflected in three of the other base units. The dashed lines to the kilogram indicate possible routes to a new definition. (See the article by T.J. Quinn "Base Units of the Système International d'Unités, their Accuracy, Dissemination and International Traceability", Metrologia **31** (1994/95) 515-527.)

As SI is imposed on the physics literature by governmental laws, the obligatory usage in textbooks of such notions as permittivity ε_0 and permeability μ_0 of vacuum, makes it difficult to appreciate the beauty of the modern electrodynamics and field theory. It corresponds to the prerelativistic stage of physics.

This list can be extended, but it seems that the above remarks are sufficient for a serious discussion. The metrological institutes and SI are of great importance for science and technology. Therefore the metrological legal documents should be to a greater degree based on modern physical concepts. Especially they should give more freedom to the usage of Gaussian and Heaviside-Lorentz systems of units in the textbooks.

Acknowledgements

I would like to thank A. Clairon, H. Fritzsch, J. Jackson, S. Karshenboim, N. Koshelyaevsky, H. Leutwyler, E. Peik, N. Sanchez, M. Tatarenko, V. Telegdi, Th. Udem, L. Vitushkin and H. Wagner for fruitful discussions. The work has been partly supported by the Russian Federal Special Scientific-Technological

Program of Nuclear Physics Fund 40.052.1.1.1112 and by A. von Humboldt award.

References

1. M. Planck: Sitz.-Ber. Preuss. Akad. Wiss. (1899) 449;
   M. Planck: Ann. d. Phys. **4**, 1, 69 (1900). Reprinted in Max Planck. Physikalische Abhandlungen und Vorträge, Band I, Fried. Vieweg. 1958, pp. 500-600, 614-667.
2. G. Gamov, D. Ivanenko, L. Landau: Journal of Russian Physicochemical Society, Ser. Phys. **LX**, 13 (1928) (in Russian), reprinted in Yad. Fiz. **65**, 1406 (2002), translated into English in Physics of Atomic Nuclei **65**, 1373 (2002).
3. M. Bronstein: Phys. Zeitschrift der Sowjetunion **9**, 140 (1936).
4. A. Zelmanov: 'Kosmologia'. In *Razvitie Astronomii v SSSR* (Nauka, Moskva, 1967) p. 323 (in Russian).
5. G. Gorelik: *Razmernost' prostranstva* (MGU, Moskva, 1983) Chapter 5 (in Russian).
6. L. Okun: Sov. Phys. Usp. **34**, 818 (1991).
7. K. Kuchar: Phys. Rev. D **22**, 1285 (1980).
8. N. Sanchez: 'Recent Progress in String Cosmology'. *International school of astrophysics "D. Chalonge", 5th Course "Current topics in astrofundamental physics", Erice, September 1996.*
9. J. Flowers and B. Petley: In: *Astrophysics, Clocks and Fundamental Constants*, ed. by S. G. Karshenboim and E. Peik, Lecture Notes in Physics Vol. 648 (Springer, Berlin, Heidelberg 2004), pp. 75–93.
10. G.J. Stoney: The Philosophical Magazine and Journal of Science **11**, 381 (1881).
11. Z. Bay, G.G. Luther, J.A. White: Phys. Rev. Lett. **29**, 189 (1972);
    Z. Bay, J.A. White: Phys. Rev. D **5**, 796 (1972).
12. T.J. Quinn: Metrologia **40**, 103 (2003).
13. Th. Udem, R. Holzwarth, T.W. Hänsch: Nature **416**, 233 (14 March 2002).
14. S.G. Karshenboim, F.S. Pavone, F. Bassani, M. Inguscio, T.W. Hänsch (Eds.): *The Hydrogen Atom. Precision Physics of Simple Atomic Systems* (Springer, 2001).
15. B.D. Josephson: Phys. Lett. **1**, 251 (1962).
16. K.V. Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980).
17. E.R. Williams, R.L. Steiner, D.B. Newell, P.T. Olsen: Phys. Rev. Lett. **81**, 2404 (1998).
18. R.N. Taylor, P.J. Mohr: Metrologia **36**, 63 (1999).
19. L.B. Okun: Physics Today, June 1989, 31; May 1990, 13, 15, 115-117; Uspekhi Fiz. Nauk **158**, 511 (1989); Sov. Phys. Usp. **32**, 629 (1989).
20. J.C. Maxwell: *A Treatise on Electricity and Magnetism.* (Oxford Univ. Press, 1873) Part IV, Chapters IX, X, XIX. 3rd edn. (1891); reprint Dover, New York (1954)
21. A. Einstein: *The meaning of relativity*, 4th edn. (Princeton, 1953)
22. L.D. Landau and E.M. Lifshitz: *The Classical Theory of Fields*, 4th edn. (Pergamon Press, Oxford, 1987)
23. L.D. Landau and E.M. Lifshitz: *Electrodynamics of continuous Media*, 2nd edn. (Addison-Wesley, Reading, MA, 1984)
24. C.Möller: *The Theory of Relativity*, 2nd edn. (Clarendon Press, Oxford, 1974)
25. P.G. Bergmann: *Introduction to the Theory of Relativity* (Dover reprint, 1976)
26. W. Pauli: *Theory of Relativity* (Pergamon Press, New York, 1958)
27. J.D. Jackson: *Classical Electrodynamics*, 3rd edn. (John Wiley and sons, Inc. 1998)

28. A. Sommerfeld: Phys. Z. **36**, 814 (1935)
29. A. Sommerfeld: *Electrodynamik* (Leipzig, 1949. Academic Press, New York, 1952)
30. N. Ashby, D. Allen: Radio Science **14**, 649 (1979).
31. B. Guinot: Metrologia **34**, 261 (1997).
32. M.J. Duff, L.B. Okun, G. Veneziano: JHEP **03**, 023 (2002), arXiv, physics/0110060.

THE VIRUS OF RELATIVISTIC MASS IN THE YEAR OF PHYSICS*

L.B. OKUN

ITEP, Moscow, 117218, Russia

The "famous formula" $E = mc^2$ and the concept of "relativistic mass" increasing with velocity, which follows from it, are historical artifacts, contradicting the basic symmetry of Einstein's Special Relativity, the symmetry of 4-dimensional space-time. The relation discovered by Einstein is not $E = mc^2$, but $E_0 = mc^2$, where E_0 is the energy of a free body at rest introduced by Einstein in 1905. The source of the longevity of the "famous formula" is the irresponsible attitude of relativity theory experts to the task of explaining it to the non-experts.
The notion of "relativistic mass" presents a kind of pedagogical virus which very effectively infects new generations of students and professors and shows no signs of decline. Moreover in the Year of Physics it threatens to produce a real pandemia.

Before writing my first article against the "Einstein famous equation $E = mc^2$" I mentioned in 1987 this intention to Volodya Gribov in one of our daily conversations. We had a complete unanimity on the issue of physics. But Volodya was a better friend of mine than I myself. And his understanding of life was better than mine. Therefore he tried to dissuade me from wasting my time on a fight against an obviously wrong cliché, which I would inevitably loose. I discarded his wise advice and wrote my first two papers on the concept of mass [1,2].

The subject seemed important to me because it concerned the proper teaching of special relativity at high schools, colleges and universities and explaining its genuine meaning to a wide audience of non-physicists, the so-called "pedestrians" in popular science magazines and books.

The task looked also not absolutely formidable because a consistent presentation of relativity existed for a long time in the world-wide accepted text-book by Landau and Lifshitz [3], which was the basis of my own understanding, and in some other text-books.

*The paper was written for the Gribov Memorial Volume dedicated to the 75th birthday of V.N. Gribov.

The existence of countless texts, in which the essence of relativity was mutilated (or semi-mutilated) had two sides. On one hand, it looked discouraging, especially because among the authors of these texts there were many famous physicists, the fathers and greatest authorities of modern physics. On the other hand, it was a challenge. So I tried to explain clearly to the readers the beauty of four-dimensional space-time approach and the ugliness and inconsistency of "relativistic mass", an illegitimate child of relativistic and non-relativistic equations.

My optimism had increased when in 1992 Taylor and Wheeler in the second edition of the influential and popular "*Spacetime Physics*" [4] included a "Dialog: Use and Abuse of the Concept of Mass", in which they supported my articles [1,2]. A copy of this book is in my bookcase with a postcard sent to me in October 1991 by John Archibald Wheeler. The postcard has a photo of the famous Albert Einstein Memorial in front of the building of the National Academy of Sciences, Washington, DC. The bronze sculpture of Einstein includes a copy book with $E = mc^2$ on an open page.

Since that time I received hundreds of letters from physicists (both professors and students) stating their adherence to the four-dimensional formulation of relativity and to the Lorentz invariant concept of mass. In a few cases I helped the authors to correct erroneous explanations of the concept of mass in preparing new editions of their textbooks. However the number of proponents of relativistic mass seemed not to decrease.

A leading role in promoting the relativistic mass have played the books by Max Jammer [5,6]. Especially aggressive the proponents of relativistic mass became in connection with the World Year of Physics, which marks the 100th anniversary of fundamental articles published by Einstein in 1905.

The campaign started by the September 2004 issue of "*Scientific American*", full with "relativistic mass" equal to $m_0/\sqrt{1 - v^2/c^2}$, where m_0 is rest mass, and "the most famous equation $E = mc^2$". A letter to the editors, defending the four-dimentional approach and invariant mass had been rejected by the editor G. Collins who in April 2005 wrote: "Most important, we believe that tackling the issue head-on in the manner you and your coauthors want in the letters column of *Sci. Am.* would be very confusing to our general audience and it would make the subject seen all the more mysterious and impenetrable to them". Thus to avoid "head-on" collision of correct and false arguments the editors of *Sci. Am.* preferred to hide from the readers the correct viewpoint.

P. Rodgers – the Editor of European "*Physics World*" wrote in January 2005 in editorial [7]: "... $E = mc^2$ led to the remarkable conclusion that mass and energy are one and the same". Unlike G. Collins, P. Rodgers published a letter criticizing this statement and partly agreed with the criticism [8].

In September 2005 the bandwagon of relativistic mass was joined by "*The New York Times*", which published an article by B. Green [9].

The journalists were supported by renowned scientists, such as R. Penrose, who in a new thousand pages thick book had written [10]:

"In a clear sense mass and energy become completely equivalent to one another according to Einstein's most famous equation $E = mc^2$."

How many students, teachers and journalists will be infected by this sentence? How many readers had been infected by the famous book by S. Hawking [11], the second edition of which appeared in 2005? On the very first page of it Hawking wrote:

"Someone told me that each equation I included in the book would halve the sales. I therefore resolved not to have any equations at all. In the end, however, I did put in one equation, Einstein's famous equation $E = mc^2$. I hope that this will not scare off half of my potential readers."

I am sure that the usage of $E = mc^2$ had doubled the sales of his book, the buyers being attracted by the famous brand. But is it possible to estimate the damage done to their understanding of relativity theory and to the general level of the literature on relativity incurred by this case of spreading the virus.

Two recent preprints by Gary Oas [12,13] written in the framework of Educational Program for Gifted Youth at Stanford University were devoted to the use of relativistic mass. The author "urged, once again, that the use of the concept at all levels to be abandoned" [12]. The manuscript has been submitted for publication to the "*Americal Journal of Physics*", but was rejected as being "too lengthy" (it contains 12 pages!). A lengthy bibliography (on 30 pages) of books referring to special and/or general relativity is provided in Ref. [13] to give a background for discussions of the historical use of the concept of relativistic mass. It is easy to forecast the aggressive reaction of the virus infected community to this attempt to cure it.

I am grateful to the grant of Russian Ministry of Education and Science No. 2328.2003.2 for partial support of this work.

References

1. L.B. Okun, *Soviet Physics-Uspekhi* **32** (July 1989) p. 629-638.
2. L.B. Okun, *Physics Today*, June 1989, p. 31-36; http://www.physicstoday.org/vol-42/iss-6/vol42no6p31_36.pdf; L.B. Okun, *Physics Today*, May 1990, p. 147.
3. L.D. Landau and E.M. Lifshitz, *"The Classical Theory of Fields"*, translated from the second Russian ed., Pergamon, New York, 1951.
4. E.F. Taylor and J.A. Wheeler, *"Spacetime Physics"*, 2nd ed. Freeman and Company. New York, 1992, p. 246-252.
5. M. Jammer, *"Concepts of Mass in Classical and Modern Physics"*, Harvard, 1961.
6. M. Jammer, *"Concepts of Mass in Contemporary Physics and Philosophy"*, Princeton, 2000.
7. P. Rodgers, *Physics World* **18**, No. **1**, 13 (2005).
8. P. Rodgers, *Physics World* **18**, No. **10**, 20 (2005).
9. B. Green, *"That Famous Equation and You"*, *The New York Times*, Sept. 30, 2005.
10. R. Penrose, *"The road to reality. A complete guide to the laws of the universe"*, A. Knopf, New York, 2004, p. 434.
11. S. Hawking, *"A Brief History of Time"*, New York, 1988.
12. G. Oas, *"On the abuse and use of relativistic mass"*, arXiv: physics/0504110.
13. G. Oas, *"On the Use of Relativistic Mass in Various Published Works"*, arXiv: physics/0504111.

THE CONCEPT OF MASS IN THE EINSTEIN YEAR

L.B. Okun[a]
State Research Center, Institute for Theoretical and Experimental Physics, 117218, Moscow, Russia

Abstract. Various facets of the concept of mass are discussed. The masses of elementary particles and the search for higgs. The masses of hadrons. The pedagogical virus of relativistic mass.

1 From "Principia" to Large Hadron Collider (LHC)

The term "mass' was introduced into mechanics by Newton in 1687 in his "*Principia*" [1]. He defined it as the amount of matter. The generally accepted definition of matter does not exist even today. Some authors of physics text-books do not consider photons – particles of light – as particles of matter, because they are massless. For the same reason they do not consider as matter the electromagnetic field. It is not quite clear whether they consider as matter almost massless neutrinos, which usually move with velocity close to that of light. Of course it is impossible to collect a handful of neutrinos similarly to a handful of coins. But in many other respects both photons and neutrinos behave like classical particles, while the electromagnetic field is the basis of our understanding of the structure of atoms. On the other hand, the so-called weak bosons W^+, W^-, Z^0 are often not considered as particles of matter because they are too heavy and too short-lived.

Even more unusual are such particles as gluons and quarks. Unlike atoms, nucleons, and leptons, they do not exist in a free state: they are permanently confined inside nucleons and other hadrons.

There is no doubt that the problem of mass is one of the key problems of modern physics. Though there is no common opinion even among the experts what is the essence of this problem. For most of particle theorists, as well as members of LHC community, the solution of the problem is connected with the quest and discovery of the higgs – scalar boson which in the Standard Model is responsible for the masses of leptons and quarks and their electroweak messengers: W and Z. The discovery of higgs and the study of higgs sector might elucidate the problem of the pattern of hierarchy of masses of leptons and quarks: from milli electron Volts for neutrinos to about 180 GeV for t-quark. For many physicists it is a QCD problem: how light quarks and massless gluons form massive nucleons and atomic nuclei. Still for majority of confused students and science journalists there is no difference between mass of a body m and its energy E divided by c^2: they believe in the "most famous formula $E = mc^2$".

If higgs exists, its discovery will depend on the funding of the particle physics. In 1993 the termination of the SSC project sent the quest for the higgs into a painful knockdown. The decision not to order in 1995 a few dozen of extra superconducting cavities prevented, a few years later, LEP II from crossing the 115 GeV threshold for the mass of the higgs.

If we are lucky and higgs is discovered around year 2010 at LHC, then the next instrument needed to understand what keeps the masses of the higgs below 1 TeV scale, is ILC (International linear collider). This machine would provide a clean environment for the study of higgs production and decays. It could also be used for discovery and study of light supersymmetric particles (SUSY). A prototype of ILC was suggested a few years ago by DESY as the project TESLA. There was no doubt that if funded, TESLA would work, but the funding was not provided by the German government. The new variant of ILC envisions increasing the maximal center of mass

[a]e-mail: okun@itep.ru

2

energy of colliding electron and positron from) 0.5 TeV to 1 TeV. If everything goes well, ILC can start before 2020.

Further increase of energy, to say, 5 TeV, would call for a machine of the type of CLIC (Compact linear collider) the project of which is under discussion at CERN for more than a decade. In this machine the role of clystrons is supposed to play a low energy but very high current "decelerator" the energy of which would be pumped into the high energy accelerator part of CLIC. Unlike situation with ILC, even the mere feasibility of CLIC is not clear now. Special experimental research to ascertain the feasibility is going on at CERN.

The discussion of higgses, neutrinos and QCD in connection with the fundamental problems of mass is often accompanied and even overshadowed by a "pseudoproblem" of the so-called "relativistic mass" (see section 5).

2 Mass in Newtonian Mechanics

The more basic is a physical notion, the more difficult to define it in words. A good example give the 1960s editions of "Encyclopedia Britannica" where energy is defined in terms of work, while the entry "work" refers to labour and professional unions. Most people have intuitive notions of space and time. Every physicist has intuitive notions of energy, mass, and momentum. But practically everybody has difficulties in casting these notions into words without using mathematics.

Though the definition of mass ("Definition I: The quantity of matter is the measure of the same, arising from its density and bulk conjointly") given by Newton in his "*Principia*" [1] was so unclear that scholars are discussing its logical consistency even today, the equations of Newtonian mechanics are absolutely self-consistent. Mass m enters in the relations of velocity $\mathbf{v} = d\mathbf{r}/dt$ and momentum $\mathbf{p}$:

$$\mathbf{p} = m\mathbf{v} \ , \tag{1}$$

as well as acceleration $\mathbf{a} = d\mathbf{v}/dt$ and force $\mathbf{F}$:

$$\mathbf{F} = d\mathbf{p}/dt = m\mathbf{a} \ . \tag{2}$$

It also enters in the equation defining the force of gravity with which a body with mass m_1 at point $\mathbf{r}_1$ attracts another body with mass m_2 at point $\mathbf{r}_2$:

$$\mathbf{F}_g = -Gm_1m_2\mathbf{r}/r^3 \ . \tag{3}$$

Here $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, $r = |\mathbf{r}|$, while G is the famous Newton constant:

$$G = 6.67 \cdot 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2} \ . \tag{4}$$

The kinetic energy of a body is defined as

$$E_k = \mathbf{p}^2/2m = m\mathbf{v}^2/2 \ . \tag{5}$$

The potential gravitational energy:

$$U_g = -Gm_1m_2/r \ , \tag{6}$$

while the total energy in this case is

$$E = E_k + U_g \ . \tag{7}$$

The total energy is conserved. When a stone falls on the earth, its potential energy decreases (becomes more negative), kinetic energy increases: so that the total energy does not change. When the stone hits the ground, its kinetic energy is shared by the ambient molecules raising the local temperature.

One of the greatest achievements of the XIX century was the formulation of the laws of conservation of energy and momentum in all kinds of processes.

At the beginning of the XX century it was realized that conservation of energy is predetermined by uniformity of time, while conservation of momentum – by uniformity of space.

But let us return to the notion of force. People strongly felt the force of gravity throughout the history of mankind, but only in XVII century the equations (3) and (6) were formulated.

An important notion in this formulation is the notion of gravitational potential φ_g. The gravitational potential of a body with mass m_1 is

$$\varphi_g = -\frac{Gm_1}{r} \tag{8}$$

Thus the potential energy of a body with mass m_2 in a potential (8) is

$$U_g = m_2\varphi_g \ , \tag{9}$$

which coincides with eq. (6).

A century later similar equations were formulated for another long-range interaction, the electrical one:

$$\mathbf{F}_e = e_1e_2\mathbf{r}/r^3 \ , \tag{10}$$

$$U_e = e_1e_2/r \ , \tag{11}$$

$$\varphi_e = e_1/r \ , \tag{12}$$

$$U_e = e_2\varphi_e \ . \tag{13}$$

In these equations, which define the Coulomb force, Coulomb potential energy, and Coulomb potential respectively, e_1 and e_2 are electrical charges of two bodies.

An important role in the theory of electricity is played by the strength of electric field $\mathbf{E}$. Charge e_1 creates field with strength

$$\mathbf{E} = e_1\mathbf{r}/r^3 \ . \tag{14}$$

Thus

$$\mathbf{F}_e = e_2\mathbf{E} \tag{15}$$

As most of matter around us is electrically neutral, the electrical interaction was known for centuries only as a kind of trifle. Unlike mass m, which is always positive, the charge e has two varieties: positive and negative. Two charges of opposite sign attract each other, while those of the same sign are repelling. Protons residing in the nucleus of an atom have positive charge, the charge of electrons, which form atomic shells, is negative. As a result the atom is electrically neutral.

Electrical interaction and its ramifications determine the main features of atoms, molecules, condensed matter, and biological cells. Gravitational interaction is too feeble to play any role at that level. To see this consider an electron and a proton. Their masses, respectively, are

$$m_e = 9.1 \cdot 10^{-31}\ \text{kg}\ , \ \ m_p = 1.7 \cdot 10^{-27}\ \text{kg}\ . \tag{16}$$

Their electric charges:

$$e_p = -e_e = e = 4.8 \cdot 10^{-10}\ \text{esu}\ \ , \tag{17}$$

4

where esu denotes electrostatic unit:

$$(1\ \mathrm{esu})^2 = 10^{-9}\ \mathrm{kg\ m^3 s^{-2}}\ . \tag{18}$$

Hence

$$-e_e e_p = 2.3 \cdot 0^{-28}\ \mathrm{kg\ m^3 s^{-2}}\ . \tag{19}$$

On the other hand

$$G m_e m_p = 1.03 \cdot 10^{-67}\ \mathrm{kg\ m^3 s^{-2}}\ , \tag{20}$$

Thus in an atom the gravitational force is $\sim 10^{-40}$ of electric one.

The importance of gravity for our every day life is caused by the huge number of atoms in the earth, and hence by its very large mass:

$$M = 5.98 \cdot 10^{24}\ \mathrm{kg}\ . \tag{21}$$

Taking into account the value of the radius of the earth

$$R = 6.38 \cdot 10^6\ \mathrm{m}\ , \tag{22}$$

we find the gravitational force of the earth acting on a body with mass m close to its surface:

$$\mathbf{F}_g = m\,\mathbf{g}\ , \tag{23}$$

where $\mathbf{g}$ is acceleration directed towards the center of the earth:

$$g = |\mathbf{g}| = \frac{6.67 \cdot 10^{-11} \cdot 5.98 \cdot 10^{24}}{(6.38 \cdot 10^6)^2}\mathrm{m\ s^{-2}} = 9.8\ \mathrm{m\ s^{-2}}\ . \tag{24}$$

Let us note that for gravity $\mathbf{g}$ plays the role of the strength of gravitational field, which is analogous to the role of $\mathbf{E}$ for electricity. The acceleration $\mathbf{g}$ does not depend on the mass or any other properties of the attracted body. In that sense it is universal. This universality was established by Galileo early in the 17th century in his experiments with balls rolling down an inclined plane. One can see this plane, with little bells ringing when a ball passes them, in Florence. (Apocryphal history tells that Galileo had discovered universality of g by dropping balls from the Tower of Pisa.)

Gravitational and electric interactions are the only long-range interactions the existence of which has been established. Two other fundamental interactions, referred to as strong and weak, have very short ranges: 10^{-15} m and 10^{-18} m respectively. Their first manifestations were discovered about a century ago in the form of radioactivity. Their further study has lead to new disciplines: nuclear physics and the physics of elementary particles.

Quite often you might find in the literature, in the discussion of Newtonian mechanics, the terms "inertial mass" m_i and "gravitational mass" m_g. The former is used in equations (1), (2), (5), defining the inertial properties of bodies. The latter is used in equations (3), (6), (23), describing the gravitational interaction. After introducing these terms a special law of nature is formulated:

$$m_i = m_g\ , \tag{25}$$

which is called upon to explain the universality of g.

However Galileo had discovered this universality before the notion of mass was introduced by Newton, while from Newton equations (1), (2), (3) the universality of g follows without additional assumptions. Thus the notions and notations m_i and m_g are simply redundant. As we will see later, their introduction is not only redundant, but contradicts the General Theory of Relativity, which explains why the same mass m enters equations (1) - (3).

The advocates of m_i and m_g argue by considering the possibility that in the future the more precise experimental tests might discover a small violation of Galilean universality. But that would mean that a new feeble long-range force exists in nature. In literature this hypothetical force is often referred to as a "fifth force" (in addition to the four established ones). When and if the "fifth force" is discovered it should be carefully studied. But at present it should not confuse the exposition of well established physical laws. Especially confusing and harmful are m_i and m_g in the text-books for students.

At that point it is appropriate to summarize the properties of mass in Newtonian mechanics:

1. Mass is a measure of the amount of matter.
2. Mass of a body is a measure of its inertia.
3. Masses of bodies are sources of their gravitational attraction to each other.
4. Mass of a composite body is equal to the sum of masses of the bodies that constitute it; mathematically that means that mass is additive.
5. Mass of an isolated body or isolated system of bodies is conserved: it does not change with time.
6. Mass of a body does not change in the transition from one reference frame to another.

3 Mass in Special Relativity

Of great conceptual importance in modern science is the principle of relativity first stated by Galileo: A rectilinear motion of a physical system with constant velocity relative to any external object is unobservable within the system itself. The essence of this principle was beautifully exposed in the famous book "Dialogue Concerning the Chief World Systems – Ptolemaic and Copernican", published in 1632 [2]:

"Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal: jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way). Have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that, you will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. In jumping, you will pass on the floor the same spaces as before, nor will you make larger jumps toward the stern than toward the prow even though the ship is moving quite rapidly, despite the fact that during the time you are in the air the floor under you will be going in a direction opposite to your jump. In throwing something to your companion, you will need no more force to get it to him whether he is in the direction of the bow or the stern, with yourself situated opposite. The droplets will fall as before into the vessel beneath without dropping toward the stern, although while the drops are in the air the ship runs many spans. The fish in their water will swim toward the front of their bowl with no more effort than toward the back, and will go with equal ease to bait placed anywhere around the edges of the bowl. Finally the butterflies and flies will continue their flights indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air. And if smoke is made by burning some incense, it will be seen going up in the form of a little cloud, remaining still and moving no more toward one side than the other. The cause of all these correspondences of effects is the fact that the ship's motion is common to all the things contained in it, and to the air also. That is why I said you should be below decks; for if this took place above in the open air, which would not follow the course of the ship, more or less noticeable differences would be seen in some of the effects noted."

Sometimes one can hear that the ship of Galileo was discussed two centuries earlier by cardinal Nicolaus Cusanus (1401 - 1464) in his book "De docta ignorata" ("On the scientific ignorance") published in 1440. Indeed, one can read in volume II, at the beginning of chapter XII "The properties of the earth":

"It is clear to us that the earth is actually moving, though we do not see this, as we feel the movement only through comparison with a point at rest. Somebody on a ship in the middle of waters, without knowing that water is flowing and without seeing the shores, how could he ascertain that the ship is moving?" [3] [b]. The relative character of motion is expressed in these lines quite clearly. But the cabin of Galileo's ship is full of various phenomena and experiments, proving that observable effects look the same in any inertial reference frame. At this point we define an inertial reference frame, as that which moves rectilinearly with constant velocity with respect to the stars. We shall give a more accurate definition when considering General Theory of Relativity.

If the velocity of the ship is $\mathbf{u}$ and it moves along the axis x, the coordinates of two inertial frames are connected by equations:

$$t' = t\ , \quad x' = x + ut\ , \quad y' = y\ , \quad z' = z\ , \tag{26}$$

where $u = |\mathbf{u}|$, the primed coordinates refer to the shore, while unprimed to the ship. From the definition of velocity $\mathbf{v} = d\mathbf{r}/dt$ one easily sees that

$$\mathbf{v}' = \mathbf{v} + \mathbf{u} \tag{27}$$

and that $\mathbf{v}'$, $\mathbf{p}'$, $\mathbf{a}'$, $\mathbf{F}'$, $\mathbf{F}'_g$, E'_k, U'_g, $\mathbf{F}'_e$, U'_e satisfy the same equations (1) - (11) as their unprimed analogues.

Galilean principle of relativity is the quintessence of Newtonian mechanics. Nevertheless the latter is called non-relativistic mechanics, as opposed to Einsteinian mechanics which is called relativistic. This is one of many examples of lack of complete consistency in the language of physics which is a natural product of its evolution. The point is that Newtonian mechanics satisfies the Galilean principle of relativity only partially. The cabin of the original Galilean ship did not contain apparatuses that were able to measure the velocity of light. This velocity was first established in 1676 by Danish astronomer O. Roemer (1644 - 1710), who deduced from the observations of the moons of Jupiter, performed by J. Cassini, that it is $2.4 \cdot 10^5$ km s^{-1}. Further measurements during three centuries established its present value: $c = 3 \cdot 10^5$ km s^{-1}.

Of greatest importance was the discovery made two centuries later by American physicists A. Michelson and E. Morley. By using a special two arm rotating interferometer they established that the velocity of light did not depend on the angle between

[b] I am grateful to Peter Zerwas for arousing my interest to Nicolaus Cusanus.

the light ray and the vector of velocity of the earth on its journey around the sun. In this experiment the earth itself played the role of Galilean ship. That result signalled that the simple law (27) of addition of velocities is not valid for light.

This, in its turn, meant that the coordinate transformations (26) - (27) should be changed when v and/or u (due to relative character of velocity) are of the order of c.

This change had been performed by H. Lorentz (1904) [4], H. Poincare (1905) [5,6] and A. Einstein (1905) [7, 8]. Lorentz considered deformation of electron moving through the so called ether, filling all the universe, and introduced primed spatial and time coordinates, as purely auxiliary quantities. Poincare and Einstein wrote transformations between primed and unprimed coordinates:

$$t' = (t + ux/c^2)\gamma \; , \;\; x' = (x + ut)\gamma \; , \;\; y' = y \; , \;\; z' = z \; , \tag{28}$$

where

$$\gamma = 1/\sqrt{1 - u^2/c^2} \; . \tag{29}$$

They were called Lorentz transformations by Poincare and later by Einstein.

Poincare believed in ether and considered that the remaining problem is to understand it. Einstein simply dispensed with ether, he considered transformations (28) - (29) as a direct expressions of properties of space and time. Galilean relativity of inertial motion resulted in relativity of simultaneity, of time, and of length.

Proceeding from his article [7] Einstein came [8] to a fundamental conclusion that a body at rest has rest-energy E_0:

$$E_0 = mc^2 \; . \tag{30}$$

Here m is the mass of the body, while index 0 in E_0 indicates that this is the energy in the body's rest frame.

In 1906 M. Planck explicitly wrote the expressions of total energy E and momentum $\mathbf{p}$ of a body with arbitrary value of its velocity $\mathbf{v}$:

$$E = mc^2\gamma \; , \;\; \mathbf{p} = m\mathbf{v}\gamma \; , \tag{31}$$

where

$$\gamma = (1 - v^2/c^2)^{-1/2} \; . \tag{32}$$

These expressions can be easily derived by assuming that E and $\mathbf{p}$ transform in the same way as t and $\mathbf{r}$, each pair $(E, \mathbf{p}c)$ and $(t, \mathbf{r}/c)$ forming a four-dimensional vector. Indeed, by applying Lorentz transformations to a body at rest, taking into account relation (30) and writing v instead of u, we come to (31) - (32). Of course, the isotropy of space should be also accounted for.

The notion of four-dimensional space-time was introduced in 1908 by G. Minkowski [10]. While four-vectors transform under Lorentz transformations (rotations in Minkowskian pseudo-Euclidian space), their squares are Lorentz-invariant:

$$\tau^2 = t^2 - (\mathbf{r}/c)^2 \; , \tag{33}$$

$$m^2c^4 = E^2 - (\mathbf{p}c)^2 \; . \tag{34}$$

Here τ is the so-called proper time, while m, as before, is the mass of a body. But now it acquires a new meaning, which was absent in Newtonian mechanics. (Note that for a body at rest ($\mathbf{p} = 0$) one recovers from eq. (34) the relation (30) between mass and rest-energy.)

It is impossible to discuss the concept of mass without explicitly basing the discussion on the achievements of XX century physics and especially on the notion of elementary particles such as electrons, photons and neutrinos, less elementary, such

8

as protons and neutrons (in which quarks and gluons are confined), or composite, such as atoms and atomic nuclei. It is firmly established that all particles of a given kind (for instance all electrons) are identical and hence have exactly the same value of mass. The same refers to protons and neutrons.

Atoms and atomic nuclei ask for further stipulations because each of these composite systems exists not only in its ground state, but can be brought to one of its numerous excited states (energy levels). For instance, a hydrogen atom is a bound system of a proton and electron, attracted to each other by Coulomb force (10). As proton is approximately two thousand times heavier than electron, one usually speaks about electron moving in the Coulomb potential (11) of proton. According to the laws of quantum mechanics this movement is quantized, forming a system of levels. The lowest level is stable, the excited ones are unstable. Electron jumps from a higher level to a lower one by emitting a quantum of light - photon. Finally it reaches the ground level.

The energy in atomic, nuclear and particle physics is measured in electron Volts: 1 eV is the energy which electron gains by traversing a potential of 1 Volt; 1 keV = 10^3 eV, 1 MeV = 10^6 eV, 1 GeV = 10^9 eV, 1 TeV = 10^{12} eV, 1 PeV = 10^{15} eV, 1 EeV = 10^{18} eV.

The binding energy of electron at the ground level of hydrogen atom is 13.6 eV. Due to relation between rest-energy and mass it is convenient to use as a unit of mass 1 eV/c^2. The mass of electron $m_e = 0.511$ MeV/c^2, the mass of proton $m_p = 0.938$ GeV/c^2. The mass of a hydrogen atom in its ground state is by 13.6 eV/c^2 smaller than the sum $m_e + m_p$. This mass difference is often referred to as defect of mass.

As c is a universal constant, it is appropriate to use it in relativistic physics as a unit of velocity and hence to put $c = 1$ in all above values of masses and defects of mass. In what follows we will use as units of mass eV and its derivatives: keV, MeV, GeV, etc.

One eV is a tiny unit when compared with Joule (J) or kilogram:

$$1\ \mathrm{J} = 6.24 \cdot 10^{18}\ \mathrm{eV}\ ,\ 1\ \mathrm{eV} = 1.6 \cdot 10^{-19}\ \mathrm{J}\ , \tag{35}$$

$$1\ \mathrm{kg} = 5.61 \cdot 10^{35}\ \mathrm{eV}\ ,\ 1\ \mathrm{eV} = 1.78 \cdot 10^{-36}\ \mathrm{kg}\ . \tag{36}$$

However it is four orders of magnitude larger than one degree of Kelvin (K).

$$1\ \mathrm{K} = 0.86 \cdot 10^{-4}\ \mathrm{eV}\ ,\ 1\ \mathrm{eV} = 1.16 \cdot 10^{4}\ \mathrm{K} \tag{37}$$

(In eq. (37) we put dimensional Boltzmann factor k equal to unity, taking into account that kT, where T is temperature, characterizes the mean energy of an ensemble of particles.) Let us estimate the relative change of mass in a few everyday processes.

The light from the sun is absorbed by vegetation on the earth to produce carbohydrates via reaction of photosynthesis:

$$\mathrm{light} + 6CO_2 + 6H_2O = 6O_2 + C_6H_{12}O_6\ .$$

The total energy of light required to produce one molecule of $C_6H_{12}O_6$ is about 4.9 eV. This does not mean that the photons are massive. They are massless, but the kinetic energy of photons is transformed into the rest energy of carbohydrates.

A combustion of methane in the gas burner of a kitchen stove:

$$CH_4 + 2O_2 \to CO_2 + 2H_2O\ . \tag{38}$$

In this reaction 35.6 MJ of heat is released per cubic meter of methane. Since the density of methane is 0.72 kg/m^3 and density of oxygen is 1.43 kg/m^3

$$\frac{\Delta m}{m} = \frac{35.6 \cdot 6.24 \cdot 10^6 \cdot 10^{18}}{(0.72 + 2 \cdot 1.43) \cdot 5.61 \cdot 10^{35}} = 1.1 \cdot 10^{-10}\ ,$$

where in the nominator eq.(35) and in denominator eq.(36) are used.

We can look at this result differently by starting from Avogadro number:

$$N_A = 6.022 \cdot 10^{23} \text{ mol}^{-1} \tag{39}$$

and molar volume (for ideal gas)

$$22.4 \cdot 10^{-3} \text{ m}^3 \cdot \text{mol}^{-1} \ .$$

This means that a cubic meter of methane contains $2.69 \cdot 10^{25}$ molecules. Thus, burning of one molecule of methane releases

$$\frac{35.6 \cdot 6.24 \cdot 10^{24}}{2.69 \cdot 10^{25}} = 8.3 \text{ eV}$$

Now we estimate the mass of one molecule of methane plus 2 molecules of oxygen: $16 \times 5 \cdot 0.94$ GeV = 75 GeV, and calculate $\Delta m/m$ at molecular level (we use 0.94 GeV as the mass of a nucleon):

$$8.3 \text{ eV}/75 \text{ GeV} = 1.1 \cdot 10^{-10} \ .$$

Thus, we see that the sum of masses of molecules on the right hand side of eq. (38) is by 8.3 eV smaller than that on the left-hand side. This mass difference is exploited in cooking.

Another example is the melting of ice. It takes $0.334 \cdot 10^6$ J to melt a kilogram of ice. That means that in this case the relative increase of mass $\Delta m/m$ is (see eqs. (35) and (36)):

$$\Delta m/m = 0.334 \cdot 10^6 \cdot 6.24 \cdot 10^{18} \cdot 1.78 \cdot 10^{-36} = 3.7 \cdot 10^{-12} \ .$$

If the temperature of a flat iron is increased by 200^0 its mass increases by $\Delta m/m = 10^{-12}$. This is readily estimated using the specific heat (25 J $\cdot$ mol^{-1} K^{-1} = 450 J kg^{-1} K^{-1}):

$$\frac{\Delta m}{m} = 450(\text{J kg}^{-1}\text{K}^{-1})200 \text{ K} = 10^{-12} \ .$$

All these mass differences are too tiny to be measured directly. Let us note that the defect of mass in a hydrogen atom 13.5 eV is also too small to be observed directly because the mass of the proton is known with large uncertainty ± 80 eV.

The tiny values of Δm in atomic transitions and chemical reactions were the basis for the statement that in non-relativistic physics mass is additive, and of the law of conservation of this additive mass.

However in nuclear and particle physics the defect of mass is much larger. For instance, in the case of deutron, which is a nuclear bound state of proton and neutron, the binding energy and hence the defect of mass is 2.2 MeV, so that $\Delta m/m \simeq 10^{-3}$.

Of special pedagogical interest is the reaction of annihilation of electron and positron into two photons (two γ-quanta). Photons are massless particles, which always move with velocity c. The latter statement follows from eqs. (31), (34):

$$|\frac{\mathbf{v}}{c}| = |\frac{\mathbf{p}c}{E}| = 1 \ , \ \text{ if } |\mathbf{p}c| = E \ . \tag{40}$$

Depending on their energy, photons are referred to as quanta of radio waves, visible and invisible light, X-rays, γ-quanta.

10

The reaction of annihilation is

$$e^+ + e^- \to \gamma + \gamma \ . \qquad (41)$$

Let us consider the case when electron and positron annihilate at rest. Then their total energy is $E = E_0 = 2m_e c^2$, while the total momentum $\mathbf{p} = 0$. Due to conservation of energy and momentum the two photons will fly with opposite momenta, so that each of them will have energy equal to $m_e c^2$. The rest frame of $e^+ + e^-$ will be obviously the rest frame of two photons. Thus, the rest energy of the system of two photons will be $2m_e c^2$ and hence the mass of this system will be $2m_e$, in spite of the fact that each of the photons is massless. We see that mass in relativity theory is conserved, but not additive.

In general case the system of two free particles with energies and momenta E_1, $\mathbf{p}_1$ and E_2, $\mathbf{p}_2$ has total energy and momentum, respectively,

$$E = E_1 + E_2 \ , \quad \mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2 \ . \qquad (42)$$

These equations follow from additivity of energy and momentum. The mass of the system is defined as before by eq. (34). Hence

$$m^2 = (E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2 = m_1^2 + m_2^2 + 2E_1E_2(1 - \mathbf{v}_1\mathbf{v}_2) \ . \qquad (43)$$

It follows from eq. (43) that the mass of a system of two particles depends not only on masses and energies of these particles, but also on the angle between their velocities. Thus for two photons m is maximal when this angle is π and vanishes when it is zero.

The mass of system has lost its Newtonian meaning of an amount of substance, its main characteristic being now rest energy (in units, where $c = 1$). Newtonian equations can be obtained from relativistic ones in the limiting case of low velocities ($v/c \ll 1$). In that case γ given by eq. (32) becomes

$$\gamma = (1 - v^2/c^2)^{-1/2} \simeq 1 + \frac{v^2}{2c^2} \ , \qquad (44)$$

so that for one particle we get:

$$E = mc^2 + \frac{mv^2}{2} = E_0 + E_{kin} \ , \quad \mathbf{p} \simeq m\mathbf{v} \ . \qquad (45)$$

For a system of two particles in the limit of vanishing v we get from eq. (43)

$$m^2 \simeq (m_1 + m_2)^2. \qquad (46)$$

Thus the approximate additivity of mass is restored.

We started the description of Newtonian mechanics by consideration of static gravitational and electric interactions, in particular, their potentials (8) and (12). For particles at rest these potentials do not depend on time. The situation is drastically changed when the velocity of particles is not negligible. Let us start with electrodynamics. First, in addition to the scalar potential φ we have now vector potential $\mathbf{A}$, so that $\varphi, \mathbf{A}$ form a four vector. Second, because of finite velocity c of propagation of electromagnetic perturbations, both φ and $\mathbf{A}$ are retarded:

$$\varphi(t_2, \mathbf{r}_2) = \frac{e}{r - \frac{\mathbf{vr}}{c}} \ , \quad \mathbf{A}(t_2, \mathbf{r}_2) = \frac{e\mathbf{v}}{c(r - \frac{\mathbf{vr}}{c})} \ , \qquad (47)$$

where

$$\mathbf{r} = \mathbf{r}_2(t_1) - \mathbf{r}_1(t_1) \ , \qquad (48)$$

while

$$\mathbf{v} = \mathbf{v}(t_1, \mathbf{r}_1) \ . \tag{49}$$

Thus defined φ and $\mathbf{A}$ allow one to calculate the strengths of electric and magnetic fields:

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \text{grad}\ \varphi \ , \ \ \mathbf{H} = \text{rot}\ \mathbf{A} \ , \tag{50}$$

where the differentiation is performed with respect to t_2, $\mathbf{r}_2$.

In a four-dimensional form the six components of antisymmetric tensor of the strength of electromagnetic field F_{ik} are expressed in terms of derivatives of four-dimensional potential:

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \ . \tag{51}$$

The lower indices are referred to as covariant, while the upper ones as contravariant.

$$A_i = (\varphi, -\mathbf{A}) \ , \ \ x^i = (ct, \mathbf{r}) \ , \ \ i = 0, 1, 2, 3 \tag{52}$$

The components of $\mathbf{E}$ and $\mathbf{H}$ are expressed in terms of components of F_{ik}:

$$E_x = F_{11} \ , \ \ E_y = F_{02} \ , \ \ E_z = F_{03} \ , \tag{53}$$

$$H_x = F_{23} \ , \ \ H_y = F_{31} \ , \ \ H_z = F_{12} \ . \tag{54}$$

The Lorentz invariant products of four-vectors are constructed by using the so-called metric tensor η_{ik}, which in an inertial reference frame is given by a diagonal 4×4 matrix:

$$\eta^{ik} = \eta_{ik} = \text{diag}(1, -1, -1, -1) \tag{55}$$

with vanishing non-diagonal elements. Multiplication of a covariant vector by η^{ik} gives a contravariant vector, e.g.:

$$p^i = \eta^{ik} p_k \ , \tag{56}$$

where summation over index k is presumed.

Up to now we considered point-like particles. If the charge is smeared over a finite volume with density ρ, the total charge of a particle is given by integral:

$$e = \int \rho(\mathbf{r}) d\mathbf{r} \ . \tag{57}$$

Similarly:

$$e\mathbf{v} = \int \mathbf{v}(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r} \ . \tag{58}$$

The four-vector $j^i = (c\rho, \mathbf{v}\rho)$ describes the density of the four-current. (In the case of a point-like particle $\rho = e\delta(\mathbf{r} - \mathbf{r}_1)$.)

The famous Maxwell's equations of classical electrodynamics have the form:

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c} j^i \ , \ \ \frac{\partial \tilde{F}^{ik}}{\partial x^k} = 0 \ , \tag{59}$$

Here

$$\tilde{F}^{ik} = \varepsilon^{iklm} F_{lm} \ , \tag{60}$$

where ε^{iklm} is Lorentz-invariant antisymmetric tensor. We see that current j^i is the source of electromagnetic field.

12

4 Mass in General Relativity

Let us now consider relativistic gravity. The role of gravitational potentials is played by 10 components of symmetric metric tensor $g_{ik}(x^i)$, four of them being diagonal, while six off-diagonal. What is very important is that in the case of gravity the ten components of g^{ik} are functions of space-time points x^i: they change from one point to another. The source of gravitational field, the analogue of vector j^i, is the density of energy-momentum tensor T^{ik}. T^{ik} is symmetric and conserved

$$\frac{\partial T^{ik}}{\partial x^k} = 0 \ . \tag{61}$$

The total 4-momentum of a system

$$p^i = \frac{1}{c} \int T^{i0}(\mathbf{r}) d\mathbf{r} \tag{62}$$

Hence T^{00} is the density of energy, while T^{10}/c, T^{20}/c and T^{30}/c represent the density of momentum. For a point-like particle with mass m the density of mass μ is given by

$$\mu = m\delta(\mathbf{r} - \mathbf{r}_1) \ . \tag{63}$$

$$T^{ik} = \mu c \frac{dx^i}{ds} \cdot \frac{dx^k}{dt} = \mu c u^i u^k \frac{ds}{dt} \ , \tag{64}$$

where u^i is contravariant velocity, while ds is an invariant interval:

$$u^i = dx^i/ds \ , \ \ ds^2 = g_{ik} dx^i dx^k \ , \ \ ds = c d\tau = \sqrt{g_{00}} dx^0 \ . \tag{65}$$

Hence

$$\tau = \frac{1}{c} \int \sqrt{g_{00}} dx^0 \ , \tag{66}$$

where τ is the proper time for a given point in space.

The connection between u^i and ordinary 3-velocity $\mathbf{v}$ is

$$u^i = (\gamma, \frac{\mathbf{v}}{c}\gamma) \ . \tag{67}$$

Thus

$$u_i u^i = 1 \ . \tag{68}$$

The most important conclusion is that the source of gravitational field is proportional to the mass of a particle.

The equation for gravitational potential g_{ik}, derived by Einstein in 1915, is more complex than the Maxwell equation for A^i:

$$R_{ik} - \frac{1}{2} g_{ik} R = 8\pi G T_{ik} \ . \tag{69}$$

Here R_{ik} is the so-called Ricci tensor, while R is scalar curvature:

$$R = g^{ik} R_{ik} \ . \tag{70}$$

The role of electromagnetic field strength F_{ik} is played in gravity by the affine connection:

$$\Gamma^i_{kl} = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right) \ , \tag{71}$$

while the role of derivative $\partial_k F^{ik}$ is played by the left-hand side of eq. (69), where the Ricci tensor is given by:

$$R_{ik} = g_{in} g^{rs} \left(\frac{\partial \Gamma^n_{ks}}{\partial x^r} - \frac{\partial \Gamma^n_{kr}}{\partial x^s} + \Gamma^n_{pr}\Gamma^p_{ks} - \Gamma^n_{ps}\Gamma^p_{kr} \right) . \tag{72}$$

The drastic difference of gravidynamics from electrodynamics is the nonlinearity of Einstein equation (69): it contains products of affine connections. This nonlinearity manifests itself at low values of v/c as a tiny effect in the precession of perihelia of planets (Mercury). However it is very important for strong gravitational fields in such phenomena as black holes.

From principal point of view of highest priority is the dual role of the tensor g_{ik}, which is both dynamical and geometrical. Dynamically g_{ik} represents the potential of gravitational field. On the other hand g_{ik} and its derivatives determine the geometry of space-time. Einstein gave to his theory of relativistic gravity the name of General Theory of Relativity (GTR).

It is clear from the above equations of GTR that in non-relativistic limit $v/c \ll 1$ the gravitational interaction is determined by only one mass m, while notations m_g and m_i are redundant and misleading. Even more useless are concepts of active and passive gravitational mass often considered by some authors.

5 The pedagogical virus of "relativistic mass"

The "famous formula $E = mc^2$" and the concept of "relativistic mass" increasing with velocity, which follows from it, are historical artifacts, contradicting the basic symmetry of Einstein's Special Relativity, the symmetry of 4-dimensional space-time. The relation discovered by Einstein is not $E = mc^2$, but $E_0 = mc^2$, where E_0 is the energy of a free body at rest introduced by Einstein in 1905. The source of the longevity of the "famous formula" is the irresponsible attitude of relativity theory experts to the task of explaining it to the non-experts.

The notion of "relativistic mass" presents a kind of pedagogical virus which very effectively infects new generations of students and professors and shows no signs of decline. Moreover in the Year of Physics it threatens to produce a real pandemia.

I published my first articles [11, 12] against the "Einstein famous equation $E = mc^2$" in 1989. The subject seemed important to me because it concerned the proper teaching of special relativity at high schools, colleges and universities and explaining its genuine meaning to a wide audience of non-physicists, the so-called "pedestrians" in popular science magazines and books.

The task looked also not absolutely formidable because a consistent presentation of relativity existed for a long time in the world-wide accepted textbook by Landau and Lifshitz [13], which was the basis of my own understanding, and in some other textbooks.

The existence of countless texts, in which the essence of relativity was mutilated (or semi-mutilated) had two sides. On one hand, it looked discouraging, especially because among the authors of these texts there were many famous physicists, the fathers and greatest authorities of modern physics. On the other hand, it was a challenge. So I tried to explain clearly to the readers the beauty of four-dimensional space-time approach and the ugliness and inconsistency of "relativistic mass", an illegitimate child of relativistic and non-relativistic equations.

My optimism had increased when in 1992 Taylor and Wheeler in the second edition of the influential and popular "Spacetime Physics" [14] included a "Dialog: Use and Abuse of the Concept of Mass", in which they supported my articles [11, 12]. A copy of this book is in my bookcase with a postcard sent to me in October 1991 by John Archibald Wheeler. The postcard has a photo of the famous Albert Einstein

Memorial in front of the building of the National Academy of Sciences, Washington, DC. The bronze sculpture of Einstein includes a copybook with $E = mc^2$ on an open page.

Since that time I received hundreds of letters from physicists (both professors and students) stating their adherence to the four-dimensional formulation of relativity and to the Lorentz invariant concept of mass. In a few cases I helped the authors to correct erroneous explanations of the concept of mass in preparing new editions of their textbooks. However the number of proponents of relativistic mass seemed not to decrease.

A leading role in promoting the relativistic mass have played the books by Max Jammer [15, 16]. Especially aggressive the proponents of relativistic mass became in connection with the World Year of Physics, which marks the 100th anniversary of fundamental articles published by Einstein in 1905.

The campaign started by the September 2004 issue of "Scientific American", full with "relativistic mass" equal to $m_0/\sqrt{1 - v^2/c^2}$, where m_0 is rest mass, and "the most famous equation $E = mc^2$". A letter to the editors, defending the four-dimentional approach and invariant mass had been rejected by the editor G. Collins who in April 2005 wrote: "Most important, we believe that tackling the issue head-on in the manner you and your coauthors want in the letters column of Sci. Am. would be very confusing to our general audience and it would make the subject seen all the more mysterious and impenetrable to them". Thus to avoid "head-on" collision of correct and false arguments the editors of Sci. Am. preferred to hide from the readers the correct viewpoint.

P. Rodgers – the Editor of European "Physics World" wrote in January 2005 in editorial [17]: "... $E = mc^2$ led to the remarkable conclusion that mass and energy are one and the same". Unlike G. Collins, P. Rodgers published a letter criticizing this statement and partly agreed with the criticism [18].

In September 2005 the bandwagon of relativistic mass was joined by "The New York Times", which published an article by B. Green [19].

The journalists were supported by renowned scientists, such as R. Penrose, who in a new thousand pages thick book had written [20]:

"In a clear sense mass and energy become completely equivalent to one another according to Einstein's most famous equation $E = mc^2$."

How many students, teachers and journalists will be infected by this sentence? How many readers had been infected by the famous book by S. Hawking [21], the second edition of which appeared in 2005? On the very first page of it Hawking wrote:

"Someone told me that each equation I included in the book would halve the sales. I therefore resolved not to have any equations at all. In the end, however, I did put in one equation, Einstein's famous equation $E = mc^2$. I hope that this will not scare off half of my potential readers."

I am sure that the usage of $E = mc^2$ had doubled the sales of his book, the buyers being attracted by the famous brand. But is it possible to estimate the damage done to their understanding of relativity theory and to the general level of the literature on relativity incurred by this case of spreading the virus?

Two recent preprints by Gary Oas [22,23] written in the framework of Educational Program for Gifted Youth at Stanford University were devoted to the use of relativistic mass. The author "urged, once again, that the use of the concept at all levels to be abandoned" [22]. The manuscript has been submitted for publication to the "Americal Journal of Physics", but was rejected as being "too lengthy" (it contains 12 pages!). A lengthy bibliography (on 30 pages) of books referring to special and/or general relativity is provided in ref. [23] to give a background for discussions of the historical use of the concept of relativistic mass. It is easy to forecast the aggressive reaction of the virus infected community to this attempt to cure it.

Acknowledgments

This work is supported by the grant of Russian ministry of education and science No. 2328.2003.2.

References

[1] I. Newton, *Philosophiae Naturalis Principia Mathematica [Mathematical Principles of Natural Philosophy]*, translated in English by A. Motte, revised and annotated by F. Cajori (University of California Press, 1966).
[2] G. Galilei, *Dialogue Concerning the Chief World Systems – Ptolematic and Copernican*, Engl. Transl. S. Drake, foreword A. Einstein, 1967 (Berkeley, CA: University of California Press) 2nd edn.
[3] Nikolas of Cusa, *On learned ignorance*: a translation and an appraisal of *De docta ignorantia* by Jasper Hopkins. 2nd ed., rev. Minneapolis: A.J. Bannning Press, 1990, [c1985].
[4] H. Lorentz, *Electromagnetic phenomena in a system moving with any velocity smaller than that of light.* Proc. Acad. Sci., Amsterdam **6**, 809 (1904).
[5] H. Poincare, *Sur la Dynamique de l'electron*, Comptes Rendues **140**, 1504 (1905).
[6] H. Poincare, *Sur la Dynamique de l'electron*, Rendiconti del Circolo Matematico di Palermo **XXI**, 129 (1906).
[7] A. Einstein, *Zur Elektrodynamik bewegter Körper*, Ann. d. Phys. **17**, 891 (1905).
[8] A. Einstein, *Ist die Trägheit eines Körpers von seinem Energiegehalt abhängig?* Ann. d. Phys., **18**, 639 (1905).
[9] M. Planck, *Das Prinzip der Relativität und die Grundgleichungen der Mechanik*, Verh. d. Deutsch. Phys. Ges. **4**, 136 (1906).
[10] H. Minkowski, *Raum und Zeit*, Phys. Z. **10**, 104 (1909).
[11] L.B. Okun, Soviet Physics-Uspekhi **32** (July 1989) 629 - 638.
[12] L.B. Okun, *Physics Today*, June 1989, pages 31-36; http://www.physicstoday.org/vol-42/iss-6/vol42no6p31_36.pdf; L.B. Okun, *Physics Today*, May 1990, page 147.
[13] L.D. Landau, E.M. Lifshitz, *The Classical Theory of Fields*, translated from the second Russian ed., Pergamon, New York (1951).
[14] E.F. Taylor, J.A. Wheeler, *Spacetime Physics*, 2nd ed. Freeman and Company. New York, 1992, pages 246 - 252.
[15] M. Jammer, *Concepts of Mass in Classical and Modern Physics*, Harvard, 1961.
[16] M. Jammer, *Concepts of Mass in Contemporary Physics and Philosophy*, Princeton, 2000.
[17] P. Rodgers, *Physics World* **18** No.1, 13 (2005).
[18] P. Rodgers, *Physics World* **18** No.10, 20 (2005).
[19] B. Green, *That Famous Equation and You*, The New York Times, Sept. 30, 2005.
[20] Roger Penrose, *The road to reality. A complete guide to the laws of the universe*, A. Knopf, New York, 2004, p. 434.
[21] S. Hawking, *A Brief History of Time*, New York, 1988.
[22] G. Oas, *On the abuse and use of relativistic mass*, arXiv: physics/0504110.
[23] G. Oas, *On the Use of Relativistic Mass in Various Published Works*, arXiv: physics/0504111.

Vol. 37 (2006) *ACTA PHYSICA POLONICA B* No 3

PHOTON: HISTORY, MASS, CHARGE*

L.B. Okun

A.I. Alikhanov Institute for Theoretical and Experimental Physics
Bolshaya Cheremushkinskaya 25, 117218 Moscow, Russia

(Received December 6, 2005)

The talk consists of three parts. "History" briefly describes the emergence and evolution of the concept of photon during the first two decades of the 20th century. "Mass" gives a short review of the literature on the upper limit of the photon's mass. "Charge" is a critical discussion of the existing interpretation of searches for photon charge. Schemes, in which all photons are charged, are grossly inconsistent. A model with three kinds of photons (positive, negative and neutral) seems at first sight to be more consistent, but turns out to have its own serious problems.

PACS numbers: 01.65.+g, 14.70.Ph

1. History

The idea that light consists of rapidly moving particles can be traced from the writings of ancient authors to Descartes and Newton. The wave theory of light was put forward by Huyghens and was later decisively proved to be correct through discovery of interference and diffraction by Young and Fresnel. Maxwell's theory of light as electromagnetic waves was one of the greatest achievements of the 19th century.

The history of the photon in the 20th century started in 1901 with the formula by Planck for radiation of a black body and introduction of what was called later the quantum of action h [1]. In 1902 Lenard discovered that energy of electrons in photo-effect does not depend on the intensity of light, but depends on the wavelength of the latter [2].

In his fundamental article "*On an Heuristic Point of View Concerning the Production and Transformation of Light*" published in 1905 Einstein pointed out that the discovery of Lenard meant that energy of light is distributed in space not uniformly, but in a form of localized light quanta [3].

* Presented at the PHOTON2005 Conference, 31 August–4 September 2005, Warsaw, Poland.

He has shown that all experiments related to the black body radiation, photoluminescence and production of cathode rays by ultraviolate light can be explained by the quanta of light.

The proof that Einstein's light quanta behave as particles, carrying not only energy, but also momentum, was given in 1923 in the experiments by Compton on scattering of X-rays on electrons [4].

The term "photon" for particles of light was coined by Lewis in 1926 in an article "*The Conservation of Photons*" [5]. His notion of a photon was different from the notion we use today. He considered photons to be "atoms" of light, which analogously to the ordinary atoms are conserved.

The term "photon" was quickly accepted by physics community. The fifth Solvay Council of Physics, which took place on October 24–29, 1927, had the name "Electrons and Photons" [6]. The term "photon" in its present meaning was first used in the talk by Compton at this meeting (see Ref. [6], p. 55).

In his talk Compton used the term "photon" as if it existed since 1905; thus on page 62 of Ref. [6] one can read: "It is known that the hypothesis of photons was introduced by Einstein in order to explain the photo-electric effect". On the other hand, on page 57 one can read:

> "When speaking of this unit of radiation, I would use the name "photon" suggested recently by G.N. Lewis (*Nature*, 18 December, 1926). ... it has the advantage of being brief without implying any relation with mechanics of quanta, more general, or the quantum theorie of atomic structure".

The Proceedings [6] open with an obituary of H.A. Lorentz who passed away in February 1928, a few months after the Fifth Solvay meeting, in which Lorentz actively participated.

The speakers at the meeting were:

W.-L. Bragg, *The Intensity of Reflected X-Rays*, pp. 1–44;

A.H. Compton, *Discordances Between the Experiment and the Electromagnetic Theory of Radiation*, pp. 55–86;

L. de Broglie, *The New Dynamics of Quanta*, pp. 105–133;

M. Born and W. Heisenberg, *The Mechanics of Quanta*, pp. 143–182;

E. Schrödinger, *The Mechanics of Waves*, pp. 185–207;

N. Bohr, *The Postulate of Quanta and the Development of Atomistics*, pp. 215–248.

Each of the talks was followed by a detailed discussion. Participated Bohr, Born, Brillouin, de Broglie, Compton, Dirac, De Donder, Ehrenfest, Fowler, Heisenberg, Kramers, Langmuir, Langevin, Lorentz, Pauli, Richardson, Schrödinger.

Einstein took part only in the "*General Discussion of the New Ideas*", expressed during the meeting. The discussion (pp. 248–289) was presented in three sections: 1. Causality, Determinism, Probability; 2. Photons; 3. Photons and Electrons.

Einstein spoke in the first section (pp. 253–256) and asked a question during the second section (p. 266). He considered a screen with a small hole in it and a spherical layer of photo-emulsion of large radius behind it. Electrons fall on the screen as De Broglie–Schrödinger plane waves normal to it and reach the emulsion as spherical waves. Einstein discussed the two possible interpretations of this thought experiment: purely statistical and purely deterministic. The term "photon" was not used in his remarks.

The term "photon" was again used by Compton on December 12, 1927, this time without any reference to Lewis, in Compton's Nobel lecture "*X-Rays as a Branch of Optics*" [7]. On page 186 one can read:

> "An X-ray photon is deflected through an angle ϕ by an electron, which in turn recoils at an angle θ, taking a part of the energy of photon".

Further on page 187:

> "... recoil electrons are in accord with the predictions of the photon theory".

The year 1927 marked the end of the history of emergence of the concept of photon. A few years later Dirac opened a new chapter in Physics by establishing Quantum Electrodynamics.

As for Einstein, he wrote in 1951:

> "All these fifty years of pondering have not brought me any closer to answering the question, What are light quanta?" [8].

2. Mass

The problem of the upper limit on the mass of the photon was raised at ITEP by Isaak Yakovlevich Pomeranchuk (20.05.1913–14.12.1966) in autumn of 1966, a few months before he lost his fight against cancer. He put this question to his former students: Igor Yuryevich Kobzarev (15.10.1932–20.01.1991) and myself. First we wrote a draft of a short research note, but then after a thorough search we discovered that most of our considerations had been already addressed in the literature by de Broglie [9, 10] (see also [11, 12]), Schrödinger [13, 14], Bass and Schrödinger [15] and by Gintsburg [16].

In particular, de Broglie [10] noticed that photon mass would lead to a larger speed of violet light than that of the red one. He concluded that during the eclipse in a double star system the color of the appearing star would change from violet to red. He also considered the dispersion of radio-waves.

Schrödinger [13, 14] pointed out that magnetic field of the Earth would be exponentially cut off at distances of the order of the photon Compton wave length $\lambda_\gamma = 1/m_\gamma$. From the observed altitude of auroras he concluded that $\lambda_\gamma > 10^4$ km.

Gintsburg [16] corrected the limit of Schrödinger and suggested that measurements of the magnetic field of Jupiter could improve the limit to $\lambda_\gamma \sim 10^6$ km. He also was the first to consider how the mass of the photon would influence the magnetohydrodynamic waves in plasma.

These results discouraged us from publishing an original article.

From the beginning of 1968 a special issue of Uspekhi Fizicheskikh Nauk was under preparation to mark 50 years of this review journal. Kobzarev and I were invited to publish our paper on the photon as a review.

In this short review [17] we corrected the estimates by de Broglie and Schrödinger. The former estimate was invalidated due to dispersion of light in the atmospheres of stars (we found that this effect was considered by Lebedev in 1908 [18]). Ref. [18] was the paper, which closed the discussion of color variation in binary stars. The effect was discovered by Belopolskii [19], Nordmann [20] and Tikhoff [21] and interpreted by them as dispersion of light in the interstellar free space. The observed minimum in red light preceded that in violet light from the variable binary stars by a few minutes. (Note that for a massive photon the violet light should be faster, not slower!) Lebedev[1] rejected this interpretation and explained the effect by the difference of pressure in the atmospheres of two stars [22, 23].

We also found that a much better limit could be extracted from the measurements by Mandelstam [24] of radio-wave dispersion in the atmosphere of the Earth.

As for the limit by Schrödinger we conservatively extended it to 30 000 km by using the data from review by Bierman [25], though these data (from rockets and satellites) indicated the spread of the geomagnetic field to 60 000 km and even to 100 000 km.

In addition to magnetic field we have interpreted in terms of λ_γ the experiment by Plimpton and Lawton [26], testing the absence of Coulomb field in the space between two concentric spheres, and derived $\lambda_\gamma < 10$ km.

[1] Petr Nikolaevich Lebedev (1866–1912) is famous by his experimental discovery of the pressure of light.

(The deviation from the Coulomb law was parametrized in Ref. [26] by $1/r^{2+\varepsilon}$.)

We also discussed why the longitudinal photons do not manifest themselves in the black body radiation, a subject considered by Bass and Schrödinger [15].

Our review [17] appeared in May 1968.

Two months later Physical Review Letters received and in August published a paper by Goldhaber and Nieto [27] *"New Geomagnetic Limit on the Mass of the Photon"*. Their geomagnetic limit was about 90 000 km. They derived $\lambda_\gamma < 10\,\text{km}$ from reference [26] and reconsidered the geomagnetic estimates by Gintsburg [16].

Three years later Goldhaber and Nieto published an extensive review [28] with about 100 references. The review by Byrne [29] published in 1977 has about 40 references. The latest review by Tu, Luo and Gillies [30] published in 2005 has about 200 references.

It is impossible to comment on all these hundreds of papers in a short review. One has to make a selection.

Since 1992 the selected references on the photon mass are cited by the Particle Data Group (PDG) in biennial Reviews of Particle Properties [31–37]. The best cited limits (in eV) were chosen by PDG:

1992: 3×10^{-27}, Chibisov [38], galactic magnetic field.

1994: 3×10^{-27}, Chibisov [38], galactic magnetic field.

1996: 6×10^{-16}, Davis *et al.* [39], Jupiter magnetic field.

1998: 2×10^{-16}, Lakes [40], torque on toroid balance.

2000: 2×10^{-16}, Lakes [40], torque on toroid balance.

2002: 2×10^{-16}, Lakes [40], torque on toroid balance.

2004: 6×10^{-17}, Ryutov [41], magnetohydrodynamics of solar wind (MHD).

If c is the unit of velocity and $\hbar$ is the unit of action, then $1\,\text{eV} = 1.78 \times 10^{-33}\,\text{g}$, $1\ \text{eV} = (1.97 \times 10^{-10}\ \text{km})^{-1}$.

Chibisov [38] considered the conditions of equilibrium of magnetic field in the smaller Magellanic cloud by applying virial theorem. This gave $\lambda_\gamma \lesssim l$, where l is the size of the cloud ($l \simeq 3$ kpc $= 3 \times 3.08 \times 10^{16}$ km $\approx 10^{17}$ km $= 10^{22}$ cm). It is not clear how reliable is this approach.

Davis *et al.* [39] used the *"Jupiter Suggestion"* of Gintsburg [16] and the new Pioneer-10 data on the magnetic field of Jupiter.

A novel idea was put forward and realized by Lakes [40]. He exploited the fact that the term $m_\gamma^2 \boldsymbol{A}^2$ in the Lagrangian breaks the gauge invariance of Maxwell's electrodynamics. In Lorenz[2] gauge one has the Maxwell–Proca equation. As a result the vector potential $\boldsymbol{A}$ becomes observable. Lakes performed an experiment with a toroid Cavendish balance to search for the torque $m_\gamma^2 \boldsymbol{A}$ produced by the ambient vector potential $\boldsymbol{A}$.

The experiment [40] disclosed that $\boldsymbol{A} m_\gamma^2 < 2 \times 10^{-9}$ Tm/m^2. If the cosmic vector potential $\boldsymbol{A}$ is 10^{12} Tm, then $\lambda\!\!\!^{-}_\gamma = m_\gamma^{-1} \gtrsim 2 \times 10^{10}$ m. This limit has been improved by other authors (see Ref. [30]). However, the estimate of the value of cosmic potential $\boldsymbol{A}$ is not reliable enough.

Ryutov [41] developed the idea of Gintsburg [34] and first derived a self-consistent and complete set of MHD equations accounting for finite photon mass. He did not put a new limit on the photon mass, but mentioned a possible way of improving it by the analysis of the sector structure of the Solar wind. In particular he noticed that the limit 6×10^{-16} eV, considered in 1996 by PDG as the best one should be reduced by approximately an order of magnitude. This is the origin of the PDG best number in 2004.

3. Charge

There exist about a dozen of papers [43–52] questioning the neutrality of photons and setting an upper limit on their charge. In all of them the upper limit follows from the non-observation of any action of external static electric or magnetic fields on photon's charge, while the fact that these fields themselves are "built from photons" is ignored. As a result all those papers [43–52] lack a self-consistent phenomenological basis. But without such a basis any interpretation of experimental data is meaningless.

In fact the authors [43–52] implicitly assumed that all photons are either neutral as in ordinary QED, or all are charged. It is obvious that the latter assumption is impossible to reconcile with the existence of classical static electric or magnetic fields. Hence the best upper limit on the value of photon charge presented by the Particle Data Group [45] seems to be meaningless.

It is clear, that for a more consistent interpretation of searches [43–52] both types of photons are necessary: charged and neutral. In such a scheme classical electric and magnetic fields are built from the latter. Hence the scattering of all charged particles (including the charged photons) by these fields occurs due to absorption of virtual neutral photons. Charge is conserved in this processes. (The failure of theoretical attempts to violate the conservation of electric charge was analyzed in references [53, 54].)

[2] Quite often the Lorenz gauge is erroneously ascribed to Lorentz (for clarification and for earlier references see Ref. [42]).

However, a scheme with both charged and neutral photons is also not without serious problems. One of them is the catastrophic infrared emission of neutral photons by massless charged ones. The other problems are connected with the emission and absorption of charged photons by ordinary charged particles, say, electrons.

Conservation of charge calls in this case for the existence of a twin electron with charge e–e', where e' is the charge of the emitted charged photon, which is assumed to be much smaller than e. The mass of the twin must be much larger than the mass of the electron in order to avoid contradiction with data on atomic, nuclear, and high energy physics.

One might consider the three photons with charges $+e'$, $-e'$, 0 as an SU(2) Yang–Mills triplet, while the electron with charge e and its twin with charge e–e' as an SU(2) doublet. The SU(2) symmetry requires mass degeneracy of particles belonging to the same multiplet. However, even in this degenerate case it is impossible to accommodate the inequality $e'/e \ll 1$. The situation is further aggravated by the breaking of SU(2) gauge symmetry, responsible for the difference of masses of particles and their twins.

I am grateful to V.P. Vizgin, K.A. Tomilin and A.D. Sukhanov for helpful discussions on the history of the concept of photon.

I am grateful to A. Buras, V. Fadin and P. Zerwas for drawing my attention to the articles on hypothetical charge of the photon and to G. Cocconi, G. Giudice and M. Vysotsky for valuable remarks.

I am grateful to M. Krawczyk for wonderful hospitality.

The work was supported by the grant of the Russian Ministry of Education and Science No. 2328.2003.2.

REFERENCES

[1] M. Planck, *Ann. Phys.* **4**, 561 (1901).

[2] P. Lenard, *Ann. Phys.* **8**, 169 (1902).

[3] A. Einstein, *Ann. Phys.* **17**, 132 (1905).

[4] A.H. Compton, *Phys. Rev.* **22**, 409 (1923).

[5] G.N. Lewis, *Nature*, No. 2981, volume. 118 (December 18, 1926) 874.

[6] *Electrons et Photons.* "Rapports et discussions du cinquiem conseil de physique tenu a Bruxelles du 24 au 29 octobre 1927 sous les auspices de l'Institut International de Physique Solvay" Paris 1928, Eds. Gauthier–Villars et c^{ie}.

[7] A.H. Compton, in "Nobel Lectures. Physics. 1901–1995" CD-ROM, World Scientific, 1998.

[8] Abraham Pais, *"Subtle in the Lord ..." The Science and the Life of Albert Einstein*, Section 19f, p. 382, Oxford 1982.

[9] L. de Broglie, *Phil. Mag.* **47**, 446 (1924).

[10] L. de Broglie, *La mechanique ondulatoire du photon. Une nouvelle theorie de la lumiere, tome premier*, Paris 1940, pp. 39.

[11] L. de Broglie, *La mechanique ondulatoire du photon et theorie quantique de champs*, Paris 1949.

[12] L. de Broglie, *La theorie generale des particule á Spin*, Paris 1943, p. 191.

[13] E. Shrödinger, *Proc. Roy. Irish Acad.* **A49**, 43 (1943).

[14] E. Shrödinger, *Proc. Roy. Irish Acad.* **A49**, 135 (1943).

[15] L. Bass, E. Shrödinger, *Proc. Roy. Soc.* **A232**, 1 (1955).

[16] M.A. Gintsburg, *Astronom. Zhurnal* **40**, 703 (1963) (in Russian);
M.A. Gintsburg, *Sov. Astron.* **7**, 536 (1964) (English translation).

[17] I.Yu. Kobzarev, L.B. Okun, *Uspekhi Fiz. Nauk* **95**, 131 (1968); I.Yu. Kobzarev, L.B. Okun, *Sov. Phys. Usp.* **11**, 338 (1968).

[18] P.N. Lebedev, Sobranie sochineniy (Collected papers), *Izd. AN SSSR*, 1963, paper 33 (in Russian).

[19] A.A. Belopolskii, *Izvestia Imper. Akademii Nauk* **21**, 153 (1904) (in Russian).

[20] C. Normann, *Comptes Rendus* **146**, 266 (1908).

[21] G.A. Tikhoff, *Comptes Rendus* **146**, 570 (1908).

[22] P.N. Lebedev, *Izvestia Imper. Akademii Nauk* **24**, 93 (1906) (in Russian).

[23] P.N. Lebedev, *Comptes Rendus* **146**, 1254 (1908); **147**, 515 (1908).

[24] L.I. Mandelshtam, *Izv. AN SSSR*, 1947, volume 2, pp. 277; volume 3, p. 238 (in Russian).

[25] L. Bierman, Sitzber. *Bayerische Akad. Wissenschaften* **37** (1965).

[26] S.J. Plimpton, W.E. Lawton, *Phys. Rev.* **50**, 1066 (1936).

[27] A.S. Goldhaber, M.M. Nieto, *Phys. Rev. Lett.* **21**, 567 (1968).

[28] A.S. Goldhaber, M.M. Nieto, *Rev. Mod. Phys.* **43**, 277 (1971).

[29] J.C. Byrne, *Astrophysics and Space Science* **46**, 115 (1977).

[30] L-C. Tu, J. Luo, G.T. Gillies, *Rep. Prog. Phys.* **68**, 77 (2005).

[31] PDG, *Phys. Rev.* **D45**, (1992).

[32] PDG, *Phys. Rev.* **D50**, 1351 (1994).

[33] PDG, *Phys. Rev.* **D54**, 207 (1996).

[34] PDG, *Eur. Phys. J.* **3**, 223 (1998).

[35] PDG, *Eur. Phys. J.* **15**, 249 (2000).

[36] PDG, *Phys. Rev.* **D66**, 281 (2002).

[37] PDG, *Phys. Lett.* **B592**, 335 (2004).

[38] G.V. Chibisov, *Usp. Fiz. Nauk* **119**, 551 (1976) (in Russian); *Sov. Phys. Usp.* **19**, 624 (1976) (English translation).

[39] L. Davis, A.S. Goldhaber, M.M. Nieto, *Phys. Rev.* **35**, 1402 (1975).

[40] R. Lakes, *Phys. Rev. Lett.* **80**, 1826 (1998).

[41] D.D. Ryutov, *Plasma Phys. Control Fusion* **39**, A73 (1997).

[42] J.D. Jackson, L.B. Okun, *Rev. Mod. Phys.* **73**, 663 (2001).

[43] L. Grodzins, D. Engelberg, W. Bertozzi, Bull. *Am. Phys. Soc.* **6**, 63 (1961).

[44] R.W. Stover, T.I. Moran, J.W. Trischka, *Phys. Rev.* **164**, 1599 (1967).

[45] PDG, "Review of Particle Physics", *Phys. Lett.* **B592**, 31 (2004).

[46] G. Cocconi, *Phys. Lett.* **B206**, 705 (1988).

[47] G. Raffelt, *Phys. Rev.* **D50**, 7729 (1994).

[48] G. Cocconi, *Am. J. Phys.* **60**, 750 (1992).

[49] V.V. Kobyshev, S.B. Popov, *Astron. Lett.* **31**, 147 (2005).

[50] C. Sivaram, *Am. J. Phys.* **63**, 1473 (1994).

[51] C. Caprini, P.G. Ferreira , *JCAP* **0502**, 006 (2005).

[52] Y.K. Semertzidis, G.T. Danby, D.M. Lazarus, *Phys. Rev.* **D67**, 017701 (2003).

[53] L.B. Okun, Ya.B. Zeldovich, *Phys. Lett.* **B78**, 597 (1978).

[54] M.B. Voloshin, L.B. Okun, *Pis'ma ZhETF* **28**, 156 (1978) (in Russian); *JETP Lett.* **28**, 145 (1978).

This page intentionally left blank

Vol. 37 (2006) ACTA PHYSICA POLONICA B No 4

FORMULA $E = mc^2$ IN THE YEAR OF PHYSICS*

L.B. Okun

A.I. Alikhanov Institute for Theoretical and Experimental Physics
Bolshaya Cheremushkinskaya 25, 117218 Moscow, Russia

(Received November 28, 2005)

The "famous formula" $E = mc^2$ and the concept of "relativistic mass" increasing with velocity, which follows from it, are historical artifacts, contradicting the basic symmetry of Einstein's Special Relativity. The relation discovered by Einstein is not $E = mc^2$, but $E_0 = mc^2$, where E_0 is the energy of a free body at rest. The source of the longevity of the "famous formula" is irresponsible attitude of relativity theory experts to the task of explaining it to the non-experts.

PACS numbers: 01.65.+g, 03.30.+p

1. Year of Physics

2005 is the first year of physics in the history of humankind. It celebrates the papers on relativity, quanta, and atoms, published in 1905 by Einstein. The whole building of modern physics has these papers at its basis.

2. Rest energy E_0

1905: Einstein introduced [1] the notion of rest energy E_0 of a body and established the connection between its change and the change of the mass of the body

$$\Delta mc^2 = \Delta E_0 \,.$$

Later on he generalized it [2] to:

$$mc^2 = E_0 \,.$$

* Presented at the PLC2005 Workshop (Kazimierz Lecture), 5–8 September 2005, Kazimierz, Poland.

3. Energy, momentum and mass

According to Special Relativity,

$$(mc^2)^2 = E^2 - \boldsymbol{p}^2c^2 = p^2\,,$$

where E — energy, $\boldsymbol{p}$ — momentum, p — 4-momentum, m — mass of a free body.

$p = (E, \boldsymbol{p}c)$ — Lorentz 4-vector, m — Lorentz scalar, relativistically invariant.

E and $\boldsymbol{p}$ can be expressed through mass m and the velocity of a body, $\boldsymbol{v} = \boldsymbol{p}c^2/E$:

$$E = mc^2\gamma\,, \quad \boldsymbol{p} = m\boldsymbol{v}\gamma\,, \quad \gamma = \frac{1}{\sqrt{1 - \boldsymbol{v}^2/c^2}}\,.$$

For a free body at rest $\boldsymbol{p} = 0$, $\boldsymbol{v} = 0$. Hence $E_0 = mc^2$, where E_0 — rest energy.

$$E_0 = mc^2\,.$$

Mass m — relativistically invariant!

4. The ship of Galileo

Why do I stress the properties of various physical quantities under Lorentz transformations? Because symmetry is the heart of physics in general and of relativity theory in particular. The principle of relativity was formulated by Galileo [3], who insisted that there is no experiment inside a ship that can tell whether the ship is at rest or is moving uniformly and rectilinearly. At the turn of 19th and 20th century it became clear, that experiments with light are not better in this respect than with any other objects. This led to Lorentz transformations of spatial and time coordinates instead of Galilean transformations according to which time is not transformed.

5. Origin of $E = mc^2$

1900: Poincaré "proved" that mass of a pulse of the light is proportional to its energy [4].

Poincaré misused non-relativistic Newton's relation

$$\boldsymbol{p} = m\boldsymbol{v}$$

at $|\boldsymbol{v}| = c$. And from Poynting relation

$$|\boldsymbol{p}| = \frac{E}{c}$$

derived $E = mc^2$.

"The famous formula" $E = mc^2$ was wrong in 1905 and even more so in 2005! But it is still very popular.

6. Longitudinal and transverse masses

1899: Lorentz [5] introduced two masses of the electron, which depend on the angle between velocity and force by using non-relativistic connection between force and acceleration

$$\boldsymbol{F} = m_t \boldsymbol{a} \text{ for } \boldsymbol{F} \perp \boldsymbol{v},$$

$$\boldsymbol{F} = m_l \boldsymbol{a} \text{ for } \boldsymbol{F} \| \boldsymbol{v},$$

$$m_t = m\gamma,$$

$$m_l = m\gamma^3,$$

$$\gamma = \frac{1}{\sqrt{1 - \boldsymbol{v}^2/c^2}}.$$

These transverse and longitudinal masses are almost forgotten today, unlike the "relativistic mass" $m = E/c^2$.

7. $E = mc^2$ in the first half of the 20th century

The relativistic mass was strongly promoted by many prominent physicists (Lewis and Tolman [5,6], Born [7], Fock [8]). In the first half of the 20th century this could be justified by the wish to preserve the role of mass in Newtonian physics, first of all as a measure of inertia. Another aim was to preserve the additivity: the mass of a system of free particles is equal to the sum of their masses. As we know today, both these properties are lost in the case of relativistic particles.

As for Einstein, he "oscillated" between $E_0 = mc^2$ and $E = mc^2$ till 1921. In 1921 he definitely chose the former [2]. But even when signing the letter to Roosvelt on atomic bomb he used the "famous formula".

8. Evolution of concept of matter

It was Newton who introduced the notion of mass. He defined mass as quantity of matter. This definition is not valid today.

At present the concept of matter includes not only atoms, but also massless and extremely light particles: the photon and neutrinos.

The masses of relativistic particles are not additive, while their energies are.

For instance, mass of the system of two photons: $m^2 = p^2 = (p_1 + p_2)^2$, where $p = (E, \boldsymbol{p})$, $p_1 = (E_1, \boldsymbol{p}_1)$, $p_2 = (E_2, \boldsymbol{p}_2)$, and we use c as a unit of velocity. The value of this mass depends on the relative momenta of photons:

$$m = 0\,, \quad \text{if} \quad \boldsymbol{p}_1 = \boldsymbol{p}_2\,;$$

$$m = 2E\,, \quad \text{if} \quad \boldsymbol{p}_1 = -\boldsymbol{p}_2\,.$$

9. Photosynthesis

The light from the sun is absorbed by vegetation on earth to produce carbohydrates via reaction of photosynthesis:

$$\text{light} + 6\text{CO}_2 + 6\text{H}_2\text{O} = 6\text{O}_2 + \text{C}_6\text{H}_{12}\text{O}_6\,.$$

The total energy of light required to produce one molecule of $C_6H_{12}O_6$ is about 4.9 eV. This does not mean that the photons are massive. They are massless, but the kinetic energy of photons is transformed into the rest energy of carbohydrates.

10. $E = mc^2$ in the Year of Physics

The Year of Physics is marked by revival of "famous equation" $E = mc^2$ and of "relativistic mass": $m_r = m_0/\sqrt{1 - v^2/c^2}$, where m_0 is "rest mass". The champion in this campaign is "Scientific American", the September 2004 issue of which is full with these notions.

Many other magazines, journals and books could be mentioned, *e.g.* "Physics World", January 2005, September 2005, October 2005. "New York Times" has recently joined the bandwagon [9]

In a brand new thousand pages thick book Penrose writes [10]:

"In a clear sense mass and energy become completely equivalent to one another, according to Einstein's most famous equation $E = mc^2$".

This book is addressed both to physicists and to "pedestrians". It is interesting to compare it with a classical monograph [11] coauthored by Penrose and Rindler and addressed to theoretical physicists, where photons and neutrinos were referred to as massless particles and which had no trace of $E = mc^2$.

11. Who is guilty?

Of course, light-minded journalists.

But first of all, renowned professors of physics, who promote $E = mc^2$ and relativistic mass as authors, lecturers, and members of editorial boards.

They try to conform the prevailing opinions of ignorant readers, instead of educating them.

12. Niels Bohr on truth and clarity

Niels Bohr once said that truth and clarity are complementary. A true statement cannot be clear, and a clear one cannot be true. This maxim is valid for the deepest truths at the front line of science, but it should be applied with some reservation in the fields, such as Special Relativity, behind the front line, where everything is firmly established.

Still many authors consider that to be clear is "politically incorrect".

It seems that this belief strengthens the longstanding confusion which surrounds the relation between energy and mass.

13. What to do?

To reach a consensus in the community of experts in Relativity Theory on the concept of unique relativistically invariant mass, m.

Experts should discard from their writings the terms "rest mass" and "relativistic mass" and the famous but wrong formula

$$E = mc^2 .$$

The rest energy should be promoted:

$$E_0 = mc^2 .$$

14. Those who are indifferent

The ongoing struggle for and against $E = mc^2$ is considered by many of physicists as a kind of lilliputian war described by J. Swift in "Gulliver's Travels" [12]. They do not consider seriously both big-endians and small-endians. Their motto is: "All true believers break their eggs at the convenient end". This attitude with respect to $E = mc^2$ is sharply criticized by Gary Oas [13]. A quasi indifferent stand of Max Jammer [14] is in fact a slightly disguised propaganda of $E = mc^2$.

REFERENCES

[1] A. Einstein, *Ann. Phys.* **27**, 132 (1905).

[2] A. Einstein, *The Meaning of Relativity: Four Lectures Delivered at Princeton University*, May 1921, fifth ed., E.P. Adams, translator, U.P., Princeton, N.J. 1970.

[3] G. Galilei, *Dialogue Concerning the Chief World Systems — Ptolemaic and Copernican*, Engl. transl. S. Drake, foreword A. Einstein, 1967, Berkeley, CA: University of California Press, 2nd edn.

[4] H. Poincaré, *Arch. Neerland* **5**, 252 (1900).

[5] G. Lewis, R. Tolman, *Philos. Mag.* **18**, 510 (1909).

[6] R. Tolman, *Philos. Mag.* **23**, 375 (1912).

[7] M. Born, *Einstein's Theory of Relativity*, New York, Dover 1962, Chapter VI (First German edition — 1920).

[8] V.A. Fock, *Teoriya Prostranstva, Vremeni i Tyagoteniya* (*Theory of Space, Time and Gravitation*), Moscow: GITTL, 1955, pp. 104, 105, 144, 145 [Translated into English, New York: Pergamon Press, 1959].

[9] B. Green, *That Famous Equation and You*, "The New York Times", September 30, 2005.

[10] R. Penrose, *The Road to Reality. A Complete Guide to the Laws of the Universe*, A. Knopf, New York 2005, p. 434.

[11] R. Penrose, W. Rindler, *Spinors and Space-Time*, v. 1, Cambridge, 1984.

[12] J. Swift, *Gulliver's Travels*, `http://www.readprint.com/chapter-8822/Jonathan-Swift`, Chapter 4.

[13] G. Oas, *On the Abuse and Use of Relativistic Mass*, `physics/0504110` v1 (2005).

[14] Max Jamer, *Concepts of Mass in Contemporary Physics and Philosophy*, Princeton, 2000.

Physics – Uspekhi **50** (4) 380 – 389 (2007)

ORAL ISSUE OF THE JOURNAL "USPEKHI FIZICHESKIKH NAUK"

PACS numbers: **01.65.+g**, 11.30.Er, **12.60.–i**

Mirror particles and mirror matter: 50 years of speculation and searching

L B Okun

DOI: 10.1070/PU2007v050n04ABEH006227

Contents

Abstract. This review describes the history of the discovery of the violation of the spatial parity P, the charge conjugation parity C, and the combined parity CP. The hypothesis of the existence of mirror particles was intended by its authors to restore the symmetry between left and right. The review presents the emergence and evolution of the concepts of 'mirror particles' and 'mirror matter' and can serve as a concise travel guide to 'mirror-land.' An important part of the review is the list of about 200 references with their titles.

1. Introduction

The terms 'mirror particles,' 'mirror matter,' and 'mirror world' currently refer to the hypothetical hidden sector of particles and interactions that compensate the mirror asymmetry of the weak interactions of ordinary particles. Mirror particles are regarded as a possible component of invisible dark matter. The history of mirror particles is a history of the intertwining of parity violation and parity degeneracy, rigorous and broken mirror symmetry, dark matter in the universe, atomic, nuclear and high-energy physics, cosmology, and astrophysics.

2. 1950s. Violation of P and C. Conservation of PC

In the middle of the 1950s, the so-called $\theta\tau$ puzzle became the most challenging problem of elementary particle physics. At that time, the decays $K^+ \to 2\pi$ and $K^+ \to 3\pi$ were assigned to two different mesons, θ^+ and τ^+, having opposite P-parities. But the masses and lifetimes of θ^+ and τ^+ were suspiciously close. Therefore, Lee and Yang put forward the idea of parity degeneracy [1].

However, at the Rochester conference in April of 1956, Feynman, referring to Block, asked the crucial question: could it be that parity is not conserved?

Here are a few excerpts from the proceedings [2]:

"J.R. Oppenheimer presiding:

There are the five objects K_{π_3}, K_{π_2}, K_{μ_2}, K_{μ_3}, K_{e_3}. They have equal, or nearly equal, masses and identical, or apparently identical, lifetimes. One tries to discover whether in fact one is dealing with five, four, three, two, or one particle...."

"Yang's introductory talk followed:

... the situation is that Dalitz's argument strongly suggests that it is not likely that $K^+_{\pi_3} (\equiv \tau^+)$ and $K^+_{\pi_2} (\equiv \theta^+)$ are the same particles".

"Dalitz discussed the $\tau\theta$ problem ... 600 events ... when plotted on the 'Dalitz diagram,' give a remarkably uniform distribution.... This would point to a τ-meson of spin-parity 0^-...".

"...Feynman brought up Block's question:

Could it be that the θ and τ are different parity states of the same particle which has no definite parity, i.e. that parity is not conserved...?"

"Yang stated that he and Lee looked into this matter without arriving at any definite conclusions."

Feynman presumably meant a special mechanism of parity violation through the mixing of degenerate scalar and pseudoscalar mesons.

It is interesting that neither Dalitz nor Michel, who also participated in the discussion, mentioned the possibility of parity violation.

A few months later, Lee and Yang suggested that parity is not conserved in weak decays and proposed experiments to search for pseudoscalar correlations of spin and momentum **sp** [3]. (Their famous paper was received by *Physical Review* on June 22, circulated as a preprint, and appeared in the

L B Okun Russian Federation State Scientific Center
'Alikhanov Institute of Theoretical and Experimental Physics',
ul. B. Cheremushkinskaya 25, 117218 Moscow, Russian Federation
Tel. (7-495) 123 31 92, (7-495) 125 96 60
E-mail: okun@itep.ru

Received 2 December 2006
Uspekhi Fizicheskikh Nauk **177** (4) 397 – 406 (2007)
Translated by L B Okun; edited by A M Semikhatov

journal on October 1, 1956.) At the end of this paper, in order to save the left–right symmetry in a more general sense, the existence of hypothetical right-handed protons, p_R, was considered, although the term 'mirror particles' was not used and p_R and p_L were assumed to interact "with the same electromagnetic field and perhaps the same pion field."

Much later I learned that already in 1952 Michel [4] considered parity-violating interactions and pseudoscalar correlations between momenta of several particles in multiparticle processes. Wick, Wightman, and Wigner considered pseudoscalar amplitudes [5]. Purcell and Ramsey suggested testing parity conservation experimentally by measuring the electric dipole moment of the neutron [6]. However, they did not realize (as Landau did subsequently) that the electric dipole moment violates the time-reversal invariance as well. Berestetsky and Pomeranchuk published a note [7] on the beta-decay of the neutron, in which they mentioned a remark by Landau that actually ten four-fermion couplings exist "if pseudo-spinors are used in addition to spinors."

As is well known, the experiments proposed by Lee and Yang were performed half a year later and found large left–right asymmetries in the β-decay of ^{60}Co [8] and in $\pi \to \mu \to e$ decays [9, 10].

Before the results of these experiments were published, Ioffe and Rudik had submitted a short paper to *ZhETF* in which they argued that the existence of a short-lived C-even K_1^0-meson and long-lived C-odd K_2^0-meson proved that C-parity was conserved and hence violation of P-parity would mean (due to the CPT theorem) violation of T-parity (time reversal invariance). This led them to the conclusion that P-odd asymmetries are impossible because they are T-even.

I vividly recall how ITEP theorists discussed these arguments with Landau after one of the traditional ITEP seminars in November 1956. (At that time, the name ITEP did not exist; the institute was called TTL, Thermo-Technical Laboratory.) The discussion took place in the room no. 9, where at that time young theorists worked and where my desk was.

At that time, Landau considered the P-parity violation impossible because space is mirror symmetric. This is analogous to the conservation of momentum and angular momentum, because space is homogeneous and isotropic. Of course, the analogy is not complete, because shifts and rotations are continuous, whereas reflections are discrete.

Half a year earlier, the Lebedev Institute hosted the first Moscow conference on elementary particles in which American physicists participated [11, 12]. I recall that Landau laughed sarcastically at Gell-Mann (the youngest of the Americans at the conference, but already very famous), when the latter, during his seminar at the Institute of Physical Problems, mentioned that parity violation could be one of the solutions to the $\theta\tau$ problem.[1]

[1] Gell-Mann repeated his talk twice: at the Institute for Physical Problems and at the Lebedev Institute, in Tamm's office. I was carefully taking notes at both of them. He stopped for a moment and asked me with a smile: "What happens if you find at home that the two records contradict each other?" In the 1980s, Telegdi published very interesting reports on the history of parity violation [13, 14]. In Ref. [14], he wrote: "Murray Gell-Mann emphasized to me ... that I.S. Shapiro most strenuously objected to the parity violation idea when M.G.M. presented the latter in 1956 in the Landau seminar as one of the possible solutions to the $\tau-\theta$ puzzle." As I have already mentioned, I remember the objections by Landau, but I do not recall that they were also raised by Shapiro at the same seminar. (See author's note to the English proofs, p. 384.)

At about the same time, Landau reacted similarly to an unpublished note by Shapiro in which a Wu-type experiment was suggested. I learned about it three years later, when Shapiro moved from Moscow State University to ITEP and showed me his unpublished note. (Later, Shapiro gave this note to the director of ITEP Alikhanov, and it was lost. There was no copying machine at ITEP.) I remember that there was an incorrect statement in this note: the value of energy is different in left- and right-handed coordinates if P is violated.[2]

But let me return to the discussion in room 9. During the discussion, I pointed out that short- and long-lived kaons might exist not due to the C-invariance, as had been proposed by Gell-Mann and Pais a year before [16], but due to the (albeit approximate) T-invariance, and hence CP-invariance. In that case, **sp** asymmetries and the decay $K_2^0 \to 3\pi^0$ would be allowed.

As a consequence of this discussion, Ioffe and Rudik decided to insert my comments into their paper and urged me to become a coauthor of an essential revision of their paper.[3] At first I refused, but conceded after Ioffe literally went down on one knee in front of me. (This took place in the same room no. 9.) Our article [18] was noticed by Yang and Lee, who, with Oehme [19], independently but later came to the same conclusions (see references to [18] in their Nobel lectures [20, 21]).

Another consequence of the discussion was that Landau abruptly changed his attitude to parity nonconservation and put forward the idea of strict CP-conservation [22] (see also [23]). At the end of this paper, he wrote: "I would like to express my deep appreciation to L Okun, B Ioffe and A Rudik for discussions from which the idea of this paper emerged." According to his idea, a mirror-reflected process cannot exist in Nature and becomes physical only after changing particles into corresponding antiparticles.[4]

An excellent example of CP-conjugated particles was presented by Landau [25] in his theory of massless longitudinally polarized neutrinos: the spin of ν is oriented opposite to its momentum, while the spin of $\bar{\nu}$ is oriented along its momentum; in other words, ν is left-handed and $\bar{\nu}$ is right-handed. In *JETP*, paper [25] immediately followed [22]. Both papers [22, 25] were later published as a single paper in English [26]. Longitudinal neutrinos were independently considered by Salam [27] and by Lee and Yang [28]. The longitudinal neutrinos lighted up the road to the theory of the universal weak $V-A$ interaction [29, 30]. According to this theory, in the relativistic limit ($v/c \to 1$), all elementary fermions become left-handed in interactions of weak charged currents and their antiparticles become right-handed. Only a few years ago the discovery of neutrino oscillations made it clear that neutrinos are not massless and hence the theory of longitudinal neutrinos is valid only approximately, although in many cases with very high accuracy.

It is worth mentioning the idea of a possible existence of baryonic photons coupled to the baryonic charge [31]. This article became an inspiration for the further search for leptonic photons, paraphotons, and mirror photons (see Sections 3, 5, and 6).

[2] Further exposition of this statement is contained in [15].

[3] The scheme in which P and T are violated but C is conserved, was discussed at length in [17] even after it was proven false by experiment.

[4] See also [24].

3. 1960s. CP-violation

I liked the idea of strict CP-conservation very much. But on the other hand, I could not understand why the coefficients in the Lagrangian could not be complex. Thus, in the lectures at ITEP [32], weak interactions of hadrons were described based on the composite model of hadrons involving the CP-conservation assumption. In lectures in Dubna [33] and in book [34], I insisted that experimental tests of CP-invariance were one of the highest priorities. A group of Dubna experimentalists led by Okonov searched for CP-forbidden decays $K_2^0 \to \pi^+\pi^-$ and established the upper bound for their branching ratio as approximately 2×10^{-3} (they did not find two-body decays among 600 three-body decays) [35]. Unfortunately, they were stopped at this stage by their lab director. The group was unlucky: two years later, several dozens two-body decays with the branching ratio almost reached in [35] were discovered by the Princeton group [36].

The discovery of the $K_2 \to 2\pi$ decay by Christenson et al. [36] put an end to Landau's idea of strict CP-conservation, according to which antiparticles look exactly like mirror images of particles. To avoid this conclusion, Nishijima and Saffouri [37] put forward the hypothesis of a 'shadow universe' to explain the two-pion decays without CP-violation. According to [37], the decays to two pions observed in [36] were decays not of CP-odd K_2^0 but of a new hypothetical long-lived CP-even 'shadow' K_1^0-meson through its transition into an ordinary K_1^0. But it was soon shown in [38] that this mechanism contradicts the results of neutrino experiments, because shadow K_1^0-mesons would penetrate through the shielding and decay into two pions in the neutrino detector, while such events were not observed.

In the next paper, Kobzarev, Okun, and Pomeranchuk [39] postulated CPA-symmetry (A from Alice) and the existence of hypothetical mirror particles and of a mirror world. [The modern terminology, in which mirror matter refers only to the duplication of all our particles (not some of them) was *in statu nascendi*, and therefore the 'mirror world' and 'mirror particles' were used in [39] practically as synonyms. It is noteworthy that the Standard Model did not exist at that time.] According to [39], mirror particles cannot participate in ordinary strong and electromagnetic interactions with ordinary particles. In this respect, they are radically different from the right-handed protons considered by Lee and Yang [3]. The hidden mirror sector must have its own strong and electromagnetic interactions. This means that mirror particles, like ordinary ones, must form mirror atoms, molecules, and, under favorable conditions, invisible mirror stars, planets, and even mirror life. Moreover, this invisible mirror world can coexist with our world in the same space. [5]

I recall a weekend hike with Igor Kobzarev in a forest near Moscow, from Firsanovka station for Leningrad-bound trains to Nakhabino station for Riga-bound trains, when I suddenly 'saw' an invisible train crossing a clearing on invisible rails, invisible and inaudible. It was argued in our paper [39] that such a situation is impossible. A mirror train needs a mirror globe, but a mirror globe would gravitationally perturb the trajectory of our globe. Gravitational coupling between two worlds seemed indispensable because in the absence of any interaction with our matter, the mirror matter was doomed to become purely fictitious. In addition to graviton exchanges, neutrino exchanges were also allowed in [39], although we did not give any specific discussion of how this could be made consistent with the $V-A$ theory of weak interactions. The coupling of two worlds via neutral kaons was considered in [43].

Mirror particles were discussed at the Fourth European Conference on Elementary Particles (September 1967) [44] and at the Moscow conference on CP-violation (January 1968) (see [45]).

Perhaps it is worth mentioning a paper on muonic photons [46], although it had no direct relation to mirror matter. It considered an additional hypothetical photon and transitions between it and the ordinary photon through a muonic loop. This gives the effective coupling $\epsilon F_{\alpha\beta} F'_{\alpha\beta}$, where F is our field and F' is the new one, while ϵ is a dimensionless constant. Because of this coupling, which is presently called 'kinetic mixing,' the muonic neutrino acquires a tiny electric charge ϵe (see also [66, 86–89, 147–153, 231–236]).

In [47], the gravitational dipole moment of a proton was examined and it was shown to be forbidden in the framework of the general theory of relativity. The gravitational interaction of so-called sterile neutrino was considered in [48].

As is well known, in the mainstream of particle physics, quarks and electroweak theory with spontaneous symmetry breaking were suggested in the 1960s. An important article by Sakharov [49] was published, linking CP-violation with the baryonic asymmetry of the universe and, ultimately, with our existence.

4. 1970s. 'Minimum.' Exotic vacua

In the 1970s, charm, beauty, and the τ-lepton were discovered and QCD was formulated, but there was a minimum of articles on mirror particles. I am aware of only one such paper, by Pavšič [50]. A relation between mirror symmetry and the structure of a particle was attempted in it: the author claimed that mirror nucleons are unconditionally necessary because the baryons are composite, while mirror leptons are necessary only if leptons also have an internal structure. This differs from the standard concept of mirror matter. In 2001, Pavšič posted his paper [50] in an electronic archive with a note: "An early proposal of 'Mirror Matter' published in 1974" [51].

Also in the 1970s, the spontaneous breaking of gauge symmetries was brought to the cosmological model of the hot universe [52–54] and the first articles were published on spontaneous violation of the CP-symmetry [55], on the domain structure of the vacuum [56, 57], and on the metastable vacuum [58]. According to [56, 57], vacuum domains are a consequence of spontaneous violation of the CP invariance. They were then to appear during cooling of the universe after the big bang. Thus, space itself could be not mirror symmetric (recall Landau's arguments). The metastable vacuum was dubbed the false vacuum three years later (see Refs [59–61]).

5. 1980s. Revival

A revival of interest in mirror particles occurred in the 1980s. In papers [62–70], various aspects of the hidden sector of particles and interactions were considered. The existence of new long-range forces and of new x and y particles was suggested in [62]. According to [62], y-particles have no

[5] We did not know of the pioneering articles on dark matter by Oort [40] and Zwicky [41, 42].

direct interactions with the ordinary ones, while x-particles serve as connectors: they are coupled to both ordinary and y-particles. In Refs [63, 64], gluon-like θ-bosons with a large confinement radius were introduced; they could form unbreakable strings with lengths measured in kilometers. The role of θ-bosons in the early universe was discussed in [65]. In Refs [66], mirror hadrons and neutral meson connectors between the ordinary and the mirror worlds were discussed. The existence of paraphotons was suggested in [67]. Their mixing could lead to oscillations of the ordinary photons, discussed in [67]. Tiny charges of particles that are usually considered neutral (atoms and neutrinos) were analyzed in [68]. A review of hypothetical phenomena was presented in the rapporteur talk "Beyond the Standard Model" [69]. Among other subjects, photon oscillations and left–right symmetric models were discussed there, but no mirror particles were considered.

In 1986, Ellis visited ITEP; together with Voloshin, we wrote review [70], whose significant part was dedicated to mirror particles. At the last moment, seeing the review as too speculative, I decided not to submit it to the Soviet review journal *Uspekhi Fiz. Nauk*, and it was published only as an ITEP preprint [70].

Voloshin continued the quest for mirror particles. He induced the ARGUS (A Russian–German–United States–Swedish) collaboration at DESY to search for decays $\Upsilon(2S) \to \pi^+\pi^-\Upsilon(1S)$, in which $\Upsilon(1S)$ due to transitions to its mirror counterpart decays into 'nothing.' The upper bound for the branching ratio of this invisible channel was established: BR $\leqslant 2.3\%$ at 90% CL [71, 72]. The search for invisible decay products of the ϕ-meson was carried out in [73].

An experimental group at ITEP measured the upper end of the electron spectrum in tritium β-decay and announced that the mass of the electron neutrino is 30 eV (this result was not confirmed later on). This prompted Zeldovich and Khlopov to publish their review [74]. Alongside other scenarios, they considered the possibility of the neutrino mass occurring due to transitions between our left-handed neutrino and the right-handed mirror neutrino, serving as a bridge to the mirror world.

A model in which connectors between the two worlds are the so-called hybrids (particles with electroweak quantum numbers of ordinary quarks, and QCD quantum numbers of mirror quarks and their mirror counterparts) was considered in [75]. Bound states with fractional charges — fractons — occur in this model.

Schwarz and Tyupkin [76] suggested the unification of mirror and ordinary particles using the SO(20) group. In this model, mirror cosmic strings appeared. After circling such a string, an ordinary particle is transformed into a mirror particle and vice versa (see also [77, 78]).

Further cosmological and astrophysical manifestations of mirror particles were discussed in [79–85]. In particular, Alice strings were considered in [79, 82].

Glashow and collaborators became interested in the mirror universe and photon oscillations [86, 87, 90]. In Ref. [87], he suggested explaining the experimental anomaly in orthopositronium decays, observed at that time, by transitions between orthopositronium and mirror orthopositronium. (See [105, 148, 150] for the later development of this suggestion)

Tiny ordinary electric charges of mirror particles, which otherwise have no ordinary charge but have a mirror electric charge, appear due to the mixing of ordinary and mirror photons [88, 89] (see also [240]).

6. 1991–2006. 'Maximum.' From cosmology and astrophysics to LHC

A flood of mirror particles articles occurred after 1990. The Australian physicist Foot became a great enthusiast of mirror particles and published dozens of articles on this subject. One can appreciate the range of his interests by looking at the titles of references [91–132] and his book [133]. In Ref. [91], a mirror-symmetric version of the standard gauge model was considered, in particular, the mixing interaction of ordinary and mirror Higgs bosons was analyzed. We note that this renormalizable model forbids strong transitions of three quarks into three mirror quarks, as well as of two gluons into two mirror gluons, because such transitions are non-renormalizable. Therefore, an experimental discovery of the transition of a neutron into the mirror neutron, $n \leftrightarrow n'$ or $\Upsilon \leftrightarrow \Upsilon'$ would disprove this theoretical model.

The fantastic new idea of grains of mirror matter embedded in ordinary matter due to the interaction caused by the mixing of ordinary and mirror photons deserves mention [114, 118, 121]. Many of Foot's coauthors — Volkas, Ignatiev, Mitra, Gninenko, Silagadze — also published their own papers on mirror particles [134–159] (see also [151]). A few of these papers were also devoted to mirror grains [145, 147, 156–159].

An impressive contribution to the field of mirror particles was made by Berezhiani, who published over 15 papers together with his coauthors [160–176] (see also [177–182]).

Most of the papers cited in this section are based on strict mirror symmetry. They ascribe the observed macroscopic disparity between mirror and ordinary particles to the inflationary stage of the universe (see [143, 166]).

Mohapatra published about 15 papers (many of them with coauthors) on various aspects of astrophysics in the framework of broken mirror symmetry [183–196].

The search [197–201] for gravitational microlenses produced by separate stars in the halos of galaxies — the so-called MACHOs (MAssive Compact Halo Objects) — has led to the discovery of an excess of MACHOs in the direction of the Large Magellanic Cloud [200, 201]. Even before this discovery, theorists had indicated [154, 162, 202] that some of the MACHOs could be mirror stars. This interpretation was developed further in [203, 204]. Although the discovery of MACHOs has been questioned [205, 206] (see the discussion in [207, 208]), many astrophysicists believe that the observed stellar dark matter cannot consist of ordinary baryons [209, 210].

Since the publication of papers by Oort [40] and Zwicky [41, 42], two alternative explanations for the anomalously high velocities of stars and galaxies (the so-called 'virial paradox') have existed: (1) invisible dark matter, (2) anomalously strong gravity at large distances. Recent observations [211, 212] of colliding clusters of galaxies seem to settle the ambiguity in favor of dark matter. Dark matter, which manifests itself through the effect of gravitational lensing, is segregated from the luminous parts of clusters in such collisions. If this dark matter is mirror matter, then the mirror stars in it must be more prominent compared to mirror gas than ordinary stars compared to ordinary gas (Blinnikov and Silagadze, private communications).

The correlation of gamma-ray bursts with the distribution of dark matter in galaxies might suggest that these bursts are produced by explosions of mirror stars, accompanied by emission of mirror neutrinos [213–215] or of mirror axions [165, 167, 181] (see also [182]).

Supernova constraints on sterile neutrino production are given in [216]. Cosmic mirror strings as sources of cosmic rays of ultra-high energies were considered in [217] (see also [218–220]). Various aspects of mirror astrophysics were discussed in articles [221, 222] and books [223, 224]. New gauge symmetry of the type of weak SU(2) was proposed in [225] and critically analyzed in [226–230]. Leptonic (muonic) photons were discussed in the 1990s in [231–236]. Upper bounds for invisible decays of B^0 mesons and η and η' mesons were established in [237, 238]. The upper bound for the branching ratio of invisible decays of the $\Upsilon(1S)$ meson $B < 2.5 \times 10^{-3}$ was established by the Belle collaboration [239].

Various 'mirror matters' were considered in [240–246]; proposals for dark matter search were formulated in [247–249].

In 2004, a physical start was made of special ring accumulator of positrons LEPTA (Low-energy particle toroidal accumulator), one of the goals being the search for mirror orthopositronium [250–255].

A very interesting discussion of invisible decay channels of the Higgs bosons that can be generated at the Large Hadron Collider at CERN can be found in [141, 256–258] (see also [259]). The invisible decays occur because of the mixing of ordinary and mirror Higgs bosons. Higgs bosons may be discovered in the near future.

7. Concluding remarks

We compare mirror symmetry with supersymmetry. The former cannot compete with the latter in the depth of its concepts and mathematics. But it can compete in the breadth and diversity of its phenomenological predictions. Without a doubt, mirror matter is much richer than the dark matter of supersymmetry.

The preliminary version of this review was prepared for a talk at the ITEP Workshop on the Future of Heavy Flavor Physics, July 24–25, 2006 (http://www.itep.ru/eng/belle-meeting) and published on June 19 as hep-ph/0606202v1. The final version (v2) was prepared for *Physics – Uspekhi* during the summer of 2006.

As a result, the number of references has doubled. It could have risen even higher. If you google for "mirror particles" (do not forget the quotation marks!), about a thousand links are found. (Typing "mirror world" or "mirror universe" returns about 200,000 links devoted mainly to "Star Trek" television episodes.) A search in Wikipedia is suggested in some of the links. But the Wikipedia articles on mirror matter may be misleading. Instead of Google, it is better to use Google Scholar, where the number of links for 'mirror universe' is about a hundred, while for 'mirror particle' it is a few hundred. The extra articles do not deal with those mirror particles that are the subject of this review. They are 'mirror' in a different sense. For instance, the terms 'mirror families' or 'mirror fermions' refer to hypothetical families of very heavy fermions with reversed isotopic quantum numbers, which are assumed to interact with ordinary photons and gluons.

Acknowledgements

I am grateful to M V Danilov for inviting me to give a talk on mirror matter at the ITEP workshop. I thank Z G Berezhiani, S I Blinnikov, O D Dalkarov, A D Dolgov, S N Gninenko, A Yu Ignatiev, M Yu Khlopov, Z K Silagadze, R R Volkas, and M B Voloshin for very valuable suggestions, and T Basaglia, E A Ilyina, and O V Milyaeva for their help in preparing this review. This work was supported in part by grant NSh-5603.2006.2.

Author's note to the English proofs

After the Russian version of this article was published, I received a letter from S S Gershtein containing the following passage concerning footnote 1 at the beginning of the article:

"I remember well what I am describing. Many interesting foreigners came to the 1956 conference. Once during the conference I saw many people following I.E. along the corridor past the FIAN library. I was told that I.E. had invited Gell-Mann in order to hear the latest news from him. G.-M. started speaking about the tau–theta problem and said that Feynman even thought of parity nonconservation as a possibility. I.E. asked: R.P. Feynman? Yes, R.P., replied G.-M. At that point, I.S. Shapiro stood up and started (as we later joked) teaching diamat to G.-M., saying that Nature must have conservation laws. I remember that G.-M. became angry and rather sharply replied that "one has to analyze the phenomena of Nature, and not to impose laws on it."

[Comments. I.E.: Igor Evgenievich Tamm, FIAN: Lebedev Institute, diamat: dialectical materialism. If one takes this testimony at its face value, one has to conclude that the note was written by I.S. Shapiro after he learned about Feynman's remark.]

References

1. Lee T D, Yang C N "Mass degeneracy of the heavy mesons" *Phys. Rev.* **102** 290 (1956)
2. *Proc. of the Rochester Conf., April, 1956* (New York: Interscience Publ., 1956) Session VIII. Theoretical interpretation of new particles
3. Lee T D, Yang C N "Question of parity conservation in weak interactions" *Phys. Rev.* **104** 254 (1956)
4. Michel L "Interactions of elementary particles", in *Progress in Cosmic Ray Physics* Vol. 1 (Ed. J G Wilson) (Amsterdam: Elsevier, 1952) [Translated into Russian (Moscow: IL, 1954)]
5. Wick G C, Wightman A S, Wigner E P "The intrinsic parity of elementary particles" *Phys. Rev.* **88** 101 (1952)
6. Purcell E M, Ramsey N F "On the possibility of electric dipole moments for elementary particles and nuclei" *Phys. Rev.* **78** 807 (1950)
7. Berestetskii V B, Pomeranchuk I Ya "O β-raspade neitrona" ("On β-decay of the neutron") *Zh. Eksp. Teor. Fiz.* **19** 75 (1949)
8. Wu C S, Ambler E, Hayward R W, Hoppes D D, Hudson R P "Experimental test of parity conservation in beta decay" *Phys. Rev.* **105** 1413 (1957)
9. Garwin R L, Lederman L M, Weinrich M "Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment of the free muon" *Phys. Rev.* **105** 1415 (1957)
10. Friedman J I, Telegdi V L "Nuclear emulsion evidence for parity nonconservation in the decay chain $\pi^+ - \mu^+ - e^{+}$" *Phys. Rev.* **105** 1681 (1957)
11. *Tezisy Dokladov na Vsesoyuznoi Konf. po Fizike Chastits Vysokikh Energii, 14–22 Maya 1956 g.* (Book of Abstracts of the All-Union Conf. on Physics of High-Energy Particles, 14–22 May 1956) (Moscow: Izd. AN SSSR, 1956)

12. Leikin E M, Burtsev A K, Danilkin I S, Lebedev A N, Okun L B "Soveshchanie po fizike chastis vysokikh energii" ("Meeting on physics of high-energy particles") *Usp. Fiz. Nauk* **61** 103 (1957)
13. Telegdi V L "Parity violation", in *Symmetries in Physics (1600–1980). Proc. of the 1st Intern. Meeting on the History of Scientific Ideas, Sant Feliu de Guíxols, Catalonia, Spain, September 20–26, 1983* (Eds M G Doncel et al.) (Bellaterra, Barcelona, Spain: Seminari d'Historia de les Ciències, Univ. Autònoma de Barcelona, 1987) p. 433
14. Telegdi V L "The early experiments leading to the $V-A$ interactions", in *Pions to Quarks: Particle Physics in the 1950s: Based on a Fermilab Symp. (2nd Intern. Symp. on the History of Particle Physics, Batavia, Ill., 1985)* (Eds L M Brown, M Dresden, L Hoddeson) (Cambridge: Cambridge Univ. Press, 1989) p. 464
15. Shapiro I S "O nesokhranenii chetnosti pri β-raspade" ("On parity non-conservation in beta-decay") *Usp. Fiz. Nauk* **61** 313 (1957)
16. Gell-Mann M, Pais A "Behavior of neutral particles under charge conjugation" *Phys. Rev.* **97** 1387 (1955)
17. Ioffe B L "O dvukh vozmozhnykh skhemakh nesokhraneniya chetnosti v slabykh vzaimodeistviyakh" *Zh. Eksp. Teor. Fiz.* **32** 1246 (1957) ["Two possible schemes of non-conservation of parity in weak interactions" *Sov. Phys. JETP* **5** 1015 (1957)]
18. Ioffe B L, Okun L B, Rudik A P "K voprosu o nesokhranenii chetnosti v slabykh vzaimodeistviyakh" *Zh. Eksp. Teor. Fiz.* **32** 396 (1957) ["The problem of parity nonconservation in weak interactions" *Sov. Phys. JETP* **5** 328 (1957)]
19. Lee T D, Oehme R, Yang C N "Remarks on possible noninvariance under time reversal and charge conjugation" *Phys. Rev.* **106** 340 (1957)
20. Lee T D "Weak interaction and nonconservation of parity" (Nobel lecture. December 11, 1957), in *Nobel Lectures: Physics: 1901–1995* (Singapore: World Scientific) (CD-ROM) p. 406; "Slabye vzaimodeistviya i nesokhranenie chetnosti" ("Weak interaction and nonconservation of parity") *Usp. Fiz. Nauk* **66** 89 (1958)
21. Yang C N "The law of parity conservation and other symmetry laws of physics" (Nobel lecture. December 11, 1957), in *Nobel Lectures: Physics: 1901–1995* (Singapore: World Scientific) (CD-ROM) p. 393; "Zakon sokhraneniya chetnosti i drugie zakony simmetrii" ("The law of parity conservation and other symmetry laws of physics") *Usp. Fiz. Nauk* **66** 79 (1958)
22. Landau L D "O zakonakh sokhraneniya pri slabykh vzaimodeistviyakh" *Zh. Eksp. Teor. Fiz.* **32** 405 (1957) ["On the conservation laws for weak interactions" *Sov. Phys. JETP* **5** 336 (1957)]
23. Anonimnyi korrespondent gazety "Pravda". "Novoe v yadernoi fizike" ("News in nuclear physics"), "Pravda", February 15, 1957, No. 46, p. 4 (Interview with L.D. Landau)
24. Wigner E P "Relativistic invariance and quantum phenomena" *Rev. Mod. Phys.* **29** 255 (1957)
25. Landau L D "Ob odnoi vozmozhnosti dlya polyarizatsionnykh svoistv neitrino" *Zh. Eksp. Teor. Fiz.* **32** 407 (1957) ["Possible properties of the neutrino spin" *Sov. Phys. JETP* **5** 337 (1957)]
26. Landau L D "On the conservation laws for weak interactions" *Nucl. Phys.* **3** 127 (1957)
27. Salam A "On parity conservation and neutrino mass" *Nuovo Cimento* **5** 299 (1957)
28. Lee T D, Yang C N "Parity nonconservation and a two-component theory of the neutrino" *Phys. Rev.* **105** 1671 (1957)
29. Feynman R P, Gell-Mann M "Theory of the Fermi interaction" *Phys. Rev.* **109** 193 (1958)
30. Sudarshan E C G, Marshak R E "Chirality invariance and the universal Fermi interaction" *Phys. Rev.* **109** 1860 (1958)
31. Lee T D, Yang C N "Conservation of heavy particles and generalized gauge transformations" *Phys. Rev.* **98** 1501 (1955)
32. Okun L B "Lektsii po teorii slabykh vzaimodeistvii" ("Lectures on theory of weak interactions"), 17 ITEP Preprints (Moscow: Institute for Theoretical and Experimental Physics of the USSR Acad. of Sci., 1960–1961) [Translated into English "Theory of weak interactions: Thirteen lectures", AEC-tr-5226 (Oak Ridge, Tenn.: US Atomic Energy Commission); Translation of lectures 14–16 and the contents: NP-10254, 10842, 10845, 10840]
33. Okun L B "Lektsii po teorii slabykh vzaimodeistvii elementarnykh chastits" ("Lectures on the theory of weak interactions of elementary particles"), Preprint P-833 (Dubna: Joint Institute for Nuclear Research, 1961)
34. Okun L B *Slaboe Vzaimodeistvie Elementarnykh Chastits* (Weak Interactions of Elementary Particles) (Moscow: Fizmatgiz, 1963) [Translated into English (Oxford: Pergamon Press, 1965)]
35. Anikina M et al. "Eksperimental'noe izuchenie nekotorykh sledstvii CP-invariantnosti v raspadakh K_2^0-mezonov" *Zh. Eksp. Teor. Fiz.* **42** 130 (1962) ["Experimental study of some consequences of CP-invariance in the decays of K_2^0-mesons" *Sov. Phys. JETP* **15** 93 (1962)]
36. Christenson J H, Cronin J W, Fitch V L, Turlay R "Evidence for the 2π decay of the K_2^0 meson" *Phys. Rev. Lett.* **13** 138 (1964)
37. Nishijima K, Saffouri M H "*CP* invariance and the shadow universe" *Phys. Rev. Lett.* **14** 205 (1965)
38. Okun L B, Pomeranchuk I Ya "'Tenevaya Vselennaya' i neitrinnyi opyt" *Pis'ma Zh. Eksp. Teor. Fiz.* **1** (6) 28 (1965) ["'Shadow Universe' and the neutrino experiment" *JETP Lett.* **1** 167 (1965)]; *Phys. Lett.* **16** 338 (1965)
39. Kobzarev I Yu, Okun L B, Pomeranchuk I Ya "O vozmozhnosti eksperimental'nogo obnaruzheniya zerkal'nykh chastits" *Yad. Fiz.* **3** 1154 (1966) ["On the possibility of observing mirror particles" *Sov. J. Nucl. Phys.* **3** 837 (1966)]
40. Oort J H "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems" *Bull. Astron. Inst. Netherlands* **6** 249 (1932)
41. Zwicky F "Die Rotverschiebung von extragalaktischen Nebeln" *Helv. Phys. Acta* **6** 110 (1933)
42. Zwicky F "On the masses of nebulae and of clusters of nebulae" *Astrophys. J.* **86** 217 (1937)
43. Nikolaev N N, Okun L B "Possible experiments on search for mirror K^0 mesons" *Phys. Lett. B* **27** 226 (1968)
44. Okun L B, Rubbia C "CP violation" (Section 6. Mirror particles), in *Proc. of the Heidelberg Intern. Conf. on Elementary Particles, Heidelberg, Germany, September 20–27, 1967* (Ed. H Filthuth) (Amsterdam: North-Holland, 1968) p. 301
45. Okun L B "Narushenie CP-invariantnosti" (Razd. V. "Zerkal'naya simmetriya i zerkal'nye chastitsy") *Usp. Fiz. Nauk* **95** 402 (1968) ["Problems related to the violation of CP invariance" (Sect. V: "Mirror symmetry and mirror particles") *Sov. Phys. Usp.* **11** 462 (1969)]
46. Okun L B "O myuonnom zaryade i myuonnykh fotonakh" *Yad. Fiz.* **10** 358 (1969) ["On muonic charge and muonic photons" *Sov. J. Nucl. Phys.* **10** 206 (1969)]
47. Kobzarev I Yu, Okun L B "O gravitatsionnom vzaimodeistvii fermionov" *Zh. Eksp. Teor. Fiz.* **43** 1904 (1962) ["Gravitational interaction of fermions" *Sov. Phys. JETP* **16** 1343 (1963)]
48. Kobzarev I Yu, Okun L B "Ob elektromagnitnom i gravitatsionnom vzaimodeistviyakh maioranovskikh chastits" ("On electromagnetic and gravitational interactions of Majorana particles"), in *Problemy Teoreticheskoi Fiziki. Pamyati Igorya Evgen'evicha Tamma* (Problems of Theoretical Physics. A Memorial Volume to Igor Evgen'evich Tamm) (Executive Ed. V I Ritus) (Moscow: Nauka, 1972) p. 219
49. Sakharov A D "Narushenie CP-invariantnosti, C-asimmetriya i barionnaya asimmetriya Vselennoi" *Pis'ma Zh. Eksp. Teor. Fiz.* **5** 32 (1967) ["Violation of CP invariance, C asymmetry, and baryon asymmetry of the Universe" *JETP Lett.* **5** 24 (1967)]
50. Pavšič M "External inversion, internal inversion, and reflection invariance" *Int. J. Theor. Phys.* **9** 229 (1974)
51. Pavšič M "External inversion, internal inversion, and reflection invariance", hep-ph/0105344
52. Kirzhnits D A "Model' Vainberga i 'goryachaya' Vselennaya" *Pis'ma Zh. Eksp. Teor. Fiz.* **15** 745 (1972) ["Weinberg model and the 'hot' Universe" *JETP Lett.* **15** 529 (1972)]
53. Kirzhnits D A, Linde A D "Macroscopic consequences of the Weinberg model" *Phys. Lett. B* **42** 471 (1972)
54. Linde A D "Phase transitions in gauge theories and cosmology" *Rep. Prog. Phys.* **42** 389 (1979)
55. Lee T D "CP nonconservation and spontaneous symmetry breaking" *Phys. Rep.* **9** 143 (1974)
56. Kobzarev I Yu, Okun L B, Zeldovich Ya B "Spontaneous *CP*-violation and cosmology" *Phys. Lett. B* **50** 340 (1974)

57. Zel'dovich Ya B, Kobzarev I Yu, Okun L B "Kosmologicheskie sledstviya spontannogo narusheniya diskretnoi simmetrii" *Zh. Eksp. Teor. Fiz.* **67** 3 (1974) ["Cosmological consequences of a spontaneous breakdown of a discrete symmetry" *Sov. Phys. JETP* **40** 1 (1974)]
58. Voloshin M B, Kobzarev I Yu, Okun L B "O puzyr'kakh v metastabil'nom vakuume" *Yad. Fiz.* **20** 1229 (1974) ["Bubbles in metastable vacuum" *Sov. J. Nucl. Phys.* **20** 644 (1975)]
59. Frampton P H "Consequences of vacuum instability in quantum field theory" *Phys. Rev. D* **15** 2922 (1977)
60. Coleman S "Fate of the false vacuum: Semiclassical theory" *Phys. Rev. D* **15** 2929 (1977); "Erratum" **16** 1248 (1977)
61. Callan C G (Jr.), Coleman S "Fate of the false vacuum. II. First quantum corrections" *Phys. Rev. D* **16** 1762 (1977)
62. Okun L B "O poiskakh novykh dal'nodeistvuyushchikh sil" *Zh. Eksp. Teor. Fiz.* **79** 694 (1980) ["On the search for new long-range forces" *Sov. Phys. JETP* **52** 351 (1980)]
63. Okun L B "Theta particles" *Nucl. Phys. B* **173** 1 (1980)
64. Okun L B "Tetony" *Pis'ma Zh. Eksp. Teor. Fiz.* **31** 156 (1979) ["Thetons" *JETP Lett.* **31** 144 (1979)]
65. Dolgov A D "O kontsentratsii reliktovykh θ-chastits" *Yad. Fiz.* **31** 1522 (1980) ["On concentration of relict theta particles" *Sov. J. Nucl. Phys.* **31** 790 (1980)]
66. Okun L B "Granitsy elektrodinamiki: parafotony?" *Zh. Eksp. Teor. Fiz.* **83** 892 (1982) ["Limits of electrodynamics: paraphotons?" *Sov. Phys. JETP* **56** 502 (1982)]; Preprint ITEP-48 (Moscow: Institute for Theoretical and Experimental Physics of the USSR Acad. of Sci., 1982)
67. Okun L B "On a search for mirror particles", Preprint ITEP-149 (Moscow: ITEP, 1983) and Addendum to the preprint
68. Okun L B, Voloshin M B, Zakharov V I "Electrical neutrality of atoms and grand unification models" *Phys. Lett. B* **138** 115 (1984)
69. Okun L "Beyond the standard model", in *Proc. of the 25th Intern. Conf. on High Energy Phys., Singapore, 2–8 August 1990* Vol. 1 (Eds K K Phua, Y Yamaguchi) (Singapore: World Scientific, 1991) p. 319
70. Voloshin M B, Okun L B, Ellis J R "O poiskakh novykh pronikayushchikh chastits" ("On searches for new penetrating particles"), Preprint ITEP-86-189 (Moscow: Institute for Theoretical and Experimental Physics of the USSR Acad. of Sci., 1986)
71. Albrecht H et al. (The ARGUS Collab.) "Search for exotic decay modes of the $\Upsilon(1s)$" *Phys. Lett. B* **179** 403 (1986)
72. Voloshin M B, Zaitsev Yu M "Fizika Υ-rezonansov: desyat' let spustya" *Usp. Fiz. Nauk* **152** 361 (1987) ["Physics of Υ resonances: ten years later" *Sov. Phys. Usp.* **30** 553 (1987)]
73. Druzhinin V P et al. "Search for rare radiative decays of φ-meson at VEPP-2M" *Z. Phys. C* **37** 1 (1988)
74. Zel'dovich Ya B, Khlopov M Yu "Massa neitrino v fizike elementarnykh chastits i kosmologii rannei Vselennoi" *Usp. Fiz. Nauk* **135** 45 (1981) ["The neutrino mass in elementary-particle physics and in big bang cosmology" *Sov. Phys. Usp.* **24** 755 (1981)]
75. Khlopov M Yu "Drobnozaryazhennye chastitsy i nevyletanie kvarkov" *Pis'ma Zh. Eksp. Teor. Fiz.* **33** 170 (1981) ["Fractionally charged particles and quark confinement" *JETP Lett.* **33** 162 (1981)]
76. Schwarz A S, Tyupkin Yu S "Grand unification and mirror particles" *Nucl. Phys. B* **209** 427 (1982)
77. Schwarz A S "Field theories with no local conservation of the electric charge" *Nucl. Phys. B* **208** 141 (1982)
78. Schwarz A S, Tyupkin Yu S "Grand unification, strings and mirror particles", in *Group Theoretical Methods in Physics: Proc. of the Intern. Seminar, Zvenigorod, 24–26 November 1982* Vol. 2 (Executive Ed. M A Markov) (Moscow: Nauka, 1983) p. 279
79. Blinnikov S I, Khlopov M Yu "O vozmozhnykh proyavleniyakh 'zerkal'nykh' chastits" *Yad. Fiz.* **36** 809 (1982) ["On possible effect of 'mirror' particles" *Sov. J. Nucl. Phys.* **36** 472 (1982)]
80. Blinnikov S I, Khlopov M Yu "Excitation of the solar oscillations by objects consisting of y-matter" *Solar Phys.* **82** 383 (1983)
81. Blinnikov S I, Khlopov M Yu "O vozmozhnykh astronomicheskikh proyavleniyakh 'zerkal'nykh' chastits" *Astron. Zh.* **60** 632 (1983) ["Possible astronomical effects of mirror particles" *Sov. Astron.* **27** 371 (1983)]
82. Sazhin M V, Khlopov M Yu "Kosmologicheskie niti i effekty gravitatsionnoi linzy" *Astron. Zh.* **66** 191 (1989) ["Cosmic strings and gravitational lens effects" *Sov. Astron.* **33** 98 (1989)]
83. Dubrovich V K, Khlopov M Yu "O domennoi strukture tenevoi materii" *Astron. Zh.* **66** 232 (1989) ["On the domain structure of shadow matter" *Sov. Astron.* **33** 116 (1989)]
84. Kolb E W, Seckel D, Turner M S "The shadow world of superstring theories" *Nature* **314** 415 (1985)
85. Goldberg H, Hall L J "A new candidate for dark matter" *Phys. Lett. B* **174** 151 (1986)
86. Georgi H, Ginsparg P, Glashow S L "Photon oscillations and cosmic background radiation" *Nature* **306** 765 (1983)
87. Glashow S L "Positronium versus the mirror universe" *Phys. Lett. B* **167** 35 (1986)
88. Holdom B "Two U(1)'s and ε charge shifts" *Phys. Lett. B* **166** 196 (1986)
89. Holdom B "Searching for ε charges and a new U(1)" *Phys. Lett. B* **178** 65 (1986)
90. Carlson E D, Glashow S L "Nucleosynthesis versus the mirror universe" *Phys. Lett. B* **193** 168 (1987)
91. Foot R, Lew H, Volkas R R "A model with fundamental improper spacetime symmetries" *Phys. Lett. B* **272** 67 (1991)
92. Foot R, He X-G "Comment on $Z-Z'$ mixing in extended gauge theories" *Phys. Lett. B* **267** 509 (1991)
93. Foot R, Lew H, Volkas R R "Possible consequences of parity conservation" *Mod. Phys. Lett. A* **7** 2567 (1992)
94. Foot R "Neutrino oscillations and the exact parity model" *Mod. Phys. Lett. A* **9** 169 (1994); hep-ph/9402241
95. Foot R "A Parity invariant $SU(3)_C \times SU(3)_L \times U(1)$ model" *Mod. Phys. Lett. A* **10** 159 (1995); hep-ph/9402244
96. Foot R, Lew H "A novel left-right symmetric model", hep-ph/9411390
97. Foot R "Experimental signatures of a massive mirror photon", hep-ph/9407331
98. Foot R, Volkas R R "Neutrino physics and the mirror world: How exact parity symmetry explains the solar neutrino deficit, the atmospheric neutrino anomaly, and the LSND experiment" *Phys. Rev. D* **52** 6595 (1995); hep-ph/9505359
99. Brumby S P, Foot R, Volkas R R "Quaternionic formulation of the exact parity model", hep-th/9602139
100. Foot R, Volkas R R "The exact parity symmetric model and big bang nucleosynthesis" *Astropart. Phys.* **7** 283 (1997); hep-ph/9612245
101. Collie M, Foot R "Neutrino masses in the $SU(5) \times SU(5)'$ mirror symmetric model" *Phys. Lett. B* **432** 134 (1998); hep-ph/9803261
102. Foot R "Have mirror planets been observed?" *Phys. Lett. B* **471** 191 (1999); astro-ph/9908276
103. Foot R "Have mirror stars been observed?" *Phys. Lett. B* **452** 83 (1999); astro-ph/9902065
104. Foot R, Volkas R R "Implications of mirror neutrinos for early universe cosmology" *Phys. Rev. D* **61** 043507 (2000); hep-ph/9904336
105. Foot R, Gninenko S N "Can the mirror world explain the orthopositronium lifetime puzzle?" *Phys. Lett. B* **480** 171 (2000); hep-ph/0003278
106. Foot R, Lew H, Volkas R R "Unbroken versus broken mirror world: a tale of two vacua" *J. High Energy Phys.* (JHEP07) 032 (2000); hep-ph/0006027
107. Foot R, Ignatiev A Yu, Volkas R R "Do 'isolated' planetary mass objects orbit mirror stars?" *Astropart. Phys.* **17** 195 (2002); astro-ph/0010502
108. Foot R, Ignatiev A Yu, Volkas R R "Physics of mirror photons" *Phys. Lett. B* **503** 355 (2001); astro-ph/0011156
109. Foot R "Are mirror worlds opaque?" *Phys. Lett. B* **505** 1 (2001); astro-ph/0101055
110. Foot R "Seven (and a half) reasons to believe in mirror matter: from neutrino puzzles to the inferred dark matter in the universe" *Acta Phys. Polon. B* **32** 2253 (2001); astro-ph/0102294
111. Foot R, Silagadze Z K "Do mirror planets exist in our solar system?" *Acta Phys. Polon. B* **32** 2271 (2001); astro-ph/0104251
112. Foot R "The mirror world interpretation of the 1908 Tunguska event and other more recent events" *Acta Phys. Polon. B* **32** 3133 (2001); hep-ph/0107132

113. Foot R, Volkas R R "A mirror world explanation for the Pioneer spacecraft anomalies?" *Phys. Lett. B* **517** 13 (2001); hep-ph/0108051
114. Foot R, Yoon T L "Exotic meteoritic phenomena: The Tunguska event and anomalous low altitude fireballs — manifestations of the mirror world?" *Acta Phys. Polon. B* **33** 1979 (2002); astro-ph/0203152
115. Foot R, Mitra S "Ordinary atom–mirror-atom bound states: A new window on the mirror world" *Phys. Rev. D* **66** 061301 (2002); hep-ph/0204256
116. Foot R "Does mirror matter exist?", hep-ph/0207175
117. Foot R, Mitra S "Mirror matter in the solar system: new evidence for mirror matter from Eros" *Astropart. Phys.* **19** 739 (2003); astro-ph/0211067
118. Mitra S, Foot R "Detecting dark matter using centrifuging techniques" *Phys. Lett. B* **558** 9 (2003); astro-ph/0301229
119. Foot R, Volkas R R "Was ordinary matter synthesized from mirror matter? An attempt to explain why $\Omega_{\rm Baryon} \approx 0.2\Omega_{\rm Dark}$" *Phys. Rev. D* **68** 021304 (2003); hep-ph/0304261
120. Foot R, Mitra S "Have mirror micrometeorites been detected?" *Phys. Rev. D* **68** 071901 (2003); hep-ph/0306228
121. Foot R, Mitra S "Detecting mirror matter on Earth via its thermal imprint on ordinary matter" *Phys. Lett. A* **315** 178 (2003); cond-mat/0306561
122. Foot R "Implications of the DAMA and CRESST experiments for mirror matter-type dark matter" *Phys. Rev. D* **69** 036001 (2004); hep-ph/0308254
123. Foot R "Experimental implications of mirror matter-type dark matter" *Int. J. Mod. Phys. A* **19** 3807 (2004); astro-ph/0309330
124. Foot R, Volkas R R "Explaining $\Omega_{\rm Baryon} \approx 0.2\Omega_{\rm Dark}$ through the synthesis of ordinary matter from mirror matter: A more general analysis" *Phys. Rev. D* **69** 123510 (2004); hep-ph/0402267
125. Foot R "Exploring the mirror matter interpretation of the DAMA experiment: Has the dark matter problem been solved?", astro-ph/0403043
126. Foot R, Silagadze Z K "Supernova explosions, 511 keV photons, gamma ray bursts and mirror matter" *Int. J. Mod. Phys. D* **14** 143 (2005); astro-ph/0404515
127. Foot R "Reconciling the positive DAMA annual modulation signal with the negative results of the CDMS II experiment" *Mod. Phys. Lett. A* **19** 1841 (2004); astro-ph/0405362
128. Foot R, Volkas R R "Spheroidal galactic halos and mirror dark matter" *Phys. Rev. D* **70** 123508 (2004); astro-ph/0407522
129. Foot R "Testing the mirror world hypothesis for the close-in extrasolar planets" *Acta Phys. Polon. B* **35** 2473 (2004); astro-ph/0406257
130. Foot R "Mirror matter-type dark matter" *Int. J. Mod. Phys. D* **13** 2161 (2004); astro-ph/0407623
131. Foot R "Generalized mirror matter models" *Phys. Lett. B* **632** 467 (2006); hep-ph/0507294
132. Foot R "Implications of the DAMA/NAI and CDMS experiments for mirror matter-type dark matter" *Phys. Rev. D* **74** 023514 (2006); astro-ph/0510705
133. Foot R *Shadowlands: Quest for Mirror Matter in the Universe* (Parkland, Fla.: Universal Publ., 2002)
134. Volkas R R, Wong Y Y Y "Energy-dependent solar neutrino flux depletion in the exact parity model and implications for SNO, SuperKamiokande, and BOREXINO" *Phys. Rev. D* **58** 113001 (1998); hep-ph/9803456
135. Bell N F, Volkas R R, Wong Y Y Y "Relic neutrino asymmetry evolution from first principles" *Phys. Rev. D* **59** 113001 (1999); hep-ph/9809363
136. Bell N F, Volkas R R "Mirror matter and primordial black holes" *Phys. Rev. D* **59** 107301 (1999); astro-ph/9812301
137. Volkas R R "Neutrino physics and the mirror world", in *8th Intern. Workshop on Neutrino Telescopes, Venezia, February 23–26, 1999* Vol. 2 (Ed. M B Ceolin) (Padova: Papergraf, 1999) p. 13; hep-ph/9904437
138. Volkas R R, Wong Y Y Y "Matter-affected neutrino oscillations in ordinary and mirror stars and their implications for gamma-ray bursts" *Astropart. Phys.* **13** 21 (2000); astro-ph/9907161
139. Volkas R R "Mirror neutrinos and the early universe", in *Cosmo-99: Proc. of the 3rd Intern. Workshop on Particle Physics and the Early Universe, Trieste, Italy, 27 September–2 October 1999* (Eds U Cotti et al.) (Singapore: World Scientific, 2000) p. 347; hep-ph/0002002
140. Bell N F "Mirror matter and heavy singlet neutrino oscillations in the early universe" *Phys. Lett. B* **479** 257 (2000); hep-ph/0003072
141. Ignatiev A Yu, Volkas R R "Discovering mirror particles at the Large Hadron Collider and the implied cold universe" *Phys. Lett. B* **487** 294 (2000); hep-ph/0005238
142. Ignatiev A Yu, Volkas R R "Geophysical constraints on mirror matter within the Earth" *Phys. Rev. D* **62** 023508 (2000); hep-ph/0005125
143. Ignatiev A Yu, Volkas R R "Mirror dark matter and large scale structure" *Phys. Rev. D* **68** 023518 (2003); hep-ph/0304260
144. Ignatiev A Yu, Volkas R R "Mirror matter", Talk given at *15th Biennial Congress of the Australian Institute of Physics, Sydney, Australia, 8–11 July 2002*; hep-ph/0306120
145. Mitra S "Applications of mirror matter", http://people.zeelandnet.nl/smitra/applications.htm
146. Mitra S "Has DAMA detected self-interacting dark matter?" *Phys. Rev. D* **71** 121302(R) (2005); astro-ph/0409121
147. Mitra S "Detecting dark matter in electromagnetic field penetration experiments" *Phys. Rev. D* **74** 043532 (2006); astro-ph/0605369
148. Gninenko S N "Limit on 'disappearance' of orthopositronium in vacuum" *Phys. Lett. B* **326** 317 (1994)
149. Gninenko S N, Krasnikov N V, Rubbia A "Invisible decay of orthopositronium vs extra dimensions", hep-ph/0205056
150. Badertscher A et al. "An apparatus to search for mirror dark matter via the invisible decay of orthopositronium in vacuum", hep-ex/0311031; Gninenko S N "An apparatus to search for mirror dark matter" *Int. J. Mod. Phys. A* **19** 3833 (2004)
151. Rubbia A "Positronium as a probe for new physics beyond the standard model" *Int. J. Mod. Phys. A* **19** 3961 (2004); hep-ph/0402151
152. Badertscher A et al. "A new experiment to search for the invisible decay of the orthopositronium", hep-ex/0404037
153. Gninenko S N, Krasnikov N V, Matveev V A, Rubbia A "Some aspects of positronium physics" *Fiz. Elem. Chastits At. Yad.* **37** 604 (2006) [*Phys. Part. Nucl.* **37** 321 (2006)]
154. Silagadze Z K "Neutrino mass and the mirror universe" *Yad. Fiz.* **60** 336 (1997) [*Phys. At. Nucl.* **60** 272 (1997)]; hep-ph/9503481
155. Silagadze Z K "Mirror World versus large extra dimensions" *Mod. Phys. Lett. A* **14** 2321 (1999); hep-ph/9908208
156. Silagadze Z K "TeV scale gravity, mirror universe, and ... dinosaurs", Extended version of the talk given at *Gran Sasso Summer Institute: Massive Neutrinos in Physics and Astrophysics, Gran Sasso, Italy, 13–24 September 1999*; *Acta Phys. Polon. B* **32** 99 (2001); hep-ph/0002255
157. Silagadze Z K "Mirror objects in the solar system?", Prepared for *Tunguska 2001: Intern. Conf., Moscow, Russia, 30 June–1 July 2001*; *Acta Phys. Polon. B* **33** 1325 (2002); astro-ph/0110161
158. Silagadze Z K "Tunguska genetic anomaly and electrophonic meteors" *Acta Phys. Polon. B* **36** 935 (2005); astro-ph/0311337
159. Silagadze Z K "Maxwell's demon through the looking glass" *Acta Phys. Polon. B* **38** 101 (2007); physics/0608114
160. Akhmedov E Kh, Berezhiani Z G, Senjanović G "Planck-scale physics and neutrino masses" *Phys. Rev. Lett.* **69** 3013 (1992); hep-ph/9205230
161. Berezhiani Z G, Mohapatra R N "Reconciling present neutrino puzzles: Sterile neutrinos as mirror neutrinos" *Phys. Rev. D* **52** 6607 (1995); hep-ph/9505385
162. Berezhiani Z G, Dolgov A D, Mohapatra R N "Asymmetric inflationary reheating and the nature of mirror universe" *Phys. Lett. B* **375** 26 (1996); hep-ph/9511221
163. Berezhiani Z G "Astrophysical implications of the mirror world with broken mirror parity" *Acta Phys. Polon. B* **27** 1503 (1996); hep-ph/9602326
164. Berezhiani Z "Unified picture of the particle and sparticle masses in SUSY GUT" *Phys. Lett. B* **417** 287 (1998)
165. Berezhiani Z, Drago A "Gamma ray bursts via emission of axion-like particles" *Phys. Lett. B* **473** 281 (2000); hep-ph/9911333
166. Berezhiani Z, Comelli D, Villante F L "The early mirror universe: inflation, baryogenesis, nucleosynthesis and dark matter" *Phys. Lett. B* **503** 362 (2001); hep-ph/0008105

167. Berezhiani Z, Gianfagna L, Giannotti M "Strong CP problem and mirror world: the Weinberg – Wilczek axion revisited" *Phys. Lett. B* **500** 286 (2001); hep-ph/0009290
168. Bento L, Berezhiani Z "Leptogenesis via collisions: leaking lepton number to the hidden sector" *Phys. Rev. Lett.* **87** 231304 (2001); hep-ph/0107281
169. Bento L, Berezhiani Z "Baryon asymmetry, dark matter and the hidden sector" *Fortschr. Phys.* **50** 489 (2002)
170. Bento L, Berezhiani Z "Baryogenesis: the lepton leaking mechanism", hep-ph/0111116
171. Berezhiani Z "Mirror world and its cosmological consequences" *Int. J. Mod. Phys. A* **19** 3775 (2004); hep-ph/0312335
172. Berezhiani Z, Ciarcelluti P, Comelli D, Villante F L "Structure formation with mirror dark matter: CMB and LSS" *Int. J. Mod. Phys. D* **14** 107 (2005); astro-ph/0312605
173. Berezhiani Z, Bento L "Neutron – mirror-neutron oscillations: How fast might they be?" *Phys. Rev. Lett.* **96** 081801 (2006); hep-ph/0507031
174. Berezhiani Z, Ciarcelluti P, Cassisi S, Pietrinferni A "Evolutionary and structural properties of mirror star MACHOs" *Astropart. Phys.* **24** 495 (2006); astro-ph/0507153
175. Berezhiani Z "Through the looking-glass: Alice's adventures in mirror world", in *From Fields to Strings: Circumnavigating Theoretical Physics: Ian Kogan Memorial Collection* Vol. 3 (Eds M Shifman et al.) (Singapore: World Scientific, 2005) p. 2147; hep-ph/0508233
176. Berezhiani Z, Bento L "Fast neutron – mirror neutron oscillation and ultra high energy cosmic rays" *Phys. Lett. B* **635** 253 (2006); hep-ph/0602227
177. Ciarcelluti P "Cosmology of the mirror universe", Ph.D. Thesis (Advisor: Z Berezhiani); astro-ph/0312607
178. Ciarcelluti P "Structure formation, CMB and LSS in a mirror dark matter scenario" *Frascati Phys. Ser.* **555** 1 (2004); astro-ph/0409629
179. Ciarcelluti P "Cosmology with mirror dark matter I: Linear evolution of perturbations" *Int. J. Mod. Phys. D* **14** 187 (2005); astro-ph/0409630
180. Ciarcelluti P "Cosmology with mirror dark matter II: Cosmic microwave background and large scale structure" *Int. J. Mod. Phys. D* **14** 223 (2005); astro-ph/0409633
181. Gianfagna L, Giannotti M, Nesti F "Mirror world, supersymmetric axion and gamma ray bursts" *J. High Energy Phys.* (JHEP10) 044 (2004); hep-ph/0409185
182. Giannotti M "Mirror world and axion: relaxing cosmological bounds" *Int. J. Mod. Phys. A* **20** 2454 (2005); astro-ph/0504636
183. Mohapatra R N, Teplitz V L "Structures in the mirror universe" *Astrophys. J.* **478** 29 (1997); astro-ph/9603049
184. Brahmachari B, Mohapatra R N "Grand unification of the sterile neutrino" *Phys. Lett. B* **437** 100 (1998); hep-ph/9805429
185. Mohapatra R N "Sterile neutrinos: phenomenology and theory", hep-ph/9808236
186. Mohapatra R N, Sciama D W "Diffuse ionization in the Milky Way and sterile neutrinos", hep-ph/9811446
187. Mohapatra R N, Teplitz V L "Mirror matter MACHOs" *Phys. Lett. B* **462** 302 (1999)
188. Mohapatra R N "Sterile neutrinos", hep-ph/9903261
189. Mohapatra R N, Nussinov S, Teplitz V L "TeV scale quantum gravity and mirror supernovae as sources of gamma-ray bursts" *Astropart. Phys.* **13** 295 (2000); astro-ph/9909376
190. Mohapatra R N "Theories of neutrino masses and mixings", in *Current Aspects of Neutrino Physics* (Ed. D O Caldwell) (Berlin: Springer-Verlag, 2001) p. 217; hep-ph/9910365
191. Mohapatra R N, Teplitz V L "Mirror dark matter and galaxy core densities" *Phys. Rev. D* **62** 063506 (2000); astro-ph/0001362
192. Mohapatra R N, Teplitz V L "Mirror dark matter", astro-ph/0004046
193. Mohapatra R N "Origin of neutrino masses and mixings" *Nucl. Phys. B: Proc. Suppl.* **91** 313 (2001); hep-ph/0008232
194. Mohapatra R N, Nussinov S, Teplitz V L "Mirror matter as self-interacting dark matter" *Phys. Rev. D* **66** 063002 (2002); hep-ph/0111381
195. Mohapatra R N, Nasri S "Avoiding BBN constraints on mirror models for sterile neutrinos" *Phys. Rev. D* **71** 053001 (2005); hep-ph/0407194
196. Mohapatra R N, Nasri S, Nussinov S "Some implications of neutron mirror neutron oscillation" *Phys. Lett. B* **627** 124 (2005); hep-ph/0508109
197. Aubourg É et al. "Microlensing optical depth of the Large Magellanic Cloud" *Astron. Astrophys.* **347** 850 (1999)
198. Lasserre T et al. "Not enough stellar mass Machos in the Galactic halo" *Astron. Astrophys.* **355** L39 (2000)
199. Alcock C et al. "Possible gravitational microlensing of a star in the Large Magellanic Cloud" *Nature* **365** 621 (1993)
200. Alcock C et al. "The MACHO project Large Magellanic Cloud microlensing results from the first two years and the nature of the Galactic Dark Halo" *Astrophys. J.* **486** 697 (1997)
201. Alcock C et al. "The MACHO project: microlensing results from 5.7 years of Large Magellanic Cloud observations" *Astrophys. J.* **542** 281 (2000)
202. Hodges H M "Mirror baryons as the dark matter" *Phys. Rev. D* **47** 456 (1993)
203. Blinnikov S I "A quest for weak objects and for invisible stars", astro-ph/9801015
204. Mohapatra R N, Teplitz V L "Mirror matter MACHOS", astro-ph/9902085
205. Belokurov V, Evans N W, Du Y L "Light-curve classification in massive variability surveys. I. Microlensing" *Mon. Not. R. Astron. Soc.* **341** 1373 (2003)
206. Belokurov V, Evans N W, Du Y L "Light-curve classification in massive variability surveys. II. Transients towards the Large Magellanic Cloud" *Mon. Not. R. Astron. Soc.* **352** 233 (2004)
207. Griest K, Thomas C L "Contamination in the MACHO data set and the puzzle of Large Magellanic Cloud microlensing" *Mon. Not. R. Astron. Soc.* **359** 464 (2005); astro-ph/0412443
208. Evans N W, Belokurov V "Is there a microlensing puzzle?", astro-ph/0505167
209. Freese K "Death of baryonic dark matter" *Phys. Rep.* **333 – 334** 183 (2000)
210. Freese K, Fields B, Graff D "Death of stellar baryonic dark matter", in *The First Stars: Proc. of the MPA/ESO Symp., Garching, Germany, 4 – 6 Aug. 1999* (ESO Astrophys. Symp., Eds A Weiss, T Abel, V Hill) (Berlin: Springer, 2000) p. 18; astro-ph/0002058
211. Bradač M et al. "Strong and weak lensing united. III. Measuring the mass distribution of the merging galaxy cluster 1E0657–56" *Astrophys. J.* **652** 937 (2006); astro-ph/0608408
212. Clowe D et al. "A direct empirical proof of the existence of dark matter" *Astrophys. J.* **648** L109 (2006); astro-ph/0608407
213. Blinnikov S "Gamma-ray bursts produced by mirror stars", astro-ph/9902305
214. Blinnikov S "Cosmic gamma-ray bursts" *Surv. High Energy Phys.* **15** 37 (2000); astro-ph/9911138
215. Tsvetkov D Yu, Blinnikov S I, Pavlyuk N N "Gamma-ray bursts, type Ib/c supernovae and star-forming sites in host galaxies", astro-ph/0101362
216. Kolb E W, Mohapatra R N, Teplitz V L "New supernova constraints on sterile-neutrino production" *Phys. Rev. Lett.* **77** 3066 (1996); hep-ph/9605350
217. Berezinsky V, Vilenkin A "Ultrahigh energy neutrinos from hidden-sector topological defects" *Phys. Rev. D* **62** 083512 (2000); hep-ph/9908257
218. Berezinsky V, Narayan M, Vissani F "Mirror model for sterile neutrinos" *Nucl. Phys. B* **658** 254 (2003); hep-ph/0210204
219. Berezinsky V "SuperGZK neutrinos: testing physics beyond the Standard Model", hep-ph/0303091
220. Berezinsky V "SuperGZK neutrinos", astro-ph/0509675
221. Khlopov M Yu, Beskin G M, Bochkareva N G, Pustil'nik L A, Pustil'nik S A "Nablyudatel'naya fizika zerkal'nogo mira" *Astron. Zh.* **68** 42 (1991) ["Observational physics of mirror world" *Sov. Astron.* **35** 21 (1991)]
222. Oknyanskij V L "QSO 0957 + 561 and other large-separated double quasars: some new results and a future observational project", astro-ph/0502087
223. Khlopov M Yu *Cosmoparticle Physics* (Singapore: World Scientific, 1999)
224. Khlopov M Yu *Osnovy Kosmofiziki* (Fundamentals of Cosmophysics) (Moscow: URSS, 2004)

225. Glashow S L "A sinister extension of the standard model to SU(3) × SU(2) × SU(2) × U(1)", in *11th Intern. Workshop on Neutrino Telescopes, Venice, Italy, 22–25 February 2005* (Ed. M B Ceolin) (Padova: Papergraf, 2005) p. 539; hep-ph/0504287
226. Khlopov M Yu "New symmetries in microphysics, new stable forms of matter around us", Invited contribution to the special issue of honorary series of "Annals of Louis de Broglie Foundations", dedicated to E. Majorana; astro-ph/0607048
227. Belotsky K M, Khlopov M Yu, Shibaev K I "Composite dark matter and its charged constituents", astro-ph/0604518
228. Khlopov M Yu, Stephan S A "Composite dark matter with invisible light from almost-commutative geometry", astro-ph/0603187
229. Fargion D, Khlopov M, Stephan C A "Cold dark matter by heavy double charged leptons?", astro-ph/0511789 (v1)
230. Fargion D, Khlopov M "Tera-leptons shadows over Sinister Universe", hep-ph/0507087
231. Çiftçi A K, Sultansoi S, Türköz S "Lepton charge as possible source of new force", Preprint AU/94-03(HEP) (Ankara: Ankara Univ., 1994)
232. Blinnikov S I, Dolgov A D, Okun L B, Voloshin M B "How strong can the coupling of leptonic photons be?" *Nucl. Phys. B* **458** 52 (1996); hep-ph/9505444
233. Okun L "Leptons and photons" *Phys. Lett. B* **382** 389 (1996); hep-ph/9512436
234. Okun L B "Leptonic photon and light element abundancies" *Mod. Phys. Lett. A* **11** 3041 (1996); hep-ph/9611360
235. Ilyin V A, Okun L B, Rozanov A N "On the search for muonic photons in neutrino experiments" *Nucl. Phys. B* **525** 51 (1998); hep-ph/9707479
236. Akkus B et al. (CHARM II Collab.) "Experimental search for muonic photons" *Phys. Lett. B* **434** 200 (1998)
237. Aubert B et al. (BABAR Collab.) "Search for B^0 decays to invisible final states and to $\nu\bar{\nu}\gamma$" *Phys. Rev. Lett.* **93** 091802 (2004); hep-ex/0405071
238. Ablikim M (BES Collab.) "Search for invisible decays of η and η' in $J/\psi \to \phi\eta$ and $\phi\eta'$" *Phys. Rev. Lett.* **97** 202002 (2006); hep-ex/0607006
239. Tajima O et al. (Belle Collab.) "Seach for the invisible decay of the $\Upsilon(1S)$" *Phys. Rev. Lett.* **98** 132001 (2007); hep-ex/0611041
240. Davidson S, Campbell B, Bailey D "Limits on particles of small electric charge" *Phys. Rev. D* **43** 2314 (1991)
241. Okun L B "Spacetime and vacuum as seen from Moscow", in *2001: A Spacetime Odyssey: Proc. of the Inaugural Conf. of the Michigan Center for Theoretical Physics, Michigan, USA, 21–25 May 2001* (Eds M J Duff, J T Liu) (River Edge, NJ: World Scientific, 2002) p. 105; *Int. J. Mod. Phys. A* **17** (S1) 105 (2002); hep-ph/0112031
242. Dubovsky S L, Sibiryakov S M "Domain walls in noncommutative gauge theories, folded D-branes, and communication with mirror world" *Nucl. Phys. B* **691** 91 (2004); hep-th/0401046
243. Evslin J, Fairbairn M "Photon mixing in domain walls and the cosmic coincidence problem" *J. Cosmol. Astropart. Phys.* (JCAP02) 011 (2006); hep-ph/0507020
244. Pokotilovski Yu N "On the experimental search for neutron → mirror neutron oscillations" *Phys. Lett. B* **639** 214 (2006); nucl-ex/0601017
245. Dolgov A D "Cosmology and new physics", hep-ph/0606230
246. Maksimenko O "V poiskakh zerkal'nogo mira" ("Search of mirror world") *Vokrug Sveta* (6) 29 (2006)
247. Rosner J L "Dark matter in many forms", astro-ph/0509196
248. Seto N, Cooray A "Search for small-mass black-hole dark matter with space-based gravitational wave detectors" *Phys. Rev. D* **70** 063512 (2004); astro-ph/0405216
249. Adams A W, Bloom J S "Direct detection of dark matter with space-based laser interferometers", astro-ph/0405266
250. Meshkov I N, Skrinsky A N "The antihydrogen and positronium generation and studies using storage rings" *Nucl. Instrum. Meth. A* **391** 205 (1997)
251. Meshkov I N, Sidorin A O "Conceptual design of the low energy positron storage ring" *Nucl. Instrum. Meth. A* **391** 216 (1997)
252. Kartavtsev O I, Meshkov I N "Method of in-flight production of exotic systems in charge-exchange reactions" *Nucl. Instrum. Meth. A* **391** 221 (1997)
253. Meshkov I N "Eksperimental'nye issledovaniya fiziki antivodoroda i pozitroniya. Problemy i vozmozhnosti" *Fiz. Elem. Chastits At. Yad.* **28** 495 (1997) ["Experimental studies of antihydrogen and positronium physics: Problems and possibilities" *Phys. Part. Nucl.* **28** 198 (1997)]
254. Meshkov I N "High-precision experiments with antihydrogen and positronium in flight" *Yad. Fiz.* **61** 1796 (1998) [*Phys. At. Nucl.* **61** 1679 (1998)]
255. Boltushkin E V et al. "Fizicheskii pusk nakopitelya LEPTA" *At. Energiya* **98** (3) 225 (2005) ["Physical startup of the LEPTA accumulator" *At. Energy* **98** (3) 217 (2005)]
256. Chacko Z, Goh H-S, Harnik R "Natural electroweak breaking from a mirror symmetry" *Phys. Rev. Lett.* **96** 231802 (2006); "The twin higgs: natural electroweak breaking from mirror symmetry", hep-ph/0506256
257. Barbieri R, Gregoire T, Hall L J "Mirror world at the Large Hadron Collider", CERN-PH-TH/2005-162, UCB-PTH-05/25, LBNL-58803; hep-ph/0509242
258. Patt B, Wilczek F "Higgs-field portal into hidden sectors", hep-ph/0605188
259. Strassler M J "Possible effects of a hidden valley on supersymmetric phenomenology", hep-ph/0607160

1

The impact of the Sakata model

L.B.OKUN

Institute of Theoretical and Experimental Physics
Moscow, Russia

The evolution of the Sakata model is described on the basis of personal recollections, proceedings of international conferences on high energy physics and some journal articles.

§1. 1956. Sakata at ITEP

Shoichi Sakata was the first foreigner who visited the ITEP theory division. He came in the spring of 1956 and compiled a list of the ITEP theorists – I.Ya. Pomeranchuk, V.B. Berestetsky, A.D. Galanin, A.P. Rudik, B.L. Ioffe, V.V. Sudakov, I.Yu. Kobzarev and myself. Sakata also took a photo of those who were present. (It would be interesting to find this picture in his archives.) I still have the three pages of thin rice paper with the Sakata model which he left with us. They correspond to his paper[1) . These three pages were crucial for all my life in physics.

Sakata[1)] considered 7 mesons ($3\,\pi$, $4\,K$) and 8 baryons ($2\,N$, Λ, $3\,\Sigma$, $2\,\Xi$) known at that time. He postulated that 3 baryons – p, n, Λ – are more fundamental than the other 5 baryons and 7 mesons and demonstrated that these 12 particles could be composed from p, n, Λ and $\bar{p}, \bar{n}, \bar{\Lambda}$. The paper had a philosophical flavor and contained no experimental predictions. In 1956 particle physicists were discussing the $\tau\theta$-puzzle and parity violation (see reference[2)] for further details). Therefore the paper[1)] as well as three accompanying papers of Sakata's students[3)–5)] had no immediate response. (S. Tanaka[4)] discussed $\tau\theta$-parity degeneracy in the Sakata model, Z. Maki[5)] attempted to calculate bound states of baryons and antibaryons, while K. Matumoto[3)] suggested a semi-empirical formula for masses of composite particles.)

§2. 1957. Padua – Venice

In the summer of 1957 I suddenly "reinvented" the Sakata model and realized its beauty and its potential. Then I recalled the three rice pages and reread them.

My first paper[6)] on the Sakata model was presented by I.I. Gurevich at the conference in Padua – Venice, September 1957. A slightly different text[7)] was published in a Russian journal. In these publications the three "sakatons" were not physical p, n, Λ, but some primary particles denoted by the same letters, so "we can assume that for the primary particles $m_\Lambda = m_N$"[7)] . Strong and weak interactions of sakatons were considered and for the latter a number of selection rules were deduced, in particular, those which are known as $|\Delta S| = 1$, $\Delta T = 1/2$ for nonleptonic decays of strange particles via the $\bar{n}\Lambda$ transition, while for the leptonic (or semi-leptonic) ones

$|\Delta S| = 1$, $\Delta Q = \Delta S$ and $\Delta T = 1/2$ via $\bar{p}\Lambda$ current.

As for the strong interactions, the existence of η- and η'-mesons was predicted in[6),7)] ; I denoted them ρ_1^0 and ρ_2^0:

"In the framework of this scheme there is a possibility of two additional neutral mesons which have not so far been observed:

$$\rho_1^0 = \Lambda\bar{\Lambda}\ , \quad \rho_2^0 = (p\bar{p} - n\bar{n})/\sqrt{2}\ .$$

The isotopic spin of the ρ-mesons is zero."[6)]

(The unconventional minus sign in the definition of ρ_2^0 was in accord with the not less unconventional definition $\pi^0 = (p\bar{p} + n\bar{n})/\sqrt{2}$.)

§3. 1957. Stanford and Berkeley

In December 1957 four Soviet particle physicists (D.I. Blokhintsev, V.P. Dzhelepov, S.Ya. Nikitin and myself) visited Palo Alto, Berkeley, Boston, New York, Brookhaven. For me it was my first trip abroad and the first flight in my life. (The next time the Soviet authorities allowed me to visit the USA was only in 1988 for the Neutrino'88 conference.)

During the 1957 trip I talked with M. Baker, H. Bethe, S. Drell, R. Feynman, R. Gatto, M. Gell-Mann, S. Goldhaber, C. Sommerfield, F. Zachariasen, C. Zemach and many others, gave a seminar at Berkeley. As a result of this seminar E. Segre invited me to write an article for the Annual Review of Nuclear Science. It appeared in 1959 (see below).

§4. 1958. Geneva

My second paper on the Sakata model "Mass reversal and compound model of elementary particles" was published in June 1958 as a Dubna preprint[8)] and I had it with me at the 1958 Rochester Conference at CERN. On the initiative of J.R. Oppenheimer and R.E. Marshak a special seminar was arranged at which I presented my paper at the start of the conference and then was asked to present it also at Session 7, "Special theoretical topics", see[9)] . (Note that selection rules for weak interactions in sections 14, 15 and references 24-28 of the Dubna preprint[8)] were deleted by the editors of the Proceedings[9)] ; see the Appendix for the deleted pages.)

In[8),9)] ρ_1^0 and ρ_2^0 became mixtures of the states discussed above. What is more important, all interactions were assumed to be γ_5-invariant following papers[10)-12)] and especially[13)] . The conservation of the vector non-strange current, postulated in[10),14)] , was shown in[8),9)] to be inevitable in the Sakata model. Unfortunately the strong interaction was written as an ugly four-fermion interaction of sakatons.

The discussion of my talk involved R. Gatto, G. Lüders, R. Adair, G. Wentzel, T.D. Lee, Y. Yamaguchi (see page 228 of the Proceedings). The discussion with Yoshio Yamaguchi continued during the lunch in the CERN canteen. In the afternoon of the same day J. Oppenheimer commented my argument that in the Sakata model conservation of the weak non-strange vector current is inevitable (see page 257).

He again at length commented the subject in his "Concluding Remarks" at the Conference (see page 293). R. Marshak stressed the novelty of chiral invariance for strong interactions (see page 257). In his talk "K_{e3} and $K_{\mu 3}$ decays and related subjects" Marshak repeatedly underlined that for these decays "$\Delta I = 1/2$ in Okun's model" (see[15] , pp. 284, 285).

In the discussion[16] I described an upper limit on $\Delta S = 2$ transitions which had been derived by B. Pontecorvo and myself[17] .

On the basis of the selection rules for weak interactions which follow from the Sakata model the lifetime of K_2^0 and its branching ratios were predicted[18] by I.Yu. Kobzarev and myself. This prediction was cited by me in December 1957 at Stanford and as reference [28] in the Dubna preprint[8] and was soon confirmed experimentally[19] .

§5. 1959. Kiev symmetry

In 1959 my paper[20] appeared as well as its Russian twin[21] . I received a hundred requests for reprints, many of them – from Japan. Strangely enough, rereading now this paper, I do not see in it the prediction of η and η' and any statement that p, n, Λ are not physical baryons, but some more fundamental particles. Both the prediction and the statement were in[6)–8)] . I cannot understand now their irrational omission in[20),21)] .

In 1959 other authors started to publish papers on the Sakata model. A. Gamba, R. Marshak and S. Okubo[22] pointed out the symmetry between the three leptons (μ, e, ν) and three baryons (Λ, n, p) "in models of Sakata[1] and Okun[7]"[*]. This symmetry has been emphasized by Marshak (in his rapporteur talk[23] at the 1959 Rochester conference in Kiev) and became known as the Kiev symmetry. (I served as a scientific secretary of R. Marshak and participated in preparation of his report.)

At the Kiev conference M. Gell-Mann told me: "If I were you, I would introduce in the Λnp model the linear superposition $(n \cos\theta + \Lambda \sin\theta)$". I do not understand why I did not follow his advice. The angle θ is known now as the Cabibbo angle. The weak current $\bar{p}(n + \varepsilon\Lambda)/(1 + \varepsilon^2)^{1/2}$ first appeared next year in the paper by M. Gell-Mann and M. Levy[24] .

In 1959 the symmetry which is now called SU(3) was introduced into Sakata model. Y. Yamaguchi[25] with reference to[9] stressed the existence of 9 pseudoscalar mesons $(9 = \bar{3} \times 3)$. O. Klein[26] and S. Ogawa[27] discussed generalizations of isotopic symmetry. In particular, S. Ogawa with reference to[25] considered 3 doublets (pn), $(n\Lambda)$, (Λp) and 3 meson triplets. M. Ikeda, S. Ogawa, Y. Ohnuki[28] with reference to[27] developed some mathematical constructs of the symmetry to which they referred as U(3). O. Klein[26] discussed the interaction between the triplet of sakatons and the octet of pseudoscalar mesons and stressed the symmetry between Λnp and $\mu e \nu$.

---

[*] Here and in other quotations the reference numbers correspond to my list of references.

§6. 1960. Rochester

In 1960 I was invited to give a rapporteur talk at the Rochester Conference in Rochester. I prepared the draft of the talk, but was not allowed by Soviet authorities to attend the conference. My draft[29] based on the Sakata model has been prepared for the Proceedings by S. Weinberg. M.L. Goldberger who "was thrown into a breach at a rather late date" served as a rapporteur on "Weak interactions (theoretical)"[30] referred to my draft. R. Feynman[31] spoke on the conserved vector current. He said that in the model of Fermi and Yang "as has been pointed out in much more detail by Okun, in any complex structure, the coupling of the beta decay is proportional to the total isotopic spin". M. Gell-Mann[32] spoke on the conserved and partially conserved currents . He said: "...there is the scheme mentioned by Feynman and favored by Okun, Marshak, and others, based on just n, p, and Λ. Of course, if that is right we do not need the elaborate machinery I just described. We simply draw an analogy". But as it is clear from their talks both Feynman and Gell-Mann at that time preferred to use the composite model only as a tool to formulate more general phenomenological approaches. Among the talks at Rochester 1960 was that by Y. Ohnuki[33] who with a reference to[7] assumed $m_\Lambda = m_N$ and the three-dimensional unitary symmetry.

An important paper of 1960 was that by J. Sakurai[34] . With a reference to[9] he mentioned that instead of N, Λ one can use as "elementary" Ξ, Λ. He considered the absence of η-meson as a serious problem: "...within the framework of Fermi–Yang–Sakata–Okun model it may be difficult to explain why the η does not exist" (see pp. 32-36).

The η-meson was discovered within a year[35] .

Further progress in SU(3) symmetric Sakata model was achieved by M. Ikeda, Y. Miyachi, S. Ogawa,[39] who applied this symmetry to weak decays. Z. Maki, M. Nakagava, Y. Ohnuki, S. Sakata published a paper on Sakata model[40] . They wrote: "... it has recently become clear that Feynman–Gell-Mann current derived from the Sakata model is quite sufficient to account for the experimental facts concerning the weak processes[7],[41]". They postulated the existence of a so-called B^+ matter. The bound state eB^+ had been identified with n, bound state μB^+ – with Λ, while νB^+ – with p.

In 1960–61 I was giving lectures[36],[37] based on the Sakata model. Subsequently they were recast into the book[38] . My major mistake at that time was that I did not consider seriously eight spin 1/2 baryons as an SU(3) octet in spite of the "eightfold way" papers by M. Gell-Mann[42],[47] and Y. Ne'eman[43] . (The former referred to papers[20],[25],[28] .)

§7. 1962. Geneva again

In 1962 the Sakata model was "falsified" for a short time by experiments,[44],[45] which discovered decays $\Sigma^+ \to n\mu^+\nu$ and $K^0 \to e^+\nu\pi^-$ forbidden by $\Delta S = \Delta Q$ rule. At the 1962 Geneva conference I tried to find a mistake in the results[44],[45] but failed. Pomeranchuk who witnessed the argument commented later that my "feathers were flying". I do not remember now how the mistake was found subsequently by experimentalists. Maybe it was a statistical fluctuation.

The authors of articles[44),45)] referred to the paper by Feynman and Gell-Mann[10)] . While in my papers[6),7)] the forbidden decays were simply listed, in[10)] the notations ΔQ and ΔS were used and the currents with $\Delta Q = \Delta S$ and $\Delta Q = -\Delta S$ ($\bar{p}\Lambda$ and $\bar{n}\Sigma^{+}$-currents) were phenomenologically considered on the same footing. The product of these currents gives transitions with $\Delta S = 2$. The limit on these transitions from the absence of decays $\Xi^{-} \to n\pi^{-}$ was not reliable because "so few Ξ particles have been seen that this is not really conclusive"[10)] . (The paper[17)] (published in June 1957) had put a much better limit on $\Delta S = 2$ processes from $K^0 \leftrightarrow \bar{K}^0$ transitions. But it was not known to Feynman and Gell-Mann when they wrote[10)] .)

In 1962 M. Gell-Mann predicted the existence of Ω-hyperon[46)] . I. Kobzarev and myself[48)] derived the SU(3) relations between semileptonic decays of π and K-mesons. Together with relations for the decays of baryons they were later derived by N. Cabibbo.[49)]

§8. 1962. From 3 to 4 sakatons

The discovery of ν_μ prompted attempts to reconcile the existence of two neutrinos with the lepton-sakaton symmetry. In order to preserve the Kiev symmetry Z. Maki, M. Makagawa, S. Sakata[50)] modified the B^+ matter model[40)] . They assumed that $p = \nu_1, B^+$, where ν_1 is one of the two orthogonal superpositions of ν_e and ν_μ. The other superposition ν_2 was assumed either not to form at all a bound state with B^+ or to form a baryon with a very large mass. On the basis of this model the paper introduced $\nu_e - \nu_\mu$ oscillations. Another way to lepton-sakaton symmetry was suggested in the paper by Y. Katayama, K. Motumoto, S. Tanaka, E. Yamada[51)] , where the fourth sakaton was explicitly introduced.

§9. 1964. Quarks

In 1964 η'-meson and Ω-hyperon were discovered[53),54)] . Earlier this year G. Zweig[55)] and M. Gell-Mann[56)] replaced the integer charged sakatons by fractionally charged particles (aces – Zweig; quarks – Gell-Mann). This allowed them to construct not only the octet and singlet of mesons, but also the octet and decuplet of baryons. When establishing the electromagnetic and weak currents in the quark model M. Gell-Mann[56)] referred to similar expressions in the Sakata model.

§10. November 2006 and afterwards

On November 3 2006 I received the following email from a colleague and a friend of mine – Valentine I. Zakharov:

"Dear Lev Borisovich,

I am now visiting Kanazawa, Japan. This month, there will be a one-day conference in Nagoya, to celebrate 50 years of the Sakata model. They invited me to come and I eagerly agreed.

One of the reasons – which you can readily guess– was that the words 'Sakata model' were among the first ones I heard about our field. (You were giving lectures to 'experimentalists', with Alikhanov in the first row; (M.I. Ryazanov from MEPHI encouraged me to attend; it was some time before I showed up later).

I will mention of course that you were developing the Sakata model at ITEP. But, unfortunately, I realized that I do not know anything else, to any extent personal about Sakata-sensei. I mean, no other papers, or their echo in Russia/USSR, nothing ... May be you can help in some way?

Excuse me, please, for bothering you and with best regards, Valya".

To answer Valentine's request I have written this brief review. Thinking that it might be of more general interest, I published it as version 1 of hep-ph/0611298.

On December 22 I received an email from K. Yamawaki who kindly invited me to publish this paper in the Proceedings of the Sakata Model Symposium. In editing the paper I benefited from email exchanges with S. Pakvasa, H. Lipkin and A. Gal. Another interpretation of the terms "Sakata model" and "Sakata symmetry" one can find in preprints[57),58)] .

Appendix

Four pages from the Dubna preprint[8)]:

– 19 –

where $\mathcal{L}^{s}_{int}$ is the Lagrangian of strong interaction in the form (10). Taking advantage of (22), we can easily prove that for the nucleon current of β -decay interaction $j^{v}_{\mu}=G^{1}_{V}\bar{\chi}\tau\gamma_{\mu}\chi$ the following relation is fulfilled:

$$\frac{\partial j^{v}_{\mu}}{\partial x_{\mu}}=0 \qquad (23)$$

It was shown in /21,23/ that relation (23) results in the conclusion that the value G^{1}_{V}, like an electrical charge, does not change when corrections connected with the strong interaction are taken into account. A similar proof can be given for G^{2}_{V}. The above-obtained result is true only in an approximation which makes no allowance for virtual slow and electromagnetic processes.

Unlike the leptonic interaction of nucleons, the leptonic interaction of Λ -hyperons (expressions (17) and (18)), as can easily be seen, does not possess the property of non--renormalizability of the vector coupling constant.

14. From expressions (17) and (18) it follows that if the strangeness of strongly interacting particles changes in leptonic decays of strange particles (as is clear from the foregoing, we attribute no strangeness to leptons), this change can only be $\Delta S = \pm 1$. Here $\Delta S = -1$ corresponds to emission of positively charged leptons, while $\Delta S = +1$ corresponds to the emission of negatively charged leptons.

Now let us enumerate the decays allowed by (15), (16), (17) and (18) and those forbidden by these interactions:

Allowed:

$$\Lambda \to p + e^-(\mu^-) + \tilde{\nu}\,, \quad \Sigma^- \to n + e^-(\mu^-) + \tilde{\nu}\,, \quad \Sigma^- \to \Lambda + e^- + \tilde{\nu}$$
$$\Sigma^+ \to \Lambda + e^+ + \nu\,, \quad \Xi^- \to \Lambda + e^-(\mu^-) + \tilde{\nu}\,, \quad \Xi^- \to \Sigma^0 + e^-(\mu^-) + \tilde{\nu}$$
$$\Xi^0 \to \Sigma^+ + e^-(\mu^-) + \tilde{\nu}\,, \quad K^0 \to \pi^- + e^+(\mu^+) + \nu\,, \quad K^0 \to \pi^+ + e^-(\mu^-) + \tilde{\nu}$$

Forbidden:

$$\Sigma^+ \to n + e^+(\mu^+) + \nu \quad \Xi^- \to n + e^-(\mu^-) + \tilde{\nu}\,, \quad \Xi^0 \to p + e^-(\mu^-) + \tilde{\nu}$$
$$\Xi^0 \to \Sigma^- + e^+(\mu^+) + \nu\,, \quad K^0 \to \pi^+ + e^-(\mu^-) + \tilde{\nu}\,, \quad \bar{K}^0 \to \pi^- + e^+(\mu^+) + \nu$$

We do not claim this list to be exhaustive. The forbiddances arising for K^0_{e3} and $K^0_{\mu 3}$-decays result in a known relationship between the number of $e^+(\mu^+)$-decays and the number of $e^-(\mu^-)$-decays in a beam of neutral K-mesons (see /24, 25, 26, 16/ on this question):

$$\frac{n_{\mu^+}}{n_{\mu^-}} = \frac{n_{e^+}}{n_{e^-}} = \frac{e^{-\frac{t}{\tau_1}} + e^{-\frac{t}{\tau_2}} + 2e^{-\frac{t}{2\tau_1} - \frac{t}{2\tau_2}} \cos \Delta m t}{e^{-\frac{t}{\tau_1}} + e^{-\frac{t}{\tau_2}} - 2e^{-\frac{t}{2\tau_1} - \frac{t}{2\tau_2}} \cos \Delta m t} \qquad (24)$$

Here n is the number of corresponding decays per unit time, τ_1 and τ_2 are the lifetimes of K_1^0 and K_2^0-mesons, the first of which has a combined parity of +1 and the second of −1. Δm is the difference of masses of K_1^0 and K_2^0-mesons.

From (17) and (18) it follows that in leptonic decays of strange particles involving strangeness changes the isotopic spin of strongly interacting particles changes by $\Delta T = 1/2$ /27, 26, 16/. This makes it possible to relate the probabilities of decay of K and K-mesons /26, 16/:

$$w(K^0 \to e^+(\mu^+) + \nu + \pi^-) = 2\,w(K^+ \to e^+(\mu^+) + \nu + \pi^0) \qquad (25)$$

which in its turn results in the relationships /28/:

$$w(K_2^\circ \to e^+(\mu^+) + V + \pi^-) = w(K_2^\circ \to e^-(\mu^-) + \tilde{V} + \pi^+) = w(K^+ \to e^+(\mu^+) + V + \pi^\circ) \quad (26)$$

In combination with the rule $\Delta T = 1/2$ for non-leptonic decays relation (26) makes it possible to relate the lifetime of the K_2° with the lifetime of the K^+-meson. Taking advantage of data on the lifetime of the K^+-meson and on the frequency of various K^+-meson decays we can estimate the lifetime of the K_2°-meson /23/ 28:

$$\tau_{K_2^\circ} \sim 4 \times 10^{-8} \text{ sec.}$$

15. It can easily be seen that interaction (19), which is responsible for the non-leptonic decays of strange particles, allows a change of strangeness of $|\Delta S| = 1$ and forbids a change of strangeness of $|\Delta S| \geq 2$. Non-leptonic decays with $\Delta T > 3/2$ are forbidden by interaction (19). It is not clear, however, how the selection rule $\Delta T = 1/2$ can be obtained from (19), if this interaction is considered as an elementary one.

16. The author wishes to acknowledge the interest taken in his work, useful discussions and critical remarks by I.Ya.Pomeranchuk, Ya.B.Zeldovich, V.B.Berestetsky and A.P.Rudik. The author is also greatly indebted to Dr. R.Gatto for an interesting discussion.

References

1. J.Tiomno. Nuovo Cimento 1, 226)(1955)
2. S.Watanabe. Progr.Theor.Phys. 15, 81 (1956)
3. S.Watanabe. Phys.Rev. 106, 1306 (1957)
4. R.Marshak, E.Sudarshan. Phys.Rev. 109, 1860 (1958)
5. J.J.Sakurai. Nuovo Cimento 7, 649 (1958)
6. W.Pauli. Nuovo Cimento, 6, 204 (1957)
7. W.Heisenberg, W.Pauli (preprint)
8. M.Gell-Mann. Suppl.Nuovo Cim. 4, N 2, 848 (1956)
9. E.Fermi, C.N.Yang. Phys.Rev. 76, 1739 (1949)
10. M.A.Markov. Doklady Ac.Sci.101, 54 (1955)
11. S.Hori, A.Wakasa.Nuovo Cimento 6, 304 (1957)
12. S.Sakata. Progr.Theor.Phys. 16, 686 (1956)
13. S.Tanaka. Progr.Theor.Phys. 16, 625, 631 (1956)
14. Z.Maki. Progr.Theor.Phys. 16, 667 (1956)
15. R.W.King, D.C.Peaslee. Phys.Rev.106,360 (1957)
16. L.B.Okun. Zhurn.Exp.Teor.Fiz. 34, 469 (1958)
17. H.Fierz. Zeits.f.Phys. 104, 553 (1937)
18. A.Baldin. Zhurn.Exp.Teor.Fiz. (in print).
19. T.D.Lee, C.N.Yang. Nuovo Cim. 3, 749 (1956)
20. M.Gell-Mann. Phys.Rev. 106, 1296 (1956)
21. M.Gell-Mann, R.Feynman. Phys.Rev.109, 193 (1958)
22. S.S.Gerstein,Ya.B.Zeldovich,Zhurn.Exp.TeorFiz.29,698(1955)
23. B.L.Ioffe. Zhurn.Exp.Teor.Fiz. (in print)
24. Ya.B.Zeldovich. Zhurn.Exp.Teor.Fiz. 30, 1168 (1956)
25. S.B.Treiman, R.G.Sachs. Phys.Rev. 103, 1545 (1956)
26. L.B.Okun. Zhurn.Exp.Teor.Fiz. 32, 400 (1957)
27. M.Gell-Mann. Proc.Rochester Conference, 1956
28. I.Yu.Kobzarev, L.B.Okun. Zhurn. Exp.Teor.Fiz.34,764(1958)

References

1) S. Sakata. On a composite model for new particles. Progr. Theor. Phys. **16** (1956) 686.
2) L.B. Okun. Mirror particles and mirror matter: 50 years of speculation and search. hep-ph/0606202, v.2.
3) K. Matumoto. Some consequences of the compound hypothesis of elementary particles. Progr. Theor. Phys. **16** (1956) 583.
4) S. Tanaka. The composite model for new unstable particles, I. Progr. Theor. Phys. **16** (1956) 625.
5) Z. Maki. On a theory of a composite model of elementary particles. Progr. Theor. Phys. **16** (1956) 667.
6) L.B. Okun. Some remarks concerning the compound model of fundamental particles. Proc. of the Intern. Conf. on mesons and recently discovered particles, Padova - Venezia, 22 - 28 Settembre 1957, p. V-55.
7) L.B. Okun. Some remarks on a compound model of elementary particles. Sov. Phys. JETP **7** (1958) 322; ZhETF **34** (1958) 469 (in Russian).
8) L.B. Okun. Mass reversal and compound model of elementary particles. Dubna preprint P-203 (1958).
9) L.B. Okun. Mass reversal and compound model of elementary particles. Proc. of 1958 Annual international conference on high energy physics at CERN, 30th June-5th July, 1958, p. 223.
10) R.P. Feynman, M. Gell-Mann. Theory of the Fermi interaction. Phys. Rev. **109** (1958) 193.
11) E.C.G. Sudarshan, R.E. Marshak. Chirality invariance and the universal Fermi interaction. Phys. Rev. **109** (1958) 1860.
12) E.C.G. Sudarshan, R.E. Marshak. The nature of the four fermion interaction. Proc. of the Intern. Conf. on mesons and recently discovered particles. Padova - Venezia, 22-28 Settembre 1957, p. V-14.
13) J.J. Sakurai. Mass reversal and weak interactions. Nuovo Cimento **7** (1958) 649.
14) S.S. Gershtein, Ya.B. Zeldovich. On meson corrections in the theory of β-decay. JETP **2** (1956) 576; ZhETF **29** (1955) 698 (in Russian).
15) R.E. Marshak. K_e3 and $K_\mu3$ decays and related subjects. Proc. of 1958 Annual international conference on high energy physics at CERN, 30th June-5th July, 1958, p. 284.
16) L. Okun. Proc. of 1958 Annual international conference on high energy physics at CERN, 30th June-5th July, 1958, p. 201.
17) L. Okun, B. Pontecorvo. Some remarks on slow processes of transition of elementary particles. Sov. Phys. JETP **5** (1957) 1297; ZhETF **32** (1957) 1587 (in Russian).
18) I.Yu. Kobzarev, L.B. Okun. Lifetime of the K_2 meson. Sov. Phys. JETP **7** (1958) 524; ZhETF **34** (1958) 763 (in Russian).
19) M. Bardon, M. Fuchs, K. Lande, L.M. Lederman, W. Chinowsky, J. Tinlot. Lifetime and decay of the K_2^0 meson. Phys. Rev. **110** (1958) 780.
20) L. Okun. Strange particles: decays. Annual Review of Nuclear Science **9** (1959) 61
21) L. Okun. Strange particles: decays. Uspekhi Fiz. Nauk **68** (1959) 449 (in Russian).
22) A. Gamba, R. Marshak and S. Okubo. On a symmetry in weak interactions. Proc. Nat. Ac. Sci. **45** (1959) 881.
23) R. Marshak. Theoretical status of weak interactions. Ninth International Annual Conference on High Energy Physics, v.2, Moscow, 1960, p. 269.
24) M. Gell-Mann, M. Levy. The axial vector current in beta decay. Nuovo Cimento **16** (1960) 705.
25) Y. Yamaguchi. A composite theory of elementary particles. Supplement to the Progr. Theor. Phys. **11** (1959) 1.
26) O. Klein. On the systematics of elementary particles. Archiv för Fysik Swed. Acad. Sci. **16** (1959) 191.
27) S. Ogawa. A possible symmetry in Sakata's composite model. Progr. Theor. Phys. **21** (1959) 209.
28) M. Ikeda, S. Ogawa, Y. Ohnuki. A possible symmetry in Sakata's model for bosons - baryons system. Progr. Theor. Phys. **22** (1959) 715.
29) L.B. Okun. Certain problems of weak interaction theory. Proceedings of Rochester Conference, August 25 - Sept. 1, 1960, p. 743.

30) M.L. Goldberger. Weak interactions (theoretical). Proceedings of Rochester Conference, August 25 - Sept. 1, 1960, p. 732.
31) R.P. Feynman. The status of the conserved vector current hypothesis. Proceedings of Rochester Conference, August 25 - Sept. 1, 1960, p, 502.
32) M. Gell-Mann. Conserved and partially conserved currents in the theory of weak interactions. Proceedings of Rochester Conference, August 25 - Sept. 1, 1960, p. 508.
33) Y. Ohnuki. Composite model of elementary particles. Proceedings of Rochester Conference, August 25 - Sept. 1, 1960, p. 843.
34) J.J. Sakurai, Theory of strong interactions. Annals of Physics **11** (1960) 1.
35) A. Pevsner et al. Evidence for three pion resonance near 550 MeV. Phys. Rev. Lett. **7** (1961) 421.
36) L.B. Okun. Lectures on the theory of weak interactions of elementary particles. 17 ITEP preprints, 1960-1961 (in Russian). English translation: Theory of weak interactions: Thirteen lectures, AEC-tr-5226. US Atomic Energy Commission. Oak Ridge, Tenn. (For the translation of lectures 14-16 and of contents see NP-10254, 10842, 10845, 10840.)
37) L.B. Okun. Lectures on the theory of weak interactions of elementary particles. JINR preprint P-833 (1961) (in Russian).
38) L.B. Okun. Slaboe vzaimodeistvie elementarnykh chastits. Moskva, Fizmatgiz, 1963 (in Russian). English translation: Weak interaction of elementary particles. Pergamon Press, 1965.
39) M. Ikeda, Y. Miyachi, S. Ogawa. Symmetry in Sakata's model and weak interactions. I. Progr. Theor. Phys. **24** (1960) 569.
40) Z. Maki, M. Nakagava, Y. Ohnuki, S. Sakata. A unified model for elementary particles. Progr. Theor. Phys. **23** (1960) 1174.
41) S.Okubo, R.E. Marshak, E.C.G. Sudarshan. V-A theory and decay of Λ hyperon. Phys. Rev. **113** (1959) 944.
42) M. Gell-Mann. The eightfold way: a theory of strong interaction symmetry. Report CTSL-20 (1961).
43) Y. Ne'eman. Derivation of strong interactions from a gauge invariance. Nucl. Phys. **26** (1961) 222.
44) R. Ely et al. Experimental test of the selection rule $\Delta S = \Delta Q$. Phys. Rev. Lett. **8** (1962) 132.
45) G. Alexander et al. Experimental test of the $\Delta I = 1/2$ rule, and $\Delta S = +\Delta Q$ rule in three-body decays of neutral K mesons. Phys. Rev. Lett. **9** (1962) 69.
46) M. Gell-Mann. Strange particle physics. Strong interactions. Proc. Intern. Conf. High Energy Phys. (CERN, 1962), p. 805.
47) M. Gell-Mann. Symmetries of baryons and mesons. Phys. Rev. **125** (1962) 1067.
48) I.Yu. Kobzarev, L.B. Okun. Unitary symmetry and universal weak interaction. JETP **15** (1962) 970; ZhETF **42** (1962) 1400 (in Russian).
49) N. Cabibbo. Unitary symmetry and leptonic decays. Phys. Rev. Lett. **10** (1963) 531.
50) Z. Maki, M. Nakagava, S. Sakata. Remarks on the unified model of elementary particles. Progr. Theor. Phys. **28** (1962) 870.
51) Y. Katayama, K. Matumoto, S. Tanaka, E. Yamada. Possible unified models of elementary particles with two neutrinos. Progr. Theor. Phys. **28** (1962) 675.
52) G. Danby, J.M. Gaillard, K. Goulianos, L.M. Lederman, N.B. Mistry, M. Schwartz, J. Steinberger. High energy neutrino interaction in matter. Proc. of ICHEP 1962 at CERN, p. 809.
53) P.M. Dauber et al. Properties of the 960-MeV boson. Phys. Rev. Lett. **13** (1964) 449.
54) V.E. Barnes et al. Observation of a hyperon with strangeness minus three. Phys. Rev. Lett. **12** (1964) 204.
55) G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Preprints CERN-TH-401, 412.
56) M. Gell-Mann. A schematic model of baryons and mesons. Phys. Lett. **8** (1964) 214.
57) H.J. Lipkin. From Sakata model to Goldberg–Ne'eman quarks and Nambu QCD phenomenology and "right" and "wrong" experiments. hep-ph/0701032 and a revised version to appear in this Proceedings.
58) A. Gal. The hypernuclear physics heritage of Dick Dalitz (1925–2006). hep-ph/0701019.

THE EVOLUTION OF THE CONCEPTS OF ENERGY, MOMENTUM, AND MASS FROM NEWTON AND LOMONOSOV TO EINSTEIN AND FEYNMAN

L.B. Okun [a]

ITEP, 117218 Moscow, Russia

*Abstract.*The talk stresses the importance of the concept of rest energy E_0 and explains how to use it in various situations.

1 Introduction

This conference is the first in a series of conferences celebrating 300 years since the birth of Mikhail Lomonosov (1711–1765).

The law of conservation of mass established in chemistry by Lomonosov and Lavoisier and seriously modified in relativistic physics two centuries later is central for understanding and teaching physics today. Therefore it is appropriate to consider the evolution of the laws of conservation of mass, energy, and momentum during this period.

The main message of the talk is the equivalence of the rest energy of a body and its mass: $E_0 = mc^2$. This equivalence is a corollary of relativity principle. The total energy of a body and its mass are not equivalent: $E \neq mc^2$.

The contents of the talk is as follows:

1. Introduction
2. XVII – XIX centuries
    - 3.1. Galileo, Newton: relativity
    - 3.2. Lomonosov, Lavoisier: conservation of mass
    - 3.3. Conservation of energy
3. The first part of the XXth century
    - 4.1. Rest energy E_0
    - 4.2. Energy and inertia
    - 4.3. Energy and gravity
    - 4.4. "Relativistic mass" *vs* mass
    - 4.5. Famous *vs* true
    - 4.6. Einstein supports $E_0 = mc^2$
4. The second part of the XXth century

[a] e-mail: okun@itep.ru

2 XVII - XIX centuries

2.1 Galileo, Newton: relativity

The concept of relativity was beautifully described by Galileo Galilei in his famous book "Dialogo" (1632) as experiments in a cabin of a ship.

The principle of relativity had been first formulated by Isaac Newton in his even more famous book "Principia" (1687), though not as a principle, but as corollary v.

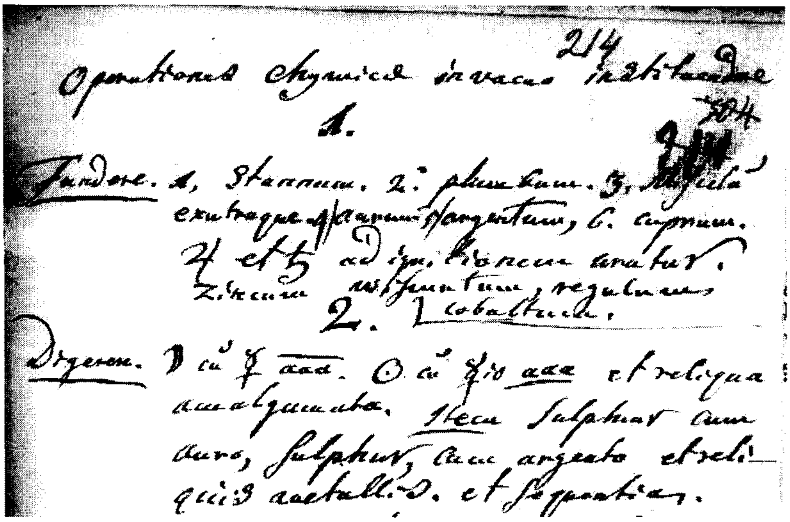
The term mass was introduced into physics by Newton in "Principia".

According to Newton, the mass is proportional to density and volume. The momentum is proportional to mass and velocity.

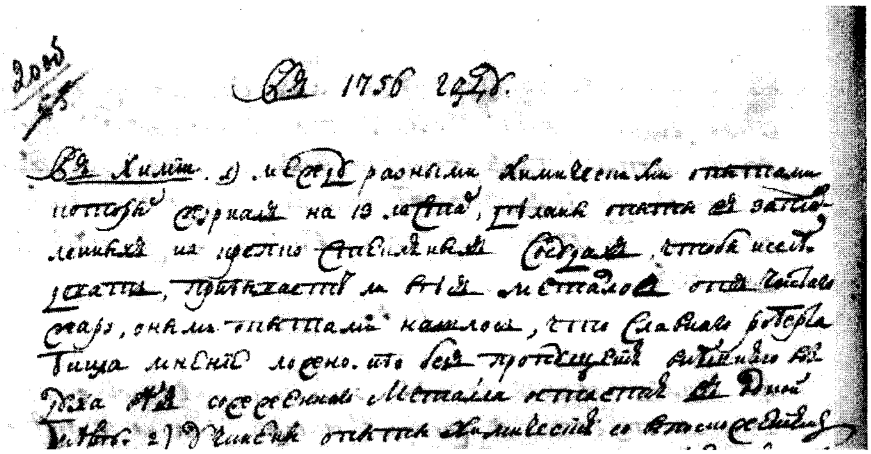
As for the term energy, Newton did not use it. He and Gottfried Leibniz called the kinetic energy vis viva – the living force.

2.2 Lomonosov, Lavoisier

In 1756 Lomonosov experimentally proved his earlier conjecture (formulated in his letter to Leonard Euler in 1748) that mass is conserved.



Lomonosov's handwriting in Latin: ignition of tin (jupiter) and lead (saturnus) in sealed retorts.



The 1756 report on Lomonosov's experiments which disproved the results of Robert Boyle on ignition of metals. (Written in Russian by a clerk.)

"... made experiments in firmly sealed glass vessels in order to investigate whether the weight of metals increases from pure heat. It was found by these experiments that the opinion of the famous Robert Boyle is false, for without letting in the external air the weight of the ignited metal remains in the same measure..."

In 1773 Antoine Lavoisier independently proved the law of conservation of mass in a series of more refined experiments.

2.3 Conservation of energy

The term energy was introduced into physics in 1807 by Thomas Young.

By the middle of the XIXth century a number of scientists and engineers, especially J.R. von Mayer and J.P. Joule, established the law of conservation of energy which included heat among the other forms of energy.

3 The first part of XX century

3.1 Rest energy E_0

The special theory of relativity was created by Hendrik Lorentz, Henri Poincaré, Albert Einstein, and Herman Minkowski.

The concept of rest energy was introduced into physics by Einstein.

In 1905 Einstein proved in the framework of special relativity that the change of the rest energy of a body is equivalent to the change of its mass.

In 1922 and especially clearly in 1935 he formulated the equivalence of mass m and rest energy E_0 – the equation $E_0 = mc^2$.

3.2 Energy and inertia

In relativity the energy E and momentum $\vec{p}$ of a body form the energy-momentum vector p_i $(i = 0, 1, 2, 3 = 0, a)$.

In the units in which $c = 1$: $p_0 = E, p_a = \vec{p}$.

The mass is a Lorentz scalar defined by the square of p_i: $m^2 = p^2 = E^2 - \vec{p}^2$.

To keep track of powers of c let us define $p_0 = E, p_a = c\vec{p}$.

Then $p^2 = E^2 - c^2\vec{p}^2 = m^2c^4$.

In Newtonian physics mass is the measure of inertia according to equations: $\vec{p} = m\vec{v}$, $\vec{F} = d\vec{p}/dt$, $\vec{F} = m\vec{a}$, where $\vec{a} = d\vec{v}/dt$.

In relativity the energy is the measure of inertia: $\vec{p} = E\vec{v}/c^2$.

If the force is defined by equation $\vec{F} = d\vec{p}/dt$, then

$$\vec{F} = m\gamma\vec{a} + m\gamma^3\vec{v}(\vec{v}\vec{a}) = m_t\vec{a} + m_l\vec{v}(\vec{v}\vec{a}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$.

In the first years of the XXth century Hendrik Lorentz who tried to define inertial mass in terms of force and acceleration ended up with the concepts of longitudinal and transverse masses : $m_l = m\gamma^3$, $m_t = m\gamma$ which later were forgotten.

3.3 Energy and gravity

In Newtonian physics the source of gravity is mass. In relativity the source of gravity is the energy-momentum tensor p_ip_k/E which serves as the "gravitational charge".

With the help of propagator of the gravitational field proportional to $g^{il}g^{km}+g^{im}g^{kl}-g^{ik}g^{lm}$, where g^{ik} is the metric tensor, the energy-momentum tensor can be reduced in a static gravitational field (when $l, m = 0$) to $(2E^2 - m^2c^4)/E$.

For a massive non-relativistic apple this expression is equal to mc^2, while for a photon it is equal to $2E$. Note the factor of 2. The energy of a photon is attracted stronger than the energy of an apple.

3.4 "Relativistic mass" vs mass

The prerelativistic commandments:

1. mass must be the measure of inertia,
2. mass must be additive.

They led to the introduction of the so-called "relativistic mass" $m = E/c^2$ which for a massive particle increases with the velocity of the particle.

The idea that mass of an electron increases with its velocity had been put forward by J.J Thomson, O. Heaviside, and G. Searle in the last decade of the XIXth century, (not so long) before relativity theory was formulated.

The idea that light with energy E has mass $m = E/c^2$ was formulated by Poincaré in 1900 and was discussed by Einstein in the first decade of the XXth century. The relativistic mass increasing with velocity was proclaimed "the mass" by G. Lewis and R. Tolman at the end of that decade. A decade later it was enthroned in books on relativity by Max Born and Wolfgang Pauli.

3.5 Famous vs true

Thus the equation $E = mc^2$ appeared and was ascribed to Einstein.

This "adopted child" is widely considered as "the famous Einstein's equation" instead of the true Einstein's equation $E_0 = mc^2$. Einstein seemed to be indifferent to this misuse.

3.6 Einstein supports $E_0 = mc^2$

In 1922 in his book "The Meaning of Relativity" Einstein formulated the equation $E_0 = mc^2$. In December 1934 Einstein delivered his Gibbs Lecture "Elementary derivation of the equivalence of mass and energy" at a joint meeting of the American Mathematical Society and the American Physical Society.

In that lecture he repeatedly stressed that mass m (with the usual time unit, mc^2) is equal to rest energy E_0.

This however did not prevent Einstein's coauthor – Leopold Infeld[b] from stating in 1955 that the main experimental confirmation of the special relativity is the dependence of mass on velocity.

4 The second half of XX century

4.1 Landau and Lifshitz

The first monograph in which special and general relativity were presented without using the notion of mass increasing with velocity was the first (1941) edition of "Field Theory" by Lev Landau and Evgeniy Lifshitz.

They wrote (in the first and the second editions in §9,§10 – in the later editions they became §8,§9) the expressions for action S, momentum $\vec{p}$, energy E and rest energy. Unfortunately for the latter they chose the same symbol E and did not introduce E_0. The latest edition still keeps this tradition.

4.2 Feynman diagrams

A major step forward in creating the present understanding of nature were diagrams introduced by Richard Feynman.

The external lines of a diagram correspond to incoming and outgoing, free, real particles. For them $p^2 = m^2$ in units of $c = 1$; they are on mass shell.

The internal lines correspond to virtual particles. For them $p^2 \neq m^2$; they are off mass shell. Energy and momentum are conserved at each vertex of a diagram.

The exchange of a virtual massless particle creates long-range force between real particles.

Thus exchange of a photon creates Coulomb force (potential).

The exchange of a virtual massive particle creates Yukawa potential – short-range force with radius $r = h/mc$.

When using Feynman diagrams, the four-dimensional momenta p and invariant masses m immensely facilitate theoretical analysis of various processes involving elementary particles.

Feynman diagrams unified matter (real particles - both massive and massless) with forces (virtual particles).

The role of Quantum Mechanics is crucial to this unification.

A nice feature of Feynman diagrams is the interpretation of antiparticles as particles moving backward in time.

[b] "A.Einstein, L.Infeld. The Evolution of Physics. 1938."

4.3 Feynman Lectures

The most famous textbook in physics is "The Feynman Lectures on Physics". Several million copies of Lectures introduced millions of students to physics. In his Lectures Feynman masterfully and enthusiastically painted the broad canvas of physics from the modern point of view. Unfortunately in this masterpiece he completely ignored the Feynman diagrams and largely ignored the covariant formulation of the relativity theory.

Lectures are based on the archaic notion of "relativistic mass" that increases with velocity and the relation $E = mc^2$. Thus millions of students were (and are!) taught that the increase of mass with velocity is an experimental fact. They sincerely believe that it is a fact, not a factoid based on a rather arbitrary definition $m = E/c^2$.

5 Conclusions

The giant figure of Newton marked the birth of modern Science. The achievements of Science since the times of Newton are fantastic. The modern views on matter differ drastically from those of Newton.

Still, even in the XXIst century many physics textbooks continue to use (incorrectly) the equations of Newton many orders of magnitude beyond the limits of their applicability, at huge ratios of kinetic energy E_k to rest energy E_0 (10^5 for electrons and 10^4 for protons at CERN), while Newton's equations are valid only for $E_k/E_0 \ll 1$.

If some professors prefer to persist in this practice, they should at least inform their students about the fundamental concept of invariant mass and the true Einstein's equation:

$$E_0 = mc^2 \ .$$

6 Acknowledgments

I am very grateful for their help to A.A. Alehina, B.M. Bolotovsky, K.G. Boreskov, M. Gottlieb, E.G. Gulyaeva, M.V. Danilov, E.A. Ilyina, O.V. Kancheli, V.I. Kisin, V.I. Kogan, M.V. Mandrik, T.S. Nosova, B.L. Okun, E.V. Sandrakova, M.B. Voloshin.

7 Discussion: FAQ about mass

7.1 Natural definition of mass

Q1: Which definition of mass is natural in the framework of the Relativity Theory?

A1: The definition according to which mass is a Lorentz invariant property of an object – the 'length' of the 4-dimensional energy-momentum vector $p = (E, c\vec{p})$. Namely $m^2 = p^2/c^4$ or in other notations $m^2 = E^2/c^4 - \vec{p}^2/c^2$.

This definition corresponds perfectly to the fundamental symmetry of special relativity and uses the minimal number of notions and symbols.

7.2 Unnatural definition of mass

Q2: Can one nevertheless introduce another definition of mass, namely, that which corresponds to the "famous Einstein's equation $E = mc^2$"? (here E is the total energy of a free body)

A2: Yes. One can do this. But this cheese is not free. People who do this refer to the ordinary mass as the "rest mass" (they denote it m_0).

They have two different symbols for energy: E and $E/c^2 = m$. This is confusing. This ignores the 4-dimensional symmetry of relativity theory: E is a component of a 4-vector, while E/c^2 is "the cat that walks by itself" – the time component of a 4-vector the space components of which are never mentioned.

Of course in any consistent theory one can introduce an arbitrary number of redundant variables by multiplying any observable by some power of a fundamental constant, like c.

With proper bookkeeping that would not produce algebraic mistakes. However, instead of creating clarity, this creates confusion.

It is like the well known Jewish joke on inserting the letter 'r' in the word 'haim':

– What for is the letter r in the word 'haim'?
– But there is no r in 'haim'
– And if to insert it?
– But what for to insert it?
– That is what I am asking: what for?

Q3: Doesn't the mass, increasing with the velocity of the body, explain why the velocity of a massive body cannot reach the velocity of light?

A3: No. It does not explain: the increase is not fast enough. This follows from the expression for longitudinal mass $m_l = m_0/(1 - v^2/c^2)^{3/2}$ derived by Lorentz from $\vec{F} = d\vec{p}/dt$.

7.3 Equivalence of mass and rest energy

Q4: Is mass equivalent to energy?

A4: Yes and no. Loosely speaking, mass and energy are equivalent. But the mass m of an object is not equivalent to its total energy E, it is equivalent to its rest energy E_0.

Q5: What is rest energy E_0?

A5: The rest energy E_0 is the greatest discovery of the XXth century.

Einstein discovered that any massive body at rest has a huge hidden energy $E_0 = mc^2$ (the subscript 0 indicates here that the velocity of the body v is equal to zero).

Q5a: How did Einstein discover $E_0 = mc^2$?

A5a: In his second 1905 paper on relativity Einstein considered a body at rest with rest energy E_0, which emits two light waves in opposite directions with the same energy $L/2$.

For an observer that moves with velocity v with respect to the body the total energy of two waves is $L/\sqrt{1 - v^2/c^2}$.

By assuming conservation of energy and by considering the case of $v \ll c$ Einstein derived that $\Delta m = L/c^2$.

In this short note two revolutionary ideas were formulated:

1. that a massive body at rest contains rest energy E_0,
2. that a system of two massless light waves with energy L has mass L/V^2. (Einstein denoted the speed of light by V.)

In his publications of 1922 and 1935 Einstein cast the relation in the form $E_0 = mc^2$.

Q5b: Is it possible to prove Einstein's relation by considering emission of one wave of light instead of two?

A5b: Yes, it is possible. But the proof is slightly more involved. In this case the rest energy of the body partly transforms into kinetic energy of light L and kinetic energy of the recoil body with mass m: $E_k = L^2/2mc^2$.

Q6: Is the relation $E_0 = mc^2$ compatible with the definition of mass given above: $m^2 = E^2/c^4 - \vec{p}^2/c^2$?

A6: Yes. It is absolutely compatible: at $\vec{v} = 0$ you have $\vec{p} = 0$, while $E = E_0$.

Q6a: You defined E_0 as the energy of a particle in its rest frame. On the other hand, photon's speed is always c. Why do you think that the concept of rest energy can be used in the case of a massless photon?

A6a: The experimental upper limit on the mass of the photon is extremely small (less than $10^{-16}eV/c^2$). Therefore it does not play any role in most cases, and we can safely and conveniently speak about massless photons. However even a tiny eventual mass, say $10^{-20}eV/c^2$, allows in principle to consider the rest frame of the photon and thus define its rest energy E_0. For all practical purposes this tiny rest energy is vanishingly small: $E_0 = mc^2 = 0$.

Q6b: How is it possible to put such a tiny upper limit on the mass of the photon which is negligible at any photon energy?

A6b: The mass m of a virtual photon would cut off the magnetic interaction at distances larger than $r = h/mc$. By observing astrophysical magnetic fields at large distances one can get the upper limit on m.

Q6c: Why not abandon the term "mass" in favor of "rest energy"? Why to have two terms instead of one, if we do know that mass is equivalent to rest energy?

A6c: "One" is not always better than "two". The word "mass" refers to a lot of phenomena which have nothing to do with the rest energy "sleeping" in massive bodies. Such a terminological reform would be a disaster not only for Newtonian mechanics for which c is alien, but for Science in general.

Q6d: Isn't it better to have both relations: $E_0 = mc^2$ and $E = mc^2$ instead of one of them? Isn't "two " always better than "one"? Recall the famous wave-particle duality.

A6d: Two relations (explanations) are better than one if both are correct and if each of them has its own realm of applicability. The relation $E = mc^2$ has no separate domain of applicability. Moreover it has no domain of applicability at all. It is a consequence of introduction, along with E, of a redundant variable of "relativistic mass" E/c^2 which usurped the throne of mass. Thus in this case "one" is much better than "two".

Q6e: Why do you dislike the relativistic mass so strongly?

A6e: I stumbled on it 20 years ago and realized how difficult it is to reeducate students and teachers brought up on the concept of mass increasing with velocity and the famous formula $E = mc^2$. It selfpropagates like a virus or a weed and prevents people from understanding the essence of relativity theory.

A century ago Max Planck said that the carriers of wrong views simply die out while new generations accept the truth. But it turned out that new generations come already infected. An important role in the mechanism of infection was played by the authors of textbooks and popular science writers, the editors of popular magazines, like "Scientific American". It is the rare case when most of the experts know the truth, but lightly preach the non-truth.

Q7: Does the mass of a box filled with gas increase with the increase of the temperature of the gas?

A7: Yes, according to relativity theory, it increases.

Q8: Doesn't it mean that the masses of molecules of gas increase with temperature, i.e. with their velocities? Or in other words, that energy and mass are equivalent?

A8: No, it does not mean that. Such an inference would presume the additivity of masses. But according to relativity theory, the total mass of the gas is not equal to the sum of the masses of its molecules: $m \neq \sum m_i$.

In fact the correct interpretation of the mass of the gas supports the relation $E_0 = mc^2$, not the relation $E = mc^2$.

This can be seen from the following reasoning.

The total energy E of the relativistic gas is equal to the the sum of total energies E_i of the individual particles of gas: $E = \sum E_i$. Each E_i increases with temperature. Hence the total energy of gas increases.

The total momentum $\vec{p}$ of the gas vanishes because the distribution of particle's momenta $\vec{p}_i$ is isotropic: $\vec{p} = \sum \vec{p}_i = 0$. Hence the total energy of gas is equal to its rest energy.

By applying the definition of invariant mass: $m^2 = E^2/c^4 - \vec{p}^2/c^2$ and taking into account that in this case $E = E_0$ one gets $E_0 = mc^2$.

This is valid both for the gas of massive particles and for massless photons.

7.4 Interconversion between rest energy and kinetic energy

Q9: Does mass convert into energy? Does energy convert into mass?

A9: No. The "mutual conversion of mass and energy" is a very loose and therefore a misleading term. The point is that energy is strictly conserved in all processes. It can neither appear, nor disappear. It can only transform from one form into another. The rest energy (mass) converts into other forms of energy (e.g. kinetic energy).

Q10: Does energy convert into mass in the processes of production of particles in accelerators?

A10: No. Various forms of energy transform into each other, but the total energy is conserved.

The kinetic energy transforms into rest energy (into masses of the produced particles) in accelerators. The colliders convert E_k into mass much more effectively than the fixed target accelerators.

Q11: Did the laws of conservation of mass and energy merge into one law of conservation of mass-energy similar to the law of conservation of energy-momentum 4-vector?

A11: Yes and No. The laws of conservation of energy and momentum of an isolated system (unified in the law of conservation of 4-momentum) correspond to the uniformity of time and space correspondingly. There is no extra space-time symmetry responsible for the conservation of mass.
The total mass of a closed (isolated) system (the rest energy of the system) is conserved due to conservation of its energy and momentum.

Q12: Doesn't the total mass change in the annihilation of positronium into two photons? Electron and positron are massive, while photons are massless.

A12: No. The total mass does not change: the rest energy of the system of two massless photons is equal, in this process, to the rest energy of positronium.

Q12a: What is the meaning of the term "rest energy of the system of two photons", if each of them has no rest energy and in a second after they were born the two photons are 600 000 km apart?

A12a: "Rest energy of the system of two photons" means here the sum of their kinetic (or total) energies in a reference frame in which the sum of their momenta is equal to zero. In this frame they fly in opposite directions with equal energies.

Q12b: Why do you refer to this rest energy as mass of the system of two photons?

A12b: Because I am applying the equation $E_0 = mc^2$.
The mass of an elementary particle has a deep physical meaning because it is an important quantum number characteristic of all elementary particles of a given sort (say, electrons or protons). The mass of a nuclear or atomic level is also a quantum number.
The mass of a macroscopic body is not as sharply defined because of overlap of huge number of quantum levels.
As for the mass of a system of free particles, it is simply their total energy (divided by c^2) in a frame in which their total momentum is equal to zero. The

value of this mass is limited only by conservation of energy and momentum, like in the case of two photons in the decay of positronium.

As a rule we are unable to measure the inertia or gravity of such a system, but the self-consistency of the relativity theory guarantees that it must behave as mass.

Q12c: Do I understand correctly that with this definition of mass the conservation of mass is not identical to the conservation of matter in the sense in which it was meant by Lomonosov and Lavoisier?

A12c: Yes. You do understand correctly. Matter now includes all particles,even very light neutrinos and massless photons. The number of particles in an isolated piece of matter is not conserved. Roughly speaking, the mass of a body is a sum of masses of constituent particles plus their kinetic energies minus the energy of their attraction to each other (of course, the energies are divided by c^2).

7.5 Binding energy in nuclei

Q13: Is the mass of a nucleus equal to the sum of the masses of the constituent nucleons?

A13: No. The mass of a nucleus is equal to the sum of the masses of the constituent nucleons minus the binding energy divided by c^2.

Thus the nucleus is lighter than the sum of the masses of its nucleons.

Q14: Can the liberation of kinetic energy in the Sun, in nuclear reactors, atomic and hydrogen bombs be explained without referring to the equation $E_0 = mc^2$?

A14: Yes. In the same way that it is explained for chemical reactions, namely, by the existence and difference of binding energies.

Rutherford considered the dependence of mass on velocity as an important fact, but neither he nor his coworkers mentioned $E = mc^2$ or $E_0 = mc^2$ in their works as a source of energy released in radioactive processes though they rejected the idea of perpetum mobile.

7.6 Mass differences of hadrons

Q15: Is the mass of a proton equal to the sum of the masses of two u quarks and one d quark which constitute the proton?

A15: No. The mass of the proton is not equal to the sum of the masses of three quarks. However, the situation here is more subtle than in the case of nucleons in a nucleus.

Q16: What is the main difference between quarks and nucleons?

A16: Nucleons can exist as free particles. (Hydrogen is the most abundant element in the universe, while free neutrons are produced in nuclear reactors.)

Quarks exist only inside hadrons. Free quarks do not exist. Mass is defined by equation $m^2 = E^2/c^4 - p^2/c^2$ only for free particles. Therefore, strictly speaking, we cannot apply this equation to quarks. However one can use the property of asymptotic freedom of QCD – Quantum Chromodynamics.

Q17: What is asymptotic freedom?

A17: According to the asymptotic freedom, the higher the momentum transfer is in interaction of quarks, the weaker their interaction is.

Thus, due to the uncertainty relation, at very short distances quarks look like almost free particles.

In units where $c = 1$ the mass of u quark is 4 MeV at such distances,while that of d quark is 7 MeV. The sum of masses of three quarks inside a proton is 15 MeV, while the mass of the proton is 938 MeV.

Q18: What constitutes the difference between 938 MeV and 15 MeV?

A18: This difference – the main part of the proton mass, as well as of the masses of other hadrons – is caused mainly by the energy of the gluon field – the vacuum condensate of gluons.

Q19: Can we speak about the values of this condensate as of binding energies?

A19: No, we cannot. The contribution of binding energy to the mass is negative, while the contribution of condensate is positive.

By supplying enough energy from outside one can liberate a nucleon from a nucleus, but one cannot liberate a quark in that way from the confinement inside a hadron.

Q20: Can we understand the source of the kinetic energy in beta decay of the neutron without invoking $E_0 = mc^2$?

A20: No, we cannot. Because we cannot express the mass difference between a neutron and a proton in terms of binding energies as we did for nuclei. This is even more so for lepton masses.

Q21: Why does the velocity of light c enter the relation between the mass and the rest energy?

A21: Because c is not only the velocity of light but also the maximal speed of propagation of any signal in Nature.

As such, it enters all fundamental interactions in Nature as well as Lorentz transformations.

Q22: Why do you claim that gravity is reducible to the interaction of energies, not masses?

A22: Because a massless photon is attracted by the gravitational field of the Sun. (The deflection of light was first observed in 1919 and brought Einstein world fame.) As for the massive particle, its mass is equal to its rest energy. Thus in both cases we deal with energy.

There is also another argument in favor of energy as a source of gravity.

I refer here to the fact established by Galileo almost four centuries ago and confirmed in the XXth century with accuracy 10^{-12}. Namely, that all bodies have the same gravitational acceleration. It does not depend on their composition, on the proportions between different terms in their rest energy. That means that only the total rest energy of a slow body determines both its gravitational attraction and its inertia.

Q23: How was this fact explained in the framework of prerelativistic physics and how is it explained by relativity theory ?

A23: In the prerelativistic physics it was formulated as a mysterious equality of inertial mass m_i and gravitational mass m_g.

In relativity theory it became trivial, because both inertia and gravity of a body are proportional to its total energy.

Q24: What are the main directions in the research on the concept of mass in the next decade?

A24: The main experimental direction is the search for higgs at LHC at CERN. According to the Standard Model, this particle is responsible for the masses of leptons and quarks as well as of W and Z bosons. Of great interest is also the experimental elucidation of the pattern of neutrino masses and mixings.

The main cosmological direction is the study of dark matter and dark energy.

Q25: What was the formulation of the Corollary v in "The Principia"?

A25: Here is the citation from "The Principia":
Sir Isaac Newton.
The Principia.
Axioms, or Laws of Motion.

COROLLARY V. The motions of bodies included in a given space are the same among themselves, whether that space is at rest, or moves uniformly forwards in a right line without any circular motion.

For the differences of the motions tending towards the same parts, and the sums of those that tend towards contrary parts, are, at first (by supposition), in both cases the same; and it is from those sums and differences that the collisions and impulses do arise with which the bodies mutually impinge one upon another. Wherefore (by Law II), the effects of those collisions will be equal in both cases; and therefore the mutual motions of the bodies among themselves in the one case will remain equal to the mutual motions of the bodies among themselves in the other. A clear proof of which we have from the experiment of a ship; where all motions happen after the same manner, whether the ship is at rest, or is carried uniformly forwards in a right line.

Physics – Uspekhi **51** (5) 513 – 527 (2008)

FROM THE HISTORY OF PHYSICS

PACS numbers: 01.10.Fv, 05.70.Fh, **75.40.–s**

The Einstein formula: $E_0 = mc^2$. "Isn't the Lord laughing?"

L B Okun

DOI: 10.1070/PU2008v051n05ABEH006538

Contents

Abstract. The article traces the way Einstein formulated the relation between energy and mass in his work from 1905 to 1955. Einstein emphasized quite often that the mass m of a body is equivalent to its rest energy E_0. At the same time, he frequently resorted to the less clear-cut statement of the equivalence of energy and mass. As a result, Einstein's formula $E_0 = mc^2$ still remains much less known than its popular form, $E = mc^2$, in which E is the total energy equal to the sum of the rest energy and the kinetic energy of a freely moving body. One of the consequences of this is the widespread fallacy that the mass of a body increases when its velocity increases and even that this is an experimental fact. As wrote the playwright A N Ostrovsky "Something must exist for people, something so austere, so lofty, so sacrosanct that it would make profaning it unthinkable."

L B Okun Russian Federation State Scientific Center
"A I Alikhanov Institute of Theoretical and Experimental Physics",
ul. B. Cheremushkinskaya 25, 117218 Moscow, Russian Federation
Tel. (7-495) 123 31 92, (7-495) 125 96 60
E-mail: okun@itep.ru

Received 21 December 2007, revised 20 March 2008
Uspekhi Fizicheskikh Nauk **178** (5) 541 – 555 (2008)
DOI: 10.3367/UFNr.0178.200805g.0541
Translated by V I Kisin; edited by A M Semikhatov and M S Aksent'eva

1. Introduction

The formula $E = mc^2$ is perhaps the most famous formula in the world. In the minds of hundreds of millions of people it is firmly associated with the menace of atomic weapons. Millions perceive it as a symbol of relativity theory. Numerous authors popularizing science keep persuading their readers, listeners, and viewers that the mass of any body (any particle) increases, as prescribed by this formula, when its velocity increases. And only a small minority of physicists — those who specialize in elementary particle physics — know that Einstein's true formula is $E_0 = mc^2$, where E_0 is the energy contained in a body at rest, and that the mass of a body is independent of the velocity at which it travels.

Most physicists familiar with special relativity know that in it, the energy E and momentum $\mathbf{p}$ of a freely moving body are related by the equation $E^2 - \mathbf{p}^2c^2 = m^2c^4$, where m is the mass of the body. Alas, not all of them realize that this formula is incompatible with $E = mc^2$. But an even smaller number of people know that it is perfectly compatible with $E_0 = mc^2$, because E_0 is the value assumed by E when $\mathbf{p} = 0$. This article is written for those who do not want to be lost in three pines* of the above three formulas and who wish to attain a better understanding of relativity theory and its history.

When Einstein first introduced the concept of rest energy in 1905 and discovered that the mass of a body is a measure of

* "To be lost in three pines" in Russian is equivalent to "loose one's way in broad daylight" in English. (*Author's note to English version of the article.*)

the energy contained in it, he felt so amazed that he wrote in a letter to a friend: "for all I know, God Almighty might be laughing at the whole matter and might have been leading me around by the nose." In what follows, we see how throughout his life Einstein returned again and again to this same question.

We shall see how the formula $E_0 = mc^2$ made its way through Einstein's writings. Also, how he carefully emphasized that the mass of a body depends on the amount of energy it contains but never stated (in contrast to his popularizers!) that mass is a function of the body's velocity. Nevertheless, it is true that he never once rejected the formula $E = mc^2$ that is believed to be 'his formula' and in the mass psyche is an icon of modern physics.

To my reader: If you feel bored with following the meticulous analysis and collation of texts, please jump to the Epilogue and the adjacent sections, where I have tried to briefly describe the results of the analysis without going into technicalities. It is possible that after you do so, reading about Einstein's many attempts to clarify the relation between energy and mass will become more interesting and compelling.

When writing this review, I used Einstein's historically first ever collected works [1]. (This four-volume edition was published in Russian in 1965 – 1967.) Where possible, I also used the multivolume Princeton Collected Papers. (Volumes with 'all papers and documents' by Einstein [2] and their translations into English [3] began to appear in Princeton in 1987. In 2007, ten volumes were published, of which five (1, 5, 8, 9, 10) contain his correspondence until 1920 and five (2, 3, 4, 6, 7) contain his works until 1921).

2. Prologue. The years 1881 – 1904

It is well known that the principle of relativity dates back to Galileo [4] and Newton [5], and that the theory of relativity was constructed in the papers of Lorentz, Einstein, Poincaré, and Minkowski [6].

The notion of velocity-dependent mass was born in the years preceding the creation of the theory of relativity and in the first years after its creation.

It was molded in the papers of Thomson [7], Heaviside [8], Searle [9], Abraham [10], and also Lorentz [11] and Poincaré [12], who tried hard to have Maxwell's equations of electromagnetism to agree with the equations of Newton's mechanics. These publications stimulated the experiments of Kaufmann [13] and Bucherer [14, 15]. They used formulas of Newton's nonrelativistic mechanics to process their experimental data and concluded that mass increases with increasing velocity.

It was the matter not only of formulas as such but also of the very spirit, the very foundations of the nonrelativistic physics in which mass is a measure of inertia of a body. It was difficult to comprehend, at the borderline between the 19th and 20th centuries, that these foundations were being replaced by a more general base: the measure of inertia of a body is not its mass but its total energy E equal to the sum of rest energy and kinetic energy. The fact that energy E entered with a factor $1/c^2$ prompted people to interpret E/c^2 as the mass. In fact, the progress in relativity theory, achieved mostly through the efforts of Einstein, Minkowski, and Noether, showed that it was necessary to connect mass not with total energy but only with rest energy.

3. 1905 — annus mirabilis

In 1905, Einstein published his three ground-breaking, fundamental papers dealing with the properties of light and matter [16 – 18].

In [16], he introduced the concept of the quantum of energy of light and, using this concept, explained the photoelectric effect, which had been experimentally discovered not long before that. (The value of the Planck constant h — the quantum of action — had been established earlier, see [19].)

In [17], Einstein considered almost the entire set of consequences of the principle of relativity and of the finite speed of light. Thus he derived in § 8 the formula for the transformation of the energy of light in the transition from one inertial reference frame to a different one that moves at a velocity v relative to the former:

$$\frac{E'}{E} = \frac{1 - (v/V)\cos\phi}{\sqrt{1 - (v/V)^2}}\,.$$

Here, V is the velocity of light and ϕ is the angle between the direction of motion of light and that of the observer. Then in § 10 he obtained the expression for the kinetic energy of the electron:

$$W = \mu V^2 \left\{ \frac{1}{\sqrt{1 - (v/V)^2}} - 1 \right\},$$

where μ is the mass of the electron and v is its velocity.

(Furthermore, in § 10, Einstein derived expressions for the so-called longitudinal m_l and transverse m_t masses of the electron that Abraham and Lorentz had earlier introduced and he obtained:

$$m_l = \frac{\mu}{\left(\sqrt{1 - (v/V)^2}\right)^3}\,,$$

$$m_t = \frac{\mu}{1 - (v/V)^2}\,.$$

The second of these expressions differs from Lorentz's m_t and is wrong, and later Einstein never insisted on it.)

As regards the formulas for the kinetic energy W of an electron and for a photon energy E', he applied both these formulas in the next paper [18] when deriving the relation between mass and energy.

There he considered 'two amounts of light,' with energy $L/2$ each, both emitted by a massive body at rest but traveling in opposite directions. In this paper, Einstein for the first time introduced the rest energy of a massive body, denoting it by E_0 before emission and by E_1 after. In view of the energy conservation law,

$$E_0 - E_1 = L\,.$$

He then looked at the same process in a reference frame moving at a velocity v relative to the body, and obtained the following expression for the difference between kinetic energies of the body before and after the act of emission:

$$K_0 - K_1 = L \left\{ \frac{1}{\sqrt{1 - (v/V)^2}} - 1 \right\}.$$

He also specially pointed out that the difference between kinetic energies contains an arbitrary additive constant C included in the expression for energy. He returned to the matter of the constant C many times during the subsequent 50 years; we discuss it later in this paper.

The left- and right-hand sides of the equality above depend on v in the same manner, as follows from the expression for W. Since the velocity v is the same before and after the emission, while the kinetic energy of the body decreased, this immediately implies that the mass of the body decreased by the amount L/V^2. From this Einstein concluded that "The mass of a body is a measure of its energy content" and remarked that it might be possible to check this conclusion in the decays of radium.

The title of the paper is noteworthy: "Does the inertia of a body depend on its energy content?". Considered together with the contents of the paper, it indicates that it was the mass that Einstein identified with the measure of a body's inertia. But this is only valid in Newton's approximation. As we know today, the measure of a body's inertia in relativity theory is its total energy E: the greater the total energy of a body, the greater its inertia. (By the 'measure of a body's inertia,' we here mean the proportionality coefficient between momentum and velocity. There is no universal proportionality coefficient between force and acceleration in relativity theory. Lorentz and Abraham had already established this when they introduced the longitudinal and transverse masses.

Einstein held to the opinion that the energy of a free body is defined in relativity theory only up to an additive constant, by analogy to potential energy in Newtonian mechanics. This may have resulted in his underestimating his own revolutionary step forward — the introduction of the concept of rest energy into physics. There is nothing special about the rest energy E_0 once energy is only defined up to C.

But as we know today, there is no place for C in the theory he created. The energy and momentum of a free particle are uniquely defined in the theory by the relation $E^2 - \mathbf{p}^2c^2 = m^2c^4$; we return to it more than once in what follows.

4. Have I been led around by the nose?

The discovery that mass depends on energy struck Einstein so forcibly that he wrote in a letter to his friend Conrad Habicht [20] (see also [21]):

"A consequence of the study on electrodynamics did cross my mind. Namely, the relativity principle, in association with Maxwell's fundamental equations, requires that the mass be a direct measure of the energy contained in a body; light carries mass with it. A noticeable reduction of mass would have to take place in the case of radium. The consideration is amusing and seductive; but for all I know, God Almighty might be laughing at the whole matter and might have been leading me around by the nose."

It looks as if God continues to lead the interpreters of the relativity theory by the nose much as He did in Einstein's time.

5. 1906 – 1910. Minkowski

1906

In 1906, Einstein published two papers on relativity theory: [22, 23]. In [22], he treated mass transfer by light in a hollow cylinder from its rear face to the front. For the cylinder not to move as a whole, he imposed the condition that light with an energy E has the mass E/V^2; he thereby reproduced Poincaré's result of 1900 [12]. Presumably, he considered it inadmissible for the energy and mass carrier to have zero mass (to be massless). In [23], he considered a method for determining the ratio of longitudinal and transverse masses of the electron previously introduced by Lorentz and Abraham. As far as mass is concerned, therefore, these papers were a step back in comparison with [18].

1907

In 1907, Einstein published four papers on relativity theory: [24 – 27]. The first of these discussed the frequency of radiation from an atom. The second emphasized the difference between the relativity principle and the relativity theory. He considered his own work as dealing with the principle of relativity, which he regarded as being analogous to those of thermodynamics. As for the theory of relativity, he believed that it was yet to be constructed.

The paper that is especially significant for us here is [26], which gave the formulation of the mass – energy equivalence (see footnote in § 4): "One should note that the simplifying assumption $\mu V^2 = \varepsilon_0$ is also the expression of the principle of the equivalence of mass and energy..." (The simplifying assumption referred to here is the choice of an arbitrary constant in the expression for energy.)

The most detailed among the papers published in 1907 was [27]. It consists of five parts: (1) Kinematics (§ 1 – § 6). (2) Electrodynamics (§ 7). (3) Mechanics of a material point (electron) (§ 8 – § 10). (4) On the mechanics and thermodynamics of systems (§ 11 – § 16). (5) Relativity principle and gravitation (§ 17 – § 20). Short note [28] with corrections of misprints and elaborations belongs to this group of papers.

Of special interest for us are parts 4 and 5. In part 4, Einstein discussed the additive constant in the energy and showed that it is not included in the relation between momentum, energy, and velocity of a body. Part 5 ended with the following words:

"Thus the proposition derived in § 11, that to an amount of energy E there corresponds a mass of magnitude E/c^2, holds not only for the *inertial* but also for the *gravitational* mass, if the assumption introduced in § 17 is correct."

On the one hand, this sentence states that energy, not mass, is both the measure of inertia and the source of gravitation. But on the other hand, it can be understood to say that a photon with an energy E has both the inertial mass and the gravitational mass equal to E/c^2. This ambiguous interpretation continues to trigger heated debates.

1908

In 1908, Einstein together with J Laub published two articles on the electrodynamics of moving macroscopic bodies: [29, 30] (see also [31, 32].) Although pertaining to relativity theory, these papers are nevertheless not relevant to the problem under discussion here, the relation between energy and mass.

The talk delivered by Hermann Minkowski in 1908 [33] was an important milestone in the history of relativity theory. Minkowski was the first to propose the four-dimensional spacetime formulation of the theory. In this formulation, as we know, the mass of a particle is a quantity independent of its velocity.

It may seem paradoxical but the first paper by Lewis [34] declaring that the mass equals E/c^2 appeared at the same time. This standpoint was further developed and spread by Lewis and Tolman in [35-38].

1909

Einstein's paper [39] published in 1909 is not concerned with the relation between mass and energy. But we find a number of statements in his articles [40-42] published at the same time that shed much light on his understanding of this problem. For instance, in [42], which contains the text of Einstein's first public speech (at a congress of German natural scientists in Salzburg), he wrote:

"The first volume of the excellent textbook [1] by Chwolson which was published in 1902, contains in the Introduction the following sentence about the ether: 'The probability of the hypothesis on the existence of this agent borders extraordinarily closely on certainty.' However, today we must regard the ether hypothesis as an obsolete standpoint."

Then: "...the inertial mass of a body decreases upon emission of light... Energy and mass appear as equivalent quantities the same way that heat and mechanical energy do... The theory of relativity has thus changed our views on the nature of light insofar as it does not conceive of light as a sequence of states of a hypothetical medium but rather as something having an independent existence just like matter."

1910

In 1910, A Einstein and L Hopf discussed the application of probability theory to the analysis of the properties of radiation [43, 44].

At the same time, Einstein published in a French journal a major review of relativity theory [45] devoted mostly to the transformations of spatial coordinates and time but also briefly outlining Minkowski's ideas about the four-dimensional world. Only at the end of this paper did he mention that "...the mass of any arbitrary body *depends on the quantity of energy* it contains... Unfortunately, the change of mass W/c^2 is so slight that one cannot hope for its detection by experiment for the time being."

Einstein did not stipulate that by "energy W contained in a body" he meant rest energy.

6. 1911-1915. On the road to General Relativity Theory

1911

In 1911, Einstein published three papers on the theory of relativity: [46-48].

In [46], he discussed the propagation of light in a gravitational field, starting with the assumption that a photon with energy E has an inertial and a gravitational mass, both of which are equal to E/c^2, and he calculated that the angle of deflection of light by the Sun's gravitational field would be 0.83 arc second — which is half the correct value that he would later derive (in 1915) using general relativity. (I should remark that the same "half value" had already been obtained and published by Soldner in 1804 (see [49, 50]). But Einstein was not aware of it: Soldner's paper was totally forgotten soon after its publication.)

[1] "The Course of Physics" by O D Chwolson (volumes 1 and 2) was published in Russian in 1897; its German translation appeared in 1902.

At the end of review paper [47] devoted mostly to clocks and rods in relativity theory, Einstein mentioned uniting the law of conservation of mass with the law of conservation of energy. "However odd this result might seem, still, in a few special cases, one can unequivocally conclude from empirically known facts, and even without the theory of relativity, that the inertial mass increases with energy content." Perhaps this sentence refers to experiments of Kaufmann and Bucherer. But this would suggest that he believed that mass increases with increasing kinetic energy and therefore with increasing velocity.

A short note [49] discussed the contraction of the length of a moving rod.

1912

Einstein's papers of this period [51-55] were mostly attempts to create a more general relativity theory that would embrace gravitation. Only lectures [51] dealt with special relativity.

His statements made during 1912 again display the above-mentioned ambiguity in the interpretation of mass as the equivalent of rest energy, on the one hand, and as a measure of inertia, on the other.

We find there a statement that m should be considered to be a characteristic constant of a 'material point' (massive point-like body), which does not vary as a function of the object's motion. On the other hand, it is also stated that the energy of a free particle is defined only up to an arbitrary additive constant. Nevertheless, mc^2 equals the rest energy (see the discussion of equation (28′) in [51]).

1913-1914

In paper [56] co-authored with M Grossmann, Einstein continued to discuss the proportionality between the inertial and gravitational masses, which had been measured with high accuracy in experiments by Eötvös, and he discussed the dependence of the speed of light c on the gravitational potential.

In 1914, Einstein published a short note expounding his point of view regarding the concept of mass [57]. A manuscript with a synopsis of his lectures on special relativity theory dates back to the same period [58].

In [57], he discussed the contribution of the gravitational field to the gravitational and inertial masses of a body and came to the conclusion that the inertia of a closed system is entirely determined by its rest energy.

Paper [58] gave an expression for the energy-momentum 4-vector and the relation $E_0/c^2 = m$, which would appear again only in 1921. We note that m was referred to in [58] as rest mass (Ruhemasse), which seems to imply that the mass of a body at rest is not the same as when the body moves.

1915

The year 1915 was marked by the completion of general relativity theory, in paper [59]. In fact, already in his preceding paper [60], Einstein had derived formulas that described two most important effects of this theory: the precession of Mercury's perihelion and the deflection of light by the gravitational field of the Sun. The secular motion of Mercury's perihelion (about 40″ per century), which could not be explained in terms of the influence of the known bodies in the solar system, was established by Le Verrier in 1859. Einstein calculated that general relativity theory predicted secular precession as 43″.

But the true world fame came from the prediction of the angle of deflection of light by 1.7″ after it had been confirmed by the British expedition that observed the solar eclipse in 1919.

7. 1917. Cosmological constant

1917

A book was published in 1917 to popularize relativity theory [61]. It dealt mostly with the joint transformation of space and time coordinates. However, § 15 mentioned the kinetic energy of a material point, which now equaled not $mv^2/2$ but $mc^2/\sqrt{1-v^2/c^2}$, and therefore incorporated both its kinetic energy proper and its rest energy.

Then we read this:

"Before the advent of relativity, physics recognized two conservation laws of fundamental importance, namely, the law of the conservation of energy and the law of the conservation of mass; these two fundamental laws appeared to be quite independent of each other. By means of the theory of relativity they have been united into one law."

And even though an attentive reader concludes from the text that follows that Einstein was speaking of $E_0 = mc^2$, a slightly less attentive reader might guess that $E = mc^2$ was meant. The fact that at times Einstein treated rest energy as part of kinetic energy did not help to clarify matters.

The most famous among Einstein's papers published in 1917 was called "Cosmological Considerations on the General Theory of Relativity" [62]. There Einstein formulated for the first time the possibility of a non-vanishing energy density of the vacuum; he denoted it by the letter λ. This energy density is the same at every point in the Universe. It is essentially a completely delocalized energy, spread over the entire Universe.

Einstein introduced this cosmological constant — the so-called λ-term — in order to be able to describe a stationary Universe in general relativity. It soon became clear, however, that a stationary solution cannot be achieved in this manner.

In 1922, Friedmann, while reading this paper by Einstein, advanced his theory of the expanding Universe [63, 64]. Einstein first dismissed Friedmann's arguments [65], but then accepted them [66]. In 1929, Hubble published the first observational data [67] supporting the expansion of the Universe.

In 1945, Einstein published the second edition of his book "The Meaning of Relativity" with a special addendum "On the Cosmological problem" devoted to the theory of the expanding Universe [68]. At the turn of the 1970s – 1980s, a model of the exponentially fast expansion (inflation) of the early Universe was suggested [69 – 71]. According to this model, the effective cosmological term forms when the Universe is created, due to a nonzero mean vacuum value of a special scalar field, which later transforms into high-energy particles. In 1998 – 1999, two groups of observers measuring the luminosity and spectra of supernovas came to a conclusion that the rate of the expansion of the Universe is increasing [72, 73] (see also [74].) The available data indicate that ordinary matter contains only 4% of the energy of the Universe, that about 24% is contained in the particles of the so-called dark matter whose nature is as yet unknown, and about 70% of the entire energy of the Universe is usually referred to as dark energy and attributed to Einstein's cosmological constant λ.

8. 1918 – 1920. Noether

1918

In 1918, the brilliant paper of Emmy Noether was published [75], in which she proved, among other things, that the dynamic conservation laws are implied by the symmetry properties of space – time. We know that conservation of energy is a consequence of the uniformity of time, and that conservation of momentum is a consequence of the uniformity of space. Angular momentum is conserved as a result of the isotropy of space: physics remains unchanged if coordinate axes undergo rotation in the planes xy, yz, zx. Similarly, Lorentz invariance follows from the fact that physics remains unchanged under pseudo-Euclidean rotations in the planes xt, yt, zt. Einstein wrote very enthusiastically about this discovery of Noether in a letter to Hilbert [76]:

"Yesterday I received a very interesting paper by Ms. Noether about the generation of invariants. It impresses me that these things can be surveyed from such general point of view. It would not have harmed the Göttingen old guard to have been sent to Miss Noether for schooling. She seems to know her trade well!"

Soon after that Einstein sent for publication a paper [77] on the conservation of energy in general relativity, which presented a statement that the energy of a closed system plays the role of both inertial and gravitational mass.

1919

Among the publications of 1919, I need to specially mention a short note "A test of the general theory of relativity" [78] on the discovery of the deflection of light rays by attraction of the Sun and an article in *The Times* entitled "What is the theory of relativity?" [79]. Among other things, Einstein wrote:

"The most important upshot of the special theory of relativity concerned the inertial masses of corporeal systems. It turned out that the inertia of a system necessarily depends on its energy-content, and this led straight to the notion that inert mass is simply latent energy. The principle of the conservation of mass lost its independence and became fused with that of the conservation of energy."

1920

In 1920, Einstein prepared a draft manuscript of an extensive popular article "Fundamental ideas and methods of the theory of relativity, presented in their developments." Einstein worked on this article as an invited publication in *Nature*, but it was never published [80].

At the same time, Einstein's letter appeared in a Berlin newspaper, "My response. On the anti-relativity company" [81]. The letter opens with the words: "Under the pretentious name "Arbeitsgemeinschaft deutscher Naturforscher," a variegated society has assembled whose provisional purpose of existence seems to be to degrade, in the eyes of nonscientists, the theory of relativity as well as me as its originator."

Then Einstein wrote: "...I have good reasons to believe that motives other than the striving for truth are at the bottom of this business. [...] I only answer because well-meaning circles have repeatedly urged me to make my opinion known.

First, I want to note that today, to my knowledge, there is hardly a scientist among those who have made substantial contributions to theoretical physics who would not admit that the theory of relativity in its entirety is founded on a logical basis and is in agreement with experimental facts which to date have been reliably established. The most important theoretical physicists — namely, H A Lorentz, M Planck, A Sommerfeld, M Laue, M Born, J Larmor, A Eddington, P Debye, P Langevin, T Levi-Civita — support the theory, and most of them have made valuable contributions to it. [...]

I have been accused of running a tasteless advertising campaign for the theory of relativity. But I can say that all my life I have been a friend of well-chosen, sober words and of concise presentation."

9. 1921. "The Meaning of Relativity"

In 1921, Einstein was invited to Princeton and delivered there a course of lectures that make up the book "The Meaning of Relativity" [82]. In this book, he described for the first time, with maximum exposure to the public and unambiguously, what he understood by the equivalence of energy and mass. His equations (41)–(43) give expressions for the components of the energy–momentum 4-vector of a body in terms of its mass and velocity. Equation (44) gives an expression for the energy of a body in terms of its mass: $E_0 = mc^2$. In equation (45), he gave an expression for energy at a low velocity q: $E = m + mq^2/2 + 3mq^4/8 + \ldots$ (in units in which $c = 1$.) The text between equations (44) and (45) reads: "Mass and energy are therefore essentially alike; they are only different expressions of the same thing. The mass of a body is not constant; it varies with changes in its energy." Then follows a footnote about energy release in radioactive decays: "The equivalence of mass at rest and energy at rest which is expressed in equation (44) has been confirmed in many cases during recent years. In radio-active decomposition the sum of the resulting masses is always less than the mass of the decomposing atom. The difference appears in the form of kinetic energy of the generated particles as well as in the form of released radiational energy."

Three aspects deserve our attention in these statements. First, while giving a clear definition of mass in the equations as a velocity-independent quantity, the term "mass at rest" is used for it, which implies that mass depends on velocity. Second, there is no explicit statement that mass changes only when the energy of a body changes, but not its velocity. Third, the ambiguous statement that mass and energy are "only different expressions of the same thing," even though mass is a relativistic invariant, i.e., a four-dimensional scalar, while energy is the fourth component of a four-dimensional vector. It is possible that these rather imprecise words accompanying perfectly precise formulas are the reason why many readers still fail to see in [82] a clear-cut statement in favor of $E_0 = mc^2$ and against $E = mc^2$.

A small popular-science brochure deserves being mentioned here: "Relativity theory" [83], whose author, I Leman, expressed his gratitude to Einstein for valuable advice. He spoke of his awe for the profundity and elegance of Minkowski's ideas and emphasized the enormous amounts of energy stored in matter as its mass.

10. 1927–1935

1927

In 1927, several conferences were dedicated to the bicentennial of the death of Isaac Newton. Einstein marked the occasion with a number of publications. He wrote in [84]: "Newton's teaching provided no explanation for the highly remarkable fact that both the weight and the inertia of a body are determined by the same quantity (its mass). The remarkableness of this fact struck Newton himself."

By 1927, mostly through the work of Einstein, it became clear that the inertia and the weight of a moving particle are determined not by its mass but by its energy E and the quantity $p_\mu p_\nu / E$, where p_μ is energy–momentum vector. In the Newtonian limit, both are reduced to the rest energy, i.e., to mass. Such is the simple explanation provided by relativity theory of the equality of the inertial and gravitating masses in Newtonian mechanics.

However, we see that Einstein continued to use the old nonrelativistic terminology.

1928

In the paper "Fundamental concepts of physics and their most recent changes" [85], Einstein formulated his attitude to the problem of causality in quantum mechanics, saying that

"Thus the field theory shook the fundamental concepts of time, space and matter. But upon one column of the edifice it made no assault: on the hypothesis of causality. From some single condition of the world at a given time, all other previous and subsequent conditions uniquely follow based upon the laws of of nature.

Today, however, serious doubts have emerged about the law of causality thus understood. This is not to be charged to the craving for new sensations on the part of the learned, but to the momentum of facts which seem irreconcilable with a theory of strict causality. It seems at this time as if the field, considered as a final reality, does not make proper allowance for the facts of radiation and atomic structure. We reach here a complication of questions with which the modern generation of physicists is struggling in a gigantic display of intellectual power."

This problem was solved twenty years later in Feynman's two papers on quantum electrodynamics (see below), but Einstein failed to notice it. This may have been caused by Einstein's belief that all of quantum physics violated causality.

1929

In his article for the Encyclopedia Britannica [86], Einstein described the four-dimensional spacetime continuum but wrote not a word about Minkowski and the energy–momentum four-dimensional space.

In his speech at the ceremony in honor of the 50th anniversary of Planck's presentation of his doctoral dissertation — at which Einstein received the Planck Medal — he returned to the problem of causality in quantum mechanics. He wrote that even though he was deeply convinced that theory would not stop at the subcausality level and would ultimately reach the supercausality in the sense discussed by him earlier, he was impressed by the work of the younger generation of physicists on quantum mechanics, and that he regarded this theory as a correct one. He only mentioned that restrictions resulting in the *statistical* nature of its laws should be eliminated with time [87].

1934–1935

On December 29, the Pittsburgh Post-Gazette published an interview with Einstein under the heading "Atom energy hope is spiked by Einstein" [88].

In December 1934, Einstein read to the joint session of the American Mathematical Society, the American Physical

Society, and the American Society for the Advancement of Science a lecture entitled "Elementary derivation of the equivalence of mass and energy." This lecture was published in 1935 in the Bulletin of the American Mathematical Society [89].

The challenge Einstein set himself was to prove that mass and energy are equivalent, on the basis of only three assumptions:

"In the following considerations, except for the Lorentz transformation, we will depend only on the assumption of the conservation principles for impulse and energy."

In its first pages, Einstein introduces the velocity 4-vector, and by multiplying it by mass m, obtains the 4-vector whose spatial components — in his opinion — can naturally be regarded as momentum and the time component as energy:

"Here it is natural to give it directly the meaning of energy, hence to ascribe to the mass-point in a state of rest the *rest-energy* m (with the usual time unit, mc^2).

Of course, ...in no way is it shown that this impulse satisfies the impulse-principle and this energy the energy-principle...

Furthermore, it is not perfectly clear as to what is meant in speaking of the *rest-energy*, as the energy is defined only to within an undetermined additive constant...

What we will now show is the following. If the principles of conservation of impulse and energy are to hold for all coordinate systems which are connected with one another by the Lorentz transformations, then impulse and energy are really given by the above expressions and the presumed equivalence of mass and rest-energy also exists."

And he undertook to prove that conservation laws indeed hold for the 4-momentum that he considered. To achieve this, he calculated the energies and momenta of two particles before and after their collision in different Lorentz reference frames and concluded:

"The rest-energy changes, therefore, in an inelastic collision (additively) like the mass. As the former, from the nature of the concept, is determined only to within an additive constant, one can stipulate that E_0 should vanish together with m. Then we have simply $E_0 = m$, which states the principle of equivalence of inertial mass and rest-energy."

It is worthy of note here that in this lecture, Einstein never mentioned Noether's theorem [75], which implies that the conservation of the 4-momentum and the Lorentz invariance follow from symmetry properties of the Minkowski space-time. He preferred to derive the properties of the 4-momentum by considering two-body collisions in the three-dimensional space and to independently assume the Lorentz invariance and conservation of energy and momentum.

On May 4, 1935, he published an obituary in The New York Times entitled "The late Emmy Noether" [90], where he spoke of his high opinion of her contributions to mathematics but failed to mention her theorem that is of such importance in physics. A self-consistent presentation of conservation laws on the basis of the symmetries of space–time in the spirit of Noether was given for the first time in 1941 by L D Landau and E M Lifshitz in their "Field theory" (see below.)

In the same year, 1935, another famous paper was published [91], written in collaboration with N Rosen and B Podolsky on the interpretation of measurements in quantum mechanics.

11. 1938–1948. Atomic bomb

1938

In 1938, the famous science-popularizing book was published, "The Evolution of Physics" [92], written by Einstein and his young assistant Leopold Infeld. The authors often returned to the concept of mass on its pages. The section "One clew remains" in chapter I "The rise of the mechanical view" introduced the concepts of inertial and gravitating masses and described their equality as a thread leading the way to general relativity. In the section "Relativity and mechanics" of chapter III "Field, relativity," the authors introduced the concept of rest mass: "A body at rest has a definite mass, called *rest mass*." Then they wrote: "radiation traveling through space and emitted from the sun contains energy and therefore has mass;" and a bit later: "According to the theory of relativity, there is no essential distinction between mass and energy. Energy has mass and mass represents energy. Instead of two conservation laws we have only one, that of mass-energy. This new view proved very successful and fruitful in the further development of physics." One might justly think that this statement is an adequate 'verbal' equivalent of the formula $E = mc^2$ and is incompatible with the formula $E_0 = mc^2$.

In the section "General relativity and its verification" of the same chapter III, we read that the elliptical orbit of Mercury precesses, completing a full cycle around the Sun in three million years. This precession of Mercury's perihelion is caused by relativistic properties of the gravitational field. The next section says this:

"We have two realities: matter and field. [...] But the division into matter and field is, after the recognition of the equivalence of mass and energy, something artificial and not clearly defined. Could we not reject the concept of matter and build a pure field physics?"

The creation of relativistically invariant quantum electrodynamics at the junction of the 1940s and 1950s, and later of the quantum field theory of the electroweak and strong interactions, as well as various models of the so-called grand unification of all interactions, can be regarded as the implementation of Einstein's dream of a unified field theory. However, all these theories are based not only on the theory of relativity but also on quantum mechanics, whose probabilistic interpretation was unacceptable to Einstein, who insisted that "God does not play dice." It was owing precisely to quantum mechanics that matter was not expelled from these theories but rather became their foundation. This is seen especially clearly in the language of Feynman diagrams, in which real particles (including photons) represent matter and virtual particles represent force fields (see below). The concluding chapter IV entitled "Quanta" is a story about quantum mechanics. The section "The quanta of light" tells the reader that light consists of grains of energy — light quanta, or photons. The section "The waves of matter" emphasizes the similarity between photons and electrons in the combination of wave and corpuscular properties. "One of the most fundamental questions raised by recent advances in science is how to reconcile the two contradictory views of matter and wave." The authors are just a stone's throw from conceding that the photon is just as much a particle of matter as the electron is. However, at the end of book, they say:

"Matter has a granular structure; it is composed of elementary particles, the elementary quanta of matter. Thus, the electric charge has a granular structure and — most important from the point of view of the quantum theory — so has energy. Photons are the energy quanta of which light is composed."

Light is therefore identified with energy and becomes an antithesis of matter. Could it be that this identification and this opposition constitute one of the roots of the mass–energy confusion?

1939

On August 2, 1939, Leo Szilard persuaded Einstein to write the famous letter to President F D Roosevelt warning that "...the element uranium may be turned into a new and important source of energy..." [88].

1941. Landau and Lifshitz

The first Russian edition of Landau and Lifshitz's "The theory of fields" [93] appeared in 1941. In § 10 "Energy and momentum" (it became § 9 in subsequent editions of the volume), they introduced the energy–momentum 4-vector and its square equal to mass squared, and discussed rest energy, although did not denote it by E_0. The nonadditivity of mass in relativity theory was mentioned as the nonconservation of mass. All conservation laws in this book were consistently obtained from the symmetry properties of space–time in accordance with Noether's theorem. However, it is very unlikely that Einstein read Russian textbooks. He likewise missed the publication of the translation into English in 1951 [94].

1942

In 1942, P G Bergmann's book was published [95] with a foreword by Einstein, which said, among other things, that:

"This book gives an exhaustive treatment of the main features of the theory of relativity which is not only systematic and logically complete, but also presents adequately its empirical basis. ...Much effort has gone into making this book logically and pedagogically satisfactory, and Dr. Bergmann has spent many hours with me which were devoted to this end."

In chapter VI, we read:

"...relativistic kinetic energy equals

$$E = \frac{mc^2}{\sqrt{1 - u^2/c^2}} + E_0 \,, \tag{6.17}$$

where E_0 is the constant of integration...

$$T = mc^2\left[\left(1 - \frac{u^2}{c^2}\right)^{-1/2} - 1\right]. \tag{6.20}$$

...The quantity mc^2 is called the 'rest energy' of a particle, while T is its 'relativistic kinetic energy."

It is not clear to me why the integration constant C had to be denoted by E_0. Neither do I understand why the 'relativistic kinetic energy' was denoted by two different symbols E and T. Could it be that a misprint crept in and Eqn (6.17) is the total, not kinetic, energy? Immediately following it is this text:

"*Relation between energy and mass.* The ratio between the momentum and the mass, the quantity μ, is often called 'the relativistic mass' of a particle, and m is referred to as 'the rest mass.' The relativistic mass is equal to the total energy divided by c^2, and likewise the rest mass is $(1/c^2)$ times the rest energy. There exists, thus, a very close correlation between mass and energy which has no parallel in classical physics."

We thus see that the additive constant in the expression for energy and the dependence of mass on velocity survived in this book. Also retained was the ambiguity connected with the definition of the relativistic kinetic energy, which dates back to a 1917 paper [61].

It looks as if all of this, including the use of the term "relativistic mass," reflected Einstein's views.

1945

On August 6, 1945, an atomic bomb was dropped on Hiroshima; another was dropped on Nagasaki on August 9.

In September, the British magazine 'Discovery' published photographs of the first atomic test explosion on July 16, 1945 and two papers, "The Progress of Science — We enter the New Age" and "The Science behind the Atomic Bomb." The latter mentioned, in the chronology of the atomic physics discoveries, "1905. Einstein's special relativity theory demonstrated the equivalence of mass and energy." However, Einstein's photograph was not among the 25 accompanying portraits of scientists from Becquerel to Oppenheimer [96].

In September 1945, the book "Atomic energy for military purposes" by H D Smyth was published [97]. The Introduction said, in the section "Conservation of mass and of energy":

"1.2 There are two principles that have been cornerstones of the structure of modern science. The first — that matter can be neither created nor destroyed but only altered in form — was enunciated in the eighteenth century and is familiar to every student of chemistry; it has led to the principle known as the law of conservation of mass. The second — that energy can be neither created nor destroyed but only altered in form emerged in the nineteenth century...; it is known as the law of conservation of energy.

1.3 ...but it is now known that they are, in fact, two phases of a single principle for we have discovered that energy may sometimes be converted into matter and matter into energy."

The section "Equivalence of mass and energy" said this:

"1.4 One conclusion that appeared rather early in the development of the theory of relativity was that the inertial mass of a moving body increased as its speed increased. This implied an equivalence between an increase in energy of motion of a body, that is, its kinetic energy, and an increase in its mass. ...He [Einstein] concluded that the amount of energy, E, equivalent to a mass, m, was given by the equation

$$E = mc^2 \,,$$

where c is the velocity of light. If this is stated in actual numbers, its startling character is apparent."

In these passages, the following deserves our attention:

1. Matter is identified with mass.
2. The law of conservation of momentum is not mentioned, although mass conservation cannot be understood without it.
3. It is stated that mass increases with velocity.
4. The rest energy and the formula $E_0 = mc^2$ are not mentioned.

We also note that H D Smyth was the chairman of the physics department of Princeton University.

1946

On July 1, 1946, *Time* magazine had Einstein's portrait on its front cover against the background of a nuclear mushroom cloud with $E = mc^2$ written on it [88].

In 1946, Einstein published two papers on the equivalence of mass and energy: "Elementary derivation of the equivalence of mass and energy" [98] and "$E = mc^2$: The most urgent problem of our time" [99].

In the first of them, he partly changed the proof given in 1905 [17]: the body at rest does not emit radiation but absorbs it; he uses the formulas of conservation not of energy but of momentum; formulas for the transformation of energy and momentum of radiation are not used, but instead Einstein uses the known angle of aberration of stellar light caused by the motion of the Earth: $\alpha = v/c$. As a result, Einstein obtains the increment to the mass of the body $M' - M = E/c^2$, where E is the energy of the absorbed radiation, and concludes: "This equation expresses the law of the equivalence of energy and mass. The energy increase E is connected with the mass increase E/c^2. Since energy according to the usual definition leaves an additive constant free, we may so choose the latter that $E = Mc^2$."

It is obvious from the derivation that E here stands for the rest energy of the body.

Einstein does not explain why the rest energy of the body is defined up to a constant.

In his brief popular-style article [99], Einstein first described the law of the conservation of energy using the kinetic and potential energy of a pendulum as an example, and then proceeded to deal with the conservation of mass:

"Now for the principle of the conservation of mass. Mass is defined by the resistance that a body opposes to its acceleration (inert mass). It is also measured by the weight of the body (heavy mass). That these two radically different definitions lead to the same value for the mass of a body is, in itself, an astonishing fact. According to the principle — namely, that masses remain unchanged under any physical or chemical changes — the mass appeared to be the essential (because unvarying) quality of matter. Heating, melting, vaporization, or combining into chemical compounds would not change the total mass.

Physicists accepted this principle up to a few decades ago. But it proved inadequate in the face of the special theory of relativity. It was therefore merged with the energy principle — just as, about 60 years before, the principle of the conservation of mechanical energy had been combined with the principle of the conservation of heat. We might say that the principle of the conservation of energy, having previously swallowed up that of the conservation of heat, now proceeded to swallow that of the conservation of mass — and holds the field alone.

It is customary to express the equivalence of mass and energy (though somewhat inexactly) by the formula $E = mc^2 \ldots$"

What deserves our attention in this passage is not only what Einstein clarified but also what he chose not to explain: namely, that the measure of inertia in relativity theory is not mass but energy and that the quantity $p_\mu p_\nu / E$ creates and feels the gravitational field (and therefore there is nothing surprising about the equality between the inertial mass and the gravitational mass in Newtonian mechanics: both are equal to E_0/c^2), that the principle of energy conservation holds the field not alone but together with the conservation of momentum, that the energy and momentum determine the mass and its conservation and/or nonconservation jointly, and that the mass is equivalent to the rest energy.

In 1949 Einstein published "Autobiographical Notes" [100], which open with these words: "Here I sit in order to write, at the age of 67, something like my own obituary." So in fact he was writing them in 1946–1947. In these notes, Einstein made an attempt to tell us what and how he had been thinking about for many years: "For me it is not dubious that our thinking goes on for the most part without use of signs (words) and beyond that to a considerable degree unconsciously."

On the creation of general relativity he wrote:

"The possibility of the realization of this program was, however, dubious from the very first, because the theory had to combine the following things:

(1) From the general considerations of special relativity theory it was clear that the *inert* mass of a physical system increases with the total energy (therefore, e.g., with the kinetic energy).

(2) From very accurate experiments (specially from the torsion balance experiments of Eötvös) it was empirically known with very high accuracy that the gravitational mass of a body is exactly equal to its *inert* mass."

These words can be interpreted, if one so wishes, as a statement that the formula $E = mc^2$ not only follows from special (partial) relativity theory but is also the cornerstone of general relativity.

1948

In June 1948, Einstein wrote about the thorny question of mass for the last time. In a letter to L Barnett, author of the book "The Universe and Dr. Einstein", he wrote [101]:

"It is not good to introduce the concept of *mass* $M = m/\sqrt{1 - v^2/c^2}$ of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the "rest mass" m. Instead of introducing M it is better to mention the expression for the momentum and energy of a body in motion."

12. 1949. Feynman diagrams

In 1949, Feynman published "The theory of positrons" [102] and "Space–time approach to quantum electrodynamics" [103]. These papers put quantum electrodynamics into a form that was completely compatible with the symmetry of the Minkowski world. In these papers, he formulated and developed a method known as Feynman diagrams.

The external lines of the diagrams correspond to real on-shell particles: for them, $p^2 = m^2$, where p is the 4-momentum of a particle and m is its mass. The internal lines correspond to virtual particles that are off-shell: for these, $p^2 \neq m^2$. Antiparticles look like particles that move backwards in time. All particles — both massive and massless — are described in the same manner, with a single difference: $m = 0$ is assumed for the latter. (Virtual photons with positive p^2 are called timelike, and those with negative p^2, spacelike.) It goes without saying that the Feynman diagram method is based on the concept of invariant mass m that is independent of the velocity of the particle.

Feynman diagrams drastically simplified calculations for processes involving elementary particles. They unified all types of matter, both for real particles and for virtual ones that replaced fields.

F Dyson, who at the time worked with Feynman, recently recalled [104]:

"During the time that the young physicists at the Institute for Advanced Study in Princeton were deeply engaged in developing the new electrodynamics, Einstein was working in the same building and walking every day past our windows on his way to and from the Institute. He never came to our seminars and never asked us about our work. To the end of his life, he remained faithful to his unified field theory."

We know that his famous aphorism—"God is subtle but He is not malicious"—was engraved above the fireplace at the Institute for Advanced Study where Einstein worked. One cannot help recalling his other pronouncement: "I have second thoughts. Maybe God *is* malicious" [105].

13. 1952 – 1955. Last years

1952

In 1952, Einstein published a new edition of his popular-science book "Relativity, The Special and the General Theory, A Popular Exposition" [106], was first published in 1917 [61]. For this new edition, he wrote a special appendix, entitled "Relativity and the problem of space," in order "to show that space-time is not necessarily something to which one can ascribe a separate existence, independently of the actual objects of physical reality... In this way the concept of 'empty space' loses its meaning." With these words, Einstein was referring not only to general relativity but also to special relativity theory. The concept of virtual particles was perhaps alien to him.

1954

Einstein's foreword to Jammer's book "Concepts of space" [107] may contain a clue to what prevented Einstein from regarding the photon as a material object: "Now as to the concept of space, it seems that this was preceded by the psychologically simpler concept of place. Place is first of all a (small) portion of the earth's surface identified by a name. The thing whose 'place' is being specified is a 'material object' or body."

From this standpoint, any particle, no matter how light, is a material object while a strictly massless particle is not.

1955

In 1895, the 16-year-old Einstein wrote his first scientific essay [108] on the propagation of light through the ether.

In 1955, in his last autobiographic notes [109], he recalled that at that time a thought experiment started to puzzle him:

"If one were to pursue a light wave with the velocity of light, one would be confronted with a time independent wave field. Such a thing doesn't seem to exist, however! This was the first childlike thought-experiment concerned with the special theory of relativity."

Thought experiments played an important role in Einstein's research during all his life.

Einstein died on April 18, 1955. A month before his death, Leopold Infeld gave a talk in Berlin at a meeting that celebrated the 50th anniversary of relativity theory [110]. He named the dependence of mass on velocity as the first of the three experimental confirmations of special relativity theory. The baton of "relativistic mass" was passed on to future generations.

14. Born, Landau, Feynman

Born's books

An important role in this passing of the baton belongs to Max Born's book "Einstein's theory of relativity." An outstanding physicist, one of the creators of quantum mechanics, Born did very much to help spread relativity theory. The first edition of his book appeared in 1920 [111] (its Russian edition was published in 1938 [112]). The next edition [113] appeared after Einstein's death in 1962 (and its translation into Russian [114] in 1964 and 1972). Unfortunately, both these editions, which greatly influenced how physics was taught in the 20th century, state without any qualifications that the increase in the mass of a body when its velocity increases is an experimental fact. It is also asserted in [115, 116].

In 1969 — a year before passing away — Born published his correspondence with Einstein [117], which lasted from 1916 till 1955. Not even one among more than a hundred letters touches on the aspect of the [in]dependence of mass on velocity. The correspondence was translated into English, its latest edition was published in 2005 [118] with a detailed foreword, which also ignored the mass controversy.

Landau and Rumer brochure

I mentioned above that the Landau – Lifshitz book "Field theory" [93] was the first monograph on relativity theory in world literature that consistently applied the idea that the mass of a body is independent of its velocity. It is all the more incomprehensible why in their popular brochure "What is relativity theory?" [119, 120], Landau and Rumer chose for the first introduction into the theory the statement that mass is a function of velocity and that this is an experimental fact. In the third edition of this brochure published in 1975, Yu B Rumer added "Pages of reminiscences about L D Landau," where he quoted a jocular characteristic of the brochure given by Landau himself: "Two con men trying to persuade the third one that for the price of a dime he would understand what relativity theory is."

The Feynman Lectures

The magnificent lectures on physics that Feynman gave to students of Caltech in 1961 – 1964 [121] instilled a love for physics in the hearts of millions of readers around the world (see, e.g. [122]). They teach readers to think independently and honestly. Alas, these lectures never mention the Feynman diagrams that he invented in 1949 [102, 103] and which brought him the Nobel prize in 1965. Furthermore, the entire relativity theory is introduced in these lectures through the formula $E = mc^2$, not through the concept of the Lorentz-invariant mass on which Feynman diagrams are based.

Feynman states already in the first chapter that the dependence of the mass of a body on its velocity is an experimental fact, in the fourth he says that Einstein discovered the formula $E = mc^2$, and in the seventh that mass is the measure of inertia. In chapter 15, we meet the formula $m = m_0/\sqrt{1 - v^2/c^2}$ and Feynman discusses the consequences of the "relativistic increase of mass"; in chapter 16, he derives this formula. This chapter ends with the words:

"That the mass in motion at speed v is the mass m_0 at rest divided by $\sqrt{1 - v^2/c^2}$, surprisingly enough, is rarely used. Instead, the following relations are easily proved, and turn out to be very useful: $E^2 - P^2c^2 = M_0{}^2c^4$ and $Pc = Ev/c$." (The original notation used by Feynman is retained in this quotation.)

Even in Chapter 17, where Feynman introduces four-dimensional space – time and uses units in which $c = 1$, he continues to speak of the rest mass m_0, not simply of the mass m.

In the course of 2007, I e-mailed a question to a number of Feynman's former students, assistants and co-authors. Not one of them was able to recall even a single occasion when Feynman used the notion of relativistic mass or the formula $E = mc^2$ in discussions he had with them. Nevertheless, several millions of readers of his lectures firmly believe that mass is a function of velocity. Why would the great physicist who gave us the language of Feynman diagrams place the notion of velocity-dependent mass at the foundation of his Feynman lectures?

Perhaps we can find an answer to this question in Feynman's Nobel lecture [123]. He described there numerous 'blind alleys' in which he had been trapped while on his way to constructing quantum electrodynamics, but still expressed the firm belief that "many different physical ideas can describe the same physical reality."

Thus he wrote about the idea of an electron moving backwards in time: "it was very convenient, but not strictly necessary for the theory because it is exactly equivalent to the negative energy sea point of view." However, without a time-reversed electron, there would be no Feynman diagrams, which introduced order and harmony into huge areas of physics.

15. Epilogue

Why is it that the weed of velocity-dependent mass is so resistant? First and foremost, because it does not lead to immediate mistakes as far as arithmetic or algebra are concerned. One can introduce additional 'quasi-physical variables' into any self-consistent theory by multiplying true physical quantities by arbitrary powers of the speed of light. The most striking example of such a 'quasi-quantity' is the so-called 'relativistic mass.' If calculations are done carefully enough, their results should be the same as in the original theory. In a higher sense, however, after the introduction of such 'quasi-quantities,' the theory is mutilated because its symmetry properties are violated. (For example, the relativistic mass is only one component of a 4-vector, while the other three components are not even mentioned.)

Some other explanations of the longevity of relativistic mass can be given here. The formula $E = mc^2$ is 'simpler' than the formula $E_0 = mc^2$ because the additional zero subscript that requires explanation is dropped. The energy divided by c^2 indeed has the dimensionality of mass. Intuition based on conventional everyday experience slips in a hint that the measure of inertia of a body is its mass, not its energy, and this prods one to 'drag' the nonrelativistic formula $\mathbf{p} = m\mathbf{v}$ into relativity theory. The same intuition suggests, with hardly less insistence, that the source of gravitation is 'our own' mass, not an 'alien' quantity $p_\mu p_\nu / E$. Everyday experience rebels particularly strongly against the idea of treating light as a type of matter.

The arguments given above may explain the 'Newtonian bias' of an ordinary person, let us say 'a pedestrian.' However, it would be too flippant to attribute them to such a great physicist as Einstein. Indeed, it was Einstein who introduced the concept of rest energy E_0 into physics and wrote about $E_0 = mc^2$ far more often than about $E = mc^2$. Still, one thing remains unexplained: why was it that during the half-century of discussing the relation between mass and energy, Einstein never once referred in either his research publications or his letters to the formula $E^2 - \mathbf{p}^2 c^2 = m^2 c^4$, which defines the Lorentz-invariant mass?

It is possible that the formulation of the total equivalence of energy and mass reflected Einstein's absolute reliance on his powerful intuition. It was without a doubt his confidence in his own intuition that resulted in his rejection of quantum mechanics. One feels that he perceived the concept of electromagnetic potential not only with his mind but with his entire body. And he 'felt' the wave function to be very much like the electromagnetic wave. His resistance to quantum mechanics prevented Einstein's world line from meeting Feynman's world line in the space of ideas — in the noosphere, so to speak. As a consequence, Einstein refused to accept the photon as a particle of matter and continued to treat it as a quantum of energy.

16. Conclusion

When shown an art exhibition in Moscow Manege in 1962, Nikita Khrushchev (1894–1971) rudely attacked the sculptures of Ernst Neizvestny. When Khrushchev died, his children requested Neizvestny to create a memorial sculpture at the grave of their father. The main part of this memorial consists of two vertical marble slabs, one white, the other black, whose protrusions penetrate each other. These slabs in a way symbolize the good and the evil.

The history of the confrontation of two concepts of mass in the 20th century resembles this sculpture. Here the light and the darkness were fighting each other in the minds of the creators of modern physics.

In the world of opinions, pluralism is considered to be politically correct. To insist on a single point of view is thought to be a manifestation of dogmatism. A good example of fruitful pluralism is the wave–particle duality in quantum mechanics. But there are cases in which a situation is ripe for establishing unambiguous terminology. The relation between energy and mass is more than ripe for this. It is high time we stopped deceiving new generations of college and high school students by inculcating into them the conviction that mass increasing with increasing velocity is an experimental fact.

Postscriptum. In memory of J A Wheeler

This review was already completed when I received the sad news that John Archibald Wheeler, an outstanding physicist and teacher who accomplished so much for establishing the spacetime interpretation of relativity theory and of the concept of Lorentz-invariant mass, died on 13 April 2008, at the age of 96. I dedicate this paper to his memory.

Acknowledgments. I am grateful to B L Okun, M B Voloshin, and V I Kisin for their invaluable help in preparing this paper. I also appreciate helpful discussions with A A Abrikosov, M S Aksent'eva, A A Alekhina, B L Altshuller, J M Bardeen, T Basaglia, S M Berman, S I Blinnikov, B M Bolotovsky, L M Brown, D K Buchwald, T L Curtright, Yu Danoyan, A D Dolgov, M A Gottlieb, M D Godfrey, Ya I Granovsky, E G Gulyaeva, D R Hofstadter, J D Jackson, M Janssen, C Jarlskog, O V Kancheli, M Karliner, V I Kogan, Ya S Kim, C Quigg, G L Landsberg, L Yu Mizrahi, V A Novikov, L I Ponomarev, P S Prokof'ev, F Ravndal, M Sands, S G Tikhodeev, K A Tomilin, V P Vizgin, M I Vysotsky, V R Zoller, G Zweig. The work was supported by the grants NSh-5603.2006.2, NSh-4568.2008.2, RFBR-07-02-00830-a.

References [2]

1. Einstein A *Sobranie Nauchnykh Trudov A. Einsteina* (The Collection of Scientific Works of Albert Einstein), Vol. 1–4 (Eds I E Tamm, Ya A Smorodinskii, B G Kuznetsov) (Moscow: Nauka, 1965–1967)
2. Einstein A *The Collected Papers of Albert Einstein* Vols 1 – 10 (Eds J Stachel (Vols 1, 2), D C Cassidy (Vols 1, 2), R Schulman (Vols 1 – 8), J Renn (Vols 2–4), M J Klein (Vols 3–6), A J Kox (Vols 3–6, 8), M Janssen (Vols 7, 8), J Illy (Vols 7 – 10), C Lehner (Vol. 7), D K Buchwald (Vol. 10), T Sauer (Vol. 10), Z Rosenkranz (Vol. 10), V I Holmes (Vol. 10)) (Princeton, NJ: Princeton Univ. Press, 1987 – 2006)
3. Einstein A *The Collected Papers of Albert Einstein* Vols 1 – 10 (Translators A Beck (Vols 1 – 3, 5), D Howard (Vol. 4), A Engel (Vols 6, 7), A Hentschel (Vols 8 – 10)) (Princeton, NJ: Princeton University Press, 1987 – 2006)
4. Galilei G *Dialogo sopra i due massimi sistemi del mondo Tolemaico e Copernicano* (Fiorenza, Per Gio: Batista Landini, 1632) [Translated into English: *Dialogue Concerning the Two Chief World Systems, Ptolemaic and Copernican* (Berkeley, Calif.: Univ. of California Press, 1967); Translated into Russian: *Dialog o Dvukh Glavneishikh Sistemakh Mira i Kopernikovoi* (Translated by A I Dolgov) (Moscow – Leningrad: OGIZ, 1948)]
5. Newton I *Philosophiae Naturalis Principia Mathematica* (Londini: Jussu Societatis Regiae, 1687) [Translated into English: *The Principia: Mathematical Principles of Natural Philosophy* (Berkeley, Calif.: Univ. of California Press, 1999); Translated into Russian: *Matematicheskie Nachala Natural'noi Filosofii* (Translated from Latin and comments by A N Krylov) (Moscow: Nauka, 1989)]
6. Lorentz G A, Poincaré H, Einstein A, Minkowski H *Printsip Otnositel'nosti. Sbornik Rabot Klassikov Relyativizma* (Principle of Relativity. Collected Papers of Classics of Relativism) (Eds V K Frederiks, D D Ivanenko) (Moscow - Leningrad: ONTI, 1935)
7. Thomson J "On the electric and magnetic effects produced by the motion of electrified bodies" *Philos. Mag.* **11** (8) 229–249 (1881)
8. Heaviside O "On the electromagnetic effects due to the motion of electrification through a dielectric" *Philos. Mag.* **27** 324–339 (1889)
9. Searle G "On the Steady Motion of an Electrified Ellipsoid" *Philos. Mag.* **44** 329 – 341 (1897)
10. Abraham M "Prinzipien der Dynamik des Elektrons" *Ann. Phys.* (Leipzig) **10** [**315**] (1) 105 – 179 (1902) [2]
11. Lorentz H "Electromagnetic phenomena in a system moving with any velocity smaller than that of light" *Proc. Acad. of Sci., Amsterdam* **6** 809–831 (1904) [Translated into Russian: "Elektromagnitnye yavleniya v sisteme, dvizhushcheisya s lyuboi skorost'yu, men'shei skorosti sveta", in *Printsip Otnositel'nosti. Sbornik Rabot Klassikov Relyativizma* (Principle of Relativity. Collected Papers of Classics of Relativism) (Eds V K Frederiks, D D Ivanenko) (Moscow–Leningrad: ONTI, 1935) p. 16]
12. Poincaré H "La théorie de Lorentz et le principe de réaction" *Arch. Néerland. Sci. Exactes Naturelles* **5** 252–278 (1900)
13. Kaufmann W "Die elektromagnetische Masse des Elektrons" *Phys. Z.* **4** 54–57 (1902)
14. Bucherer A "On the principle of relativity and on the electromagnetic mass of the electron. A Reply to Mr. E. Cunningham" *Philos. Mag.* **15** 316–318 (1908)
15. Bucherer A "Messungen an Becquerelstrahlen. Die experimentelle Bestätigung der Lorentz–Einsteinschen Theorie" *Phys. Z.* **9** 755–762 (1908)
16. Einstein A "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt" *Ann. Phys.* (Leipzig) **17** [**322**] 132–148 (1905) [Translated into English "On a heuristic point of view concerning the production and transformation of light", in *CPAET* Vol. 2, doc. 14; Translated into Russian: "Ob odnoi evristicheskoi tochke zreniya, kasayushcheisya vozniknoveniya i prevrashcheniya sveta", in *SNT* Vol. 3, p. 92]

[2] For most references, I tried to provide bibliographical information about the original source, and its English and Russian translations. In what follows, *Sbornik Nauchnykh Trudov* (*SNT*) refers to [1], *CPAE* refers to [2], and *CPAET* refers to [3]. For the journal *Annalen der Physik*, the volume number in square brackets refers to the new numbering scheme of the publisher according to the Wiley InterScience website.

17. Einstein A "Zur Elektrodynamik bewegter Körper" *Ann. Phys.* (Leipzig) **17** [**322**] 891–921 (1905) [Translated into English: "On the electrodynamics of moving bodies", in *CPAET* Vol. 2, doc. 23; Translated into Russian: "K elektrodinamike dvizhushchikhsya tel", in *SNT* Vol. 1, p. 7]
18. Einstein A "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" *Ann. Phys.* (Leipzig) **18** [**323**] 639–641 (1905) [Translated into English: "Does the inertia of a body depend on its energy content?", in *CPAET* Vol. 2, doc. 24; Translated into Russian: "Zavisit li inertsiya tela ot soderzhashcheisya v nem energii?", in *SNT* Vol. 1, p. 36]
19. Planck M "Über irreversible Strahlungsvorgänge" *Ann. Phys.* (Leipzig) **1** [**306**] 69–122 (1900) [Translated into Russian: "O neobratimykh protsessakh izlucheniya" ("On irreversible radiation processes"), in Planck M *Izbrannye Trudy* (Selected Works) (Moscow: Nauka, 1975) p. 191]
20. Einstein A "Letter to Conrad Habicht, 30 June–22 September 1905", in *CPAET* Vol. 5, doc. 28
21. Seelig C *Albert Einstein. Leben und Werk eines Genies unserer Zeit* (Zürich: Europa Verlag, 1952) [Translated into English: *Al'bert Einstein* (abridged translation from German) (Moscow: Atomizdat, 1964) p. 65; 2nd ed. (Moscow: Atomizdat, 1966) p. 73]
22. Einstein A "Das Prinzip von der Erhaltung der Schwerpunktsbewegung und die Trägheit der Energie" *Ann. Phys.* (Leipzig) **20** [**325**] 627–633 (1906) [Translated into English: "The principle of conservation of motion of the center of gravity and the inertia of energy", in *CPAET* Vol. 2, doc. 35; Translated into Russian: "Zakon sokhraneniya dvizheniya tsentra tyazhesti i inertsiya energii", in *SNT* Vol. 1, p. 39]
23. Einstein A "Über eine Methode zur Bestimmung des Verhältnisses der transversalen und longitudinalen Masse des Elektrons" *Ann. Phys.* (Leipzig) **21** [**326**] 583–586 (1906) [Translated into English: "On a method for the determination of the ratio of the transverse and the longitudinal mass of the electron", in *CPAET* Vol. 2, doc. 36; Translated into Russian: "O metode opredeleniya sootnosheniya mezhdu poperechnoi i prodol'noi massami elektrona", in *SNT* Vol. 1, p. 45]
24. Einstein A "Über die Möglichkeit einer neuen Prüfung des Relativitätsprinzips" *Ann. Phys.* (Leipzig) **23** [**328**] 197–198 (1907) [Translated into English: "On the possibility of a new test of the relativity principle", in *CPAET* Vol. 2, doc. 41; Translated into Russian: "O vozmozhnosti novogo dokazatel'stva printsipa otnositel'nosti", in *SNT* Vol. 1, p. 49]
25. Einstein A "Bemerkungen zu der Notiz von Hrn. Paul Ehrenfest: 'Die Translation deformierbarer Elektronen und der Flächensatz'" *Ann. Phys.* (Leipzig) **23** [**328**] 206–208 (1907) [Translated into English: "Comments on the note by Mr. Paul Ehrenfest: 'The translatory motion of deformable electron and the area law'", in *CPAET* Vol. 2, doc. 44; Translated into Russian "Po povodu zametki Paulya Erenfesta 'Postupatelnoe dvizhenie deformiruemykh elektronov i teorema ploshchadei'", in *SNT* Vol. 1, p. 51]
26. Einstein A "Über die vom Relativitätsprinzip geforderte Trägheit der Energie" *Ann. Phys.* (Leipzig) **23** [**328**] 371–384 (1907) [Translated into English: "On the inertia of energy required by the relativity principle", in *CPAET* Vol. 2, doc. 45; Translated into Russian: "Ob inertsii energii, trebuemoi printsipom otnositel'nosti", in *SNT* Vol. 1, p. 53
27. Einstein A "Über das Relativitätsprinzip und die aus demselben gezogene Folgerungen" *Jahrbuch Radioaktivität Elektron.* **4** 411–462 (1907) [Translated into English: "On the relativity principle and the conclusions drawn from it", in *CPAET* Vol. 2, doc. 47; Translated into Russian: "O printsipe otnositel'nosti i ego sledstviyakh", in *SNT* Vol. 1, p. 65]
28. Einstein A "Berichtigung zu der Arbeit: 'Über das Relativitätsprinzip und die aus demselben gezogene Folgerungen'" *Jahrbuch Radioaktivität Elektron.* **5** 98–99 (1908) [Translated into English: "Corrections to the paper 'On the relativity principle and the conclusions drawn from it'", in *CPAET* Vol. 2, doc. 49]
29. Einstein A, Laub J "Über die elektromagnetischen Grundgleichungen für bewegte Körper" *Ann. Phys.* (Leipzig) **26** [**331**] 532–540 (1908) [Translated into English: "On the fundamental electromagnetic equations for moving bodies", in *CPAET* Vol. 2, doc. 51;

Translated into Russian: "Ob osnovnykh elektrodinamicheskikh uravneniyakh dvizhushchegosya tela", in *SNT* Vol. 1, p. 115]

30. Einstein A, Laub J "Über die im elektromagnetischen Felde auf ruhende Körper ausgeübten ponderomotorischen Kräfte" *Ann. Phys.* (Leipzig) **26** **[331]** 541–550 (1908) [Translated into English: "On the ponderomotive forces exerted on bodies at rest in the electromagnetic field", in *CPAET* Vol. 2, doc. 52; Translated into Russian: "O ponderomotornykh silakh, deistvuyushchikh v elektromagnitnom pole na pokoyashchiesya tela", in *SNT* Vol. 1, p. 126]
31. Einstein A, Laub J "Berichtigung zur Abhandlung: 'Über die elektromagnetischen Grundgleichungen für bewegte Körper'" *Ann. Phys.* (Leipzig) **27** **[332]** 232–232 (1908) [Translated into English: "Correction to the paper: 'On the fundamental electromagnetic equations for moving bodies'", in *CPAET* Vol. 2, doc. 53; Included in *SNT* Vol. 1, p. 115]
32. Einstein A, Laub J "Bemerkungen zu unserer Arbeit: 'Über die elektromagnetischen Grundgleichungen für bewegte Körper'" *Ann. Phys.* (Leipzig) **28** **[333]** 445–447 (1909) [Translated into English: "Remarks on our paper 'On the fundamental electromagnetic equations for moving bodies'", in *CPAET* Vol. 2, doc. 54; Translated into Russian: "Zamechaniya k nashei rabote 'Ob osnovnyh elektrodinamicheskih uravneniyah dvizhuschegosya tela'", in *SNT* Vol. 1, p. 123]
33. Minkowski H "Raum und Zeit" *Phys. Z.* **10** 104–111 (1909) [Translated into English: "Space and time", in: Einstein A, Lorentz H A, Minkowski H, Weyl H *The Principle of Relativity: a Collection of Original Memoirs on the Special and General Theory of Relativity* (London: Methuen, 1923); reprinted (New York: Dover, 1952); Translated into Russian: Minkowski H "Prostranstvo i vremya" (Space and time) *Usp. Fiz. Nauk* **69** 303 (1959); see also: Lorentz G A, Poincaré A, Einstein A, Minkowski H *Printsip Otnositel'nosti. Sbornik Rabot Klassikov Relyativizma* (Principle of Relativity. Collected Papers of Classics of Relativism) (Eds V K Frederiks, D D Ivanenko) (Moscow–Leningrad: ONTI, 1935) p. 181]
34. Lewis G "A revision of the fundamental laws of matter and energy" *Philos. Mag.* **16** 705–717 (1908)
35. Lewis G, Tolman R "The principle of relativity & non-newtonian mechanics" *Philos. Mag.* **17** 510–523 (1909)
36. Tolman R "Note on the derivation from the principle of relativity of the fifth fundamental equation of the Maxwell–Lorentz theory" *Philos. Mag.* **21** 296–301 (1911)
37. Tolman R "Non-newtonian mechanics: The direction of force and acceleration" *Philos. Mag.* **22** 458–463 (1911)
38. Tolman R "Non-newtonian mechanics — the mass of a moving body" *Philos. Mag.* **23** 375–380 (1912)
39. Einstein A "Bemerkung zu der Arbeit von D. Mirimanoff: 'Über die Grundgleichungen...'" *Ann. Phys.* (Leipzig) **28** **[333]** 885–888 (1909) [Translated into English: "Comment on the paper of D Mirimanoff 'On the fundamental equations...'", in *CPAET* Vol. 2, doc. 55; Translated into Russian: "Zamechanie k rabote Mirimanova 'Ob osnovnykh uravneniyakh...'", in *SNT* Vol. 1, p. 135]
40. Einstein A "Zum gegenwärtigen Stand des Strahlungsproblems" *Phys. Z.* **10** 185–193 (1909) [Translated into English: "On the present status of the radiation problem", in *CPAET* Vol. 2, doc. 56; Translated into Russian: "K sovremennomu sostoyaniyu problemy izlucheniya", in *SNT* Vol. 3, p. 164]
41. Ritz W, Einstein A "Zum gegenwärtigen Stand des Strahlungsproblems" *Phys. Z.* **10** 323–324 (1909) [Translated into English: "On the present status of the radiation problem", in *CPAET* Vol. 2, doc. 57; Translated into Russian: "K sovremennomu sostoyaniyu problemy izlucheniya", in *SNT* Vol. 3, p. 180]
42. Einstein A "Über die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung" *Verhandl. Deutsche Phys. Gesellschaft* **11** 482–500 (1909) [Translated into English: "On the development of our views concerning the nature and constitution of radiation", in *CPAET* Vol. 2, doc. 60; Translated into Russian: "O razvitii nashikh vzglyadov na sushchnost' i strukturu izlucheniya", in *SNT* Vol. 3, p. 181]
43. Einstein A, Hopf L "Über einen Satz der Wahrscheinlichkeitsrechnung und seine Anwendung in der Strahlungstheorie" *Ann. Phys.* (Leipzig) **33** **[338]** 1096–1104 (1910) [Translated into English: "On a theorem of the probability calculus and its application in the theory of radiation", in *CPAET* Vol. 3, doc. 7; Translated into Russian: "Ob odnoi teoreme teorii veroyatnostei i ee primenenii v teorii izlucheniya", in *SNT* Vol. 3, p. 196]
44. Einstein A, Hopf L "Statistische Untersuchung der Bewegung eines Resonators in einem Strahlungsfeld" *Ann. Phys.* (Leipzig) **33** **[338]** 1105–1115 (1910) [Translated into English: "Statistical investigation of a resonator's motion in a radiation field", in *CPAET* Vol. 3, doc. 8; Translated into Russian: "Statisticheskoe issledovanie dvizheniya rezonatora v pole izlucheniya", in *SNT* Vol. 3, p. 205]
45. Einstein A "Le principe de relativité et ses conséquences dans la physique moderne" *Arch. Sci. Phys. Naturelles* **29** 5–28 (1910) [Translated into English: "The principle of relativity and its consequences in modern physics", in *CPAET* Vol. 3, doc. 2; Translated into Russian: "Printsip otnositel'nosti i ego sledstviya v sovremennoi fizike", in *SNT* Vol. 1, p. 138]
46. Einstein A "Über den Einfluβ der Schwerkraft auf die Ausbreitung des Lichtes" *Ann. Phys.* (Leipzig) **35** **[340]** 898–908 (1911) [Translated into English: "On the influence of gravitation on the propagation of light", in *CPAET* Vol. 3, doc. 23; Translated into Russian: "O vliyanii sily tyazhesti na rasprostranenie sveta", in *SNT* Vol. 1, p. 165]
47. Einstein A "Die Relativitätstheorie" *Vierteljahrsschrift Naturforsch. Gesellschaft* (Zürich) **56** 1–14 (1911) [Translated into English: "The theory of relativity", in *CPAET* Vol. 3, doc. 17; Translated into Russian: "Teoriya otnositel'nosti", in *SNT* Vol. 1, p. 175]
48. Einstein A "Zum Ehrenfestschen Paradoxon. Bemerkung zu V. Varicak's Aufsatz" *Phys. Z.* **12** 509–510 (1911) [Translated into English: "On the Ehrenfest paradox. Comment on V. Varicak's paper", in *CPAET* Vol. 2, doc. 23; Translated into Russian: "K paradoksu Erenfesta", in *SNT* Vol. 1, p. 187]
49. von Soldner J "Über die Ablenkung eines Lichtstrahls von seiner geradlinigen Bewegung durch die Attraktion eines Weltkörpers an welchem er nahe vorbeigeht", in *Astronomisches Jahrbuch für das Jahr 1804* (Berlin, 1801) pp. 161–172 [Translated into English: "On the deviation of a light ray from its motion along straight line through the attraction of a celestial body which it passes close by" *Found. Phys.* **8** 939–950 (1978)]
50. Jaki S "Johan Georg von Soldner and the gravitational bending of light, with an english translation of his essay on it published in 1801" *Found. Phys.* **8** 927–938 (1978)
51. Einstein A "Manuscript on the special theory of relativity (1912–1914)", in *CPAET* Vol. 4, doc. 1
52. Einstein A "Lichtgeschwindigkeit und Statik des Gravitationsfeldes" *Ann. Phys.* (Leipzig) **38** **[343]** 355–369 (1912) [Translated into English: "The speed of light and the statics of the gravitational field", *CPAET* Vol. 4, doc. 3; [Translated into Russian: "Skorost' sveta i staticheskoe gravitatsionnoe pole", in *SNT* Vol. 1, p. 189]
53. Einstein A "Zur Theorie des statischen Gravitationsfeldes" *Ann. Phys.* (Leipzig) **38** **[343]** 443–458 (1912) [Translated into English: "On the theory of the static gravitational field", in *CPAET* Vol. 4, doc. 4; Translated into Russian: "K teorii staticheskogo gravitatsionnogo polya", in *SNT* Vol. 1, p. 202]
54. Einstein A "Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham" *Ann. Phys.* (Leipzig) **38** **[343]** 1059–1064 (1912) [Translated into English: "Relativity and gravitation. Reply to a comment by M. Abraham", in *CPAET* Vol. 4, doc. 8; Translated into Russian: "Otnositel'nost' i gravitatsiya. Otvet na zamechanie M. Abrahama", in *SNT* Vol. 1, p. 217]
55. Einstein A "Gibt es eine Gravitationswirkung die der elektrodynamischen Induktionswirkung analog ist?" *Vierteljahrsschrift gericht. Med. offentliches Sanitätswesen* **44** 37–40 (1912) [Translated into English: "Is there a gravitational effect which is analogous to electrodynamic induction?", in *CPAET* Vol. 4, doc. 7; Translated into Russian: "Sushchestvuet li gravitatsionnoe vozdeistvie, analogichnoe elektrodinamicheskoi induktsii?", in *SNT* Vol. 1, p. 223]
56. Einstein A, Grossmann M "Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation" *Z. Math. Phys.* **62** 225–261 (1913) [Translated into English: "Outline of a generalized theory of relativity and of a theory of gravitation", in *CPAET* Vol. 4, doc. 13; Translated into Russian: "Proekt obobschennoi teorii otnositel'nosti i teorii tyagoteniya", in *SNT* Vol. 1, p. 227]

57. Einstein A "Nachtragliche Antwort auf eine Frage von Reißner" *Phys. Z.* **15** 108–110 (1914) [Translated into English: "Supplementary response to a question by Mr. Reißner", in *CPAET* Vol. 4, doc. 24; Translated into Russian: "Dopolnitel'nyi otvet na vopros Reisnera", in *SNT* Vol. 1, p. 229]
58. Einstein A "Lecture notes for course on relativity at the University of Berlin, winter semester 1914/1915", in *CPAET* Vol. 6, doc. 7
59. Einstein A "Die Feldgleichungen der Gravitation" *Sitzungsber. Königl. Preuß. Akad. Wissenschaft.* (Berlin) **48** 844–847 (1915) [Translated into English: "The field equations of gravitation", in *CPAET* Vol. 6, doc. 25; Translated into Russian: "Uravnenie gravitatsionnogo polya", in *SNT* Vol. 1, p. 448]
60. Einstein A "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie" *Sitzungsber. Königl. Preuß. Akad. Wissenschaft.* (Berlin) **47** (2) 831–839 (1915) [Translated into English: "Explanation of the perihelion motion of Mercury from the general theory of relativity", in *CPAET* Vol. 6, doc. 24; Translated into Russian: "Ob"yasnenie dvizheniya perigeliya Merkuriya v obshchei teorii otnositel'nosti", in *SNT* Vol. 1, p. 439]
61. Einstein A Über die spezielle und die allgemeine Relativitätstheorie, gemeinverständlich Vol. 38 (Braunschweig, Germany: Friedrich Vieweg und Sohn, 1917) [Translated into English: "On the special and general theory of relativity (a popular account)", in *CPAET* Vol. 6, doc. 42; Translated into Russian: "O spetsial'noi i obshchei teorii otnositel'nosti (obshchedostupnoe izlozhenie)", in *SNT* Vol. 1, p. 530]
62. Einstein A "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie" *Sitzungsber. Königl. Preuß. Akad. Wissenschaft.* (Berlin) **1** 142–152 (1917) [Translated into English: "Cosmological considerations in the general theory of relativity", in *CPAET* Vol. 6, doc. 43; Translated into Russian: "Voprosy kosmologii i obshchaya teoriya otnositel'nosti", in *SNT* Vol. 1, p. 601]
63. Friedmann A "Über die Krümmung des Raumes" *Z. Phys.* **10** (1) 377–386 (1922) [Translated into Russian: "O krivizne prostranstva" ("On curvature of space") *Zh. Russk. Fiz.-Khim. Obshch.* **56** 59 (1924)]
64. Friedmann A "Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes" *Z. Phys.* **21** (1) 326–332 (1924) [Translated into Russian: "O vozmozhnosti mira s postoyannoi otritsatel'noi kriviznoi prostranstva" ("On the possibility of a world with constant negative curvature of space") *Usp. Fiz. Nauk* **80** 447 (1963)]
65. Einstein A "Bemerkung zu der Arbeit von A. Friedmann: Über die Krümmung des Raumes" *Z. Phys.* **11** 326 (1922) [Translated into Russian: "Zamechanie k rabote A. Fridmana 'O krivizne prostranstva'" ("Comment on a paper by A. Fridman 'On curvature of space'"), in *SNT* Vol. 2, p. 118; *Usp. Fiz. Nauk* **80** 454 (1963)]
66. Einstein A "Notiz zu der Bemerkung zu der Arbeit von A. Friedmann: Über die Krümmung des Raumes" *Z. Phys.* **16** 228 (1922) [Translated into Russian: "K rabote A. Fridmana 'O krivizne prostranstva'" ("On a paper by A. Fridman 'On curvature of space'"), in *SNT* Vol. 2, p. 119; *Usp. Fiz. Nauk* **80** 454 (1963)]
67. Hubble E "A relation between distance and radial velocity among extra-galactic nebulae" *Proc. Natl. Acad. Sci.* **15** 168–173 (1929)
68. Einstein A "Appendix to the second edition. On the 'cosmologic problem'", in *The Meaning of Relativity* 2nd ed. (Princeton, NJ: Princeton Univ. Press, 1945) [Translated into Russian: "O 'kosmologicheskoi probleme'" ("On cosmological problem"), in *SNT* Vol. 2, p. 597]
69. Starobinsky A A "Spektr reliktovogo gravitatsionnogo izlucheniya i nachal'noe sostoyanie Vselennoi" *Pis'ma Zh. Eksp. Teor. Fiz.* **30** 719–723 (1979) ["Spectrum of relict gravitational radiation and the early state of the universe" *JETP Lett.* **30** 682–685 (1979)]
70. Guth A H "The inflationary universe: a possible solution to the horizon and flatness problems" *Phys. Rev. D* **23** 347–356 (1981)
71. Linde A D "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems" *Phys. Lett. B* **108** 389–393 (1982)
72. Riess A G et al. (Supernova Search Team) "Observational evidence from supernovae for an accelerating universe and a cosmological constant" *Astron. J.* **116** 1009–1038 (1998)
73. Perlmutter S et al. (The Supernova Cosmology Project) "Measurements of Ω and Λ from 42 high-redshift supernovae" *Astrophys. J.* **517** 565–586 (1999)
74. Astier P et al. (Supernova Legacy Survey) "The supernova legacy survey: Measurement of Ω_M, Ω_Λ and w from the first year data set" *Astron. Astrophys.* **447** 31–48 (2006)
75. Noether E "Invariante Variationsprobleme" *Nachricht. Königlichen Gesellschaft Wissenschaft. Göttingen. Math.-phys. Kl.* 235–257 (1918)
76. Einstein A "Letter to David Hilbert, 24 May 1918", in *CPAET* Vol. 8, doc. 548
77. Einstein A "Der Energiesatz in der allgemeinen Relativitätstheorie" *Sitzungsber. Königlich Preuß. Akad. Wissenschaft.* (Berlin) **1** 448–459 (1918) [Translated into English: "The law of energy conservation in the general theory of relativity", in *CPAET* Vol. 7, doc. 9; Translated into Russian: "Zakon sokhraneniya energii v obshchei teorii otnositel'nosti", in *SNT* Vol. 1, p. 650]
78. Einstein A "Prüfung der allgemeinen Relativiätstheorie" *Naturwissenschaft.* **7** 776 (1919) [Translated into English: "A test of the general theory of relativity", in *CPAET* Vol. 7, doc. 23; Translated into Russian: "Dokazatel'stvo obshchei teorii otnositel'nosti", in *SNT* Vol. 1, p. 663]
79. Einstein A "Was ist die Relativitäts-Theorie?" *The Times* (28 November 1919) [Translated into English: "What is the theory of relativity?", in *CPAET* Vol. 7, doc. 25; Translated into Russian: "Chto takoe teoriya otnositel'nosti?", in *SNT* Vol. 1, p. 677]
80. Einstein A "Grundgedanken und Methoden der Relativitätstheorie, in ihrer Entwicklung dargestellt" (1920) [Translated into English: Fundamental ideas and methods of the theory of relativity, presented in their developments", in *CPAET* Vol. 7, doc. 31]
81. Einstein A "Meine Antwort. Ueber die anti-relativitätstheoretische G.m.b.H." *Berliner Tageblatt* (27 August 1920) [Translated into English: "My response. On the anti-relativity company", in *CPAET* Vol. 7, doc. 45; Translated into Russian: "Moi otvet. Po povodu antirelyativistskogo aktsionernogo obshchestva", in *SNT* Vol. 1, p. 693]
82. Einstein A Vier Vorlesungen "Über Relativitätstheorie, Gehalten im Mai, 1921, an der Universität Princeton" (Braunschweig, Germany: Friedrich Vieweg und Sohn, 1922) [Translated into English: "Four lectures on the theory of relativity, held at Princeton University in May 1921, in *CPAET* Vol. 7, doc. 71; Translated into Russian: "Sushchnost' teorii otnositel'nosti" ("The meaning of relativity theory"), in *SNT* Vol. 2, p. 5]
83. Leman I *Teoriya Otnositel'nosti*. Populyarnoe Izlozhenie bez Matematicheskikh Formul. (Relativity Theory. Popular Presentation Without Mathematical Formulas) (Moscow: Rabotnik prosveshcheniya, 1922)
84. Einstein A "Newtons Mechanik und ihr Einfluß auf die Gestaltung der theoretischen Physik" *Naturwissenschaft* **15** (12) 273–276 (1927) [Translated into English: "The mechanics of Newton and their influence on the development of theoretical physics", in: Einstein A *Essays in Science* (New York: Philosophical Library, 1934) p. 28; Translated into Russian: "Mekhanika Neuton'a i ee vliyanie na formirovanie teoreticheskoi fiziki", in *SNT* Vol. 4, p. 82]
85. Einstein A "Fundamental concepts of physics and their most recent changes" *St. Louis Post-Dispatch* (9 December 1928) [Translated into Russian: "Fundamental'nye ponyatiya fiziki i izmeneniya, kotorye proizoshli v nikh za poslednee vremya", in *SNT* Vol. 4, p. 103]
86. Einstein A "Space-time", in *Encyclopaedia Britannica* Vol. XXI, 14th edn. (1929) pp. 105–108 [Translated into Russian: "Prostranstvo-vremya", in *SNT* Vol. 2, p. 234]
87. Einstein A "Ansprache von Prof. Einstein an Prof. Planck" *Forschung. Fortschr.* **5** 248–249 (1929) [Translated into Russian: "Rech na yubilee Professora Planka" ("Address on the Jubilee anniversary of Professor Planck"), in *SNT* Vol. 4 p. 109]
88. Friedman A J, Donley C C *Einstein as Myth and Muse* (Cambridge: Cambridge Univ. Press, 1985)
89. Einstein A "Elementary derivation of the equivalence of mass and energy" *Bull. Am. Math. Soc.* **41** 223–230 (1935) [Translated into Russian: "Elementarnyi vyvod ekvivalentnosti massy i energii", in *SNT* Vol. 2, p. 416]

90. Einstein A "The late Emmy Noether" *New York Times* (4 May 1935) [Translated into Russian: "Pamyati Emmi Neter" ("In memory of Emmy Noether"), in *SNT* Vol. 4, p. 198]
91. Einstein A, Podolsky B, Rosen N "Can quantum mechanical description of physical reality be considered complete?" *Phys. Rev.* **47** (10) 777–780 (1935) [Translated into Russian: "Mozhno li schitat' kvantovomekhanicheskoe opisanie fizicheskoi real'nosti polnym?", in *SNT* Vol. 3, p. 604]
92. Einstein A, Infeld L *The Evolution of Physics: the Growth of Ideas from Early Concepts to Relativity and Quanta* (New York: Simon and Schuster, 1938)
93. Landau L D, Lifshitz E M *Teoriya Polya* (Theory of Fields) (Moscow–Leningrad: Gostekhizdat, 1941)
94. Landau L D, Lifshitz E M *The Classical Theory of Fields* (Cambridge, MA: Addison-Wesley Press, 1951)
95. Bergmann P G *Introduction to the Theory of Relativity* (New York: Prentice-Hall, 1942) [Translated into Russian: *Vvedenie v Teoriyu Otnositel'nosti* (Introduction to the Theory of Relativity (With A. Einstein's foreword); Translated from English, Ed. V L Ginzburg) (Moscow: IL, 1947)]
96. *Discovery* (September 1945)
97. Smyth H *Atomic Energy for Military Purposes: The Official Report on the Development of the Atomic Bomb under the Auspices of the United States Government, 1940–1945* (Princeton, NJ: Princeton Univ. Press, 1945) [Translated into Russian: *Atomnaya Energiya dlya Voennykh Tselei. Ofitsial'nyi Otchet o Razrabotke Atomnoi Bomby pod Nablyudeniem Pravitel'stva SShA* (Translated from English, Ed. G N Ivanov) (Moscow: Transzheldorizdat, 1946)]
98. Einstein A "Elementary derivation of the equivalence of mass and energy" *Technion J.* **5** 16 (1946)
99. Einstein A "$E = mc^2$: The most urgent problem of our time" *Sci. Illustrated* **1** 16–17 (1946)
100. Einstein A "Autobiographical notes", in *Albert Einstein, Philosopher-Scientist* (The Library of Living Philosophers, Vol. 7, Ed. P A Schilpp) (Evanston, IL: Library of Living Philosophers, 1949) [Translated into Russian: "Avtobiograficheskie zametki", in *SNT* Vol. 4, p. 259]
101. Einstein A "Letter to L. Barnett, 19 June 1948", in: Okun L B "The concept of mass" *Phys. Today* **42** (6) 31–36 (1989)
102. Feynman R P "The theory of positrons" *Phys. Rev.* **76** 749–759 (1949)
103. Feynman R P "Space-time approach to quantum electrodynamics" *Phys. Rev.* **76** 769–789 (1949)
104. Dyson F "The world on a string" *The New York Review of Books* (51) (May 13, 2004)
105. Calaprice A *The Quotable Einstein* (Princeton, NJ: Princeton Univ. Press, 1996)
106. Einstein A "Appendix V. Relativity and the problem of space", in *Relativity, the Special and the General Theory, A Popular Exposition* 15th ed. (New York: Crown Publ., 1952) [Translated into Russian: "Otnositel'nost' i problema prostranstva" ("Relativity and problem of space"), in *SNT* Vol. 2, p. 744]
107. Jammer M *Concepts of Space: the History of Theories of Space in Physics* (Cambridge, MA: Harvard Univ. Press, 1954)
108. Einstein A "On the investigation of the state of the ether in a magnetic field", in *CPAE* Vol. 1, doc. 4 (1895)
109. Einstein A "Erinnerungen — Souvenirs" *Schweiz. Hochschulzeitung, Sonderheft 100 Jahre ETH* **28** 145–153 (1955); Reprinted as "Autobiographicsche Skizze", in *Helle Zeit-dunkle Zeit; in Memoriam Albert Einstein* (Ed. C Seelig) (Zürich: Europa, 1956) p. 9 [Translated into Russian: "Avtobiograficheskie nabroski", in *SNT* Vol. 4, p. 350]
110. Infeld L "Istoriya razvitiya teorii otnositel'nosti" ("History of development of relativity theory") *Usp. Fiz. Nauk* **57** 193 (1955)
111. Born M *Die Relativitätstheorie Einsteins und ihre physikalische Grundlagen gemeinverständlich dargestellt* (Berlin: Springer, 1920)
112. Born M *Teoriya Otnositel'nosti Einshteina i ee Fizicheskie Osnovy* (Einstein's Theory of Relativity and its Physical Foundation) (Moscow: ONTI, 1938)
113. Born M *Einstein's Theory of Relativity* (New York: Dover, 1962)
114. Born M *Einshteinovskaya Teoriya Otnositel'nosti* (Einstein's Theory of Relativity) (Moscow: Mir, 1964, 2nd ed. — 1972)
115. Born M *Atomic Physics* (London: Blackie, 1963)
116. Born M *Atomnaya Fizika* (Atomic Physics) (Moscow: Mir, 1967)
117. Einstein A, Born M, Born H *Albert Einstein, Hedwig und Max Born: Briefwechsel, 1916–1955* (München: Nymphenburger Verlagshandlung, 1969)
118. Born M, in *The Born-Einstein Letters 1916–1955: Friendship, Politics and Physics in Uncertain Times* (New York: Mcmillan, 2005)
119. Landau L D, Rumer G B *What is Relativity?* (Dover: Basic Books Inc., 1959)
120. Landau L D, Rumer Yu B *Chto Takoe Teoriya Otnositelnosti?* (What is Relativity?) (Moscow: Sovetskaya Rossiya 1960, 2nd ed. — 1963, 3rd ed. — 1975)
121. Feynman R, Leighton R, Sands M *The Feynman Lectures on Physics* (Reading, Mass.: Addison-Wesley, 1963–1965)
122. Feynman R, Leighton R, Sands M *Feinmanovskie Lektsii po Fizike* (The Feynman Lectures on Physics) (Moscow: Mir, 1st ed. — 1965; 4th ed. — 2004)
123. Feynman R P "The development of the space-time view of quantum electrodynamics", in *Nobel Lectures, Physics 1963–1970* (Amsterdam: Elsevier, 1972) [Translated into Russian: "Razvitie prostranstvenno-vremennoi traktovki kvantovoi elektrodinamiki" (Nobelevskie lektsii po fizike 1965) *Usp. Fiz. Nauk* **91** 29 (1967)]

The theory of relativity and the Pythagorean theorem*

L. B. Okun

Abstract

It is shown that the most important effects of special and general theory of relativity can be understood in a simple and straightforward way. The system of units in which the speed of light c is the unit of velocity allows to cast all formulas in a very simple form.The Pythagorean theorem graphically relates energy, momentum and mass. The paper is addressed to those who teach and popularize the theory of relativity

1 Introduction

The report "Energy and mass in the works of Einstein, Landau and Feynman" that I was preparing for the Session of the Division of Physical Sciences of the Russian Academy of Sciences (DPS RAS) on the occasion of the 100th anniversary of Lev Davidovich Landau's birth was to consist of two parts, one on history and the other on physics. The history part was absorbed into the article "Einstein's formula: $E_0 = mc^2$. 'Isn't the Lord laughing?"' that appeared in the May issue of Uspekhi Fizicheskikh Nauk [Physics-Uspekhi] journal [1]. The physics part is published in the present article. It is devoted to various, so to speak, technical aspects of the theory, such as the dimensional analysis and fundamental constants c and $\hbar$; the kinematics of a single particle in the entire velocity range from 0 to c; systems of two or more free particles; and the interactions between particles: electromagnetic, gravitational, etc. The text uses the slides of the talk at the session of the Section of Nuclear Physics of the DPS RAS in November 2007 at the Institute for Theoretical and Experimental Physics (ITEP). My goal was to present the main formulas of the theory of relativity in the simplest possible way, using mostly the Pythagorean theorem.

2 Relativity

2.1. The advanced standpoint.

The history of the concept of mass in physics runs to many centuries and is very interesting, but I leave it aside here. Instead, this will be an attempt to look at mass from an advanced standpoint. I borrowed the words from the famous title of Felix Klein's *Elementary Mathematics from an Advanced Standpoint* (traditionally translated into

*This is a slightly corrected version of the paper published in *Physics - Uspekhi* **51** *622 (2008)*. In particular, the short subsections of the text are numbered as had been suggested by one of the readers.

Russian incorrectly as *Elementary Mathematics from the Standpoint of Higher Mathematics.* See V.G. Boltyanskii's foreword to the 4th Russian edition). The advanced modern standpoint based on principles of symmetry in general and on the theory of relativity in particular makes it possible to avoid inevitable terminological confusion and paradoxes.

2.2. The principle of relativity. Ever since the time of Galileo and Newton, the concept of relativity has been connected with the impossibility of detecting, by means of any experiment, a translational (uniform and rectilinear) motion of a closed space (for instance, inside a ship) while remaining within this space. At the turn of XIX and XX centuries Poincaré gave to this idea the name 'the principle of relativity' . In 1905 Einstein generalized this principle to the case of the existence of the limiting velocity of propagation of signals. (The finite velocity of propagation of light has been discovered by Römer already in 1676). Planck called the theory constructed in this way 'Einstein's theory of relativity'.

2.3. Mechanics and optics. Newton tried to construct a unified theory uniting the theory of motion of massive objects (mechanics) and the theory of propagation of light (optics). In fact, it became possible to create the unified theory of particles of massive matter and of light only in the XX century. It was established on the road to the vantage ground of truth (I am using here the ironical wording of Francis Bacon) that light is also a sort of matter, just like the massive stuff, but that its particles are massless. This interpretation of particles of light — photons — continues to face resistance from many students of physics, and even more from physics teachers.

3 Dimensions

3.1. Units in which c = 1. The maximum possible velocity is known as the speed of light and is denoted by c. When dealing with formulas of the theory of relativity it is convenient to use a system of units in which c is chosen as a unit of velocity. Since $c/c = 1$, using this system means that we set $c = 1$ in all formulas, thus simplifying them greatly. If time is measured in seconds, then distance in this system of units should be measured in light seconds: one light second equals $3 \cdot 10^{10}$ cm.

3.2. Poincaré and c. One of the creators of the theory of relativity, Henri Poincare, when discussing in 1904 the fact that c is found in every equation of electrodynamics, compared the situation with the geocentric theory of Ptolemy's epicycles in which every relation between motions of celestial bodies included the terrestrial year. Poincare expressed his hope that the future Copernicus would rid electrodynamics of c [3]. However, Einstein showed already in 1905 that c was to play the key role as the limit for the velocity of signal propagation.

3.3. Two systems of units: SI and c = 1. The unit of velocity in the International System of Units SI, 1 m/s, is forced on us by convenience arguments and by standardization of manufacturing and commerce but not by the laws of Nature. In contrast to this, c as a unit of velocity is imposed by Nature itself when we wish to consider fundamental processes of Nature.

3.4. Dimensional factors. Consider some physical quantity a. Let us denote by $[a]$ the dimension of the quantity a. The dimension of a definitely changes if it is multiplied by any power of the universal constant c but its physical meaning remains unaffected. In what follows I explain why this is so.

3.5. Velocity, momentum, energy, mass. The dimensions of momentum, mass, and velocity of a particle are usually related by the formula $[\mathbf{p}] = [\mathbf{m}][\mathbf{v}]$ while the dimensions of energy, mass, and velocity are related by the formula $[E] = [m][\mathbf{v}^2]$. Let us introduce dimensionless velocity $\mathbf{v}/c$ and from now on denote this ratio as $\mathbf{v}$. Likewise, referring to momentum $\mathbf{p}$ we actually mean the ratio $\mathbf{p}/c$. When speaking of energy, we actually mean the ratio $e = E/c^2$. Obviously, the dimensions of $\mathbf{p}$, e, and m become identical and therefore, these quantities can be measured in the same units, for example, in grams or electron-volts, as is customary in elementary particle physics.

3.6. On the letter e denoting energy. Choosing e as the notation for energy may invite the reader's ire since this symbol traditionally stands for electron and electric charge. However, this choice cannot cause confusion and, importantly, it will lead to a compact form of formulas for a single particle, always reminding us that these formulas were written using the system of units in which $c = 1$. On the other hand, it will be clear a little later that the letter E is a convenient notation for the energy of two or more particles. I happened to see Einstein's formula with a lower-case e on a billboard on Rublevskoye highway in Moscow. I wonder, why should this e irritate physicists?

3.7. On the difference between energy and frequency. Two paragraphs ago I insisted that $e = E/c^2$ is energy even though its dimension is that of mass. In that case it is logical to ask why $\omega = E/\hbar$ is not energy but frequency? Indeed, the quantum of action $\hbar$, like the speed of light c, is a universal constant. The answer to this question can be found by considering how e and ω are measured. E and e are measured by the same procedure, say, using a calorimeter, while frequency is measured in a drastically different manner, say, using clocks. Therefore, the equality $\omega = E/\hbar$ informs us of the link between two different types of measurement, while the equality $e = E/c^2$ carries no such information. Arguments similar to those concerning frequency hold equally well for wavelength. I have to emphasize that these metrological distinctions are mostly of a historical nature since in our day atomic clocks operate on the difference between atomic energy levels.

4 Single particle

4.1. Relative and absolute quantities. The kinetic energy of any body is a relative quantity: it depends on the reference frame in which it is measured. The same is true for the momentum of a body and its velocity. In contrast to them, the mass of a body is an absolute quantity: it characterizes the body as such, irrespective of the observer. The rest energy of a body (see below) is also an absolute quantity since the frame of reference is fixed in it once and for all — 'nailed to it'.

4.2. Invariant mass. The mass of a body is defined in the theory of relativity by the formula

$$m^2 = e^2 - p^2. \tag{1}$$

Here and in what follows $p = |\mathbf{p}|$. Likewise, $v = |\mathbf{v}|$. Note that energy and momentum of a given body are not bounded from above while the mass of the body is fixed. Formula (1) is the simplest relation between energy, momentum, and mass that one could write 'off the top of one's head'. (The relation between e, $\mathbf{p}$, and m cannot be linear since $\mathbf{p}$ is a vector while e and m are scalars in three-dimensional space.) We shall see now that formula (1) has another, much more profound theoretical foundation.

4.3. The 4-momentum. Minkowski was the first to point out that the theory of relativity gains the simplest form if considered in four-dimensional spacetime [4]. Energy and momentum in the theory of relativity form a four-dimensional energy-momentum vector $p_i (i = 0, a)$, where $p_0 = e$, $p_a = \mathbf{p}$, and $a = 1, 2, 3$. Mass is the Lorentz scalar that characterizes the length of the 4-vector p_i: $m^2 = p_i{}^2 = e^2 - \mathbf{p}^2$; four-dimensional space is pseudo-Euclidean, which explains the minus sign in the formula for length squared. (The reader will recall that $\mathbf{p}^2 = p^2$.) Another way to clarify why the sign is negative is by introducing the imaginary momentum $i\mathbf{p}$. Then $m^2 = e^2 + (i\mathbf{p})^2$ and we are dealing with the Pythagorean theorem for such a pseudo-Euclidean right triangle in which the hypotenuse m is shorter than the cathetus e.

4.4. Relation between momentum and velocity. The momentum of a body is related to its velocity $\mathbf{v}$ by the formula

$$\mathbf{p} = e\mathbf{v}. \tag{2}$$

This formula satisfies in the simplest manner the requirement that the momentum 3-vector be proportional to the velocity 3-vector and that the dimensional proportionality coefficient not vanish for the massless photon. Conservation of the thus defined momentum in the theory of relativity is implied by the uniformity of 3-space while conservation of energy is implied by the uniformity of time (Noether's theorem).

4.5. The Pythagorean theorem. Formula (1) is shown in Fig. 1 by an ordinary right triangle in which m and p are catheti and e is the hypotenuse.

4.6. Transition from $\mathbf{m \neq 0}$ to $\mathbf{m = 0}$. Formula (1) is obviously valid at $m = 0$ while formula (2) holds for $v = 1$. This implies that there is a smooth transition from massless particles to massive, when the energy of the latter particles greatly exceeds their mass.

4.7. Physics from $\mathbf{p = 0}$ to $\mathbf{p = e}$. Let us consider formulas (1) and (2) first at zero momentum, then in the limit of very low momenta (when $p \ll m$), and then in the limit of very high momenta when $p \sim e \gg m$, and finally in the case of massless photons. We will call the case of very small momenta and velocities the Newtonian case, and that of very high momenta and velocities close to the speed of light, the ultrarelativistic case. We will start with zero momentum.

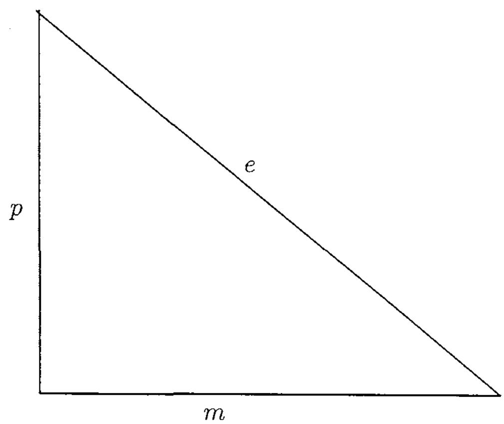


Fig. 1

5 Rest energy

5.1. Zero momentum. If momentum is zero, then in the case of a massive particle the velocity is also zero and energy e is by definition equal to the rest energy e_0. (The subscript 0 reminds us that here we are dealing not with the energy of a given body in general but with its energy precisely in the case when its momentum is zero!) Hence equation (1) implies

$$e_0 = m. \tag{3}$$

If, however, the particle is massless, then equation (1) at $p = 0$ implies that $e = e_0 = 0$ (see 7.6).

5.2. Horizontal 'biangle'. If $m \neq 0$ and $p = 0$, then the triangle shown in Fig.1 'collapses' to a horizontal 'biangle' (Fig. 2).

5.3 Einstein's great discovery. In units in which $c \neq 1$, equation (3) has the form

$$E_0 = mc^2\,. \tag{4}$$

The realization that ordinary matter at rest stores an enormous amount of energy in its mass was Einstein's great discovery.

5.4. The 'famous formula'. Equation (4) is very often written (especially in popular physics literature) in the form of 'Einstein's famous equation' that drops the subscript 0:

$$E = mc^2\,. \tag{5}$$

This simplification, to which Einstein himself sometimes resorted, might seem innocuous at first glance, but it results in unacceptable confusion in understanding the foundations of physics. In particular, it generates a totally false idea that 'according to the theory of relativity' the mass of a body is equivalent to its total energy and, as an inevitable result, depends on its velocity. ('Wished to make it simpler, got it as always'.[1])

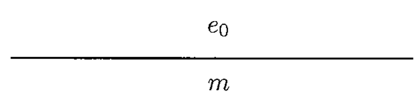


Fig. 2

5.5. No experiment can disprove the 'famous formula'. Very clever people thought up this formula in such a way that it never contradicts experiments. However, it contradicts the essence of the theory of relativity. In this respect, the situation with the 'famous formula' is unique — I do not know another case that could be compared with this one.

5.6. This is not a matter of taste but of understanding. You hear time and again that the introduction of momentum-dependent mass is 'a matter of taste'. Of course, one can write the letter m instead of E/c^2 and even call it 'mass', although

[1] A paraphrase of former Russian Prime Minister Chernomyrdin's 'statement of the day': "Wished to make it better, got as always." (Note added by the Author in translation.)

it is no more sensible than writing p instead of E/c and calling it 'momentum'. Alas, this 'dress changing' introduces unnecessary and bizarre notions — relativistic mass and rest mass m_0— and creates an obstacle to understanding the theory of relativity. A well-known Russian proverb comes to mind: "Call me a pot if you wish but don't push me into the oven." Unfortunately, people who call E/c^2 'mass' do place this 'pot' into the 'oven' of physics teaching.

5.7. Longitudinal and transverse masses. In addition to relativistic mass, concepts of intense use at the beginning of the XX century were the transverse and longitudinal masses: m_t and m_l. This longitudinal masses increased as $(e^3/m^3)\,m$ and 'explained' — in terms of Newton's formula $F = ma$ — why a massive body cannot be accelerated to the speed of light. Then it was forgotten and such popularizers of the theory of relativity as Stephen Hawking started to persuade their readers that even much gentler growth of mass with velocity $((e/m)\,m)$ could explain why the velocity of a massive body cannot reach c. I single out Hawking only because, printed on the dust jacket of the Russian edition of his latest popular science book [5], which advertises the formula $E = mc^2$, we see this text: "Translated into 40 languages. More than 10 million copies sold worldwide."

5.8. False intuition. After my talk at the ITEP A N Skrinsky told me that the notion of relativistic mass hampered a well-known physicist's understanding that a relativistic electron colliding with an electron at rest can transfer all its energy to the latter. Indeed, how could a heavy baseball bat transfer all its energy to the lightest ping-pong ball? In physics, as in daily life, people very often rely on intuition. This is why it is so important, when studying the theory of relativity, to work out the relativistic intuition and mistrust nonrelativistic intuition. (In order to 'feel' how an electron at rest can receive the entire energy of a moving electron it is sufficient to use their center-of-inertia frame to consider scattering by 180 degrees, and then return back to the laboratory frame.)

6 Newtonian mechanics

6.1. Momentum in Newtonian mechanics. Newtonian mechanics describes with high accuracy the motion of macroscopic bodies in a terrestrial environment and of massive celestial bodies because their velocities are much smaller than the speed of light. For instance, the velocity of a bullet is of the order of 1 km/s, which corresponds to $v = 1/300000$ and $v^2 = 10^{-11}$. In this situation equation (2) reduces to

$$\mathbf{p} = m\mathbf{v}\,. \tag{6}$$

Equation (1) is schematically shown in the Newtonian limit in Fig. 3. The side of the triangle representing p in Fig. 3 is far too long. Scaled correctly, it should be a few microns.

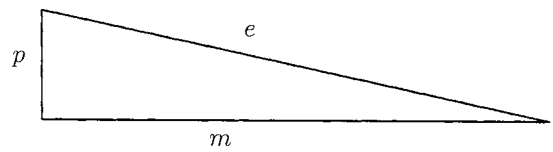


Fig. 3

6.2. Kinetic energy e_k. It is reasonable to rewrite formula (1) for low velocities so as to isolate the contribution of the short cathetus:

$$e^2 - m^2 = p^2 \tag{7}$$

and then to present it in the form

$$(e-m)(e+m) = p^2 . \tag{8}$$

This allows us to obtain a nonrelativistic expression for kinetic energy without resorting to the conventional series expansion of the square root. We take into account that the total energy e is the sum of rest energy e_0 and kinetic energy e_k and therefore $e = m + e_k$.

6.3. Energy in Newtonian mechanics. In the Newtonian limit we have $e_k \ll m$ (e.g. for a bullet $e_k/m = 10^{-11}$). Energy can therefore be replaced with high accuracy by mass m in formula (2) for momentum and in the factor $(e+m)$ in equation (8). This last equation immediately implies an expression for kinetic energy e_k in Newtonian mechanics:

$$e_k = \frac{p^2}{2m} = \frac{mv^2}{2} . \tag{9}$$

6.4. Potential energy. In addition to velocity-dependent kinetic energy, an important role in nonrelativistic mechanics is played by potential energy, which depends only on the position (coordinate) of the body. The sum of kinetic and potential energy is conserved at any instance of time. The potential energy of a body placed in an external field of force is defined to within an arbitrary additive constant because the force acting on the body equals the gradient of potential energy. In a similar manner, the potential energy of interaction of several bodies depends only on their positions at the moment of interaction. However, in the theory of relativity any interaction propagates at a finite velocity. Hence, potential energy is an essentially nonrelativistic concept.

6.5. Newton and modern physics. Newton's flash of genius marked the birth of modern science. The post-Newtonian progress of science is fantastic. Today's understanding of the structure of matter is radically different from Newton's. Nevertheless, even in the XXI century many physics textbooks continue to use Newton's equations at energies $e_k \gg e_0$, which exceed the limits of applicability of Newton's mechanics ($e_k \ll e_0$) by many orders of magnitude. If some professors prefer to insist on keeping up with this tradition of velocity-dependent mass, they ought to at least familiarize their students with the fundamental concepts of mass and rest energy, and with the true Einstein equation $E_0 = mc^2$.

7 Ultrarelativism

7.1. High energy physics. Let us now consider in some detail the limiting case in which $e/m \gg 1$. The ratio of energy and mass characteristic for high energy physics is precisely this. For example, this ratio for electrons in the LEP (Large Electron-Positron) Collider at CERN was $e/m = 10^5$, since $m = 0.5$ MeV and $e = 50$ GeV. For protons in the LHC (Large Hadron Collider), which is located in the same tunnel where the LEP was in previous years, we find $e/m \sim 10^4$. (Here, $m \sim 938$ MeV, $e \sim 7$ TeV.)

7.2. A vertical triangle. The triangle for protons in the LHC is drawn highly schematically in Fig. 4. Its base is in fact shorter than its hypotenuse by four orders of magnitude.

7.3. The neutrino. Neutrinos are even more ultrarelativistic particles: their masses are a fraction of one electron-volt and their energies reach several MeV for neutrinos emerging from the Sun and nuclear reactors, and several GeV for neutrinos generated in particle decays in cosmic rays and in accelerators. The base of the triangle shown schematically in Fig. 4 is much shorter at these energies than its vertical cathetus and its hypotenuse.

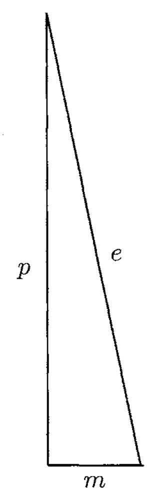


Fig. 4

7.4. Neutrino oscillations and $\mathbf{m^2/2e}$. Equation $(e-p)(e+p) = m^2$ immediately implies that $e - p \simeq m^2/2e$. The differences between the masses of three neutrinos ν_1, ν_2, ν_3 possessing definite masses in a vacuum result in oscillations between neutrinos having no well-defined masses but possessing certain flavors: ν_e, ν_μ, ν_τ. (This phenomenon is similar to well-known beats that occur when several frequencies interfere.) The neutrino oscillation data give

$$\Delta m_{21}^2 = (0.77 \pm 0.04) \times 10^{-4}\ \mathrm{eV}^2$$

$$|\Delta m_{32}^2| = (24 \pm 3) \times 10^{-4}\ \text{eV}^2\,.$$

7.5. The photon. The photon mass is so small that no experiment has been able to detect it. Hence, it is usually assumed that the photon mass equals zero. This means that for a photon $e = p$, where $p = |\mathbf{p}|$, and the triangle shown in Fig. 4 collapses to a vertical biangle (Fig. 5).

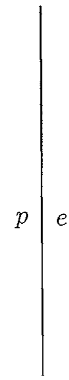


Fig. 5

7.6 The photon and rest energy? It is logical to conclude the discussion of single-particle mechanics by returning to the question: is the concept of rest energy e_0 applicable to massless photon? It may seem at first glance that it is not, since a photon propagates at the speed c, however small its energy is, so that 'a rest for it is but a dream' [2]. This being so, how can we use the equality $e_0 = 0$ if the photon is never at rest? We can because our e_0 is defined as the energy corresponding to zero momentum, not velocity. Obviously this energy is zero for the photon with $p = 0$: this is implied by equation (1). If a particle has $m = 0, p = 0, e = 0$ and biangle of Fig. 5 collapses to a point, we can say that it 'passed away to the state of eternal rest'. Looking at the limiting transition to zero mass, we can show that the reference frame in which a photon is 'eternally at rest' has to be rigidly connected to another 'eternally resting' photon. Consequently, the value $e_0 = 0$ at $m = 0$ is in perfect agreement with the limiting transition.

8 Two free particles

8.1. Collision of two particles. Colliders. If two particles collide at relativistic

[2]This is a paraphrase of the famous line from Alexander Block. (Note added in translation).

energies, a comparison of the reference frame in which one of them is at rest with a reference frame in which their common center of inertia is at rest demonstrates the advantages of the latter. We already saw this in the case commented on by A N Skrinsky. If the momenta of the colliding particles are equal and oppositely directed, as for example in the LHC or LEP collider, then practically the entire energy of the colliding particles may be spent on the creation of new particles.

8.2. Mass of a system of particles. The total energy E and the total momentum $\mathbf{P}$ of an isolated system of particles are conserved. Energy and momentum being additive, for two free particles we have

$$E = e_1 + e_2 \tag{10}$$

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2. \tag{11}$$

We now define the quantity M by the formula

$$M^2 = E^2 - \mathbf{P}^2. \tag{12}$$

8.3. Masses are additive at $\mathbf{v} = \mathbf{0}$. Equation (12) is invariant under Lorentz transformations, as is equation (1). Therefore, it is logical to refer to M as the mass of a system of two particles. In the static limit, when p_1 and p_2 equal zero, equation (12) implies that

$$M = e_{01} + e_{02} = m_1 + m_2. \tag{13}$$

In the Newtonian limit, M equals the sum of the masses of the two particles with an accuracy of $(v/c)^2$, i.e. the masses are practically additive.

8.4. Masses are not additive at $\mathbf{v} \neq \mathbf{0}$. However, M and the masses m_1 and m_2 are practically unrelated at high velocities. For instance, M exceeds the electron mass in the LEP collider or the proton mass in the LHC by four orders of magnitude (see section 7). The value of M is crucially dependent on the relative directions of the momenta of two particles, since the sum of two vectors is a function of the angle between them. Thus, we have for two photons moving in the same direction

$$P = |\mathbf{P}| = |\mathbf{p}_1 + \mathbf{p}_2| = p_1 + p_2. \tag{14}$$

8.5. Collinear photons. For photons $p_1 = e_1$, and $p_2 = e_2$. Therefore, for two photons moving in the same direction we can write

$$P = p_1 + p_2 = e_1 + e_2 = E. \tag{15}$$

Equation (12) then implies that in this case the mass of a pair of photons $M = 0$. And this means that the mass of a 'needle' light beam is zero.

8.6. What if photons fly away from each other? However, if photons fly away in opposite directions with equal energies, then $\mathbf{p}_1 = -\mathbf{p}_2$ and $\mathbf{P} = 0$. In that case, the rest energy of two photons simply equals the sum of their energies and the mass of this system is

$$M = E_0 = 2e. \tag{16}$$

8.7. Shock. Of course, the statement that a pair of two massless or very light particles might have an enormous mass may shock the unprepared reader. Is there any sense in speaking of the rest energy of two photons if 'rest is but a dream' to either of them? What is at rest in this case?

8.8. The answer is obvious. The entity at rest is the geometric point – the center of inertia of the two photons. While the rest energy for one particle is the energy hidden in its mass, for two photons it is simply the sum of their energies (kinetic energies!) in the reference frame in which their momenta are equal in magnitude and opposite in direction. There is no hidden energy in this case!

8.9. What does it mean to be conserved? When saying that energy is conserved, we mean that the sum of the energies of particles entering a reaction equals the sum of the energies of particles created as a result of this reaction. The statement on the conservation of momentum has a similar meaning. However, since momentum is a vector quantity, now we are dealing with a vector sum of momenta. (In the case of momenta we speak about three independent conservation laws: conserved are the sums of projections of momenta on three mutually orthogonal directions.) The conserved quantities are thus $E = \sum e_i$ and $\mathbf{P} = \sum \mathbf{p}_i$. As for the energies of individual particles e_i their momenta $\mathbf{p}_i$ in the laboratory reference frame, they are conserved only in elastic forward scattering. Here, it is important to stress the difference between the concepts of additivity and conservation. The former concept refers to the state of a system of free particles, the latter refers to the process of interaction of the particles.

8.10. Is mass conserved? With E and $\mathbf{P}$ conserved, the mass M of a system (a set) of particles, defined by the formula $M^2 = E^2 - \mathbf{P}^2$, must be conserved as well. In contrast to energy and momentum, however, mass is not additive: $M \neq \sum m_i$. Some authors talk about the non-additivity of mass as if it were identical to its non-conservation (e.g. we find this statement in §9 of *Field Theory* by Landau and Lifshitz [6].) In fact, as I emphasized above, in general neither masses nor energies or momenta are conserved for individual particles participating in a reaction; not even the particles themselves are. Hence, it is incorrect to speak of mass nonconservation as something in contrast to conservation of energy and momentum.

8.11. Einstein's thought experiment. Of course, the concept of the mass of two photons flying away from each other looks rather strange. However, it was by using this very idea that Einstein came to discover the rest energy of a massive body in 1905. He noticed that having emitted 'two amounts of light' in opposite directions, the body at rest continues to stay at rest but that its mass in this thought experiment diminishes. In the laboratory reference frame both the body and the center of inertia of the two photons are at rest. Consequently, the mass of the initial body equals the sum of two masses: that of the resulting body and that of the system of two photons.

8.12. Positronium annihilation. *Nihil* in Latin means *nothing.* A positronium is an 'atom' consisting of a positron and an electron. The reaction in which a positronium converts to two photons $e^+e^- \to \gamma\gamma$ was given the name annihilation, perhaps because at that time photons were not considered particles of matter. Annihilation conserves M because E and $\mathbf{P}$ are conserved. In the initial state M equals the sum of masses of the electron and the positron [minus the binding energy, which is small and in this context irrelevant (see below)]. In the final state M equals the sum of energies of two photons in the positronium's rest frame. The rest energy of the electron and the positron thus transforms completely into the energy (kinetic) of the photons, but the masses of the initial and final states are identical in this process, exactly as follows from the conservation of total energy and total momentum.

8.13. Meson decays. Likewise, when a K meson decays into two or three π mesons, the kaon's rest energy transforms into the sum of total energies of the pions, each of which has the form $e = e_k + m$. However, the mass of a system of two or three pions produced in the decay of a kaon equals the kaon mass.

8.14. What do we call 'matter'? In any decay the rest energy transforms into the energy of motion, while the total energy of an isolated system remains conserved. The mass of the system is conserved but the masses of its individual particles are not. Massive particles decay into less massive particles, or sometimes into massless ones. In elementary particle physics we call 'particles of matter' not only massive particles such as protons and electrons, but also very light neutrinos and massless photons, and even gravitons (see below). Today's quantum field theory treats all of them on an equal basis.

8.15. Energy without particles? Matter does not disappear in decay and annihilation reactions leaving behind only energy like the Cheshire cat would leave behind only its smile. In all these processes the carriers of energy are particles of matter. Energy without matter ('pure energy') has never been observed in any process studied so far. True, this might be not so for so-called dark energy, which was discovered in the last years of the XX century. Dark energy manifests itself in the accelerating expansion of the Universe. (The evidence for this accelerating expansion is found in recession velocities of remote supernovas.) Three-fourths of the entire energy in the Universe is dark energy and its carrier appears to be the vacuum. The remaining quarter is carried by ordinary matter (5%) and dark matter (20%). Dark energy does not affect processes with ordinary matter observed in laboratories. In a laboratory experiment energy is always carried by particles.

9 Non-free particles

9.1. Bodies and particles. All physical bodies consist of elementary particles. Such elementary particles as the proton and the neutron are themselves made up of 'more elementary particles' — quarks and gluons. Such particles as the electron and the neutrino appear at our current level of understanding as truly elementary particles. The feature common for the proton and the electron is that the masses of all protons in the world are strictly identical, as are the masses of all electrons. In contrast to this, the masses of all macroscopic bodies of the same type, say, of all 10-cent coins, are only approximately equal. Practically the difference between two coins arises because the process of minting coins is far from being ideal. What is more important here is that the mass of a coin is not well defined because different energy levels of a coin are practically degenerate, while the mass of the nearest excited state of a proton exceeds the proton mass by several hundred MeV. Therefore Nature mints ideally identical protons.

9.2. Mass of a gas. In all the cases discussed above, particles moved away freely when the mass of the system of particles was greater than the sum of their masses. Let us turn now to a situation in which they are not free to move away. This situation is found, for example, in the frequently discussed thought experiment with a gas of molecules or photons in a closed vessel at rest. The total momentum of this gas is zero because the gas is isotropic: $\mathbf{P} = \sum \mathbf{p}_i = 0$. Hence, the total mass M of this gas equals its total energy E (and in this case it is identical to E_0) and hence to the sum of energies of individual particles: $M = E = \sum e_i$.

9.3. Mass of a heated gas. When gas in a nonmoving vessel is heated, its total momentum remains unchanged and equal to zero while the total energy increases because the kinetic energy of every particle increases. As a result, the mass of the gas as a whole increases, while the mass of each individual particle remains unchanged. (Sometimes a wrong statement may be encountered in the literature that the masses of particles (or photons) increase as their kinetic energies are increased.)

9.4. Mass of a hot iron. In the same manner, the mass of an iron must increase as it heats up, even though the masses of the vibrating atoms remain the same. However, the set of formulas (10)–(12) written for a system of free particles cannot be applied to the iron since the particles (atoms in this case) are not free but are tied into the crystal lattice of the metal. Obviously, an increase in the iron mass is too small to be measurable.

10 Atoms and atomic nuclei

10.1. On formulas (10)–(12). Why are formulas (10)–(12) unsuitable for dealing with such non-free particles as electrons in atoms and nucleons in atomic nuclei? First and foremost, on account of the uncertainty relation these particles do not possess precisely defined momenta. The smaller the volume to which they are confined, the greater is the uncertainty of their momenta.

10.2. Uncertainty relation. The laws of quantum mechanics, and the uncertainty relation as one among them, are very important both for atoms and for nuclei. As we know, the product of the momentum uncertainty Δp and the coordinate uncertainty Δx must be not smaller than the quantum of action $\hbar$. Hence, particles within atoms have no definite momenta and only possess a certain total momentum.

10.3. Energy of the field. Another reason why formulas (10)–(12) are not valid inside atoms is the fact that the space between individual particles in an atom is essentially not empty but filled with a material medium, i.e. physical fields. The space inside the atom is filled with an electromagnetic field and the space inside a nucleus, by a much denser and stronger field, often described as the meson field.

10.4. Real and virtual particles. In classical theory particles and fields are concepts that cannot be reduced to one another. In quantum field theory we use the language of Feynman diagrams, which reduce the concept of a field to that of a virtual particle for which $e^2-\mathbf{p}^2 \neq m^2$. We say about such particles that they are off mass shell. (Particles that are called on mass shell are real particles and for them $e^2 - \mathbf{p}^2 = m^2$.) Also, the 4-momentum $p_i = (e, \mathbf{p})$ is conserved at each vertex of the diagram.

10.5. Binding energy. As a result of the presence of the field, we need to take into account in formula (10), $E = e_1 + e_2$, the field energy of two closely interacting particles, say, in the deuteron, the nucleus of heavy hydrogen. Consequently, $M < m_1 + m_2$. The quantity $\varepsilon = m_1 + m_2 - M$ is known as the binding energy. The mass of the deuteron is less than the mass of the proton plus that of the neutron of which deuteron consists. The binding energy of nucleons in deuteron is 2.2 MeV. To break deuteron into nucleons we need to spend an amount of energy equal to or greater than the binding energy. The atomic nuclei of all other elements of the periodic Mendeleyev table also owe their existence to the binding energy of their nucleons in the nucleus.

10.6. Fusion and fission of nuclei. We know that the binding energy per nucleon rises to a maximum at the beginning of the periodic Mendeleyev table for the helium nucleus and in the middle of the Table for the iron nucleus. This is why huge amounts of kinetic energy are released when helium is formed from hydrogen in fusion reactions in the Sun and in hydrogen bombs. In nuclear reactors and atomic bombs, kinetic energy is released by fission reactions when heavy nuclei of uranium and plutonium break into lighter nuclei from the middle of the periodic Mendeleyev Table.

10.7. Chemical reactions. Substantially lower energy, on the order of electron-volts, is released in chemical reactions. It is caused by differences in binding energies in various chemical compounds. However, the source of kinetic energy in both chemical and nuclear reactions is the difference between the masses of initial and final particles (molecules or nuclei) that take part in these reactions. Since molecules and even atomic nuclei are nonrelativistic bound systems and the concept of potential energy is applicable to their components, the corresponding mass differences can be calculated using this concept. Thus, one can explain the released energy in terms of potential energy transforming into kinetic energy.

10.8. Coulomb's law. The binding energy of electrons in atoms is much lower than the electron mass. Hence, the concept of binding energy in atoms can be explained in terms of the nonrelativistic concept of potential energy. The binding energy ε equals (with a minus sign) the sum of positive kinetic energy of the bound particle and its negative potential energy. The potential energy of, say, an electron in a hydrogen atom is given by Coulomb's law (in units, in which $\hbar$, $c = 1$):

$$U = -\frac{\alpha}{r}, \tag{17}$$

where $\alpha = e^2/\hbar c = 1/137$ and e is the electron charge.

10.9. More about potential energy. The concept of potential energy is defined only in the Newtonian limit (see Landau and Lifshitz,*Mechanics*, [7]: §5 "The Lagrange function of a system of material points" and §6 "Energy"). The sum of kinetic and potential energies is conserved. If one of the two interacting particles is essentially relativistic, or both are, the concept of potential energy is inapplicable.

10.10. Electromagnetic field. The Coulomb field in the theory of relativity is the 0th component of the 4-potential of the electromagnetic field $A_i (i = 0, 1, 2, 3)$. The source of the field of a particle with electric charge e is the 4-dimensional electromagnetic current given in the next paragraph. The interaction between two moving particles works through propagation of the field from one charge to the other. It is described by the so-called Green's function or the propagator of an electromagnetic field. (In quantum electrodynamics, we speak of propagation of virtual photons. The potential A_i is a 4-vector because the spin of the photon equals unity.)

10.11. Important clarification. If a virtual photon carries away a 4-momentum q, then 4-momenta of the charged particle prior to the emission of a photon $p_{\rm in}$ and after its emission $p_{\rm fi}$ satisfy the condition $p_{\rm in} - p_{\rm fi} = q$. The 4-vector p in the expression ep_i/E for the conserved current is $p = (p_{\rm in} + p_{\rm fi})/2$, and $E = \sqrt{E_{\rm in} E_{\rm fi}}$. As $p_{\rm in}^2 = p_{\rm fi}^2 = m^2$, so $qp = 0$. (I denoted energy here by the letter E because e in the expression for current stands for charge. We are clearly short of letters.)

10.12. Gluons and quarks. A gluon's spin also equals unity. At first glance, the interaction between gluons and quarks is completely analogous to the interaction between photons and electrons. Not at second glance, though. The point is that all electrons carry the same electric charge while quarks have three different color charges. A quark emitting or absorbing a gluon may change its color. Clearly, this means that gluons must themselves be colored. It can be shown that there must be eight different color species of gluons. While photons are electrically neutral, gluons carry color charges.

10.13. Quantum chromodynamics. It might seem that color-charged gluons must be intense emitters of gluons, being a sort of 'luminous light'. In fact, quantum chromodynamics — the theory of interaction between quarks and gluons — has a spectacular property known as confinement. In contrast to electrons and photons, colored quarks and gluons do not exist in a free state. These colored particles are locked 'for life' inside colorless (white) hadrons. They can only change their incarceration locality. There are no Feynman diagrams with lines of free gluons or free quarks.

11 Gravitation

11.1. Gravitational orbits. Various emblems often show the orbits of electrons in atoms resembling the orbits of planets. It should be clear from the above that according to quantum mechanics, there are no such orbits in atoms. On the other hand, quantum effects are absolutely infinitesimal for macroscopic bodies, all the more so for such heavy ones as planets. Consequently, their orbits are excellently described by classical mechanics.

11.2. Newton's constant. The potential energy of the Earth in the gravitational field of the Sun is given by Newton's law

$$U = -\frac{GMm}{r}, \tag{18}$$

where M is the solar mass, m is the mass of the Earth, r is distance between their centers, and G is Newton's constant:

$$G = 6.71 \times 10^{-39}\hbar c\ [\mathrm{GeV}/c^2]^{-2}. \tag{19}$$

(Here we use units in which $c \neq 1$.)

11.3. The quantity $p_i p_k/e$. The source of gravitation in Newtonian physics is mass. In the theory of relativity the source of gravitation is the quantity $p_i p_k/e$, which plays the role of a kind of 'gravitational current'. (The reader will recall that p_i is the energy-momentum 4-vector, and $i = 0, 1, 2, 3$. Consequently, the 'gravitational current' has four independent components instead of the ten that a most general symmetrical four-dimensional tensor would have.) The propagation of the field from the source to the 'sink' is described by Green's function of the gravitational field or the propagator of the graviton — a massless spin-2 particle. This propagator is proportional to $g^{il}g^{km} + g^{im}g^{kl} - g^{ik}g^{lm}$, where g^{ik} is a metric tensor. (As in the case of the photon discussed above, the 4-momentum of the graviton is $q = p_{\mathrm{in}} - p_{\mathrm{fi}}$ and the 4-momentum in the expression for current is $p = (p_{\mathrm{in}} + p_{\mathrm{fi}})/2$, while $e = \sqrt{e_{\mathrm{in}} e_{\mathrm{fi}}}$. We are again short of letters! This time, letters for indices.)

11.4. The graviton. Like the photon, the graviton is a massless particle. This is the reason why Newton's and Coulomb's potentials have the form $1/r$. However, in contrast to the photon, which cannot emit photons, the graviton can and must emit gravitons. In this respect the graviton resembles gluons, which emit gluons.

11.5. The Planck mass. Elementary particle physics often uses the concept of the Planck mass:

$$m_{\mathrm{P}} = \sqrt{\frac{\hbar c}{G}}. \tag{20}$$

In units in which $c, \hbar = 1$ we have $m_{\mathrm{P}} = 1/\sqrt{G} = 1.22 \cdot 10^{19}$ GeV.

The gravitational interaction between two ultrarelativistic particles increases as the square of their energy E in the center-of-inertia reference frame. It reaches maximum strength at $E \sim m_{\mathrm{P}}$ as the distance between the particles approaches $r \sim 1/m_{\mathrm{P}}$. However, let us return from these fantastically large energies and short distances to apples and photons in gravitational fields of the Earth and the Sun.

11.6. An apple and a photon. Consider a particle in a static gravitational field, for instance, that of the Sun. The source of the field is the quantity $P_l P_m/E$ where P_l is the 4-momentum of the Sun and E is its energy. In the rest frame of the Sun l, $m = 0$ and $P_l P_m/E = M$, where M is the solar mass. In this case the numerator of the propagator of the gravitational field $g^{il}g^{km} + g^{im}g^{kl} - g^{ik}g^{lm}$ is $2g^{i0}g^{k0} - g^{ik}g^{00}$, and the tensor quantity $p_i p_k$ times the numerator of the propagator reduces to a simple expression $2e^2 - m^2$. Hence, for a nonrelativistic apple of mass m the 'gravitational charge' equals m while for a photon with energy e it equals $2e$. Note the coefficient 2. Kinetic energy is attracted twice as strongly as the hidden energy locked in mass. This simple derivation of the coefficient 2 makes unnecessary the complicated derivation of paper [8] using isotropic coordinates.

11.7. A photon in the field of the Sun. The interaction of photons with the gravitational field must cause a deflection of a ray of light propagating from a remote star and passing close to the solar disk. In 1915 Einstein calculated the deflection angle and showed that it must be $4GM/c^2R \simeq 1.75''$. (Here, M and R denote the solar mass and solar radius, respectively.) This prediction was confirmed during the solar eclipse of 1919, which stimulated a huge surge of interest in the theory of relativity.

11.8. An atom in the field of the Earth. As a nonrelativistic body on the Earth moves upwards, its potential energy increases in proportion to its mass. Correspondingly, the difference between energies of two levels of an atomic nucleus must be the higher, the higher the floor of the building in which this nucleus is located.

11.9. A photon's energy is conserved. On the other hand, the frequency ω of a photon propagating through a static gravitational field, and correspondingly its total energy $e = \hbar\omega$, should remain unchanged.

As a result, a photon emitted on the ground floor of a building from a transition between two energy levels of a nucleus will be unable to produce a reverse transition in the same nucleus on the upper floor. This theoretical prediction was confirmed in the 1960s by Pound and Rebka [9] who used the just discovered Mössbauer effect, which makes it possible to measure the tiniest shifts in nuclear energy levels.

However, the wavelength changes. A photon propagating through a static gravitational field like a stone has its total energy e and frequency ω conserved. However, its momentum and therefore wavelength change as the distance to the gravitating body changes.

11.10. Refractive index. As a photon moves away from the source of a gravitational field, its velocity increases and tends to c, and when it approaches the source, it decreases. Hence, the gravitational field, like a transparent medium, has a refractive index. This is a visually clear explanation of the deflection of light in the field of the Sun and in the gravitational lenses of galaxies. Shapiro experimentally discovered the decrease in the velocity of photons near the Sun when measuring the delay of the radar echo returned by planets.

11.11. Clocks and gravitation. Ordinary clocks, like atomic clocks, are ticking the faster, the higher they are lifted. Let two synchronized clocks A and B be placed on the first floor. If we move clock A to the second floor and then, say, a day later, move clock B to the second floor as well, clock A will be ahead of B as A has been ticking faster than B for 24 hours. Nevertheless, both A and B will continue to serve as identically reliable stopwatches. When every point in space is assigned an individual clock, one in fact assumes that all clocks tick at a rate that is independent of the distance to gravitating bodies (in our case, on which floor of the building they are). However, this is not true for ordinary clocks. In order to distinguish extraordinary clocks from ordinary clocks, we will refer to extraordinary ones as 'cloned'. As we saw above, the frequency of light measured using clocks placed on various floors is independent of the floor number. If, however, it is measured with 'cloned' local clocks, we discover that it is the lower, the higher is the floor. One interpretation of the Pound–Rebka experiment, stating that the energy of a vertically moving photon decreases with height, like the kinetic energy of a stone thrown upwards, is based on precisely this argument. However, a drop in kinetic energy of the stone is accompanied with an increase in its potential energy, so that the total energy is conserved. Now, a photon has no potential energy, so that its energy in a static gravitational field remains constant.

12 Epistemology and linguistics

12.1. Physics and epistemology. *Episteme* in Greek means knowledge. Epistemics is the science of knowledge, a relatively young branch of epistemology, the theory of knowledge and cognition. Obviously, the problems I discuss in this talk concern not only physics but epistemology, too.

12.2. Physics and semantics. The Greek attribute 'semanticos' (signifying) was used in linguistics already by Aristotle. However, what are the links tying the science of languages — linguistics — and semantics — the science of words and symbols, an element of linguistics — to physics? This is the right moment to recall the words allegedly said by V A Fock: "Physics is an essentially simple science. The most important problem in it is to understand what each letter denotes." XX century physics drastically changed our understanding of what a vacuum and matter are, and connected in a new way such properties of matter as energy, momentum, and mass. The elaboration of the fundamental concepts of physics has not been completed and is unlikely to end in the foreseeable future. This is one of the reasons why it is so important to choose the adequate words and letters when discussing physical phenomena and theories.

12.3. Concepts glued together'. Newton's *Principia* 'glued together' the concepts of mass and matter (substance): "mass is proportional to density and volume." In Einstein's papers mass is 'glued together' with inertia and gravitation (the inertial and gravitational masses). And energy is glued to matter.

12.4. The archetype. According to dictionaries, an archetype is the historically original form (the protoform), the original concept or word, or the original type (prototype). The concept of the archetype keenly interested Pauli, who in 1952 published a paper on the effect of archetypical notions on the creation of natural-science theories by Kepler. It is possible that the concept of mass is just the archetypical notion that glued together the concepts of matter, inertia, and weight.

12.5. Atom and archetype. *Atom and Archetype* — that was the title chosen for the English translation from German of the book [10] presenting the correspondence between Wolfgang Pauli and the leading German proponent of psychoanalysis Carl Jung, covering the period from 1932 to 1958. W. Pauli and C. Jung discussed, among other things, the material nature of time and the possibility of communicating with people who lived several centuries or millennia before us. It is widely known that Pauli treated rather seriously the effect named after him: when he walked into an experimental laboratory, measuring equipment broke down.

12.6. Poets on terminology. David Samoilov on words: "We wipe them clean as we clean glass. This is our trade." Vladimir Mayakovskii: "The street is writhing for want of tongue. It has no nothing for yelling or talking." (*Translated by Nina Iskandaryan.*) Many an author responds to the dearth of precise terms and inability to use them by resorting to meaningless words like 'rest mass' which impart smoothness and 'energetics' to texts, just as 'blin' [3] does to ordinary speech.

[3] 'Blin' is a slang euphemism for a 'four-letter word' in vulgar Russian.

12.7. How to teach physics. Terms need 'wiping clean' and 'unglueing'. The 'umbilical cord' connecting the modern physical theory with the preceding 'mother theory' needs careful cutting in teaching. (In the case of the theory of relativity the mother was the 'centaur' composed of Maxwell's field theory and Newton's mechanics, with relativistic mass serving as the umbilical cord.) Let us recall the title of F Klein's famous book *Elementary Mathematics from an Advanced Standpoint.* The landscape of modern physics must be contemplated from an advanced standpoint: not from a historical gully but from the pinnacle of symmetry principles. I firmly believe that it is unacceptable to claim that the dependence of mass on velocity is an experimental fact and thus hide from the student that it is a mere interpretational 'factoid'. (Dictionaries explain that a factoid looks very much like a fact but is trusted only because we find it in printed texts.)

13 Concluding remarks

13.1. The '$\mathbf{E = mc^2}$ problem': could it be avoided? One is tempted to think that the '$E = mc^2$ problem' would not arise from the first place if the quantity E/c^2 — the proportionality coefficient between velocity and momentum — were identified with a new physical quantity christened as, say, 'inertia' or 'iner'; it would be identical to mass as momentum tended to zero. As a result, mass would become 'rest inertia'. Likewise, another new quantity could be introduced — 'heaviness' or 'grav' — $p_i p_k/E$ reducing to mass at zero momentum. But physicists preferred 'to refrain from multiplying entities' and from introducing new physical quantities. They formulated instead new, more general relations between old quantities, for example $E^2 - \mathbf{p}^2c^2 = m^2c^4$ and $\mathbf{p} = \mathbf{v}E/c^2$.

Unfortunately, many authors attempt to retain even in relativistic physics such nonrelativistic equations as $\mathbf{p} = m\mathbf{v}$, and such nonrelativistic glued-up concepts as 'mass is a measure of inertia' and 'mass is a measure of gravitation'; as a result, they prefer to use the notion of velocity-dependent mass. It is amazing how again and again a physicist would choose the first of these paths (new equations) in his research papers and the second one (old glued-up concepts) in science-popularizing and pedagogical activities. This could of course only produce unbelievable confusion in the minds of those who read popular texts and blindly follow the authority.

13.2. On the reliability of science. An opinion that has become widely publicized recently is that science in general and physics in particular are untrustworthy. Many popularizers of science create the impression that the theory of relativity proved Newton's mechanics wrong just as chemistry proved alchemy wrong and astronomy proved astrology wrong. Such declarations are a crude distortion of the essence of scientific revolutions. Newton's mechanics remains a correct science today, in the XXI century, and will continue to be correct forever. The discovery of the theory of relativity only put bounds on the domain of applicability of Newton's mechanics to velocities much smaller than the speed of light c. It also demonstrated its approximate nature in this domain (to within corrections of the order of v^2/c^2). Similarly, the discovery of quantum mechanics put bounds on the domain of applicability of classical mechanics to phenomena for which the quantity of action is large in comparison with the quantum of action $\hbar$. Quite to the contrary, the domain where astrology and alchemy exist is that of prejudice, superstition, and ignorance. It is rather funny that those who compare Newton's mechanics with astrology typically believe that mass depends on velocity.

13.3. Recent publications. Additional information on the aspects discussed above can be found in [11, 12].

13.4. On the title. My good friend and expert in the theory of relativity read the slides of this talk and advised me to drop Pythagoras's name from the title. I chose not to follow his advice as in the relativity-related literature I had never come across a discussion of right-angled triangles without the approximate extraction of square roots.

13.5. Acknowledgment I am grateful to T Basaglia, A Bettini, S I Blnnikov, V F Chub, M A Gottlieb E G Gulyaeva, E A Ilyina, C Jarllscog, V I Kisin, B A Klumov, B L Okun, S G Tikhodeev, M B Voloshin, V R Zoller for their advice and help. The work was supported by the grants NSh-5603.2006.2, NSh-4568.2006.2 and RFBR-07-02-00830-a.

References

[1] Okun L B "Formula Einshteina: $E_0 = mc^2$. 'Ne smeetsya li Gospod' Bog'?" *Usp. Fiz. Nauk* **178** 541 (2008) ["The Einstein formula $E_0 = mc^2$ 'Isn't the Lord laughing'?" *Phys. Usp.* **51** 513 (2008)], arXiv:0808.0437

[2] Klein F Elementarmathematik vom höeheren Standpunkte aus. Erster Band, Arithmetik, Algebra, Analysis 3 Auflage (Berlin: Verlag fon Julius Springer, 1924) [Translated into English: Elementary Mathematics from an Advanced Standpoint, Arithmetics, Algebra, Analysis (New York: Dover Publ., 2007); Translated into Russian (Moscow: Nauka, 1987)]

[3] Poincaré H "Sur la dynamique de l'electron", *Rendiconti del Circolo Matematico di Palermo* **21** 129 (1906) [Translated into Russian: "O dinamike elektrona" ("On the dynamics of electron") Izbrannye Trudy (Selected Works) Vol. 3 (Moscow: Nauka, 1974) p. 433]

[4] Minkowski H "Raum und Zeit" *Phys. Z.*, **10**, 104–111 (1909) [Translated into Russian: "Prostranstvo i Vremya" ("Space and time"), in Lorentz H A, Poincare H, Einstein A, Minkowski H *Printsip Otnositel'nosti. Sbornik Rabot Klassikov Relyativizma* (The Principle of Relativity. Collected Papers of Classics of Relativism) (Eds V K Frederiks, D D Ivanenko) (Moscow - Leningrad: ONTI, 1935) pp. 181 - 203]

[5] Hawking S *The Universe in a Nutshell* (New York: Bantam Books, 2001) [Translated into Russian (Translated from English by A Sergeev) St.-Petersburg: Amfore, 2007)]

[6] Landau L D, Lifshitz E M *Teoriya polya* (The Classical Theory of Fields) (Moscow: Nauka, 1988) [Translated into English (Amsterdam: Reed Elsevier, 2000)]

[7] Landau L D, Lifshitz E M *Mekhanika* (Mechanics) (Moscow: Nauka, 1988) [Translated into English (Amsterdam: Elsevier Sci., 2003)]

[8] Okun L B "The concept of mass" *Phys. Today* **42** (6) 31–36 (1989); Okun L B "Putting to rest mass misconceptions" *Phys. Today* **43** (5) 13, 15, 115, 117 (1990)

[9] Pound R V, Rebka G A (Jr.) "Apparent weight of photons" *Phys. Rev. Lett.* **4** 337–341 (1960)

[10] Pauli W, Jung C *Atom and Archetype: Pauli/Jung Letters. 1932–1958.* (Ed. C A Meyer) (Princeton, NJ: Princeton Univ. Press, 2001); Meier C A (Herausgegeben) Wolfgang Pauli, C. G. Jung *Ein Briefwechsel 1932–1958* Berlin: Springer-Verlag, (1992)

[11] Okun L B "Chto takoe massa? (Iz istorii teorii otnositel'nosti)" ('What is mass? (From the history of relativity theory)'), in *Issledovaniya po Istorii Fiziki i Mekhaniki.* 2007 (Research on the History of Physics and Mechanics. 2007) (Executive Ed. G M Idlis) (Moscow: Nauka, 2007)

[12] Okun L B "The evolution of the concepts of energy, momentum and mass from Newton and Lomonosov to Einstein and Feynman", in *Proc. of the 13th Lomonosov Conf. August 23, 2007* (Singapore: World Scientific) (in press)

This page intentionally left blank

Name Index

Energy and Mass in Relativity Theory

Energy and Mass in Relativity Theory presents about 30 pedagogical papers published by the author over the last 20 years. They deal with concepts central to relativity theory: energy E, rest energy E_0, momentum $\mathbf{p}$, mass m, velocity $\mathbf{v}$ of particles of matter, including massless photons for which $v = c$. Other related subjects are also discussed.

According to Einstein's equation $E_0 = mc^2$, a massive particle at rest contains rest energy which is partly liberated in the nuclear reactions in the stars and the Sun, as well as in nuclear reactors and bombs on the Earth. The mass entering Einstein's equation does not depend on velocity of a body. This concept of mass is used in the physics of elementary particles and is gradually prevailing in the modern physics textbooks.

This is the first book in which Einstein's equation is explicitly compared with its popular though not correct counterpart $E = mc^2$, according to which mass increases with velocity. The book will be of interest to researchers in theoretical, atomic and nuclear physics, to historians of science as well as to students and teachers interested in relativity theory.

World Scientific
www.worldscientific.com
6833 hc

ISBN-13 978-981-281-411-1
ISBN-10 981-281-411-6