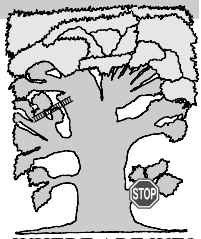
Chapter 3

BEYOND THE SCHRÖDINGER EQUATION



WHERE ARE WE?

Where are we?

The problems considered in the present chapter are shown as a small side-branch at the base of the TREE.[1]

An example

Copper, silver and gold – many people want to know everything about them (especially about the latter). The yellow shine of this metal has hypnotized humanity for centuries. Few people know that the colour of gold, as calculated assuming the infinite velocity of light (as it is in the Schrödinger equation), would be silver.[2] The Schrödinger equation fails. Here is an example of three diatomics: Cu_2, Ag_2, Au_2 ($Z_{Cu} = 29$, $Z_{Ag} = 47$, $Z_{Au} = 79$).

BOND LENGTH (Å)	Cu	Ag	Au
Non-relativistic calculations	2.26	2.67	**2.90**
Relativistic calculations	2.24	2.52	2.44
Experimental results	2.22	2.48	**2.47**

It is seen that the error of calculations within relativistic theories is of the order of 1%–2%, while the non-relativistic calculations lead to errors of the order of 2%, 8%, 20%, respectively. This is a *huge* discrepancy for such a quantity as bond length.

What is it all about

A glimpse of classical relativity theory (⧫) **p. 93**

- The vanishing of apparent forces
- The Galilean transformation
- The Michelson–Morley experiment
- The Galilean transformation crashes

[1] This chapter owes much to the presentation given by L. Pisani, J.-M. André, M.-C. André, E. Clementi, *J. Chem. Educ*. 70 (1993) 894–901, as well as to the work of my friends J.-M. André, D.H. Mosley, M.-C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, *Exploring Aspects of Computational Chemistry: Vol. I, Concepts*, Presses Universitaires de Namur, pp. 150–166 (1997), *Vol. II, Exercises*, Presses Universitaires de Namur, pp. 249–272 (1997), and J.-M. André, M.-C. André, "*Une introduction à la théorie de la relativité classique et quantique à l'usage des chimistes*", Namur, 1999.

[2] P. Pyykkö, *Chem. Rev*. 88 (1988) 563; also P. Pyykkö, *ibid*. 97 (1997) 597.

The greater the velocity of an object the greater the errors in Newton dynamics. Electrons have a greater velocity when close to nuclei of large electric charges.[3] This is why relativistic corrections may turn out to be important for heavy elements.

relativistic mass effect

The Schrödinger equation is incompatible with special relativity theory. This has to be corrected somehow. This is far from being solved, but progress so far shows the Schrödinger equation, the spin of a particle, etc. in a new light.

Why is this important?

The subject of the present chapter addresses the very foundations of physics, and in principle has to be treated on an equal footing with the postulates of quantum mechanics. The Schrödinger equation of Chapter 2 does not fulfil (as will be shown in the present chapter) the requirements of relativity theory, and therefore is in principle "illegal". In the present chapter, Dirac's attempt to generalize the Schrödinger equation to adapt it to relativity theory will be described. If one assumes that particle velocities are small compared to that of light, then from this more general theory one obtains the Schrödinger equation. Also the

[3] This is easy to estimate. From Appendix H on p. 969 it follows that the mean value of the kinetic energy of an electron described by the 1*s* orbital in an atom of atomic number Z is equal to $\bar{T} = \frac{1}{2}Z^2$ (in a.u.). On the other hand, for a rough estimation of the electron velocity v, one may write $\bar{T} = \frac{mv^2}{2}$. This results in the expression $v = Z$ valid in a.u., while the velocity of light $c = 137.036$ a.u. The largest Z known hardly exceeds a hundred. It is seen, therefore, that if an atom with $Z > 137$ existed, then 1s electrons would attain velocities exceeding the velocity of light. Even if this calculation is nothing but a rule of thumb, there is no doubt that when Z increases a certain critical Z value is approached (the so called *relativistic mass effect*).

notion of the spin, which was introduced as a postulate in Chapter 1, follows as a natural consequence of the relativistic theory. One may draw the conclusion that the present chapter addresses "the foundations of foundations" and therefore should occupy a prominent position in the TREE, instead of representing a small side branch (as it does now). However, the relativistic effects, even if visible in chemistry, do not play an important role in the case of the light elements (covering almost the whole of organic chemistry as well as almost the whole of biology). This is why I have chosen a rather pragmatic ("non-fundamental") way of presentation. This chapter is mainly for those readers who are interested in:

- "the foundations of foundations"
- very accurate calculations for small atoms and molecules
- calculations for systems containing heavy elements

What is needed?

- The postulates of quantum mechanics (Chapter 1, necessary).
- Operator algebra (Appendix A, p. 889, necessary).
- Vector and scalar potentials (Appendix G, p. 962, necessary).

Classical works

In 1881 the American physicist Albert Michelson and in 1887 with Edward Morley carried out some experiments showing that the speed of light is the same in the directions perpendicular and parallel to the Earth's orbit, i.e. the Earth's orbital velocity did not change the speed of light with respect to the Earth. The results were published in the *American Journal of Science*, 22 (1881) 120 under the title "*The Relative Motion of the Earth and the Luminiferous Aether*", and *ibid.*, 34 (1887) 333 (with similar title). ★ In 1889 the Irish physicist George Francis FitzGerald made the conjecture that, if all moving objects were foreshortened in the direction of their motion, this would account for the strange results of the Michelson–Morley experiment. This was published in *Science*, 13 (1889) 390 with the title "*The Ether and the Earth's Atmosphere*". ★ The revolutionary special relativity theory (that explained this in detail) was developed by Albert Einstein in an article entitled "*Zur Elektrodynamik bewegter Körper"* published in *Annalen der Physik (Leipzig)*, 17 (1905) 891. ★ The article is based largely on the ideas of the Dutchman Hendrik Antoon Lorentz, who independently of FitzGerald[4] proposed the Lorentz transformation (of space and time) in 1904. The transformation accounted for the contraction of moving objects, as predicted by FitzGerald. The paper "*Electromagnetic Phenomena in a System Moving with any Velocity less than that of Light*" was published in *Proceedings of the Academy of Sciences of Amsterdam*, 6 (1904) 809. ★ The German mathematician Hermann Minkowski realized that the work of Lorentz and Einstein could best be understood using a non-Euclidean space of the space and time variables. His first paper on this subject was *"Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körper*" published in *Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen* (1908). ★ The Soviet physicist Vladimir A. Fock derived the first relativistic wave equation for a particle (published in *Zeitschrift für Physik*, 39 (1926) 226),

[4] It was pointed out to Lorentz in 1894 that FitzGerald had published something similar. He wrote to FitzGerald, but the latter replied that indeed he has sent a half-page article to *Science*, but he did not know "*whether they ever published it*". Afterwards Lorentz took every opportunity to stress that FitzGerald was first to present the idea.

then the German Walter Gordon did the same and also published in *Zeitschrift für Physik*, 40 (1926) 117. Finally, a similar theory was proposed independently by the Swede Oskar Klein in *Zeitschrift für Physik*, 41 (1927) 407. The Austrian Erwin Schrödinger also derived the same equation, and this is why it is sometimes called "*the equation with many fathers*". ★ A more advanced quantum mechanical theory (for a single particle) adapted to the principles of relativity was given by the British Paul Adrien Maurice Dirac in several articles in *Proceedings of the Royal Society (London)* entitled "*The Fundamental Equations of Quantum Mechanics*", A109 (1926) 642, "*Quantum Mechanics and a Preliminary Investigation of the Hydrogen Atom*", *ibid*., A110 (1926) 561, "*The Quantum Theory of Radiation*", *ibid*., A114 (1927) 243, "*The Quantum Theory of the Electron*", *ibid*., A117 (1928) 610, and "*The Quantum Theory of the Electron. Part II*" *ibid*., A118 (1928) 351. ★ An extension of relativistic quantum theory to many-electron problems (still approximate) was published by the American Gregory Breit in *Physical Review* with the title "*The Effect of Retardation on the Interaction of Two Electrons*", 34 (1929) 553, and then in two other papers entitled "*Fine Structure of He as a Test of the Spin Interaction of Two Electrons*", *ibid.*, 36 (1930) 383, and "*Dirac's Equation and the Spin–Spin Interactions of Two Electrons*", *ibid.*, 39 (1932) 616. ★ In 1948 the Americans Richard Feynman and Julian Schwinger as well as the Japanese Shinichiro Tomonaga independently invented the quantum electrodynamics (QED), which successfully combined quantum theory with the special theory of relativity and produced extremely accurate results.

3.1 A GLIMPSE OF CLASSICAL RELATIVITY THEORY

3.1.1 THE VANISHING OF APPARENT FORCES

The three principles of Newtonian[5] dynamics were taught to us in school. The first principle, that a free body (with no acting force) moves uniformly along a straight line, seems to be particularly simple. It was not so simple for Ernest Mach though.

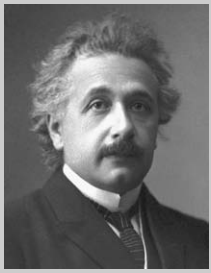
Ernest Mach (1838–1916), Austrian physicist and philosopher, professor at the Universities of Graz, Prague, and Vienna, godfather of Wolfgang Pauli. Mach investigated supersonic flows. In recognition of his achievements the velocity of sound in air (1224 km/hour) is called Mach 1.

Mach wondered how one recognizes that no force is acting on a body. The contemporary meaning of the first principle of Newton dynamics is the following. First, we introduce a Cartesian coordinate system x, y, z to the Universe, then remove all objects except one from the Universe, to avoid any interactions. Then, we measure equal time intervals using a spring clock and put the corresponding positions of the body in the coordinate system (we are thus there with our clock and our ruler…). The first principle says that the positions of the body are along a straight line and equidistant. What a crazy procedure! The doubts and dilemmas of Mach were implanted in the mind of Albert Einstein.

[5] For Newton's biography see Chapter 7.

Albert Einstein (1879–1955) born in Ulm (Germany) studied at the ETH, Zurich. Considered by many as genius of all times. As a teenager and student, Einstein rejected many social conventions. This is why he was forced to begin his scientific career at a secondary position in the Federal Patent Office. Being afraid of his supervisor, he used to read books hidden in a drawer (he called it the "Department of Physics").

The year of his 26th birthday was particularly fruitful ("miraculous year" 1905). He published three fundamental papers: about relativity theory, about Brownian motion and about the photoelectric effect. For the last, Einstein received the Nobel Prize in 1921. After these publications he was appointed professor at the University of Zurich and then at the University of Prague. From 1914 Einstein headed the Physics Institute in Berlin, which was founded especially for him. He emigrated to the USA in 1933, because of menacing persecution, because of his Jewish origin. Einstein worked at the Institute for Advanced Study in Princeton in the USA. He died there in 1955. According to his will, his ashes were dispersed over America from the air.



This Bern Patent Office employee also knew about the dramatic dilemmas of Lorentz, which we will talk about in a moment. Einstein recalls that there was a clock at a tram stop in Bern. Whenever his tram moved away from the stop, the modest patent office clerk asked himself what would the clock show, if the tram had the velocity of light. While other passengers probably read their newspapers, Einstein had questions which led humanity on new pathways.

Let us imagine two coordinate systems (each in 1D): O "at rest" (we assume it inertial[6]) while the coordinate system O' moves with respect to the first in a certain way (possibly very complicated). The position of the moving point may be measured in O giving the number x as the result, while in O' on gets the result x'. These numbers are related one to another (f is a function of time t):

$$x' = x + f(t). \tag{3.1}$$

If a scientist working in a lab associated with the coordinate system O would like to calculate the force acting on the above mentioned point body, he would get a result proportional to the acceleration, i.e. to $\frac{\mathrm{d}^2x}{\mathrm{d}t^2}$. If the same were done by

[6]That is, in which the Newton equation is satisfied. A coordinate system associated with accelerating train is not inertial, because there is a non-zero force acting on everybody in the train, while the acceleration with respect to the train coordinate system is zero.

another scientist working in a lab in O', then he would obtain *another force*, this time proportional to the acceleration computed as $\frac{d^2x'}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2f}{dt^2}$. The second term in this force is the *apparent* force. Such apparent forces (from the point of view of an observer on the ground) are encountered in lifts, on a carousel, etc.

Let us note an important consequence: if one postulates the same forces (and therefore the same dynamics) in two coordinate systems, $f(t)$ has to be a *linear* function (because its second derivative is equal to zero). This means that a family of all coordinate systems that moved uniformly with respect to one another would be characterized by the same description of phenomena because the forces computed would be the same (*inertial systems*).

Physics textbooks written in the two laboratories associated to O and O' would describe all the phenomena in the same way.

inertial systems

The linearity condition gives $x' = x + vt$, if at $t = 0$ O and O' coincide. Let us take a fresh look at this equation: x' represents a linear combination of x and t, which means that *time and the linear coordinate mix together*. One has two coordinates: one in the O coordinate system and the other in the O' coordinate system. Wait a minute! Since the time and the coordinate are on an equal footing (they mix together), maybe one may also have the time (t) appropriate for (i.e. running in) the O and the time (t') running in the O' coordinate system?

Now, a crucial step in the reasoning. Let us write *in a most general way* a linear transformation of coordinates and time that ensures the two systems equivalent (no apparent forces):

$$x' = Ax + Bt,$$
$$t' = Cx + Dt.$$

First of all the corresponding transformation matrix *has to be* invertible (i.e. non-singular), because inversion simply means exchanging the roles of the two coordinate systems and of the observers flying with them. Thus, one has:

$$x = \bar{A}x' + \bar{B}t',$$
$$t = \bar{C}x' + \bar{D}t'.$$

Next, A has to be equal to $\bar{A}$, because the measurements of length in O and O', i.e. x and x', cannot depend on whether one looks at the O coordinate system from O', or at O' from O. If the opposite were true, then one of the coordinate systems would be privileged (treated in a special way). This, however, is impossible, because the two coordinate systems differ *only* in that O' flies from O with velocity

v, while O flies from O' with velocity $-v$, but the space is isotropic. The same *has to happen* with the time measurements: on board O, i.e. t, and on board O', i.e. t', therefore $D = \bar{D}$. Since (from the inverse transformation matrix) $\bar{A} = \frac{D}{AD-BC}$ and $\bar{D} = \frac{A}{AD-BC}$, therefore we have

$$\frac{D}{AD-BC} = A, \qquad \frac{A}{AD-BC} = D.$$

From this $\frac{D}{A} = \frac{A}{D}$ follows, or:

$$A^2 = D^2. \tag{3.2}$$

From the two solutions: $A = D$ and $A = -D$, one has to choose only $A = D$, because the second solution would mean that the times t and t' have opposite signs, i.e. when time run forwards in O it would run backwards in O'. Thus, we have

$$A = D. \tag{3.3}$$

3.1.2 THE GALILEAN TRANSFORMATION

The equality condition $A = D$ is satisfied by the Galilean transformation, in which the two coefficients are equal to 1:

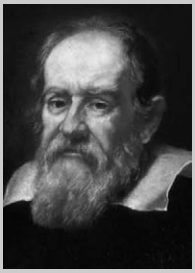
$$x' = x - vt,$$
$$t' = t,$$

where position x and time t, say, of a passenger in a train, is measured in a platform-fixed coordinate system, while x' and t' are measured in a train-fixed coordinate system. There are no apparent forces in the two coordinate systems related by the Galilean transformation. Also, the Newtonian equation is consistent with our intuition, saying that time flows at the same pace in any coordinate system.

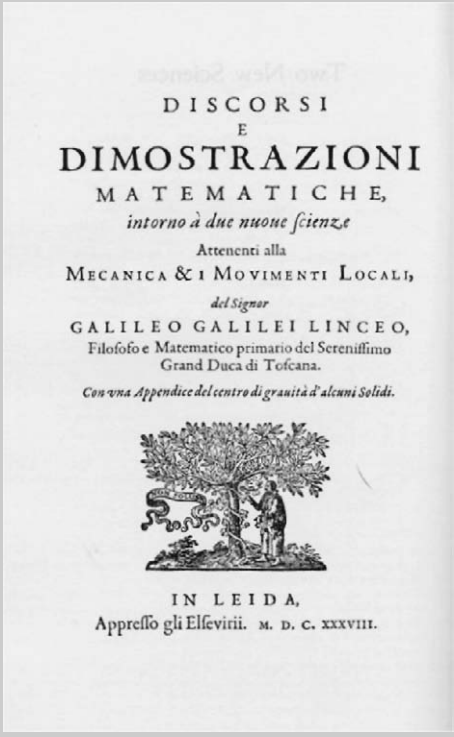
3.1.3 THE MICHELSON–MORLEY EXPERIMENT

Hendrik Lorentz indicated that the Galilean transformation represents only *one possibility* of making the apparent forces vanish, i.e. assuring that $A = D$. Both constants need not be equal to 1. As it happens that such a generalization is forced by an intriguing experiment performed in 1887.

Michelson and Morley were interested in whether the speed of light differs, when measured in two laboratories moving with respect to one another. According to the Galilean transformation, the two velocities of light should be different, in the same way as the speed of train passengers (measured with respect to the platform)



Galileo Galilei (1564–1642), Italian scientist, professor of mathematics at the University of Pisa. Only those who have visited Pisa are able to appreciate the inspiration (for studying the free fall of bodies of different materials) from the *incredibly* leaning tower. Galileo's *opus magnum* (right-hand side) has been published by Elsevier in 1638. Portrait by Justus Sustermans (XVII century).



DISCORSI
E
DIMOSTRAZIONI
MATEMATICHE,
intorno à due nuoue ſcienze
Attenenti alla
MECANICA & I MOVIMENTI LOCALI,
del Signor
GALILEO GALILEI LINCEO,
Filoſofo e Matematico primario del Sereniſſimo
Grand Duca di Toſcana.
Con vna Appendice del centro di grauità d'alcuni Solidi.

IN LEIDA,
Appreſſo gli Elſevirii. M. D. C. XXXVIII.

Hendrik Lorentz (1853–1928), Dutch scientist, professor at Leiden. Lorentz was very close to formulating the special theory of relativity.

Albert Michelson (1852–1931), American physicist, professor in Cleveland and Chicago, USA. He specialized in the precise measurements of the speed of light.

His older colleague Edward Williams Morley was American physicist and chemist, professor of chemistry at Western Reserve University in Cleveland, USA.

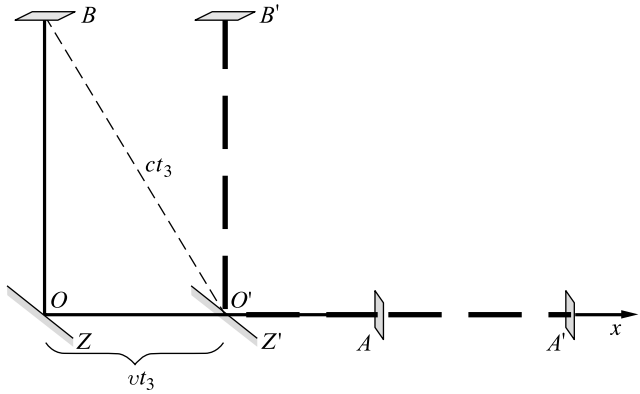


Fig. 3.1. The Michelson–Morley experimental framework. We have two identical V-shaped right-angle objects, each associated with a Cartesian coordinate system (with origins O and O'). The first is at rest, while the second moves with velocity v with respect to the first (along coordinate x). We are going to measure the velocity of light in two laboratories rigidly bound to the two coordinate systems. The mirrors are at the ends of the objects: A, B in O and A', B' in O', while at the origins two semi-transparent mirrors Z and Z' are installed. Time $2t_3 \equiv t_\downarrow$ is the time for light to go down and up the vertical arm.

differs depending on whether they walk in the same or the opposite direction with respect to the train motion. Michelson and Morley replaced the train by Earth, which moves along its orbit around the Sun with a speed of about 40 km/s. Fig. 3.1 shows the Michelson–Morley experimental framework schematically. Let us imagine two identical right-angle V-shaped objects with all the arm lengths equal to L.

Each of the objects has a semi-transparent mirror at its vertex,[7] and ordinary mirrors at the ends. We will be interested in how much time it takes the light to travel along the arms of our objects (back and forth). One of the two arms of any object is oriented along the x axis, while the other one must be orthogonal to it. The mirror system enables us to overlap the light beam from the horizontal arm (x axis) with the light beam from the perpendicular arm. If there were any difference in phase between them we would immediately see the interference pattern.[8] The second object moves along x with velocity v (and is associated with coordinate system O') with respect to the first ("at rest", associated with coordinate system O).

3.1.4 THE GALILEAN TRANSFORMATION CRASHES

In the following we will *suppose that the Galilean transformation is true.* In coordinate system O the time required for light to travel (round-trip) the arm along the

[7] Such a mirror is made by covering glass with a silver coating.

[8] From my own experience I know that interference measurement is very sensitive. A laser installation was fixed to a steel table 10 cm thick concreted into the foundations of the Chemistry Department building, and the interference pattern was seen on the wall. My son Peter (then five-years-old) just touched the table with his finger. Everybody could see immediately a large change in the pattern, because the table bent.

x axis ($T_{\rightarrow}$) and that required to go perpendicularly to axis ($T_{\downarrow}$) are the same:

$$T_{\rightarrow} = \frac{2L}{c}, \qquad T_{\downarrow} = \frac{2L}{c}.$$

Thus, in the O coordinate system, there will be no phase difference between the two beams (one coming from the parallel, the other from the perpendicular arm) and therefore no interference will be observed. Let us consider now a similar measurement in O'. In the arm co-linear with x, when light goes in the direction of $\boldsymbol{v}$, it has to take more time (t_1) to get to the end of the arm:

$$ct_1 = L + vt_1, \tag{3.4}$$

than the time required to come back (t_2) along the arm:

$$ct_2 = L - vt_2. \tag{3.5}$$

Thus, the total round-trip time $t_{\rightarrow}$ is[9]

$$t_{\rightarrow} = t_1 + t_2 = \frac{L}{c-v} + \frac{L}{c+v} = \frac{L(c+v)+L(c-v)}{(c-v)(c+v)} = \frac{2Lc}{c^2-v^2} = \frac{\frac{2L}{c}}{1-\frac{v^2}{c^2}}. \tag{3.6}$$

What about the perpendicular arm in the coordinate system O'? In this case the time for light to go down (t_3) and up will be the same (let us denote total flight time by $t_{\downarrow} = 2t_3$, Fig. 3.1). Light going down goes along the hypotenuse of the rectangular triangle with sides: L and $\frac{vt_{\downarrow}}{2}$ (because it goes down, but not only, since after $\frac{t_{\downarrow}}{2}$ it is found at $x = \frac{vt_{\downarrow}}{2}$). We will find, therefore, the time $t_{\downarrow}$ from Pythagoras' theorem:

$$\left(c\frac{t_{\downarrow}}{2}\right)^2 = L^2 + \left(v\frac{t_{\downarrow}}{2}\right)^2, \tag{3.7}$$

or

$$t_{\downarrow} = \sqrt{\frac{4L^2}{c^2-v^2}} = \frac{2L}{\sqrt{c^2-v^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.8}$$

The times $t_{\downarrow}$ and $t_{\rightarrow}$ do not equal each other for the moving system and there will be the interference, we were talking about a little earlier.

However, there is absolutely no interference!
Lorentz was forced to put the Galilean transformation into doubt
(apparently the foundation of the whole science).

[9] Those who have some experience with relativity theory, will certainly recognize the characteristic term $1 - \frac{v^2}{c^2}$.

3.1.5 THE LORENTZ TRANSFORMATION

The interference predicted by the Galilean transformation is impossible, because physical phenomena would experience the two systems in a different way, while they differ only by their relative motions ($\boldsymbol{v}$ has to be replaced by $-\boldsymbol{v}$).

To have everything back in order, Lorentz assumed that, when a body moves, its length (measured by using the unit length at rest in the coordinate system O) along the direction of the motion, *contracts* according to equation

$$l = L\sqrt{1-\frac{v^2}{c^2}}. \tag{3.9}$$

length contraction

If we insert such a length l, instead of L, in the expression for $t_{\rightarrow}$, then we obtain

$$t_{\rightarrow} = \frac{\frac{2l}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L\sqrt{1-\frac{v^2}{c^2}}}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.10}$$

and everything is perfect again: $t_{\downarrow} = t_{\rightarrow}$. No interference. This means that x' (i.e. the position of a point belonging to a rigid body as measured in O') and x (the position of the same point measured in O) have to be related by the following formula. The coordinate x measured by an observer in his O is composed of the intersystem distance OO', i.e. vt plus the distance O' – point, but measured using the length unit of the observer in O, i.e. the unit that resides in O (thus, *non-contracted* by the motion). Because of the contraction 1: $\sqrt{1-\frac{v^2}{c^2}}$ of the rigid body the latter result will be *smaller* than x' (recall, please, that x' is what the observer measuring the position in his O' obtains), hence:

$$x = x'\sqrt{1-\frac{v^2}{c^2}} + vt, \tag{3.11}$$

or:

$$x' = \frac{x}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{vt}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.12}$$

which means that in the linear transformation

$$A = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.13}$$

$$B = -\frac{v}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.14}$$

As we have already shown, in linear transformation $(x', t') \to (x, t)$ the diagonal coefficients have to be equal ($A = D$), therefore

$$t' = Cx + Dt, \tag{3.15}$$

$$D = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{3.16}$$

To complete determination of the linear transformation we have to calculate the constant C. Albert Einstein assumed, that if Professors Oconnor and O'connor began (in their own coordinate systems O and O') measurements on the velocity of light, then despite the different distances gone (x and x') and different flight times[10] (t and t'), both scientists would get *the same velocity of light* (denoted by c).

In other words $x = ct$ and $x' = ct'$.

Using this assumption and eqs. (3.12) and (3.16) we obtain:

$$ct' = Dct - vDt, \tag{3.17}$$

while multiplying equation (3.15) for t' by c we get:

$$ct' = cCx + Dct. \tag{3.18}$$

Subtracting both equations we have

$$0 = -vDt - cCx \tag{3.19}$$

or

$$C = -\frac{vtD}{cx} = -\frac{vtD}{cct} = -\frac{vD}{c^2}. \tag{3.20}$$

Thus we obtain the full Lorentz transformation, which assures that no of the systems is privileged, and the same speed of light in *both* systems:

$$x' = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}x - \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}}t,$$

$$t' = -\frac{v}{c^2}\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}x + \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}t.$$

[10] At the moment of separation $t = t' = 0$.

Let us check first of all, whether if $v = 0$, then everything is OK. Yes it is. Indeed, the denominator equals 1 and we have $t' = t$ and $x' = x$. Let us see what would happen if the velocity of light were equal to infinity. Then, the Lorentz transformation becomes identical with the Galilean. In general, after expanding t' and x' in a power series of v^2/c^2 we obtain

$$x' = -vt + x + \frac{1}{2}(-vt + x)\frac{v^2}{c^2} + \cdots,$$
$$t' = t + \left(-\frac{x}{v} + \frac{t}{2}\right)\frac{v^2}{c^2} + \cdots$$

This means that only at very high velocity v, may we expect differences between both transformations.

Contraction is relative

Of course, Professor O'connor in his laboratory O' would not believe in Professor Oconnor (sitting in his O lab) saying that he (O'connor) has a contraction of the rigid body. And indeed, if Professor O'connor measured the rigid body using his standard length unit (he would not know his unit is contracted), then the length measured would be exactly the same as that measured just before separation of the two systems, when both systems were at rest. In a kind of retaliation, Professor O'connor could say (smiling) that it is certainly not him who has the contraction, but his colleague Oconnor. He would be right, because for him, his system is at rest and his colleague Oconnor flies away from him with velocity $-v$. Indeed, our formula (3.11) makes that very clear: expressing in (3.11) t by t' from the Lorentz transformation leads to the point of view of Professor O'connor

$$x' = x\sqrt{1 - \frac{v^2}{c^2}} - vt', \tag{3.21}$$

and one can indeed see an evident contraction of the rigid body of Professor Oconnor. This way, neither of these two coordinate systems is privileged. That is very, very good.

3.1.6 NEW LAW OF ADDING VELOCITIES

Our intuition was worked out for small velocities, much smaller than the velocity of light. The Lorentz transformation teaches us something, which is against intuition. What does it mean that the velocity of light is constant? Suppose we are flying with the velocity of light and send the light in the direction of our motion. Our intuition tells us: the light will have the velocity equal to $2c$. Our intuition has to be wrong. How it will happen?

Let us see. We would like to have the velocity in the coordinate system O, but first let us find the velocity in the coordinate system O', i.e. $\frac{\mathrm{d}x'}{\mathrm{d}t'}$. From the Lorentz

transformation one obtains step by step:

$$\frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\mathrm{d}x - \frac{v}{\sqrt{1-\frac{v^2}{c^2}}}\mathrm{d}t}{-\frac{v}{c^2}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\mathrm{d}x + \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\mathrm{d}t} = \frac{\frac{\mathrm{d}x}{\mathrm{d}t} - v}{1 - \frac{v}{c^2}\frac{\mathrm{d}x}{\mathrm{d}t}}. \tag{3.22}$$

By extracting $\frac{\mathrm{d}x}{\mathrm{d}t}$ or using the symmetry relation (when $O' \to O$, then $v \to -v$) we obtain:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\frac{\mathrm{d}x'}{\mathrm{d}t'} + v}{1 + \frac{v}{c^2}\frac{\mathrm{d}x'}{\mathrm{d}t'}} \tag{3.23}$$

or

VELOCITY ADDITION LAW

$$V = \frac{v' + v}{1 + \frac{vv'}{c^2}}. \tag{3.24}$$

In this way we have obtained a new rule of adding the velocities of the train and its passenger. Everybody naively thought that if the train velocity is v and, the passenger velocity with respect to the train corridor is v', then the velocity of the passenger with respect to the platform is $V = v + v'$. It turned out that this is not true. On the other hand when both velocities are small with respect to c, then indeed one restores the old rule

$$V = v' + v. \tag{3.25}$$

Now, let us try to fool Mother Nature. Suppose our train is running with the velocity of light, i.e. $v = c$, and we take out a torch and shine the light forward, i.e. $\frac{\mathrm{d}x'}{\mathrm{d}t'} = v' = c$. What will happen? What will be the velocity V of the light with respect to the platform? $2c$? From (3.24) we have $V = \frac{2c}{2} = c$. This is precisely what is called the universality of the speed of light. Now, let us make a bargain with Nature. We are hurtling in the train with the speed of light $v = c$ and walking along the corridor with velocity $v' = 5$ km/h. What will our velocity be with respect to the platform? Let us calculate again:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{5 + c}{1 + \frac{c}{c^2}5} = \frac{5 + c}{1 + \frac{5}{c}} = c\frac{5 + c}{5 + c} = c. \tag{3.26}$$

Once more we have been unable to exceed the speed of light c. One last attempt. Let us take the train velocity as $v = 0.95c$, and fire along the corridor a powerful

missile with speed $v' = 0.10c$. Will the missile exceed the speed of light or not? We have

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{0.10c + 0.95c}{1 + \frac{0.95c}{c^2}0.10c} = \frac{1.05c}{1 + 0.095} = \frac{1.05}{1.095}c = 0.9589c. \tag{3.27}$$

c is not exceeded. Wonderful formula.

3.1.7 THE MINKOWSKI SPACE-TIME CONTINUUM

The Lorentz transformation may also be written as:

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \begin{bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}.$$

What would happen if the roles of the two systems were interchanged? To this end let us express x, t by x', t'. By inversion of the transformation matrix we obtain[11]

$$\begin{bmatrix} x \\ ct \end{bmatrix} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \begin{bmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x' \\ ct' \end{bmatrix}. \tag{3.28}$$

relativity principle

We have perfect symmetry, because it is clear that the sign of the velocity has to change. Therefore:

none of the systems is privileged (relativity principle).

Now let us come back to Einstein's morning tram meditation[12] about what he would see on the tramstop clock if the tram had the velocity of light. Now we have the tools to solve the problem. It concerns the two events – two ticks of the clock observed in the coordinate system associated with the tramstop, i.e. $x_1 = x_2 \equiv x$, but happening at two different times t_1 and t_2 (differing by, say, one second, i.e. $t_2 - t_1 = 1$, this is associated with the corresponding movement of the clock hand).

[11] You may check this by multiplying the matrices of both transformations – we obtain the unit matrix.

[12] Even today Bern looks quite provincial. In the centre Albert Einstein lived at Kramgasse 49. A small house, squeezed by others, next to a small café, with Einstein's achievements on the walls. Einstein's small apartment is on the second floor showing a room facing the backyard, in the middle a child's room (Einstein lived there with his wife Mileva Marić and their son Hans Albert; the personal life of Einstein is complicated), and a large living room facing the street. A museum employee with oriental features says the apartment looks as it did in the "miraculous year 1905", everything is the same (except the wall-paper, she adds), and then: "*maybe this is the most important place for the history of science*".

What will Einstein see when his tram leaves the stop with velocity v with respect to the stop, or in other words when the tramstop moves with respect to him with velocity $-v$? He will see the same two events, but in his coordinate system they will happen at

$$t_1' = \frac{t_1}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{and} \quad t_2' = \frac{t_2}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}},$$

i.e. *according to the tram passenger* the two ticks at the tramstop will be separated by the time interval

$$t_2' - t_1' = \frac{t_2 - t_1}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Thus, when the tram ran through the streets of Bern with velocity $v = c$, the hands on the tramstop clock when seen from the tram would not move at all, and this second would be equivalent to eternity.

This is known as *time dilation*. Of course, for the passengers waiting at the tramstop (for the next tram) and watching the clock, its two ticks would be separated by exactly one second. If Einstein took his watch out of his waistcoat pocket and showed it to them through the window *they* would be amazed. The seconds will pass at the tramstop, while Einstein's watch would seem to be stopped. The effect we are describing has been double checked experimentally many times. For example, the meson lives such a short time (in the coordinate system associated with it), that when created by cosmic rays in the stratosphere, it would have no chance of reaching a surface laboratory before decaying. Nevertheless, as seen from the laboratory coordinate system, the meson's clock ticks very slowly and mesons are observable.

time dilation

Hermann Minkowski introduced the seminal concept of the four-dimensional space-time continuum (x, y, z, ct).[13] In our one-dimensional space, the elements of the Minkowski space-time continuum are events, i.e. vectors (x, ct), something happens at space coordinate x at time t, when the event is observed from coordinate system O. When the same event is observed in two coordinate sys-

Hermann Minkowski (1864–1909), German mathematician and physicist, professor in Bonn, Königsberg, Technische Hochschule Zurich, and from 1902 professor at the University of Göttingen.

[13] Let me report a telephone conversation between the PhD student Richard Feynman and his supervisor Prof. Archibald Wheeler from Princeton Advanced Study Institute (according to Feynman's Nobel

tems, then the corresponding x, t and x', t' satisfy the Lorentz transformation. It turns out that in both coordinate systems *the distance of the event from the origin of the coordinate system is preserved*. The square of the distance is calculated in a strange way as:

$$(ct)^2 - x^2 \tag{3.29}$$

for the event (x, ct). Indeed, let us check carefully:

$$\begin{aligned}(ct')^2 - (x')^2 &= \frac{1}{1-\frac{v^2}{c^2}}\left(-\frac{v}{c}x + ct\right)^2 - \frac{1}{1-\frac{v^2}{c^2}}\left(x - \frac{v}{c}ct\right)^2 \\ &= \frac{1}{1-\frac{v^2}{c^2}}\left[\frac{v^2}{c^2}x^2 + c^2t^2 - 2vxt - x^2 - \frac{v^2}{c^2}c^2t^2 + 2vxt\right] \\ &= \frac{1}{1-\frac{v^2}{c^2}}\left[\frac{v^2}{c^2}x^2 + c^2t^2 - x^2 - \frac{v^2}{c^2}c^2t^2\right] = (ct)^2 - (x)^2. \end{aligned} \tag{3.30}$$

There it is! This equation enabled Hermann Minkowski to interpret the Lorentz transformation (3.28) as a *rotation* of the event (x, ct) in the Minkowski space about the origin of the coordinate system (since any rotation preserves the distance from the rotation axis).

3.1.8 HOW DO WE GET $E = mc^2$?

The Schrödinger equation is invariant with respect to the Galilean transformation. Indeed, the Hamiltonian contains the potential energy which depends on inter-particle distances, i.e. on the differences of the coordinates, whereas the kinetic energy operator contains the second derivative operators which are invariant with respect to the Galilean transformation. Also, since $t = t'$, the time derivative in the time-dependent Schrödinger equation does not change.

> Unfortunately, both Schrödinger equations (time-independent and time-dependent) are not invariant with respect to the Lorentz transformation and therefore are illegal.

As a result, one cannot expect the Schrödinger equation to describe accurately objects that move with velocities comparable to the speed of light.

Lecture, 1965): "Wheeler: "*Feynman, I know why all electrons have the same charge and the same mass!*" F: "*Why?*" W: "*Because they are all the same electron!*" Then, Wheeler explained: "*suppose that the world lines which we were ordinarily considering before in time and space – instead of only going up in time were a tremendous knot, and then, when we cut through the knot by the plane corresponding to a fixed time, we would see many, many world lines and that would represent many electrons (...)*".

Let us consider a particle moving in the potential V. The Schrödinger equation has been "derived" (see p. 70) from the total energy expression:

$$E = \frac{p^2}{2m} + V, \tag{3.31}$$

where $\boldsymbol{p}$ is the momentum vector and m is the mass.

Einstein was convinced that nothing could be faster than light.[14] Therefore, what would happen if a particle were subject to a constant force? It would eventually attain the velocity of light and what would happen afterwards? There was a problem, and Einstein assumed that in the laboratory coordinate system in which the particle is speeded up, the particle will increase its ... mass. In the coordinate system fixed on the particle no mass increase will be observed, but in the laboratory system it will. We have to admire Einstein's courage. For millions of people the mass of a body represented an invariant characteristic of the body. How was the mass supposed to increase? Well, the increase law – Einstein reasoned – should be such that the particle was able to absorb *any* amount of the kinetic energy. This means that when $v \to c$, then we have to have $m(v) \to \infty$. One of the possible formulae for $m(v)$ may contain a factor typical of relativity theory [cf. eq. (3.16)]:

$$m(v) = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{3.32}$$

where m_0 is the so called rest mass of the particle (i.e. its mass measured in the coordinate system residing on the particle).[15] It is seen that if v/c were zero (as it is in the non-relativistic world), then m would be equal to m_0, i.e. to a constant as it is in non-relativistic physics.[16]

For the time being, the legitimacy of eq. (3.32) is questionable as being just one of the possible *ad hoc* suppositions. However, Einstein has shown that this

[14] Maybe this is true, but nothing in the special theory of relativity compels us to think that c is the maximum speed possible.

[15] Because of the speed-dependent mass in relativity theory it is impossible to separate the centre-of-mass motion.

[16] And therefore no corrections to the Schrödinger equation are needed. At the beginning of this chapter we arrived at the conclusion that electron velocity in an atom is close to its atomic number Z (in a.u.). Hence, for the hydrogen atom ($Z_H = 1$) one may estimate $v/c \simeq 0.7\%$, i.e. v of the electron in the $1s$ state represents a velocity of the order of 2100 km/s, which is probably very impressive for a car driver, but not for an electron. However, for gold ($Z_{\mathrm{Au}} = 79$) we obtain $v/c \simeq 51\%$. This means that in the atom of gold the electron mass is larger by about 15% with respect to its rest mass and therefore the relativistic effects are non-negligible. For such important elements as C, O, N (biology) the relativistic corrections may be safely neglected. A young graduate student, Grzegorz Łach, posed an interesting purely academic question (such questions and the freedom to discuss them represent the cornerstone and the beauty of university life): will the human body survive the switching off of relativistic effects? Most of the biomolecules would function practically without significant changes, but the heavy metal atoms in enzyme active sites might react differently in the chemical reactions in which they are involved. Would they indeed? Would the new direction be destructive for the body? Nobody knows. On the other hand, we have forgotten about the spin concept that follows in a consequent way only from relativistic quantum theory (see below). Without spin no world similar to ours is conceivable.

particular formula fits the existing equation of motion. First, after expanding the mass into the Taylor series one obtains something interesting

$$m(v) = m_0\left\{1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots\right\}, \tag{3.33}$$

kinetic energy

especially after multiplying the result by c^2:

$$mc^2 - m_0c^2 = \frac{m_0v^2}{2} + \text{smaller terms}. \tag{3.34}$$

It looks as if indeed the kinetic energy was stored directly in the mass m. Einstein deduced that it may be that the total kinetic energy of the body is equal to:

$$E = mc^2.$$

He convinced himself about this after calculating its time derivative. After assuming that eq. (3.32) is correct, one obtains:

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= c^2\frac{\mathrm{d}m}{\mathrm{d}t} = c^2\frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = m_0c^2\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \\
&= m_0c^2\left(-\frac{1}{2}\right)\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}\frac{-2v}{c^2}\frac{\mathrm{d}v}{\mathrm{d}t} = m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}v\frac{\mathrm{d}v}{\mathrm{d}t} \\
&= \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\frac{1}{1-\frac{v^2}{c^2}}v\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\left(1+\frac{\frac{v^2}{c^2}}{1-\frac{v^2}{c^2}}\right)v\frac{\mathrm{d}v}{\mathrm{d}t} \\
&= \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}v\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{v^2}{c^2}m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}v\frac{\mathrm{d}v}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} + v^2\frac{\mathrm{d}m}{\mathrm{d}t} = v\frac{\mathrm{d}(mv)}{\mathrm{d}t}.
\end{aligned}$$

Precisely the same equation is satisfied in non-relativistic mechanics, if E denotes the *kinetic energy*:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}(\frac{mv^2}{2})}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} = v\frac{\mathrm{d}(mv)}{\mathrm{d}t}. \tag{3.35}$$

Therefore in relativity theory

$$E_{\mathrm{kin}} = mc^2. \tag{3.36}$$

This formula has been verified in laboratories many times. For example, it is possible nowadays to speed electrons in cyclotrons up to a velocity that differs from c by $\frac{1}{8000000}c$. That corresponds to $1-\frac{v^2}{c^2}=\frac{1}{4000000}$, and the electron's mass m becomes 2000 times larger than its m_0. This means that the electron is pumped up with energy to such an extent that its mass is similar to that of the proton. The energy stored in mass is huge. If, from the mass of a 20000 TNT atomic bomb, one subtracted the mass of its ashes after explosion,[17] then one would obtain only about 1g! The energy freed from this 1g gives an effect similar to the apocalypse.

3.2 RECONCILING RELATIVITY AND QUANTUM MECHANICS

"The equation with many fathers" (Klein–Gordon, also Fock and Schrödinger...)

We would like to express the kinetic energy E_{kin} through the particle's momentum p, because we would then know how to obtain the corresponding quantum mechanical operators (Chapter 1, p. 18). To this end let us consider the expression

$$E_{\text{kin}}^2-\left(m_0c^2\right)^2=m^2c^4-m_0^2c^4=m_0^2c^4\left(\frac{1}{1-v^2/c^2}-1\right)$$
$$=\frac{m_0^2c^4}{1-v^2/c^2}\frac{v^2}{c^2}=m^2v^2c^2=p^2c^2. \tag{3.37}$$

Therefore,

$$E_{\text{kin}}=c\sqrt{p^2+m_0^2c^2} \tag{3.38}$$

and the total energy E in the external potential V

$$E=c\sqrt{p^2+m_0^2c^2}+V. \tag{3.39}$$

What if the particle is subject to an electromagnetic field, given by the electric field $\boldsymbol{\mathcal{E}}$ and the magnetic field $\boldsymbol{H}$ (or, the magnetic induction $\boldsymbol{B}$) in every point of the space? Instead of $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{H}$ (or $\boldsymbol{B}$) we may introduce two other quantities: the vector field $\boldsymbol{A}$ and the scalar field ϕ (see Appendix G). As we can show in classical electrodynamics,[18] the kinetic energy of the particle subject to an electromagnetic field is very similar to the same expression without the field (eq. (3.38)), namely, for a particle of charge q, the momentum $\boldsymbol{p}$ is to be replaced by $\boldsymbol{p}-\frac{q}{c}\boldsymbol{A}$ and the potential V by $q\phi$. Therefore, we obtain the following expression for the total

[17] R. Feynman, R.B. Leighton, M. Sands, "*Feynman Lectures on Physics*", Addison-Wesley Publishing Company, 1964.

[18] For example, H.F. Hameka, *Advanced Quantum Chemistry*, Addison-Wesley, Reading, MA, 1965, p. 40.

energy of the particle in an electromagnetic field

$$E = c\sqrt{\left(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2} + q\phi, \tag{3.40}$$

where $\boldsymbol{A}$ and ϕ represent functions of the particle's position.

If we wanted to use the last expression to construct the Hamiltonian, then we would find serious difficulty, namely, the momentum operator $\hat{\boldsymbol{p}} = -i\hbar\nabla$ (replacing $\boldsymbol{p}$ according to the appropriate postulate, Chapter 1) is under the square root sign, thus leading to non-linear operators. Brave scientists noted, however, that if you squared both sides, the danger would disappear. We would obtain

$$(E - q\phi)^2 = c^2\left[\left(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right]. \tag{3.41}$$

All this has been, and still is, a sort of groping and guessing from some traces or indications.

> The equations corresponding to physical quantities will be transformed to the corresponding operator equations, and it will be assumed that both sides of them will act on a wavefunction.

Oskar Klein (1894–1977) was the youngest son of the chief rabi of Sweden and professor of mathematics and physics at Stockholm Högskola. Walter Gordon (1893–1940) until 1933 was a professor at the University of Hamburg, and afterwards resided in Sweden.

What to insert as the operator $\hat{H}$ of the energy E? This was done by Schrödinger (even before Fock, Klein and Gordon). Schrödinger inserted what he had on the right-hand side of his time-dependent equation

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi, \quad \text{i.e.} \quad \hat{H} = i\hbar\frac{\partial}{\partial t}.$$

This way

$$\left(i\hbar\frac{\partial}{\partial t} - q\phi\right)^2 = c^2\left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right], \tag{3.42}$$

or after acting on the wave function we obtain the equation known as Klein–Gordon:

$$\left(i\hbar\frac{\partial}{\partial t} - q\phi\right)^2\Psi = c^2\left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right]\Psi. \tag{3.43}$$

This equation has at least one advantage over the Schrödinger equation: ct and x, y, z enter the equation on an equal footing, which is required by special relativity. Moreover, the Klein–Gordon equation is invariant with respect to the Lorentz transformation, whereas the Schrödinger equation is not. This is a prerequisite of any relativity-consistent theory and it is remarkable that such a simple derivation made the theory invariant. The invariance does not however mean that the equation is accurate. The Klein–Gordon equation describes a boson particle.

3.3 THE DIRAC EQUATION

3.3.1 THE DIRAC ELECTRONIC SEA

Paul Dirac used the Klein–Gordon equation to derive a Lorentz transformation invariant equation[19] for a single fermion particle. The Dirac equation is solvable only for several very simple cases. One of them is the free particle (Dirac), the other is an electron in the electrostatic field of a nucleus (Darwin, not *that* one).

One may add here a few other systems, e.g., the harmonic oscillator and that's it.

From eq. (3.38), in the case of a free particle $V = 0$, one obtains two sets of energy eigenvalues, one corresponding to the negative energies

$$E = -\sqrt{p^2c^2 + m_0^2c^4} \tag{3.44}$$

and the other corresponding to the positive energies

$$E = +\sqrt{p^2c^2 + m_0^2c^4}. \tag{3.45}$$

Dirac was not worried by the fact that both roots appear after an *ad hoc* decision to square the expression for the energy (eqs. (3.40) and (3.41)). As we can see,

Paul Adrien Maurice Dirac (1902–1984), British physicist theoretician, professor at universities in Cambridge, and then Oxford. Dirac was keen on hiking and climbing. He used to practise before expeditions by climbing trees near Cambridge, in the black outfit in which always gave his lectures.

He spent his last years at the University of Tallahassee (Florida, USA). On being guided through New York City, Dirac remembered old times. The guide remarked that there were visible changes, among others the buses had been painted pink. Dirac quietly agreed, adding that indeed they had, at least from one side…

[19] See J.D. Bjorken, S.D. Drell, "*Relativistic Quantum Mechanics*", McGraw-Hill, 1964.

Charles Galton Darwin (1887–1962), British physicist and mathematician, professor at University of Edinburgh, Scotland, grandson of the evolutionist Sir Charles Robert Darwin. Darwin investigated the scattering of α particles on atoms.

since the momentum may change from 0 to ∞ ($p = mv$, and for $v \to c$, we have $m \to \infty$), we therefore have the *negative energy continuum* and symmetrically located *positive energy continuum*, both continua separated by the energy gap $2m_0c^2$ (Fig. 3.2).

Dirac (when 26 years old) made the apparently absurd assumption that what people call a vacuum is in reality a sea of electrons occupying the negative energy continuum ("*Dirac electronic sea*"). The sea was supposed to consist of an infinite number of electrons, which had to imply catastrophic consequences concerning, for example, the mass of the Universe (infinite), but Dirac did not doubt or discourage: "*We see only those electrons, that have the positive energy"* – said Dirac. Then, he asked himself, whether one could somehow see those electrons that occupy the sea and answered yes, it is possible. According to Dirac it is sufficient to excite such an electron by providing the energy

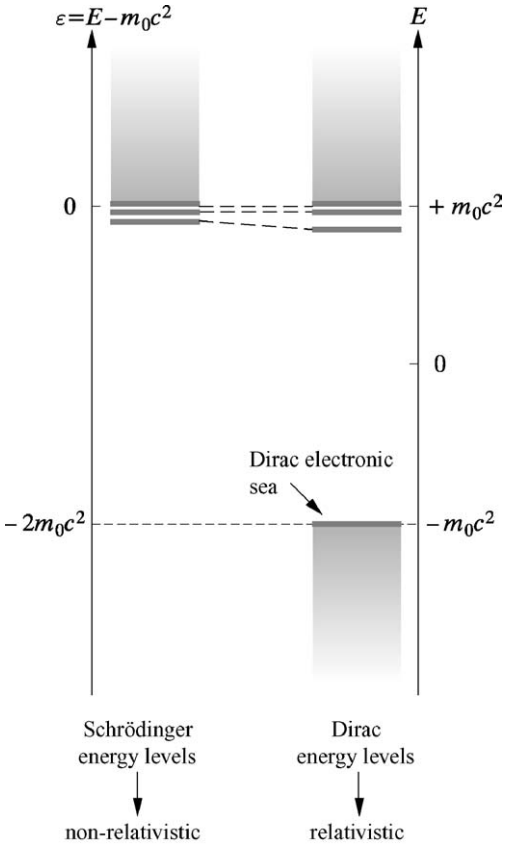


Fig. 3.2. Energy levels for the hydrogen atom according to Schrödinger (left hand side) and Dirac (right hand side). Shadowed areas correspond to the positive and negative energy continua.

of the order of $2m_0c^2$ to cover the energy gap (the energy $2m_0c^2$ is very large, of the order of 1 MeV). Then the sea electron would have positive energy and could be observed as others electrons with positive energy. However, besides the electron one, would leave a hole in the Dirac sea. Dirac has been severely molested about what this strange hole would correspond to in experimental physics. Once, when pushed too strongly, he said desperately that this is a ... proton. Some seemed to be satisfied, but others began to attack him furiously. However, then Dirac has demonstrated that the hole would have the dynamical and electrical properties of an electron, except that its sign would be opposite.[20] This has been nothing but a hypothesis for the existence of antimatter, a state of matter unknown at that time. Please imagine the shock, when three years later Carl Anderson reported the creation of electron–positron pairs from a vacuum after providing energy $2m_0c^2$. This was a day of glory for quantum theory.

positron

antimatter

Carl David Anderson (1905–1991), American physicist, professor at the Pasadena Institute of Technology. In 1932 Anderson discovered the positron, for which he received the Nobel Prize in 1936. He was also a co-discoverer of the muon.

In a moment we will see the determination with which Dirac attacked the Klein–Gordon equation, which we will write down a little differently:

$$\left[\frac{i\hbar\frac{\partial}{\partial t}-q\phi}{c}\right]^2-\left[\left(-i\hbar\nabla-\frac{q}{c}\boldsymbol{A}\right)^2+m_0^2c^2\right]=0. \tag{3.46}$$

Let us first introduce the following abbreviations:

$$\pi_0=\frac{i\hbar\frac{\partial}{\partial t}-q\phi}{c},\qquad \pi_\mu=-i\hbar\frac{\partial}{\partial\mu}-\frac{q}{c}A_\mu, \tag{3.47}$$

for $\mu=x,y,z$ or $1,2,3$.

[20] Paul Dirac, when a pupil in primary school, made his reputation after solving a riddle which goes very well with the person who thought out the positively charged electron (positron).
Three fishermen went overnight fishing and camping at a lake. After heavy fishing, around evening they put the fish in a bucket, and tired, fell asleep in the tent. At midnight one of the fishermen woke up and, tired of the whole escapade decided to take $\frac{1}{3}$ of all the fish, leave the tent quietly and go home. When he counted the fish in the bucket, it turned out that the number of fish was not divisible by 3. However, when he had thrown one fish back to the lake, the number was divisible by 3, he took his $\frac{1}{3}$ and went away. After a while a second fisherman woke up and did the same, and then the third. The question was, how many fish were in the bucket. Dirac's answer was: -2. Indeed, the number is indivisible by 3, but after the first fisherman threw away one fish the number was -3. He took his $\frac{1}{3}$, i.e. -1 fish, wrapped it up using a newspaper and went away leaving -2 fish splashing in the bucket. The same happened to each of the other two fishermen.

Dirac persisted in treating eq. (3.46) as $a^2 - b^2$ and therefore rewriting it in the form $(a+b)(a-b)$, i.e.

$$\left(\pi_0 + \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) = 0. \tag{3.48}$$

He was so self-assured, that he said eq. (3.48) has to be satisfied at any price by finding suitable unknowns α_i (independent of coordinates and time). The α's have to satisfy the following relations (anti-commutation relations)

$$\alpha_\mu^2 = 1, \tag{3.49}$$

$$\alpha_\mu \alpha_\nu + \alpha_\nu \alpha_\mu = 0 \quad \text{for } \mu \neq \nu. \tag{3.50}$$

anti-commutation

Indeed, using the anti-commutation relations one recovers the Klein–Gordon equation:

$$\begin{aligned}
&\left(\pi_0 + \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) \\
&= \pi_0^2 - \left[\sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right]^2 = \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} \alpha_\mu \alpha_\nu \pi_\mu \pi_\nu \\
&\quad - \sum_{\mu=x,y,z}^{3} (\alpha_\mu \alpha_0 + \alpha_0 \alpha_\mu) \pi_\mu m_0 c - \alpha_0 m_0^2 c^2 \\
&= \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} (\alpha_\mu \alpha_\nu + \alpha_\mu \alpha_\nu) \pi_\mu \pi_\nu - m_0^2 c^2 = \pi_0^2 - \sum_{\mu=x,y,z}^{3} \pi_\mu^2 - m_0^2 c^2.
\end{aligned}$$

Note that the α's cannot be just numbers, because no numbers can satisfy the anticommutation relation. They have to be matrices. Since we have four variables x, y, z, t, then we may expect matrices of order 4, but they could be larger. Here is one of the consistent choices of matrices:

$$\alpha_x = \begin{pmatrix} \mathbf{0} & \sigma_x \\ \sigma_x & \mathbf{0} \end{pmatrix}, \qquad \alpha_y = \begin{pmatrix} \mathbf{0} & \sigma_y \\ \sigma_y & \mathbf{0} \end{pmatrix},$$

$$\alpha_z = \begin{pmatrix} \mathbf{0} & \sigma_z \\ \sigma_z & \mathbf{0} \end{pmatrix}, \qquad \alpha_0 \equiv \beta = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}.$$

Please note the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$, defined on p. 28, determine electron spin. This is the first sign of what will happen later on: the Dirac equation will automatically describe the spin angular momentum.

3.3.2 THE DIRAC EQUATIONS FOR ELECTRON AND POSITRON

After the factorization described above Dirac obtained two operator equations. The Dirac equations (for the positron and electron) correspond to these operators acting on the wave function Ψ. Thus, we obtain the equation for the negative electron energies (positron)

$$\left(\pi_0 + \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\Psi = 0 \tag{3.51}$$

and for the positive electron energies (electron)

$$\left(\pi_0 - \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right)\Psi = 0. \tag{3.52}$$

These two equations are coupled together through the same function Ψ which has to satisfy both of them. This coupling caused a lot of trouble in the past. First, people assumed that the equation with the negative electron energies (positron equation) may be ignored, because the energy gap is so large that the Dirac sea is occupied whatever a chemist does with a molecule. This assumption turned out to cause some really vicious or weird performances of numerical procedures (see later on). The electron equation alone reads as

$$i\hbar \frac{\partial \Psi}{\partial t} = \left(q\phi + c \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c^2\right)\Psi. \tag{3.53}$$

If one is interested in *stationary states* (cf. p. 21), the wave function has the form $\Psi(x, y, z, t) = \Psi(x, y, z)e^{-i\frac{E}{\hbar}t}$, where we have kept the same symbol for the time independent factor $\Psi(x, y, z)$. After dividing by $e^{-i\frac{E}{\hbar}t}$ we obtain

stationary states

THE DIRAC EQUATION FOR STATIONARY ELECTRONIC STATES

$$\left(E - q\phi - \beta m_0 c^2 - c\boldsymbol{\alpha} \cdot \boldsymbol{\pi}\right)\Psi(x, y, z) = 0, \tag{3.54}$$

where $\beta = \alpha_0$.

The quantity $q\phi = V$ in future applications will denote the Coulomb interaction of the particle under consideration with the external potential.

3.3.3 SPINORS AND BISPINORS

The last equation needs a comment. Because the matrices α have dimension 4, then Ψ has to be a four component vector (known as *bispinor*, its connection to the

spinors and bispinors

spin concept will be shown later on)

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix},$$

where the first two components (ψ_1 and ψ_2, functions of class Q), for reasons that will become clear in a moment, are called *large components*, are hidden in vector $\boldsymbol{\psi}$, while the two *small components* (ϕ_1 and ϕ_2, functions of class Q)[21] are labelled by vector $\boldsymbol{\phi}$. Vectors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are called the *spinors*.

How to operate the N-component spinor (for $N = 4$ we have called them bispinors)? Let us construct the proper Hilbert space for the N-component spinors. As usual (p. 895), first, we will define the sum of two spinors in the following way:

$$\begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \cdots \\ \Psi_N \end{pmatrix} + \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \cdots \\ \Phi_N \end{pmatrix} = \begin{pmatrix} \Psi_1 + \Phi_1 \\ \Psi_2 + \Phi_2 \\ \cdots \\ \Psi_N + \Phi_N \end{pmatrix},$$

and then, the product of the spinor by a number γ:

$$\gamma \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \cdots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} \gamma\Psi_1 \\ \gamma\Psi_2 \\ \cdots \\ \gamma\Psi_N \end{pmatrix}.$$

Next, we check that the spinors form an Abelian group with respect to the above defined addition (cf. Appendix C, p. 903) and, that the conditions for the vector space are fulfilled (Appendix B). Then, we define the scalar product of two spinors

$$\langle \Phi | \Psi \rangle = \sum_{i=1}^{N} \langle \Phi_i | \Psi_i \rangle,$$

where the scalar products $\langle \Phi_i | \Psi_i \rangle$ are defined as usual in the Hilbert space of class Q functions. Then, using the scalar product $\langle \Phi | \Psi \rangle$ we define the distance between two spinors: $\|\Phi - \Psi\| \equiv \sqrt{\langle \Phi - \Psi | \Phi - \Psi \rangle}$ and afterwards the concept of the Cauchy series (the distances between the consecutive terms tend to zero). The Hilbert space of spinors will contain all the linear combinations of the spinors together with the limits of all the convergent Cauchy series.

[21] It will be shown that in the non-relativistic approximation the large components reduce to the wave function known from the Schrödinger equation, and the small components vanish. In eq. (3.54) the constant E as well as the function V individually multiply each component of the bispinor Ψ, while $\boldsymbol{\sigma} \cdot \boldsymbol{\pi} \equiv \alpha_x \pi_x + \alpha_y \pi_y + \alpha_z \pi_z$ denotes the "dot product" of the matrices α_μ, $\mu = x, y, z$, by the operators π_μ (in the absence of the electromagnetic field, it is simply the momentum operator component, see p. 962). The matrix β is multiplied by the constant $m_0 c^2$, then by the bispinor Ψ.

An operator acting on a spinor means a spinor with each component resulting from action on the corresponding component

$$\hat{A}\begin{pmatrix}\Psi_1\\ \Psi_2\\ \dots\\ \Psi_N\end{pmatrix}=\begin{pmatrix}\hat{A}\Psi_1\\ \hat{A}\Psi_2\\ \dots\\ \hat{A}\Psi_N\end{pmatrix}.$$

Sometimes we will use the notation, in which a matrix of operators acts on a spinor. In this case the result corresponds to multiplication of the matrix (of operators) and the vector (spinor)

$$\begin{pmatrix}\hat{A}_{11} & \hat{A}_{12} & \dots & \hat{A}_{1N}\\ \hat{A}_{21} & \hat{A}_{22} & \dots & \hat{A}_{2N}\\ \dots & \dots & \dots & \dots\\ \hat{A}_{N1} & \hat{A}_{N2} & \dots & \hat{A}_{NN}\end{pmatrix}\begin{pmatrix}\Psi_1\\ \Psi_2\\ \dots\\ \Psi_N\end{pmatrix}=\begin{pmatrix}\sum_j \hat{A}_{1j}\Psi_j\\ \sum_j \hat{A}_{2j}\Psi_j\\ \dots\\ \sum_j \hat{A}_{Nj}\Psi_j\end{pmatrix}.$$

3.3.4 WHAT NEXT?

In the following we will show

1. that the first two components of the bispinor are *much larger* than the last two
2. that in the limit $c \to \infty$ the Dirac equation gives the *Schrödinger equation*
3. that the Dirac equation accounts for the spin angular momentum of the electron
4. how to obtain, in a simple way, an approximate solution of the Dirac equation to the electron in the field of a nucleus ("*hydrogen-like atom*").

3.3.5 LARGE AND SMALL COMPONENTS OF THE BISPINOR

Using matrix multiplication rules, the Dirac equation (3.54) with bispinors can be rewritten in the form of *two equations with spinors* $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$:

$$\big(E - V - m_0c^2\big)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.55}$$

$$\big(E - V + m_0c^2\big)\boldsymbol{\phi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.56}$$

The quantity m_0c^2 represents the energy. Let us use this energy to *shift the energy scale* (we are always free to choose 0 on this scale): $\varepsilon = E - m_0c^2$, in order to make ε comparable in future to the eigenvalue of the Schrödinger equation (p. 70). We obtain

$$(\varepsilon - V)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.57}$$

$$\big(\varepsilon - V + 2m_0c^2\big)\boldsymbol{\phi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.58}$$

This set of equations corresponds to a single matrix equation:

$$\begin{pmatrix} V & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2 \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix}. \tag{3.59}$$

3.3.6 HOW TO AVOID DROWNING IN THE DIRAC SEA

When, in the past, the above equation was solved and the energy ε minimized (routine practice in the non-relativistic case) with respect to the variational parameters (see Chapter 5) in the trial spinors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$, then some serious numerical problems occurred. Either the numerical procedures diverged or the solutions obtained were physically unacceptable. The reason for this was that the existence of the Dirac sea had been totally ignored by neglecting eq. (3.51) for the positron and taking solely eq. (3.52) for electron motion. The variational trial functions felt, however, the presence of the Dirac sea electronic states (there was nothing in the theory that would prevent the electron from attempting to occupy negative energies) and the corresponding variational energies dived down the energy scale towards the abyss of the sea without a bottom.[22] The presence of the Dirac sea makes the Dirac theory, in fact, a theory of an infinite number of particles, whereas formally it was only a theory of a single particle in an external field. This kind of discomfort made people think of the possibility of describing the electron from the Dirac electronic sea by replacing the bispinors by the exact spinor (two components) theory.[23] Such exact separation has been reported by Barysz and Sadlej.[24]

An approximate (and simple) prescription was also invented to avoid the catastrophic drowning described above. Indeed, eq. (3.58) can be transformed without any problem to

$$\boldsymbol{\phi} = \left(1+\frac{(\varepsilon-V)}{2m_0c^2}\right)^{-1}\frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}.$$

Since $2m_0c^2$ represents a huge energy when compared to the kinetic energy $\varepsilon - V$, then the first parenthesis on the right-hand side is to a good approximation

[22] How severe the problem might be has been shown by M. Stanke, J. Karwowski, "*Variational Principle in the Dirac Theory: Spurious Solutions, Unexpected Extrema and Other Traps*" in "*New Trends in Quantum Systems in Chemistry and Physics*", vol. I, p. 175–190, eds. J. Maruani et al., Kluwer Academic Publishers. Sometimes an eigenfunction corresponds to a quite different eigenvalue. Nothing of that sort appears in non-relativistic calculations.

[23] Exact within the Dirac model.

[24] M. Barysz, A.J. Sadlej, J.G. Snijders, *Int. J. Quantum Chem.* 65 (1997) 225; M. Barysz, *J. Chem. Phys*. 114 (2001) 9315; M. Barysz, A.J. Sadlej, *J. Mol. Struct. (Theochem)* 573 (2001) 181; M. Barysz, A.J. Sadlej, *J. Chem. Phys.* 116 (2002) 2696. In the latter paper an exact solution to the problem was given. The two-component theory, although more appealing, both from the point of view of physics as well as computationally, implies a change in definition of the operators, e.g., the position operator is replaced by a quite complex expression. This fact, ignored in computations using two-component theories, has been analysed in the articles: V. Kellő, A.J. Sadlej, B.A. Hess, *J. Chem. Phys*. 105 (1996) 1995; M. Barysz, A.J. Sadlej, *Theor. Chem. Acc*. 97 (1997) 260; V. Kellő, A.J. Sadlej, *Int. J. Quantum Chem*. 68 (1998) 159; V. Kellő, A.J. Sadlej, *J. Mol. Struct. (Theochem)* 547 (2001) 35.

equal to 1. This means however that

$$\boldsymbol{\phi} \approx \frac{1}{2m_0 c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi},$$

which is known as "kinetic balancing". It was shown that the "*kinetically balanced*" trial function achieves the miracle[25] of the energy not tending to $-\infty$. The kinetic balancing indicates some fixed relation between $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$.

kinetic balancing

Let us focus now on $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$. This is a 2×2 matrix and in the absence of an electromagnetic field ($\boldsymbol{\pi}=\boldsymbol{p}$) one has:

$$\begin{aligned}\boldsymbol{\sigma}\cdot\boldsymbol{\pi} &= \sigma_x\hat{p}_x + \sigma_y\hat{p}_y + \sigma_z\hat{p}_z \\ &= \begin{pmatrix} 0 & \hat{p}_x \\ \hat{p}_x & 0 \end{pmatrix} + \begin{pmatrix} 0 & -i\hat{p}_y \\ i\hat{p}_y & 0 \end{pmatrix} + \begin{pmatrix} \hat{p}_z & 0 \\ 0 & -\hat{p}_z \end{pmatrix} = \begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}.\end{aligned}$$

It is seen that $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ is of the order of momentum mv, and for the small velocities of the order of $m_0 v$.

Hence, one obtains $\boldsymbol{\phi} \approx \frac{1}{2m_0 c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} \approx \frac{v}{2c}\boldsymbol{\psi}$, therefore the component $\boldsymbol{\phi}$ is for small v much smaller than the component $\boldsymbol{\psi}$,

which justifies the terms "small" and "large" components.[26]

3.3.7 FROM DIRAC TO SCHRÖDINGER – HOW TO DERIVE THE NON-RELATIVISTIC HAMILTONIAN?

The *approximate* relation ("kinetic balance") between the large and small components of the bispinor (that holds for small v/c) may be used to *eliminate* the small components[27] from (3.57) and (3.58). We obtain

$$(\varepsilon - V)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\frac{1}{2m_0 c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \tag{3.60}$$

$$(\varepsilon - V)\boldsymbol{\psi} - \frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.61}$$

[25] This remedy has not only an *ad hoc* character, but moreover does not work for the heaviest atoms, which are otherwise the most important target of relativistic computations.

[26] These terms refer to the positive part of the energy spectrum. For the negative continuum (Dirac sea) the proportion of the components is reversed.

[27] A more elegant solution was reported by Andrzej W. Rutkowski, *J. Phys. B.* 9 (1986) 3431, *ibid.* 19 (1986) 3431, *ibid.* 19 (1986) 3443. For the one-electron case, this approach was later popularized by Werner Kutzelnigg as Direct Perturbation Theory (DPT).

Let us take a closer look at the meaning of the expression

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) = \begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}\begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}$$
$$= \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} = \hat{p}^2\mathbf{1}.$$

Now please look carefully. Let us insert this into the last equation. We obtain what is sometimes called the Schrödinger equation with spin (because it is satisfied by a two-component spinor)

$$\left(\frac{\hat{p}^2}{2m_0} + V\right)\boldsymbol{\psi} = \varepsilon\boldsymbol{\psi}.$$

Recalling that $\hat{\boldsymbol{p}}$ represents the momentum operator, we observe each of the large components satisfies the familiar Schrödinger equation

$$\left(-\frac{\hbar^2}{2m_0}\Delta + V\right)\psi = \varepsilon\psi.$$

Therefore, the non-relativistic equation has been obtained from the relativistic one, assuming that the velocity of particle v is negligibly small with respect to the speed of light c. The Dirac equation remains valid even for larger particle velocities.

3.3.8 HOW DOES THE SPIN APPEAR?

It will be shown that the Dirac equation for the free electron in an external electromagnetic field is leading to the spin concept. Thus, in relativistic theory, the spin angular momentum appears in a natural way, whereas in the non-relativistic formalism it was the subject of a postulate of quantum mechanics, p. 25.

First let us introduce the following identity:

$$(\boldsymbol{\sigma}\cdot\boldsymbol{a})(\boldsymbol{\sigma}\cdot\boldsymbol{b}) = (\boldsymbol{a}\cdot\boldsymbol{b})\mathbf{1} + i\boldsymbol{\sigma}\cdot(\boldsymbol{a}\times\boldsymbol{b}),$$

where, on the left-hand side, we have a product of two matrices, each formed by a "scalar product" of matrices[28] $\boldsymbol{\sigma}$ and a vector, whereas on the right-hand side we have the scalar product of two vectors multiplied by a unit matrix plus the scalar

[28] That is, $\boldsymbol{\sigma}\cdot\boldsymbol{a} = \sigma_x a_x + \sigma_y a_y + \sigma_z a_z$.

product of the matrix $i\boldsymbol{\sigma}$ and the vector $\boldsymbol{a}\times\boldsymbol{b}$. The left-hand side:

$$\begin{aligned}
&\left[\begin{pmatrix}0 & a_x\\ a_x & 0\end{pmatrix}+\begin{pmatrix}0 & -ia_y\\ ia_y & 0\end{pmatrix}+\begin{pmatrix}a_z & 0\\ 0 & -a_z\end{pmatrix}\right]\\
&\quad\times\left[\begin{pmatrix}0 & b_x\\ b_x & 0\end{pmatrix}+\begin{pmatrix}0 & -ib_y\\ ib_y & 0\end{pmatrix}+\begin{pmatrix}b_z & 0\\ 0 & -b_z\end{pmatrix}\right]\\
&=\begin{pmatrix}a_z & a_x-ia_y\\ a_x+ia_y & -a_z\end{pmatrix}\begin{pmatrix}b_z & b_x-ib_y\\ b_x+ib_y & -b_z\end{pmatrix}\\
&=\begin{pmatrix}\boldsymbol{a}\cdot\boldsymbol{b}+i(\boldsymbol{a}\times\boldsymbol{b})_z & (\boldsymbol{a}\times\boldsymbol{b})_y+i(\boldsymbol{a}\times\boldsymbol{b})_x\\ -(\boldsymbol{a}\times\boldsymbol{b})_y+i(\boldsymbol{a}\times\boldsymbol{b})_x & \boldsymbol{a}\cdot\boldsymbol{b}-i(\boldsymbol{a}\times\boldsymbol{b})_z\end{pmatrix}
\end{aligned}$$

is therefore equal to the right-hand side, which is what we wanted to show.

Now, taking $\boldsymbol{a}=\boldsymbol{b}=\boldsymbol{\pi}$ one obtains the relation

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})=(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\mathbf{1}+i\boldsymbol{\sigma}(\boldsymbol{\pi}\times\boldsymbol{\pi}).$$

If the vector $\boldsymbol{\pi}$ had numbers as its components, the last term would have had to be zero, because the vector product of two parallel vectors would be zero. This, however, need not be true when the vector components are *operators* (as it is in our case). Since $\boldsymbol{\pi}=\boldsymbol{p}-\frac{q}{c}\boldsymbol{A}$, then $(\boldsymbol{\pi}\cdot\boldsymbol{\pi})=\boldsymbol{\pi}^2$ and $(\boldsymbol{\pi}\times\boldsymbol{\pi})=iq\frac{\hbar}{c}\operatorname{curl}\boldsymbol{A}$. To check this, we will obtain the last equality for the x components of both sides (the proof for the other two components looks the same). Let the operator $(\boldsymbol{\pi}\times\boldsymbol{\pi})$ act on an arbitrary function $f(x,y,z)$. As a result we expect the product of f and the vector $iq\frac{\hbar}{c}\operatorname{curl}\boldsymbol{A}$. Let us see:

$$\begin{aligned}
(\boldsymbol{\pi}\times\boldsymbol{\pi})_x f&=(\hat{p}_y-q/cA_y)(\hat{p}_z-q/cA_z)f-(\hat{p}_z-q/cA_z)(\hat{p}_y-q/cA_y)f\\
&=[\hat{p}_y\hat{p}_z-q/c\hat{p}_yA_z-q/cA_y\hat{p}_z+(q/c)^2A_yA_z\\
&\quad-\hat{p}_z\hat{p}_y+q/c\hat{p}_zA_y+q/cA_z\hat{p}_y-(q/c)^2A_zA_y]f\\
&=-q/c(-i\hbar)\left\{\frac{\partial}{\partial y}(A_zf)-A_z\frac{\partial f}{\partial y}+A_y\frac{\partial f}{\partial z}-\frac{\partial}{\partial z}(A_yf)\right\}\\
&=i\hbar q/c\left\{\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right\}f=\frac{iq\hbar}{c}(\operatorname{curl}\boldsymbol{A})_x f.
\end{aligned}$$

This is what we expected to get. From the Maxwell equations (p. 962), we have $\operatorname{curl}\boldsymbol{A}=\boldsymbol{H}$, where $\boldsymbol{H}$ represents the magnetic field intensity. Let us insert this into the Dirac equation (valid for kinetic energy much smaller than $2m_0c^2$, see eq. (3.60))

$$(\varepsilon-V)\boldsymbol{\psi}=\frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}$$

$$= \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\boldsymbol{\psi} + \frac{i}{2m_0}\boldsymbol{\sigma}\cdot(\boldsymbol{\pi}\times\boldsymbol{\pi})\boldsymbol{\psi}$$
$$= \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\boldsymbol{\psi} + \frac{i}{2m_0}\frac{iq\hbar}{c}(\boldsymbol{\sigma}\cdot\boldsymbol{H})\boldsymbol{\psi}$$
$$= \left[\frac{\pi^2}{2m_0} - \frac{q\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\boldsymbol{H}\right]\boldsymbol{\psi} = \left[\frac{\pi^2}{2m_0} + \frac{e\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\boldsymbol{H}\right]\boldsymbol{\psi}.$$

In the last parenthesis, beside the kinetic energy operator (first term), there is a strange second term. The term has the appearance of the interaction energy $-\boldsymbol{M}\cdot\boldsymbol{H}$ of a mysterious magnetic dipole moment $\boldsymbol{M}$ with magnetic field $\boldsymbol{H}$ (cf. interaction with magnetic field, p. 659). The operator of this electronic dipole moment $\boldsymbol{M} = -\frac{e\hbar}{2m_0c}\boldsymbol{\sigma} = -\mu_B\boldsymbol{\sigma}$, where μ_B stands for the Bohr magneton equal to $\frac{e\hbar}{2m_0c}$. The spin angular momentum operator of the electron is denoted by (cf. p. 28) $\boldsymbol{s}$. Therefore, one has $\boldsymbol{s} = \frac{1}{2}\hbar\boldsymbol{\sigma}$. Inserting $\boldsymbol{\sigma}$ to the equation for $\boldsymbol{M}$ we obtain

$$\boldsymbol{M} = -2\frac{\mu_B}{\hbar}\boldsymbol{s} = -\frac{e}{m_0c}\boldsymbol{s}. \tag{3.62}$$

It is exactly twice as much as we get for the *orbital* angular momentum and the corresponding *orbital* magnetic dipole (hence the anomalous magnetic spin moment of the electron), see eq. (12.53).

When two values differ by an integer factor (as in our case) this should stimulate our mind, because it may mean something fundamental that might depend on, e.g., the number of dimensions of our space or something similar. However, one of the most precise experiments ever made by humans gave[29] $2.0023193043737 \pm 0.0000000000082$ instead of 2. Therefore, our excitement must diminish. A more accurate theory (quantum electrodynamics, some of the effects of this will be described later) gave a result that agreed with the experiment within an experimental error of ± 0.0000000008. The extreme accuracy achieved witnessed the exceptional status of quantum electrodynamics, because no other theory of mankind has achieved such a level of accuracy.

3.3.9 SIMPLE QUESTIONS

How to interpret a bispinor wave function? Does the Dirac equation describe a single fermion, an electron, a positron, an electron and a Dirac sea of other electrons (infinite number of particles), an effective electron or effective positron (interacting with the Dirac sea)? After eighty years these questions do not have a clear answer.

[29] R.S. Van Dyck Jr., P.B. Schwinberg, H.G. Dehmelt, *Phys. Rev. Letters* 59 (1990) 26.

Despite the glorious invariance with respect to the Lorentz transformation and despite spectacular successes, the Dirac equation has some serious drawbacks, including a lack of clear physical interpretation. These drawbacks are removed by a more advanced theory – quantum electrodynamics.

3.4 THE HYDROGEN-LIKE ATOM IN DIRAC THEORY

After this short escapade we are back with Dirac theory. The hydrogen-like atom may be simplified by immobilizing the nucleus and considering a single particle – the electron[30] moving in the *electrostatic field of the nucleus*[31] $-Z/r$. This problem has an exact solution first obtained by Charles Galton Darwin, cf. p. 112. The electron state is described by four quantum numbers n, l, m, m_s, where $n = 1, 2, \ldots$ stands for the principal, $0 \leqslant l \leqslant n-1$ for the angular, $|m| \leqslant l$ for the magnetic and $m_s = \frac{1}{2}, -\frac{1}{2}$ for the spin quantum number. Darwin obtained the following formula for the relativistic energy of the hydrogen-like atom (in a.u.):

$$E_{n,j} = -\frac{1}{2n^2}\left[1 + \frac{1}{nc^2}\left(\frac{1}{j+\frac{1}{2}} - \frac{3}{4n}\right)\right],$$

where $j = l + m_s$, and c is the speed of light (in a.u.). For the ground state (1s, $n = 1$, $l = 0$, $m = 0$, $m_s = \frac{1}{2}$) we have

$$E_{1,\frac{1}{2}} = -\frac{1}{2}\left[1 + \left(\frac{1}{2c}\right)^2\right].$$

Thus, instead of the non-relativistic energy equal to $-\frac{1}{2}$, from the Dirac theory we obtain -0.5000067 a.u., which means a very small correction to the non-relativistic energy. The electron energy levels for the non-relativistic and relativistic cases are shown schematically in Fig. 3.2.

3.4.1 STEP BY STEP: CALCULATION OF THE GROUND STATE OF THE HYDROGEN-LIKE ATOM WITHIN DIRAC THEORY

Matrix form of the Dirac equation

We will use the Dirac equation (3.59). First, the basis set composed of two bispinors will be created: $\boldsymbol{\Psi}_1 = \begin{pmatrix} \boldsymbol{\psi} \\ \mathbf{0} \end{pmatrix}$ and $\boldsymbol{\Psi}_2 = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\phi} \end{pmatrix}$, and the wave function $\boldsymbol{\Psi}$ will be

[30] In the Dirac equation $\boldsymbol{A} = \mathbf{0}$ and $-e\phi = V = -\frac{Ze^2}{r}$ were set.

[31] The centre-of-mass motion can be easily separated from the Schrödinger equation, Appendix I. Nothing like this has been done for the Dirac equation. The atomic mass depends on its velocity with respect to the laboratory coordinate system, the electron and proton mass also depend on their speeds, and there is also a mass deficit as a result of binding between both particles. All this seems to indicate that centre of mass separation is not possible. Nevertheless, for an energy expression accurate to a certain power of c^{-1}, such a separation is, at least in some cases, possible.

sought as a linear combination $\mathbf{\Psi} = c_1\mathbf{\Psi}_1 + c_2\mathbf{\Psi}_2$, which represents an approximation. Within this approximation the Dirac equation looks like this

$$\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}(c_1\mathbf{\Psi}_1 + c_2\mathbf{\Psi}_2) = \mathbf{0},$$

which gives

$$c_1\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_1 + c_2\begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_2 = \mathbf{0}.$$

By making a scalar product first with $\mathbf{\Psi}_1$ and then with $\mathbf{\Psi}_2$ we obtain two equations:

$$\begin{aligned} &c_1\left\langle \mathbf{\Psi}_1 \middle| \begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_1 \right\rangle \\ &\quad + c_2\left\langle \mathbf{\Psi}_1 \middle| \begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_2 \right\rangle = 0 \\ &c_1\left\langle \mathbf{\Psi}_2 \middle| \begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_1 \right\rangle \\ &\quad + c_2\left\langle \mathbf{\Psi}_2 \middle| \begin{pmatrix} V-\varepsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\varepsilon \end{pmatrix}\mathbf{\Psi}_2 \right\rangle = 0. \end{aligned}$$

Taking into account the particular structure of the bispinors $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$, we obtain the same equations expressed in spinors (two component spinors)

$$\begin{aligned} &c_1\langle\boldsymbol{\psi}\,|(V-\varepsilon)\boldsymbol{\psi}\rangle + c_2\langle\boldsymbol{\psi}|\,c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle = 0, \\ &c_1\langle\boldsymbol{\phi}\,|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle + c_2\langle\boldsymbol{\phi}|\,(V-2m_0c^2-\varepsilon)\boldsymbol{\phi}\rangle = 0. \end{aligned}$$

This is a set of homogeneous linear equations. To obtain a non-trivial solution,[32] the determinant of the coefficients multiplying the unknowns c_1 and c_2 has to be zero (the secular determinant, cf. variational method in Chapter 5)

$$\begin{vmatrix} \langle\boldsymbol{\psi}|(V-\varepsilon)\boldsymbol{\psi}\rangle & \langle\boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle \\ \langle\boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle & \langle\boldsymbol{\phi}|(V-2m_0c^2-\varepsilon)\boldsymbol{\phi}\rangle \end{vmatrix} = 0.$$

The potential V in the above formula will be taken as $-Z/r$, where r is the electron–nucleus distance.

[32] It is easy to give a trivial one, but not acceptable (the wave function cannot equal zero everywhere): $c_1 = c_2 = 0$.

The large component spinor

It is true that we have used an extremely poor basis, however, we will try to compensate for it by allowing a certain flexibility within the large component spinor: $\boldsymbol{\psi} = \begin{pmatrix} 1s \\ 0 \end{pmatrix}$, where the hydrogen-like function $1s = \sqrt{\frac{\zeta^3}{\pi}} \exp(-\zeta r)$. The parameter ζ will be optimized in such a way as to minimize the energy ε of the electron. This idea is similar to the variational method in the non-relativistic theory (Chapter 5 and Appendix H, p. 969), however, it is hardly justified in the relativistic case. Indeed, as proved by numerical experience the variational procedure very often fails. As a remedy we will use kinetic balancing already used to describe the large and small components of the bispinor (p. 119). The spinor of the small components is therefore obtained automatically from the large components (approximation):

$$\boldsymbol{\phi} = \mathcal{N}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} = \mathcal{N} \begin{pmatrix} \hat{p}_z & \hat{p}_x + i\hat{p}_y \\ \hat{p}_x - i\hat{p}_y & \hat{p}_z \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix}$$

$$= \mathcal{N} \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix},$$

where $\mathcal{N}$ is a normalization constant. In the above formula $\hat{\boldsymbol{p}}$ represents the momentum operator. The normalization constant $\mathcal{N}$ will be found from

$$\langle \boldsymbol{\phi} | \boldsymbol{\phi} \rangle = 1 = |\mathcal{N}|^2 \left\{ \left\langle \hat{p}_z(1s) \,\middle|\, \hat{p}_z(1s) \right\rangle + \left\langle (\hat{p}_x + i\hat{p}_y)(1s) \,\middle|\, (\hat{p}_x + i\hat{p}_y)(1s) \right\rangle \right\}$$

$$= |\mathcal{N}|^2 \cdot \left\{ \begin{array}{c} \langle \hat{p}_z(1s) | \hat{p}_z(1s) \rangle + \langle \hat{p}_x(1s) | \hat{p}_x(1s) \rangle + i \langle \hat{p}_x(1s) | \hat{p}_y(1s) \rangle \\ -i \langle \hat{p}_y(1s) | \hat{p}_x(1s) \rangle + \langle \hat{p}_y(1s) | \hat{p}_y(1s) \rangle \end{array} \right\}.$$

In the above formula, integrals with the imaginary unit i are equal to zero, because the integrand is an odd function. After using the Hermitian character of the momentum operator we obtain $1 = |\mathcal{N}|^2 \langle 1s | \hat{p}^2 1s \rangle = \zeta^2$. The last equality follows from Appendix H, p. 969. Thus, one may choose $\mathcal{N} = 1/\zeta$.

Calculating integrals in the Dirac matrix equation

We will calculate one by one all the integrals that appear in the Dirac matrix equation. The integral $\langle \boldsymbol{\psi} | -\frac{Z}{r} \boldsymbol{\psi} \rangle = -Z\zeta$, because the scalar product leads to the nuclear attraction integral with a hydrogen-like atomic orbital, and this gives the result above (Appendix H, p. 969). The next integral can be computed as follows

$$\left\langle \boldsymbol{\phi} \left| \frac{1}{r} \boldsymbol{\phi} \right. \right\rangle = |\mathcal{N}|^2 \left\langle \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \left| \frac{1}{r} \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \right. \right\rangle$$

$$= |\mathcal{N}|^2 \left\langle \hat{p}_z(1s) \left| \frac{1}{r} \hat{p}_z(1s) \right. \right\rangle + \left\langle (\hat{p}_x + i\hat{p}_y)(1s) \left| \frac{1}{r} (\hat{p}_x + i\hat{p}y)(1s) \right. \right\rangle$$

$$= |\mathcal{N}|^2 \left\langle (1s) \left| \hat{p}_z \frac{1}{r} \hat{p}_z (1s) \right\rangle + \left\langle (1s) \left| (\hat{p}_x - i\hat{p}_y) \frac{1}{r} (\hat{p}_x + i\hat{p}_y)(1s) \right\rangle \right.$$

$$= |\mathcal{N}|^2 \left\langle (1s) \left| \left(\hat{p}_z \frac{1}{r} \right) \hat{p}_z (1s) \right\rangle + \left\langle (1s) \left| \left[(\hat{p}_x - i\hat{p}_y) \frac{1}{r} \right] (\hat{p}_x + i\hat{p}_y)(1s) \right\rangle \right.$$

$$+ \left\langle (1s) \left| \frac{1}{r} \hat{p}_z \hat{p}_z (1s) \right\rangle + \left\langle (1s) \left| \frac{1}{r} (\hat{p}_x - i\hat{p}_y) (\hat{p}_x + i\hat{p}_y) (1s) \right\rangle \right. . \tag{3.63}$$

In the second row, the scalar product of spinors is used, in the third row, the Hermitian character of the operator $\hat{p}$. Further,

$$\left\langle \boldsymbol{\phi} \left| \frac{1}{r} \boldsymbol{\phi} \right\rangle = |\mathcal{N}|^2 \left[\left\langle (1s) \left| \left(\hat{p}_z \frac{1}{r} \right) \hat{p}_z (1s) \right\rangle + \left\langle (1s) \left| \frac{1}{r} \left(\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2 \right) (1s) \right\rangle \right. \right.$$

$$\left. \left. + \left\langle (1s) \left| \left[(\hat{p}_x - i\hat{p}_y) \frac{1}{r} \right] (\hat{p}_x + i\hat{p}_y) (1s) \right\rangle \right. \right]$$

$$= |\mathcal{N}|^2 \left[\left\langle (1s) \left| \left(\hat{p}_z \frac{1}{r} \right) \hat{p}_z (1s) \right\rangle - \left\langle (1s) \left| \frac{1}{r} \Delta (1s) \right\rangle + \left\langle (1s) \left| \left(\hat{p}_x \frac{1}{r} \right) \hat{p}_x (1s) \right\rangle \right. \right.$$

$$+ \left\langle (1s) \left| \left(\hat{p}_y \frac{1}{r} \right) \hat{p}_y (1s) \right\rangle - i \left\langle (1s) \left| \left(\hat{p}_y \frac{1}{r} \right) \hat{p}_x (1s) \right\rangle \right. \right.$$

$$\left. \left. + i \left\langle (1s) \left| \left(\hat{p}_x \frac{1}{r} \right) \hat{p}_y (1s) \right\rangle \right. \right] . \tag{3.64}$$

We used the atomic units and therefore $\hat{p}^2 = -\Delta$, and the momentum operator is equal to $-i\nabla$. The two integrals at the end cancel each other, because each of the integrals does not change when the variables are interchanged: $x \leftrightarrow y$.

Finally, we obtain the following formula

$$\left\langle \boldsymbol{\phi} \left| \frac{1}{r} \boldsymbol{\phi} \right\rangle = -|\mathcal{N}|^2 \left\{ \left\langle 1s \left| \frac{1}{r} \Delta (1s) \right\rangle + \left\langle 1s \left| \left(\nabla \frac{1}{r} \right) \nabla (1s) \right\rangle \right. \right\} \right.$$

$$= -\zeta^{-2} \left\{ \left(-3\zeta^3 + 2\zeta^3 \right) \right\} = \zeta ,$$

where the equality follows from a direct calculation of the two integrals.[33]

The next matrix element to calculate is equal to $\langle \boldsymbol{\phi} | c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \boldsymbol{\psi} \rangle$. We proceed as follows (please recall kinetic balancing and we also use Appendix H, p. 969):

$$\langle \boldsymbol{\phi} | c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \boldsymbol{\psi} \rangle = \mathcal{N} c \left\langle (\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right| (\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle$$

[33] In the first integral we have the same situation as a while before. In the second integral we write the nabla operator in Cartesian coordinates, obtain a scalar product of two gradients, then we get three integrals equal to one another (they contain x, y, z), and it is sufficient to calculate one of them by spherical coordinates by formula (H.2) in Appendix H, p. 969.

$$
\begin{aligned}
&= \mathcal{N}c\left\langle \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \middle| \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix} \right\rangle \\
&= \mathcal{N}c\left[\langle \hat{p}_z(1s) \mid \hat{p}_z(1s)\rangle + \langle (\hat{p}_x + i\hat{p}_y)(1s) \mid (\hat{p}_x + i\hat{p}_y)(1s)\rangle\right] \\
&= \mathcal{N}c\langle 1s|\hat{p}^2(1s)\rangle = \frac{1}{\zeta}c\zeta^2 = c\zeta.
\end{aligned}
$$

The last matrix element reads as

$$
\begin{aligned}
\langle \boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle &= \mathcal{N}c\left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle \\
&= \mathcal{N}c\left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle = \mathcal{N}c\langle 1s \mid \hat{p}^2 1s\rangle = c\frac{1}{\zeta}\zeta^2 = c\zeta.
\end{aligned}
$$

Dirac's secular determinant

We have all the integrals needed and may now write the secular determinant corresponding to the matrix form of the Dirac equation:

$$
\begin{vmatrix} \langle \boldsymbol{\psi}|V\boldsymbol{\psi}\rangle - \varepsilon & \langle \boldsymbol{\psi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle \\ \langle \boldsymbol{\phi}|c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi}\rangle & \langle \boldsymbol{\phi}|(V - 2c^2))\boldsymbol{\phi}\rangle - \varepsilon \end{vmatrix} = 0,
$$

and after inserting the calculated integrals

$$
\begin{vmatrix} -Z\zeta - \varepsilon & c\zeta \\ c\zeta & -Z\zeta - 2c^2 - \varepsilon \end{vmatrix} = 0.
$$

Expanding the determinant gives the equation for the energy ε

$$
\varepsilon^2 + \varepsilon\left(2Z\zeta + 2c^2\right) + \left[Z\zeta\left(Z\zeta + 2c^2\right) - c^2\zeta^2\right] = 0.
$$

Hence, we get two solutions

$$
\varepsilon_{\pm} = -\left(c^2 + Z\zeta\right) \pm \sqrt{c^4 + \zeta^2 c^2}.
$$

Note that the square root is of the order of c^2 (in a.u.), and with the (unit) mass of the electron m_0, it is of the order of m_0c^2. Therefore, the minus sign before the square root corresponds to a solution with energy of the order of $-2m_0c^2$, while the plus sign corresponds to energy of the order of zero. Let us recall that we have shifted the energy scale in the Dirac equation and the last solution ε_+ (hereafter denoted by ε) is to be compared to the energy of the non-relativistic hydrogen-like atom

$$
\varepsilon = -\left(c^2 + Z\zeta\right) + \sqrt{c^4 + \zeta^2 c^2} = -\left(c^2 + Z\zeta\right) + c^2\sqrt{1 + \frac{\zeta^2}{c^2}}
$$

$$= -(c^2 + Z\zeta) + c^2\left(1 + \frac{\zeta^2}{2c^2} - \frac{\zeta^4}{8c^4} + \ldots\right)$$

$$= -Z\zeta + \frac{\zeta^2}{2} + \left(-\frac{\zeta^4}{8c^2} + \ldots\right). \tag{3.65}$$

Non-relativistic solution

If $c \to \infty$, i.e. we approach the non-relativistic limit, then $\varepsilon = -Z\zeta + \frac{\zeta^2}{2}$. Minimization of this energy with respect to ζ gives its optimum value $\zeta_{\text{opt}}^{\text{nonrel}} = Z$. In this way one recovers the result known from non-relativistic quantum mechanics (Appendix H) obtained in the variational approach to the hydrogen atom with the $1s$ orbital as a trial function.

3.4.2 RELATIVISTIC CONTRACTION OF ORBITALS

Minimizing the relativistic energy equation (3.65) leads to an equation for optimum $\zeta \equiv \zeta_{\text{opt}}^{\text{rel}}$:

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}\zeta} = 0 = -Z + \frac{1}{2}\left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} 2\zeta c^2 = -Z + \left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} \zeta c^2,$$

giving

$$\zeta_{\text{opt}}^{\text{rel}} = \frac{Z}{\sqrt{1 - \frac{Z^2}{c^2}}}.$$

The result differs remarkably from the non-relativistic value $\zeta_{\text{opt}}^{\text{nonrel}} = Z$, but approaches the non-relativistic value when $c \to \infty$. Note than the difference between the two values increases with atomic number Z, and that the relativistic exponent is always *larger* that its non-relativistic counter-part. This means that the relativistic orbital *decays faster* with the electron–nucleus distance and therefore

> the relativistic orbital $1s$ is *smaller* (contraction) than the corresponding non-relativistic one.

Let us see how it is for the hydrogen atom. In that case $\zeta_{\text{opt}}^{\text{rel}} = 1.0000266$ as compared to $\zeta_{\text{opt}}^{\text{nonrel}} = Z_H = 1$. And what about $1s$ orbital of gold? For gold $\zeta_{\text{opt}}^{\text{rel}} = 96.68$, while $\zeta_{\text{opt}}^{\text{nonrel}} = Z_{Au} = 79$! Since for a heavy atom, the effective exponent of the atomic orbitals decreases when moving from the low-energy compact $1s$ orbital to higher-energy outer orbitals, this means that the most important relativistic orbital contraction occurs for the inner shells. The chemical properties of an atom depend on what happens to its outer shells (valence shell). Therefore, we

may conclude that the relativistic corrections are expected to play a secondary role in chemistry.[34]

If we insert $\zeta_{\text{opt}}^{\text{rel}}$ in eq. (3.65) we obtain the minimum value of ε

$$\varepsilon_{\min} = -\left(c^2 + Z\zeta\right) + \sqrt{c^4 + \zeta^2}. \tag{3.66}$$

Since Z^2/c^2 is small with respect to 1, we may expand the square root in the Taylor series, $\sqrt{1-x} = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \cdots$. We obtain

$$\begin{aligned}\varepsilon_{\min} &= -c^2 + c^2\left\{1 - \left(\frac{1}{2}\right)\left(\frac{Z^2}{c^2}\right) - \frac{1}{8}\left(\frac{Z^2}{c^2}\right)^2 - \cdots\right\} \\ &= -\frac{Z^2}{2}\left(1 + \left(\frac{Z}{2c}\right)^2 + \cdots\right).\end{aligned} \tag{3.67}$$

In the case of the hydrogen atom ($Z = 1$) we have

$$\varepsilon_{\min} = -\frac{1}{2}\left(1 + \left(\frac{1}{2c}\right)^2 + \cdots\right), \tag{3.68}$$

where the first two terms shown give Darwin's exact result[35] (discussed earlier). Inserting $c = 137.036$ a.u. we obtain the hydrogen atom ground-state energy $\varepsilon = -0.5000067$ a.u., which agrees with Darwin's result.

3.5 LARGER SYSTEMS

The Dirac equation represents an approximation[36] and refers to a single particle. What happens with larger systems? Nobody knows, but the first idea is to construct the total Hamiltonian as a sum of the Dirac Hamiltonians for individual particles plus their Coulombic interaction (the *Dirac–Coulomb approximation*). This is practised routinely nowadays for atoms and molecules. Most often we use the mean-field approximation (see Chapter 8) with the modification that each of the one-electron functions represents a four-component bispinor. Another approach is extremely pragmatic, maybe too pragmatic: we perform the non-relativistic calculations with a pseudopotential that mimics what is supposed to happen in a relativistic case.

Dirac–Coulomb approximation

[34] We have to remember, however, that the relativistic effects also propagate from the inner shells to the valence shell through the orthogonalization condition, that has to be fulfilled after the relativistic contraction. This is why the gold valence orbital 6*s* shrinks, which has an immediate consequence in the relativistic shortening of the bond length in Au_2, which we cited at the beginning of this chapter.

[35] I.e. the exact solution to the Dirac equation for the electron in the external electric field produced by the proton.

[36] Yet it is strictly invariant with respect to the Lorentz transformation.

3.6 BEYOND THE DIRAC EQUATION. . .

How reliable is the presented relativistic quantum theory? The Dirac or Klein–Gordon equations, as is usual in physics, describe only some aspects of reality. The fact that both equations are invariant with respect to the Lorentz transformation indicates *only* that the space-time *symmetry* properties are described correctly. The physical machinery represented by these equations is not so bad, since several predictions have been successfully made (antimatter, electron spin, energy levels of the hydrogen atom). Yet, in the latter case an assumption of the external field $V = -\frac{Ze^2}{r}$ is a positively desperate step, which in fact is unacceptable in a fair relativistic theory for the proton and the electron (and not only of the electron in the external field of the nucleus). Indeed, the proton and the electron move. At a given time their distance is equal to r, but such a distance might be inserted into the Coulombic law if the speed of light were infinite, because the two particles would feel their positions instantaneously. Since, however, any perturbation by a positional change of a particle needs time to travel to the other particle, we have to use another distance somehow taking this into account (Fig. 3.3). The same pertains, of course, to any pair of particles in a many-body system (the so-called *retarded potential*).

retarded potential

There is certainly a need for a more accurate theory.

3.6.1 THE BREIT EQUATION

Breit constructed a many-electron relativistic theory that takes into account such a retarded potential in an approximate way. Breit explicitly considered only the electrons of an atom, nucleus of which (similar to Dirac theory) created only an external field for the electrons. This ambitious project was only partly success-

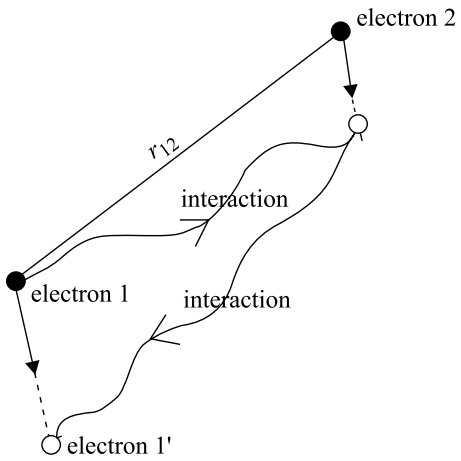


Fig. 3.3. Retardation of the interaction. The distance r_{12} of two particles in the interaction potential (as in Coulomb's law) is bound to represent an approximation, because we assume an instantaneous interaction. However, when the two particles catch sight of each other (which takes time) they are already somewhere else.

ful, because the resulting theory turned out to be approximate not only from the point of view of quantum theory (some interactions not taken into account) but also from the point of view of relativity theory (an *approximate* Lorentz transformation invariance).

For two electrons the Breit equation has the form (r_{12} stands for the distance between electron 1 and electron 2)

Gregory Breit (1899–1981), American physicist, professor at the universities New York, Wisconsin, Yale, Buffalo. Breit with Eugene Wigner introduced the resonance states of particles, and with Condon created the proton–proton scattering theory.

$$\left\{\hat{H}(1)+\hat{H}(2)+\frac{1}{r_{12}}-\frac{1}{2r_{12}}\left[\boldsymbol{\alpha}(1)\boldsymbol{\alpha}(2)+\frac{[\boldsymbol{\alpha}(1)\cdot\boldsymbol{r}_{12}][\boldsymbol{\alpha}(2)\cdot\boldsymbol{r}_{12}]}{r_{12}^2}\right]\right\}\Psi=E\Psi, \tag{3.69}$$

where (cf. eq. (3.54) with E replaced by the Hamiltonian)

$$\hat{H}(i)=q_i\phi(\boldsymbol{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2=-e\phi(\boldsymbol{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2$$

is the Dirac Hamiltonian for electron i pointed by vector $\boldsymbol{r}_i$, whereas the Dirac matrices for electron i: $\boldsymbol{\alpha}(i)=[\alpha_x(i),\alpha_y(i),\alpha_z(i)]$ and the corresponding operators $\pi_\mu(i)$ have been defined on p. 114, $\phi(\boldsymbol{r}_i)$ represents the scalar potential calculated at $\boldsymbol{r}_i$. The wavefunction Ψ represents a 16-component spinor (here represented by a square matrix of rank 4), because for each electron we would have the usual Dirac bispinor (four component) and the two-electron wavefunction depends on the Cartesian product of the components.[37]

The Breit Hamiltonian (in our example, for two electrons in an electromagnetic field) can be approximated by the following useful formula[38] known as the *Breit–Pauli Hamiltonian*

Breit–Pauli Hamiltonian

$$\hat{H}(1,2)=\hat{H}_0+\hat{H}_1+\cdots+\hat{H}_6, \tag{3.70}$$

where:

- $\hat{H}_0=\frac{\hat{p}_1^2}{2m_0}+\frac{\hat{p}_2^2}{2m_0}+V$ represents the familiar non-relativistic Hamiltonian.
- $\hat{H}_1=-\frac{1}{8m_0^3c^2}(\hat{p}_1^4+\hat{p}_2^4)$ comes from the velocity dependence of mass, more precisely from the Taylor expansion of eq. (3.38), p. 109, for small velocities.
- $\hat{H}_2=-\frac{e^2}{2(m_0c)^2}\frac{1}{r_{12}}[\hat{\boldsymbol{p}}_1\cdot\hat{\boldsymbol{p}}_2+\frac{\boldsymbol{r}_{12}\cdot(\boldsymbol{r}_{12}\cdot\hat{\boldsymbol{p}}_1)\hat{\boldsymbol{p}}_2}{r_{12}^2}]$ stands for the correction[39] that accounts in part for the above mentioned retardation. Alternatively, the term may be viewed as the interaction energy of two magnetic dipoles, each resulting from the orbital motion of an electron (*orbit–orbit term*).

orbit–orbit term

[37] In the Breit equation (3.69) the operators in {} act either by multiplying the 4×4 matrix Ψ by a function (i.e. each element of the matrix) or by a 4×4 matrix resulting from α matrices.

[38] H.A. Bethe, E.E. Salpeter, "*Quantum Mechanics of One- and Two-Electron Atoms*", Springer, 1977, p. 181.

[39] For non-commuting operators $\hat{\boldsymbol{a}}(\hat{\boldsymbol{a}}\cdot\hat{\boldsymbol{b}})\hat{\boldsymbol{c}}=\sum_{i,j=1}^{3}\hat{a}_i\hat{a}_j\hat{b}_j\hat{c}_i$.

- $\hat{H}_3 = \frac{\mu_B}{m_0 c}\{[\boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) \times \hat{\boldsymbol{p}}_1 + \frac{2e}{r_{12}^3}\boldsymbol{r}_{12} \times \hat{\boldsymbol{p}}_2] \cdot \boldsymbol{s}_1 + [\boldsymbol{\mathcal{E}}(\boldsymbol{r}_2) \times \hat{\boldsymbol{p}}_2 + \frac{2e}{r_{12}^3}\boldsymbol{r}_{21} \times \hat{\boldsymbol{p}}_1] \cdot \boldsymbol{s}_2\}$ is the interaction energy of the electronic magnetic moments (resulting from the above mentioned orbital motion) with the spin magnetic dipole moments (*spin–orbit coupling*), μ_B stands for the Bohr magneton, and $\boldsymbol{\mathcal{E}}$ denotes the electric field vector. Since we have two orbital magnetic dipole moments and two spin orbital dipole moments, there are four spin–orbit interactions. The first term in square brackets stands for the spin–orbit coupling of the same electron, while the second term represents the coupling of the spin of one particle with the orbit of the second.

spin–orbit coupling

- $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}[\hat{\boldsymbol{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) + \hat{\boldsymbol{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_2)]$ is a non-classical term peculiar to the Dirac theory (also present in the one-electron Dirac Hamiltonian) called the *Darwin term*.

Darwin term

- $\hat{H}_5 = 4\mu_B^2\{-\frac{8\pi}{3}(\boldsymbol{s}_1 \cdot \boldsymbol{s}_2)\delta(\boldsymbol{r}_{12}) + \frac{1}{r_{12}^3}[\boldsymbol{s}_1 \cdot \boldsymbol{s}_2 - \frac{(\boldsymbol{s}_1 \cdot \boldsymbol{r}_{12})(\boldsymbol{s}_2 \cdot \boldsymbol{r}_{12})}{r_{12}^2}]\}$ corresponds to the spin dipole moment interactions of the two electrons (*spin–spin term*). The first term is known as the *Fermi contact term*, since it is non-zero only when the two electrons touch one another (see Appendix E, p. 951), whereas the second term represents the classical *dipole–dipole interaction* of the two electronic spins (cf. the multipole expansion in Appendix X, p. 1038 and Chapter 13), i.e. the interaction of the two spin magnetic moments of the electrons (with the factor 2, according to eq. (3.62), p. 122).

spin–spin
Fermi contact term

- $\hat{H}_6 = 2\mu_B[\boldsymbol{H}(\boldsymbol{r}_1) \cdot \boldsymbol{s}_1 + \boldsymbol{H}(\boldsymbol{r}_2) \cdot \boldsymbol{s}_2] + \frac{e}{m_0 c}[\boldsymbol{A}(\boldsymbol{r}_1) \cdot \hat{\boldsymbol{p}}_1 + \boldsymbol{A}(\boldsymbol{r}_2) \cdot \hat{\boldsymbol{p}}_2]$ is known as the *Zeeman interaction*, i.e. the interaction of the spin (the first two terms) and the orbital (the second two terms) electronic magnetic dipole moments with the external magnetic field $\boldsymbol{H}$ (cf. eq. (3.62)).

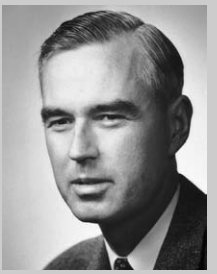
Zeeman term

The terms listed above are of prime importance in the theory of the interaction of matter with the electromagnetic field (e.g., in nuclear magnetic resonance).

3.6.2 A FEW WORDS ABOUT QUANTUM ELECTRODYNAMICS (QED)

The Dirac and Breit equations do not account for several subtle effects.[40] They are predicted by quantum electrodynamics, a many-particle theory.

Willis Eugene Lamb (b. 1913), American physicist, professor at Columbia, Stanford, Oxford, Yale and Tucson universities. He received the Nobel Prize in 1955 "*for his discoveries concerning the fine structure of the hydrogen spectrum*".



[40] For example, an effect observed in spectroscopy for the first time by Willis Lamb.

The QED energy may be conveniently developed in a series of $\frac{1}{c}$:

- in zero order we have the non-relativistic approximation (solution to the Schrödinger equation);
- there are no first order terms;
- the second order contains the Breit corrections;
- the third and further orders are called the *radiative corrections*.

radiative corrections

Radiative corrections

The radiative corrections include:

- *Interaction with the vacuum* (Fig. 3.4.a). According to modern physics the perfect vacuum does not just represent nothing. The electric field of the vacuum itself fluctuates about zero and these instantaneous fluctuations influence the motion of any charged particle. When a strong electric field operates in a vacuum, the latter undergoes a polarization (*vacuum polarization*), which means a *spontaneous creation of matter*, more specifically, of particle-antiparticle pairs.

vacuum polarization

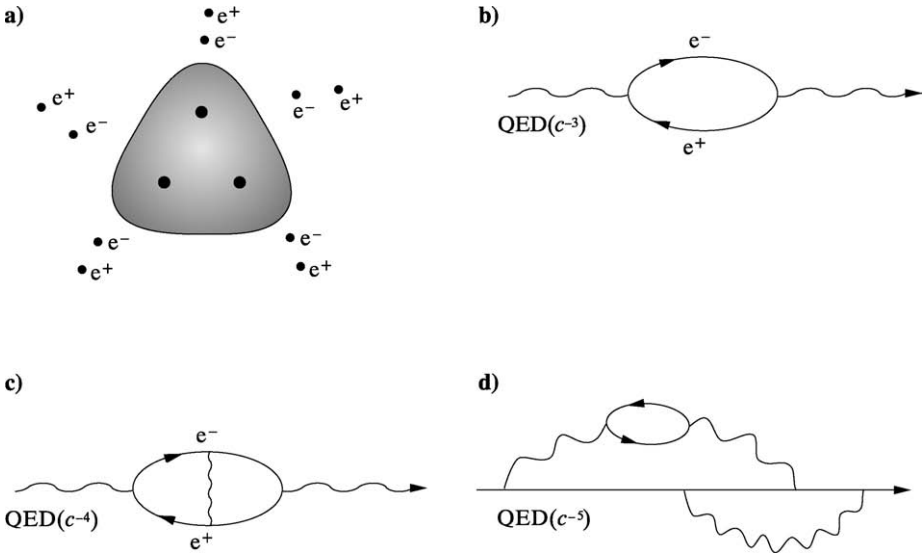


Fig. 3.4. (a) The electric field close to the proton (composed of three quarks) is so strong that it *creates matter and antimatter* (shown as electron–positron pairs). The three quarks visible in scattering experiments represent the valence quarks. (b) One of the radiative effects in the QED correction of the c^{-3} order (see Table 3.1). The pictures show the sequence of the events from left to the right. A photon (wavy line on the left) polarizes the vacuum and an electron–positron pair (solid lines) is created, and the photon vanishes. Then the created particles annihilate each other and a photon is created. (c) A similar event (of the c^{-4} order in QED), but during the existence of the electron–positron pair the two particles interact by exchange of a photon. (d) An electron (horizontal solid line) emits a photon, which creates an electron–positron pair, that annihilates producing another photon. Meanwhile the first electron emits a photon, then first absorbs the photon from the annihilation, and afterwards the photon emitted by itself earlier. This effect is of the order c^{-5} in QED.

The probability of this event (per unit volume and time) depends[41] (Fig. 3.4.a–d) on the particle mass m and charge q:

$$w = \frac{\mathcal{E}^2}{c\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \exp\left(-\frac{n\pi m^2}{|q\mathcal{E}|}\right), \tag{3.71}$$

where $\mathcal{E}$ is the electric field intensity. The creation of such pairs in a static electric field has never yet been observed, because we cannot yet provide sufficient $\mathcal{E}$. Even for the electron on the first Bohr orbit, the $|q\mathcal{E}|$ is small compared to m^2 (however, for smaller distances the exponent may be much smaller).

creation of matter

- *Interaction with virtual photons*. The electric field influences the motion of electron. What about its own electric field? Does it influence its motion as well? The latter effect is usually modelled by allowing the electron to emit photons and then to absorb them ("*virtual photons*")[42] (Fig. 3.4.d).

The QED calculations performed to date have been focused on the energy. The first calculations of atomic susceptibilities (helium) within an accuracy including the c^{-2} terms were carried out independently[43] by Pachucki and Sapirstein[44] and by Cencek and coworkers,[45] and with accuracy up to c^{-3} (with estimation of the c^{-4} term) by Łach and coworkers (see Table 3.1). To get a flavour of what subtle effects may be computed nowadays, Table 3.1 shows the components of the first ionization energy and of the dipole polarizability (see Chapter 12) of the helium atom.

Comments to Table 3.1

- $\hat{H}_0$ denotes the result obtained from an accurate solution of the Schrödinger equation (i.e. the non-relativistic and finite nuclear mass theory). Today the solution of the equation could be obtained with greater accuracy than reported here. Imagine, that here the theory is limited by the *precision of our knowledge of the helium atom mass*, which is "only" 12 significant figures.
- The effect of the non-zero size of the nucleus is small, it is practically never taken into account in computations. If we enlarged the nucleus to the size of an apple, the first Bohr orbit would be 10 km from the nucleus. And still (sticking to our analogy) the electron is able to distinguish a point from an apple? Not quite. It sees the (tiny) difference because the electron knows the region close to the nucleus: it is there that it resides most often. Anyway the theory is able to compute such a tiny effect.

[41] C. Itzykson, J.-B. Zuber, "*Quantum Field Theory*", McGraw-Hill, 1985, p. 193.

[42] As remarked by Richard Feynman (see Additional Literature in the present chapter, p. 140) for unknown reasons physics is based on the interaction of objects of spin $\frac{1}{2}$ (like electrons or quarks) mediated by objects of spin 1 (like photons, gluons or W particles).

[43] With identical result, that increases enormously the confidence one may place in such results.

[44] K. Pachucki, J. Sapirstein, *Phys. Rev. A* 63 (2001) 12504.

[45] W. Cencek, K. Szalewicz, B. Jeziorski, *Phys. Rev. Letters* 86 (2001) 5675.

Table 3.1. Contributions of various physical effects (non-relativistic, Breit, QED and beyond QED) to the ionization energy and the dipole polarizability α of the helium atom as well as comparison with the experimental values (all quantities in atomic units, i.e. $e = 1$, $\hbar = 1$, $m_0 = 1$, where m_0 denotes the rest mass of electron). The first column gives the symbol of the term in the Breit–Pauli Hamiltonian (3.70) as well as of the QED corrections given order by order (first corresponding to the electron–positron vacuum polarization (QED), then, beyond quantum electrodynamics, to other particle–antiparticle pairs (non-QED): μ, π, ...) split into several separate effects. The second column contains a short description of the effect. The estimated error (third column) is given in parentheses in the units of the last figure reported

Term	Physical interpretation	Ionization energy [MHz]	α [a.u.$\times 10^{-6}$][1]
$\hat{H}_0$	**Schrödinger equation**	**5 945 262 288.62(4)**	**1 383 809.986(1)**
δ	**non-zero size of the nucleus**	**−29.55(4)**	**0.022(1)**
$\hat{H}_1$	**p^4 term**	**1 233 305.45(1)**	**−987.88(1)**
$\hat{H}_2$(el-el)	electron–electron retardation (Breit interaction)	48 684.88(1)	−23.219(1)
$\hat{H}_2$(el-n)	electron–nucleus retardation (Breit interaction)	319.16(1)	−0.257(3)
$\hat{H}_2$	**Breit interaction (total)**	**49 004.04(1)**	**−23.476(3)**
$\hat{H}_3$	**spin–orbit**	0	0
$\hat{H}_4$(el-el)	electron–electron Darwin term	117 008.83(1)	−66.083(1)
$\hat{H}_4$(el-n)	electron–nucleus Darwin term	−1 182 100.99(1)	864.85(2)
$\hat{H}_4$	**Darwin term (total)**	**−1 065 092.16(1)**	**798.77(2)**
$\hat{H}_5$	**spin–spin (total)**	**−234 017.66(1)**	**132.166(1)**
$\hat{H}_6$	**spin-field**	0	0
QED(c^{-3})	vacuum polarization correction to electron–electron interaction	−72.48(1)	0.41(1)
QED(c^{-3})	vacuum polarization correction to electron–nucleus interaction	1 463.00(1)	−1.071(1)
QED(c^{-3})	Total vacuum polarization in c^{-3} order	1 390.52(1)	−1.030(1)
QED(c^{-3})	**vac.pol. + other c^{-3} QED correction**	**−40 483.98(5)**	**30.66(1)**
QED(c^{-4})	vacuum polarization	12.26(1)	0.009(1)
QED(c^{-4})	**Total c^{-4} QED correction**	**−834.9(2)**	**0.56(22)**
QED-h.o.	**Estimation of higher order QED correction**	**84(42)**	**−0.06(6)**
non-QED	**contribution of virtual muons, pions, etc.**	**0.05(1)**	**−0.004(1)**
$\sum$	**Theory (total)**	**5 945 204 223(42)**[2]	**1 383 760.79(23)**
	Experiment	**5 945 204 238(45)**[3]	**1 383 791(67)**[4]

[1] G. Łach, B. Jeziorski, K. Szalewicz, *Phys. Rev. Letters* 92 (2004) 233001.
[2] G.W.F. Drake, W.C. Martin, *Can. J. Phys*. 76 (1998) 679; V. Korobov, A. Yelkhovsky, *Phys. Rev. Letters* 87 (2001) 193003.
[3] K.S.E. Eikema, W. Ubachs, W. Vassen, W. Hogervorst, *Phys. Rev. A* 55 (1997) 1866.
[4] F. Weinhold, *J. Phys. Chem.* 86 (1982) 1111.

- *The term p^4 and the total Darwin effect nearly cancel each other for unclear reasons*. This cancellation is being persistently confirmed also in other systems. Mysteriously enough, this pertains not only to the ionization energy, but also to the polarizability.
- After the above mentioned cancellation (of p^4 and Darwin terms), retardation becomes one of the most important relativistic effects. As seen from the Table, the effect is about a 100 times larger (both for the ionization energy and the polarizability) for the electron–electron retardation than for that of the nucleus–electron. This is quite understandable, because the nucleus represents a "massive rock" (it is about 7000 times heavier) in comparison to an electron, it moves slowly and in the nucleus–electron interaction only the electron contributes to the retardation effect. Two electrons make the retardation much more serious.
- Term $\hat{H}_3$ (spin–orbit coupling) is equal to zero for symmetry reasons (for the ground state).
- In the Darwin term, the nucleus–electron vs electron–electron contribution have reversed magnitudes: about 1 : 10 as compared to 100 : 1 in retardation). Again this time it seems intuitively correct. We have the sum of the particle–particle terms in the Hamiltonian $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}[\hat{\boldsymbol{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_1) + \hat{\boldsymbol{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\boldsymbol{r}_2)]$, where $\boldsymbol{\mathcal{E}}$ means an electric field created by two other particles on the particle under consideration. Each of the terms is proportional to $\nabla_i \nabla_i V = \Delta_i V = 4\pi q_i \delta(\boldsymbol{r}_i)$, where δ is the δ Dirac delta function (Appendix E, p. 951), and q_i denotes the charge of the particle "i". The absolute value of the nuclear charge is twice the electron charge.
- In term $\hat{H}_5$ the spin–spin relates to the electron–electron interaction because the helium nucleus has spin angular momentum of 0.
- The Coulombic interactions are modified by the polarization of vacuum (similar to the weaker interaction of two charges in a dielectric medium). Table 3.1 reports such corrections[46] to the electron–electron and the electron–nucleus interactions [QED(c^{-3})] taking into account that electron–positron pairs jump out from the vacuum. One of these effects is shown in Fig. 3.4.a. As seen from Table 3.1, the nucleus polarizes the vacuum much more easily (about ten times more that the polarization by electrons). Once again the larger charge of the nucleus makes the electric field larger and qualitatively explains the effect. Note that the QED corrections (corresponding to e-p creation) decrease quickly with their order. One of such higher order corrections is shown in Fig. 3.4.d.
- What about the creation of other (than e-p) particle-antiparticle pairs from the vacuum? From (3.71) we see that the larger the rest mass the more difficult it is to squeeze out the corresponding particle-antiparticle pair. And yet, we have some tiny effect (see non-QED entry) corresponding to the creation of such pairs as muon-antimuon (μ), pion-antipion[47] (π), etc. This means that the helium atom is composed of the nucleus and the two electrons only, when we look

[46] However, these effects represent a minor fraction of the total QED(c^{-3}) correction.

[47] Pions are π mesons, the subnuclear particles with mass comparable to that of the muon, a particle about 200 times more massive than an electron. Pions were discovered in 1947 by C.G. Lattes, G.S.P. Occhialini and C.F. Powell.

at it within a certain approximation. To tell the truth, the atom contains also photons, electrons, positrons, muons, pions, and whatever you wish, but with smaller and smaller probability. All that *silva rerum* has only a minor effect of the order of something like the seventh significant figure (both for the ionization potential and for the polarizability).

Summary

The beginning of the twentieth century has seen the birth and development of two revolutionary theories: relativity and quantum mechanics. These two theories turned out to be incompatible, and attempts were made to make them consistent. This chapter consists of two interrelated parts:

- introduction of the elements of relativity theory, and
- attempts to make quantum theory consistent with relativity (relativistic quantum mechanics).

ELEMENTS OF SPECIAL RELATIVITY THEORY

- If experiments are to be described in the same way in two laboratories that move with respect to the partner laboratory with constant velocities v and $-v$, respectively, then the *apparent forces have to vanish*. The same event is described in the two laboratories (by two observers) in the corresponding coordinate system (in one the event happens at coordinate x and time t, in the second – at x' and t'). A sufficient condition that makes the apparent forces vanish is based on linear dependence: $x' = Ax + Bt$ and $t' = Cx + Dt$, where A, B, C, D denote some constants.
- In order to put both observers on the same footing, we have to have $A = D$.
- The Michelson–Morley experiment has shown that each of the observers will note that in the partner's laboratory there is a *contraction* of the dimension pointing to the partner. As a consequence there is a time *dilation*, i.e. each of the observers will note that time flows slower in the partner's laboratory.
- Einstein assumed that in spite of this, any of the observers will measure the same speed of light, c, in his coordinate system.
- This leads to the *Lorentz transformation* that says where and when the two observers see the same event. The Lorentz transformation is especially simple after introducing the Minkowski space (x, ct):

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{Bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{Bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}.$$

  None of the two coordinate systems is privileged (relativity principle).
- Finally, we derived Einstein's formula $E_{\text{kin}} = mc^2$ for the kinetic energy of a body with mass m (this depends on its speed with respect to the coordinate system where the mass is measured).

RELATIVISTIC QUANTUM DYNAMICS

- Fock, Klein and Gordon found the total energy for a particle using the Einstein formula for kinetic energy $E_{\text{kin}} = mc^2$, adding the potential energy and introducing the momen-

tum[48] $p = mv$. After introducing an external electromagnetic field (characterized by the vector potential $\boldsymbol{A}$ and the scalar potential ϕ) they obtained the following relation among operators

$$\left[\frac{i\hbar\frac{\partial}{\partial t} - q\phi}{c}\right]^2 - \left[\left(-i\hbar\nabla - \frac{q}{c}\boldsymbol{A}\right)^2 + m_0^2c^2\right] = 0,$$

where m_0 denotes the rest mass of the particle.

- Paul Dirac factorized the left hand side of this equation by treating it as the difference of squares. This gave two continua of energy separated by a gap of width $2m_0c^2$. Dirac assumed that the lower (negative energy) continuum is fully occupied by electrons ("vacuum"), while the upper continuum is occupied by the single electron (our particle). If we managed to excite an electron from the lower continuum to the upper one, then in the upper continuum we would see an electron, while the hole in the lower continuum would have the properties of a positive electron (positron). This corresponds to the creation of the electron–positron pair from the vacuum.
- The Dirac equation for the electron has the form:

$$\left(i\hbar\frac{\partial}{\partial t}\right)\Psi = \left(q\phi + c\sum_{\mu=x,y,z}\alpha_\mu\pi_\mu + \alpha_0 m_0c^2\right)\Psi,$$

where π_μ in the absence of magnetic field is equal to the momentum operator $\hat{p}_\mu$, $\mu = x, y, z$, while α_μ stand for the square *matrices* of the rank 4, which are related to the Pauli matrices (cf. introduction of spin, Chapter 1). In consequence, the wavefunction Ψ has to be a four-component vector composed of square integrable functions (*bispinor*).

- The Dirac equation demonstrated "pathological" behaviour when a numerical solution was sought. The very reason for this was the decoupling of the electron and positron equations. The exact separation of the negative and positive energy continua has been demonstrated by Barysz and Sadlej, but it leads to a more complex theory. Numerical troubles are often removed by an *ad hoc* assumption called *kinetic balancing*, i.e. fixing a certain relation among the bispinor components. By using this relation we prove that there are two *large* and two *small* (smaller by a factor of about $\frac{v}{2c}$) components of the bispinor.[49]
- The kinetic balance can be used to eliminate the small components from the Dirac equation. Then, the assumption $c = \infty$ (non-relativistic approximation) leads to the *Schrödinger equation* for a single particle.
- The Dirac equation for a particle in the electromagnetic field contains the interaction of the spin magnetic moment with the magnetic field. In this way spin angular momentum appears in the Dirac theory in a natural way (as opposed to the non-relativistic case, where it has had to be postulated).
- The problem of an electron in the external electric field produced by the nucleus (the hydrogen-like atom) has been solved exactly. It turned out that the relativistic corrections are important only for systems with heavy atoms.
- It has been demonstrated in a step-by-step calculation how to obtain an approximate solution of the Dirac equation for the hydrogen-like atom. One of the results is that the relativistic orbitals are contracted compared to the non-relativistic ones.

[48] They wanted to involve the momentum in the formula to be able to change the energy expression to an operator ($\boldsymbol{p} \to \hat{\boldsymbol{p}}$) according to the postulates of quantum mechanics.

[49] For solutions with negative energies this relation is reversed.

- Finally, the Breit equation has been given. The equation goes beyond the Dirac model, by taking into account the *retardation* effects. The Pauli–Breit expression for the Breit Hamiltonian contains several easily interpretable physical effects.
- Quantum electrodynamics (QED) provides an even better description of the system by adding *radiative* effects that take into account the interaction of the particles with the vacuum.

Main concepts, new terms

apparent forces (p. 93)
inertial system (p. 95)
Galilean transformation (p. 96)
Michelson–Morley experiment (p. 96)
length contraction (p. 100)
Lorentz transformation (p. 100)
velocity addition law (p. 103)
relativity principle (p. 104)
Minkowski space-time (p. 104)
time dilation (p. 105)
relativistic mass (p. 107)
Einstein equation (p. 108)
Klein–Gordon equation (p. 109)
Dirac electronic sea (p. 111)
Dirac vacuum (p. 112)
energy continuum (p. 112)
positron (p. 113)
anticommutation relation (p. 114)
Dirac equation (p. 115)
spinors and bispinors (p. 115)
kinetic balance (p. 119)
electron spin (p. 122)
Darwin solution (p. 123)
contraction of orbitals (p. 128)
retarded potential (p. 130)
Breit equation (p. 131)
spin–orbit coupling (p. 132)
spin–spin coupling (p. 132)
Fermi contact term (p. 132)
Zeeman effect (p. 132)
vacuum polarization (p. 133)
particle–antiparticle creation (p. 134)
virtual photons (p. 134)

From the research front

Dirac theory within the mean field approximation (Chapter 8) is routinely applied to molecules and allows us to estimate the relativistic effects even for large molecules. In the computer era, this means, that there are commercial programs available that allow anybody to perform relativistic calculations.

Much worse is the situation with more accurate calculations. The first estimation for molecules of relativistic effects beyond the Dirac approximation has been carried out by Janos Ladik[50] and then by Jeziorski and Kołos[51] while the first calculation of the interaction with the vacuum for molecules was done by Bukowski et al.[52] Besides the recent computation of the Lamb shift for the water molecule,[53] not much has been computed in this area.

Ad futurum...

In comparison with typical chemical phenomena, the relativistic effects in almost all instances, remain of marginal significance for biomolecules or for molecules typical of tradi-

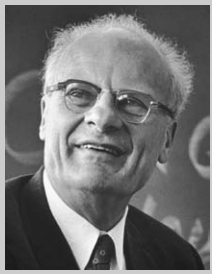
[50] J. Ladik, *Acta Phys. Hung.* 10 (1959) 271.

[51] The calculations were performed for the hydrogen molecular ion H_2^+, B. Jeziorski, W. Kołos, *Chem. Phys. Letters* 3 (1969) 677.

[52] R. Bukowski, B. Jeziorski, R. Moszyński, W. Kołos, *Int. J. Quantum Chem*. 42 (1992) 287.

[53] P. Pyykkö, K.G. Dyall, A.G. Császár, G. Tarczay, O.L. Polyansky, J. Tennyson, *Phys. Rev. A* 63 (2001) 24502.

Hans Albrecht Bethe (1906–2005), American physicist, professor at Cornell University, student of Arnold Sommerfeld. Bethe contributed to many branches of physics, e.g., crystal field theory, interaction of matter with radiation, quantum electrodynamics, structure and nuclear reactions of stars (for the latter achievement he received the Nobel Prize in 1967).



tional organic chemistry. In inorganic chemistry, these effects could however be much more important. Probably the Dirac–Coulomb theory combined with the mean field approach will for a few decades remain a satisfactory standard for the vast majority of researchers. At the same time there will be theoretical and computational progress for small molecules (and for atoms), where Dirac theory will be progressively replaced by quantum electrodynamics.

Additional literature

H. Bethe, E. Salpeter, "Quantum Mechanics of One- and Two-Electron Atoms", Springer, Berlin, 1957.

This book is absolutely exceptional. It is written by excellent specialists in such a competent way and with such care (no misprints), that despite the lapse of many decades it remains the fundamental and best source.

I.M. Grant, H.M. Quiney, "Foundations of the Relativistic Theory of Atomic and Molecular Structure", *Adv. At. Mol. Phys*., 23 (1988) 37.

Very good review.

L. Pisani, J.M. André, M.C. André, E. Clementi, *J. Chem. Educ*., **70**, 894–901 (1993), also J.M. André, D.H. Mosley, M.C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, Exploring Aspects of Computational Chemistry: *Vol. I, Concepts*, Presses Universitaires de Namur, pp. 150–166 (1997), Vol. II, Exercises, Presses Universitaires de Namur, p. 249–272 (1997).

Fine article, fine book, written clearly, its strength is also in very simple examples of the application of the theory.

R.P. Feynman, "QED – The Strange Theory of Light and Matter", Princeton University Press, Princeton, 1988.

Excellent book written by one of the celebrities of our times in the style "*quantum electrodynamics not only for poets*".

Questions

1. In the Lorentz transformation the two coordinate systems:
   a) are both at rest; b) move with the same velocity; c) are related also by Galilean transformation; d) have x' and t' depending linearly on x and t.

2. The Michelson–Morley experiment has shown that when an observer in the coordinate system O measures a length in O' (both coordinate systems fly apart; $v' = -v$), then he obtains:

a) the same result that is obtained by an observer in O'; b) contraction of lengths along the direction of the motion; c) expansion of lengths along the direction of the motion; d) contraction of lengths in any direction.

3. An observer in O measures the times a phenomenon takes in O and O' (both coordinate systems fly apart; $v' = -v$):
a) the time of the phenomenon going on in O will be shorter; b) time goes with the same speed in O'; c) time goes more slowly in O' only if $|v| > \frac{c}{2}$; d) time goes more slowly in O' only if $|v| < \frac{c}{2}$.

4. In the Minkowski space, the distance of any event from the origin (both coordinate systems fly apart; $v' = -v$) is:
a) equal to vt; b) equal to ct; c) the same for observers in O and in O'; d) equal to 0.

5. A bispinor represents:
a) a two-component vector with functions as components; b) a two-component vector with complex numbers as components; c) a four-component vector with square integrable functions as components; d) a scalar square integrable function.

6. Non-physical results of numerical solutions to the Dirac equation appear because:
a) the Dirac sea is neglected; b) the electron and positron have the same energies; c) the electron has kinetic energy equal to its potential energy; d) the electron has zero kinetic energy.

7. The Schrödinger equation can be deduced from the Dirac equation under the assumption that:
a) $v = c$; b) v/c is small; c) all components of the bispinor have equal length; d) the magnetic field is zero.

8. In the Breit equation there is an approximate cancellation of:
a) the retardation effect with the non-zero size of the nucleus effect; b) the retardation effect electron–electron with that of electron–nucleus; c) the spin–spin effect with the Darwin term; d) the Darwin term with the p^4 term.

9. Dirac's hydrogen atom orbitals when compared to Schrödinger's are:
a) more concentrated close to the nucleus, but have a larger mean value of r; b) have a larger mean value of r; c) more concentrated close to the nucleus; d) of the same size, because the nuclear charge has not changed.

10. The Breit equation: a) is invariant with respect to the Lorentz transformation; b) takes into account the interaction of the magnetic moments of electrons resulting from their orbital motion; c) neglects the interaction of the spin magnetic moments; d) describes only a single particle.

Answers

1d, 2b, 3a, 4c, 5c, 6a, 7b, 8d, 9c, 10b