It should be noted that $\bar{\psi}^\dagger \alpha_m \psi$ is invariant, since

$$\bar{\psi}^{*\dagger} \alpha_m \psi^* = \bar{\psi}^\dagger (1 - M\alpha_m)\alpha_m (1 + \alpha_m M)\psi$$
$$= \bar{\psi}^\dagger \alpha_m \psi.$$

We must verify finally the conservation law, that the divergence

$$\frac{\partial}{\partial x_\mu} (\bar{\psi}^\dagger \alpha_\mu \psi) \tag{22}$$

vanishes. To prove this, multiply equation (13) by $\bar{\psi}^\dagger$ on the left. The result is

$$\bar{\psi}^\dagger \alpha^\mu \left(i\hbar \frac{\partial \psi}{\partial x^\mu} + \frac{e}{c} A_\mu \psi \right) - \bar{\psi}^\dagger \alpha_m mc\psi = 0.$$

The conjugate imaginary equation is

$$\left(-i\hbar \frac{\partial \bar{\psi}^\dagger}{\partial x^\mu} + \bar{\psi}^\dagger \frac{e}{c} A_\mu \right) \alpha^\mu \psi - \bar{\psi}^\dagger \alpha_m mc\psi = 0.$$

Subtracting and dividing by $i\hbar$, we get

$$\bar{\psi}^\dagger \alpha^\mu \frac{\partial \psi}{\partial x^\mu} + \frac{\partial \bar{\psi}^\dagger}{\partial x^\mu} \alpha^\mu \psi = 0,$$

which just expresses the vanishing of (22). In this way we complete the proof that our theory gives consistent results in whichever frame of reference it is applied.

69. The motion of a free electron

It is of interest to consider the motion of a free electron in the Heisenberg picture according to the above theory and to study the Heisenberg equations of motion. These equations of motion can be integrated exactly, as was first done by Schrödinger.‡ For brevity we shall omit the suffix t which the notation of § 28 requires to be inserted in dynamical variables that vary with time in the Heisenberg picture.

As Hamiltonian we must take the expression which we get as equal to cp_0 when we put the operator on ψ in (10) equal to zero, i.e.

$$H = c\rho_1(\boldsymbol{\sigma}, \mathbf{p}) + \rho_3 mc^2 = c(\boldsymbol{\alpha}, \mathbf{p}) + \rho_3 mc^2. \tag{23}$$

We see at once that the momentum commutes with H and is thus a constant of the motion. Further, the x_1-component of the velocity is

$$\dot{x}_1 = [x_1, H] = c\alpha_1. \tag{24}$$

This result is rather surprising, as it means an altogether different

‡ Schrödinger, *Sitzungsb. d. Berlin. Akad.*, 1930, p. 418.

relation between velocity and momentum from what one has in classical mechanics. It is connected, however, with the expression $\bar{\psi}^{\dagger} c\alpha_1 \psi$ for a component of the probability current. The $\dot{x}_1$ given by (24) has as eigenvalues $\pm c$, corresponding to the eigenvalues ± 1 of α_1. As $\dot{x}_2$ and $\dot{x}_3$ are similar, we can conclude that *a measurement of a component of the velocity of a free electron is certain to lead to the result* $\pm c$. This conclusion is easily seen to hold also when there is a field present.

Since electrons are observed in practice to have velocities considerably less than that of light, it would seem that we have here a contradiction with experiment. The contradiction is not real, though, since the theoretical velocity in the above conclusion is the velocity at one instant of time while observed velocities are always average velocities through appreciable time intervals. We shall find upon further examination of the equations of motion that the velocity is not at all constant, but oscillates rapidly about a mean value which agrees with the observed value.

It may easily be verified that a measurement of a component of the velocity must lead to the result $\pm c$ in a relativistic theory, simply from an elementary application of the principle of uncertainty of § 24. To measure the velocity we must measure the position at two slightly different times and then divide the change of position by the time interval. (It will not do to measure the momentum and apply a formula, as the ordinary connexion between velocity and momentum is not valid.) In order that our measured velocity may approximate to the instantaneous velocity, the time interval between the two measurements of position must be very short and hence these measurements must be very accurate. The great accuracy with which the position of the electron is known during the time-interval must give rise, according to the principle of uncertainty, to an almost complete indeterminacy in its momentum. This means that almost all values of the momentum are equally probable, so that the momentum is almost certain to be infinite. An infinite value for a component of momentum corresponds to the value $\pm c$ for the corresponding component of velocity.

Let us now examine how the velocity of the electron varies with time. We have

$$i\hbar\dot{\alpha}_1 = \alpha_1 H - H\alpha_1.$$

Now since α_1 anticommutes with all the terms in H except $c\alpha_1 p_1$,

$$\alpha_1 H + H\alpha_1 = \alpha_1 c\alpha_1 p_1 + c\alpha_1 p_1 \alpha_1 = 2cp_1,$$

and hence

$$\left.\begin{aligned} i\hbar\dot{\alpha}_1 &= 2\alpha_1 H - 2cp_1, \\ &= -2H\alpha_1 + 2cp_1. \end{aligned}\right\} \quad (25)$$

Since H and p_1 are constants, it follows from the first of equations (25) that

$$i\hbar\ddot{\alpha}_1 = 2\dot{\alpha}_1 H. \quad (26)$$

This differential equation in $\dot{\alpha}_1$ can be integrated immediately, the result being

$$\dot{\alpha}_1 = \dot{\alpha}_1^0 e^{-2iHt/\hbar}, \quad (27)$$

where $\dot{\alpha}_1^0$ is a constant, equal to the value of $\dot{\alpha}_1$ when $t = 0$. The factor $e^{-2iHt/\hbar}$ must be put to the right of the factor $\dot{\alpha}_1^0$ in (27) on account of the H occurring to the right of the $\dot{\alpha}_1$ in (26). The second of equations (25) leads in the same way to the result

$$\dot{\alpha}_1 = e^{2iHt/\hbar}\dot{\alpha}_1^0.$$

We can now easily complete the integration of the equation of motion for x_1. From (27) and the first of equations (25)

$$\alpha_1 = \tfrac{1}{2}i\hbar\dot{\alpha}_1^0 e^{-2iHt/\hbar}H^{-1} + cp_1 H^{-1}, \quad (28)$$

and hence the time-integral of equation (24) is

$$x_1 = -\tfrac{1}{4}c\hbar^2\dot{\alpha}_1^0 e^{-2iHt/\hbar}H^{-2} + c^2 p_1 H^{-1}t + a_1, \quad (29)$$

a_1 being a constant.

From (28) we see that the x_1 component of velocity, $c\alpha_1$, consists of two parts, a constant part $c^2 p_1 H^{-1}$, connected with the momentum by the classical relativistic formula, and an oscillatory part

$$\tfrac{1}{2}ic\hbar\dot{\alpha}_1^0 e^{-2iHt/\hbar}H^{-1},$$

whose frequency is high, being $2H/h$, which is at least $2mc^2/h$. Only the constant part would be observed in a practical measurement of velocity, such a measurement giving the average velocity through a time-interval much larger than $h/2mc^2$. The oscillatory part secures that the instantaneous value of $\dot{x}_1$ shall have the eigenvalues $\pm c$. The oscillatory part of x_1 is small, being, according to (29),

$$-\tfrac{1}{4}c\hbar^2\dot{\alpha}_1^0 e^{-2iHt/\hbar}H^{-2} = \tfrac{1}{2}ic\hbar(\alpha_1 - cp_1 H^{-1})H^{-1},$$

which is of the order of magnitude $\hbar/mc$, since $(\alpha_1 - cp_1 H^{-1})$ is of the order of magnitude unity.

70. Existence of the spin

In § 67 we saw that the correct wave equation for the electron in the absence of an electromagnetic field, namely equation (7) or (10), is equivalent to the wave equation (6) which is suggested from analogy

with the classical theory. This equivalence no longer holds when there is a field. The wave equation to be expected from analogy with the classical theory in this case is

$$\left\{\left(p_0+\frac{e}{c}A_0\right)^2-\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2-m^2c^2\right\}\psi=0, \tag{30}$$

in which the operator is just the classical relativistic Hamiltonian. If we multiply (11) by some factor on the left to make it resemble (30) as closely as possible, namely the factor

$$p_0+\frac{e}{c}A_0+\rho_1\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)+\rho_3 mc,$$

we get

$$\left\{\left(p_0+\frac{e}{c}A_0\right)^2-\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2-m^2c^2-\rho_1\left[\left(p_0+\frac{e}{c}A_0\right)\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)-\right.\right.$$
$$\left.\left.-\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)\left(p_0+\frac{e}{c}A_0\right)\right]\right\}\psi=0. \tag{31}$$

We now use the general formula that, if $\mathbf{B}$ and $\mathbf{C}$ are any two three-dimensional vectors that commute with $\boldsymbol{\sigma}$,

$$(\boldsymbol{\sigma},\mathbf{B})(\boldsymbol{\sigma},\mathbf{C})=\sum_{123}\{\sigma_1^2 B_1C_1+\sigma_1\sigma_2 B_1C_2+\sigma_2\sigma_1 B_2C_1\},$$

the summation referring to cyclic permutations of the suffixes 1, 2, 3, or

$$(\boldsymbol{\sigma},\mathbf{B})(\boldsymbol{\sigma},\mathbf{C})=(\mathbf{B},\mathbf{C})+i\sum_{123}\sigma_3(B_1C_2-B_2C_1)$$
$$=(\mathbf{B},\mathbf{C})+i(\boldsymbol{\sigma},\mathbf{B}\times\mathbf{C}). \tag{32}$$

Taking $\mathbf{B}=\mathbf{C}=\mathbf{p}+e/c.\mathbf{A}$, we find, since

$$\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)\times\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)=\frac{e}{c}\{\mathbf{p}\times\mathbf{A}+\mathbf{A}\times\mathbf{p}\}$$
$$=-i\hbar e/c.\operatorname{curl}\mathbf{A}=-i\hbar e/c.\mathcal{H},$$

where $\mathcal{H}$ is the magnetic field, that

$$\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2=\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2+\frac{\hbar e}{c}(\boldsymbol{\sigma},\mathcal{H}). \tag{33}$$

Also we have

$$\left(p_0+\frac{e}{c}A_0\right)\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)-\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)\left(p_0+\frac{e}{c}A_0\right)$$
$$=\frac{e}{c}(\boldsymbol{\sigma},p_0\mathbf{A}-\mathbf{A}p_0+A_0\mathbf{p}-\mathbf{p}A_0)$$
$$=\frac{i\hbar e}{c}\left(\boldsymbol{\sigma},\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}+\operatorname{grad}A_0\right)=-i\frac{\hbar e}{c}(\boldsymbol{\sigma},\mathcal{E}),$$

where $\mathcal{E}$ is the electric field. Thus (31) becomes

$$\left\{\left(p_0+\frac{e}{c}A_0\right)^2-\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2-m^2c^2-\frac{\hbar e}{c}(\boldsymbol{\sigma},\mathcal{H})+i\rho_1\frac{\hbar e}{c}(\boldsymbol{\sigma},\mathcal{E})\right\}\psi=0. \tag{34}$$

This equation differs from (30) through having two extra terms in the operator. These extra terms involve some new physical effects, but since they are not real they do not lend themselves very directly to physical interpretation.

To get an understanding of the physical features involved in the difference between (34) and (30) it is better to work with the Heisenberg picture, this picture being always the more suitable one for comparisons between classical and quantum mechanics. The Heisenberg equations of motion are determined by the Hamiltonian

$$H=-eA_0+c\rho_1\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)+\rho_3 mc^2, \tag{35}$$

the generalization of (23) to the case when there is a field. Equation (35) gives

$$\begin{aligned}\left(\frac{H}{c}+\frac{e}{c}A_0\right)^2&=\left\{\rho_1\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)+\rho_3 mc\right\}^2\\&=\left(\boldsymbol{\sigma},\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2+m^2c^2\\&=\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2+m^2c^2+\frac{\hbar e}{c}(\boldsymbol{\sigma},\mathcal{H})\end{aligned} \tag{36}$$

with the help of (33). We have here the real part of the extra terms in (34) appearing without the pure imaginary part. For an electron moving slowly (i.e. with small momentum), we may expect the Heisenberg equations of motion to be determined by a Hamiltonian of the form mc^2+H_1, where H_1 is small compared with mc^2. Putting mc^2+H_1 for H in (36) and neglecting H_1^2 and other terms involving c^{-2}, we get, on dividing by $2m$,

$$H_1+eA_0=\frac{1}{2m}\left(\mathbf{p}+\frac{e}{c}\mathbf{A}\right)^2+\frac{\hbar e}{2mc}(\boldsymbol{\sigma},\mathcal{H}). \tag{37}$$

The Hamiltonian H_1 given by (37) is the same as the classical Hamiltonian for a slow electron, except for the last term

$$\frac{\hbar e}{2mc}(\boldsymbol{\sigma},\mathcal{H}).$$

This term may be considered as an additional potential energy

which a slow electron has in the quantum theory and may be interpreted as arising from *the electron having a magnetic moment* $-\hbar e/2mc.\boldsymbol{\sigma}$. This magnetic moment is the one assumed in §§ 41 and 47 for dealing with the Zeeman effect and is in agreement with experiment.

The spin angular momentum does not give rise to any potential energy and therefore does not appear in the result of the preceding calculation. The simplest way of showing the existence of the spin angular momentum is to take the case of the motion of a free electron or an electron in a central field of force and determine the angular momentum integrals. This means working with the Hamiltonian (23), or with the Hamiltonian (35) with $\mathbf{A} = 0$ and A_0 a function of the radius r, i.e.

$$H = -eA_0(r)+c\rho_1(\boldsymbol{\sigma}, \mathbf{p})+\rho_3 mc^2, \tag{38}$$

and obtaining the Heisenberg equations of motion for the angular momentum. With either Hamiltonian we find for the rate of change of the x_1-component of orbital angular momentum, $m_1 = x_2 p_3 - x_3 p_2$, with the help of commutation relations proved in § 35,

$$\begin{aligned} i\hbar\dot{m}_1 &= m_1 H - H m_1 \\ &= c\rho_1\{m_1(\boldsymbol{\sigma}, \mathbf{p})-(\boldsymbol{\sigma}, \mathbf{p})m_1\} \\ &= c\rho_1(\boldsymbol{\sigma}, m_1\,\mathbf{p}-\mathbf{p}m_1) \\ &= i\hbar c\rho_1\{\sigma_2 p_3-\sigma_3 p_2\}. \end{aligned}$$

Thus $\dot{m}_1 \neq 0$ and the orbital angular momentum is not a constant of the motion. This result is to be expected from the integrated equation of motion (29), the oscillatory part of the motion here displayed giving rise to an oscillatory term in the angular momentum.

We have further

$$\begin{aligned} i\hbar\dot{\sigma}_1 &= \sigma_1 H - H\sigma_1 \\ &= c\rho_1\{\sigma_1(\boldsymbol{\sigma}, \mathbf{p})-(\boldsymbol{\sigma}, \mathbf{p})\sigma_1\} \\ &= c\rho_1(\sigma_1\boldsymbol{\sigma}-\boldsymbol{\sigma}\sigma_1, \mathbf{p}) \\ &= 2ic\rho_1\{\sigma_3 p_2-\sigma_2 p_3\} \end{aligned}$$

with the help of equations (51) of § 37. Hence

$$\dot{m}_1+\tfrac{1}{2}\hbar\dot{\sigma}_1 = 0,$$

so that the vector $\mathbf{m}+\frac{1}{2}\hbar\boldsymbol{\sigma}$ is a constant of the motion. This result one can interpret by saying *the electron has a spin angular momentum* $\frac{1}{2}\hbar\boldsymbol{\sigma}$, which must be added to the orbital angular momentum $\mathbf{m}$ before one gets a constant of the motion. The spin angular momentum

could alternatively be obtained from the rotation operators for states of spin in accordance with the general method of § 35.

The same vector $\boldsymbol{\sigma}$ fixes the directions of both the spin magnetic moment and the spin angular momentum. If an electron in a certain state of spin has a spin angular momentum of $\frac{1}{2}\hbar$ in a particular direction, it will have a magnetic moment $-e\hbar/2mc$ in the same direction.

We were led to the value $\frac{1}{2}\hbar$ for the spin of the electron by an argument depending simply on general principles of quantum theory and relativity. One could apply the same argument to other kinds of elementary particle and one would be led to the same conclusion, that the spin angular momentum is half a quantum. This would be satisfactory for the proton and the neutron, but there are some kinds of elementary particle (e.g. the photon and certain kinds of meson) whose spins are known experimentally to be different from $\frac{1}{2}\hbar$, so we have a discrepancy between our theory and experiment.

The answer is to be found in a hidden assumption in our work. Our argument is valid only provided the position of the particle is an observable. If this assumption holds, the particle must have a spin angular momentum of half a quantum. For those particles that have a different spin the assumption must be false and any dynamical variables x_1, x_2, x_3 that may be introduced to describe the position of the particle cannot be observables in accordance with our general theory. For such particles there is no true Schrödinger representation. One might be able to introduce a quasi wave function involving the dynamical variables x_1, x_2, x_3, but it would not have the correct physical interpretation of a wave function—that the square of its modulus gives the probability density. For such particles there is still a momentum representation, which is sufficient for practical purposes.

71. Transition to polar variables

For the further study of the motion of an electron in a central field of force with the Hamiltonian (38), it is convenient to make a transformation to polar coordinates, as was done in § 38 in the non-relativistic case. We can introduce r and p_r as before, but instead of k, the magnitude of the orbital angular momentum $\mathbf{m}$, which is no longer a constant of the motion, we must now use the magnitude of the total angular momentum $\mathbf{M} = \mathbf{m} + \frac{1}{2}\hbar\boldsymbol{\sigma}$. Let us put

$$j^2\hbar^2 = M_1^2 + M_2^2 + M_3^2 + \tfrac{1}{4}\hbar^2. \tag{39}$$

The eigenvalues of m_3 are integral multiples of $\hbar$, those of $\frac{1}{2}\hbar\sigma_3$ are $\pm\frac{1}{2}\hbar$, and hence those of M_3 must be half odd integral multiples of $\hbar$. It follows from the theory of § 36 that the eigenvalues of $|j|$ must be integers greater than zero.

If in formula (32) we take $\mathbf{B} = \mathbf{C} = \mathbf{m}$, we get

$$\begin{aligned}(\boldsymbol{\sigma}, \mathbf{m})^2 &= \mathbf{m}^2 + i(\boldsymbol{\sigma}, \mathbf{m}\times\mathbf{m}) \\ &= \mathbf{m}^2 - \hbar(\boldsymbol{\sigma}, \mathbf{m}) \\ &= (\mathbf{m}+\tfrac{1}{2}\hbar\boldsymbol{\sigma})^2 - 2\hbar(\boldsymbol{\sigma}, \mathbf{m}) - \tfrac{3}{4}\hbar^2.\end{aligned}$$

Hence $$\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\}^2 = \mathbf{M}^2 + \tfrac{1}{4}\hbar^2.$$

Thus $(\boldsymbol{\sigma}, \mathbf{m})+\hbar$ is a quantity whose square is $\mathbf{M}^2+\frac{1}{4}\hbar^2$ and we could, consistently with equation (39), define $j\hbar$ as $(\boldsymbol{\sigma}, \mathbf{m})+\hbar$. This would not be the most convenient definition for j, however, since we would like to have j a constant of the motion and $(\boldsymbol{\sigma}, \mathbf{m})+\hbar$ is not constant. We have, in fact, from applications of (32),

$$(\boldsymbol{\sigma}, \mathbf{m})(\boldsymbol{\sigma}, \mathbf{p}) = i(\boldsymbol{\sigma}, \mathbf{m}\times\mathbf{p})$$

and

$$(\boldsymbol{\sigma}, \mathbf{p})(\boldsymbol{\sigma}, \mathbf{m}) = i(\boldsymbol{\sigma}, \mathbf{p}\times\mathbf{m}),$$

so that

$$\begin{aligned}(\boldsymbol{\sigma}, \mathbf{m})(\boldsymbol{\sigma}, \mathbf{p}) + (\boldsymbol{\sigma}, \mathbf{p})(\boldsymbol{\sigma}, \mathbf{m}) &= i\sum_{123}\sigma_1\{m_2p_3 - m_3p_2 + p_2m_3 - p_3m_2\} \\ &= i\sum_{123}\sigma_1 . 2i\hbar p_1 = -2\hbar(\boldsymbol{\sigma}, \mathbf{p}),\end{aligned}$$

or

$$\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\}(\boldsymbol{\sigma}, \mathbf{p}) + (\boldsymbol{\sigma}, \mathbf{p})\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\} = 0.$$

Thus $(\boldsymbol{\sigma}, \mathbf{m})+\hbar$ anticommutes with one of the terms in the expression (38) for H, namely the term $c\rho_1(\boldsymbol{\sigma}, \mathbf{p})$, and commutes with the other two. It follows that $\rho_3\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\}$ commutes with all the three terms in H and is a constant of the motion. But the square of $\rho_3\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\}$ is also $\mathbf{M}^2+\frac{1}{4}\hbar^2$. We can therefore take

$$j\hbar = \rho_3\{(\boldsymbol{\sigma}, \mathbf{m})+\hbar\}, \tag{40}$$

which gives us a convenient rational definition for j which is consistent with (39) and makes j a constant of the motion. The eigenvalues of this j are all positive and negative integers, excluding zero.

By a further application of (32), we get

$$\begin{aligned}(\boldsymbol{\sigma}, \mathbf{x})(\boldsymbol{\sigma}, \mathbf{p}) &= (\mathbf{x}, \mathbf{p}) + i(\boldsymbol{\sigma}, \mathbf{m}) \\ &= rp_r + i\rho_3 j\hbar - i\hbar,\end{aligned} \tag{41}$$

with the help of (40) and also of equation (58) of § 38. We introduce the linear operator ϵ defined by

$$r\epsilon = \rho_1(\boldsymbol{\sigma}, \mathbf{x}). \tag{42}$$

Since r commutes with ρ_1 and with $(\boldsymbol{\sigma}, \mathbf{x})$, it must commute with ϵ. We thus have

$$r^2\epsilon^2 = [\rho_1(\boldsymbol{\sigma}, \mathbf{x})]^2 = (\boldsymbol{\sigma}, \mathbf{x})^2 = \mathbf{x}^2 = r^2,$$

or

$$\epsilon^2 = 1.$$

Now $\rho_1(\boldsymbol{\sigma}, \mathbf{p})$ commutes with j, and since there is symmetry between $\mathbf{x}$ and $\mathbf{p}$ so far as angular momentum is concerned, $\rho_1(\boldsymbol{\sigma}, \mathbf{x})$ must also commute with j. Hence ϵ commutes with j. Further, ϵ must commute with p_r, since we have

$$(\boldsymbol{\sigma}, \mathbf{x})(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})(\boldsymbol{\sigma}, \mathbf{x}) = (\boldsymbol{\sigma}, \mathbf{x}(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})\mathbf{x}) = i\hbar(\boldsymbol{\sigma}, \mathbf{x}),$$

which gives

$$r\epsilon r p_r - r p_r r\epsilon = i\hbar r\epsilon,$$

or

$$r^2\epsilon p_r - r^2 p_r \epsilon = 0.$$

From (41) and (42) we obtain

$$r\epsilon\rho_1(\boldsymbol{\sigma},\mathbf{p}) = r p_r + i\rho_3 j\hbar - i\hbar,$$

or

$$\rho_1(\boldsymbol{\sigma}, \mathbf{p}) = \epsilon(p_r - i\hbar/r) + i\epsilon\rho_3 j\hbar/r.$$

Thus (38) becomes

$$H/c = -e/c\,.A_0 + \epsilon(p_r - i\hbar/r) + i\epsilon\rho_3 j\hbar/r + \rho_3 mc.$$

This gives our Hamiltonian expressed in terms of polar variables. It should be noticed that ϵ and ρ_3 commute with all the other variables occurring in H and anticommute with one another. This means that we can take a representation with ρ_3 diagonal in which ϵ and ρ_3 are represented respectively by the matrices

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{43}$$

If r is also diagonal in the representation, the representative $\langle r'\rho_3'|\rangle$ of a ket will have two components, $\langle r', 1|\rangle = \psi_a(r')$ and $\langle r', -1|\rangle = \psi_b(r')$ say, referring to the two rows and columns of the matrices (43).

72. The fine-structure of the energy-levels of hydrogen

We shall now take the case of the hydrogen atom, for which $A_0 = e/r$, and work out its energy-levels, given by the eigenvalues H' of H. The equation $(H' - H)|\rangle = 0$ which defines these eigenvalues, when written in terms of representatives in the representation discussed

above with ϵ and ρ_3 represented by the matrices (43), gives the equations

$$\left(\frac{H'}{c}+\frac{e^2}{cr}\right)\psi_a+\hbar\left(\frac{\partial}{\partial r}+\frac{1}{r}\right)\psi_b+\frac{j\hbar}{r}\psi_b-mc\psi_a=0,$$

$$\left(\frac{H'}{c}+\frac{e^2}{cr}\right)\psi_b-\hbar\left(\frac{\partial}{\partial r}+\frac{1}{r}\right)\psi_a+\frac{j\hbar}{r}\psi_a+mc\psi_b=0.$$

If we put
$$\frac{\hbar}{mc-H'/c}=a_1, \qquad \frac{\hbar}{mc+H'/c}=a_2, \tag{44}$$

these equations reduce to

$$\left.\begin{aligned}\left(\frac{1}{a_1}-\frac{\alpha}{r}\right)\psi_a-\left(\frac{\partial}{\partial r}+\frac{j+1}{r}\right)\psi_b=0,\\ \left(\frac{1}{a_2}+\frac{\alpha}{r}\right)\psi_b-\left(\frac{\partial}{\partial r}-\frac{j-1}{r}\right)\psi_a=0,\end{aligned}\right\} \tag{45}$$

where $\alpha=e^2/\hbar c$, which is a small number. We shall solve these equations by a similar method to that used for equation (73) in § 39.

Put
$$\psi_a=r^{-1}e^{-r/a}f, \qquad \psi_b=r^{-1}e^{-r/a}g, \tag{46}$$

introducing two new functions, f and g, of r, where

$$a=(a_1a_2)^{\frac{1}{2}}=\hbar(m^2c^2-H'^2/c^2)^{-\frac{1}{2}}. \tag{47}$$

Equations (45) become

$$\left.\begin{aligned}\left(\frac{1}{a_1}-\frac{\alpha}{r}\right)f-\left(\frac{\partial}{\partial r}-\frac{1}{a}+\frac{j}{r}\right)g=0,\\ \left(\frac{1}{a_2}+\frac{\alpha}{r}\right)g-\left(\frac{\partial}{\partial r}-\frac{1}{a}-\frac{j}{r}\right)f=0.\end{aligned}\right\} \tag{48}$$

We now try for a solution in which f and g are in the form of power series,

$$f=\sum_s c_s r^s, \qquad g=\sum_s c'_s r^s, \tag{49}$$

in which consecutive values of s differ by unity though these values need not be integers. Substituting these expressions for f and g in (48) and picking out coefficients of r^{s-1}, we obtain

$$\left.\begin{aligned}c_{s-1}/a_1-\alpha c_s-(s+j)c'_s+c'_{s-1}/a=0,\\ c'_{s-1}/a_2+\alpha c'_s-(s-j)c_s+c_{s-1}/a=0.\end{aligned}\right\} \tag{50}$$

By multiplying the first of these equations by a and the second by a_2 and subtracting, we eliminate both c_{s-1} and c'_{s-1}, since from (47) $a/a_1=a_2/a$. We are left with

$$[a\alpha-a_2(s-j)]c_s+[a_2\alpha+a(s+j)]c'_s=0, \tag{51}$$

a relation which shows the connexion between the primed and unprimed c's.

The boundary condition at $r = 0$ requires that $r\psi_a$ and $r\psi_b \to 0$ as $r \to 0$, so from (46) f and $g \to 0$ as $r \to 0$. Thus the series (49) must terminate on the side of small s. If s_0 is the minimum value of s for which c_s and c'_s do not both vanish, we obtain from (50), by putting $s = s_0$ and $c_{s_0-1} = c'_{s_0-1} = 0$,

$$\left.\begin{aligned} \alpha c_{s_0} + (s_0+j)c'_{s_0} &= 0, \\ \alpha c'_{s_0} - (s_0-j)c_{s_0} &= 0, \end{aligned}\right\} \quad (52)$$

which give $$\alpha^2 = -s_0^2 + j^2.$$

Since the boundary condition requires that the minimum value of s shall be greater than zero, we must take

$$s_0 = +\sqrt{(j^2-\alpha^2)}.$$

To investigate the convergence of the series (49) we shall determine the ratio c_s/c_{s-1} for large s. Equation (51) and the second of equations (50) give approximately, when s is large,

$$a_2 c_s = a c'_s$$

and $$s c_s = c_{s-1}/a + c'_{s-1}/a_2.$$

Hence $$c_s/c_{s-1} = 2/as.$$

The series (49) will therefore converge like

$$\sum_s \frac{1}{s!}\left(\frac{2r}{a}\right)^s,$$

or $e^{2r/a}$. This result is similar to that obtained in § 39 and allows us to infer, as in § 39, that all values of H' are permissible for which a is pure imaginary, i.e. from (47), for which $H' > mc^2$, while for $H' < mc^2$ we take a to be positive and then find that only those values of H' are permissible for which the series (49) terminate on the side of large s.

If the series (49) terminate with the terms c_s and c'_s, so that $c_{s+1} = c'_{s+1} = 0$, we obtain from (50) with $s+1$ substituted for s

$$\left.\begin{aligned} c_s/a_1 + c'_s/a &= 0, \\ c'_s/a_2 + c_s/a &= 0. \end{aligned}\right\} \quad (53)$$

These two equations are equivalent on account of (47). When combined with (51), they give

$$a_1[a\alpha - a_2(s-j)] = a[a_2\alpha + a(s+j)],$$

which reduces to $$2a_1a_2s = a(a_1-a_2)\alpha,$$

or $$\frac{s}{a} = \frac{1}{2}\left(\frac{1}{a_2}-\frac{1}{a_1}\right)\alpha = \frac{H'}{c\hbar}\alpha,$$

with the help of (44). Squaring and using (47), we obtain

$$s^2(m^2c^2-H'^2/c^2) = \alpha^2H'^2/c^2.$$

Hence $$\frac{H'}{mc^2} = \left(1+\frac{\alpha^2}{s^2}\right)^{-\frac{1}{2}}.$$

The s here, which specifies the last term in the series, must be greater than s_0 by some integer not less than zero. Calling this integer n, we have

$$s = n+\sqrt{(j^2-\alpha^2)}$$

and thus $$\frac{H'}{mc^2} = \left\{1+\frac{\alpha^2}{\{n+\sqrt{(j^2-\alpha^2)}\}^2}\right\}^{-\frac{1}{2}}. \tag{54}$$

This formula gives the discrete energy-levels of the hydrogen spectrum and was first obtained by Sommerfeld working with Bohr's orbit theory. There are two quantum numbers n and j involved, but owing to α^2 being very small the energy depends almost entirely on $n+|j|$. Values of n and $|j|$ that give the same $n+|j|$ give rise to a set of energy-levels lying very close to one another, and to the energy-level given by the non-relativistic formula (80) of § 39 with $s = n+|j|$, apart from the constant term mc^2.

We used equations (53) by combining them with (51), but this does not make full use of (53) since the coefficients of c_s and c'_s in (51) may both vanish. In this case we get, multiplying the first coefficient by a_1 and the second by a and adding,

$$a(a_1+a_2)\alpha+2a_1a_2j = 0.$$

Thus j must be negative in this case. With the help of (44) and (47) we get further

$$-\frac{2j}{\alpha} = \frac{a}{a_2}+\frac{a}{a_1} = \frac{2mca}{\hbar} = \frac{2mc}{(m^2c^2-H'^2/c^2)^{\frac{1}{2}}},$$

or $$\frac{H'^2}{m^2c^4} = 1-\frac{\alpha^2}{j^2}.$$

Since H' must be positive, this leads to

$$\frac{H'}{mc^2} = \frac{\sqrt{(j^2-\alpha^2)}}{|j|}, \tag{55}$$

which is the value of H' given by (54) when $n = 0$. The case $n = 0$

with j negative thus needs further investigation to see whether the conditions (53) are then fulfilled.

With $n = 0$, the maximum value of s is the same as the minimum, so equations (53) with s_0 substituted for s should agree with (52). Now (55) gives, from (44) and (47),

$$\frac{1}{a_1} = \frac{mc}{\hbar}\left(1 - \frac{\sqrt{(j^2-\alpha^2)}}{|j|}\right), \qquad \frac{1}{a} = \frac{mc}{\hbar}\frac{\alpha}{|j|},$$

so the first of equations (53) with s_0 substituted for s gives

$$c_{s_0}\{|j| - \sqrt{(j^2-\alpha^2)}\} + c'_{s_0}\alpha = 0.$$

This agrees with the second of equations (52) only if j is positive. We can conclude that, for $n = 0$, j must be a positive integer, while for the other values of n all non-zero integral values of j are allowed.

73. Theory of the positron

It has been mentioned in § 67 that the wave equation for the electron admits of twice as many solutions as it ought to, half of them referring to states with negative values for the kinetic energy $cp_0 + eA_0$. This difficulty was introduced as soon as we passed from equation (5) to equation (6) and is inherent in any relativistic theory. It occurs also in classical relativistic theory, but is not then serious since, owing to the continuity in the variation of all classical dynamical variables, if the kinetic energy $cp_0 + eA_0$ is initially positive (when it must be greater than or equal to mc^2), it cannot subsequently be negative (when it would have to be less than or equal to $-mc^2$). In the quantum theory, however, discontinuous transitions may take place, so that if the electron is initially in a state of positive kinetic energy it may make a transition to a state of negative kinetic energy. It is therefore no longer permissible simply to ignore the negative-energy states, as one can do in the classical theory.

Let us examine the negative-energy solutions of the equation

$$\left\{\left(p_0 + \frac{e}{c}A_0\right) - \alpha_1\left(p_1 + \frac{e}{c}A_1\right) - \right.$$
$$\left. - \alpha_2\left(p_2 + \frac{e}{c}A_2\right) - \alpha_3\left(p_3 + \frac{e}{c}A_3\right) - \alpha_m mc\right\}\psi = 0 \quad (56)$$

a little more closely. For this purpose it is convenient to use a representation of the α's in which all the elements of the matrices representing α_1, α_2, and α_3 are real and all those of the matrix representing