Recent Developments in the Theory of the Electron

VICTOR F. WEISSKOPF
Massachusetts Institute of Technology, Cambridge, Massachusetts

THE application of the microwave technique to spectroscopy has greatly increased the accuracy of spectroscopic measurements. Recent experiments on the spectrum of hydrogen and other simple elements have revealed that the results are not exactly in agreement with our fundamental theories of the mechanics of the electron within the atom. Small deviations were found in checking the values of the energy levels in hydrogen given by the Sommerfeld formula.[1] The measured value of the magnetic moment of the electron deviated by about 1 in 1000 from the value given by Dirac's fundamental equation of the electron.[2]

These experimental findings led to a reinvestigation of the theory, and especially of its weakest point—the interaction of the electron with radiation. This interaction was treated by a theory named "quantum electrodynamics" which, since its inception by Dirac in 1926, suffered from some internal inconsistencies connected with the old problem of the internal structure of the electron. These inconsistencies make it impossible in this theory to calculate radiation phenomena in a rigorous way.

In the last years, however, some theoretical work has been carried out[3] in an attempt to isolate the unsolved problems and inconsistencies within the theory and to increase the accuracy of the predictions of the theory, in spite of the fact that the structure of the electron and its effects are not understood. The results of this development have been quite successful. The theoretical predictions were in complete agreement with the new experiments. The confidence in the fundamental concepts of "quantum electrodynamics" was greatly enhanced.

It is attempted in this article to present an account of this new development in a form which, I hope, is understandable to the physicist who is not specialized in this field. Only a qualitative and very incomplete picture of the underlying problems can be given. It seemed advisable not to restrict this report to the newest achievements, but to recapitulate shortly the development of our ideas about the electron beginning with H. A. Lorentz's classical electron theory and including the theory of the positron. The significance of the present problems cannot be evaluated without referring to the most important steps in this development.

I. THE CLASSICAL ELECTRON THEORY

There was hardly any other discovery which led to the understanding of so many and varied phenomena as the discovery of the electron.

Many topics which were thought to be unrelated, as optics, electricity, and chemistry, were understood by the same fundamental mechanism on the basis of the electron theory. It was mainly H. A. Lorentz who brought the classical electron theory into a consistent frame. These were his fundamental assumptions: The electron is an elementary particle with a charge e and a mass m; the motion of the electrons is determined by classical mechanics if the force acting on the electron is given by the expression:

$$F=e\mathfrak{E}+(e/c)(\mathbf{v}\times\mathfrak{H}),$$

where e and $\mathbf{v}$ are the charge and velocity of the electron and $\mathfrak{E}$ and $\mathfrak{H}$ are the electric and magnetic field strengths. The electromagnetic field in turn is given by the Maxwell equations

$$(1/c)(\partial\mathfrak{E}/\partial t)-\mathrm{curl}\mathfrak{H}=4\pi\mathbf{i},\quad \mathrm{div}\mathfrak{E}=4\pi\rho;$$
$$(1/c)(\partial\mathfrak{H}/\partial t)+\mathrm{curl}\mathfrak{E}=0,\quad \mathrm{div}\mathfrak{H}=0.$$

The sources of the field strengths are the charge density ρ and a current density $\mathbf{i}$, which are produced by the electrons.

In most cases it is possible to consider the electron as a point charge. The field created by the electron can then be expressed in a simple manner. We quote only one trivial example: an electron at rest is surrounded by an electric field

$$\mathfrak{E}=e/r^2, \qquad (1)$$

[1] W. E. Lamb, Jr., and R. C. Retherford, Phys. Rev. **72**, 241 (1947).

[2] Nafe, Nelson, and Rabi, Phys. Rev. **71**, 914 (1497). Nagel, Julian, and Zacharias, Phys. Rev. **72**, 971 (1947). P. Kusch and H. M. Foley, Phys. Rev. **72**, 1256 (1947).

[3] This development started at a conference of theoretical physicists in June, 1947, on Shelter Island, New York, sponsored by the National Academy of Sciences. The same development has been carried out completely independently by a group of Japanese physicists around Professor Tomonaga. The following papers have been published so far: S. Tomonaga, Prog. Theor. Phys. **1**, 27 (1946); Koba, Tati, and Tomonaga, Prog. Theor. Phys. **2**, 101, 198 (1947); S. Kanesawa and S. Tomonaga, Prog. Theor. Phys. **3**, 1 (1948); S. Tomonaga, Phys. Rev. **74**, 224 (1948); H. A. Bethe, Phys. Rev. **72**, 339 (1947); H. W. Lewis, Phys. Rev. **73**, 173 (1948); H. A. Kramers, Solvay Report, 1948. J. Schwinger, Phys. Rev. **73**, 415 (1948); Phys. Rev. **74**, 1439 (1948); Phys. Rev. **75**, 651 (1949). R. P. Feynman, Phys. Rev. **74**, 939, 1430 (1948). F. J. Dyson, Phys. Rev. **73**, 617 (1948); Phys. Rev. **75**, 486 (1949). N. M. Kroll and W. Lamb, Phys. Rev. **75**, 388(1949); T. Welton, Phys. Rev. **74**, 1157 (1948); J. B. French and V. F. Weisskopf, Phys. Rev. **75**, 1240 (1949).

where r is the distance from the electron. The expressions for the field surrounding an electron in motion are somewhat more complicated.

Some additional assumptions had to be made regarding the conditions under which electrons move in matter: Lorentz assumed that there are several electrons in each atom, which are elastically bound to an equilibrium position and thus are able to perform harmonic vibrations with given frequencies. In electric conductors additional electrons were assumed to move freely about. With these fundamental theoretical tools it was possible to explain a great number of phenomena, as for example, the absorption, scattering, and refraction of light by matter, the Zeeman effect, the optical properties of metals for infra-red radiation and many more. In many cases the explanation was only qualitative. Some of the detailed features were not understood. The main assumption of the elastic binding of electrons within atoms was unexplained, especially in view of the planetary structure of the atom. The frequencies of the electron within the atom were neither understood nor determined by the theory.

Lorentz also investigated another fundamental problem: How far is it possible to consider the electron as a point charge? He was forced to make some assumptions about the internal structure of the electron in order to apply the electrodynamic equations within the electron. We quote from his book *The Theory of the Electron:* "While I am speaking so boldly of what goes on in the interior of an electron, as if I had been able to look into these small particles, I fear one will feel inclined to think I had better not try to enter into all these details. My excuse must be, that one can scarcely refrain from doing so, if one wishes to have a perfectly definite system of equations, moreover, as we shall see later on, our experiments can really teach us something about the dimensions of the electrons. In the second place, it may be observed that in those cases in which the internal state of the electrons can make itself felt, speculations like those we have now entered upon, are at all events interesting, be they right or wrong, whereas they are harmless as soon as we may consider the internal state as a matter of little importance."

The main point of interest in the question of the structure of the electron can be formulated very simply today, since the equivalence of mass and energy has become common place: the total energy E_{st} of the electrostatic field (1) of the electron is given by

$$E_{st}=(1/8\pi)\int \mathcal{E}^2 dv,$$

where the integration is extended over the whole space. $\mathcal{E}$ is given by (1) outside of the electron, but (1) is, of course, no longer valid "inside" the electron; it is convenient to assume that the charge of the electron is concentrated on the surface of a sphere with the radius a. In this case E would vanish inside and we would get:

$$E_{st}=(e^2/2)\int_a^\infty (dr/r^2)=e^2/2a. \qquad (2)$$

Any other assumption as to the charge distribution does not change the general character of (2): the energy of the electric field depends critically upon the radius of the electron. It necessarily contributes to the mass m of the electron, and we obtain from the Einstein relation

$$m=m_0+(E_{st}/c^2)=m_0+(e^2/2c^2a),$$

where m_0 is the "mechanical" mass of the electron, by which we understand all contributions to the mass which are not of electromagnetic origin. Since the total mass m is known experimentally, there is a lower limit for the radius, corresponding to the assumption that all the mass is of electric origin (we exclude the rather artificial choice of a negative value for m_0):

$$a\geq e^2/2mc^2=r_0. \qquad (3)$$

The electron radius is at least as large as r_0 which is usually called "the classical electron radius." Thus we are forced to abandon the notion of an exact point charge.

Lorentz, Abraham, and Poincaré have studied at length the consequences of this new picture. It is not of very great interest to discuss the detailed consequences of the assumption of a finite classical electron. Later developments have brought into the picture new features which completely overshadow these classical considerations.

One point should be mentioned, however. At what energy should one expect in the classical theory the radius of the electron to change significantly the results expected with a point electron? It is easy to see that scattering cross sections of electrons by electrons, or of electrons by equal charges (as protons), should be influenced if the energy is high enough, so that the particles could approach to distances smaller than a. This would happen at energies larger than $2\ mc^2$, that is, larger than one Mev. One may remark that the physicists of that time would have been very much surprised if they had been able to perform these experiments. Instead of finding an effect of the finite extension of the electron, they would have observed the creation of electron-positron pairs. The fundamental connection of the pair creation with the problem of the structure of the electron will be discussed later in this article.

II. THE QUANTUM THEORY OF THE ELECTRON

The problems of the structure of the electron were soon removed from the focus of interest by the successful development of the quantum theory of the electron. The discovery of the quantum of action, Bohr's theory of the quantum orbits in the atom, and the duality of wave and particle properties of the electron, led eventually to the development of quantum mechanics. A new interpretation of the mechanical concepts of momentum, energy, position, and velocity was introduced to describe consistently the facts that appear to be contradictory as, for example, the wave and particle properties of the electron, or the stability against collisions of planetary orbits in atoms. The new theory is best known in the mathematical form of the Schroedinger wave equation.

The success of quantum mechanics was overwhelming. Many unsolved problems of classical electron theory were solved. One can now understand and calculate the resonance frequencies of atoms, the stability of electron orbits, and many other facts which cannot be explained in classical mechanics. There is scarcely any phenomenon within the realm of atoms and molecules which, at least in principle, cannot be accounted for by quantum-mechanical description. It is worth while to point out that the quantum theory of the electron could explain all forces between atoms, molecules and electrons as purely electromagnetic phenomena.

Quantum mechanics can answer all questions as to the behavior of the electron (or other particles) in electromagnetic fields, if these fields are given as functions of space and time. Most of the problems in atomic physics can be put into this form by asking: How does the electron move in an electric or magnetic field of a well defined character? Difficulties do arise, however, if the question, "What fields are created by the moving charges themselves?," is asked. For example, it could not be explained by the theory in this stage that an atom in its ground state does not radiate light, in spite of the fact that charges are in rapid motion.

Nevertheless, it was possible to construct a number of unambiguous rules to calculate the radiation of atomic systems. This was done by means of two principles. One is the *light quantum hypothesis:* light of frequency ν can only be emitted and absorbed in quanta of an energy $h\nu$.† Thus its emission or absorption must be accompanied by a transition from one quantum state to another, whose energy difference is $h\nu$. The other is the *correspondence principle:* Quantum states of very high excitation show the same mechanical properties as one would obtain from a classical calculation of the same problem. Their radiation should then also be equal to the one which is calculated classically. It was possible to derive rules from these two principles with which one could calculate successfully emission, absorption and scattering of light by atomic systems. If the wave-length of light is large compared to the dimensions, a system in a quantum state n is, in many respects, equivalent to an assembly of classical electric oscillators with frequencies given by

$$h\nu_{nk}=(E_n-E_k),$$

where k is some other state of the system. The effective charge ϵ of these oscillators is given by $\epsilon^2=e^2f_{nk}$ where f_{nm} is the so-called oscillator strength:

$$f_{nk}=(2m/h)\nu_{nk}\left|\int\psi_n^*\mathbf{r}\psi_k dv\right|^2, \qquad (4)$$

where the integral represents the matrix element of r between the states n and k.

The problem of the structure of the electron does not enter into this theory. The theory admits the construction of an electronic wave packet with arbitrarily small diameter, even smaller than a if only wave-lengths smaller than a are used. The difficulty arising from the field created by such a packet did not arise since the creation of fields by quantum-mechanical systems was not yet clearly defined.

It is worth while mentioning, however, that it is no longer possible to measure effects of an electron radius $a=r_0$ by having two electrons collide with an energy of the order mc^2. The wave-length corresponding to this energy is h/mc, which is much larger than r_0. Thus it is impossible at that energy to locate the electron better than within h/mc. During a collision their average distance will be h/mc, and they practically never will be within a distance comparable with the radius.

III. THE RELATIVISTIC WAVE EQUATION AND QUANTUM ELECTRODYNAMICS

The quantum theory of the electron needed improvement in two directions: it needed a generalization for high energies in conformity with the theory of relativity and it needed a consistent treatment of the interaction of matter with radiation. It was Dirac who initiated both steps. He was able to devise a wave equation for the electron which fulfilled the relativistic requirements. He made use of the fact that the electron has an intrinsic spin moment whose state, much like the polarization of light, can always be described as a superposition of a spin parallel and opposite a given direction of reference. Thus, the electron wave had to be considered as a "spinor" wave with two components

† Here, and in what follows, we understand by ν the frequency in 2π seconds, and by h the magnitude usually referred to as $\hbar$. $h=1.04\times10^{-27}$ g cm^2 sec.$^{-1}$

corresponding to the two spin directions. Dirac has shown that, for a relativistic wave equation, one has to introduce two more components which, for low velocities, are very much smaller than the others. An electron wave is fully described by giving all four components. Dirac's relativistic wave equation determines the mechanics of this four-component wave. For low kinetic energies (small compared to mc^2) two of the components become very small; the two large ones are themselves solutions of the non-relativistic (Schroedinger) wave equations, each of them corresponding to one of the two directions of the spin.

The non-relativistic theory had to ascribe arbitrarily a magnetic moment μ to the spin, whose value it took from the experimental results. Dirac's relativistic equation contains implicitly an interaction of the spin with a magnetic field. The resulting magnetic moment of the electron $\mu=eh/2mc$ is in almost exact agreement with the experiment.

The relativistic wave equation of the electron exhibits, however, several fundamentally unacceptable features. The equation admits solutions which correspond to states of a particle with negative rest mass. The kinetic energy in these states is negative; the particle moves opposite to the motion in ordinary states. For example: a particle of electronic charge is repelled by the field of a proton. These states are, of course, not realized in nature and the most obvious trouble comes from the fact that their energy is negative and, therefore, below the energy of the actual lowest state with positive rest mass. There should be radiative transitions with the emission of light quanta from the regular states to the irregular ones. No regular state could be stable since there are an infinite number of states of negative energy to which it could go with the emission of a suitable quantum of light.

These states cannot be excluded simply by stating that they do not exist in nature. The regular states alone are not what one calls a complete set of solutions. Physically speaking, if by a certain measurement the electron is put into some arbitrary state, it will very probably be a combination of states containing some of the irregular ones. Especially if an electron is localized in a region smaller than the Compton "wave-length" $\lambda\!\!\!^{-}_c=h/mc$, the states of negative mass will be strongly represented.

We now proceed to Dirac's treatment of the radiation. In order to describe in a consistent way the interaction between matter and radiation, it is necessary to "quantize" not only the motion of the material particles, but also the electromagnetic field. We understand by "quantizing," the consistent application of certain rules, which led from classical mechanics to quantum mechanics. It is relatively simple to apply these rules to the electromagnetic field in an empty space. The field can be decomposed into its "Fourier components;" it can be thought as a superposition of monochromatic waves. Each of these waves has dynamical properties very much like those of a harmonic oscillator. Thus the "quantization" of the electromagnetic field is equivalent to the quantization of a set of harmonic oscillators and, hence, the energy in one monochromatic wave can change only by multiples of $h\nu$. Thus electromagnetic energy of a frequency ν must appear always in portions of the size $h\nu$. This is the light quantum hypothesis. A further important consequence is the zero-point fluctuations: a harmonic oscillator in its state of lowest energy still has a finite amplitude of vibration. Applying this to the electromagnetic field, we conclude that even in the state of lowest energy the electromagnetic vibrations in space are not zero. The state of lowest energy is the state in which *no* light quanta are present. Hence, in this state the mean squares of the field strengths do not vanish.

We now give an estimate of the strength of the field fluctuations averaged over a volume V of linear dimensions a: $V=a^3$. The amplitude B of the zero-point oscillation of an oscillator of frequency ν is given by $B\sim(h/2m\nu)^{\frac{1}{2}}$; it corresponds to a vibration with an energy $h\nu/2$. The main contribution to the field fluctuations in the volume a^3 comes from waves of a wave-length $\lambda\!\!\!^{-}=c/\nu\sim a$.‡ The amplitude should correspond to an energy of $h\nu/2$, one-half light quantum. Now $(1/4\pi)\langle\mathcal{E}^2\rangle_{\text{Av}}a^3$ is the field energy content in a^3; this must be put equal to $h\nu/2=hc/2a$, so that we get approximately

$$\mathcal{E}^2_{\text{fluct.}}\sim hc/a^4. \tag{5}$$

It is larger, the smaller the volume chosen.

The interaction between light and matter can now be described as an interaction between two quantized systems: the electromagnetic field, on one hand, and the electron in the atoms, on the other. Such interaction can be treated by the current methods of quantum mechanics. The interaction energy is given by the classical expression:

$$\int(\mathbf{i}\cdot\mathbf{A})dv,$$

where $\mathbf{i}$ is the current density in the atom and $\mathbf{A}$ is the vector potential in the field. The integral is taken over the space. The two variables i and A are now physical magnitudes, which must be dealt with according to the rules of quantum mechanics. Dirac has shown that by this method absorption, emission, and scattering of light can be calculated and that the result is equal to the one which was obtained by the correspondence principle. The

‡ We use here the term "wave-length" for the length $\lambda\!\!\!^{-}$ which is $1/2\pi$ times the conventional wave-length.

emission of light in a transition from the state n to the state k, for example, in this theory is described in the following way. At a given time, say $t=0$, the emitting atom is in an excited state n and all electromagnetic vibrations are in their ground states. Because of the interaction, the excitation energy E_n-E_k goes over into one of the vibrations; it must, of course, be a vibration whose frequency fulfills the condition $h\nu=E_n-E_k$. The probability P that after a time t the excitation energy has gone into the field turns out to have an exponential time dependence: $P=1-e^{-\Gamma t}$. Γ is then the emission probability per unit time. The value of Γ is given by

$$\Gamma=(2e^2\nu_{nk}^2/3mc^3)f_{nk},$$

in conformity with the probability of radiation of an oscillator with the strength f_{nk}, as defined in (4).

Dirac's quantum electrodynamics gave a more consistent derivation of the results of the correspondence principle, but it also brought about a number of new and serious difficulties. The structure and size of the electron appeared again in the theory. The trouble arose from the interaction with the electron of the zero-point fluctuations of the field. Let us consider a free electron under the influence of an oscillatory field strength $\mathfrak{E}=\mathfrak{E}_0e^{i\nu t}$: it performs forced oscillations of frequency ν with a displacement x_ν. The average square $\langle x_\nu^2\rangle_{Av}$ of this displacement and the average square of the velocity $\langle\dot{x}_\nu^2\rangle_{Av}$ of a free electron are given by

$$\langle x_\nu^2\rangle_{Av}=\tfrac{1}{2}(e^2\mathfrak{E}_0^2/m^2\nu^4),\quad \langle\dot{x}^2\rangle_{Av}=\tfrac{1}{2}(e^2\mathfrak{E}_0^2/m^2\nu^2). \quad (6)$$

The kinetic energy of the electron in these oscillations is

$$E_\nu=\tfrac{1}{2}m\langle\dot{x}_\nu^2\rangle_{Av}=e^2\mathfrak{E}_0^2\lambda^2/4mc^2, \quad (6a)$$

where λ is the wave-length belonging to the frequency ν. Hence, the zero-point oscillations of the field contribute to the electron a certain amount of energy. Let us assume for a moment that the electron is a sphere with a radius a. Then only waves with a wave-length $\lambda>a$ will act upon the electron; the ones with $\lambda\gg a$ are not very important, so that we are allowed to put in (6a) $\lambda=a$. If we then enter the value (5) for $\langle\mathfrak{E}_0^2\rangle_{Av}$ over a volume a^3, we obtain for the energy E_{fl} of the electron due to the zero-point field fluctuations:

$$E_{fl}\sim e^2h/4mca^2. \quad (7)$$

In a more accurate calculation we start with expression (6), which gives the effects on the electron induced by an oscillatory field strength of amplitude $\mathfrak{E}_0$ and frequency ν. In order to calculate the value of $\mathfrak{E}_0^2$ for the zero-point oscillations, we include the electromagnetic field and the electron into a big volume Ω. The zero-point amplitude $\mathfrak{E}_0e^{i\nu t}$ of one proper vibration can be calculated by putting the total energy of the oscillation equal to:

$$(1/8\pi)\int(\mathfrak{E}^2+\mathfrak{H}^2)d\nu=(1/8\pi)\mathfrak{E}_0^2\Omega$$
$$=h\nu/2;\quad \mathfrak{E}_0^2=4\pi h\nu/\Omega.$$

We use the well-known formula that there are

$$z(\nu)d\nu=\Omega(\nu^2/\pi^2c^3)d\nu$$

proper vibrations in the frequency interval $d\nu$. Since the zero-point oscillations of different frequencies are statistically independent, their contributions to the average square of the displacement and of the velocity add up and we get for the total of these magnitudes:

$$\langle x^2\rangle_{Av}=\int\langle x_\nu^2\rangle_{Av}z(\nu)d\nu=(2e^2h/\pi m^2c^3)\int_{\nu_0}^{\infty}(d\nu/\nu), \quad (8)$$

$$\langle\dot{x}^2\rangle_{Av}=(2e^2h/\pi m^2c^3)\int_{\nu_0}^{\infty}\nu d\nu. \quad (9)$$

The integrals are extended between a lower limit ν_0 and infinity. The frequency ν_0 depends on the state of binding of the electron. $h\nu_0$ is of the order of the binding energy. If the frequency of the field oscillations falls below the frequency ν_0, the electron can no longer be considered as free and (6) is no longer valid. The resulting effect is equivalent to an omission of the frequencies below ν_0.

Both expressions (8) and (9) lead to infinite results. This is especially troublesome in the case of the velocity square because it gives rise also to an infinite kinetic energy E_{fl} of the electron due to the zero-point fluctuations:

$$E_{fl}=(m/2)\langle\dot{x}^2\rangle_{Av}=(e^2h/\pi mc^3)\int_{\nu_0}^{\infty}\nu d\nu. \quad (10)$$

This expression contains a quadratically divergent integral. Since this energy is an inseparable part of the total energy of an electron, it must appear as part of its mass energy mc^2. In order to keep the mass finite, one therefore is forced to assume some structural properties of the electron which prevent the interaction with high frequencies of the field. We can do this by introducing an upper limit ν_{max} to the interaction which cuts off the integral in (10) at that limit. The fluctuation energy assumes the form

$$E_{fl}=(e^2h/2\pi mc^3)\nu_{max}^2, \quad (7a)$$

and we can determine an upper bound for ν_{max} by setting E_{fl} equal to mc^2:

$$h\nu_{max}\leq(2\pi hc/e^2)^{\frac{1}{2}}mc^2\approx 15\text{ Mev}. \quad (7b)$$

This would remove the interaction with an electron

at rest of a quantum of an energy >15 Mev, a rather improbable result. The introduction of $\nu_{\max}$ is equivalent to the assumption of an electron radius $a=c/\nu_{\max}$, which shows the equivalence of (7) and (7a). Equation (7b) gives rise to a value of $a\approx(hc/e^2)^{\frac{1}{3}}r_0$, which is larger than the classical limit (3). Thus the fluctuation energy seemingly pushes the electron radius to even greater values than the one which we obtained from the energy of the electrostatic field. It should be noted, however, that in interactions with light of an energy of more than $2mc^2$, the irregular solutions with negative mass play an essential role. Thus the significance of these states will have an essential bearing upon the problem of the self-energy of the electron.

Dirac's two generalizations of quantum mechanics, the relativistic wave equation and the quantum electrodynamics, were very successful in some respects: the explanation of the magnetic moment of the electron, the derivation of the Sommerfeld fine structure formula, and the consistent derivation of the expressions for the absorption, emission, and scattering of light. Two fundamental difficulties were introduced simultaneously:

(1) The existence of states of the electron of negative mass. They cause an instability of a normal bound state by the emission of a quantum of high energy and subsequent transitions into a state of negative mass. Thus, the "normal" states of the electron have a very strong "resonance" interaction with light quanta of high energy.

(2) The quantization of the electromagnetic field introduces infinite fluctuations of the electron. In order to keep their contribution to the energy within the observed mass energy value, the interaction of the electron with light quanta of an energy $h\nu>(137mc^2)^{\frac{1}{2}}$ would have to be basically altered. It will be shown in the next section that the positron theory removed the first difficulty and completely changes the aspect of the second.

IV. THE POSITRON THEORY

The phenomenon of creation of a positron and an electron by a light quantum introduces a new aspect into the theory of the electron. The fundamental process can be described as follows: a light quantum of an energy larger than $2mc^2$ (1 Mev) can be absorbed by the empty space, in the presence of strong electric fields. The energy is then transformed into a pair consisting of a positive and a negative electron.

Two outstanding facts are shown in this phenomenon: the existence of a positive electron, and the fact that the vacuum has physical properties, which enables it to absorb light and to produce electrons. Hence, the physical description of the vacuum is bound to be more complicated than hitherto and must contain the latent electron pairs which can be created.

It was again Dirac who, turning a vice into a virtue, used the unacceptable states of negative mass for the description of the vacuum. A reinterpretation of these states gives an almost perfect description of the vacuum and the existence of positrons: the states of negative mass correspond in some respects to the states of a particle of opposite charge since they move in opposite directions in any electromagnetic field. They are, however, still unacceptable because of their negative kinetic energy. The reinterpretation which removes this difficulty can be formulated as follows: According to the Pauli exclusion principle, any state can be either occupied by one single electron, or unoccupied. The occupation of a state of energy E_i increases the total energy of the system by the amount E_i, the removal of an electron from the state decreases the total energy by E_i. Dirac's reinterpretation of the states of negative mass consists in the exchange of "occupation" and "removal." We decide to call an occupied state of negative mass "empty" and an empty state "occupied." The transition from "empty" to "occupied" is then connected with an energy change of $-E_i$. Since E_i is negative itself, the energy actually increases by $+|E_i|$. The trouble with the negative energy is thus removed.

The vacuum can then be described formally by assuming that all states of negative mass are occupied by electrons. They are not "actually" occupied, because of our reinterpretation, so that one need not be bothered by the infinite charge density which one would get if all states of negative mass were really occupied. The wave functions, which represent the *absence* of positrons are the same functions which would have represented the *presence* of electrons of negative mass. It is a new feature that the "absence" of a particle is described by a wave function. This is, however, an expression of the fact that the vacuum has the physical properties described above; it is filled with latent electrons.

This reinterpretation removes at once the difficulty which the states of negative mass have introduced. Since in the vacuum these states are occupied, no electron in the regular states can jump into them. Thus the regular states are no longer unstable against decay into the irregular ones. They no longer are in "resonance" interaction with arbitrarily high light quanta.

The pair creation is then described as follows: a light quantum produces a transition from an occupied state of negative mass to a state of positive mass. The result is an electron in a state of positive mass and an unoccupied state of negative mass. The latter must be interpreted as an occupied state

of a positron with positive mass. Thus the light quantum has created two particles positive and negative with positive mass.

Such transition can only occur in the presence of external fields. Without those fields energy and momentum cannot be conserved. The transition probability can be calculated and the results reproduce excellently the experimental material. The opposite process is the annihilation of a positive and a negative electron, with the emission of either one quantum in an electric field or of two quanta in the field-free space. It can be described by our picture as the transition of the electron into the "unoccupied" state by which the positron is represented. This transition is accompanied by the emission of light quanta.

The new aspect of the vacuum has a decisive effect upon the problem of the self-energy of the electron. The properties of the vacuum with respect to the electrons are now, in some aspects, analogous to its properties in respect to the electromagnetic field. There exist also zero-point fluctuations of the electric charge and the electric current in the vacuum. These fluctuations are very small when averaged over a volume of a size larger than the Compton "wave-length" $\lambda\!\!\!^{-}_c = h/mc$. They represent the latent electron pairs which, by means of light quanta, could be brought into real existence.

Let us now consider the properties of the "vacuum" in the neighborhood of an actual electron. There will be an interaction between this electron and the latent charges, mainly because of the Pauli exclusion principle. According to this principle, electrons tend to keep distance from one another. Two electrons (of equal spin) do not come nearer than a distance d which is determined by their relative momentum p: $d \sim h/p$. (They must not be in the same cell of the phase space.) The presence of one actual electron in the vacuum introduces some changes in the "charge distribution" of the vacuum. This charge distribution would be zero on the average if undisturbed. The wave functions which represent the electrons of negative mass are slightly removed from the place of the actual electron. This change of charge distribution, compared to the undisturbed vacuum, appears as an addition to the "actual" electron. This manifests itself in form of a spread in the charge distribution of an electron, since the vacuum electrons are slightly pushed away from the actual electron. The calculation shows that this spread is enough to change the classical electrostatic self-energy to $(e^2/hc)mc^2 \log(\lambda\!\!\!^{-}_c/a)$. Here a is the "radius" of the electron, or, as a better definition, a is a limit of wave-length so that fields with $\lambda\!\!\!^{-} < a$ are no longer assumed to interact with the electron.

The effects of the zero-point field oscillations are even more drastically changed by our new concept of the vacuum. This comes from the fact that the field oscillations also interact with the latent electron pairs in the vacuum. As long as their frequency is much smaller than $2mc^2/h$ (the minimum frequency of pair creation), the "vacuum" is very little influenced and the old calculation (6) of the displacement $\langle x_\nu^2\rangle_{Av}$ and the velocity $\langle \dot{x}_\nu^2\rangle_{Av}$ are still valid. For frequencies higher than $2mc^2/h$, however, the field oscillations have a strong effect on the latent electron pairs and the induced charge and current fluctuations in the vacuum interfere with the induced fluctuation of the electron itself. This interference is destructive and reduces to some extent values of the induced displacement and velocity. The reduction can be roughly approximated in its main features by a factor $(mc^2/h\nu)^2$ to the expressions (6) for $h\nu > 2mc^2$:

$$\left.\begin{aligned}\langle x_\nu^2\rangle_{Av} &= \tfrac{1}{2}(e^2\mathfrak{E}_0^2/m\nu^4)(mc^2/h\nu)^2\\ \langle \dot{x}_\nu^2\rangle_{Av} &= \tfrac{1}{2}(e^2\mathfrak{E}_0^2/m\nu^2)(mc^2/h\nu)^2\end{aligned}\right\} \text{for } h\nu > 2mc^2. \quad (6')$$

This effect is difficult to explain in qualitative language. It is connected with the Pauli exclusion principle, according to which electron have a tendency to keep apart from one another. Thus the charge and current fluctuations of the vacuum in the neighborhood of the electron tend to be in opposite phase to the fluctuations of the electron itself and therefore cause the destructive interference.

These effects represent a definite improvement. The average displacement $\langle x^2\rangle_{Av}$ does no longer lead to infinities. The divergent integral in (8) converges now because of the reduced contribution (6′) of the frequencies above $2mc^2/h$, and we obtain

$$\langle x^2\rangle_{Av} = (2e^2h/\pi m^2c^3)\log(fmc^2/h\nu_0), \quad (11)$$

where f is a factor of the order unity, which can be determined if the effect of the higher frequencies is exactly taken into account. The average velocity square (9) is still infinite but the divergence is only logarithmic. We get from (6′):

$$\langle \dot{x}^2\rangle_{Av} \sim (2e^2h/\pi m^2c^3)\left[\int_0^{2mc^2/h} \nu d\nu + (mc^2/h)^2\int_{2mc^2/h}^{\infty} (d\nu/\nu)\right].$$

The fluctuation energy $E_{fl} = (m/2)\langle \dot{x}^2\rangle_{Av}$ is reduced to $E_{fl} = (e^2/\pi hc)mc^2 \log(fh\nu_{\max}/mc^2)$ where f is a numerical factor and $\nu_{\max}$ the cut-off frequency. In order to keep this energy below the total mass energy mc^2 of the electron, it is now sufficient to keep $a = c/\nu_{\max}$ larger than $(h/mc)\exp[-(\hbar c/e^2)]$. This lower limit is very much smaller than any length considered so far. It is no longer necessary to tamper with the interaction of the electron with light quanta of an energy of a few Mev. It is still

unsatisfactory, of course, that the limit cannot be chosen to be infinity without obtaining infinite self-energies; thus the internal structure of the electron will appear somewhere in the theory. However, some changes in the interaction between light and matter are certain to occur at very high energy values where we have good reason to expect the appearance of new phenomena (nuclear or meson type).

So far we have discussed the influence of an actual electron on the vacuum due to the Pauli-exclusion principle. There is also an influence, although weaker, in the form of a displacement of the vacuum electrons due to electric interaction. It is easier to describe this effect, not for an actual electron, but for a *proton*, which is embedded in the vacuum. The wave functions of the states of negative mass are all deformed because of the presence of the proton. Since the vacuum is described by the undeformed states, the difference between the deformed and undeformed ones should give rise to an actual charge density. This is called the polarization of the vacuum by an external charge (the proton).

The proton induces a charge density ρ_i in the vacuum. The calculation shows that $\rho_i(\mathbf{r})$ as function of the location $\mathbf{r}$ has the following form:

$$\rho_i(\mathbf{r}) = A\rho_0(\mathbf{r}) + \int G(\mathbf{r}-\mathbf{r}')\rho_0(\mathbf{r}')d\mathbf{r}'. \qquad (12)$$

Here $\rho_0(\mathbf{r})$ is the external charge density; in our case, ρ_0 is the charge density of the proton. A is a constant and $G(\mathbf{r}-\mathbf{r}')$ is a function of the distance between the points $\mathbf{r}$ and $\mathbf{r}'$. The integral is extended over all points $\mathbf{r}'$. The expression for the induced charge consists of two parts: the first term is exactly proportional to the inducing charge density ρ_0; the second part is an effect at a distance. According to this term a point charge at $r=0$ (like a proton) would give rise to a charge distribution $G(r)$. $G(r)$ is different from zero only over distances up to the Compton wave-length λ_c. The effect is the same as if the dielectric coefficient of the vacuum was different from unity by about 1/137 over a region of the order λ_c. It is important to note that the first part is unobservable in principle. Its effect is undistinguishable from the original charge density ρ_0, since it is always induced by it. What is actually measured in nature as the charge of the proton would not be e, but $(1+A)e$. It thus represents nothing but a renormalization of the charge. The second term only has physical significance.

There is one serious difficulty with this interpretation: the factor A turns out to be logarithmically infinite: $A \sim (e^2/hc)\log(\lambda_c/a)$ if the "cut-off" radius a is put equal to zero. This would mean that the external charge ρ_0 of the proton induces a charge in the vacuum at the same place, which changes its value by an infinite amount. It is true that this change is in itself unobservable, since one always observes the total charge, external plus induced, in nature. However, the fact that the induced charge is infinite for $a=0$ represents a serious difficulty of the theory.

The vacuum is polarized not only by a proton but also by an electron. The situation is somewhat more complicated in this case because of exchange phenomena between the electron and the vacuum electrons. The fact remains, however, that the electron, if considered as a point ($a=0$), also induces a charge in the vacuum which adds an infinite contribution to its original charge. Thus the internal structure of the electron is relevant not only for its mass but also for its charge.

One can make these infinite additions finite without changing the second term in (12) by arbitrarily removing the interaction of the field with electrons whose wave-length is smaller than a. Here, as in the self-energy, the infinity comes from the interaction at very high energies, and there is hope that a future theory will change this interaction so that the constant A remains finite and small.

In spite of these difficulties, the theory of the positron can be regarded as a big step forward in our understanding of the electron: By means of Dirac's reinterpretation of the states of negative mass it was possible to explain the new phenomena of pair creation and annihilation and to remove several fundamental difficulties of the Dirac equation:

(1) The radiative transitions from the ordinary states into states of negative mass are removed.

(2) The fluctuation energy is much less sensitive to the structure of the electron because of its logarithmic dependence on the electron radius.

(3) The average square displacement of the electron by the field fluctuations is finite and independent of the radius or the structure.

V. THE EXPERIMENTAL TEST OF QUANTUM ELECTRODYNAMICS

The quantization of the electromagnetic field so far has not brought much reward. It is true that it made it possible to derive the expressions for the absorption, emission, and scattering of light, which before were based only upon a recipe contrived by means of the correspondence principle. On the other hand, new difficulties came about, all connected with the zero-point oscillations of the electromagnetic field and their effect on the self-energy of the electron. Quantum electrodynamics has not yet shown its superiority over the correspondence principle. On the contrary, its actual expressions for the electromagnetic phenomena be-

come senseless, since a consistent interpretation of the theory would force us to put the mass m of the electron equal to infinity at all places where it occurs.

Encouraged by some new experiments, which will be discussed later on, a new attempt was made recently to find observable effects, which are directly connected with the new features introduced by quantum electrodynamics. The main theoretical difficulty consisted in the problem of how to separate the infinities of mass and charge from the rest of the theory, in order to obtain results that can be applied to nature. This was done by isolating the expressions for the infinite mass and charge within the theory, in the hope that mass and charge will be made finite by a future improvement. Such procedure is possible since the self-energy terms and the infinite charge come mostly from the interaction with very high energy light quanta and are, therefore, largely independent of the state of binding of the electron in fields normally occurring in nature. Hence, they can be split off as an additional mass and charge of the electron. This has been shown already for the charge in the last section by discussing expression (12). The separation of the mass term is mathematically much more complicated but can be performed in an analogous way. The relativistic transformation properties of the terms occurring in the calculation proved to be of great importance for finding an unambiguous rule as to what parts of the expression of the self-energy can be considered as a mass term. It was necessary to reformulate quantum electrodynamics so that the relativistic invariance of the theory was more explicit than before. This very laborious task was performed by J. Schwinger and independently by S. Tomanaga.

There is, however, a small part of the self-energy which is not contained in the mass and which is due to the interaction with oscillations of lower frequencies. This part depends on the external conditions and may give rise to a slight shift of energy levels, depending on the conditions of binding, and a slight change in some of the fundamental properties of the electron. It is due mainly to the effect of the displacement x of the electron by the zero-point oscillations, whose square average $\langle x^2\rangle_{\text{Av}}$ turned out to be finite and due entirely to the interaction with lower frequencies. This can be demonstrated by means of quite elementary calculations[4] in a case which corresponds to an actual experiment, namely, the shift of the levels in hydrogen-like atoms.

Let us consider a stationary state n of the electron in a Coulomb field, whose wave function is given by ψ_n. The Coulomb field is described by the potential energy $V(r)=Ze^2/r$, where r is the distance from the nucleus. The average potential energy $\bar{V}$ in the state n can be written in the form

$$\bar{V}=\int V(r)\,|\psi_n(\mathbf{r})|^2 dv, \tag{13}$$

where $|\psi_n(\mathbf{r})|^2$ is the well-known probability of finding the electron at a point $\mathbf{r}$: the integral is extended over the volume. This expression must be changed in view of the existence of the zero-point oscillations. The effect of these oscillations on the electromagnetic mass is already assumed to be contained in the observed electron mass m. There is, however, also an influence on the potential energy, since the electron is forced to oscillate around the position $\mathbf{r}$. It will be shown that this oscillation changes the average value of the potential energy by a small amount. This change gives rise to a shift of the energy levels.

In order to calculate this change we replace $V(\mathbf{r})$ in (13) by $V(\mathbf{r}+\mathbf{x})$, where $\mathbf{x}$ is the zero-point oscillation of the electron. We use a Taylor expansion because of the smallness of $\mathbf{x}$.*

$$V(\mathbf{r}+\mathbf{x})=V(\mathbf{r})+\text{grad}\,V\cdot\mathbf{x}+\tfrac{1}{2}\Delta V\cdot(\mathbf{x}^2/3), \tag{14}$$

where ΔV is the Laplace operation on V: $\Delta V=[(\partial^2/\partial x^2)+(\partial^2/\partial y^2)+(\partial^2/\partial z^2)]V$. The second term is zero in the average, since $\mathbf{x}$ is an oscillation. Thus the addition δE_n to the average potential energy of the state n may be written:

$$\delta E_n=\frac{1}{6}\int \Delta V\cdot\langle x^2\rangle_{\text{Av}}\,|\psi_n(\mathbf{r})|^2 dv.$$

The Laplacian of the Coulomb potential is proportional to the charge density ρ_0 which produces it: $\Delta V=4\pi e\rho_0$, ρ_0 is the charge density of the nucleus, which we approximate by a δ-function:** $\rho_0=Ze\delta(r)$, where Ze is the charge of the nucleus. Hence we obtain for δE_n:

$$\delta E_n=(2\pi/3)Ze^2\,|\psi_n(0)|^2\langle x^2\rangle_{\text{Av}}, \tag{15}$$

where $|\psi_n(0)|^2$ is the intensity of the wave function at the nucleus, and we insert the value (11) which we found for $\langle x^2\rangle_{\text{Av}}$ into (15) to calculate the level shift. The frequency ν_0 which occurs in (11) depends on the binding of the electron and is of the order of the Rydberg frequency ν_R for an electron in a hydrogen-like atom. Since ν_0 appears only under a logarithm, its exact value is not of great importance. It has been shown by Bethe[3] that, for

[4] We are following here a calculation outlined by T. Welton, Phys. Rev. **74**, 1157 (1948).

* The simple form of the third term in (14) comes from the fact that, in the average: $\langle x_x x_y\rangle_{\text{Av}}=0$, $x_x^2=x_y^2=x_z^2=(\mathbf{x})^2/3$.

** The δ-function $\delta(r)$ is zero everywhere except at $r=0$. It is normalized such that the volume integral $\int\delta(r)dv$ is equal to unity.

a quantum state $\mathbf{n}$, ν_0 is given by the formula

$$\log h\nu_0=\frac{\sum_m |p_{nm}|^2(E_m-E_n)\log|(E_m-E_n)|}{\sum_m |p_{nm}|^2(E_m-E_n).},$$

where E_n is the energy of the state n and the sums are extended over all other quantum states m. p_{nm} is the matrix element of the momentum between the states n and m.

We observe that $|\psi(0)|^2$ vanishes for all states except S states (states with the orbital angular momentum zero), for which simple relation holds:

$$|\psi_n(0)|^2=Z^3|\pi l^3 n^3, \tag{16}$$

where $l=h^2/me^2$ is the Bohr radius. Thus the level shift vanishes for states with an angular momentum different from zero.

We finally get the level shift for S states from (15), (11), and (16). It is practical to express it in form of a relative shift by dividing δE_n by the energy E_n of the level which is given by the Balmer relation $E_n=Z^2me^4/2h^2n^2$:

$$\delta E_n/E_n=(8/3\pi)(e^2/hc)^3(Z^2/n)\log(fmc^2/h\nu_0). \tag{17}$$

The exact calculation for the $2S_{\frac{1}{2}}$ term yields the values $\nu_0=18\nu_R$, $f=1.3$. Thus the S levels of hydrogen-like atoms should be shifted upwards (δE_n is positive) by small amounts, relative to the values given by the Sommerfeld formula. This is a direct effect of the zero-point oscillation and its experimental verification constitutes a strong support of quantum electrodynamics.

The polarization of the vacuum by the proton produces also a shift which has to be added to (16). According to the discussions of the last section, the only observable effect is a small polarization around the proton of the extension λ_c. The calculation shows that this causes a line shift $\delta E_n'$:

$$\delta E_n'/E_n=-(8/15\pi)(e^2/hc)^3Z^2/n. \tag{18}$$

It amounts to only about 1/40 of the shift δE_n.

The most reliable experiment on the lineshift was performed by Lamb and Retherford[1] on hydrogen. According to the Sommerfeld formula the $2S_{\frac{1}{2}}$ level and the $2P_{\frac{1}{2}}$ level of the hydrogen atom should coincide in energy, and the $2P_{\frac{3}{2}}$ level should lie 10,000 megacycles higher. Lamb and Retherford have measured the $2S_{\frac{1}{2}}$ level relative to the two other levels and have found that the $2S_{\frac{1}{2}}$ level is shifted upwards by about 1060 mc, a value which is in good agreement with the theoretical formula (17). Similar shifts have been found by J. Mack[5] and Kopfermann and Paul[6] in helium. The present measurements are not accurate enough to prove the existence of shifts as small as the one given by (18), caused by the polarization of the vacuum. Future experiments will show whether this additional effect can be considered as real.

Another important result obtained by these methods is the correction to the g factor of the electron. According to Dirac's equation, the magnetic moment of the electron μ_e is equal to $he/2mc$. The ratio between this value and the mechanical moment $\hbar/2$ of the electron is $g(e/2mc)$ with $g=2$, in contrast to the value of this ratio for orbital motions in which $g=1$. If the interaction of the electron with the radiation field is properly taken into account, one obtains the result that g is not accurately equal to 2 but $g=2+e^2/\pi hc$.

Unfortunately, it is impossible to give a qualitative description of this effect along the lines in which the level shift was explained. The spin of the electron is in itself a phenomenon which is not amenable to a simple pictorial understanding. A way to understand the effect may be found by remembering that the magnetic moment of the Dirac electron is due to circular currents of the radius h/mc. The zero-point oscillations of the electromagnetic field influence these currents to a certain extent, and so do the current fluctuations induced in the "vacuum." These interactions cause the slight change of the magnetic moment. The numerical result is in excellent agreement with recent experimental measurements.[2] The magnetic moment of the electron was determined with great accuracy from the Zeeman effect of some fine structure doublets. Although the correction of the g factor cannot be understood in simple terms, it represents the most important result of quantum electrodynamics since it deals with one of the fundamental properties of the free electron—its magnetic moment.

The great success in these two instances of the quantum-electrodynamical concepts proves that the fundamental ideas must contain a great deal of truth. The main achievement of the recent development consisted in finding an unambiguous and relativistically invariant way of separating those effects of the interaction between light and electron which can be interpreted as additional mass and charge, from the other effects which give rise to observable phenomena. The additional mass and charge are contained in the observed values of m and e and can never be observed independently. It must not be forgotten, however, that these magnitudes are still infinite in this theory. This constitutes a warning that the interaction of the electron with light quanta of very high energy is not yet understood. Somewhere at very high energies, the internal structure of the electron must play an essential role in a future theory in a way which is completely unknown. This structure appears at present in the form of the arbitrary length a which

[5] J. Mack, Phys. Rev. **73**, 1233 (1948).
[6] Kopfermann and Paul, Naturwiss. (1948).

we have introduced as a radius of the electron in order to make the mass and the charge of the electron finite magnitudes.

The importance of the recent developments lies in the recognition of the following fact: for problems dealing with atomic energies only mass and charge of the electron are "structure dependent" (meaning dependent on the value of a and going to infinity if a is chosen zero), whereas all other effects, such as scattering cross sections, energy levels, magnetic moments, etc., can be calculated without making any assumption regarding the structure of the electron.

There is perhaps some significance in the fact that the theory of the electron cannot be brought into a completely satisfactory form without introducing some new elements into the theory at high energies. It cannot be a pure accident that the charge of the proton and of the meson is equal to the electronic charge, or that the classical electron radius r_0 is almost equal to the range of nuclear forces. There must be a connection between quantum electrodynamics and the future theory of mesons and of the nuclear forces, which at present exists only in very rudimentary form. The tie between these theories should be of importance for the electron only at energies of the order of the meson rest mass or higher. This would be high enough (>100 Mev) to leave unchanged the results of the theory for atomic energies. One may hope that the understanding of this tie will solve the problem of the electromagnetic mass and of the induced charge of the electron.

In discussing the classical electron theory, we remarked that a scattering experiment testing the limits of the classical theory would have revealed the existence of positrons, a phenomenon which was of fundamental significance for the further development of the theory. An experiment trying to test the present theory at high energies (100 Mev and over) will probably give rise to meson production. This is perhaps an indication of the important role of the mesons in a future theory of the electron. Future experiments with the new accelerating machines which are now under construction will reach energies of these critical values. It is hoped that the phenomena found by means of these new tools will shed new light upon the fundamental problem of the relation between elementary particles.