

# Development of Quantum Electrodynamics

## Personal recollections

Sin-Itiro Tomonaga - Nobel Lecture

May 6, 1966

(1) In 1932, when I started my research career as an assistant to Nishina, Dirac published a paper in the Proceedings of the Royal Society, London[1]. In this paper, he discussed the formulation of relativistic quantum mechanics, especially that of electrons interacting with the electromagnetic field. At that time a comprehensive theory of this interaction had been formally completed by Heisenberg and Pauli<sup>2</sup>, but Dirac was not satisfied with this theory and tried to construct a new theory from a different point of view. Heisenberg and Pauli regarded the (electromagnetic) field itself as a dynamical system amenable to the Hamiltonian treatment; its interaction with particles could be described by an interaction energy, so that the usual method of Hamiltonian quantum mechanics could be applied. On the other hand, Dirac thought that the field and the particles should play essentially different roles. That is to say, according to him, "the role of the field is to provide a means for making observations of a system of particles" and therefore "we cannot suppose the field to be a dynamical system on the same footing as the particles and thus be something to be observed in the same way as the particles".

Based on such a philosophy, Dirac proposed a new theory, the so-called many-time theory, which, besides being a concrete example of his philosophy was of much more satisfactory and beautiful form than other theories presented up to then. In fact, from the relativistic point of view, these other theories had a common defect which was inherent in their Hamiltonian formalism. The Hamiltonian dynamics was developed on the basis of non-relativistic concepts which make a sharp distinction between time and space. It formulates a physical law by describing how the state of a dynamical system changes with time. Speaking quantum-mechanically, it is a formalism to describe how the probability amplitude changes with time  $t$ . Now, as an example, let us consider a system composed of  $N$  particles, and let the coordinates of each particle be  $r_1, r_2, \dots, r_N$ . Then the probability amplitude of the system is a function of the  $N$  variables  $r_1, r_2, \dots, r_N$ , and in addition, of the time  $t$  to which the amplitude is referred. Thus this function contains only one time variable in contrast to  $N$  space variables. In the theory of relativity, however, time and space must be treated on an entirely equal footing so that the above imbalance is not satisfactory. On the other hand, in Dirac's theory which does not use the Hamiltonian formalism, it

becomes possible to consider different time variables for each particle, so that the probability amplitude can be expressed as a function of  $r_1, t_1, r_2, t_2, \dots, r_N, t_N$ . Accordingly, the theory satisfies the requirement of the principle of relativity that time and space be treated with complete equality. The reason why the theory is called the many-time theory is because  $N$  distinct time variables are used in this way.

This paper of Dirac's attracted my interest because of the novelty of its philosophy and the beauty of its form. Nishina also showed a great interest in this paper and suggested that I investigate the possibility of predicting some new phenomena by this theory. Then I started computations to see whether the Klein-Nishina formula could be derived from this theory or whether any modification of the formula might result. I found out immediately however, without performing the calculation through to the end, that it would yield the same answer as the previous theory. This new theory of Dirac's was in fact mathematically equivalent to the older Heisenberg-Pauli theory, and I realized during the calculation that one could pass from one to the other by a unitary transformation. The equivalence of these two theories was also discovered by Rosenfeld[3] and by Dirac-Fock-Podolsky[4] and was soon published in their papers.

Though Dirac's many-time formalism turned out to be equivalent to the Heisenberg-Pauli theory, it had the advantage that it gave us the possibility of generalizing the former interpretation of the probability amplitude. Namely, while one could calculate the probability of finding particles at points with coordinates  $r_1, r_2, \dots, r_N$ , all at the time  $t$  according to the previous theory, one could now compute more generally the probability that the first particle is at  $r_1$  at time  $t_1$ , the second at  $r_2$  at time  $t_2$ , and the  $N$ -th at  $r_N$  at time  $t_N$ . This was first discussed by Bloch[5] in 1934.

(2) In this many-time theory developed by Dirac, electrons were treated according to the particle picture. Alternatively, in quantum theory, any particle should be able to be treated according to the wave picture. As a matter of fact, electrons were also treated as waves in the Heisenberg-Pauli theory, and it was well known that this wave treatment was frequently more convenient than the particle treatment. So the question arose as to whether one could reformulate the Heisenberg-Pauli theory in a way which would be more satisfactory relativistically, when electrons were treated as waves as well as the electromagnetic field.

As Dirac already pointed out, the Heisenberg-Pauli theory is built upon the Hamiltonian formalism and therefore the probability amplitude contains only one time variable. That is to say, the probability amplitude is given as a function of the field strength at different space points and of one common time variable. However, the concept of a common time at different space points does not have a relativistically covariant meaning.

Around 1942, Yukawa[6] wrote a paper emphasizing this unsatisfactory aspect of the quantum field theory. He thought it necessary to use the idea of the g.t.f. (generalized transformation function) proposed by Dirac[7] to correct this defect of the theory. Here I shall omit talking about the g.t.f., but, briefly,

Yukawa's idea was to introduce as the basis of a new theory a concept which generalized the conventional conception of the probability amplitude. However, as pointed out also by Yukawa, we encounter the difficulty that, in doing this, cause and effect can not be clearly separated from each other. According to Yukawa, the inseparability of cause and effect would be an essential feature of quantum field theory, and without abandoning the causal way of thinking which strictly separates cause and effect, it would not be possible to solve various difficulties appearing in quantum field theory about which I will talk later. I thought however, that it might be possible (without introducing such a drastic change as Yukawa and Dirac tried to do) to remedy the unsatisfactory, unpleasant aspect of the Heisenberg-Pauli theory of having a common time at different space points. In other words, it should be possible, I thought, to define a relativistically meaningful probability amplitude which would be manifestly covariant, without being forced to give up the causal way of thinking. In having this expectation I was recalling Dirac's many-time theory which had enchanted me 10 years before.

When there are  $N$  particles in Dirac's many-time theory, we assign a time  $t_1$  to the first particle,  $t_2$  to the second, and so on, thus introducing  $N$  different times,  $t_1, t_2, \dots, t_N$ , instead of the one common time  $t$ . Similarly, I tried in quantum field theory to see whether it was possible to assign different times, instead of one common time, to each space point. And in fact I was able to show that this was possible[8].

As there are an infinite number of space points in field theory in contrast to the finite number of particles in particle theory, the number of time variables appearing in the probability amplitude became infinite. But it turned out that no essential difficulty appeared. An interpretation quite analogous to the one discussed by Bloch in connection with Dirac's many-time theory could be given to our probability amplitude containing an infinite number of time variables. Further, it was found that the theory thus formulated was completely covariant and that this covariant formulation was equivalent in its whole content to the Heisenberg-Pauli theory: it was shown, just as in the case of the many-time theory, that we could pass from one to the other by a unitary transformation. I began this work about 1942, and completed it in 1946.

(3) As I mentioned a little while ago, there are many difficulties in the quantum mechanics of fields. In particular, infinite quantities always arise which are associated with the presence of field reactions in various processes. The first phenomenon which attracted our attention as a manifestation of field reactions was the electromagnetic mass of the electron. The electron, having a charge, produces an electromagnetic field around itself. In turn, this field, the so-called self-field of the electron, interacts with the electron. We call this interaction the field reaction. Because of the field reaction the apparent mass of the electron differs from the original mass. The excess mass due to this field reaction is called the electromagnetic mass of the electron and the experimentally observed mass is the sum of the original mass and this electromagnetic mass. The concept of the electromagnetic mass had already appeared in the classical theory of the electron by Lorentz, who computed the electromagnetic mass by applying the classical

theory and obtained the result that the mass becomes infinite for the point (zero size) electron. On the other hand, the electromagnetic mass was computed in quantum theory by various people, and here I mention particularly the work of Weisskopf[9]. According to him, the quantum-mechanical electromagnetic mass turned out to be infinite, and although the order of the divergence was much weaker than in the case of the Lorentz theory, the observed mass, which included this additional mass, would be infinite. This would be, of course, contrary to experiment.

In order to overcome the difficulty of an infinitely large electromagnetic mass, Lorentz considered the electron not to be point-like but to have a finite size. It is very difficult, however, to incorporate a finite sized electron into the framework of relativistic quantum theory. Many people tried various means to overcome this problem of infinite quantities, but nobody succeeded.

In connection with field reactions, the next problem which attracted the attention of physicists was determining what kind of influence the field reaction exerts in electron-scattering processes. Let us consider, as a concrete example, a problem in which an electron is scattered by an external field. In the ordinary treatment, we neglect the effect of field reactions on the scattered electron, assuming that it is negligibly small. Then the behavior of the scattering obtained by calculation (e.g. the Rutherford formula) fits very well with experiment. But what will happen if the influence of field reaction is taken into account? This theoretical problem was examined non-relativistically by Braunbeck-Weinmann[10] and Pauli-Fierz[11] and relativistically by Dancoff[12].

While Dancoff applied an approximation method, the perturbation method, in his relativistic calculation, Pauli and Fierz treated the problem in such a way that the most important part of the field reaction was first separated out exactly by employing a contact transformation method which was similar to the one which Bloch-Nordsieck[13] had published a year before. Since Pauli and Fierz adopted a non-relativistic model, and further simplified the problem by using the so-called dipole approximation, their calculation was especially transparent. At any rate, both non-relativistic and relativistic calculations exhibited several infinities in the scattering processes\*.

The conclusions of these people were fatal to the theory. That is, the influence of the field reaction becomes infinite in this problem. The effect of field reaction on a quantity called the scattering cross section, which expresses quantitatively the behavior of the scattering, rather than becoming negligibly small, becomes infinitely large. This does not, of course, agree with experiment.

This discouraging state of affairs generated in many people a strong distrust of quantum field theory. There were even those with the extreme view that the concept of field reaction itself had nothing to do at all with the true law of nature.

On the other hand, there was also the view that the field reaction might not be altogether meaningless but would play an essential role in the scattering processes, though the appearance of divergences revealed a defect of the theory. Heisenberg[14], in his paper published in 1949, emphasized that the field reaction would be crucial in meson-nucleon scattering. Just at that time I was studying at

Leipzig, and I still remember vividly how Heisenberg enthusiastically explained this idea to me and handed me galley proofs of his forthcoming paper. Influenced by Heisenberg, I came to believe that the problem of field reactions far from being meaningless was one which required a frontal attack.

Thus, after coming back to Japan from Leipzig, I began to examine the nature of the infinities appearing in scattering processes at the same time that I was engaged in the above-mentioned work of formulating a covariant field theory. What I wanted to know was what kind of relationship exists between the infinity associated with the scattering process and that associated with the mass. If you read the above-mentioned papers of Bloch-Nordsieck and Pauli-Fierz, you will see that one of the terms containing infinite quantities is first separated out by a contact transformation and this term turns out to be just the term modifying the mass. Besides this kind of infinity there appeared, according to Pauli-Fierz, another kind of infinity characteristic of the scattering process. I further investigated a couple of simple models which were not realistic, but could be solved exactly. What was understood from these models, was that the most strongly divergent terms in the scattering process had the same form as the expression giving the modification of the particle mass due to field reactions, and therefore both should be manifestations of the same effect. In other words, at least a portion of the infinities appearing in the scattering process could be amalgamated into the infinity associated with the particle mass, leaving infinities proper to the scattering process alone. These turned out to be more weakly divergent than the infinity associated with the mass.

Since these conclusions were derived from non-relativistic or unrealistic models, it was still doubtful whether the same thing would occur in the case of relativistic electrons interacting with the electromagnetic field. Dancoff tried to answer this question. He calculated relativistically the infinities appearing in the scattering process and determined which of them could be amalgamated into the mass and which remained as infinities proper to the scattering process alone. He found that there remained, in the latter group of infinite terms, one which was at least as divergent as the infinity of the mass, a finding which differed from the conclusion based on fictitious models.

Actually, there are two kinds of field reactions in the case of the relativistic electron and electromagnetic field. One of them ought to be called "of mass type" and the other "of vacuum polarization type". The field reaction of mass type changes the apparent electronic mass from its original value by the amount of the electromagnetic mass as was calculated by Weisskopf. On the other hand, the field reaction of vacuum polarization type changes the apparent electronic charge from its original value. As was discussed in further papers by Weisskopf[15] and others, infinite terms appear in the apparent electronic charge if the effect of vacuum polarization is taken into account. However, in this talk, for simplicity, I will mention only briefly the divergence of the vacuum polarization type.

(4) In the meantime, in 1946, Sakata[16] proposed a promising method of eliminating the divergence of the electron mass by introducing the idea of a field of cohesive force. It was the idea that there exists unknown field, of the

type of the meson field which interacts with the electron in addition to the electromagnetic field. Sakata named this field the cohesive force field, because the apparent electronic mass due to the interaction of this field and the electron, though infinite, is negative and therefore the existence of this field could stabilize the electron in some sense. Sakata pointed out the possibility that the electromagnetic mass and the negative new mass cancel each other and that the infinity could be eliminated by suitably choosing the coupling constant between this field and the electron. Thus the difficulty which had troubled people for a long time seemed to disappear insofar as the mass was concerned. (It was found later that Pais[17] proposed the same idea in the U.S. independently of Sakata.) Then what concerned me most was whether the infinities appearing in the electron-scattering process could also be removed by the idea of a plus-minus cancellation.

An example of a computation of how the field reaction influences the scattering process was already given by Dancoff. What we had to do was just to replace the electromagnetic field by the cohesive force field in Dancoff's calculation. I mobilized young people around me and we performed the computation together[18] Infinities with negative sign actually appeared in the scattering cross-section as was expected. However, when we compared these with the infinities with positive sign which Dancoff calculated for the electromagnetic field, the two infinities did not cancel each other completely. That is to say, according to our result, the Sakata theory led to the cancellation of infinities for the mass but not for the scattering process. It was also known that the infinity of vacuum polarization type was not cancelled by the introduction of the cohesive force field.

Unfortunately, Dancoff did not publish the detailed calculations in his paper, and while we were engaged in the above considerations, we felt it necessary to do Dancoff's calculation over again for ourselves in parallel with the computation of the influence of the cohesive force field. At the same time I happened to discover a simpler method of calculation.

This new method of calculation was to use the technique of contact transformations based on the previously mentioned formalism of the covariant field theory and was in a sense a relativistic generalization of the Pauli-Fierz method. This method had the advantage of separating the electromagnetic mass from the beginning, just as was shown in their paper.

Our new method of calculation was not at all different in its contents from Dancoff's perturbation method, but had the advantage of making the calculation more clear. In fact, what took a few months in the Dancoff type of calculation could be done in a few weeks. And it was by this method that a mistake was discovered in Dancoff's calculation ; we had also made the same mistake in the beginning. Owing to this new, more lucid method, we noticed that, among the various terms appearing in both Dancoff's and our previous calculation, one term had been overlooked. There was only one missing term, but it was crucial to the final conclusion. Indeed, if we corrected this error, the infinities appearing in the scattering process of an electron due to the electro-magnetic and cohesive force fields cancelled completely, except for the divergence of vacuum

polarization type.

(5) When this unfortunate error of Dancoff's was discovered, we had to re-examine his conclusions concerning the relation between the divergence of the scattering process and the divergence of the mass, in particular, the conclusion that there remained a portion of the infinities of the scattering process which could not be amalgamated into the modification of the mass. In fact, it turned out that after correcting the error, the infinity of mass type appearing in the scattering process could be reduced completely to the modification of the mass, and the remaining field reaction belonging to the scattering proper was not divergent[19]. In other words, the highest divergence part of the infinities appearing in the scattering process, in the relativistic as well as in the non-relativistic case, could be attributed to the infinity of mass. The reason why the remaining part became finite in the relativistic case was due to the fact that the order of the highest divergence was only  $\log co$ , and after amalgamating the divergence into the mass term, the remainder was convergent. The great value of this method of contact transformations was that once the infinity of the mass was separated out, we obtained a divergence-free theoretical framework.

In this way the nature of various infinities became fairly clear. Though I did not describe here the infinity of vacuum polarization type, this too appears in the scattering process, as mentioned earlier. However, Dancoff had already discovered that this infinity could be amalgamated into an apparent change in the electronic charge. To state the conclusion, therefore, all infinities appearing in the scattering process can be attributed either to the infinity of the electromagnetic mass or to the infinity appearing in the electronic charge - there are no other divergences in the theory.

It is a very pleasant thing that no divergence is involved in the theory except for the two infinities of the electronic mass and charge. We cannot say that we have no divergences in the theory, since the mass and charge are in fact infinite. It is to be noticed, however, that if we reduce the infinities appearing in the scattering process to modifications of mass and charge, the remaining terms all become finite. Further, if we examine the structure of the theory, after the infinities are amalgamated into the mass and charge terms, we see that the only mass and charge appearing in the theory are the values modified by field reactions - the original values and excess ones due to field reactions never appear separately.

This situation gives rise to the following possibility. The theory does not of course yield a resolution of the infinities. That is, since those parts of the modified mass and charge due to field reactions contain divergence, it is impossible to calculate them by the theory. However, the mass and charge observed in experiments are not the original mass and charge but the mass and charge as modified by field reactions, and they are finite. On the other hand, the mass and charge appearing in the theory are, as I mentioned above, after all the values modified by field reactions. Since this is so, and particularly since the theory is unable to calculate the modified mass and charge, we may adopt the procedure of substituting experimental values for them phenomenologically. When a theory is incompetent in part, it is a common procedure to rely on experiment

for that part. This procedure is called the renormalization of mass and charge, and our method has brought the possibility that the theory will lead to finite results by the renormalization even if it contains defects.

The idea of renormalization is far from new. Many people used explicitly or implicitly this idea, and we find the word renormalization already in Dancoff's paper. In his calculation it appeared, because of an error that there still remained a divergence in the scattering even after the renormalization of the electron mass. This error was very unfortunate; if he had performed the calculation correctly, the history of renormalization theory would have been completely different.

(6) This period, around 1946-1948, was soon after the second world war, and it was quite difficult in Japan to obtain information from abroad. But soon we got the news that in the U.S., Lewis and Epstein[20] found Dancoff's mistake and gave the same conclusions as ours, Schwinger[21] constructed a covariant field theory similar to ours, and he was probably performing various calculations making use of it. In particular, little by little news arrived that the so-called Lamb-shift was discovered[22] as a manifestation of the electromagnetic field reaction and that Bethe[23] was calculating it theoretically. The first information concerning the Lamb-shift was obtained not through the Physical Review, but through the popular science column of a weekly U.S. magazine. This information about the Lamb-shift prompted us to begin a calculation more exactly than Bethe's tentative one.

The Lamb-shift is a phenomenon in which the energy levels of a hydrogen atom show some shifts from the levels given by the Dirac theory. Bethe thought that the field reactions were primarily responsible for this shift. According to his calculation, field reactions give rise to an infinite level shift, but he thought that it should be possible to make it finite by a mass renormalization and a tentative calculation yielded a value almost in agreement with experiments.

This problem of the level shift is different from the scattering process, but it was conceivable that the renormalization which was effective in avoiding infinities in the scattering process would be workable in this case as well. In fact, the contact transformation method of Pauli and Fierz devised to solve the scattering problem could be applied to this case, clarifying Bethe's calculation and justifying his idea. Therefore the method of covariant contact transformations, by which we did Dancoff's calculation over again would also be useful for the problem of performing the relativistic calculation for the Lamb-shift. This was our prediction.

The calculation of the Lamb-shift was done by many people in the U.S.[24]. Among others, Schwinger, commanding powerful mathematical techniques, and by making thorough use of the method of covariant contact transformations, very skilfully calculated not only the Lamb-shift but other quantities such as the anomalous magnetic moment of the electron. After long, laborious calculations, less skilful than Schwinger's, we[25] obtained a result for the Lambshift which was in agreement with Americans'. Furthermore, Feynman[26] devised a convenient method based on an ingenious idea which could be used to extend the approximation of Schwinger and ours to higher orders, and Dyson[27] showed



that all infinities appearing in quantum electrodynamics could be treated by the renormalization procedure to an arbitrarily high order of approximation. Furthermore, this method devised by Feynman and developed by Dyson was shown by many people to be applicable not only to quantum electrodynamics, but to statistical mechanics and solid-state physics as well, and provided a new, powerful method in these fields. However, these matters will probably be discussed by Schwinger and Feynman themselves and need not be explained by me. So far I have told you the story of how I played a tiny, partial role in the recent development of quantum electrodynamics, and here I would like to end my talk.

\* The main purpose of the work of Bloch-Nordsieck and Pauli-Fierz was to solve the so-called infrared catastrophe which was one of a number of divergences. Since this difficulty was resolved in their papers we confine ourselves here to a discussion of the other divergences which are of the so-called ultraviolet type.

- [1] Dirac, Proc. Roy. Soc. (London), 136 (1932) 453.
- [2] Heisenberg and Pauli, Z. Physik, 56 (1929) I.
- [3] Rosenfeld, Z. Physik, 76 (1932) 729.
- [4] Dirac, Fock and Podolsky, Phys. U.S.S.R., 2 (1932) 468.
- [5] Bloch, Phys. Z. U.S.S.R., 5 (1943) 301.
- [6] Yukawa, Kagaku (Tokyo), 12 (1943) 249.
- [7] Dirac, Phys.Z. U.S.S.R., 3 (1933) 64.
- [8] Tomonaga, Progr. Theoret. Phys. (Kyoto), I (1946) 27; Koba, Tati and Tomonaga, *ibid.*, 2 (1947) IOI, 198; Kaneshawa and Tomonaga, *ibid.*, 3 (1948) I, IOI.
- [9] Weisskopf, Phys. Rev., 56 (1939) 72.
- [10] Braunbeck and Weinman, Z. Phys., 110(1938) 369.
- [11] Pauli and Fierz, Nuovo Cimento, 15 (1938) 267.
- [12] Dancoff, Phys. Rev., 55 (1939) 959.
- [13] Bloch and Nordsieck, Phys. Rev., 52 (1937) 54.
- [14] Heisenberg, Z. Phys., 113 (1939) 61.
- [15] Weisskopf, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd., 14 (1936) No. 6.
- [16] Sakata, Progr. Theoret. Phys. (Kyoto), I (1946) 143.
- [17] Pais, Phys. Rev., 73 (1946) 173.
- [18] Ito, Koba and Tomonaga, Progr. Theoret. Phys. (Kyoto), 3 (1948) 276; Koba and Takeda, *ibid.*, 3 (1948) 407.
- [19] Koba and Tomonaga, Progr. Theoret. Phys. (Kyoto), 3 (1948) 290; Tati and Tomonaga, *ibid.*, 3 (1948) 391.
- [20] Lewis, Phys. Rev., 73 (1948) 173; Epstein, *ibid.*, 73 (1948) 179.
- [21] Schwinger, Phys. Rev., 73 (1948) 416, 74 (1948) 1439, 75 (1949) 651, 76 (1949) 790.
- [22] Lamb and Rutherford, Phys. Rev., 72 (1947) 241.
- [23] Bethe, Phys. Rev., 72 (1947) 339.
- [24] Kroll and Lamb, Phys. Rev., 75 (1949) 388; French and Weisskopf, *ibid.*, 75 (1949) 1241.

[25] Fukuda, Miyamoto and Tomonaga, *Progr. Theoret. Phys. (Kyoto)*, 4 (1949) 47, 121.

[26] Feynman, *Phys. Rev.*, 74 (1948) 1430, 76 (1949) 769.

[27] Dyson, *Phys. Rev.*, 75 (1949) 486, 1736.

From *Nobel Lectures, Physics 1963-1970*, Elsevier Publishing Company, Amsterdam, 1972

Copyright ©The Nobel Foundation 1965