# Zero-charge and asymptotic freedom

V. B. Berestetskiĭ

*Institute of Theoretical and Experimental Physics, Moscow*
Usp. Fiz. Nauk **120**, 439–454 (November 1976)

PACS numbers: 01.10.Jj

1. The problems of present-day theory are the problems of field theory. Whereas classical mechanics and quantum mechanics are logically faultless schemes (the limitations on the region of applicability of the former have been dictated solely by experiment, and, with regard to the latter, up to now there are no facts in sight that would give us cause to reconsider it), field theory has contained internal contradictions from the very beginning.

Why is field theory necessary? It is required by the theory of relativity, in which every interaction propagates with a finite velocity and, therefore, a carrier of the interaction—a field—is required. A field is a certain dynamical system having an infinite number of degrees of freedom. On the other hand, by virtue of the fact that in relativistic mechanics there is no particle-number conservation law—particles can be created and annihilated—every system actually has an infinite number of degrees of freedom, i.e., acquires the character of a field. Thus, relativistic theory is a theory of interacting fields.

It is well known that internal contradictions were already contained in classical electrodynamics. The Maxwell-Lorentz equations give the possibility of determining the field from a given distribution of charges and currents. But the distribution of charges and currents is itself determined by the field. The equations of motion of the charges give the possibility of finding this distribution, but for a given field. However, the solution of the consistent system of equations for the combined dynamical system "field and charges" comes up against insuperable difficulties.

In classical electrodynamics only one step was taken. It is possible to find first the motion of an electron in a given field, then find the field created by the electron with the given motion, and then take into account the action of this field on the motion of the electron. Thus is the radiation reaction calculated. But for this method of successive approximations to be carried through it is necessary that the following condition be fulfilled: the force of the radiation reaction should be smaller than the original force. This is fulfilled for a field with wavelength greater than the classical electron radius $r_e = e^2/mc^2$. Classical electrodynamics does not know the answer for shorter lengths.

Here the difficulties of classical electrodynamics are removed by external means. A quantum treatment is already required for wavelengths significantly greater than $r_e$—of the order of the electron Compton wavelength $\lambda_e = \hbar/mc = 137 r_e$.

Classical electrodynamics remains fundamentally unsatisfactory, but in practice the difficulties are transferred to quantum electrodyamics.

The solutions of the equations of quantum electrodynamics are also constructed by the method of successive approximations (perturbation theory); the solutions appear as expansions in the small parameter $e^2 = 1/137$.[1)] And in the first approximation quantum electrodynamics led to excellent results, describing a wide range of phenomena: emission of photons, scattering of photons and electrons, and the formation and annihilation of electron-positron pairs. But attempts to find corrections to these results led to absurdities. The next term of the expansion did indeed contain an extra power of the small parameter $e^2$, but multiplying it there appeared an integral over the wavelengths, of the type

$$\int \frac{d\lambda}{\lambda},$$

in which the upper limit was given by a wavelength characteristic of the given process (determined, e.g., by the transferred momentum or energy), and the lower limit was unbounded, i.e., the integral was logarithmically divergent at the lower limit. The situation here is similar to that which obtained in classical electrodynamics in the attempt to calculate the electromagnetic mass of the electron. Indeed, the Coulomb field at a distance $r$ from a charge at rest is proportional to $1/r^2$, and, therefore, the energy $\sim \int r^{-2} dr$. We note that the entire electron self-energy $mc^2$ is already contained in the region $r > r_e/2$ in this integral, so that, in this problem too, the physical restriction of the range of applicability to the region $r > \lambda_e$ hands over the problem to quantum electrodynamics.

Such a strange situation, in which the theory gives answers in agreement with experiment by an approximate method whose validity cannot be justified within the theory itself by the small size of the corrections to it, could not fail to irritate the pioneers of quantum electrodynamics. Therefore, they, and in particular Heisenberg, expected the same resolution of the problem as there had been for classical electrodynamics, i.e., that facts contradicting the results of quantum electrodynamics would appear and then it would be necessary to assume that a radically new theory was needed and quantum electrodynamics would be restricted (no longer from within, but from without) to the region of not very short distances or not very high energies. Therefore, each new experimental fact was treated first of all from this point of view. (An idea of the frame of mind of theorists at this time is given by Heitler's book,

[1)] Here and in the following, by $e^2$ we mean the dimensionless quantity $e^2/\hbar c$.

which came out in 1936.) Such was the situation when showers were discovered in cosmic rays. This was regarded as contradicting the theory—in quantum electrodynamics the probability of production of a large number of pairs and photons in one event is extremely small. Later it turned out that the showers are beautifully described by cascade theory, without multiple production and in agreement with quantum electrodynamics. Next a contradiction arose between cascade theory and the intensity of the soft component of cosmic rays in the lower layers of the atmosphere. The contradiction was resolved by the discovery of muons and pions and by a revision of the scheme of the origin of cosmic rays, and cascade theory, i.e., quantum electrodynamics, was found to be correct.

The next stage in the development of the theory starts from a new experimental fact—the discovery of the Lamb shift of levels. Its order of magnitude was such that it would have been possible to interpret it as the second-order corrections of perturbation theory, if the theory had made it possible to calculate this second approximation. And then it was calculated. How? Surely the second approximation contains an infinity—a divergent integral! Crudely speaking, the prescription is as follows. We close our eyes to the divergence of the integral. We pretend not to notice it. For this it is best to denote a certain (divergent) part of it, e.g., $\int^{\lambda_e}$, by some symbol, $a$ or $\delta$. Then we recall that multiplying $a$ or $\delta$ there appears the small parameter $e^2$. We note that in the expression for any amplitude calculated to a given accuracy in powers of $e^2$, $a$ and $\delta$ appear only in the form $e^2(1+\delta) = e'^2$ and $m+\delta = m'$, where $e$ and $m$ are the charge and mass of the electron. We then do a renormalization, i.e., we assume that it is $e'$ and $m'$ which are the electron charge and mass that are measured in the standard way in experiment. After this there appears before us a formula of the second or a higher approximation, and there are no infinities. The precise procedure of the calculations differs little from this rough outline. On a fundamental level it is desirable to convince oneself of the uniqueness of the procedure and of the fact that, in any approximation, everything can be reduced to a renormalization of the mass and charge. This has been done.

Is this theory a good one? Yes—there is excellent agreement with experiment. At the present time the most sensitive check consists in measuring the anomalous magnetic moment of the electron and muon. For the electron the calculations have been taken up to the third approximation and surpass experiment in accuracy. For the muon, in accordance with the theory, experiment has reached that level of accuracy beyond which it is necessary to take into account the contribution of hadrons. But how are we to understand this handling of the divergent integrals? The simplest way is to imagine that the lower limit is in fact finite. Something external to the theory itself cuts off the integrals; precisely where this happens is not very important for us—it is somewhere at very short distances.

**2.** The students of Landau, knowing how highly he estimated concrete physical results and how little he liked conversations on general "fundamental" topics, were somewhat surprised by the comparative restraint with which Landau greeted the outstanding successes of quantum electrodynamics in calculating the radiative corrections. But they displayed insufficient understanding of the character of their teacher's attitude to science. In fact, Landau could not work outside an atmosphere of ideological clarity. To his way of thinking, the latter concept was by no means the same thing as formal rigor. He did indeed dislike debates on themes concerned with the fundamental basis of the sciences, but this was only for those sciences whose fundamental principles he considered to be clear, such as, e.g., quantum mechanics or statistical physics. His attitude to those subjects in which there was no clarity was entirely different. A good example is provided by his earlier papers with Peierls, and also by Landau's attitude to the parity-nonconservation problem that arose later.

The same situation obtained in the era of the radiative corrections. He valued the results, of course, but regarded the methods by which they were obtained as formal prescriptions. So Landau and his students devoted themselves to seeking the foundations.

The result was four articles by Landau, Abrikosov and Khalatniko.[1] The formulation of the problem in these papers is as follows. Insofar as the divergences are associated with short distances, i.e., with the point nature of the charges, is it not possible to find first the solution for charges of finite dimensions $r_0$ and then see what will happen when $r_0 \to 0$? In order to elucidate this approach, we shall consider a simple example, which, as it turns out, is entirely adequate for the general problem. Suppose that we have two equal charges, situated a distance $r$ apart. If the charges are in a vacuum, their potential energy has the form of Coulomb's law

$$V = e^2/r$$

irrespective of whether the charge is a point charge or has an arbitrary spherically symmetric distribution with radius $r_0 < r$. This expression is the solution of the electrostatics equation

$$\operatorname{div} \mathbf{E} = 0 \quad \text{or} \quad \Delta V = 0, \quad r > r_0.$$

If the charges are in a dielectric, the electrostatics equation takes the form

$$\operatorname{div} \mathbf{D} = 0,$$

where

$$\mathbf{D} = (1 + 4\pi\chi)\,\mathbf{E},$$

($\chi$ is the polarizability of the medium). In the general case the quantity $\chi$ is an integral operator. The expression for $V$ can be written in the form

$$V = \frac{e^2 d}{r}.$$

To determine $d$ it is necessary to know the polarizability of the medium. We note that the form of $d$ depends on the dimensions $r_0$ of the charge and that we need to know the solution of the electrostatics equation only in the region $r > r_0$. The quantity $r_0$ can be varied arbitrarily. For example, even if the charge is a point charge, we can cut out a sphere of radius $r_0$ around it, but include in $e(r_0)$ not only the original charge but also all the associated charges induced by the polarization of the medium inside the sphere of radius $r_0$.[2)] If we choose $r_0 = r$, the simple Coulomb law is restored, i.e., for $r = r_0$, $d = 1$. It may be said that the problem reduces to the determination of $e^2(r)$, i.e., of the charge as a function of the radius:

$$e^2(r) = e^2(r_0)\, d. \tag{1}$$

In quantum electrodynamics, because of the phenomenon of polarization of the vacuum, charges placed in a vacuum do not interact in accordance with Coulomb's law. The problem consists in determining the polarization $\chi$ of the vacuum, and the function $d$. The solution of this problem by the method of perturbation theory implies that we take Coulomb's law as the initial (zeroth) approximation, the next approximation then considers the polarization of the vacuum in the given field, and so on. Thus, we obtain the solution in the form of a series

$$d = 1 + e^2 d_1 + \dots$$

The function $d_1$ contains the divergent integral $\int \lambda^{-1} d\lambda$ spoken of earlier. But if we start not from a point charge but from a charge of finite dimensions, an additional boundary condition is imposed on $d$: $d = 1$ for $r = r_0$. This means that the lower limit of the integral is now not 0 but $r_0$, and $\int_{r_0}^{r} \lambda^{-1} d\lambda = \ln(r/r_0)$. In place of the divergent integral we have obtained a finite integral, cut off at $r_0$. We have achieved this by restricting ourselves: we do not know what happens in the region $r < r_0$, and we have denoted by $e$ the entire charge within the sphere, including the primary ("bare") charge and the associated vacuum charges. It then turns out that, although we have based the perturbation theory on the small value of $e^2$, the expansion parameter is $e^2 \ln(r/r_0)$. Therefore, we can vary $r_0$ only within limits such that the condition

$$e^2 \ln \frac{r}{r_0} \ll 1.$$

is fulfilled.

We note also that this simple expression for $d_1$ holds for $r \ll \lambda_e$, while for $r$ of the order of $\lambda_e$ or $r > \lambda_e$ we must replace $r$ by $\lambda_e$ in this formula (furthermore, terms independent of $r_0$, which will not be of interest to us, are added; for simplicity we confine ourselves to distances $r < \lambda_e$).

It should be said that the interpretation, used in these papers, of the cutoff of the divergent integrals was not fundamentally new. What was essentially new was the second part of the formulation of the problem; this reduces to the determination of the function $d$ without the use of perturbation theory. The point is that the formal system of equations of quantum electrodynamics is more symbolic than real. From it we can obtain a perturbation-theory series for the amplitudes. But there is no closed system of equations for the amplitudes or Green functions—an infinite system of equations (the Dyson equations) is obtained for them. In the work of Landau, Abrikosov and Khalatnikov it was shown that for small $e^2$ a closed system of equations can be formulated. The authors solved it for the asymptotic limit $r_0 \ll r \leqslant \lambda_e$ and found that

$$d = \frac{1}{1 + (2/3\pi)\, e^2 \ln (r/r_0)}, \tag{2}$$

i.e.,

$$e^2(r) = \frac{e^2(r_0)}{1 + (2/3\pi)\, e^2(r_0) \ln (r/r_0)}. \tag{3}$$

We note that in the case when $(2e^2/3\pi)\ln(r/r_0) < 1$ the expression (2) can be expanded in the series

$$d = 1 - \frac{2e^2}{3\pi} \ln \frac{r}{r_0} + \left(\frac{2e^2}{3\pi} \ln \frac{r}{r_0}\right)^2 + \dots,$$

which coincides with the series that can be obtained in perturbation theory if terms containing $e^2$ and $\ln(r/r_0)$ to equal powers are kept. The formulas (2) and (3) are valid under lesser restrictions: for $e^2 \ll 1$ and an arbitrary value of $r/r_0 \gg 1$ (but, of course, $r < \lambda_e$; in practice, they are valid up to $r \sim \lambda_e$).

Thus, in these papers it was shown that if we start from a primary charge that is not a point but has dimensions $r_0$, then for $e^2 \ll 1$ a solution for $d$ exists. This solution makes it possible to carry out the renormalization of the charge. What we usually call the physical charge is the constant appearing in Coulomb's law for long-wavelength photons. In the terms used, this is $e(\lambda_e)$. Thus, we can carry out the calculations by perturbation theory and express the results in terms of $e(\lambda_e)$; the primary ("bare") charge $e(r_0)$ never enters our formulas, if we are interested in the region $r > r_0$. The renormalization of the mass can be justified analogously.

Formula (3) was obtained taking into account only the contribution of the electrons to the polarization of the vacuum. Apart from the contribution of the electrons, it is necessary also to take into account that of the muons, of other heavy leptons if they exist, and perhaps, of other particles (e.g., quarks). Therefore, in the limit of small $r$ the coefficient of the logarithm in (2) and (3) must be increased by a factor of $\nu$, where $\nu$ is the effective number of fermions participating in the polarization of the vacuum at short distances.

Results analogous to (2) and (3) were obtained at almost the same time by Gell-Mann and Low.[2] They

[2)] One need not visualize both charges as spheres of radius $r_0$. It is only necessary to regard the "active" charge, i.e., the one "creating" the field, as such. Of course, it does not matter which of the two charges is assigned the "active" role and which the "passive." It is important only that their product be measurable. In the following we shall regard $e^2$ rather than $e$ as a function of $r_0$.

devised a method that was developed later as the renormalization-group method (cf.[9]). The starting point of the method is the relation (1). Since $d$ is a dimensionless quantity, for $r \ll \lambda_s$ it is a function only of the charge $e$ and the ratio $r/r_0$. Thus,

$$e^2(r) = e^2(r_0)\, d\left(e^2(r_0),\ \frac{r}{r_0}\right). \tag{4}$$

The relation (4) remains valid for an arbitrary choice of $r_0$. Therefore, putting $r_0 = \rho$ and $r = \rho + d\rho$ and taking into account that $d(e^2, 1) = 1$, we obtain from (4), having denoted $e^2(\rho) = \alpha$, the following differential equation

$$d\alpha = \varphi(\alpha)\, d\rho/\rho,$$

where

$$\varphi(\alpha) = \alpha \left(\frac{\partial d(\alpha, x)}{\partial x}\right)_{x=1} \tag{5}$$

or

$$\frac{d\alpha}{\varphi(\alpha)} = \frac{d\rho}{\rho}. \tag{6}$$

Integrating (6) from $\rho = r_0$ to $\rho = r$, we obtain

$$\ln \frac{r}{r_0} = \int_{e^2(r_0)}^{e^2(r)} \frac{d\alpha}{\varphi(\alpha)}. \tag{7}$$

To perform the integration in (7) we need to know $\varphi(\alpha)$, a function of only one variable (the charge). Therefore, assuming the charge to be small, we can make use of perturbation theory. We then obtain

$$d(\alpha, x) = -1 - \frac{2\alpha}{3\pi} \ln x, \quad \varphi(\alpha) = -\frac{2\alpha^2}{3\pi} \tag{8}$$

and from (7) we obtain the expression (3) for $e^2(r)$.

The essence of the method can be elucidated in the language of the equivalent electrostatics problem. We solve the equation for the potential exactly, and determine the polarizability of the medium by perturbation theory.

The method of Gell-Mann and Low gives the possibility, in principle, of improving the approximation by taking into account terms of higher order in $\alpha$ in $\varphi(\alpha)$. For this it is necessary to calculate for $d(\alpha, x)$ in perturbation theory the terms of order $\alpha^{n+1} \ln x$ that do not appear in the expansion of the formula (2).

**3.** We turn now to the analysis of the formula (3):

$$e^2(r) = \frac{e^2(r_0)}{1 + \beta e^2(r_0) \ln (r/r_0)}, \tag{9}$$

where

$$\beta = \frac{2\nu}{3\pi}.$$

We shall trace what happens to $e^2(r)$ on decrease of $r_0$ with constant $e^2(r_0)$. The formula (9) was obtained under the assumption that $e^2(r_0)$ is small, but the magnitude of the logarithm was not restricted. Consequently, by decreasing $r_0$ we can reach a value of it such that

$$\beta e^2(r_0) \ln \frac{r}{r_0} \gg 1.$$

Then we can neglect unity in the denominator of (9), and $e^2(r)$ is found to be a quantity independent of $e(r_0)$:

$$e^2(r) = \frac{1}{\beta \ln \frac{r}{r_0}}. \tag{10}$$

We can now pass to the point-charge limit

$$e^2(r) \to 0 \quad \text{as} \quad r_0 \to 0. \tag{11}$$

This result was obtained by Landau and Pomeranchuk[3] and Fradkin.[4]

But we are not obliged to assume that $e^2(r_0)$ is small. From experiment we know only that the "renormalized" or "physical" charge $e^2(\lambda_s)$ is small. We are not in a position to solve the problem for large $e^2(r_0)$ exactly. However, Landau and Pomeranchuk adduced arguments that the result (11) should also remain valid for arbitrary $e^2(r_0)$. These arguments are based on the formally exact expression for $e^2(r)$. For example, the latter can be represented in the form of the following functional integral (cf., e.g., [12]):

$$e^2(r) = \int A(0)\, A(r)\, f(A) \exp\left[-\frac{1}{2e^2} \int A(x)\, \Box\, A(x)\, d^4x\right] dA;$$

here $e^2$ (i.e., $e^2(r_0)$) appears only in the exponential factor that expresses the role of the free field in the Lagrangian, while the factor $f(A)$ expressing the role of the interaction does not contain $e$. But $e^2$ appears in the denominator in the exponent, and in this case the function $e^2(r)$ is found to be independent of $e$, even for small $e^2$. Therefore, it seems natural to think that with increase of $e$ the role of this factor, i.e., the role of the free field relative to the role of the interaction, will be still smaller.

"We arrive at the fundamental conclusion that it apparently follows from formal quantum electrodynamics that the electron charge is equal to zero. The caveat "apparently" refers to a certain lack of rigor in the reasoning expounded above." (This is a quotation from the article by Landau and Pomeranchuk.)

The meaning of the results (9)–(11) is simple. A charge placed in a medium that becomes polarized decreases on account of the polarization. At short distances this polarization is so strong that, irrespective of the magnitude of the charge, at a certain distance the residual charge no longer depends on the original charge. In the limit of a point primary charge (even an infinite one), at any finite distance nothing of it remains.

This result substantially changes our conception of the content of the equations of quantum electrodynamics. Before, the situation appeared to be as follows: there are some formal equations, and their solution is known in the form of a perturbation-theory series, each term of which, except the first, contains infinities. Now, there is a solution, obtained by passing to the limit of a point charge from a charge of finite dimensions, but this solution gives zero charge, i.e., absence of all interaction and absence of all processes. Such a theory is not meaningless, but is physically unsatisfactory.

If the field equations have solutions which can be obtained by passing to the limit of a point charge from a charge of finite dimensions, it is possible to try to find solutions by means of a more general limiting process, in which there are two lengths—the radius $r_0$ of the electron and the range $r_0'$ of the interaction of the electron with the field. Such a method was proposed by Abrikosov and Khalatnikov.[10] Pomeranchuk showed[5] that, using the method, it is possible to escape from the assumption that the charge $e^2(r_0)$ is small and obtain the results (10), (11) directly for any value of $e(r_0)$, in confirmation of the arguments expressed in the paper of Landau and Pomeranchuk. Thus, the way was open to the study of fields with large coupling constant, e.g., the theory of the interaction of nucleons with the $\pi$-meson field. Strictly, it was this which constituted the main interest in Pomeranchuk's investigations.

In a series of papers[5] Pomeranchuk showed that, in this case too, field theory leads to zero value of the effective coupling constant $g(r)$ at finite distances (the important problem of taking meson-meson interactions into account was considered by Dyatlov, Sudakov and Ter-Martirosyan[11]).

The conclusion that Pomeranchuk drew from these results was radical: field theory in its existing form was useless for the description of strong interactions.

How can the practical successes of quantum electrodynamics in its domain be understood? We rewrite the relation (9) in the form solved for $e^2(r_0)$, and choose $r=\lambda_e$:

$$e^2(r_0)=\frac{e^2(\lambda_e)}{1-\beta e^2(\lambda_e)\ln(\lambda_e/r_0)}; \qquad (12)$$

here $e^2(\lambda_e)$ is, as already noted above, the "physical" charge of the electron, i.e., the charge that is manifested at large distances, outside the effective region of polarization of the vacuum. When we go inside this region ($r_0<\lambda_e$), the charge increases. However, we cannot, on the basis of formula (12), reach a large charge, since we cannot use (12) near those values of $r_0$ for which the denominator vanishes. In fact, in practice one never needs to use (12) in this region, because this is the region of values $r_0\sim\exp(-137/\beta)$. Quantum electrodynamics is valid in practice precisely because we use approximate solutions with a point interaction, leaving open the question of what occurs at short distances.

Thus, there are two possible ways of interpreting quantum electrodynamics—the "theoretical" and the "pragmatic." The theoretical way consists in taking its equations seriously, solving them by the method of taking the limit $r_0\to 0$, and arriving at the trivial solution $e=0$ ("zero-charge"). The pragmatic way assumes that the "physical" charge $e(\lambda_e)$ is given empirically; then there are solutions in the form of perturbation-theory series, but the theory is not applicable to short distances.

For the strong interactions (i.e., for a constant $g(\lambda_e)$ of order unity) the pragmatic approach will give nothing, since in the attempt to make use of the formula of the type (12) we immediately find ourselves outside its region of applicability. There is no region in which it would be possible to disregard short distances.

4. It cannot be said that the radical conclusions of Pomeranchuk found wide recognition or sympathy. This, however, was not connected with the fact that important defects were found in these papers. True, it was pointed out that the Landau-Pomeranchuk solution was found for small $e^2$ and that taking higher approximations into account could alter the character of the Gell-Mann–Low function (5) in such a way that from formula (7) would follow not the relation (9), but another that does not lead to zero value of the charge. The requirements which it is necessary to impose on the behavior of $\varphi(\alpha)$, and the possible consequences, have been analyzed. But nobody has indicated a real mechanism that would lead to such a change of $\varphi(\alpha)$. It has also been correctly pointed out that a situation in which the limiting process has nothing in common with the exact solution is mathematically possible. But nobody has pointed to the existence of a solution of the equations of quantum electrodynamics with a point interaction, outside perturbation theory, other than that which is obtained by the physically intuitive method of passing to the limit of a point charge from a finite charge. And it is not known whether there is any real meaning in any other formulation of the problem. Thus, the criticism did not go further than that degree of doubt which was contained in the quotation cited earlier from the article of Landau and Pomeranchuk.

Rather, it may be said that the conclusions of Pomeranchuk were ignored by virtue of their purely negative character (Pomeranchuk himself thought this). With regard to electrodynamics they changed practically nothing, while with regard to the strong interaction they gave no indications of how we must work with them. Landau and Pomeranchuk[3] put forward the hypothesis that, since field theory is inapplicable to the strong interaction, i.e., at distances of the order of the Compton wavelength $\lambda_\pi$ of the $\pi$-meson, a new universal length of the order of $\lambda_\pi$ should appear in physics. However, the most sensitive experiment pertaining to such an effect—the sufficiently accurate measurement of the anomalous magnetic moment of the muon (since $\lambda_\mu$ is close to $\lambda_\pi$)—did not confirm this assertion. The value of the anomalous magnetic moment agrees with that calculated by quantum electrodynamics.

Nevertheless, most of the leading theorists independently developed a sense of deadlock in their attempts to obtain concrete physical results from field theory outside the framework of perturbation theory. This feeling was shared, for example, by Feynman, who had made great efforts in this direction and had developed for this a method of ordering of operators and the method of path integrals. He expressed his point of view in a letter to Landau dated about 1955, in which he characterized the attempts to create a theory of the strong interactions as a child-like imitation of quantum electrodynamics (with a simple replacement of the vector interaction by a pseudoscalar interaction) and expressed the opinion that Nature is "not so stupid" as not to invent something more subtle.

This feeling led to the result that the actual development of the theory of strong interactions in the following decade turned substantially away from field theory. Starting from the ideas expressed by Heisenberg in 1943, people began to consider the basic elements of the theory to be not the fields, but amplitudes (the elements of the scattering matrix) closer to directly measurable quantities. The scattering matrix satisfies the unitarity condition; this is a basic requirement of quantum mechanics.[3)] The unitarity relations indicate the singularities of the amplitudes, regarded as functions of complex variables. To a certain extent, the analytic properties of the amplitudes can be formally justified in field theory. Together, unitarity and analyticity produce a series of relations, which can be regarded as the analog of a dynamical system of equations. Unfortunately, this system is infinite, and the analytic properties of the amplitudes are complicated. Only in the case when there is a small parameter, as in quantum electrodynamics, does this actually enable us to solve practically all the problems without using field theory. Nevertheless, in this way it has been possible to obtain a number of results that are important for the theory of strong interactions, including such results as Pomeranchuk's theorem on the asymptotic equality of particle and antiparticle cross-sections and Froissart's theorem on the maximum growth of cross-sections.

The ideas of $S$-matrix theory have also been promoted by the fact that there became too many hadrons for each of them to be assigned its own field. It is natural to believe that it is impossible to regard some of them as elementary and others as composite, and the idea of the "democracy of the hadrons" arose. The idea of the "bootstrap" emerged, according to which practically any hadrons with suitable quantum numbers can be taken as the initial ones, and the analyticity and unitarity requirements will themselves give the entire spectrum of hardons.

The peak of this development was the method of complex angular momenta (Regge poles), developed for application to the theory of strong interactions in 1961–1962. Enthusiasts, e.g., Chew, considered that the theory of strong interactions was very near to completion.

However, at precisely this time the development of the theory was given a backward tilt in the direction of field theory. This began when it became clear that the complex angular-momentum plane was not so simple. Branch points appeared on it. The indication that they exist emerged from field theory after Mandelstam had pointed out the types of Feynman diagram that lead to branch points in the complex angular-momentum plane. In order to investigate the properties of the amplitudes with allowance for the branch points it was necessary to construct a field-theoretical form of the reggeon, and "ladders," "combs," etc., appeared. The original simplicity of the theory of Regge poles, analogous to the simplicity of the first approximation in field theory, was transformed into the complexity of an infinite series, in which, in the general case, all the terms were of the same order. Gribov developed a diagrammatic scheme, analogous to the Feynman-diagram scheme, in which each element of a diagram was effectively already an aggregate of certain Feynman diagrams. Working with this scheme became just as difficult an art as solving problems in field theory. And for all that, lying at the basis of it was a somewhat abstracted field theory, of a field that did not correspond to definite initial particles but possessed properties which cannot be obtained from a concrete model—properties ensuring that the transverse momentum transfers are small, a basic fact in the interaction of high-energy hadrons.

From another direction, the idea of higher symmetries of the hadrons appeared and could be interpreted intuitively by the hypothesis of quarks. Facts about the inelastic scattering of leptons emerged which seemed to indicate the existence of primary point objects within the hadrons.

Thus we returned to the domain of field theory. There are primary fermion fields (e.g., quarks), analogous to the electron-positron field. The democracy of the hadrons is preserved—they all consist of unobservable quarks. The fermion fields somehow interact amongst themselves, e.g., through the medium of certain boson fields. From this everything should follow.

But how are we to understand this, if field theory has led to the self-switching-off of the interaction—the vanishing of the charge?

**5.** It has turned out that this knot is being cut through rather than untied. For the fields and interactions that have been considered by everybody, including Pomeranchuk, since the 1930's, the zero-charge theorem is valid. But it is possible to construct fields and interactions to which the theorem is inapplicable and which possess nontrivial (not zero-charge) solutions. Such fields were first considered by Yang and Mills in 1954.[6]

It has been found that, for this, it is not necessary to decline to imitate quantum electrodynamics—rather, we must, in a certain sense, strengthen it. The electromagnetic field possesses two properties that were regarded as not very important in general field theory. The first is that the photon mass is equal to zero. The second is that the vector potential $A_\mu$ appearing in the expression for the interaction has four components, whereas the photon has only two polarizations. This led to formal difficulties. It was necessary either to reject an explicitly Lorentz-covariant description, associating only the transverse components of $A_\mu$ with the photons and introducing a Coulomb interaction between the charges (this is what was done in the earlier stage of the development of quantum electrodynamics, and it made it difficult to study the radiative corrections and carry out renormalizations), or, using, e.g., the invariant perturbation theory of Feynman, to introduce into the treatment longitudinally polarized and time-like

[3)]The unitarity condition expresses two fundamental features of the quantum-mechanical description: the probability interpretation of the amplitudes and the principle of superposition of states.

photons, the latter possessing a property that is not normal in quantum mechanics—they have negative norm ("negative probabilities"). These difficulties disappeared when the interactions of the fermions with other fields (e.g., the $\pi$-meson field) associated with particles with nonzero mass and zero spin were considered. In the development of general formal field theory—in particular, of axiomatic field theory, the case of electrodynamics was even ignored altogether. The requirement of positive norm and the existence of an energy gap between the vacuum and the first excited state of the system of fields were postulated. At the same time, owing to the zero mass of the photon, in electrodynamics the energy spectrum of the system abuts continuously upon the vacuum. The presence in quantum electrodynamics of the "infrared catastrophe"—the phenomenon of emission of an infinite number of long-wavelength photons in collisions—is connected with this circumstance. Owing to this circumstance, in quantum electrodynamics it is even very difficult to introduce with sufficient formal rigor the concept of the scattering matrix—the basic quantity in the calculation of probabilities and cross-sections. In physical applications these difficulties are fairly easily overcome, but in formal field theory they are a barrier.

The special properties of quantum electrodynamics indicated above are a consequence of the presence of an additional symmetry in its Lagrangian. In addition to invariance under Lorentz transformations, it possesses invariance under gauge transformations. If $\psi$ is the Dirac field of the electron and $A_\mu$ is the vector potential, the gauge transformations have the form

$$\psi \to e^{ie\chi}\psi, \quad A_\mu \to A_\mu - \partial_\mu \chi,$$

where the parameter $\chi$ is an arbitrary function of the coordinates and time, $e$ is the electron charge, and $\partial_\mu = \partial/\partial x^\mu$. The gauge transformations form a group: two successive transformations with parameters $\chi_1$ and $\chi_2$ give a transformation of the same type, with parameter $\chi = \chi_1 + \chi_2$. Since two successive transformations performed in opposite orders give the same result (the transformation operations commute), this group is called Abelian. It is designated in group theory as the group $U(1)$. The number 1 signifies that the group is a one-parameter group. It is characterized by only one generator (infinitesimal-transformation operator)—the unit operator. This group is also called local, since $\chi$ has a value at each point of space-time.

The existence of the gauge group leads to a simple prescription for introducing an interaction (the "minimal" interaction) between the fields. We take the free-field Lagrangian, i.e., the sum of the Lagrangians of the free electromagnetic and electron fields:

$$L^0 = L^0_A + L^0_\psi,$$

where

$$L^0_A = \frac{1}{4} F_{\mu\nu}F^{\mu\nu}, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

$$L^0_\psi = i\bar{\psi}(\gamma^\mu \partial_\mu - m)\psi,$$

and make in it the replacement

$$\partial_\mu \to D_\mu = \partial_\mu + ieA_\mu,$$

i.e.,

$$L(\psi, A_\mu, \partial_\mu) = L^0(\psi, A_\mu, D_\mu).$$

We thus obtain the Lagrangian $L$ of the interacting fields:

$$L^0 \to L = L^0 - j^\mu A_\mu, \quad j^\mu = e\bar{\psi}\gamma^\mu\psi.$$

The Lagrangian $L$ possesses invariance under gauge transformations, which was not possessed by $L^0$ (more precisely, by $L^0_\psi$).

Yang and Mills carried out a generalization of quantum electrodynamics, introducing in place of the Abelian gauge group $U(1)$ the nonabelian gauge group $SU(2)$, i.e., the group of rotations in isotopic space with local parameters (angles of rotation).[4] The group $SU(n)$ is also introduced entirely analogously. Since the group $SU(n)$ has $N = n^2 - 1$ parameters $\chi^i$ and the same number of generators $\tau^i$ (in the case $n = 2$, these are the isospin operators), it is necessary to introduce the same number of vector fields $A^i_\mu$ ($i = 1, 2, \ldots, N$) (in the case of the group $SU(2)$, $N = 3$, so that $A^i$ is a three-dimensional isovector; in the case of $SU(3)$, $A^i$ is an octet), and define $D_\mu$ as the matrix

$$D_\mu = \partial_\mu + igA_\mu,$$

where $A_\mu$ is the matrix

$$A_\mu = \sum_{i=1}^{N} \tau^i A^i_\mu,$$

and $g$ is the constant by which we have replaced the constant $e$.

Now, applying the previous prescription, we obtain the Lagrangian of the system of fields $A^i_\mu$ interacting with the Dirac fields $\psi^k$ in the form

$$L = -\frac{1}{4}\operatorname{Sp} F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}^k(\gamma^\mu D_\mu - m)\psi^k; \tag{13}$$

here summation over $k$ is implied; if $k$ corresponds to a spinor representation (isospinor in the case $n = 2$, superspinor for $n = 3$), $k = 1, \ldots, n$. The explicit form of the matrices $\tau^i$ corresponds to this representation. We note that the tensor field is equal to

$$F_{\mu\nu} = D_\mu A_\nu - D_\nu A_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu],$$

where the brackets denote the commutator. In the case of electrodynamics (the group $U(1)$), $[A_\mu, A_\nu] = 0$ and $F_{\mu\nu}$ does not differ from the free field. But in the case

[4] The group $U(1)$ is equivalent to the group of rotations of a plane about an axis perpendicular to it, and the group $SU(2)$ is equivalent to the rotation group in three-dimensional space. Since two rotations performed about different axes in opposite orders lead to different results, this is a nonabelian group.

of the group $SU(n)$ the commutator is nonzero and, therefore, the Lagrangian contains interaction of the fields $A_\mu^i$ with each other (terms in $L$ of third and fourth order in $A$). Thus, the corresponding "photons" radiate each other and scatter each other even in the absence of other fields.

At first, the Yang-Mills fields did not excite great interest. This is explained by their abstract character. They are a set of vector fields corresponding to massless particles, and we know only one such particle—the photon. In 1962 the theory of Yang-Mills fields attracted the interest of Feynman, who, while examining them as an intermediate model having a certain analogy with the quantum theory of gravitation, discovered that the theory requires a modification of the usual Feynman rules. This stimulated a large number of papers, which led to the construction of a perturbation-theory scheme for nonabelian gauge fields. Popular interest in Yang-Mills fields arose after Weinberg and Salam showed in 1967 that they can serve as a basis for constructing a unified theory of the weak and electromagnetic interactions; here we shall not be interested in this aspect. Nonabelian gauge fields became a serious basis for the theory of strong interactions in 1973, after Gross and Wilczek, and Politzer, [7] had shown that they do not obey the zero-charge theorem but possess a property that is in a certain sense opposite and has acquired the name of asymptotic freedom.

Gross and Wilczek and Politzer posed for Yang-Mills fields the problem analogous to that solved in quantum electrodynamics by Landau, Abrikosov and Khalatnikov and by Gell-Mann and Low—the problem of the effective charge $g(r)$ as a function of radius. They obtained an expression analogous to the expression (9) in electrodynamics, but with an important difference: the coefficient of $g^2(r_0)\ln(r/r_0)$ in the denominator can have the opposite sign. Writing it in the form solved for $g^2(r_0)$ (in analogy with (12)), we obtain

$$g^2(r_0)=\frac{g^2(r)}{1+\gamma g^2(r)\ln(r/r_0)}, \quad r>r_0, \tag{14}$$

where

$$\gamma=\frac{11}{6\pi}n-\beta; \tag{15}$$

here the first term is determined by the interaction of the fields $A_\mu^i$ amongst themselves and $n$ is the dimensionality of the group, i.e., the label in the group name $SU(n)$. The second term $\beta$ is determined by those fields with which the Yang-Mills fields additionally interact. In particular, if these are Dirac fermion fields forming a superspinor of $SU(n)$ (an $n$-plet), then $\beta=1/3\pi$. If the number of such independent $n$-plets is $\nu$, then

$$\beta=\frac{\nu}{3\pi}$$

and

$$\gamma=\frac{11n-2\nu}{6\pi}.$$

We see that for

$$\nu<\nu_{max}=\frac{11}{2}n$$

the coefficient $\gamma$ is positive. For $n=2$, $\nu_{max}=11$, and for $n=3$, $\nu_{max}=16$. For $n=2$ and $\nu=1$, $\gamma=10/3\pi$. For $n=3$ and $\nu=3$, $\gamma=27/6\pi$. For $n=3$ and $\nu=4$, $\gamma=25/6\pi$.

We now decrease $r_0$. For $\gamma g^2\ln(r/r_0)\gg 1$ we obtain from (14)

$$g^2(r_0)=\frac{1}{\gamma\ln(r/r_0)}$$

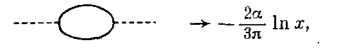
and

$$g^2(r_0)\to 0 \quad \text{as} \quad r_0\to 0. \tag{16}$$

This result is the direct opposite of (11). What happens is not that the charge at a finite distance vanishes for any value of the original point charge, but that a zero point charge corresponds to a finite charge at a finite distance. We note that it is impossible to obtain this result if, being formally guided by the field equations, we consider only point charges. It is necessary to proceed by a limit process, as was suggested by Landau.

The property (16) has acquired the name of asymptotic freedom—at short distances the interaction weakens and the particles become free. The polarization of the medium leads not to a decrease of the charge as in electrostatics, but to an increase. How does this come about?

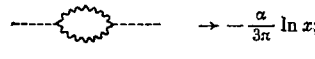
First we shall return to quantum electrodynamics. If we make use of the method of Gell-Mann and Low, then, to obtain $e^2(r)$, it is necessary, as described earlier, to calculate the function $\varphi(\alpha)$ by formula (5) and substitute it into (7). But to calculate $\varphi(\alpha)$ it is necessary to find the function $d(\alpha, x)$ by perturbation theory to first order in $\alpha$. For this we must consider the following Feynman diagram:




in which a dashed line corresponds to the Coulomb field and a solid line to the electron field. The electron loop expresses the virtual production and annihilation of a pair, giving rise to polarization of the vacuum. In this way we arrive at the expression (8). Diagrams of an analogous type also occur in the case of a Yang-Mills field, and describe the polarization of the vacuum by those particles with which it interacts. They lead to a term $-\beta\alpha\ln x$ and, as in the case of electrodynamics, have negative sign. But, besides this, the Yang-Mills field contains self-charged components and it is necessary to take into account their contribution to the vacuum polarization.

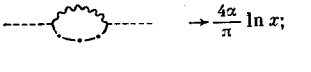
In electrodynamics, as already mentioned earlier, different descriptions of the interaction are possible. The old description is that the electromagnetic interaction is realized in two ways: the instantaneous (the Coulomb interaction) and the retarded—through emission and absorption of quanta that can have two transverse polarizations. A more modern description, which goes back to Feynman, is that the entire interaction is realized as if by emission and absorption of quanta of four polarizations. We shall use the old way,

following the paper of Khriplovich, [8] who calculated the function $d$ in Yang-Mills theory for the case $n=2$, $\nu=1$. The field has three components—two charged and one neutral. Each of these components consists of the Coulomb field and transverse quanta. Accordingly, two diagrams will give a contribution to $d$. The first is



$$\rightarrow -\frac{\alpha}{3\pi}\ln x;$$

here a dashed line corresponds to the neutral Coulomb field, and wavy lines to the charged quanta. This diagram, like the previous one, describes the production and annihilation of a pair of charged particles and gives a contribution to $d$ that again has negative sign. It can be shown that every diagram of this type leads to a negative quantity—this is a consequence of the unitarity relation. The second diagram, while outwardly similar, is different in character. It cannot be interpreted as the virtual creation and annihilation of particles, and it is this diagram which leads to a positive value:



$$\rightarrow \frac{4\alpha}{\pi}\ln x;$$

here the dashed-dotted line is the charged Coulomb field. This is not a propagating particle—it is an instantaneous interaction with charge transfer. The sum of these two diagrams gives the value $11/3\pi$ for the first term in $\gamma$, in agreement with (15) (for $n=2$, $\nu=1$).

**6.** The ideas about the strong interaction that hold sway at the present time reduce to the following. There exist several ($\nu$) kinds of quarks $q^{(f)}(f=1, 2, \ldots, \nu)$. Until recently it was assumed that $\nu=3$. At the present time data that indicate the necessity to introduce a fourth quark have appeared. It is possible that the number of types of quark will have to be increased. Each of the quarks $q^{(f)}$ forms a triplet of the group $SU(3)$, i.e., the quark of each kind $q^{(f)}$ can be found in three states $q_k^{(f)}$, $k=1, 2, 3$. The index $k$ is customarily called "color," e.g., $k=1$ is "blue," $k=2$ is "yellow," $k=3$ is "red." The local group $SU(3)$ is usually called color $SU(3)$ (it must not be confused with the $SU(3)$ group that groups together the three kinds of quark $q^{(f)}$ with different $f=1, 2, 3$; we shall not be concerned with this group below). The colored quarks $q_k^{(f)}$ interact with a vector gauge field $A_\mu^i$ that forms an octet (in accordance with the number of parameters or generators of the group $SU(3)$)—these are the Yang-Mills fields. They are called gluon fields. All the quarks possess a charge, determined by their interaction with the gluons. For the theory of such a system of colored quarks and gluon fields, describable by the Lagrangian (13), Gell-Mann has proposed the name "quantum chromodynamics."

The existence of asymptotic freedom implies that chromodynamics can lead to nontrivial solutions when arbitrarily short distances are taken into account. In the limit of short distances ($r_0 \to 0$) has effective charge $g(r_0) \to 0$; this means that inside the hadrons, in the region of small $r$, the quarks behave as free particles. This explains the experiments on deep-inelastic scattering of electrons and neutrinos, which, in a certain region of energies and momentum transfers, can be interpreted as scattering by independent point objects. The particular region of distances for which $g^2(r_0) \ll 1$ can be taken into account by using perturbation theory and the renormalization-group method. The converse aspect of asymptotic freedom is the growth of the interaction at large distances. Indeed, if we rewrite formula (14) in the form

$$g^2(r)=\frac{g^2(r_0)}{1-\gamma g^2(r_0)\ln(r/r_0)}, \quad r>r_0, \tag{17}$$

it is easy to see that $g(r)$ grows with increase of $r$. We do not know how to fix $r_0$ and $g(r_0)$ in this formula, or up to what values of $r$ it can be used. It is clear that there is a limit, since on decreasing the denominator we leave the region of applicability of the formula. Qualitatively, however, one can conceive that the interaction increases with distance. If this increase continues without limit, the quarks are locked in a potential well and cannot be torn out of it. This is the explanation of the unobservability fo free quarks. They can only be in a bound state, and such states are the hadrons. If the interaction potential of the atoms in a molecule were an exact oscillator potential, the atoms would also be unobservable.

The unobservability of quarks is intimately connected with the unobservability of "color." All the hadrons are "white" states, i.e., singlets under the group $SU(3)_{\text{col}}$.

At the present time great efforts are being made to find a serious justification for these ideas and for the confinement of quarks. In such situations, of course, illusions are possible. The difficulty of treating large distances (the "infrared catastrophe") remains.

But the problem of treating short distances is resolved. We imagine that any "true," i.e., dynamical, field theory is a theory using gauge vector fields. The other field theories that directly describe the hadrons and their interactions are phenomenological and should reduce, at a certain stage, to the theory of the interactions of quarks with gluons. If we compare the theory of the strong interactions with the nonrelativistic quantum mechanics of atomic systems, we can draw the following analogy. The dynamics of atomic systems is the dynamics of electrons with Coulomb interaction. The dynamics of hadron systems is the dynamics of quarks and gluons. In a certain approximation, from the dynamics of electrons there arise phenomenological potentials for the interactions of atoms, molecular forces, quasi-particles, etc. From the dynamics of quarks there arise the phenomenological theories of the interaction of hadrons, reggeons, etc.

Electrodynamics is also such a phenomenological theory, although it sounds blasphemous to characterize it in this way. Electrodynamics is valid in the pragmatic sense mentioned above—everywhere except at very short distances, where its application becomes invalid by virtue of the zero-charge theorem. But before this happens electrodynamics merges with the theory of the weak interaction, into a unified theory of interactions with nonabelian gauge fields. This happens at

distances at which the weak interaction is comparable in strength with the electromagnetic interaction.


[1]L. D. Landau, A. A. Abrikosov and I. M. Khalatnikov, Dokl. Akad. Nauk SSSR **95**, 497, 773, 1177; **96**, 261 (1964) (English translation in the "Collected Works of L. D. Landau," Gordon and Breach, N. Y., 1965).

[2]M. Gell-Mann and F. E. Low, Phys. Rev. **95**, 1300 (1954).

[3]L. D. Landau and I. Ya. Pomeranchuk, Dokl. Akad. Nauk SSSR **102**, 489 (1955) (English translation in the "Collected Works of L. D. Landau," Gordon and Breach, N. Y., 1965); I. Ya. Pomeranchuk, Sobranie nauchnykh trudov (Collected Scientific Works), Vol. 1, p. 163, "Nauka," M., 1972.

[4]E. S. Fradkin, Zh. Eksp. Teor. Fiz. **28**, 750 (1955) [Sov. Phys. JETP **1**, 604 (1955)].

[5]I. Ya. Pomeranchuk, Sobranie nauchnykh trudov (Collected Scientific Works), Vol. 2, p. 168, "Nauka," M., 1972.

[6]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[7]D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).

[8]I. B. Khriplovich, Yad. Fiz. **10**, 409 (1969) [Sov. J. Nucl. Phys. **10**, 235 (1970)].

[9]N. N. Bogolyubov and D. V. Shirkov, Vvedenie v teoriyu kvantovannykh poleĭ (Introduction to the Theory of Quantized Fields), "Nauka," M., 1973 (English translation of earlier edition published by Interscience, N. Y., 1959).

[10]A. A. Abrikosov and I. M. Khalatnikov, Dokl. Akad. Nauk SSSR **95**, 497 (1954).

[11]I. T. Dyatlov and K. A. Ter-Martirosyan, Zh. Eksp. Teor. Fiz. **30**, 416 (1956) [Sov. Phys. JETP **3**, 454 (1956)]; I. Dyatlov, V. Sudakov and K. Ter-Martirosyan, Zh. Eksp. Teor. Fiz. **32**, 767 (1957) [Sov. Phys. JETP **5**, 631 (1957)]; V. V. Sudakov, Dokl. Akad. Nauk. SSSR **111**, 338 (1956) [Sov. Phys. Dokl. **1**, 662 (1957)].

[12]A. I. Akhiezer and V. B. Berestetskiĭ, Kvantovaya elektrodinamika (Quantum Electrodynamics), "Nauka," M., 1969 (English translation of earlier edition published by Consultants Bureau N. Y., 1957).


Translated by P. J. Shepherd