SOVIET PHYSICS USPEKHI VOLUME 9, NUMBER 5 MARCH-APRIL 1967

530.145.1

*NONLOCAL QUANTUM FIELD THEORY**

D. A. KIRZHNITS

P. N. Lebedev Physics Institute, Academy of Sciences, U.S.S.R.

Usp. Fiz. Nauk 90, 129–142 (September, 1966)

NONLOCAL quantum field theory, which arose in its initial form [1] soon after the foundations of field quantization were laid and the difficulties with divergences became obvious, has always been regarded as one of the approaches in the theory of elementary particles most lacking in consistency and most remote from reality. This opinion has been based on the many difficulties inherent in nonlocal theory, along with the lack of any apparent points of contact of the idea of nonlocality with experiment. Except for a small number of enthusiasts for nonlocal theory, people have mentioned it only when the difficulties with the divergences of local theory have taken a particularly sharp form (cf., e.g., [2]).

This situation has changed decidedly in the last few years. First, we cannot fail to point out the appearance of a definite practical interest in nonlocal theory, which is stimulated by experiments being done and being planned to test the dispersion relations and to determine the limits of applicability of quantum electrodynamics. In particular, there has been wide discussion of the question as to whether the form of the dispersion relations is changed when we go over to a nonlocal theory; in other words, whether an affirmative result of the test of these relations will mean an actual proof of the local character of the interaction of elementary particles.[3–8] Nonlocal theory is used especially often—though in a rather imperfect form—in the planning and in the processing of results from experiments to test quantum electrodynamics.[9–12]† Finally, special attention is directed to nonlocal theory by the very possibility that the results of these experiments may fail to agree with the predictions of local theory (of course when all possible masking effects are taken into account).‡

In parallel with the awakening interest in nonlocal theory there has been a change in the estimation of its inherent difficulties. It is now supposed that these difficulties are, at least predominantly, not matters of principle, but that they arise because of a too straightforward extension of the apparatus of local field theory. The possibility of overcoming these difficulties can be illustrated with models of nonlocal theory which are internally selfconsistent and compatible with the principles of quantum field theory.[15–17] The only problem not fully solved is that of macroscopic causality. There is still no positive solution for it, though it is possible to refute the direct arguments that have been given in the literature against the possibility of satisfying the conditions in question (for details see Secs. 3 and 9).

This article contains a brief survey of the present state of nonlocal field theory.

1. The local theory of elementary particles is based on the following three fundamental postulates. First, there is the relativistic postulate, which defines geometrical and kinematical aspects of the description of elementary particles and their interactions. Next, the quantum nature of elementary particles is postulated and a probabilistic interpretation of the processes of interaction and interconversion among them is formulated. The third postulate stands somewhat apart—the postulate of the local (point) character of the interaction.* Being borrowed from the classical theory of point particles, in its simplest formulation it requires that the field operators which occur in a product in the operator for a physical quantity be referred to the same point of space-time. For example, the local action function for the interaction of electrons with the electromagnetic field is written in the form

$$S = e \int dx\, \overline{\psi}(x)\, \gamma_\mu \psi(x)\, A_\mu(x). \tag{1}$$

In a more general sense the postulate of locality expresses the requirement of microscopic causality (the cause event must always precede the effect event). It must be pointed out, however, that it is much inferior to the first two postulates in clarity of physical content, and moreover shows a definite lack of physical correspondence to them (for details see Sec. 3). Therefore the locality postulate, being to a large extent a formal requirement, is one of the most

*Expanded text of a survey report at a session of the Nuclear Physics Section of the Academy of Sciences of the U.S.S.R. (November, 1965).

†A comparison of the nonlocal correction to the matrix element with the quantity characterizing the accuracy of the experiment allows us to compare the effectiveness of various experiments, and also to determine an upper limit on the elementary length from the experimental data that are already available.

‡The first communications about this sort of disagreements to appear in the literature [13,14] are still in need of serious checking and improvements in accuracy.

*In the S-matrix method developed in recent years the postulate of locality is replaced by the requirement of maximum analyticity of the matrix elements.

vulnerable of the assumptions forming the basis of local elementary-particle theory.

The opinion is often expressed that the locality postulate cannot be regarded as independent of the relativistic postulate, since violation of locality is in logical contradiction with relativistic kinematics. The argument usually given for this point of view is that renunciation of locality means (at least in non-quantum theory) the appearance of signals faster than light; this is regarded as in contradiction with the kinematics of relativity theory.

Actually the requirement that there be no signals faster than light follows not from relativistic kinematics, but from the additional condition of causality, which is largely empirically based. In itself relativistic kinematics contains no limitations on the speed of a signal.* In this connection it is important to note that there does not exist any reference system in which the speed of an ultralight signal would be zero (as is also true of the speed of a photon). Therefore the speed of an inertial reference system, as involved in the Lorentz transformation, is always less than the speed of light.

We refer to the detailed discussion of this range of questions in [19], and give here a relevant statement by Einstein himself. In discussing the possible change of the time order of events by Lorentz transformations when the events are connected by an ultralight signal ($w > c$), Einstein formulates his conclusion that there are no ultralight motions in the following way: "Although from a purely logical point of view this result does not, in my opinion, contain any contradictions, it nevertheless so contradicts the nature of all our experience that the impossibility of the conjecture $w > c$ seems sufficiently well established."[20] It is superfluous to say that Einstein could have in mind only the macroscopic experience that had been accumulated up to the beginning of our century.

2. Under the general head of nonlocal field theory there are quite a number of theoretical schemes, which are very different in their initial ideas and in the way they are realized. They are unified by the common assumption that the divergence difficulties of local theory indicate the faultiness of the third of the postulates we have listed—the postulate of locality.†

The departure from the point character of the interaction can be accomplished in various ways. We here distinguish two main approaches. The first is characterized by the bringing in of additional postulates, which make the very concept of locality of the interaction meaningless, in the same way as the concept of a trajectory loses its meaning for a quantum-mechanical object. This approach includes in particular theories which start from the impossibility of giving an exact meaning to the concept of a field referred to a point of space-time; the corresponding postulate is that field operators and coordinates do not commute.[21,22] Another possibility, connected with the assumption that the very concept of a definite point of space-time is without exact meaning (non-commutation of the components of the coordinate operator), leads to the theory of quantized space-time.[23-26] There are as yet no clear prospects for such "physical" nonlocal theories, and they will not be considered in this survey (for a rather complete bibliography of papers on this question see [27]).

Over against the "physical" nonlocal theories there are the phenomenological schemes in which the departure from locality is made without bringing in new physical ideas, and which are based as before on the space-time ideas of the theory of relativity and the probability ideas of quantum mechanics. The models of such phenomenological nonlocal theories are usually constructed by introducing into the local expressions some prescribed functions of the coordinates or of the momenta—form-factors. The simplest way to introduce a form-factor is artificially to "shift apart" the field operators—to refer them to different space-time points. For example, the action function (1) is replaced by the quantity

$$e \int dx\, dy\, dz F(x, y, z)\, \bar{\psi}(x)\, \gamma_\mu \psi(y)\, A_\mu(z), \tag{2}$$

where F is the form-factor.

If the future consistent theory of elementary particles is indeed a nonlocal theory (and in favor of this we have, besides the divergences, also the analysis of the procedure for measuring coordinates), then it is extremely likely that besides changes in the postulate of locality changes in the other two postulates will also be required. Therefore phenomenological nonlocal theory can primarily pretend only to be a crude preliminary description of the nonlocal effects. A virtue of this sort of theory, as of all phenomenology, lies in a certain generality, its independence of the concrete realizations of the idea of nonlocality, whose true content is still in the highest degree mysterious. It is important that the typical difficulties inherent in nonlocal theory are of rather general character and are already manifested in a phenomenological theory. It is natural to conduct an analysis of these difficulties and a search for ways to overcome them in the framework of a general phenomenological approach.

3. Denial of the locality of the interaction inevitably leads to violation of the condition of microscopic causality, which requires that there be no influence of a point event in the future on an event which is al-

*This is already evidenced by the very fact that is is possible to give relativistic formulations of quite a number of nonlocal and nonlinear schemes which lead to ultralight signals in the nonquantum case.

†As is well known, the divergences of local theory are explicitly connected with the equality of the arguments of the field operators in the intersection Lagrangian.

ready past, and accordingly that there be no mutual influence between events separated by a spacelike interval. The simplest formulation of this condition is

$$[\varphi(x), \varphi(y)] = 0, \quad (x_0 - y_0)^2 < (\mathbf{x} - \mathbf{y})^2, \qquad (3)$$

where φ is the Heisenberg field operator (for definiteness, of a Bose field).

It is essential to point out that the condition of microscopic causality has no direct physical meaning in relativistic quantum theory and is essentially no more than a mathematical extrapolation of the classical causality condition. It suffices to note that the very concept of a point event, which appears in the formulation of the condition in question, is incompatible with relativistic quantum ideas.* In other words, it is impossible even in imagination to realize any experimental situation which would allow us to establish a violation of causality in arbitrarily small regions of space-time (on this point see [28-31]).

Therefore the impossibility of securing the condition of microscopic causality in a nonlocal theory provides no physical arguments against such a theory (to say nothing of the fact that in a future theory such concepts as "earlier" and "later" and so on may lose their simple meanings altogether). The only important thing is that there be no violations of causality on a large scale in space and time, where causality is in all cases verified by experience. The possibility of satisfying this last condition (the condition of macroscopic causality) in a nonlocal theory will be discussed below.

At first glance the principle of macroscopic causality is deprived of value by the fact that the values of space and time intervals taken separately are not relativistically invariant. Indeed, if we confine the violation of causality to an invariant region

$$|(x-y)^2| = |(x_0 - y_0)^2 - (\mathbf{x} - \mathbf{y})^2| < l^2, \qquad (4)$$

where l is some small length, then in directions close to the light cone the spatial and temporal extents of the acausal region, taken separately, are arbitrarily large.

This fact has led a number of authors (cf., e.g.,[34-36,14]) to introduce into nonlocal theory a supplementary timelike vector P_μ, by means of which one can "localize" the violation of causality. If, for example, we replace (4) by the condition

$$\left|(x-y)^2 - \frac{[(x-y, P)]^2}{2P_\mu^2}\right| < l^2, \qquad (5)$$

then the quantities $x_0 - y_0$ and $\mathbf{x} - \mathbf{y}$ will also be limited when taken separately. The vector P_μ can be related either to empty space, to the vacuum, or to the system of interacting particles itself. The former case leads to a violation of relativity (the reference system in which $\mathbf{P} = 0$ is singled out) and goes beyond the framework of the relativistic nonlocal theories considered in this article.* In the latter case P_μ can be identified, for example, with the energy-momentum vector of the system of particles. Then, however, we encounter a strong violation of macroscopic causality. In fact, the merely conceptual inclusion in the composition of the system of any additional particles, located at no matter how large distances, changes the quantity P_μ, and thus changes the nature of the interaction of the particles of the original system. It is clear that such action at a distance is physically inadmissible.[16]

Meanwhile, even without introducing an auxiliary vector we can make the acausal influence we spoke of in connection with Eq. (4) fall off rapidly enough as we go away from the vertex of the light cone. It follows from a special analysis of this question[32,33] that this decrease is exponential if certain conditions imposed on the actual matrix element of a process in the momentum representation are satisfied. Namely, it is sufficient that the expression in question not contain any singularities on the real axis with non-Feynman (acausal) rules for avoiding them.† In explanation we note that the appearance of a pole on the real axis means that the matrix element, like an ordinary propagation function, is capable of describing the transfer of an interaction over macroscopic distances and times.

4. The demarking of space-time into "small" regions, where causality is violated, and "large" regions, where it is satisfied, is impossible without the appearance in the nonlocal theory of a new constant of the dimension of length—an elementary length. In phenomenological nonlocal theory the form-factor must contain this quantity.

The elementary length l should become the third fundamental physical constant, after the speed of light c and the Planck constant $\hbar$. The necessity of its appearance in the theory of fundamental particles is generally recognized, but we can as yet only guess about the value of this constant.[37] ‡

It is interesting to put the three fundamental constants c, $\hbar$, and l in correspondence with the three basic postulates of local theory, of which we spoke in Sec. 1. Since the constants c and $\hbar$ are associated with the first two postulates, which arose through

*It is well known that one cannot construct from particle states on the mass surface and with the positive sign of the energy a wave packet which is a point packet not only in the spatial but also in the temporal sense.

*Formally it is possible to preserve the corresponding group properties if we assign to each reference system its own vector P_μ.[36] Such a scheme, however, will differ rather strongly in its conclusions from orthodox relativistic theory.

†See also [30]; regarding a possible weaker form of this condition see below, Sec. 7.

‡From experiments made to test quantum electrodynamics we have the estimate $l < 10^{-14}$ cm.

abandonment of the corresponding classical ideas, it seems natural to suppose that the constant l must appear when the third postulate, the postulate of locality, is abandoned. It is this assumption that is the basis of nonlocal field theory.

There is usually associated with the appearance of the elementary length the hope for convergence of the integrals that become infinite in local theory: it is expected that these integrals will be automatically "cut off" at a momentum value $\lambda \sim \hbar/l$. Therefore nonlocal theory is often regarded as a method for regularizing divergences. Such a regularization is necessary for a consistent treatment of problems of elementary-particle physics, in particular the physics of weak and other nonrenormalizable interactions.

In reality "cutting off" of integrals in nonlocal theory by no means always occurs. Although in this theory the field operators are indeed referred to noncoincident points of space-time, this leads to convergence only for integrals with respect to the four-dimensional squares of the momenta. There still remains an integration over the (infinite) angular region of the pseudoeuclidean momentum space. In local theory this integration always leads to a finite result; the presence of the Feynman rules for going around singularities allows us to go over to a Euclidean space, where there are clearly no angular divergences. In nonlocal theory, however, there microscopic causality is violated, the transition to Euclidean space is impossible, and there is an extremely sharp question of angular divergences [32] (cf. Sec. 5).

5. The *questions of macroscopic causality* and convergence which have already been touched on in the preceding sections give some idea of the characteristic difficulties inherent in nonlocal theory.

The difficulties of nonlocal theory, which have often even led to the conclusion that it is completely inconsistent, penetrate literally into all aspects of the theory. Among them are questions connected with relativistic invariance (ordering with respect to time,[38] the condition of mathematical compatibility [39, 38]), questions relating to the quantum-mechanical description (unitarity of the scattering matrix,[40,41] definiteness of the metric [42]), questions of the convergence of matrix elements,[32,43-45] gauge invariance (in electrodynamics),[32,46] questions of macroscopic causality,[32,33,47-49] and a number of others.

At present there is every reason to believe that these difficulties are not as serious as they had usually been thought to be. One of the general reasons for their appearance is that mutually equivalent formulations of local theory have turned out to be very different in regard to their nonlocal extensions (the introduction of form-factors). The roots of most of the difficulties are precisely in an unhappy choice of the initial local formulation. Accordingly it has been possible to overcome these difficulties owing to a special choice of this formulation.

A detailed study of the causes for the appearance of these difficulties [15-17] has led to the finding of ways to overcome them and construct examples of noncontradictory nonlocal schemes; in Sec. 9 we shall discuss the problem of macroscopic causality, which still remains without a positive solution. It not being possible to go into details on the various questions here, we shall give the examples in question in Secs. 6—8.

It is important to point out that a version of nonlocal theory which is discussed in the literature more often than others—the introduction of a form-factor into the vertex part of the Feynman diagram (cf. [2])—is inconsistent. In the matrix elements in question singularities appear at certain fixed values (which do not go to infinity for $l \to 0$) of the external momenta, and the rules for passing around these points are nonfeynman.[15] These singularities are "vestiges" remaining from the divergences of the local expression, and arise because for these values of the momenta the form-factor ceases to "cut off" the divergent integrals. As has already been indicated in Sec. 3, such singularities lead to violation of macroscopic causality, and it is their appearance that led to the well known conclusion that it is impossible to apply the principles of causality and unitarity in nonlocal theory. The refutation of this conclusion[17] consisted precisely in the construction of a version of nonlocal theory in which there are no such singularities.

In addition to the difficulties with macroscopic causality, there are angular divergences in the type of nonlocal theory we have been discussing. They are typical, in particular, of those matrix elements which in a local nonrenormalizable theory diverged more rapidly than logarithmically (cf. Sec. 4). Paradoxical as it may be, the situation with divergences is here even worse than in the case of a local renormalizable theory. If, for example, in the local theory the main divergence of a polarization operator took the form of an infinite constant, easily removable by a mass renormalization, in the nonlocal case the corresponding expression converges for timelike values of the boson momentum and diverges for spacelike values.[32] This fact was regarded earlier as a most serious difficulty of nonlocal theory, and even instigated a rejection of relativism.[50] This is too high a price, however, for the convergence of matrix elements, the more so because in reality the choice of a different way of constructing the nonlocal theory suffices to get rid of the difficulties in question.

6. We shall give the simplest example of a nonlocal theory free from the difficulties enumerated above.[16] Although we do not possess a proof that the condition of macroscopic causality is satisfied in its space-time sense, here a different condition is satisfied which can be expected to have nearly the same meaning. Namely, the analytic properties of the

matrix elements are exactly the same as in the local theory, as long as the values of the energy, the momentum transfer, and so on do not exceed a limiting value determined by the elementary length.

As was already emphasized in the preceeding section, the introduction of a form-factor in the vertex part does not lead to success. Let us try to introduce it in the propagation function. For this purpose it is convenient at an intermediate stage to use states with an indefinite metric. Following the method developed in [51,49], we extend the Hilbert space of the states, including besides the physical in-states φ_{in} states with negative norm $\chi^{\kappa}_{\text{in}}$, corresponding to the mass κ. The complete in-field operator is written in the form

$$\tilde{\varphi}_{\text{in}} = \varphi_{\text{in}} + \int\limits_{\Lambda^2}^{\infty} d\kappa^2 C(\kappa^2)\, \chi^{\kappa}_{\text{in}}, \quad \Lambda \sim \frac{\hbar}{l}.$$

By imposing the well known Pauli-Villars conditions on the function $C(\kappa^2)$, we can easily assure that the commutators and Green's functions of the operator $\tilde{\varphi}_{\text{in}}$ have no singularities on the light cone.

Let us now specialize the type of interaction, and take the interaction Lagrangian for simplicity in the form $L_{\text{in}} = g\colon \varphi^3_{\text{in}}\colon$ (g is the charge, and renormalization terms are omitted). In the complete space the scattering matrix S is local and is given by the usual expression

$$\tilde{S} = T \exp\left(i \int dx\, L_{\text{in}}(x) \right). \tag{6}$$

Differentiating it with respect to the charge, we arrive at the equation

$$g \frac{d\tilde{S}}{dg} = i\tilde{S} \int dx\, L(x), \tag{7}$$

where $L(x) = \tilde{S}^{+} T[L_{\text{in}}(x)\tilde{S}]$ is the local Lagrangian in the Heisenberg representation. It is easily seen that it satisfies the equation

$$g \frac{dL(x)}{dg} = i \int dy\, \theta(x-y)\, [L(x), L(y)] \tag{8}$$

with the initial condition $L|_{g\to 0} \to L_{\text{in}}$ (cf. [52]). We emphasize that we are so far dealing with a purely local theory including states with an indefinite metric.

To obtain the scattering matrix in the physical space we introduce the operator of projection onto this space, P,* and in analogy with (7) we define the physical scattering matrix S by the equation

$$g \frac{dS}{dg} = iS \int dx\, PL(x) P.$$

Integrating this equation with respect to the charge and introducing the symbol $\tilde{T}_g$, which means that the quantities to which it is prefixed are "antiordered" with respect to charge (arranged in order of charges increasing from left to right), we finally get

$$S = \tilde{T}_g \exp\left[i \int\limits_0^g \frac{dg}{g} \int dx\, PL(x) P \right]. \tag{9}$$

The replacement of ordering with respect to time by ordering with respect to charge has a meaning in principle in the nonlocal theory, since in this case equations of the type of (6) are not relativistically invariant.[38]

After the projection on the physical space the states with negative norm drop out, but then the theory becomes nonlocal. This can be seen from the fact that the commutator of operators PL(x)P is

The scattering matrix (9) is relative to the mass shell and describes the complete evolution of the system in the interval $-\infty < t < \infty$. By bringing this matrix beyond the limits of the mass shell in one way or another, one can get an even more detailed space-time description.[15] If, for example, we add a small classical term to the operator φ_{in}, then, by introducing the current operator

$$j(x) = iS^{+} \frac{\delta S}{\delta \varphi_{\text{in}}(x)},$$

we can easily get the Lagrange equations of the nonlocal theory.

If, on the other hand, we put an upper limit on the integration over x in (9) with a spacelike hypersurface σ, then, defining the interaction Hamiltonian by the condition

$$H(x) = i \frac{\delta S}{\delta \sigma(x)} S^{+} = -\int\limits_0^g \frac{dg}{g} SL(x) S^{+}, \tag{10}$$

we can arrive at the Hamiltonian description. It is important that this Hamiltonian automatically satisfies the Bloch compatibility condition

$$[H(x), H(y)] = i \left[\frac{\delta H(y)}{\delta \sigma(x)} - \frac{\delta H(x)}{\delta \sigma(y)} \right],$$

which is not the same as the condition of macroscopic causality, $[H(x), H(y)] = 0$ [cf. (3)], which is always violated in nonlocal theory, if the Hamiltonian depends explicitly on the surface σ; and indeed this is the case with the Hamiltonian (10). Hamiltonians which do not depend on σ are of no use in nonlocal theory, and this explains the lack of success of many earlier attempts to construct such a theory.

7. Returning to the expression (9) for the scattering matrix, let us convince ourselves that it is free from the difficulties enumerated earlier.

Its relativistic invariance follows directly from its construction. It is a nontrivial point that the function $\theta(x-y)$ is encountered only in combination with the commutator of the local operator $L(x)$, which vanishes outside the light cone [cf. (8)]. Therefore no Lorentz transformation can change the time ordering of the operators that occur in (8) (cf. [38]).

The unitarity of the matrix (9) in the physical space is a consequence of the Hermitian character of the operator PL(x)P. The indefinite metric is in-

*The quantity $P : \tilde{\varphi}_{\text{in}} \ldots \tilde{\varphi}_{\text{in}} : P$ is by definition equal to the normal product of the physical operators, $: \varphi_{\text{in}} \ldots \varphi_{\text{in}} :$.

troduced only in the intermediate developments and is eliminated from the final expression. An expression which is in a certain sense complementary to (9) is

$$S = P\widetilde{T}_g \exp\left(i \int_0^g \frac{dg}{g} \int dx\, L(x) \right) P,$$

which is (with unimportant changes) the expression for the scattering matrix as regularized by the Pauli-Villars method. This expression is local, but not unitary.

As is well known, the divergences of the local theory have their origin in the singularity of the retarded Green's function. As can be seen from (8), in our case this function appears always in regularized form. The absence of angular divergences can also be illustrated by noting that the Green's function of the operator $\widetilde{\varphi}_{in}$, which in our case plays the role of an effective form-factor, has the Feynman rules for going around singularities. Therefore in the part of the matrix element which is dangerous from the point of view of divergences we can make the transition to Euclidean space (see Sec. 4 and [16]). For the same reasons the matrix elements of (9) will not contain any singularities for fixed momenta, such as were spoken of in Sec. 5.*

We shall not here go into narrower questions (conservation laws, dynamical variables, and so on), which can be given positive solutions,[16] nor into the problem of macroscopic causality.

Although the matrix (9) is defined by extremely simple and compact equations, there is a different, though indeed very similar, version of nonlocal theory which is more convenient for practical purposes. We shall formulate directly a diagram technique corresponding to this version.

The real part of the matrix element is identical with the real part of the local matrix element regularized by the Pauli-Villars method. The imaginary part, which is chosen according to the unitarity condition, is obtained from the imaginary part of the regularized matrix element by striking out all terms that are of a threshold character in the auxiliary masses κ.† For example, in the diagram for the self energy of the particle we must strike out terms of the forms

$$\int_{\Lambda^2}^{\infty} d\varkappa^2 C(\varkappa^2)\, \theta(k^2 - 4\varkappa^2)(\ldots), \qquad \int_{\Lambda^2}^{\infty} d\varkappa^2 C(\varkappa^2)\, \theta(k^2 - (\varkappa+\mu)^2)(\ldots).$$

For both of these types of nonlocal theory the analytic properties of the matrix elements are the same; these properties are the same as in the local theory, as long as none of the kinematical invariants of the process, the quantities s, t, etc., exceeds a certain limiting value determined by the quantity $\Lambda^2 \sim \hbar^2/l^2$. When this condition is violated additional singularities appear, which were not present in the local theory (in the simplest diagrams, additional "distant" cuts with non-Feynman rules for the paths).

In a corresponding way the dispersion relations at energies less than the threshold energy differ from the usual ones only in that the absorptive part is not equal to the imaginary part (nor, in particular, for forward scattering, to the total cross section) for energies larger than the threshold value. With a reasonable choice of the elementary length such dispersion relations are not in contradiction with existing experimental data (cf. the analogous analysis in [35]).

8. In the construction of a nonlocal electrodynamics there are additional difficulties associated with gauge invariance.

The roots of these difficulties lie in the fact that to secure gauge invariance the operators for the momentum $\hat{p}$ and the potential $\hat{A}$ must appear in physical quantities only in the form of the combination $\hat{p}_\mu - eA_\mu$. Since the form-factor $F(x - y)$ introduces into the theory an additional dependence on the momentum,* it must be accompanied by an additional function of the potential [53]

$$E(x, y) = \exp\left[ie \int_x^y d\xi_\mu A_\mu(\xi) \right].$$

When this condition is taken into account the expression for the action function for electrons interacting with the electromagnetic field is of the form [46]

$$-\int dx\, dy \bar{\psi}(x) E(x, y)(\gamma\hat{p}_x - m) F(x-y)\psi(y).$$

Here, however, there is such a strong nonlinearity that additional divergences appear, caused precisely by this nonlinearity (on this point see [54, 55]). This in turn has led to much lack of confidence in the possibility of constructing a gauge-invariant nonlocal electrodynamics, and has even led to a discussion of the question as to the experimental observation of the emission of longitudinal and scalar photons at large energies.[56]

Recently it has been common practice to formulate gauge-invariant local electrodynamics by directly replacing the quantity A_μ in the expression $\int dx\, j_\mu(x) A_\mu(x)$ by its transverse part $A_\mu - A_\mu^l$ [$A_\mu^l = (k_\mu k_\nu/k^2) A_\nu$]. Then because of the conservation of the current j_μ the quantity A_μ^l indeed gives no contribution. Recently a number of papers have appeared [57-59] in which this method is carried over to

*At the same time, there appear in the theory "distant" cuts (which recede to infinity for $l \to 0$) with nonfeynman rules for the path. As is shown in [49], by a suitable choice of the function $C(\kappa^2)$ we can nevertheless bring it about that the acausal influence function falls off exponentially (cf. Sec. 3).

†These terms describe the contribution of nonphysical intermediate states and violate unitarity.

*The quantity $\int dy F(x-y)\,\psi(y)$ can always be represented in the form $\tilde{F}(p)\,\psi(x)$, where $\tilde{F}(p)$ is the Fourier transform of the function $F(x-y)$.

nonlocal electrodynamics. Here, however, because of the explicit nonconservation of the nonlocal current the quantity A^l_μ does not drop out of the result. Therefore the remaining singularity of A^l_μ at $k^2 = 0$ leads to violation either of macroscopic causality or of unitarity, (depending on the choice of the rules for the path around this singularity).

The construction of a nonlocal electrodynamics satisfying the condition of gauge invariance is possible in the framework of the scheme expounded in the previous sections. It suffices to set

$$L_{\text{in}} = e\widetilde{A}_{\mu,\,\text{in}}\widetilde{j}_{\mu,\,\text{in}},$$

$$\widetilde{A}_{\text{in}} = A_{\text{in}} + \int_{\Lambda^2}^{\infty} d\varkappa^2 C(\varkappa^2)\, A^{\varkappa}_{\text{in}}; \quad \widetilde{j}_{\text{in}} = j_{\text{in}} + \int_{\Lambda}^{\infty} dM\, C'(M)\, j^M_{\text{in}}$$

(here the quantity e plays the role of the charge g in the formulas of Sec. 6). It is important that the current is regularized as a whole, and not the separate operators $\overline{\psi}$, ψ; this leads to the conservation of the nonlocal current. In the diagram technique of the version considered in Sec. 7, matters reduce to the carrying out of the Pauli-Villars regularization for entire electron path with a single auxiliary mass for all the lines of the path (cf. [60]).

The most conspicuous consequence of such a theory, and one that is of interest from the point of view of testing quantum electrodynamics, is the decided difference between processes of the type of the Compton effect and pair creation and annihilation and processes of the type of the scattering of electrons by electrons or positrons.[61] In processes of the first type there are no virtual photon lines in the lowest order of perturbation theory, and according to what has just been said a virtual electron line must have the same mass as the free ends. Therefore the corresponding matrix element differs from the local expression only in higher orders of perturbation theory (through radiative corrections). The situation is different for processes of the second type, where there is a virtual photon line and where the difference appears in the lowest order of perturbation theory. This fact can be regarded as an argument in favor of the greater effectiveness of this type of scattering processes from the point of view of testing quantum electrodynamics. Of course, this argument is not extremely convincing; it depends, in particular, on the possibility of constructing other gauge-invariant nonlocal schemes (cf. in this connection [62,63]).

As is well known, in the development and planning of experiments to test quantum electrodynamics different ways of introducing form-factors are used in parallel—in the vertex part, in the propagation function, and so on. A consistent nonlocal electrodynamics rejects some of these possibilities. For example, for experiments to determine the anomalous magnetic moment of the muon the only consistent way is to introduce the form-factor in the photon propagator; this gives a somewhat larger upper limit on the elementary length than the other possibilities (cf. [11]).

9. In conclusion we shall consider briefly the problem of macroscopic causality. The point is that a nonlocal theory must not lead to experimentally observable consequences which contradict causality on a large scale of space and time—appearance of a scattered wave before the incident wave has got to the scatterer, ultralight signals,* and so on.

At one time there were assertions that it is impossible to reconcile the conditions of macroscopic causality and unitarity in a nonlocal theory.[47,49] The basis of that discussion was a criterion of macroscopic causality which used the concept of the scattering matrix outside the mass shell, and in which the interaction was localized in an artificial way in a small space-time region (in this connection see also [65]). The very difficulty of understanding the connection of this sort of approach with the description of actual conditions of a physical experiment makes the conclusions not too convincing. Moreover, it has been shown directly [17] that the conclusions of [47] are due to an unfortunate choice of the real part of the matrix element considered there (see also Sec. 5). As for [49], which contains a sufficiently general treatment, it involves an essentially superfluous requirement that the scattering matrix be unitary when taken beyond the limits of the mass shell in the way indicated. Meanwhile, it is always possible to add to a scattering matrix which is unitary on the mass shell terms which vanish on the shell and are such as to secure the validity of the condition of macroscopic causality in the formal form in which it was used in the papers in question.

The actual proof of the macroscopic causality property of nonlocal theory is a very difficult question and still far from a complete solution. Here it is first of all necessary to convince ourselves that there are no long-range acausal influences such as we spoke of in Sec. 3. This in turn requires a study of the analytic properties of the nonlocal matrix elements, which in itself is a rather complicated task. A study of this sort for the model of Sec. 6, which led to affirmative results, was made for diagrams containing not more than four external lines. The hope of an affirmative solution of this question for diagrams of more complicated types is based in the last analysis on the fact that as the elementary length goes to zero the theory in question goes over into the local theory, and consequently into a causal theory, while the analytic properties of local theory are recovered already for finite l (cf. beginning of Sec. 6).

If the acausal influences in nonlocal theory are

*A special investigation of the possibility of the macroscopic propagation of ultralight signals is contained in [64].

actually of short range, the quantitative question arises as to the permissible degree of their damping with distance. The solution of this problem is very much impeded by the fact that there is so far no opposite criterion expressed in the language of observable quantities only and with proper correspondence to the conditions of physical experiments (cf. discussion in[66]). A very similar and also still unsolved problem is that of deriving the analytic properties of local matrix elements from the physical criterion of macroscopic causality. This problem has recently been receiving more and more attention.[29,30,67-69]

If we take an optimistic view of the possibility of satisfying the condition of macroscopic causality in nonlocal theory—and no grounds can be perceived for the opposite view—then we can evaluate nonlocal theory as a consistent apparatus capable of competing with local theory and free from the difficulties of the latter.

Of course the answer to the main question, as to how nature is actually constructed—in a local or in a nonlocal way—can be given only by nature itself. The deciding word in settling this question, the most important in the physics of elementary particles, belongs to experiment.

[1] G. Wataghin, Z. Physik 88, 92 (1934).
[2] L. D. Landau, in: Niels Bohr and the Development of Physics, New York, McGraw-Hill, 1955.
[3] R. Oehme, Phys. Rev. **100**, 1503 (1955).
[4] B. L. Ioffe, JETP **31**, 583 (1956), Soviet Phys. JETP **4**, 534 (1957).
[5] G. Allcock, Nucl. Phys. **14**, 177 (1959).
[6] A. Wightman, in: Les Problémes mathêmatiques de la théorie quantique des champs, Paris, 1959.
[7] D. Ya. Petrina, Ukr. matem. zh. **13**, 109 (1961).
[8] L. V. Prokhorov, JETP **45**, 791 (1963), Soviet Phys. JETP **18**, 542 (1964).
[9] S. Drell, Ann. Phys. (N. Y.) **4**, 75 (1958).
[10] V. N. Baĭer, UFN **78**, 619 (1962), Soviet Phys. Uspekhi **5**, 976 (1963).
[11] Proc. Conf. on Photon Interaction MIT, 1963.
[12] Proc. Symp. on Electron and Photon Interactions, Hamburg, 1965.
[13] R. Blumental, D. Ehn, W. Faissler, P. Joseph, L. Lanzerotti, F. Pipkin, and D. Stairs, Phys. Rev. Letters **14**, 660 (1965).
[14] D. I. Blokhintsev, Preprint P-2422, Dubna, 1965. V. S. Barashenkov, Preprint P-2397, Dubna, 1965.
[15] D. A. Kirzhnits, JETP **41**, 551 (1961) and **45**, 144, 2024 (1963), Soviet Phys. JETP **14**, 395 (1962) and **18**, 103, 1390 (1964).
[16] A. N. Leznov and D. A. Kirzhnits, JETP **48**, 622 (1965), Soviet Phys. JETP **21**, 411 (1965).
[17] A. N. Leznov, JETP **49**, 784 (1965), Soviet Phys. JETP **22**, 545 (1966).
[18] D. I. Blokhintsev, UFN **61**, 137 (1957).
[19] D. A. Kirzhnits and V. L. Polyachenko, JETP **46**, 755 (1964), Soviet Phys. JETP **19**, 514 (1964).
[20] A. Einstein, Collected Works (Russian transl.), Nauka, 1965, Vol. I, page 76.
[21] M. A. Markov, JETP **10**, 1311 (1940).
[22] H. Yukawa, Phys. Rev. **77**, 219 (1950).
[23] H. Snyder, Phys. Rev. **71**, 38 (1947).
[24] Yu. A. Gol'fand, JETP **37**, 504 (1959) and **43**, 256 (1962), Soviet Phys. JETP **10**, 356 (1960) and **16**, 184 (1963).
[25] V. G. Kadyshevskiĭ, JETP **41**, 1885 (1961), Soviet Phys. JETP **14**, 1340 (1962); DAN SSSR **147**, 588 (1962), Soviet Phys. Doklady **7**, 1031 (1963).
[26] I. E. Tamm, Report at 12th International Conf. on High-energy Physics, Dubna, 1964.
[27] A. N. Vyal'tsev, Diskretnoe prostranstvo-vremya (Discrete Spacetime), Moscow, Nauka, 1965.
[28] M. Fierz, Helv. Phys. Acta **23**, 731 (1950).
[29] A. Bohr, Lectures in Theoret. Phys., vol. 3, N. Y., 1961.
[30] G. Wanders, Nuovo cimento **14**, 168 (1959).
[31] A. Ferretti, Nuovo cimento **27**, 1503 (1963).
[32] C. Bloch, Kgl Danske Mat.-Fys. **27**, 8 (1952).
[33] M. Chrétion and R. Peierls, Nuovo cimento **10**, 688 (1953).
[34] G. Wataghin, Nuovo cimento **25**, 1383 (1962).
[35] D. I. Blokhintsev and G. I. Kolerov, Nuovo Cimento **34**, 163 (1964).
[36] R. Ingraham, Nuovo cimento **39**, 361 (1965).
[37] M. A. Markov, Neĭtrino (Neutrinos), Moscow, Nauka, 1964.
[38] M. A. Markov, Giperony i K-mezony (Hyperons and K Mesons), Moscow, Fizmatgiz, 1958.
[39] M. A. Markov, UFN **51**, 317 (1953).
[40] B. V. Medvedev, DAN SSSR **100**, 433 (1955).
[41] C. Hayashi, Progr. Theor. Phys. **10**, 533 (1953); **11**, 266 (1954).
[42] A. Pais and G. Uhlenbeck, Phys. Rev. **79**, 145 (1950).
[43] J. Rzewnski, Acta Phys. Polon. **11**, 9 (1951).
[44] J. Rayski, Acta Phys. Polon. **11**, 25 (1951).
[45] C. Moller, Proc. Conf. Theor. Phys., Kyoto—Tokyo, 1953.
[46] M. Chrétien and R. Peierls, Proc. Roy. Soc. **A223**, 468 (1954).
[47] E. Stueckelberg and G. Wanders, Helv. Phys. Acta **27**, 607 (1954).
[48] M. Ebel, Kgl Danski Mat.-Fys. **29**, 2 (1954).
[49] D. A. Slavnov and A. D. Sukhanov, JETP **36**, 1472 (1959), Soviet Phys. JETP **9**, 1044 (1959).
[50] W. Heitler, XII Conseil de Solvay, 1962.
[51] N. N. Bogolyubov, B. V. Medvedev, and M. K. Polivanov, Nauchn. dokl. vyssh. shkoly (fiz.-matem. nauki) **2**, 137 (1958).
[52] D. A. Kirzhnits, JETP **49**, 1544 (1965), Soviet Phys. JETP **22**, 1059 (1966).
[53] D. A. Kirzhnits, JETP **21**, 949 (1951).

[54] E. S. Fradkin, Nucl. Phys. **49**, 624 (1963).
[55] G. V. Efimov, Phys. Letts. **4**, 314 (1963).
[56] V. S. Barashenkov, JETP **31**, 694 (1956), Soviet Phys. JETP **4**, 599 (1957).
[57] M. Lévy, Nucl. Phys. **57**, 152 (1964).
[58] H. Ferwerda, Physica **29**, 999 (1963); **30**, 665 (1964).
[59] M. Arons, M. Han, and E. Sudarshan, Phys. Rev. **137**, 1085 (1965).
[60] N. N. Bogolyubov and D. V. Shirkov, Vvedenie v teoriyu kvantovannykh poleĭ (Introduction to the Theory of Quantized Fields), Moscow, Gostekhizdat, 1957.
[61] A. N. Leznov, Dissertation (OIYaI, 1965) (Joint Inst. Nucl. Res.).
[62] P. Budini and T. Weber, Nuovo cimento **22**, 1321 (1961).
[63] J. McClure and S. Drell, Nuovo cimento **37**, 1638 (1965).
[64] D. I. Blokhintsev, JETP **22**, 254 (1952).
[65] B. V. Medvedev, DAN SSSR **103**, 37 (1955).
[66] Proc. of Discussion at CERN Conference on July 7, 1962.
[67] R. Eden and P. Landshoff, Ann. Phys. **31**, 370 (1965).
[68] G. Wanders, Helv. Phys. Acta **38**, 142 (1965).
[69] H. Stapp, Phys. Rev. **139**, 257 (1965).

Translated by W. H. Furry