On the Construction of the Scattering Matrix. I. Integral Causality Condition in Bogoliubov's Method

B. V. MEDVEDEV
Mathematical Institute, Academy of Sciences, USSR

(Submitted to JETP editor June 28, 1955)
J. Exptl. Theoret. Phys. (U.S.S.R.) 31, 791-802 (November, 1956)

An integral causality condition is formulated in Bogoliubov's method for the construction of the scattering matrix. Starting from this method, the compatibility of the conditions imposed on the scattering matrix and the existence theorem for the matrix are demonstrated; the scattering matrix is obtained in explicit form for the local theory.

IN most works on quantum field theory, the scattering matrix is constructed by starting either from Schrödinger's equation in the interaction representation[1-3] or from the equations of motion for the field operator in the Heisenberg representation[4,5]. In other words, the S–matrix is obtained as the solution of certain equations within the framework of the Hamiltonian method. However it was shown[6] some time ago that an analysis of the scattering matrix opens the way for the construction of a wider class of theories than the Hamiltonian. There is therefore considerable interest in a method recently proposed by Bogoliubov[7,8] (developing an idea of Stueckelberg[9]) for constructing the scattering matrix without reference to the equations of motion. We would especially like to emphasize that this method offers new possibilities for non-local theories which are apparently irreconcilable with the Hamiltonian scheme.

In order to preserve the possibility of a space-time description, one introduces in Bogoliubov's* method a function $g(x)$ ranging over the interval $[0, 1]$ which describes inclusion or exclusion of an interaction; the effective interaction $L(x)\,dx$ is changed to $L(x)\,g(x)\,dx$, and one seeks the scattering matrix for such an interaction, considering it as a function of g, through a formal decomposition in "powers" of $g(x)$:

$$S(g) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \times \int S_n(x_1, \ldots, x_n)\, g(x_1) \ldots g(x_n)\, dx_1 \ldots dx_n. \quad (1)$$

Instead of using the equation of motion, one applies obvious physical requirements to the scattering matrix: A—correspondence with the classical theory, B—relativistic invariance, C—unitarity $S(g)\,S^+(g) = S^+(g)\,S(g) = 1$, D—causality. The last, which shall be called the differential causality condition, may be written in the form

$$\frac{\delta}{\delta g(x)}\left(\frac{\delta S(g)}{\delta g(y)} S^+(g)\right) = 0 \quad \text{for} \quad x \lesssim y, \quad (2)$$

where the sign $\gtrsim$ signifies "later or space-like."

The purpose of the present article is to investigate (within the framework of Bogoliubov's method) the causality condition in a form which differs from (2) and which, it is believed, permits construction of the scattering matrix in a simpler and more natural fashion. Furthermore, the new causality condition which will be introduced here has the substantial advantage that it may be relatively easyly generalized to a non-local theory whose analysis by non-Hamiltonian methods proves of the greatest interest.

2. THE INTEGRAL CAUSALITY CONDITION

In order to formulate the causality condition, it is necessary to make use of the concepts "later" and "earlier." Considering two events in classical (non-relativistic) theory, the meaning of these concepts is graphically specified by the sign of the time interval between these events. It is well known that such a specification is not invariant in relativistic theory, and in order to specify invariantly the time order of two events, it is necessary to introduce three concepts: "later," "earlier," and "space-like." Since we are seeking here the scattering matrix in the form of a functional obtained by integrating over certain space-time regions, it is natural to start with certain concepts regarding the possible forms of 4-dimensional time order.

Consider two space-time regions G_1 and G_2. Let us construct at each point of G_1 a half-light cone, directed into the future, which we shall call the extended G_1 region, and let us denote by $\overline{G}_1$ the set of points which lie inside or at the boundary of at least one of these cones (Fig. 1). The region $\overline{G}_2$ is constructed in a similar fashion.

Now if the regions G_1 and $\overline{G}_2$ have no common point $G_1 \cap \overline{G}_2 = 0$, we shall say that the region G_1

---

*Reported by N. N. Bogoliubov at the meeting of the physico-mathematical division of the Jubilee meeting of the Academy of Sciences of the USSR in April 1954.

is situated *not later* than G_2, and write $G_1 \lesssim G_2$. In the opposite case, if $G_1 \cap G_2 \neq 0$, we shall write $G_1 \not\lesssim G_2$. Similar statements will hold for G_2 and G_1.

The relative positions of G_1 and G_2 can now be described by the following four logical possibilities.

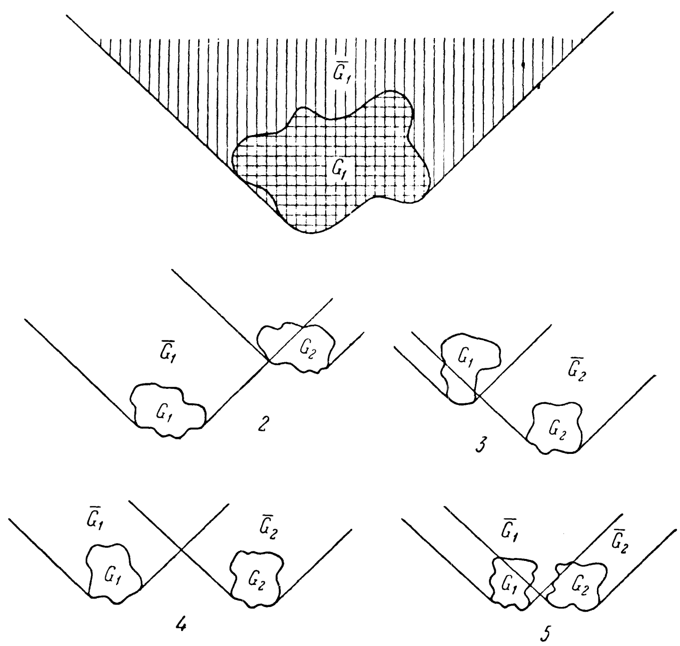
1. The region G_1 is situated not later than the region G_2, the region G_2 is not situated not later than the region G_1: $G_1 \lesssim G_2$, $G_2 \not\lesssim G_1$. In this case we shall say that the region G_2 is situated *later* than the region G_1 and write (Fig. 2) $G_1 < G_2$.

2. The region G_2 is situated not later than G_1, G_1 is not situated not later than G_2: $G_2 \lesssim G_1$, $G_1 \not\lesssim G_2$. We shall say that G_1 is situated later than G_2 and write (Fig. 3) $G_2 < G_1$.

3. The region G_1 is situated not later than G_2, the region G_2 is situated not later than G_1 We shall say that G_1 and G_2 are space-like with respect to each other and write (Fig. 4) $G_1 \sim G_2$.

If at least one of these possibilities is present, we shall say that the regions G_1 and G_2 are separable.

4. The region G_1 is not situated not later than G_2, the region G_2 is not situated not later than G_1: $G_1 \not\lesssim G_2$, $G_2 \not\lesssim G_1$. We shall say that the region G_1 and G_2 are unseparable and write (Fig. 5) $G_1 \gtrless G_2$.



FIGS. 1, 2, 3, 4, 5

It is clear that $G_2 \not\lesssim G_1$ implies that G_1 is situated later than or is space-like with respect to G_2 as reflected by the chosen notation.

In addition to the concepts of chronological order which we have just defined in 4-dimensions, we shall need later a representation of chronological order for two sets of space-time points $\{x_1, \ldots, x_l\}$ and $\{y_1, \ldots, y_m\}$. In order to do this we shall place each set of points $\{x_1, \ldots, x_m\}$ in a suitable 4-dimensional region G, a disconnected region consisting only of the points of the set; now we shall say that the set $\{x_1, \ldots, x_l\}$ is situated later than the set $\{y_1, \ldots, y_m\}$, etc., if the regions G_1, and G_2 are so situated with respect to each other.

Consider now two regions G_1 and G_2 such that G_1 is situated later than or is space-like with respect to G_2, $G_1 \gtrsim G_2$, and define a class of functions $g_1(x)$ and $g_2(x)$ such that

$$g_1(x) \neq 0, \quad \text{only if} \quad x \in G_1 \text{ and } g_2(x) \neq 0, \quad \text{only if } x \in G_2. \tag{3}$$

It is evident that the principle of causality as it is understood in the usual sense requires that no event in G_1 can have any influence on events in G_2.

It follows that the interaction localized in G_2 [i. e. described in terms of an arbitrary function belonging to the class $g_2(x)$] must act as if there was no interaction at all in the region G_1, i. e. [see Ref. 8, Eq. (3.13)], it must transfer the system from the state described by the constant initial state amplitude Φ_0, to a state of amplitude

$$\Phi_{g_2} = S(g_2)\Phi_0.$$

As for the effect on the system of the interaction localized in G_1, it must be independent of the explicit form of the interaction arising in G_2 but must be solely dependent on the final state described by the amplitude Φ_{g_2}, arising as a result of the interaction, i.e., the system must be transferred from the state Φ_{g_2} to the state

$$\Phi_{g_1 g_2} = S(g_1)\Phi_{g_2} = S(g_1)S(g_2)\Phi_0. \quad (4)$$

But on the other hand it is possible to reach the same final state by considering at once the sum of the effects of the interactions localized in the two regions G_1 and G_2. Application of the general rule [see Ref. (8)] must then yield a state with amplitude

$$\Phi_{g_1+g_2} = S(g_1+g_2)\Phi_0. \quad (5)$$

But amplitudes (4) and (5) describe the same state of the system; therefore, $\Phi_{g_1 g_2} = \Phi_{g_1+g_2}$, i.e.,

$$S(g_1+g_2) = S(g_1)S(g_2), \quad \text{if } G_1 \gtrsim G_2. \quad (6)$$

Equation (6) may be considered as the mathematical formulation of the causality condition. We shall call it the *integral causality condition.*

Let us restate the integral causality condition in the formalism of the operator functions S_n appearing in Eq. (1) (see Ref. 8, Sec. 1). We substitute expansion (1) for $S(g_1)$, $S(g_2)$, and $S(g_1+g_2)$ in Eq. (6), equate coefficients of equal powers in $g_1(x)$ and $g_2(x)$, and make use of the arbitrariness of the functions $g_1(x)$ and $g_2(x)$ within the boundary conditions imposed on them by Eq. (3) and the condition $G_1 \gtrsim G_2$; the integral causality condition for the operator function $S_n \times (x_1, \ldots, x_n)$, equivalent to the one imposed on the functionals in the total integrity causality condition (6), then becomes*

$$S_n(x_1, \ldots, x_n) \quad (7a)$$
$$= S_l(x_1, \ldots, x_l)\, S_{n-l}(x_{l+1}, \ldots, x_n),$$

$$\text{if} \quad \{x_1, \ldots, x_l\} \gtrsim \{x_{l+1}, \ldots, x_n\}. \quad (7b)$$

Comparing Eq. (7) with the analytical form [Ref. 8, Eqs. (4.11) and (4.12)] of the differential causality condition (2), we see that the former is considerably simpler. This reflects the fact that the integral condition (6) is imposed on the *matrix* $S(g)$ *itself*, while the differential condition (2) is imposed on its *variational derivative*.** Thus the integral causality condition seems more natural if the direct construction of the scattering matrix forms the basis of the theory.

We shall now establish the relationship between (6) and the differential condition (2), and show that they are equivalent. Consider a system with an interaction of intensity

$$g(x) = g_1(x) + g_2(x), \quad (8)$$

where $g_1(x)$ and $g_2(x)$ belong to the class (3), and assume that $G_1 \gtrsim G_2$. Then, applying (6)

$$S(g_1+g_2)$$
$$= S(g_1)S(g_2) \text{ and } S^+(g_1+g_2) = S^+(g_2)S^+(g_1).$$

Now give the function $g(x)$ an infinitesimal increment $\delta_y g$ in the infinitesimal neighborhood of the

---

* Because of the symmetry of the function S_n in all its arguments, the order of enumeration of the variables in (7) is unessential, and it may be written in the more general form

$$S_n(x_1, \ldots, x_n) \quad (7'a)$$
$$= S_l(x_{\lambda_1}, \ldots, x_{\lambda_l})\, S_{n-l}(x_{\lambda_{l+1}}, \ldots, x_{\lambda_n}),$$

$$\text{if} \quad \{x_{\lambda_1}, \ldots, x_{\lambda_l}\} \gtrsim \{x_{\lambda_{l+1}}, \ldots, x_{\lambda_n}\}. \quad (7'b)$$

Later on we shall sometimes write formulas in non-symmetrical form without mentioning the particular possibility of so doing in analogy with (7').

**Condition (2) takes on a simpler form if it is imposed on the Hamiltonian $H(x; g)$ [Ref. 8, Eq. (4.10)]; it therefore seems more natural if one starts to construct the theory not from the scattering matrix but from the analogue of an equation of the Tomonaga-Schwinger form. Making use of the latter method for constructing the theory it was shown by Bogoliubov that condition (2) physically describes the need for a local Hamiltonian in the interaction representation.

point $y \epsilon G_1$. The operator $1 + \delta_y S(g) S^+(g)$ which transforms $\Phi(g)$ into $\Phi(g) + \delta_y \Phi(g)$, then equals

$$1 + \delta_y S(g_1)\cdot S(g_2) S^+(g_2) S^+(g_1) = 1 + \delta_y S(g_1)\cdot S^+(g_1)$$

and is of course independent of the behavior of $g_2(x)$, i.e., of the behavior of the function $g(x)$ for points $z \epsilon G_2$. The function $g(x)$ having the particular form (8) will then satisfy the equation

$$\frac{\delta}{\delta g(z)}\left(\frac{\delta S(g)}{\delta g(y)}\cdot S^+(g)\right) = 0 \quad \text{for} \quad y \in G_1;\ z \in G_2, G_1 \gtrsim G_2, \tag{9}$$

which is the same as the differential condition (2).

Consider now some arbitrary function $g(x)$ and two points $z \lesssim y$. Then it is always possible to construct a space-like hypersurface σ, such that $z < \sigma < y$. Let us construct a family of functions $g_\epsilon(x)$ equal to $g(x)$ if $|x^4 - \sigma| > \epsilon$, equal to zero if $x \epsilon \sigma$, and increasing smoothly on both sides of σ into strips of width ϵ. For any value of ϵ, the functions $g_\epsilon(x)$ obviously belong to the class (8) for which Eq. (9) holds. The original function $g(x)$ may therefore be considered as the limit of the functions $g_\epsilon(x)$ as $\epsilon \to 0$. Accordingly, Eq. (9) applies to any $g(x)$ and $z \lesssim y$.

Thus the differential causality condition (2) follows from the integral causality condition (6).* The reverse statement follows from the fact that equation (2) leads to an expression for $S(g)$ which obviously satisfies Eq. (6) [see Ref. 8, Eq. (4.34)].

3. COMPATIBILITY OF THE CONDITIONS IMPOSED ON THE SCATTERING MATRIX

We shall show that the conditions A–D imposed on the scattering matrix are compatible. Returning to the representation (7) obtained from the causality condition, we note first of all that if the operator functions S_l and S_{n-l} appearing on the right-hand side are correctly transformed [see Ref. 8, Eq. (4.4)], then the same thing will automatically happen to S_n. Now if the variables $x_1, \ldots, x_n$ can be divided into two sets $\{x_1, \ldots, x_l\}$ and $\{x_{l+1}, \ldots, x_n\}$ which are space-like with respect to each other, then it follows from (7) that there are two simultaneous representations for S_n:

$$S_n(x_1, \ldots, x_n) = S_l(x_1, \ldots, x_l) S_{n-l}(x_{l+1}, \ldots, x_n) = S_{n-l}(x_{l+1}, \ldots, x_n) S_l(x_1, \ldots, x_l),$$
$$\text{if} \quad \{x_1, \ldots, x_l\} \sim \{x_{l+1}, \ldots, x_n\}. \tag{10}$$

Therefore, in order to preserve the relativistic invariance of the theory, it is necessary that the operator functions for two sets space-like with respect to each other commute. It is easily seen that this condition is fulfilled if and only if the elementary commutators of the field operators vanish for space-like intervals, as in the case for the ordinary invariant functions D and S. The compatibility of conditions B and D will then be guaranteed.

If the set of arguments $\{x_1, \ldots, x_n\}$ of the operator functions can be divided in more than one way into two subsets satisfying (7b) [or (7′b)], then the causality condition leads to several expressions for the same S_n through operator functions of lower order. We shall show that these expressions are essentially identical, i.e., that the several conditions imposed in such a case on a single S_n are compatible.

Assume that the set of arguments of the function $S_n(x_1, \ldots, x_n)$ can be divided in two ways into two sets satisfying (7b):

$$A \equiv \{x_{\lambda_1}, \ldots, x_{\lambda_l}\} \gtrsim \{x_{\lambda_{l+1}}, \ldots, x_{\lambda_n}\} \equiv \mathbf{A}, \tag{11.1}$$

$$B \equiv \{x_{\mu_1}, \ldots, x_{\mu_m}\} \gtrsim \{x_{\mu_{m+1}}, \ldots, x_{\mu_n}\} \equiv \mathbf{B}. \tag{11.2}$$

According to (7′) these divisions lead to the following expressions for S_n:

$$S_n(x_1, \ldots, x_n) = S(A) S(\mathbf{A}), \tag{12.1}$$

$$S_n(x_1, \ldots, x_n) = S(B) S(\mathbf{B}). \tag{12.2}$$

In order to show their equivalence, consider the intersection $A \cap B$ of the sets A and B. Since $A \cap B \subset A$, then from (11.1):

$$A \cap B \gtrsim \mathbf{A}. \tag{13.1}$$

On the other hand, $A \cap B \subset B$; therefore, it follows similarly that:

*The fact that the differential causality condition imposed on the operator functions S_n follows from (7) may be shown, without recourse to limiting processes, by means of a direct though rather cumbersome combinatorial transformation.

$$A \cap B \gtrsim \mathbf{B}. \qquad (13.2)$$

The set of arguments $A - A \cap B$ contained in A but not in B must be contained in B

$$A - A \cap B \subset \mathbf{B}. \qquad (14.1)$$

Similarly,

$$B - A \cap B \subset \mathbf{A}. \qquad (14.2)$$

But it follows from (14.1) and (13.2) that $A \cap B \gtrsim A - A \cap B$, and from (14.2) and (13.1) that $A \cap B \gtrsim B - A \cap B$. Therefore, it is possible to carry out a further division of the factors $S(A)$ and $S(B)$ and write

in (12.1): $S(A) = S(A \cap B)\, S(A - A \cap B)$

and in (12.2): $S(B) = S(A \cap B)\, S(B - A \cap B)$.

Repeating the same procedure for the sets A and B, we are led to two series of ordered relations*:

$$A \cap B \gtrsim A - A \cap B \gtrsim \mathbf{A} - \mathbf{A} \cap \mathbf{B} \gtrsim \mathbf{A} \cap \mathbf{B}, \qquad (15.1)$$

$$A \cap B \gtrsim B - A \cap B \gtrsim \mathbf{B} - \mathbf{A} \cap \mathbf{B} \gtrsim \mathbf{A} \cap \mathbf{B}, \qquad (15.2)$$

from which are obtained the following representations for S_n:

$$S_n = S(A \cap B)\, S(A - A \cap B)\, S \times (\mathbf{A} - \mathbf{A} \cap \mathbf{B})\, S(\mathbf{A} \cap \mathbf{B}), \qquad (16.1)$$

$$S_n = S(A \cap B)\, S(B - A \cap B)\, S \times (\mathbf{B} - \mathbf{A} \cap \mathbf{B})\, S(\mathbf{A} \cap \mathbf{B}), \qquad (16.2)$$

which only differ in the two inside factors. But it is easily seen that $A - A \cap B = \mathbf{B} - \mathbf{A} \cap \mathbf{B}$ and $B - A \cap B = \mathbf{A} - \mathbf{A} \cap \mathbf{B}$. Calling the first of these sets C, and the second D, it appears that (16.2) is obtained from (16.1) by the exchange $S(C)\, S(D) \to S(D)\, S(C)$. But it follows in particular from (15.1) that $C \gtrsim D$, and from (15.2) that $D \gtrsim C$. Therefore $C \sim D$ and according to (10), $S(C)$ and $S(D)$ commute. The compatibility of various expressions of the causality condition is then demonstrated if this condition can be repeatedly applied to a single S_n.

*Whenever we write a series of the form

$$A_1 \gtrsim A_2 \gtrsim \cdots \gtrsim A_k,$$

it implies not only the relations $A_1 \gtrsim A_2$, $A_2 \gtrsim A_3$, but also all the relations $A_i \gtrsim A_j$, if $i < j$. This proviso must be added since the relation $\gtrsim$ is not transitive.

We shall consider now how expression (7), which follows from the integral causality condition, is related to the condition of unitarity [Ref. 8, Eq. (4.9)]. We shall first examine the simple case wherein one of the points $x_1, \ldots, x_n$ (for instance, x_n) is situated earlier (later) than or is spacelike with respect to all the others; we shall then show that the operator functions appearing in Eq. (7) automatically satisfy the unitarity condition for $n > 1$ if it is satisfied for $n = 1$.

Let us rewrite the left-hand side of the unitarity condition [Ref. 8, Eq. (4.9)], separating it into summations over k and over the permutations of terms in which x_n appears in S_k and S^+_{n-k}, and let us apply (7). We obtain

$$S_1^+(x_n)\, S_{n-1}^+(x_1, \ldots, x_{n-1}) + \sum_{k=1}^{n-1} P\left(\frac{x_1, \ldots, x_{k-1}}{x_k, \ldots, x_{n-1}}\right) \times S_{k-1}(x_1, \ldots, x_{k-1})\, S_1(x_n)\, S_{n-k}^+(x_k, \ldots, x_{n-1})$$
$$+ \sum_{k=1}^{n-1} P\left(\frac{x_1, \ldots, x_k}{x_{k+1}, \ldots, x_{n-1}}\right) \times S_k(x_1, \ldots, x_k)\, S_1^+(x_n)\, S_{n-k-1}^+(x_{k+1}, \ldots, x_{n-1}) + S_{n-1}(x_1, \ldots, x_{n-1})\, S_1(x_n).$$

Now carrying out in the first sum the transformation $k \to k+1$, removing from it the term $k = 0$, and from the second one the term $k = n - 1$, it is found that all the terms occur in pairs whose only difference is the exchange of $S_1(x_n)$ into $S_1^+(x_n)$; the whole expression can then be written in the compact form

$$\sum_{k=0}^{n-1} P\left(\frac{x_1, \ldots, x_k}{x_{k+1}, \ldots, x_{n-1}}\right) S_k(x_1, \ldots, x_k)\, (S_1(x_n) + S_1^+(x_n))\, S_{n-k+1}^+(x_{k+1}, \ldots, x_{n-1}),$$

where it is assumed that $S_0 = S_0^+ = 1$. The last expression has a common factor $S_1(x_n) + S_1^+(x_n)$ which is equal to zero due to the unitarity condition for $n = 1$, and thus the whole expression vanishes.

Our assertion has therefore been proven. We note that the correspondence principle requires that[8] $S_1(x) = iL(x)$ and therefore the unitarity condition for $n = 1$ requires a Hermitian Lagrangian $L(x)$. If $L(x)$ is a Hermitian, the conditions A and C are compatible.

Consider now the general case wherein the arguments $x_1, \ldots, x_n$ can be divided into two groups

$$\{x_1, \ldots, x_{n-m}\} \geqslant \{x_{n-m+1}, \ldots, x_n\}. \tag{17}$$

Again we transform the sums over the permutations

$$P\left(\frac{x_1, \ldots, x_k}{x_{k+1}, \ldots, x_n}\right) S_k(x_1, \ldots, x_k)\, S^+_{n-k}(x_{k+1}, \ldots, x_n) \tag{18}$$

occuring in Eq. (4.9) of Ref. (8) so as to distinguish permutations within each set $\{x_1, \ldots, x_{n-m}\}$ and $\{x_{n-m+1}, \ldots, x_n\}$, from permutation between sets; we single out in a sum over λ the arguments of the first sets whose subscripts appear in the arguments of S_k. The remaining sum over the permutations $P\left(\frac{x_1, \ldots, x_k}{x_{k+1}, \ldots, x_n}\right)$ is then reduced to a sum over permutations within the first group of arguments $P\left(\frac{x_1, \ldots, x_\lambda}{x_{\lambda+1}, \ldots, x_{n-m}}\right)$ and within the second group

$$P\left(\frac{x_{n-m+1}, \ldots, x_{n-m+k-\lambda}}{x_{n-m+k-\lambda+1}, \ldots, x_n}\right).$$

Expression (18) then becomes

$$\sum_{(\lambda)} P\left(\frac{x_1, \ldots, x_\lambda}{x_{\lambda+1}, \ldots, x_{n-m}}\right) \times P\left(\frac{x_{n-m+1}, \ldots, x_{n-m+k-\lambda}}{x_{n-m+k-\lambda+1}, \ldots, x_n}\right) \times S_k(x_1, \ldots, x_\lambda, x_{n-m+1}, \ldots, x_{n-m+k-\lambda}) \times S^+_{n-k}(x_{\lambda+1}, \ldots, x_{n-m}, x_{n-m+k-\lambda+1}, \ldots, x_n), \tag{19}$$

where the summation limits over λ are given by the inequalities

$$\lambda \geqslant 0;\ \lambda \geqslant k-m;\ \lambda \leqslant k;\ \lambda \leqslant n-m.$$

Now we use the causality condition (7) in order to split each of the operator functions entering in (19) into the product of two factors, and we substitute expression (18), thus transformed, into the equation stating the unitarity condition [Ref. 8, Eq. (4.9)]. We obtain

$$\sum_{k=0}^{n} \sum_{(\lambda)} P\left(\frac{x_1, \ldots, x_\lambda}{x_{\lambda+1}, \ldots, x_{n-m}}\right) \times P\left(\frac{x_{n-m+1}, \ldots, x_{n-m+k-\lambda}}{x_{n-m+k-\lambda+1}, \ldots, x_n}\right) S_\lambda(x_1, \ldots, x_\lambda) \times S_{k-\lambda}(x_{n-m+1}, \ldots, x_{n-m+k-\lambda}) \times S^+_{m-k+\lambda}(x_{n-m+k-\lambda+1}, \ldots, x_n) \times S^+_{n-m-\lambda}(x_{\lambda+1}, \ldots, x_{n-m}). \tag{20}$$

Exchanging the summation variables k and λ for the variables $\mu = k-\lambda$ and λ, the summation region is transformed within the plane μ, λ into a rectangle bounded by the lines $\lambda = 0$, $\lambda = n-m$, $\mu = 0$, and $\mu = m$, i.e., the sums over μ and λ can be carried out independently. Thus we can carry out in (20) an internal sum over μ

$$\sum_{\mu=0}^{m} P\left(\frac{x_{n-m+1}, \ldots, x_{n-m+\mu}}{x_{n-m+\mu+1}, \ldots, x_n}\right) S_\mu(x_{n-m+1}, \ldots, x_{n-m+\mu})\, S^+_{m-\mu}(x_{n-m+\mu+1}, \ldots, x_n),$$

which will enter in the factor of each term of the sum over λ. But it is obvious that this sum equals the left-hand side of the unitarity condition equation for $n=m$; if we therefore assume that the unitarity condition is satisfied for $m = 1, \ldots, n-1$ and for arbitrary values of the arguments, then (2) automatically vanishes for $0 < m < n$.

Thus if the unitarity condition is satisfied for $m = 1, \ldots, n-1$ and arbitrary values of the arguments, then it is atomatically satisfied for $m=n$ by virtue of the integral causality condition, as long as there exists at least one way to partition the arguments $x_1, \ldots, x_n$ into two non-empty sets satisfying (17). If on the other hand such a partition does not exist, the causality condition generally does not impose any limitations on S_n and the unitarity condition can always be satisfied. Thus the compatibility of the unitarity and causality conditions is demonstrated.

Let us finally note that the compatibility of conditions A and B is guaranteed if $L(x)$ is a scalar; the compatibility of A and D is guaranteed by a local Lagrangian; and the compatibility of B and C follows from the fact that Eq. (4.4) of Ref. 8 requires that S_n transforms identically with all the products $S_n S^+_{n-k}$ appearing in the unitary condition [Ref. 8, Eq. (4.9)]. The compatibility of the conditions A–D is thus demonstrated.

The relative simplicity of this demonstration must be ascribed to the use of the integral form (6) of the causality condition. The use of the differential form [Ref. 8, Eqs. (4.11) and (4.12)] considerably complicates the demonstration and it was found easier in Ref. 8 to obtain first an explicit form for $S(g)$ and then demonstrate that it satisfies the conditions A–D.

The existence theorem for the scattering matrix $S(g)$ follows immediately from the proof of the compatibility of conditions A–D. Assume that the quantities $S_1, \ldots, S_n$ constructed above satisfy conditions A–D. Now construct S_{n+1}. If the values of the arguments $x_1, \ldots, x_n$ are such that it is possible to carry out at least one partition of the form (7′b), then the value of S_{n+1} is determined from the causality condition without violating (as demonstrated) any of the other conditions. If the values of the arguments negate the existence of even one such partition, then the value of S_{n+1} may be obtained (even though non-uniquely) by starting from the unitarity condition again without conflicting with the remaining conditions. Thus the existence theorem for the scattering matrix $S(g)$ satisfying conditions A–D, is demonstrated.*

4. THE FORM OF THE SCATTERING MATRIX

We shall show that the method chosen for constructing the scattering matrix leads in a natural fashion to explicit expressions for the coefficients of the functions S_n and thereby to a simple closed form expression for $S(g)$. First we introduce some definitions.

If a set $\{g_1, \ldots, g_n\}$ of space-time regions (in particular the arguments $x_1, \ldots, x_n$ of the operator functions may be such regions) is divided into two non-empty sets $\{g_{\lambda_1}, \ldots, g_{\lambda_l}\}$ and $\{g_{\lambda_{l+1}}, \ldots, g_{\lambda_n}\}$ satisfying a condition of the type (7′b)

$$\{g_{\lambda_1}, \ldots, g_{\lambda_l}\} \gtrsim \{g_{\lambda_{l+1}}, \ldots, g_{\lambda_n}\}, \tag{21}$$

then we shall say that a *section* has been carried out which has separated the original set into the sum of two sets $\{g_{\lambda_1}, \ldots, g_{\lambda_l}\}$ and $\{g_{\lambda_{l+1}}, \ldots, g_{\lambda_n}\}$.

*The demonstration is of course a formal one inasmuch as we do not concern ourselves with questions of convergence.

It is clear that every section is realized by means of some space-like hypersurface σ, such that every region of the first set is situated later than the hypersurface and every region of the second is situated earlier.

If it is impossible to carry out any sections in the set of regions $\{g_1, \ldots, g_n\}$, we shall say that the regions ("points") $g_1, \ldots, g_n$ are *completely unseparable* or the set of these regions is *indecomposable*. In the opposite case we shall say that the set of regions $\{g_1, \ldots, g_n\}$ is *decomposable*.

If a set of regions $\{g_1, \ldots, g_n\}$ can be divided into m groups $\{g_{\lambda_1}, \ldots, g_{\lambda_{\nu_1}}\}, \ldots,$

$$\{g_{\lambda_{\nu_1+\ldots+\nu_{m-1}+1}}, \ldots, g_{\lambda_n}\},$$

such that: a) any two groups

$$\{g_{\lambda_{\nu_1+\ldots+\nu_{j-1}+1}}, \ldots, g_{\lambda_{\nu_1+\ldots+\nu_j}}\}$$

$$\text{and}\ \{g_{\lambda_{\nu_1+\ldots+\nu_{i-1}+1}}, \ldots, g_{\lambda_{\nu_1+\ldots+\nu_i}}\}$$

are separate with respect to each other and b) the regions contained within each group are completely unseparable (i.e., every group of regions is indecomposable), then we shall say that the set of regions $\{g_1, \ldots, g_n\}$ is *completely decomposable*.

Let us consider a decomposable set of regions G. We can by definition perform a section in that set. Let us assume that such a section divides G into the sum of G_1 and G_2 such that $G_1 \gtrsim G_2$. Consider now the set G_1. It may be either indecomposable or decomposable. In the first case we shall consider the set G_2, in the second case there exists a section which divides G_1 into the sum of two sets G_{11} and G_{12}, $G_{11} \gtrsim G_{12}$. Since $G_1 \gtrsim G_2$, such a relationship is also satisfied for each part G_{11} and G_{12} of the set G_1: $G_{11} \gtrsim G_{12}$ and $G_{12} \gtrsim G_2$. We can therefore write (see footnote on page 677):

$$G_{11} \gtrsim G_{12} \gtrsim G_2, \tag{22}$$

i.e., G is divided into three sets, G_{11}, G_{12} and G_2 satisfying a series of relations (21). (If we have to contend with the first case mentioned above, and G_2 is decomposable, then $G_1 \gtrsim G_{21} \gtrsim G_{22}$; if G_2 is indecomposable, then the problem of representing a decomposable set into a sum of indecomposable sets is immediately solved.) Each of these sets may be either indecomposable—then we no longer consider it—or decomposable. In the latter case there exists a section which decomposes the set G_α into two sets $G_{\alpha 1} \gtrsim G_{\alpha 2}$. By the same arguments which were used to deduce (22), $G_{\alpha 1}$ and $G_{\alpha 2}$ will form another term in the series just like G_α; thus the performance of a section reduces simply to the exchange

$$G_\alpha \to G_{\alpha 1} \gtrsim G_{\alpha 2}$$

in expression (22). Such a process can obviously be continued until the remaining sets are all indecomposable and one obtains the series of relations

$$G_{\nu_1} \gtrsim G_{\nu_2} \gtrsim \ldots \gtrsim G_{\nu_m}; \quad (\Sigma \nu_i = n) \qquad (23)$$

($m \leq n$, where n is the number of regions in G wherein one only considers indecomposable sets of regions).

It has therefore been shown that every decomposable set may be decomposed into a sum of indecomposable sets, i.e., *every decomposable set is completely decomposable*,

It has further been shown that every decomposable set can be represented in the form of an *ordered* succession of indecomposable sets satisfying (23).

It is easily seen that every decomposition is unique. Indeed, let us assume that there exists another decomposition of G into indecomposable sets some of which at least do not coincide with the sets appearing in (23). Let G_β be such a set. There are two logical possibilities: either it consists of regions contained in various sets of (23), but such a set cannot be indecomposable since the regions it contains can be divided by one of the sections utilized in (23); or else it consists of regions which are all contained in one of the indeccomposable sets of (23), although they do not exhaust it; but this implies decomposition of an indecomposable set, which is impossible by definition.

On the other hand the representation of a decomposable set in the form of an *ordered* succession of indecomposable sets is not unique. This may be seen, for example, by considering a set $\{g_1, g_2\}$ of two regions space-like with respect to each other $g_1 \sim g_2$. It is obvious that the decomposable set $\{g_1, g_2\}$ may be represented in the form of two non-identically ordered series

$$g_1 \gtrsim g_2 \text{and} g_2 \gtrsim g_1.$$

In the above example, non-uniqueness follows from the possibility of exhanging in (23) two regions which are space-like with respect to each other. We shall show that this is the case in general. It is obvious from the way in which the ordered series (23) was constructed that there are indeed several series of sections which may be chosen from those avaliable in the original set G, and the non-uniqueness of the ordered series of sets follows from the available non-unique choices of such series of sections. Therefore a transition from one ordering (23) to another is accomplished by a transition from one series of sections to another. But the latter transition can always be carried by successive exchange of one section for another. We have already examined such exchanges when we deduced the compatibility of the various forms of the causality conditions, and we have found that they always reduce to permutations in series of sets of regions of the type (23), which proves our contention.

We shall now stop our geometrical digression and we shall concern ourselves with applying the results which we have obtained to the explicit construction of the coefficients of the functions of the scattering matrix.

In order to do this we shall consider a set of regions G which turn out to be the arguments $x_1, \ldots, x_n$ of the operator functions S_n (the regions consist of one point each). According to the theorem we have just demonstrated, this set may be uniquely decomposed into a sum of m ($1 \leq m \leq n$). indecomposable sets

$$\{x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}\}; \ \{x_{\lambda_{\nu_1+1}}, \ldots, x_{\lambda_{\nu_1+\nu_2}}\}, \ldots,$$

$$\{x_{\lambda_{\nu_1+\ldots+\nu_{n-1}+1}}, \ldots, x_{\lambda_n}\},$$

which may be (non-uniquely) ordered into a chronological series

$$\{x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}\} \gtrsim \{x_{\lambda_{\nu_1+1}}, \ldots, x_{\lambda_{\nu_1+\nu_2}}\} \qquad (24)$$

$$\gtrsim \ldots \gtrsim \{x_{\lambda_{\nu_1+\ldots+\nu_{m-1}+1}}, \ldots, x_{\lambda_n}\}.$$

Applying now the integral causality condition (7′a) to each section figuring in (24), we obtain the following representation for S_n

$$S_n(x_1, \ldots, x_n)$$
$$= S_{\nu_1}(x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}) S_{\nu_2}(x_{\lambda_{\nu_1+1}}, \ldots, x_{\lambda_{\nu_1+\nu_2}})$$
$$\times \ldots S_{\nu_m}(x_{\lambda_{\nu_1+\ldots+\nu_{m-1}+1}}, \ldots, x_{\lambda_n});$$

$$\nu_1 + \ldots + \nu_m = n. \qquad (25)$$

Considering now the symmetry of S_n in all its arguments,[8] it is easily seen that (25) can be written in the form of the T-product

$$S_n(x_1, \ldots, x_n) \tag{26}$$

$$= T\,[S_{\nu_1}(x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}})\, S_{\nu_2}(x_{\lambda_{\nu_1+1}}, \ldots, x_{\lambda_{\nu_1+\nu_2}})$$

$$\times \ldots S_{\nu_m}(x_{\lambda_{\nu_1+\ldots+\nu_{m+1}+1}}, \ldots, x_{\lambda_n})],$$

where it should be kept in mind that chronological ordering is only carried out among distinct indecomposable groups of points. Representation (26) is valid if the set of arguments $\{x_1, \ldots, x_n\}$ may be decomposed into a sum of indecomposable sets $\{x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}\}, \ldots, \{x_{\lambda_{\nu_1+\ldots+\nu_{m-1}+1}}, \ldots, x_{\lambda_n}\}$. It is not necessary to discuss here orderings of the type appearing in (24) inasmuch as it will be considered in the evaluation of the T-product [the proof of the theorem on the ordering of sums of decomposable sets guarantees that the T-product (26), carries a well-defined meaning].

If in particular all the arguments $x_1, \ldots, x_n$ are pairwise unseparable, then (26) becomes

$$S_n = T\,[S_1(x_1) \ldots S_1(x_n)].$$

Furthermore if the principle of correspondence [Ref. 8, Eq. (4.23) is taken into account, then it is found that

$$S_n = i^n T\,[L(x_1) \ldots L(x_n)], \tag{27}$$

coinciding exactly with the particular solution given in Eq. (4.23) of Ref. 8, which thus appears as a general solution of S_n for pairwise unseparable arguments.

We have therefore expressed all the operator functions S_n in the form of chronological products of operator functions of indecomposable sets of points and a knowledge of these is now sufficient to completely determine the scattering matrix. But in local theory a set of points can only be nondecomposable if all these points coincide. Accordingly, the signfiicance of the operator function of indecomposable point sets appears to be that of quasi-local operators in the sense defined in Ref. 8, Sec. 2.

The causality condition no longer imposes any limitations on these quasi-local operators [apart from the commutation requirement on two such operators for space-like indecomposable sets of points, cf. Eq. (10)] , we need therefore only concern ourselves with fulfilling the unitarity condition. It is easily seen that it [Ref. 8, Eq. (4.9)] defines uniquely their Hermitian parts, while their anti-Hermitian parts remain arbitrary. They must be specified during the formulation of a theory. Denoting them by $i\,\tilde{L}_\nu(x_1, \ldots, x_\nu)$, where $\tilde{L}_\nu$ is Hermitian (we depart here from the notation of Ref. 8, where quasi-local operators are denoted by Λ_ν), it may be said that (26) uniquely determines all the operator functions S_n by means of the following series of Hermitian quasi-local operators

$$\tilde{L}_1(x) = L(x), \tag{28}$$

$$\tilde{L}_2(x_1, x_2), \ldots, \tilde{L}_n(x_1, \ldots, x_n), \ldots,$$

which must be supplied by a particular theory.

Such a formulation, however, suffers from the disadavantage that it does not assign unique expressions to the functions S_n over the whole range of the arguments. In order to avoid this, it is necessary to represent the value of S_n for an indecomposable group of arguments in the form of a sum of expressions, extending its value in a continuous fashion over a decomposable group of arguments, and certain new quasi-local operators.

In other words we must generalize the definition of the T-product to include the case of coinciding arguments whereby an intuitive chronological order loses significance. This may be done by starting from Wick's theorem; however, in order to give meaning to the products of singular functions with coinciding singularities arising in this case, it is necessary to have recourse to some limiting process (regularization). The result will depend of course on the form of the limiting process. The usual methods of regularization possess the advantage that when they are used to determine the value of the T-product for coinciding arguments, the operator functions S_n automatically satisfy the condition of unitarity (in local theory this is a natural consequence of continuity). The new quasi-local operators will then be anti-Hermitian and we shall denote them by $i\,L_\nu(x_1, \ldots, x_\nu)$.

Expression (26) may then be written in the form

$$S_n(x_1, \ldots, x_n) \tag{29}$$

$$= \sum T\,[i\,L_{\nu_1}(x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}) \cdot \ldots \cdot i\,L_{\nu_m}(\ldots, x_{\lambda_n})],$$

where the summation is to be carried out over all possible divisions of the set $\{x_1, \ldots, x_n\}$ into subsets

$$\{x_{\lambda_1}, \ldots, x_{\lambda_{\nu_1}}\}, \ldots, \{x_{\lambda_{\nu_1+\ldots+\nu_{m-1}+1}}, \ldots, x_{\lambda_n}\},$$

wherein the order of the factors appearing within the T-product is unimportant. Writing out the sum appearing in (29) in explicit form yields the expression

$$S_n(x_1, \ldots, x_n) = i^n T(L(x_1) \ldots L(x_n)) \quad (30)$$

$$+ \sum_{\substack{2 \leqslant m \leqslant n-1 \\ \Sigma\nu_i = n}} \frac{i^m}{m!} P(x_1, \ldots, x_{\nu_1} \; x_{\nu_1+1}, \ldots,$$

$$x_{\nu_1+\nu_2} | \ldots | x_{\nu_1+\ldots+\nu_{m-1}+1}, \ldots, x_n)$$

$$\times T(\mathrm{L}_{\nu_1}(x_1, \ldots, x_{\nu_1}) \ldots \mathrm{L}_{\nu_m}$$

$$\times (x_{\nu_1+\ldots+\nu_{m-1}+1}, \ldots, x_n))] + i\mathrm{L}_n(x_1, \ldots, x_n),$$

which coincides with the expression obtained in Ref. 8, Eq. (4.32). Thus the integral causality condition leads automatically (and without the need for any artificial considerations or conjectures) to the general expression (30) obtained by Bogoliubov, for the operator functions S_n.

Combining the whole series of quasi-local operators

$$\mathrm{L}_1(x) = L(x),\ \mathrm{L}_2(x_1, x_2), \ldots, \mathrm{L}_n(x_1, \ldots, x_n), \ldots, \quad (31)$$

which must be specified in order to determine (30), in the form of a single functional

$$L(x; g) = L(x) \quad (32)$$

$$+ \sum_{\nu=2}^{\infty} \frac{1}{\nu!} \int \mathrm{L}_\nu(x, x_1, \ldots, x_{\nu-1})$$

$$\times g(x_1) \ldots g(x_{\nu-1})\, dx_1 \ldots dx_{\nu-1},$$

we obtain [see Ref. 8, Eqs. (4.30) and (4.34)] the final closed-form expression

$$S(g) = T \exp\left\{i \int L(x; g)\, g(x)\, dx\right\} \quad (33)$$

for the matrix $S(g)$ as a function of the "generalized Lagrangian" $L(x;g)$, and coinciding with the expression obtained in Eq. (4.34), Ref. 8.

The author wishes to take this opportunity to express his deep gratitude to N. N. Bogoliubov for his constant interest in this work and for many valuable suggestions and demonstrations, and also to D. V. Shirkov and V. L. Bonch-Bruevich for fruitful discussions.

[1] S. Tomonaga, Progr. Theoret. Phys. **1**, 27 (1946); Koba, Tati, and Tomonaga, Progr. Theoret. Phys. **2**, 101 (1947).

[2] J. Schwinger, Phys. Rev. **74**, 1439 (1948); **76**, 790 (1949).

[3] F. J. Dyson, Phys. Rev. **75**, 486, 1736 (1949).

[4] G. Källen, Ark. for Fys. **2**, 187 (1950); 371 (1951).

[5] C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).

[6] W. Heisenberg, Zeis. f. Phys. **120**, 513, 673 (1943).

[7] N. N. Bogoliubov, Dokl. Akad. Nauk (SSSR) **81**, 757, 1015 (1951); **82**, 217 (1952); Izv. Akad. Nauk (SSSR), seria fizhika (in press).

[8] N. N. Bogoliubov and D. V. Shirkov, Usp. Fiz. Nauk **55**, 151; **57**, 3 (1955).

[9] E. C. G. Stueckelberg, Phys. Rev. **81**, 130 (1951); Stueckelberg *et al.*, Helv. Phys. Acta **22**, 215 (1949); **24**, 317 (1951); **26**, 499 (1953).

Translated by M. A. Melkanoff
172