Causality and the Dispersion Relation: Logical Foundations*†

JOHN S. TOLL
Princeton University, Princeton, New Jersey and University of Maryland, College Park, Maryland

(Received January 31, 1956; revised manuscript received August 24, 1956)

"Strict causality" is the assumption that no signal whatsoever can be transmitted over a space-like interval in space-time, or that no signal can travel faster than the velocity of light in vacuo. In this paper a rigorous proof is given of the logical equivalence of strict causality ("no output before the input") and the validity of a dispersion relation, e.g., the relation expressing the real part of a generalized scattering amplitude as an integral involving the imaginary part. This proof applies to a general linear system with a time-independent connection between the output and a freely variable input and has the advantage over previous work that no tacit assumptions are made about the analytic behavior or single-valuedness of the amplitude, but, on the contrary, strict causality is shown to imply that the generalized scattering amplitude is analytic in the upper half of the complex frequency plane. The dispersion relations are given first as a relation between the real and imaginary parts of the generalized scattering amplitude and then in terms of the complex phase shift.

1. INTRODUCTION

In this paper we shall discuss the logical equivalence of the dispersion relation and a condition of "strict causality." By strict causality we mean the condition "no output can occur before the input." For particular physical systems, the condition can be given more specific form. Thus, in the case of a scattering system, the causality condition becomes that "no scattered wave can appear until the primary wave has reached some part of the scatterer;" for a homogeneous refractive medium the condition is that "no signal can be transmitted faster than c" or "a source at the time $t=0$ can produce no electromagnetic disturbance whatsoever at the plane $x=x_0$ in advance of the time $t=x_0/c$." We shall give a rigorous proof of the logical equivalence of strict causality and the validity of the dispersion relations.

A dispersion relation is a simple integral formula relating a dispersive process to an absorption process. Such relations occur in many fields of physics. The relation is perhaps best known in the theory of dispersion of light[1] in a dielectric, where the complex refractive index $n(\omega)$ is expressed as an integral over all frequencies involving the linear absorption coefficient

$$n(\omega_r)=1+\frac{c}{\pi}\lim_{\epsilon\to 0+}\int_0^\infty \frac{\alpha(\omega_a)d\omega_a}{\omega_a^2-(\omega_r+i\epsilon)^2}. \qquad (1.1)$$

The dispersion relation can also be expressed as an integral relation between the total interaction cross section and the forward coherent scattering amplitude for a single scattering center. More generally, it is a relation connecting real and imaginary parts of the diagonal elements of the scattering matrix[2] or between the absorption and the phase shift, and it occurs in the scattering of nuclear particles[3] and numerous other fields. Similar relations are well known in electrical network theory[4] where the resistance as a function of frequency can be obtained as an integral involving the reactance function. Thus the dispersion relation is of wide generality and usefulness. As we shall see, we can expect such a general connection in any theory where the "output" function of the time (e.g., a scattered wave) is a *linear* functional of an "input" function (e.g., the primary wave), where the interaction law is time-independent, and where the output function cannot begin before the input function is applied (causality condition).

There has been considerable study of the logical foundations of the dispersion relation. Sommerfeld[5] and Brillouin[6] proved that in an idealized dielectric no signal travels faster than c even though there may be frequencies for which both the phase velocity and the group velocity exceed c. Kramers[7] used the notion of the complex refractive index defined by analytic continuation in the complex frequency plane to show that a signal cannot travel faster than c in any medium for which the dispersion relation is satisfied; hence the deeper reason why Sommerfeld and Brillouin had found strict causality to be satisfied was that they had chosen a complex refractive index function which

* The first part of a dissertation of John S. Toll, which was submitted to Princeton University in partial fulfillment of the requirements for the Doctor of Philosophy degree (1952).

† Most of this work was carried out during 1949–1951 under the auspices of Princeton University; the author gratefully acknowledges a Proctor Fellowship and a grant from the Friends of Elementary Particle Physics during this period. The author is now at the University of Maryland where this work was prepared for publication with support from a National Science Foundation Grant.

[1] For reviews of this development and further references, see K. L. Wolf and K. F. Herzfeld, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1928), Vol. 20, Chap. 10, p. 480; A. Korff and G. Breit, Revs. Modern Phys. 4, 471 (1932).

[2] See, for example, Jost, Luttinger, and Slotnick, Phys. Rev. **80**, 189 (1950), Appendix A.

[3] See, for example, W. Schutzer and J. Tiomno, Phys. Rev. **83**, 349 (1951). For an excellent general discussion see E. P. Wigner, Am. J. Phys. **23**, 371 (1955) where further references are given.

[4] See, for example, H. W. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, Inc., New York, 1940); Y. W. Lee, J. Math Phys. **11**, 83 (1932).

[5] A. Sommerfeld, Ann. Physik 44, 177 (1914).

[6] L. Brillouin, Ann. Physik 44, 203 (1914).

[7] H. A. Kramers, *Estratto dagli Atti del Congresso Internazionale de Fisici Como* (Nicolo Zonichelli, Bologna, 1927).

satisfied the dispersion relation. Krönig[8] then gave the first proof of the equivalence of causality and dispersion, showing that the dispersion relation is the necessary, as well as sufficient, condition for strict causality to be satisfied. The present paper is also concerned with proving the logical equivalence of causality and dispersion, and the essential ideas for our proof were already contained in Krönig's work. However, Krönig made tacit assumptions about analytic behavior of the dispersion function which are avoided in the proof given here; and we find that the dispersion relation must in fact be trivally extended before it is logically equivalent to strict causality. Schutzer and Tiomno[3] studied the equivalence of the dispersion relation and causality for nonrelativistic particles. Van Kampen has investigated the equivalence of causality and the dispersion relation both for light[9] and for nonrelativistic particles[10]; his work is a beautiful and fully rigorous treatment of the special case of spherical waves impinging on spherically symmetric scattering centers.

2. PRELIMINARY DISCUSSION

This section will give a brief heuristic discussion of dispersion relations and causality; a rigorous proof of their connection is then given in Sec. 3.

It is easy to see how a general relation like the dispersion relation arises in a scattering system. To each absorption process there corresponds a higher order contribution to the coherent scattering, which can be visualized as occurring in two steps: first, the absorption of the incident particle and then its re-emission with the whole system of absorber and incident particle returning to its initial state. The absorption can be "virtual" or "real," where by the latter we mean that the conservation laws are satisfied in the intermediate state as well as in the initial and final states. The virtual processes contribute to the real part, and the real processes contribute to the imaginary part, of the coherent scattering amplitude. Thus the imaginary part of the coherent scattering amplitude at a particular energy can be found directly from the knowledge of the total interaction cross section, for incident particles of that energy, simply by an application of the principle of microscopic reversibility to get the re-emission probability from the absorption probability. This connection of the imaginary part of the coherent scattering amplitude with the total cross section is an immediate result of the principle of conservation of probability.[11] On the other hand, the real part of the coherent scattering amplitude gives the contribution of virtual processes or of transitions between states of the scattering system whose energy difference is not equal to the energy $\hbar\omega$ of the incident particle. These matrix elements thus occur in the calculation of the absorption at frequencies other than that of the incident particle. Thus the knowledge of the interaction cross section at the frequency ω does not determine the real part of the coherent scattering amplitude at ω. However, we expect that the amplitude might be determined by an integral over all frequencies of the total interaction cross section. It is just this integral relation which constitutes the dispersion relation.

The logical connection of the dispersion relation and the causality principle can be indicated as follows. Let the input be given in terms of a real variable t (which we shall regard as the time) by the function $F(t)$, and let the resulting output be $G(t)$. In order to relate causality and dispersion, we have to postulate certain general properties for our connection between input and output. A superposition principle is essential in our discussion, so we must assume that the input is a linear functional of the output. Let $T(t)$ be the response to an instantaneous unit pulse input at the time $t=0$ (i.e., Dirac delta function input). Next we assume that the system is not explicitly time-dependent; that is, we assume that a displacement in time of the whole input signal will cause a corresponding shift in the output [i.e., the response to $\delta(t-t_0)$ will be $T(t-t_0)$]. Then the superposition principle yields that an arbitrary input $F(t)$ will produce an output $G(t)$ given by

$$G(t)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{+\infty}T(t-t')F(t')dt'. \tag{2.1}$$

This resultant or "faltung" relation takes its simplest form[12] when expressed in terms of the Fourier transforms $f(\omega_r)$, $g(\omega_r)$, and $A(\omega_r)$ of $F(t')$, $G(t)$, and $T(t-t')$, respectively, and becomes

$$g(\omega_r)=A(\omega_r)f(\omega_r). \tag{2.2}$$

Thus the connection between input and output is characterized either by the time-delay distribution function $T(\tau)$ or by its Fourier transform $A(\omega_r)$, which we call the generalized scattering amplitude.

In general, the time delay distribution $T(\tau)$ may not be a well-defined function and a rigorous formulation of the above discussion requires the theory of distributions. However, in the case that $T(\tau)$ and $F(t)$ are square-integrable functions, the integrals and Fourier transforms are well defined[13,14] and we therefore limit

[8] R. Krönig, Ned. Tijdschr. Natuurk **9**, 402 (1942).
[9] N. G. van Kampen, Phys. Rev. **89**, 1072 (1953).
[10] N. G. van Kampen, Phys. Rev. **91**, 1267 (1953).
[11] E. Feenberg, Phys. Rev. **40**, 40 (1932); Bohr, Peierls, and Placzek, Nature **144**, 200 (1939); M. F. Mott and H. S. W. Massey, *Theory of Atomic Collisions* (Oxford University Press, New York, 1949), second edition, p. 133; M. Lax, Phys. Rev. **80**, 299 (1950).

[12] See, for example, N. Wiener, *The Fourier Integral and Certain of Its Applications* (Cambridge University Press, New York, 1933), pp. 1–2, 46–71.
[13] For convenience, we shall say that a function $F(t)$ is "square-integrable" if and only if $\int_{-\infty}^{+\infty}|F(t)|^2dt$ is finite. Furthermore, the only Fourier transforms that we will consider in this paper are the transforms of square-integrable functions. In this case the

our preliminary discussions to this case. The principle of causality that no output can occur before the input is clearly equivalent to the requirement that $T(\tau)$ must vanish for negative τ. To discuss the consequences of this limitation on $T(\tau)$, we introduce the concept of a *casual transform* which is defined as the boundary value of an analytic function belonging to a certain class:

(i) Let $\varphi(\omega)$ be a function of the complex variable $\omega=\omega_r+i\omega_i$ which is defined in the upper half of the complex plane and is analytic there. (The real axis is *not* assumed to belong to the domain of analyticity.)

(ii) There exists a positive number K such that, for all $\omega_i>0$,

$$\int_{-\infty}^{+\infty}|\varphi(\omega_r+i\omega_i)|^2d\omega_r\leq K. \tag{2.3}$$

Under these conditions the boundary value $\lim_{\omega_i\to 0+}\times\varphi(\omega_r+i\omega_i)$, which we will denote as $\varphi(\omega_r)$, exists almost everywhere on the real axis and is square-integrable.[14] Such a function $\varphi(\psi_r)$ will be called a "causal transform." We note here the fact, which we will need later on, that for $\omega_i>0$,

$$\int_{-\infty}^{+\infty}|\varphi(\omega_r+i\omega_i)|^2d\omega_r\leq\int_{-\infty}^{+\infty}|\varphi(\omega_r)|^2d\omega_r. \tag{2.4}$$

In other words, *a causal transform is a square-integrable function of a real variable which can be extended almost everywhere in the above sense to give a function which is analytic in the upper half of the complex plane and which is of uniformly bounded square integral along any line parallel to and above the real axis.* Titchmarsh[14] gives a beautiful criterion which can be stated in our terminology as: "A function of integrable square is zero for all negative values of its argument if and only if its Fourier transform is a causal transform." Thus the causality condition that $T(\tau)$ vanish for negative τ is equivalent to the requirement that $A(\omega_r)$ be a causal transform. However, Titchmarsh has also proved[14] another useful necessary and sufficient condition for a causal transform, namely that $\varphi(\omega_r)=\varphi_r(\omega_r)+i\varphi_i(\omega_r)$ is a causal transform if and only if its real and imaginary parts are Hilbert transforms of each other, that is

$$\varphi_r(\omega_r)=\frac{P}{\pi}\int_{-\infty}^{+\infty}\frac{\varphi_i(\nu)d\nu}{\nu-\omega_r}, \tag{2.5a}$$

and

$$\varphi_i(\omega_r)=\frac{-P}{\pi}\int_{-\infty}^{+\infty}\frac{\varphi_r(\nu)d\nu}{\nu-\omega_r}, \tag{2.5b}$$

where P implies that the principal part is to be taken at the point $\nu=\omega_r$. It can further be shown[14] that each of the relations (2.5) implies the other, so that a necessary and sufficient condition that $T(\tau)=0$ for $\tau<0$ is that $A=A_r+iA_i$ satisfies[15]

$$A_r(\omega)=\frac{P}{\pi}\int_{-\infty}^{+\infty}\frac{A_i(\nu)d\nu}{\nu-\omega_r}. \tag{2.6}$$

Thus in this case the causality condition is equivalent to the dispersion relation (2.6) [or to the alternative relation for $A_i(\omega_r)$ in terms of $A_r(\nu)$]. We see that the imaginary part of generalized scattering amplitude determines the real part (and vice versa). Either the real or the imaginary part can be chosen as an *arbitrary* square integrable function, but then the companion function is determined by the causality condition, and the analytic continuation into the upper half-plane then also exists (and can be calculated by applying Cauchy's integral formula, where the contour enclosing the point in the upper half-plane can be replaced by the real axis).

This example of the special case for any square-integrable $A(\omega_r)$ illustrates the power of the causality condition and shows how causality implies both dispersion relations and analyticity of $A(\omega)$ in the upper half of the complex frequency plane. However, in most physical problems the generalized scattering amplitude is not square-integrable and the more general discussion given in the next section is then required.

3. PROOF OF LOGICAL EQUIVALENCE OF STRICT CAUSALITY AND THE VALIDITY OF THE THE DISPERSION RELATION

We shall now show how the logical equivalence of strict causality and a dispersion relation can be expected in any proplem in which an "output" function is related to a freely variable "input" by a linear, bounded, time-invariant connection. (See Fig. 1.) From the invariance of the connection under time displacement, it follows[12] that each frequency component is mapped onto itself with only a change in magnitude and phase as given by Eq. (2.2) in terms of a generalized scattering amplitude $A(\omega_r)$. Hence we shall formulate our general connection between input and output by assuming that, if the input is

$$F(t)=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_{\infty}^{+\infty}f(\omega_r)e^{-i\omega_rt}d\omega_r, \tag{3.1}$$

Fourier transform is defined almost everywhere as a limit in the mean (see, for example, reference 14, p. 69); we shall always assume in this paper, without further remark, that the equalities and Fourier transforms are to be understood in this sense. Furthermore, two functions wich differ only over a set of measure zero will be identified without further remark since their Fourier transforms are necessarily equal and are defined only within arbitrary changes which can be made on any point set of measure zero.

[14] E. C. Titchmarsh, *Theory of Fourier Integrals* (Clarendon Press, Oxford, 1948), second edition, pp. 119–128.

[15] We shall use the subscripts r and i to designate real and imaginary parts, respectively, throughout this paper. In order not to confuse the functions defined originally on the real axis with their analytic extensions, we shall use ω_r as a real variable throughout and $\omega=\omega_r+i\omega_i$ for the point in the upper half of the complex plane. ν is real throughout this paper. For brevity, we shall hereafter use the word "frequency" to refer to circular frequency ω or to 2π multiplied by the usual frequency.

then the output is given by

$$G(t)=\frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}A(\omega_r)f(\omega_r)e^{-i\omega_r t}d\omega_r. \qquad (3.2)$$

We shall restrict consideration to input functions such that the integral over all time of $|F(t)|^2$ is finite. (In most problems, $|F|^2$ represents an intensity, so that we are limiting our consideration to input signals for which the total energy is finite.) Then the Fourier transform $f(\omega_r)$ is well defined and Eq. (3.1) and its reciprocal formula are valid. We shall assume that the connection is such that the integral over all time of $|G|^2$ is a uniformly bounded multiple of the integral of $|F|^2$; this implies that $|A(\omega_r)|$ is bounded and, if we choose the units of G appropriately, this bound can be set equal to unity. (In most examples this is equivalent to the natural assumption that the total output energy cannot exceed the total input energy.) It is remarkable that these are all the assumptions needed to prove the equivalence of causality and a dispersion relation.

An example of such a connection between input and output is a scattering problem, in which case the input is the "primary wave," the output is the "scattered wave" and the connection is determined by the scattering matrix. If we apply these considerations to an electric network, the input can be an impressed current as a function of time and the output can be any resulting voltage as a function of time, with the connection given by a complex impedance function. In other fields of physics, F and G can have other interpretations and they can depend on other parameters in addition to t, for example, on space coordinates or spin or vector indices; A can be a general matrix giving the dependence of an array of G's on an array of F's. However, only the time-dependence of the signals is essential to the present discussion.

Thus we shall leave the nature of our generalized scattering system unspecified, assuming only that its scattering amplitude $A(\omega_r)$ connects to any square integrable input $F(t)$ an output given by Eqs. (3.1) and (3.2), where $A(\omega_r)$ is an arbitrary function of the real circular frequency ω_r, subject only to the restriction $|A(\omega_r)|\leq 1$.[16,17] We shall prove the following basic equivalence theorem.

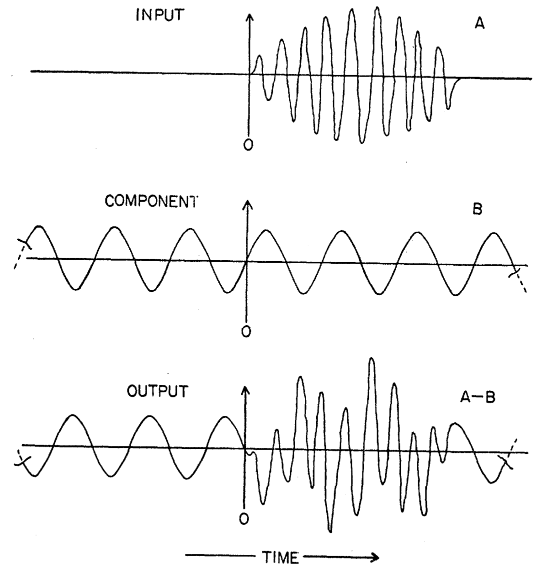

FIG. 1. This figure illustrates schematically the basic reason for the logical connection of causality and dispersion. An input A which is zero for times t less than zero is formed as a superposition of many Fourier components such as B, each of which extends from $t=-\infty$ to $t=\infty$. These components produce the zero-input signal by destructive interference for $t<0$. It is impossible to design a system which absorbs just the component B without affecting other components, for in this case the output would contain the complement of B during times before the onset of the input wave, in contradiction with causality. Thus causality implies that absorption of one frequency must be accompanied by a compensating shift of phase of other frequencies; the required phase shifts are prescribed by the dispersion relation.

Theorem.—Let $A(\omega_r)$ be any complex Lebesgue-measurable function of bounded absolute value (≤ 1) for all values of the real variable ω_r. Then, if $A(\omega_r)$ is the generalized scattering amplitude which connects to any freely variable input $F(t)$ of integrable square an output given by Eqs. (3.2) and (3.1), the following seven statements are logically equivalent[18]:

(i) Strict causality: No output before the input, or $F(t)=0$ for $t<0$ implies $G(t)=0$ for $t<0$.

(ii) (Upper half-plane) bounded regularity condition, hereafter called for simplicity the *regularity condition*: $A(\omega_r)$ is the boundary value function for almost all real ω_r of a function which is analytic and of

---

[16] If we wish to limit consideration to real-valued inputs F, then $f(-\omega_r)=f^*(\omega_r)$; however two real inputs can always be combined as the real and imaginary parts of a complex input to give a general $f(\omega_r)$, so we will hereafter consider $F(t)$ or $f(\omega_r)$ to be arbitrary square integrable functions. Similarly, if our connection relates a real output to a real input, $A(\omega_r)$ must satisfy the symmetry condition: $A(-\omega_r)=A^*(\omega_r)$; by choice of a real input and separate consideration of the real and imaginary parts of $G(t)$, a general $A(\omega_r)$ can be reduced to $B+iC$, where B and C are amplitudes satisfying the causality principle and each of which fulfills the symmetry condition. In this sense the symmetry condition can always be introduced. We note that the symmetry condition becomes on analytic continuation to complex values: $A(-\omega)=A^*(\omega^*)$, so that singularities of A occur in pairs at points correlated by reflection in the imaginary ω-axis.

[17] In the case when $A(\omega_r)$ is a matrix connecting an array of F's to an array of G's, the condition $|A(\omega_r)|\leq 1$ need be required only for each element of the matrix $A(\omega_r)$ separately. Then, if each element of F is independently variable, the proof in this section can be carried out for each element of $A(\omega_r)$ separately.

[18] The proof in this section is similar in some respects to the proof given independently by N. G. van Kampen, Phys. Rev. **90**, 1072–1079 (1953), in his discussion of the causality condition for individual angular momenta in the case of the Maxwell field scattered by a spherically symmetric scattering center. However, van Kampen's proof assumes that his $S(k)$ or our $A(\omega_r)$ is of absolute value 1 for all real frequencies. Our proof is more general in that we assume only $|A(\omega_r)|\leq 1$. Indeed, since $|A(\omega_r)|=1$ implies that the imaginary part of the phase shift is zero (see Sec. 4), this is equivalent in most problems to stating that the total energy of the output is equal to the total energy of the input or that the absorption is zero. In most physical problems the variation of the absorption with frequency is significant, and the more general proof given in this section is then required.

absolute value less than unity throughout the upper half of the complex ω plane.

(iii) Generalized dispersion relation in terms of real part: $A(\omega_r)$ is given almost everywhere by

$$A(\omega_r)=A_r(\omega_r)+iA_i(\omega_r)=\frac{1}{\pi i}\lim_{\omega_i\to 0+}\int_{-\infty}^{+\infty}\times\frac{1+\nu(\omega_r+i\omega_i)}{1+\nu^2}A_r(\nu)\frac{d\nu}{\nu-\omega_r-i\omega_i}+iA_{i0},\quad(3.3)$$

where A_{i0} is a real number.[19] In case $A(\omega_r)$ is chosen to satisfy the symmetry relation $A(-\omega_r)=A^*(\omega_r)$, this equation becomes

$$A(\omega_r)=\frac{2}{\pi i}\lim_{\omega_i\to 0+}(\omega_r+i\omega_i)\int_0^\infty\frac{A_r(\nu)d\nu}{\nu^2-(\omega_r+i\omega_i)^2}.\quad(3.3a)$$

(iv) Generalized dispersion relation in terms of imaginary part: $A(\omega_r)$ is given almost everywhere in terms of its imaginary part by

$$A(\omega_r)=\frac{1}{\pi}\lim_{\omega_i\to 0+}\int_{-\infty}^{+\infty}\frac{1+\nu(\omega_r+i\omega_i)}{1+\nu^2}\times A_i(\nu)\frac{d\nu}{\nu-(\omega_r+i\omega_i)}+A_{r0},\quad(3.3b)$$

where A_{r0} is a real constant. If $A(-\omega_r)=A^*(\omega_r)$, this equation becomes

$$A(\omega_r)=\frac{2(1+\omega_r^2)}{\pi}\lim_{\omega_i\to 0+}\times\int_0^\infty\frac{\nu A_i(\nu)d\nu}{(\nu^2+1)(\nu^2-(\omega_r+i\omega_i)^2)}+A_{r0}.\quad(3.3c)$$

(v) Integral criterion for all points in lower half-plane: The function $A(\omega_r)$ is such that $\int_{-\infty}^{+\infty}d\nu[A(\nu)/(\nu-\tilde{\omega})^2]$ vanishes for *all* complex numbers $\tilde{\omega}$ of negative imaginary part ($\tilde{\omega}_i<0$).

(vi) Integral criterion for any fixed point in the lower half-plane: The function $A(\omega_r)$ is such that, for one particular fixed complex number $\tilde{\omega}$ of negative imaginary part, $\int_{-\infty}^{+\infty}d\nu[A(\nu)/(\nu-\tilde{\omega})^{n+1}]$ vanishes for *all* positive integers n.

(vii) Strict causality for an exponentially decaying input: For some complex number $\tilde{\omega}$ of negative imaginary part, $A(\nu)/(\nu-\tilde{\omega})$ is a causal transform.

Proof.—First we will prove that (i) and (ii) are equivalent. For this we note that (i) is equivalent to the statement: *$A(\omega_r)f(\omega_r)$ is a causal transform whenever $f(\omega_r)$ is a causal transform.* It is clear that (ii) implies (i), for, if $A(\omega_r)$ extends into an analytic and bounded function in the upper half complex plane, multiplication of $f(\omega_r+i\omega_i)$ by this $A(\omega_r+i\omega_i)$ cannot change either its analyticity or the uniform bound on its absolute square integral, and thus $A(\omega_r)f(\omega_r)$ remains a causal transform if $f(\omega_r)$ is a causal transform. To show that (i) implies (ii), we note that we can choose for $f(\omega_r)$ the particular causal transform $(\omega_r-\beta+i\gamma)^{-1}$ where β and γ are both real and γ is positive.[20] If strict causality holds, $g(\omega_r)=A(\omega_r)\cdot(\omega_r-\beta+i\gamma)^{-1}$ must then be a causal transform; hence it is the boundary value of a function which is analytic in the upper half-plane. We can then define the analytic function $A(\omega_r+i\omega_i)$ in the upper half-plane by the product of analytic functions $g(\omega_r+i\omega_i)$ and $(\omega_r+i\omega_i-\beta+i\gamma)$. Hence it remains only to show $|A(\omega_r+i\omega_i)|\leq 1$. For this purpose, we use the fact that $A(\omega_r)f(\omega_r)$ is a causal transform; hence by the theorem of Titchmarsh,[14] its analytic extension in the upper half-plane is given by

$$A(\omega)f(\omega)=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{A(\nu)f(\nu)d\nu}{(\nu-\omega)}.\quad(3.4)$$

Hence,

$$|A(\omega)|\leq\frac{1}{2\pi|f(\omega)|}\int_{-\infty}^{+\infty}|A(\nu)|\cdot\left|\frac{f(\nu)}{\nu-\omega}\right|d\nu.\quad(3.5)$$

This equation must hold for *any* causal transform $f(\nu)$. We use the fact that $|A(\nu)|$ is less than unity along the axis and then choose the causal transform $f(\nu)$ to be $(\nu-\omega^*)^{-1}=(\nu-\omega_r+i\omega_i)^{-1}$. [This can be shown to be the choice of $f(\nu)$ which minimizes the right-hand side when $|A|=1$.] Then the inequaltity (3.5) becomes

$$|A(\omega)|\leq\frac{\omega_i}{\pi}\int_{-\infty}^{+\infty}\frac{d\nu}{(\nu-\omega_r)^2+\omega_i^2}=1.\quad(3.6)$$

Hence we have shown that strict causality implies the regularity condition.

Next we must show that the regularity condition is equivalent to the generalized dispersion relation. For this purpose we use the fact that any function which is analytic and of bounded absolute value in the upper half of the complex plane is given (to within an imaginary constant) by an "analytic Poisson formula" involving only the value of its real part on the real axis.[21,22] This formula is

For $\omega_i>0$,

$$A(\omega)=\frac{1}{\pi i}\int_{-\infty}^{+\infty}\frac{1+\nu\omega}{1+\nu^2}A_r(\nu)\frac{d\nu}{\nu-\omega}+iA_{i0}.\quad(3.7)$$

[19] The integrand of Eq. (3.3) can be written in dimensional form so as to contain the factor $(\kappa^2+\nu\omega)/(\kappa^2+\nu^2)$, where κ is an arbitrary reference frequency; for simplicity, we have assumed throughout that the units are so chosen that the reference frequency becomes unity.

[20] This choice for $f(\omega_r)$ corresponds to an idealized input which is zero for negative t and which equals the decreasing and oscillating signal $\exp(-\gamma t-i\beta t)$ for positive t.

[21] For the case when the function is analytic and bounded in the unit circle, the corresponding result follows essentially from Fatou's theorem and his work on the Poisson integral (see

Here A_{i0} is a real constant which can vary arbitrarily within the range permitted by our original assumption that $|A(\omega_r)|\leq 1$ (This range will depend on the particular function $A_r(\omega_r)$.) By taking the limit of Eq. (3.7) as ω_i approaches zero from above, we obtain Eq. (3.3). [The similar relation (3.3b) for $A_i(\omega_r)$ can be obtained by considering the function $w(z)=iA(\omega)$.] Thus we see that the regularity condition does imply the generalized dispersion relations. By considering real inputs and treating real and imaginary parts of the output separately, we can always reduce our problem to functions $A(\omega_r)$ which satisfy the symmetry condition, so that the real and imaginary parts are, respectively, even and odd functions of the frequency. Then A_{i0} vanishes and the generalized dispersion relations take the form of Eqs. (3.3a) and (3.3c).

Now we will show that the generalized dispersion relation Eq. (3.3) implies the regularity condition. We define $A(\omega_r+i\omega_i)$ in terms of the given $A(\omega_r)$ by Eq. (3.7); this function is analytic in the upper half complex plane and Eq. (3.3) just states that it approaches almost everywhere on the boundary our given $A(\omega_r)$. It can be shown from the properties of the Poisson integral from which formula (3.7) was obtained that $|A(\omega_r+i\omega_i)|$ is also uniformly bounded.[23] Hence the generalized dispersion relation (3.3) implies condition (ii). In an entirely similar way, it can be shown that (ii) and (iv) are equivalent and our proof of the logical equivalence of the strict causality, the regularity condition, and each of the generalized dispersion relations is then complete.

The generalized dispersion relations (3.3) [or (3.3a) and (3.3b), or (3.3c)] are therefore each the necessary and sufficient condition for strict causality. The arbitrary constant A_{r0} in the determination of $A_r(\omega_r)$ from $A_i(\omega_r)$ cannot be eliminated by arguments of causality alone, for the simple physical meaning of this constant is that the addition to a causal output of a constant multiple of the input can never violate the causality condition [i.e., for any constant A_0, $G'(t)=A_0F(t)+G(t)$ is just as acceptable an output as $G(t)$, as far as the causality principle alone is concerned].

Each of the generalized dispersion relations (iii) and (iv) is a necessary and sufficient criterion which the amplitude $A(\omega_r)$ must satisfy in order for strict causality to be valid, and is of a form which is especially useful in many physical problems. However, many other equivalent mathematical criteria can be stated; as examples of such alternative criteria, we give the statements (v) and (vi), since these are both succinct and help to illustrate the mathematical consequences of strict causality.

We will sketch now the proof of the equivalence of the regularity condition (ii) and the integral criterion for the whole lower half-plane (v). For this we will use the fact[14] that a square-integrable function $g(\nu)$ is a causal transform if and only if the following equation holds for all complex numbers $\tilde{\omega}$ with negative imaginary part:

$$\int_{-\infty}^{+\infty} d\nu[g(\nu)/(\nu-\tilde{\omega})]=0. \tag{3.8}$$

[This result follows rigorously from the analytic properties of causal transforms; see, for example, reference 14, especially page 128. The result is easily remembered by imagining the contour of integration closed at $+i\infty$, and noting that Eq. (3.8) then resembles the Cauchy integral, where $\tilde{\omega}$ is outside the contour of integration.] Suppose the regularity condition (ii) is valid. Then $g(\nu)=A(\nu)f(\nu)$ is a causal transform whenever $f(\nu)$ is a causal transform. For each $\tilde{\omega}$ in the lower half-plane we choose for $f(\nu)$ the particular causal transform $(\nu-\tilde{\omega})^{-1}$ and substitute in Eq. (3.8) to obtain

$$\int_{-\infty}^{+\infty} d\nu[A(\nu)/(\nu-\tilde{\omega})^2]=0. \tag{3.9}$$

This shows (ii) implies (v). Conversely, we will show (v) implies condition (vii) which in turn implies (ii). Assume that (v) or that Eq. (3.9) holds for all points $\tilde{\omega}$ in the lower half-plane. Select any particular point $\tilde{\mu}$ in the lower half-plane; then we integrate Eq. (3.9) over $\tilde{\omega}$ from $\tilde{\mu}$ to $\tilde{\lambda}$ along a smooth contour of finite length that remains a finite distance from the real axis. The interchange of the order of the integrations can be proved to be allowable (since the integrand is uniformly continuous) and we obtain, after replacing $\tilde{\lambda}$ by $\tilde{\omega}$,

$$\int_{-\infty}^{+\infty} \frac{A(\nu)d\nu}{(\nu-\tilde{\mu})(\nu-\tilde{\omega})}=0. \tag{3.10}$$

---

reference 22). For the unit circle, the "analytic Poisson formula" is

$$w(z)=\frac{1}{2\pi}\int_0^{\infty}\frac{e^{i\theta}+z}{e^{i\theta}-z}u(e^{i\theta})d\theta+iA_{i0}.$$

The analytic kernel figuring in this formula has the ordinary Poisson kernel for its real part. We have mapped this formula from the unit circle in the complex z plane into the complex ω plane by the identification:

$$\omega=i(1+z)/(1-z)$$

and

$$A(\omega)=w(z)=u(z)+iv(z).$$

[22] R. Nevanlinna, *Eindeutige Analytische Funktionen* (Verlag Julius Springer, Berlin, 1936).

[23] This is done by mapping the upper half of the ω plane into the unit circle in the z plane as in footnote 21. The boundedness of $|u(z)|$ follows immediately by taking the absolute value signs inside the integral for u; this yields that $|u(z)|$ cannot exceed the least upper bound of $|u[\exp(i\theta)]|$ on the perimeter, which in turn is not greater than 1. For the imaginary part a more delicate argument is required. The function u is square integrable both on the perimeter and in the interior. By Parseval's relation, the conjugate function v is then also square integrable and is therefore given by an ordinary Poisson formula in terms of its boundary values. The boundedness of $|v|$ on the perimeter then implies that it is bounded in the interior. Then, since $|u|$ and $|v|$ are both bounded, $w(z)$ is analytic and of bounded absolute value in the interior; it is therefore given by the ordinary Poisson formula in terms of w on the perimeter, and it then follows that $|w(z)|$ in the interior cannot exceed its least upper bound on the perimeter.

Equation (3.10) holds for all $\tilde{\omega}$ in the lower half-plane; comparing with Eq. (3.8), we see that Eq. (3.10) implies that $A(\nu)/(\nu-\tilde{\mu})$ is a causal transform or that statement (vii) holds for any point $\tilde{\mu}$. If we assume (vii) to hold, then Eq. (3.10) holds for one particular $\tilde{\mu}$ and all $\tilde{\omega}$ in the lower half-plane. If we multiply Eq. (3.10) by $\tilde{\mu}-\tilde{\omega}$ and then differentiate with respect to $\tilde{\omega}$, the interchange of integration and differentiation can be shown to be valid and we obtain Eq. (3.9) as a result. Hence we have shown that (vii) implies (v). By this equivalence we have shown that, if $A(\nu)/(\nu-\tilde{\omega})$ is a casual transform for any particular value of $\tilde{\omega}$ in the lower half-plane, then it is a causal transform for *every* $\tilde{\omega}$ in the lower half-plane. However, in this case Eq. (3.4) holds with $f(\nu)=(\nu-\tilde{\omega})^{-1}$, where ω is an arbitrary point in the upper half-plane. Since $\tilde{\omega}$ can now be chosen to be ω^*, Eq. (3.4) leads to Eq. (3.6), thus demonstrating that $A(\omega)$ is analytic and of modulus less than unity in the upper half-plane. Hence we have shown that statements (v) and (vii) are equivalent to the regularity condition (ii).

It remains only to show that statement (vi) is equivalent to (v). This is easily done by noting that the integral in Eq. (3.9) defines an analytic function in the lower half-plane which is zero throughout this region if and only if it is zero in a small neighborhood of a particular point $\tilde{\omega}$. But the function is zero in this neighborhood if and only if all coefficients of its Taylor series expansion about $\tilde{\omega}$ vanish, that is, if all its derivatives at $\tilde{\omega}$ vanish. The differentiation can be performed under the integral sign and the $(n-1)$th derivative is proportional to $\int_{-\infty}^{+\infty}d\nu[A(\nu)/(\nu-\tilde{\omega})^{n+1}]$. Hence the vanishing of these integrals for a particular $\tilde{\omega}$ for all positive integers n is a necessary and sufficient condition that the Taylor series will give zero and thus yield an analytic extension of zero. Therefore (v) and (vi) are equivalent.

This concludes our proof of the basic equivalence theorem. We have shown that the seven statements given above are logically equivalent in the sense that, if a bounded measurable function $A(\omega_r)$ satisfies one of these conditions, then it satisfies all of them. We call a function $A(\omega_r)$ satisfying the conditions of this theorem a *causal factor*; this name is chosen to emphasize that an essential feature of such a function is that its product with any causal transform is another causal transform.

The condition (vii) is a useful facet of the theorem and has a simple physical interpretation. In order to test whether a given $A(\nu)$ is a causal transform, we must test whether for *every* causal transform $f(\nu)$ the product $A(\nu)f(\nu)$ is also a causal transform. However (vii) tells us that any particular causal transform of the form $(\nu-\tilde{\omega})^{-1}$ is sufficient to test $A(\nu)$ and that, if $A(\nu)(\nu-\tilde{\omega})^{-1}$ is a causal transform, then $A(\nu)$ is a causal factor and $A(\nu)f(\nu)$ is a causal transform for every causal transform $f(\nu)$. The transform $(\nu-\tilde{\omega})^{-1}$ corresponds to an input which vanishes for $t<0$ and which for $t>0$ is the exponentially decreasing signal $\exp[-|\tilde{\omega}_i|t+i\omega_r t]$. The simplest input of this class is given by $\tilde{\omega}=i$, or is the signal e^{-t} for $t>0$. Condition (vii) tells us that, if our system is causal for this particular input, then it will be causal for every input. In particular, if our system is designed to delay any input of this class by a time T, then it will delay every input by at least a time T.

4. PHASE-SHIFT FORM OF THE DISPERSION RELATION

In the last section we showed that under rather general conditions strict causality implies that the imaginary part of a generalized scattering amplitude can be determined from its real part (or vice versa) by a dispersion relation. However, in many physical problems the real and imaginary parts of the scattering amplitude may be of less direct interest than the absolute value and the phase of the scattering amplitude. We will now investigate the relationship between these quantities that results from strict causality.

We define the complex phase shift $\eta(\omega)=\eta_r-i\eta_i$ by:

$$A(\omega)=\exp[i\eta(\omega)]=\exp[-\eta(\omega)]\cdot\exp[i\eta_r(\omega)], \quad (4.1)$$

$$\eta_i(\omega)=-\ln|A(\omega)|; \quad \eta_r(\omega)=\arg A(\omega). \quad (4.2)$$

Thus the imaginary part of the phase shift determines the absolute values of the scattering amplitude (and thus is normally related to the absorption), while the real part of the phase shift determines the argument of the scattering amplitude. $|A(\omega)|\leq 1$ implies $\eta_i(\omega)\geq 0$. The symmetry condition $A(-\omega_r)=A^*(\omega_r)$ implies that $\eta_i(\omega_r)$ is an even function and $\eta_r(\omega_r)$ an odd function of the circular frequency ω_r.[24] Let us assume the conditions of the theorem of the last section are satisfied and that strict causality is valid; so that the regularity condition and the dispersion relations hold. Then we ask whether knowledge of $\eta_i(\omega_r)$ completely determines the total phase shift. We will find that it does not, for there is in fact a large infinite family of real phase shifts permitted with a given imaginary phase shift. However, the real phase shifts are still far from arbitrary; we will characterize this family of solutions and show that there exists a canonical phase shift given by a dispersion relation which has the minimum real phase shift conjugate to the given imaginary phase shift, and all other permissible real phase shifts must increase with frequency at least as rapidly as the canonical phase shift.

The fact that the absolute value of the scattering amplitude still leaves considerable freedom in the scattering phase can be seen simply as follows. From the regularity condition we know that the scattering amplitude $A(\omega_r)$ is consistent with strict causality if and only if it is the boundary value function of a

[24] Since the symmetry condition can be introduced at will in problems of the type we are considering with a freely variable input, we will assume it to hold throughout this section.

function $A(\omega)$ which is bounded and analytic in the upper half of the complex ω plane ($\omega_i>0$). Suppose we multiply $A(\omega)$ by a factor $B_n(\omega)=(\omega-\mu_n)/(\omega-\mu_n^*)$, where μ_n is an arbitrary point in the upper half complex plane. Since $|B_n(\omega_r)|=1$ for all real ω_r and $|B_n(\omega)|<1$ throughout the upper half-plane, we see that $A(\omega)B_n(\omega)$ also satisfies the regularity condition and yields the same $|A(\omega)|$ on the boundary. Hence the absolute value of $A(\omega_r)$ does not determine $A(\omega_r)$ or the phase shift at all uniquely, for we are at liberty to introduce a zero at any point in the upper half-plane.

Suppose that the non-negative function $\eta_i(\omega_r)$ is given and we wish to find the most general $A(\omega)$ satisfying the regularity condition which reduces almost everywhere on the boundary to a function of absolute value $\exp[\eta_i(\omega_r)]$. Fortunately, essentially this mathematical problem has been solved.[25] First, in order for the problem to have any solution at all, the function must satisfy the inequality[26]

$$\int_0^\infty \frac{\eta_i(\omega_r)d\omega_r}{\omega_r^2+1}<\infty. \tag{4.3}$$

Then, if this integral is finite, the most general solution $A(\omega)$ is of the following form[25]:

$$A(\omega)=\tilde{A}(\omega)C(\omega), \tag{4.4}$$

where $\tilde{A}(\omega)$ is a bounded, analytic, *nonzero* function in the upper half-plane and $C(\omega)$ is any bounded, analytic function in the upper half-plane whose boundary value has the absolute value of unity almost everywhere on the real axis. [Thus $|A(\omega_r)|=|\tilde{A}(\omega_r)|$ for almost all ω_r.] These factors can be written

$$\tilde{A}(\omega)=\exp[i\tilde{\eta}(\omega)], \tag{4.5}$$

where

$$\tilde{\eta}(\omega)=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1+\omega\nu}{1+\nu^2}\eta_i(\nu)\frac{d\nu}{\nu-\omega}=\frac{2\omega}{\pi}\int_0^\infty\frac{\eta_i(\nu)d\nu}{\nu^2-\omega^2}, \tag{4.6}$$

$$C(\omega)=B(\omega)D(\omega), \tag{4.7}$$

where

$$D(\omega)=\exp\left[\frac{2i\omega}{\pi}\int_0^\infty\frac{d\alpha(\nu)}{\nu^2-\omega^2}+id_0\omega\right], \tag{4.8}$$

where d_0 is a non-negative number and $\alpha(\nu)$ is a nondecreasing bounded function of ν with a derivative that exists and *vanishes* for almost all ν. $B(\omega)$ is a "Blaschke product," which is a product of a denumerable number of terms, each representing a zero of $A(\omega)$ in the upper half-plane:

$$B(\omega)=\prod_n B_n(\omega)=\prod_n(\omega-\mu_n)/(\mu_n^*-\omega),\quad \text{where each } \mu_{ni}>0. \tag{4.9}$$

This term $B(\omega)$ is of the form already studied by van Kampen,[27] who gives an excellent discussion of the properties of these products including the fact that, in order that this product converge, the zeros must be distributed so that

$$\sum_n\mu_{ni}/|\mu_n|^2<\infty. \tag{4.10}$$

Thus, from these results the phase shift $\eta(\omega)$ can be defined in the upper half-plane by

$$\eta(\omega)=\tilde{\eta}(\omega)+\vartheta(\omega)=\tilde{\eta}(\omega)+d_0\omega+\frac{2\omega}{\pi}\int_0^\infty\frac{d\alpha(\nu)}{\nu^2-\omega^2}+\xi(\omega), \tag{4.11}$$

where $\xi(\omega)$ is the phase shift of the Blaschke product[22]:

$$\xi(\omega)=-i\sum_n\ln[(\omega-\mu_n)/(\omega-\mu_n^*)]=-\tfrac{1}{2}i\sum_n\ln[(\omega-\mu_n)(\omega+\mu_n^*)/(\omega-\mu_n^*)(\omega+\mu_n)]. \tag{4.12}$$

Nevanlinna shows that the boundary values of $\xi(\omega)$ are real.[22] We call the function $\tilde{\eta}(\omega)$ the "canonical phase shift"; it is given by the dispersion integral (4.6) which defines a function which is analytic throughout the upper half-plane and of non-negative imaginary part there. The terms involving d_0 and $\alpha(\nu)$ in Eq. (4.11) are just anomalous terms that can be given the physical interpretation of infinitely narrow absorption lines at various finite frequencies [$\alpha(\nu)$ term] and at infinite frequency (d_0 term). Thus, in this interpretation, a complete specification of the absorption requires the determination of $\eta_i(\omega_r)$ and the sources of the anomalies, $\alpha(\nu)$ and d_0; if the absorption is given in this completeness, the phase shift is then determined by Eq. (4.11) up to $\xi(\omega_r)$, the contribution of an arbitrary Blaschke product. If we insist that the boundary value function $\eta(\omega_r)$ be continuous, the term involving $\alpha(\nu)$ will be eliminated. As ω approaches ω_r on the real axis, Eq. (4.11) becomes the "phase shift form" of the generalized dispersion relation, and the term $\vartheta(\omega)$ approaches a real limit $\vartheta_r(\omega_r)$ almost everywhere while $\tilde{\eta}_i(\omega)$ reduces almost everywhere to the originally given $\eta_i(\omega_r)$. We note that each of the three terms in $\vartheta_r(\omega)$ is a non-

[25] For a discussion of related work, see reference 22. I owe the essential features of the present formulation of this theorem to Professor Marcel Riesz. See also J. A. Shohat and J. D. Tamarkin, *Problem of Moments, Mathematical Surveys, No. 1* (American Mathematical Society, New York, 1943), p. 23.

[26] It is clear that there must be some restriction on $\eta_i(\omega r)$; for example, if $|A(\omega_r)|=0$ over an interval of real ω_r, then the analytic continuation $A(\omega)$ and then $A(\omega_r)$ would be zero everywhere. The necessary and sufficient condition that $A(\omega_r)$ is not "too close to zero too often" is given by the finiteness of the integral (4.3).

[27] N. G. van Kampen, Phys. Rev. 90, 1072 (1953), Eqs. (14) and (18). In case the symmetry condition is not fulfilled, an additional factor q_n must be combined with each term B_n in order to guarantee the convergence of the real part of the phase of the infinite product $B(\omega)$. This factor q_n is of absolute value 1 and is given by $q_n=-|1+\mu_n^2|/(1+\mu_n^2)$. If we assume the symmetry condition, we find that we can omit these terms, since the convergence is guaranteed without them provided we combine each pair of terms correlated by reflection in the imaginary axis. Then the remaining terms are on the imaginary axis and only a finite number of these can give $q_n=-1$, and $q_n=1$ for the remainder.

negative, nondecreasing function of frequency. Thus, of all the possible real phase shifts $\eta_r(\omega_r)$ correlated with $\eta_i(\omega_r)$ by strict causality, the canonical phase shift $\tilde{\eta}_r(\omega_r)$ given by Eq. (4.6) is the one of minimum value and of minimum rate of increase with frequency. The usual dispersion integral (4.6) gives us in this sense a minimal solution; this is the phase shift corresponding to the amplitude $\tilde{A}(\omega)$ which is the solution of maximum possible absolute value throughout the upper half of the complex ω plane. As we have seen, other solutions are obtained from this solution by introduction of the absorption anomalies given by d_0 and $\alpha(\nu)$ and by the introduction of a Blaschke product of zeros. The phase shifts due to the zeros are far from arbitrary [e.g., in addition to its nondecreasing character, $\xi(\omega_r)$ satisfies the condition that its total variation over all real frequencies$=2\pi N$, where N is the number of zeros in the product, which integer can be either finite of infinite]. But causality does not locate these zeros; other information (e.g., energy of bound states in particle scattering problems) is needed to determine them. However, even if this term $\xi(\omega_r)$ is not determined, the fact that $\tilde{\eta}_r(\omega_r)$ is minimal gives powerful inequalities restricting the phase shift once the absorption is given.

One convenient feature of the phase shift formulation of the dispersion relations is that it is easily extended to the case when a limited time advance, a, is permitted, i.e., when the causality condition is replaced by: $G(t)=0$ for $t<-a<0$ when $F(t)=0$ for $t<0$. (Such a weakened causality condition is appropriate for scattering by a target of finite range.) In this case $A(\omega)$ need not be a causal factor, for it can in fact diverge exponentially in the upper half of the complex frequency plane; but $A(\omega)\cdot\exp(i\omega a)$ is a causal factor. Thus $(\eta+\text{a})$ satisfies the same conditions as η did before, or the dispersion relations (4.11) are unchanged except for the addition of a constant term $-a$.

5. RELAXATION OF THE BOUNDEDNESS CONDITION

The proof of Sec. 3 used repeatedly the condition that the absolute value of the scattering amplitude is bounded. While this condition is normally equivalent to conservation of energy and is therefore to be expected in any physically sensible problem, there are idealized problems in physics in which the scattering amplitude is unbounded. We shall show in this section how dispersion relations can still be derived in such cases, provided the divergence is no worse than a power of the frequency; these methods are adequate to cover most problems of physical interest.

When the generalized scattering amplitude is square-integrable (Sec. 2), causality implies that the imaginary part determines the real part completely. When this integrability condition was relaxed to boundedness (Sec. 3), we found that causality implies that the real and imaginary parts determine each other to within a constant. As the integrability condition is further relaxed, we shall find that additional arbitrary constants enter the dispersion relations.

It is easy to illustrate why causality does not give a complete connection between real and imaginary parts in the general case. Consider the output $G(t)$ to be the nth derivative of $F(t)$; then clearly $G(t)$ is zero so long as $F(t)$ remains zero, so causality is satisfied. The generalized scattering amplitude in this case is $(-i\omega_r)^n$. Similarly, the n-fold integral of $F(t)$ from $-\infty$ to t yields a causal output, with $A(\omega_r)=(-i\omega_r)^{-n}$. A more general scattering amplitude satisfying the causality principle is thus given by a finite sum of these functions:

$$A(\omega_r)=\sum_{n=-M}^{N} a_n\omega_r^n, \tag{5.1}$$

where the a_n's are arbitrary complex numbers; hence the real and imaginary parts of the sum are entirely independent of one another, with no dispersion relation between them. We shall see that the arbitrary constants a_n in this example are typical of the general case.

(a) Functions Diverging at Infinity

Of particular interest in physics are idealized problems for which $A(\omega_r)$ may diverge at high frequencies. For, example, in the limit of an infinitely distant scattering center, the forward scattering amplitude diverges at least as fast as the product of the frequency and the total cross section σ (and σ often approaches a constant or diverges logarithmically).[28] To derive a dispersion relation for amplitudes which diverge at high frequency, it is sufficient to assume that $A(\omega_r)$ is integrable over any finite frequency interval, that for some positive integer j the function $A(\omega_r)/\omega_r^{j-1}$ is bounded in the limit of infinite frequency, and that there is some real frequency λ such that the jth derivative of $A(\omega_r)$ exists in some neighborhood about λ.[29]

[28] The idealization of infinite distance has produced this divergence of the amplitude; for any actual observation of scattering at a distance r from a scatterer of lateral dimension d, the difference in path lengths to the observer from points at the center of the scatterer and at the edge leads to destructive interference of these infinitesimal contributions which damps the forward scattering amplitude for frequencies greater than cr/d^2.

[29] The proof in this section is less elegant mathematically than that in Sec. 3 in that in Sec. 3 no mathematical assumptions were made whose physical content was not clear and reasonable, while here we introduce the assumption that there exists some neighborhood in which the jth derivative of A exists. This condition is not necessary for the existence of a dispersion relation, but it greatly shortens the proof and leads to the simple form of the dispersion relation given in Eq. (5.6). The existence of such a neighborhood on the real axis is valid in all physically interesting cases that we have encountered, and Eq. (5.6), or the equation that results from it when the symmetry condition holds, is expected to be adequate for any physical application. It should be noted that all assumptions have been made only on the real frequency axis where $A(\omega_r)$ has direct physical meaning and no assumptions are implied about the possibility of analytic continuation into the complex frequency plane; instead the analytic continuation is a result of the proof of the dispersion relation from the causality principle.

Since the generalized scattering amplitude may now be unbounded, a square-integrable input need not always result in a square-integrable output, and the Fourier transform of the output may not be well defined in such cases. To avoid these difficulties, we can limit consideration to those square-integrable inputs such that the output is square-integrable. In particular, we consider the input signal whose Fourier transform $f(\omega_r)$ is $(\omega_r+\mu)^{-j}$, where μ is any complex number of positive imaginary part. Then the output's Fourier transform $g(\omega_r)=A(\omega_r)/(\omega_r+\mu)^j$ is square-integrable; since $(\omega_r+\mu)^{-j}$ is a causal transform, the input vanishes for $t<0$ and hence the causality principle requires that the output vanish for $t<0$ or that $A(\omega_r)/(\omega_r+\mu)^j$ be a causal transform. We now define a new amplitude function $B(\omega_r)$ by

$$B(\omega_r)=A(\omega_r)-\sum_{p=0}^{j-1}\left(\frac{d^pA}{d\omega_r{}^p}\right)_{\omega_r=\lambda}\frac{(\omega_r-\lambda)^p}{p!}. \quad (5.2)$$

Then $B(\omega_r)/(\omega_r+\mu)^j$ is a causal transform, since the sum of a finite number of causal transforms is a causal transform and $A(\omega_r)/(\omega_r+\mu)^j$ and $(\omega_r-\lambda)^p/(\omega_r+\mu)^j$ for $p\leq j-1$ are causal transforms. Furthermore, $B(\omega_r)/(\omega_r-\lambda)^j$ is bounded near $\omega_r=\lambda$ (here we use the existence of the jth derivative of A or B) and is a square-integrable function. We shall now show that $B(\omega_r)/(\omega_r-\lambda)^j$ is a causal transform.

This proof is done in steps by introducing the functions

$$C_m(\omega_r)=\frac{B(\omega_r)}{(\omega_r-\lambda)^{m-1}(\omega_r+\mu)^{j-m}} \quad \text{for} \quad m=1,2,\cdots,j. \quad (5.3)$$

$B(\omega_r)$ was so constructed that $C_m(\lambda)=0$, and each $C_m(\omega_r)$ and $C_m(\omega_r)/(\omega_r-\lambda)$ is square-integrable. We shall prove that, if $C_m(\omega_r)/(\omega_r+\mu)$ is a causal transform, then $C_m(\omega_r)/(\omega_r-\lambda)=C_{m+1}(\omega_r)/(\omega_r+\mu)$ is a causal transform. For $C_m(\omega_r)/(\omega_r+\mu)$ is a causal transform if and only if it satisfies almost everywhere the Hilbert transform relations (2.5), that is

$$\frac{C_m(\omega_r)}{\omega_r+\mu}=\frac{P}{\pi i}\int_{-\infty}^{+\infty}\frac{C_m(\nu)d\nu}{(\nu+\mu)(\nu-\omega_r)}. \quad (5.4)$$

Since

$$\frac{\omega_r+\mu}{(\nu+\mu)(\nu-\omega_r)}=\frac{(\omega_r-\lambda)}{(\nu-\lambda)(\nu-\omega_r)}+\frac{(\mu+\lambda)}{(\nu+\mu)(\nu-\omega_r)},$$

we can obtain directly from Eq. (5.4):

$$C_m(\omega_r)=(\omega_r-\lambda)\frac{P}{\pi i}\int_{-\infty}^{+\infty}\frac{C_m(\nu)d\nu}{(\nu-\lambda)(\nu-\omega_r)} +\frac{(\mu+\lambda)P}{\pi i}\int_{-\infty}^{+\infty}\frac{C_m(\nu)d\nu}{(\nu+\mu)(\nu-\lambda)}. \quad (5.5)$$

Comparison of the last term in Eq. (5.5) with Eq. (5.4) shows that this term is just $C_m(\lambda)$. [Since Eq. (5.4) can fail to hold on a set of points of measure zero, it is possible *a priori* that Eq. (5.4) fails at $\omega_r=\lambda$. However, in this case we can replace λ by choosing a new point from within the neighborhood in which the jth derivative of $A(\omega_r)$ exists; since this neighborhood has a positive measure, it is always possible to choose λ so that the Hilbert transforms that we need do exist at $\omega_r=\lambda$, and we shall henceforth assume that this has been done.] Since $C_m(\lambda)=0$, the second integral in Eq. (5.5) vanishes and Eq. (5.5) is then recognized to be the Hilbert transform formula for $C_m(\omega_r)/(\omega_r-\lambda)$; since it satisfies the Hilbert transform formula connecting its real and imaginary parts, $C_m(\omega_r)/(\omega_r-\lambda)=C_{m+1}(\omega_r)/(\omega_r+\mu)$ is a causal transform.

We can now apply repeatedly the proof of the last paragraph, beginning with the fact that $C_1(\omega_r)/(\omega_r+\mu)=B(\omega_r)/(\omega_r+\mu)^j$ is a causal transform, to obtain that $C_j(\omega_r)/(\omega_r-\lambda)=B(\omega_r)/(\omega_r-\lambda)^j$ is a causal transform and therefore satisfies

$$B(\omega_r)=\frac{(\omega_r-\lambda)^j}{\pi i}P\int_{-\infty}^{+\infty}\frac{B(\nu)d\nu}{(\nu-\lambda)^j(\nu-\omega_r)}. \quad (5.6)$$

Therefore, the real and imaginary parts of $A(\omega_r)$ determine each other through the dispersion relation (5.6) except for the j-arbitrary complex numbers in (5.2) which give the value of A and its first $(j-1)$ derivatives at $\omega_r=\lambda$.

If we are given only $A_i(\omega_r)$, we can find $B_i(\omega_r)$ in (5.2) and can then always find $B_r(\omega_r)$ from (5.6), and thus determine $A_r(\omega_r)$ except for the j-arbitrary constants. Thus (5.6) is a useful dispersion relation in the usual sense; usually λ is conveniently chosen as zero although it may happen that the necessary derivatives diverge at the origin, in which case another point must be selected for λ.

(b) Functions Diverging at Finite Point

In some physical theories the generalized scattering amplitude may diverge at a finite frequency. For example, the impedance of an ideal capacitance diverges at the origin as the inverse of the frequency.[30] Suppose $A(\omega_r)$ diverges at $\omega_r=a$, but is such that $(\omega_r-a)^qA(\omega_r)$ remains finite. Then, if this is the only point of divergence, a dispersion relation for $(\omega_r-a)^qA(\omega_r)$ such as in Eq. (5.6) can be derived. Similarly any finite number of divergences, each no worse than a finite power, can be treated. The factor $(\omega_r-a)^q$ increases the power of divergence near infinite frequency by q and thus adds q arbitrary constants to the dispersion relation; these constants can in turn be regarded as determining the q

[30] If the nonzero conductance that is inevitably in parallel with any actual condenser is included, the pole in the impedance functions lies in the lower half of the complex plane and the impedance is a causal factor.

arbitrary coefficients in the expansion of $A(\omega_r)$ about $\omega_r = a$.

In addition to the integrability or boundedness condition, one might consider altering other assumptions of the proof of Sec. 3. The linearity and time-invariance of the connection are essential conditions in this discussion of the equivalence of causality and the dispersion relations, and we are doubtful if these restrictions can be relaxed. However, the above proof also assumed that the input was freely variable and could be chosen to be any function of integrable square. In many interesting physical problems, the input is constrained; for example, a certain range of frequencies may be excluded as in the Klein-Gordon or Dirac equations. In such a case, the restrictions placed on the generalized scattering amplitude by strict causality are less stringent; we hope to return in a later paper to a discussion of the consequences of strict causality for such a constrained input system.

6. DISCUSSION

In this paper we have investigated analytic conditions and dispersion relations which are necessary and sufficient for a hypothesis of strict causality. It is an interesting open question whether or not strict causality is a valid physical hypothesis. As shown by Dirac[31] and by Wheeler and Feynman,[32] strict causality is not satisfied in the simpler forms of classical electrodynamics. However, this classical result is irrelevant to the question of whether such acausal effects actually exist, for the precursor effects of the classical theory are due to high-energy phenomena and to radiative reaction which we know are not correctly described by classical theory. Stueckelberg[33] and Fierz[34] has shown shown that the present form of quantum electrodynamics is strictly causal if and only if we follow the rule for the time ordering of all factors in the S matrix. Gell-Mann, Goldberger, and Thirring[35] and many others have recently shown how to obtain dispersion relations for quantum field theories as consequences of strict causality. Whether in a future better theory strict causality would be exactly valid, or could even be accurately defined, is an unsettled problem. However, we adopt the attitude that we should employ the dispersion relation, at least until it is shown to be faulty, and we shall proceed to apply it to special problems in following papers.

Even if strict causality should prove to be invalid or undefineable in future theories, it is to be expected that macroscopic acausality involving propagation of appreciable energy over large space-like intervals would still be forbidden. We have investigated a conceivable alternative to strict causality, a weaker "principle of limiting velocity": "No energy can be transmitted to infinite distance at a velocity greater than c." This requirement can be shown to be satisfied if and only if the group velocity does not exceed c for any frequency at which the absorption coefficient is zero; it provides no connection between the refractive index and absorption at other frequencies and hence is an empty restriction in cases of physical interest.

7. ACKNOWLEDGMENTS

The author wishes to express his gratitude to many colleagues, especially at Princeton University and the University of Maryland, for enlightening discussions of this work. He is especially thankful to Professor John A. Wheeler for suggesting this problem and for direction, collaboration, and encouragement.

[31] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938).
[32] J. A. Wheeler and R. P. Feynman, Revs. Modern Phys. **17**, 157 (1945).
[33] E. C. G. Stueckelberg, Helv. Phys. Acta **19**, 241 (1946).
[34] M. Fierz, Helv. Phys. Acta **23**, 731 (1950).
[35] Gell-Mann, Goldberger, and Thirring, Phys. Rev. **95**, 1612 (1054); M. L. Goldberger, Phys. Rev. **99**, 979 (1955). Further references and discussion will be included in later papers.