Optical theorem and beyond*

Roger G. Newton
Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 1 December 1975)

The history of the so-called *optical theorem* in scattering theory is traced from its beginning over 100 years ago to recent applications of its generalizations.

The curious history of the optical theorem, its generalizations and their consequences, are beautiful examples of how a scientific idea may occur in various contexts independently, be forgotten, and rediscovered. This paper will trace that history from its beginnings more than a century ago, to some relatively recent ideas concerning the use to which the generalizations may be put with advantage.

The essential thought behind the optical theorem was originally the idea that the optical index of refraction of a medium depends upon its absorption coefficient. The resulting influence of the latter on the dispersion of light propagating through the material was discussed first by Sellmeier[1] in 1871 and, in a more quantitative manner, independently by Strutt[2] (better known as Lord Rayleigh) during the same year. Rayleigh used the idea in his famous studies of the color and polarization of the sky,[2] both in his first paper on this subject in 1871 and when he returned to it 28 years later, in 1899.[3] The context in which it occurs in these works is that of a complex index of refraction, the real part of which is the geometrical index of refraction proper, and the imaginary part, the absorption coefficient. Rayleigh realizes that the index of refraction n is linearly related to what is now called the forward scattering amplitude $A(0)$ of the constituents of the medium,

$$n=1+2\pi N k^{-2} A(0), \tag{1}$$

where N is the number of particles per unit volume and $k=\omega/c=l/\lambda\!\!\!^{-}$ is the (reduced) wave number of the light, and that the absorption coefficient is proportional to what later became known as the extinction cross section, σ_{ex}, of the constituents. He then makes explicit reference to the conservation of energy to equate the latter, that is, the blockage of light, to the total amount of light scattered in all directions, that is in modern language, the total cross section, σ_{tot}. The result is what we now write in the form

$$\sigma_{\mathrm{tot}}=\sigma_{\mathrm{ex}}=4\pi k^{-1}\,\mathrm{Im}A(0). \tag{2}$$

Of course, the statement of the theorem is not quite that explicit even in Rayleigh's papers, but the central idea is certainly there.

Note that in the propagation problem for a cloud of many randomly situated particles Eq. (2) is the reason why extinction cross sections are additive,

$$\sigma_{\mathrm{ex}}=\sum_j \sigma_{j,\mathrm{ex}}, \tag{3}$$

just as, for quite different reasons, the total scattering cross sections are

$$\sigma_{\mathrm{scatt}}=\sum_j \sigma_{j,\mathrm{scatt}}. \tag{4}$$

During the next 10 years, after 1899, the theory of the scattering of light by particles was further developed and the optical theorem is stated rather explicitly in 1908 in the famous paper[4] by Mie on the scattering of light by spherical particles, and again by Gans and Happel,[5] citing Mie, in 1909.

Historically the next context in which Eq. (2) is used is that of the theory of dispersion of light by Kronig[6] and Kramers[7] in 1926/27. Here the idea of *causal* propagation of radiation in a medium, with no possible effects before a signal has arrived, leads to a Hilbert-transform relation between the real and imaginary parts of the index of refraction as a function of the frequency ω. If the refractive index is then expressed as in (1) in terms of the forward scattering amplitude, and (2) is used, one gets the dispersion relation

$$\mathrm{Re}\,A(0,\omega)=\frac{1}{2\pi^2 c}\,P\int_0^\infty d\omega'\,\frac{\omega'^2\sigma_{\mathrm{tot}}(\omega')}{\omega'^2-\omega^2}, \tag{5}$$

in which P indicates Cauchy's principal value of the integral. Causality, in the sense of "no effect before its cause," together with Eq. (2) therefore leads to the possibility of inferring the dispersion, that is, the frequency dependence of the optical index of refraction, from the absorption, and similarly, the phase of the forward scattering amplitude from the total cross section.

The first appearance of Eq. (2) in the quantum-theoretical description of the scattering of particles occurs completely independently in Feenberg's Ph.D. thesis, and in his paper.[8] His method of derivation used a phase-shift expansion (which was then called a Faxen–Holtzmark expansion) of the scattering amplitude and it is restricted to elastic scattering by a spherically symmetric potential. Neither Feenberg's Ph.D. examiners nor, so far as we can tell, the readers of the *Physical Review* found the relation (2) particularly interesting at the time.

The S matrix as a tool in the theory of scattering was invented[9] by Wheeler in 1937, and it was shown by him to be unitary. He did not, however, draw (2) as a conclusion from that unitarity. Two years later Bohr, Peierls, and Placzek[10] stated the equation without explicit derivation, and for the next twenty years and longer, it continued to be referred to by many as the Bohr–Peierls–Placzek relation.

In the meantime the second world war had broken out and communications between German and Western scientists were disrupted. In 1943 Heisenberg[11] independently invented the S matrix as a useful concept for the description of scattering and he proved its unitarity. As an observable consequence of the rather abstract unitarity property of the scattering matrix, he derived, for the first time, the *generalized optical theorem*. If we call $\mathbf{k}=k\hat{\mathbf{k}}$ the momentum of the incident particles, $\mathbf{k}'=k\hat{\mathbf{k}}'$ that of the final ones, and $A_k(\hat{\mathbf{k}}',\hat{\mathbf{k}})$ the elastic scattering amplitude, then the generalized optical theorem (as it is now called, but was not by Heisenberg) is the equation

$$\operatorname{Im} A_k(\hat{\mathbf{k}}', \hat{\mathbf{k}}) = (k/4\pi) \int d\hat{\mathbf{k}}'' A_k(\hat{\mathbf{k}}', \hat{\mathbf{k}}'') A_k^*(\hat{\mathbf{k}}'', \hat{\mathbf{k}}). \quad (6)$$

It is valid in this form provided that the interaction is invariant under time reversal. When $\hat{\mathbf{k}}' \to \hat{\mathbf{k}}$ it goes over into (2) as a special case. In a footnote Heisenberg refers to the paper[10] by Bohr, Peierls, and Placzek in *Nature* as having been mentioned to him by Giancarlo Wick. However, he had no access to it, as the war had been going since 1939.

The years after the war saw a considerable expansion in work on scattering theory and as a result, of the area of applicability of the optical theorem. In 1949, Wick[12] extends it to particles with spin, in 1950 Lax[13] includes inelastic processes, and in 1954 Schiff[14] extends that area of use further. A 1953 paper[15] by Glauber and Schomaker on electron diffraction contains the generalized optical theorem without use of the S matrix and with no references to earlier work.

It is interesting to examine the textbooks on quantum mechanics for the first appearance of the theorem. None of the prewar books contain it, nor does Pauli's 1946 book,[16] or even the second edition[17] of Mott and Massey (1949). The first edition of Schiff's book,[18] which was published in 1949, does not mention it either. However, it is noteworthy that the books by Messiah[19] (1963), Gottfried[20] (1966), and Sakurai[21] (1967), still call Eq. (2) the Bohr–Peierls–Placzek relation. Others who mention it during this period usually refer to it as "a well-known relation." Rohrlich and Gluckstern[22] in 1952 refer to it as "a well-known theorem of optics," and approximately from then on it appears to begin to be called the *optical theorem* by more and more physicists working in quantum particle scattering theory. Bethe and de Hoffman[23] refer to it as the "so-called 'Optical Theorem' " in 1955, the earliest reference in print by that name that I have run across. By 1961 it is very often referred to by that name.[24–26] References to Eugene Feenberg's paper still do not appear to exist at all in the literature of particle physics.

Meanwhile, back in electromagnetic theory, an amusing development took place. On the one hand, particle physicists working in that area or in general diffraction theory, were aware of (2). For example, Levine and Schwinger[27] derive it (without use of the S matrix) and say that it is generally valid for scattering phenomena. They give no reference and it may be assumed that they regarded it as well known. On the other hand, van de Hulst rediscovers[28] Eq. (2) in 1949 for the scattering of light, totally unaware that it was already well known both in optics and in quantum scattering theory. There is a paper[29] by Jones, of 1955, on what he calls "van de Hulst's theorem." However, Jones also cites Feenberg and Levine and Schwinger. By 1957 van de Hulst knew of the earlier references and in his book[30] on the scattering of light, he cites, among other, Feenberg's 1932 paper. The optics treatise[31] by Born and Wolf nevertheless credit van de Hulst with the theorem, while also mentioning the papers by Jones, Feenberg, and Lax.

Van de Hulst's derivation of the optical theorem is, from a physical point of view, very nice and intuitive and well worth reproducing. It provides somewhat more physical insight than a formal derivation by means of the unitarity of the S matrix (or the reality of the phase shifts, which is the same thing).

For simplicity, let us take a scalar wave with scattering boundary conditions. At large distances,

$$r \gg \lambda, \quad (7)$$

it goes as[32]

$$\psi(r) \simeq \exp(ikz) + \exp(ikr) r^{-1} A(\theta) + \cdots \quad (8)$$

if the incident wave travels along the positive z axis and θ is the angle between $\mathbf{r}$ and the z axis. Let us look in the almost-forward direction, where

$$\theta \ll 1 \quad (9)$$

and the plane-wave term is indistinguishable from the spherical wave. There

$$r = (x^2 + y^2 + z^2)^{1/2} = z + (x^2 + y^2)/2z + \cdots,$$

and hence

$$|\psi|^2 \simeq 1 + 2z^{-1} \operatorname{Re}\{A(0) \exp[ik(x^2 + y^2)/2z]\} + \cdots.$$

The total energy collected on a screen of radius R is obtained by integrating the above expression over x and y, up to $x^2 + y^2 = R^2$. If we assume that the screen is so large that

$$kR^2/z \gg 2\pi \quad (10)$$

while still

$$R/z \ll 1 \quad (11)$$

so as to preserve (9), then the integral can be easily done with negligible error as a complete Gaussian and we obtain

$$\int ds\, |\psi|^2 = \pi R^2 - (4\pi/k) \operatorname{Im} A(0). \quad (12)$$

This means that the energy captured on the screen is diminished from what it would be in the absence of a scatterer by the amount

$$\sigma_{\mathrm{ex}} = 4\pi k^{-1} \operatorname{Im} A(0). \quad (13)$$

Energy conservation now dictates that this must equal the total amount either scattered or absorbed:

$$\sigma_{\mathrm{tot}} = \sigma_{\mathrm{scatt}} + \sigma_{\mathrm{abs}} = \sigma_{\mathrm{ex}} = 4\pi k^{-1} \operatorname{Im} A(0). \quad (14)$$

Note the need for both inequalities (10) and (11), i.e.,

$$(\lambda z)^{1/2} \ll R \ll z,$$

which are compatible because of (7). The physical meaning of (10) is that, in order for (13) to be valid, the screen must be so large that it captures all the main diffraction peaks. If it collects just the central diffraction spot, then (12) does not hold.

Historical interest apart, the more important question is what is Eq. (2) good for. The first answer to this is, simply as a calculational device. By the middle of the 1950's it became quite common for theorists to calculate total cross sections by means of (2) rather than direct integration. If that were its only use, of course, its value would be very limited.

The utility of (2) in the dispersion relation (5) raises it to a much higher level of importance. The use of (2) in that context is the main reason why the "forward dispersion relation" (5) has more experimental content than analogous dispersion relations that have been derived for nonforward scattering amplitudes. This leads to the question to what extent (2) is experimentally verified. The fact that dispersion relations have not been found violated in a variety of contexts is probably the best measure of its agreement with experience.

The more general question raised by any attempt to verify the optical theorem is: To what extent does the phase of the scattering amplitude have any experimental meaning at all? An answer is provided by the fact that, in principle, this phase can be measured. The most direct method for such a measurement is that of the effect of intensity-fluctuation correlations between two independent beams scattered by the same target, as discovered by Hanbury-Brown and Twiss.[33] For particle beams, these fluctuation correlations were discussed by Goldberger, Lewis, and Watson.[34,35] The corresponding experiments are very difficult and, at least in the particle context, have never been done. Although this method makes the phase of the scattering amplitude, in principle, experimentally meaningful, it leads to its measurement only to within an additive angle-independent constant. In other words, it allows the measurement only of the difference between phases at different angles.

There is a good reason for asserting that a constant (angle-independent) addition to the phase of a scattering amplitude is experimentally not detectable. The very definition of the amplitude, after all, makes it appear in the wave function multiplied by $\exp(ikr)$, as in (8). Hence to give experimental meaning to a constant phase addition requires measuring the distance r from the scattering center with an accuracy equal to a fraction of the wavelength $\lambda = 2\pi/k$. This is never feasible. As a consequence, the optical theorem itself cannot be expected to be experimentally verified. In fact, it may be regarded as *defining* the phase from an experimental point of view. This remark does not, of course, apply to the generalized optical theorem (6).

Let us note as an application that the optical theorem all by itself implies that, if the total cross section is large compared to the wavelength, then there must be a large and sharp forward diffraction peak. Equation (2) implies that

$$\sigma(0)/\bar{\sigma}_{\rm el} \geqslant \pi\sigma_{\rm tot}/\lambda^2 \qquad (15)$$

and

$$\Theta \lesssim 2\lambda(\pi\sigma_{\rm tot})^{-1/2} \qquad (16)$$

if $\bar{\sigma}_{\rm el}$ is the average elastic differential cross section, $\sigma(0)$ is the forward differential cross section, and Θ is the width of the diffraction peak.

Another important use of the general equation (6) was initiated first by Puzikov, Ryndin, and Smorodinskii[36] in 1957, with a more detailed analysis by Klepikov[37] in 1964, and then independently by this author[38] in 1968 and by Martin[39] in 1969. The idea is to consider (6) as a (nonlinear) integral equation for the phase of the scattering amplitude if the differential cross section is experimentally given as a function of the scattering angle.

In order to explain the principle, suppose the scatterer is spherically symmetric and the scattered particles have zero spin. Then the scattering amplitude depends on the scattering angle θ only, or on $x = \cos\theta$. Write

$$A(\theta) = k^{-1}G(x)\exp[i\varphi(x)],$$

where $G \geqslant 0$. Then (6) reduces to the form

$$\sin\varphi(x) = \int_I dy\, dz\, H(x, y, z)\cos[\varphi(y) - \varphi(z)], \qquad (17)$$

where

$$H(x, y, z) = \frac{G(y)G(z)}{2\pi G(z)(1 - x^2 - y^2 - z^2 + 2xyz)^{1/2}},$$

and the integration extends over the interior I of the ellipse

$$1 - x^2 - y^2 - z^2 + 2xyz = 0.$$

Since the differential cross section is related to A by

$$\frac{d\sigma}{d\Omega} = |A|^2 = G^2,$$

G may be considered to be experimentally given.

The questions raised by (17), regarded as an integral equation for $\varphi(x)$ [40] are the following: (i) Does a solution always exists? (ii) Is the solution unique? (iii) How can (17) actually be solved?

As for the first question, it is not as physically uninteresting as physicists usually consider mathematical existence questions. It is equivalent to asking whether the unitarity of the S matrix imposes any restrictions on the possible angle dependence of a differential cross section. The optical theorem, of course, does impose such a restriction, namely

$$\left[\frac{d\sigma}{d\Omega}(\theta = 0)\right]^{1/2} \geqslant \frac{k}{4\pi}\sigma_{\rm tot}.$$

But is this the only restriction? As for uniqueness, there is an obvious ambiguity: If $\varphi(x)$ solves (17), then so does $\pi - \varphi(x)$. In other words, the sign of the real part of A cannot be determined. One may refer to uniqueness *apart from this,* or to uniqueness of $\sin\varphi(x)$, as *essential* uniqueness.

What has been shown is that whenever

$$Q(x) = \int_I dy\, dz\, H(x, y, z) < 1 \qquad (18)$$

and $G(x)$ is continuous, then (17) always has a solution.[38,39,41,42] If

$$Q(x) < 0.79, \qquad (19)$$

then the solution is essentially unique and can be constructed by iteration.

These results have all been generalized[42,43] and the spherical symmetry of the scatterer is not important. For $Q > 1$ no answers are known, but explicit examples of nonuniqueness have been constructed.[44–48] There now exists an extensive literature[49] on the use of (6) for the construction of the phase of the scattering amplitude from a knowledge of the differential cross section. Unfortunately, the restriction to (18) or (19), or to analogous limitations for other norms, makes all the known results on existence, uniqueness, and constructibility of very little experimental use. Most of the interesting cross sections violate these inequalities.

We may, finally, raise the question of the generalization of (6) to collisions in which, initially and/or finally, more than two particles participate. Since (6) is equivalent to the unitarity of the S matrix, and the latter is known to be unitary in general collision problems, (6) certainly has such a generalization. In fact, Heisenberg[11] already stated it in such a general context. One may, however, look at the partial-wave expansion of elastic amplitudes as the most useful method of satisfying unitarity (by making the phase shifts real) and hence (6). The question of extending (6) then may be asked in the form: Is there a generalization of the phase-shift expansion to more than two particles?

Since the three (or more) body problem is very complicated, I can give here only the briefest answer. When the "ionization" channel is open, i.e., when there is enough energy for three particles to move freely, the question is further complicated by the fact that the amplitude blows up at those final momenta for which it is energetically possible for two pairs of particles to collide in succession. It has nevertheless been proved[50] that if the S matrix for three particles is factored into three pair-wise two-particle S matrices times a remainder, as in

$$S = S_{12} S_{23} S_{31} S',$$

then S' has a discrete spectrum. Hence one may define a denumerable set of (real) eigenphase shifts of the unitary S'. An expansion of S in terms of these eigenphase shifts was shown to converge in a well-defined sense. However, the physical meaning of the eigenfunctions and their construction, are still quite obscure.

This takes us to the frontier of our present knowledge of the main ramifications of the optical theorem and its generalizations. Its tortuous history shows that the progress of scientific ideas is not always as straight as our textbooks would have us think.

ACKNOWLEDGMENT

I thank E. Feenberg for informative conversations and correspondence.

*Supported in part by the National Science Foundation. This paper was presented by invitation to the symposium Concepts and Methods in Microscopic Physics, in honor of Eugene Feenberg, 20–21 March 1974, at Washington University, St. Louis, MO.

[1] E. Sellmeier, Ann. Phys. (Leipz.) **143**, 272 (1871).
[2] J. W. Strutt, Philos. Mag. **41,** 107, (1871); see also Lord Rayleigh, *Scientific Papers* (Dover, New York, 1964), Vol. 1, p. 87.
[3] J. W. Strutt, Philos. Mag. **47,** 375 (1899); see also Lord Rayleigh, Ref. 2, Vol. 4, p. 397.
[4] M. Mie, Ann. Phys. (Leipz.) **25**, 377 (1908), Eq. (98).
[5] R. Gans and H. Happel, Ann. Phys. (Leipz.) **29**, 294 (1909).
[6] R. Kronig, J. Am. Opt. Soc. **12**, 547 (1926).
[7] H. A. Kramers, Atti Congr. Int. Fis. Como **2,** 545 (1927); see also *Collected Scientific Papers* (North-Holland, Amsterdam, 1956), p. 333.
[8] E. Feenberg, Phys. Rev. **40**, 40 (1932). This paper is also remarkable for the fact that it was published on 1 April 1932, with a receipt date stated as 8 September 1932.
[9] J. A. Wheeler, Phys. Rev. **52**, 1107 (1937).
[10] N. Bohr, R. E. Peierls, and G. Placzek, Nature **144**, 200 (1939).
[11] W. Heisenberg, Z. Phys. **120**, 513 (1943); **120,** 673 (1943).
[12] G. Wick, Phys. Rev. **75**, 1459 (1949).
[13] M. Lax, Phys. Rev. **78**, 306 (1950).
[14] L. I. Schiff, Prog. Theor. Phys. **11**, 288 (1954).
[15] R. Glauber and V. Schomaker, Phys. Rev. **89**, 667 (1953).
[16] W. Pauli, *Meson Theory of Nuclear Forces* (Interscience, New York, 1946).
[17] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Clarendon, Oxford, 1949).
[18] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1949).
[19] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1961).
[20] K. Gottfried, *Quantum Mechanics* (Benjamin, New York, 1966).
[21] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, MA, 1967).
[22] F. Rohrlich and R. L. Gluckstern, Phys. Rev. **86**, 1 (1952).
[23] H. A. Bethe and F. de Hoffmann, *Mesons and Fields* (Row, Peterson, Evanston, IL, 1955), Vol. II.
[24] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961).
[25] J. D. Jackson, in *Disperson Relations*, edited by G. R. Screaton, Oliver and Boyd, Edinburgh, 1961).
[26] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson, Evanston, IL, 1961).
[27] H. Levine and J. Schwinger, Phys. Rev. **74**, 958 (1948).
[28] H. C. van de Hulst, Physica **15**, 740 (1949).
[29] D. S. Jones, Philos. Mag. **46**, 957 (1955).
[30] H. C. van de Hulst, *Light Scattering by Small Particles* (Wiley, London, 1957); see particularly p. 39.
[31] M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1959).
[32] For notational simplicity we assume here that the scatterer has rotational symmetry and hence A depends on θ only. However, the argument is quite independent of this assumption.
[33] R. Hanbury-Brown and R. Q. Twiss, Philos. Mag. **45**, 663 (1954); Nature **177**, 27 (1956); **177**, 178 (1956); **177**, 1447 (1956); Proc. R. Soc. Lond. A **242**, 300 (1957); **243**, 291 (1957).
[34] M. L. Goldberger, H. W. Lewis, and K. M. Watson, Phys. Rev. **132**, 2764 (1963).
[35] M. L. Goldberger and K. M. Watson, Phys. Rev. **134,** B919 (1964); **137,** B1396 (1965); **140,** B500 (1965).
[36] L. Puzikov. R. Ryndin, and Ia. Smorodinskii, Zh. Eksp. Teor. Fiz. **32,** 592 (1957) [Sov. Phys.–JETP **5,** 489 (1957)].
[37] N. P. Klepikov. Zh. Eksp. Teor. Fiz. **47,** 757 (1964) [Sov. Phys.–JETP **20,** 505 (1965)].
[38] R. G. Newton, J. Math. Phys. **9**, 2050 (1968); **16**, 1358 (1975).
[39] A. Martin, Nuovo Cimento A **59**, 131 (1969); **26**, 94 (1975).
[40] On the domain $-1 \leqslant x \leqslant 1$ and with the range, say, $-\pi \leqslant \varphi \leqslant \pi$.
[41] D. Atkinson, P. W. Johnson, and R. L. Warnock, Commun. Math. Phys. **28**, 133 (1972).
[42] M. Tortorella, J. Math. Phys. **15**, 745 (1974); **15**, 1153 (1974).
[43] M. Tortorella, J. Math. Phys. **16,** 823 (1975).
[44] J. H. Crichton, Nuovo Cimento A **45**, 256 (1966).
[45] A. Gersten, Nucl. Phys. B **12**, 537 (1969).
[46] D. Atkinson, P. W. Johnson, N. Mehta, and M. de Roo, Nucl. Phys. B **55**, 125 (1973).
[47] F. A. Behrends and S. N. M. Ruysenaars, Nucl. Phys. B **56,** 525 (1973).
[48] H. Cornille and J. M. Drouffe, Nuovo Cimento A **20**, 401 (1974).
[49] For references, see M. de Roo, Ph.D. thesis (Groningen, 1974) (unpublished); D. Atkinson, M. Kaekebeke, and M. de Roo, J. Math. Phys. **16**, 685 (1975).
[50] R. G. Newton, J. Math. Phys. **15**, 338 (1974).