Anthony Duncan

Statement

and

Readings

The Analyticity Principle in 20th Century Physics: From the Kramers Dispersion Formula to Dual Amplitudes and Strings

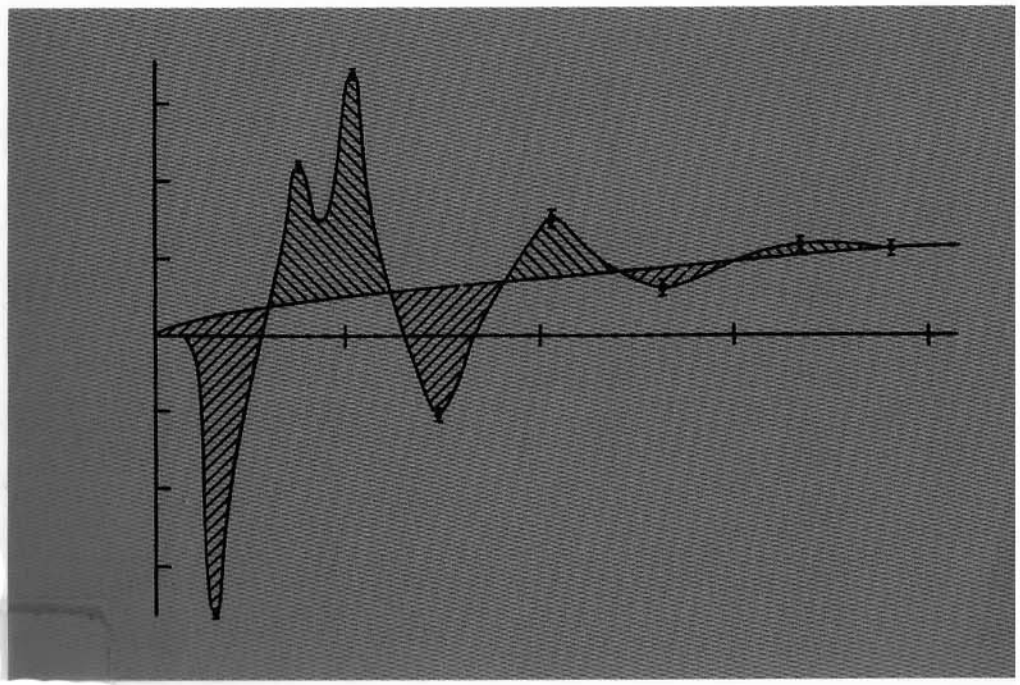
Anthony Duncan
Department of Physics and Astronomy, University of Pittsburgh
Pittsburgh, PA, USA

Abstract

Two formulas of great historical resonance in twentieth century theoretical physics- the Kramers dispersion formula of 1924, the immediate precursor of Heisenberg's matrix mechanics, and the Veneziano dual amplitude of 1968, the direct ancestor of the Nambu–Goto–Nielsen string- are continuously connected by a conceptual strand of enormous importance. The requirement of analyticity of quantum-mechanical amplitudes, intimately related to the demands of causality, as a constraint on fundamental physics, is perhaps second only to the emergence of the local gauge symmetry principle in guiding the development of modern microphysical theories from relativistic quantum field theories to string theory. In this talk, some of the important signposts along the path connecting these two seminal formulas will be discussed.

Theory construction and selection in modern physics

THE S MATRIX



James T. Cushing

Theory construction and selection in modern physics

THE *S* MATRIX

James T. Cushing
Professor of Physics
University of Notre Dame

CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York Port Chester
Melbourne Sydney

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521381819

First published 1990
This digitally printed first paperback version 2005

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Cushing, James T., 1937–
Theory construction and selection in modern physics: the *S* Matrix
James T. Cushing
p. cm.
Includes bibliographical references
ISBN 0-521-38181-9
1. *S*-matrix theory. 2. Physics–philosophy. 3. Physics–
Methodology. I. Title
QC174.35.S2C87 1990
530. 1'22–dc20

ISBN-13 978-0-521-38181-9 hardback
ISBN-10 0-521-38181-9 hardback

ISBN-13 978-0-521-01730-5 paperback
ISBN-10 0-521-01730-0 paperback

1 Introduction and background

In recent years there has been a move to 'naturalize' the philosophy of science. This has meant basing work in the philosophy of science upon the actual historical record of real scientific practice and stressing (in varying degrees) the use of the methods of science in studying the scientific enterprise. This attention to actual scientific practice has been supported by traditional realists, an example being Ernan McMullin (1984) who early on (1976a) argued for a central role for the history of science in the philosophy of science; by philosophers of various anti-theoretical bents, such as Nancy Cartwright (1983) and Ian Hacking (1983); and by empiricists, like Bas van Fraassen (1980, 1985). In the early 1960s, it was attention to the historical record that led Thomas Kuhn (1970), in his *The Structure of Scientific Revolutions*, to stress the importance of social factors in the practice of real science. The spirit of the present work is that careful and detailed study must be made of the actual development of science *before* conclusions are drawn about the appropriateness of any particular methodology of science. Ours is mainly a story about theory, but not one uncoupled from its relation to experiment.

We claim that many of the philosophically interesting questions in science, especially in regard to possible changes in the methodology and goals of science, can be seen and appreciated only upon examination of the technical details of that practice. So, in this chapter we discuss some motivations for studying *current* scientific practice and then set the problem background out of which the *S*-matrix program arose. This sketch is somewhat ahistorical since we mention some field theory developments that occurred *after* 1943 (when Heisenberg introduced his *S*-matrix program). The formalism of classical particle mechanics and of wave phenomena have been, in this century, successively

reinterpreted to yield (nonrelativistic) quantum mechanics, relativistic quantum field theory for electromagnetic phenomena and, finally, relativistic quantum field theory for strong (or 'nuclear') phenomena. The prototypical experimental arrangement (of either the real or *gedanken* type) for studying the fundamental interactions of nature has been a scattering process in which a target (such as a nucleus) is probed by a projectile (such as an electron or a proton).

However, the development of these quantum field theories has not always been a smooth one. In particular, in the 1930s and early 1940s, the quantum field theory program had run into considerable technical and experimental difficulties (Cassidy, 1981; Galison, 1983a). Mathematical inconsistencies, most notably divergences or infinities produced by calculations with the formalism, occurred for quantum electrodynamics (QED) at short distances (or, equivalently, at high energies) of the order of the size of the electron. Similar problems plagued Fermi's theory of β decay and Yukawa's meson theory of nuclear forces. Here we have examples of three of the four basic forces in nature: the electromagnetic, which is responsible for the atomic phenomena producing those features of the world we commonly encounter; the weak, which accounts for the spontaneous decay or conversion of a free neutron into a proton; and the strong, which predominates at very short distances for nuclear processes. The last is the gravitational force, which plays no essential role in our story here. During the same period of confusion in the arena of theory, experimental results from cosmic ray data seemed to contradict expectations based on quantum field theory (QFT). It appeared as though cosmic ray showers, or 'explosive' events, occurred (in contrast to the cascades built up from many essentially pairwise events, which could be readily accounted for by Dirac's hole theory). Heisenberg took the existence of these multiple processes to signal a breakdown of conventional quantum field theory and to require the introduction of a fundamental length into the theory. The field theory situation was further complicated by the confusion (in the 1930s) caused when the mesons observed in cosmic-ray interactions were at first identified with Yukawa's nuclear-interaction π meson (pion), before they were finally identified as μ mesons (muons), which are essentially 'heavy' electrons. These difficulties encountered by the quantum field theory program provided a significant part of the motivation for Heisenberg's proposing his *S*-matrix theory (SMT).

By the late 1940s, a mathematical technique (renormalization) had been formulated which allowed one to circumvent the divergences of QED and to make accurate predictions confirmed by experiment. This

is the first cycle of the oscillation of theory between QFT and SMT. Others occurred when QFT was stymied by the strong interactions, from which it subsequently recovered with gauge field theories. This back and forth between formalisms, with their corresponding paradigms, is an important feature of the episodes we present. It will be especially relevant for our evaluation of methodology in science (in Chapter 10).

1.1 Internal history of recent science

This work is largely, but not exclusively, an *internal* history of an extended episode in modern high-energy physics. That is, the published physics literature is a major source for the technical developments we present. Nevertheless, interviews and correspondence are also employed. The primary interest in and motivation for doing the research necessary for this case study are philosophical. Some obvious questions, that arise about the value and wisdom of doing an internal history of a *current* (and hence not completed) episode in a *highly technical* (or specialized) subject area, must be addressed.

Schweber (1984, p. 41), in his history of the early developments of quantum field theory, has stated one of the problems of internal history as follows:

> [I]nternal history faces the problem common to all good history: how to avoid the pitfalls of Whiggish history, that is, the writing of history with the final, culminating event or set of events in focus, with all prior events selected and polarized so as to lead to that climax.

So, while the philosophy of science must be based on history (i.e., events as they actually occurred), it can be important not to focus exclusively on the form and content of 'successful' scientific theories alone. The arguments and contingent events that inform the course of development and selection of theories are essential. That is, how things might have gone a very different way at certain crucial junctures and why they did not may be as important as the reasons for the 'right' choices that science has made. The present case study focuses on a 'failed' program that has never been proven to be incorrect. It is evident that one cannot explain its rejection just in terms of falsification. Perhaps there is something to be learned from the history of such a dead-end theory. The relevance of such ('sociological') factors as the previous interests and expertise of the participants becomes apparent enough.

Historians may extol the virtues (in fact, necessity) of doing the history of an episode only long after the clamor of the day has settled. They can argue that once time and events have produced a stable picture of the past, one feels some confidence that one may be able to find 'the objective truth' in those long-dead events (or 'corpse') (Burckhardt, 1963, pp. 74–76).[1] There is an old tradition of this attitude in the history of science. Thus, in Whewell's *History of the Inductive Sciences* (1857, Vol. II, p. 434) we find: 'It is only at an interval of time after such events have taken place that their history and character can be fully understood, so as to suggest lessons in the Philosophy of Science.' Even if one accepts that thesis (and it *can* be debated), he can still feel that something (perhaps important, Burckhardt and Whewell to the contrary notwithstanding) has slipped away. The detailed dynamics of the events and the motivations of the protagonists have been lost (in large measure, at least) behind the veil of time. Now if one believes that all final scientific positions are reached ultimately through rational judgments *alone*, then there is probably even virtue in waiting until the flotsam has been swept away by time to leave a residue of objective truth. But it is not clear that science operates (*even* in the long run) quite as objectively as we might like to think. An examination of the record of actual scientific practice may shed some light on that question. The goal is not to clear up the rules and mechanisms that regulate the eternal ups and downs of fashions and fads in theoretical physics. It is not certain that there is *a* set of rules and mechanisms, but we can learn what some of them might plausibly be.

There are inherent dangers in studying fairly recent episodes in physics (or in anything else) (Brinkley, 1984). But, since the final 'verdict' is not yet in and since many of the participants are still alive, there are opportunities here that are not available in more traditional 'corpse dissections'. Most obviously, one can ask the major figures involved what their motivations were, how they saw events at the time, and what they recall about the interactions of other scientists. An obvious danger in gathering such recollections is that people sometimes feel (rightly at times, but often not) that their own contributions have been slighted. This 'interview' approach can be taken too far, as when sociologists of science monitor the day-to-day routine activities of scientists. An additional useful dimension can be added by examining a recent episode in science.

Many of the most interesting questions in the philosophy of science come from studying actual scientific practice, rather than from armchair *a priori* reasoning that philosophers sometimes engage in. I

have chosen the dispersion-theory and *S*-matrix theory program of theoretical physics because I had some familiarity with the technical literature of that program in the 1960s and 1970s, because the major activity in that area was confined to a time period of several years and that activity was reasonably localized (around relatively few central theorists), and because several philosophically significant issues, such as the origin, development and selection of theories, can be illuminated with specific instances from a history of that program.

A difficulty in doing a case study of a major episode in modern theoretical physics is that one of the traditional sources of corroboration – an extensive personal correspondence among the major creators of the theories – is by and large no longer available. That is, historians of science are wary of taking at face value and relying solely upon the personal recollections of individuals. While such recollections are an invaluable source for leads about what actually went on behind the 'story' as reconstructed from the published physics literature, those recollections must be checked for support against other documents, usually the published literature and the private correspondence among key theorists. While the leading theorists of an earlier era (e.g., Einstein, Bohr, Schrödinger, Heisenberg, Pauli) did correspond frequently and extensively with one another (and much of that correspondence survives), markedly increased use of the telephone and relatively easily-available travel to many topical conferences have obviated the need for such correspondence among already busy individuals. The situation for the history of recent experimental physics is not so bad since laboratory notebooks and, more often, research proposals to funding agencies and the internal memoranda of large groups give details of what was going on in the major experiments (cf., Galison, 1987). However, research proposals for theoretical work provide a less reliable guide to what a theorist actually ends up doing.

This problem of a missing record of corrspondence among theorists is especially bad with those generations of theorists who have begun working since the end of the Second World War. Lacking a large body of such correspondence, the only resource available appears to be getting as many independent recollections as possible of key episodes in the development of the disperison-theory and *S*-matrix programs and then looking for the common overlap among these.

One can also question the value of studying a frontier area involving the creation of new physical theories, since this may be a singular exercise in science. Rigden (1987) has characterized such creative developments as follows.

> When first-rate minds are engaged in the intellectual activity called physics, as was the case in February 1927 when Heisenberg was struggling with the 'pq–qp swindle', it is an activity with no equivalent in any other natural science. In fact, there is no equivalent in any intellectual arena except, possibly, first-rate theological thinking. These special times in physics do not come often, but when they do, physicists must often create new constructs for which neither previous experience nor previous thought patterns provide guidance. New words representing entirely new concepts must be created, words whose meaning cannot be rendered even by the most deliberate use of older words. The new meaning takes form slowly, but with a groping awkwardness. Soon the new ideas become the basis for empirical predictions and, in the process, a 'sense of understanding' emerges. However, in the end, the basic concepts of physics are aloof, they remain outside our ability to convey their meaning.

Rather than taking this to mean that such activities in theoretical physics are largely irrelevant to the philosophy of science, we can see in these episodes a unique opportunity to examine how foundational theories are created – perhaps at a time of singular flexibility and underdetermination of the outcome.

1.2 Philosophical issues and the Forman thesis

Since a primary interest of ours here is certain philosophical questions, references are not given to *every* technical development in S-matrix and dispersion theory. By examining in detail a major episode in contemporary physics, we hope to illuminate somewhat the processes by which theories are generated and selected by the scientific community. A question of central interest for us is the relative importance of internal versus external factors in the development of a scientific theory. Forman (1971) initially raised this issue with regard to the origin and acceptance of the concept of acausality in physics in Germany after the First World War when modern quantum mechanics was being formulated. We shall often use the expression 'Forman thesis' to refer more generally to the role of social and sociological influences in the development and acceptance of a scientific concept or theory. It does seem evident that, once we 'buy' into a set of starting assumptions, then the 'internal' logic of a formalism can largely take over (Raine and Heller, 1981). However, the origin of hypotheses central to a theory often lies in very specific and technical developments, having little, if anything, to do with overarching philosophical schemes. For that reason some fairly extended discussion of technical details is necessary. (The reader can get an overview of the philosophical issues and conclusions from Chapters 1 and 10 alone, aided perhaps by the introductions and brief summaries at the end of each intervening chapter.) Retrospectively, the central tenets of a theory may be put into or associated with a particular philosophical world view. Furthermore, the acceptance (or the effective infectivity) of a theory can be greatly influenced by the social environment and by generally accepted overarching principles. This case study does not support the radical Forman (1971) thesis that the social milieu plays a central role in the *creation* of scientific theories, but it is consonant with the more modest Forman-type thesis (Forman, 1979; Hendry, 1980) that social factors are relevant for the *acceptance* of a theory. There does remain an important distinction for science between internal factors (such as formalism, logic and experiment) and external ones (such as group interests and social influences).

Another set of issues to be discussed in the context of this episode in physics is the interplay between the discovery and the justification of a scientific theory (both initially and later in the program) and the symbiosis between theory and experiment in the development of a program. This is just *one* case study and its conclusions may or may not have any general applicability to the way other scientific theories have developed. It is by no means clear that there is *a* (i.e., one) scientific rationality that applies usefully to all science in all eras (in spite of some claims made, for example, by Popper, with his emphasis on falsification, and Lakatos, with his representation of the dynamics of science in terms of progressive and degenerating research programs).

1.3 The purview of this case study

Because this case study is intended mainly for historians and philosophers of science who have an interest in the modern scientific enterprise, I have attempted to give an essentially accurate representation on technical matters, but have usually avoided telling the *whole* truth (i.e., giving all the technical details). Rather, the central concepts and techniques are often illustrated with simple mathematical examples. Although I do not want to reconstruct past developments from the biased vantage of today's state of knowledge, I have nevertheless employed a unified notation in these mathematical examples in order to make the line of argument more accessible to a

wider audience. Along the way I do point out important notational and conceptual differences between my illustrative examples and the original presentations found in the physics literature.

The last introductory comment concerns my use of the expression '*S*-matrix program'. I do not mean to equate the dispersion-theory program and the *S*-matrix theory (SMT) program[2] nor do I wish to obfuscate the distinction between the 'bootstrap condition' as a uniqueness criterion and the much broader implications that term has in the program associated with Geoffrey Chew and his collaborators. This case study should make it clear (1) that at any given time the term 'bootstrap' has not had a unique, universally accepted meaning among theoretical physicists (if, indeed, it has any specific meaning at all) and (2) that within a given group or school of theorists the meaning of that term has evolved over the years. To respect this caveat, I shall use the designation 'autonomous *S*-matrix program' to distinguish the radical or fundamentally revisionary conjecture from the more general *S*-matrix and dispersion-theory program[3].

Much of the *S*-matrix program discussed in this study will appear as a largely American project and this may give the entire project too much of an American, even 'Chewian', flavor. It is true that major developments in several areas took place in Europe or have been made by Europeans. I attempt to point this out in the narration that follows. Nevertheless, it does remain that much of the major activity of the *S*-matrix program was centered around Geoffrey Chew and his collaborators in Berkeley.

1.4 Quantum field theory (QFT) background

A brief sketch of the development of quantum field theory (QFT) is necessary in order to place *S*-matrix theory (SMT) in some historical context for the reader. More complete discussions of the history of QFT can be found in Cushing (1982), Darrigol (1984, 1986) and Schweber (1984, 1990). Philosophical problems associated with QFT have been addressed by Stöckler (1984). In fact, the following is largely a summary from my (1982) paper on high-energy theoretical physics. Detailed references to the relevant physics literature can be found there, so that we shall not repeat those references here. This condensed outline is essentially ahistorical in that it presents only the central ideas of QFT without any pretense at maintaining a strict historical sequence. Let us begin by reproducing (Cushing, 1987a) a chronological outline of the

sequence of developments (Schweber, 1987) we shall take as background for the following chapters in this case study. Some of these topics will be elaborated here, others in subsequent chapters.

1925–1927: formulation of nonrelativistic quantum mechanics
1927–1947: formulation of relativistic quantum field theory (QFT)
1947–1950: renormalization program for quantum electrodynamics (QED)
late 1950s–1970: a period of serious problems for perturbative QFT, with various alternative avenues pursued
- (*a*) axiomatic QFT (Wightman school)
- (*b*) local, asymptotic QFT (LSZ formalism)
- (*c*) dispersion relations and *S*-matrix theory (SMT) ↓ duality program

1970s–present: gauge QFT ↘ ↙ strings

The transition from classical mechanics to (nonrelativistic) quantum mechanics in the period 1925-27 can be seen (at least now, retrospectively) as a reinterpretation of the equations and of the formalism of classical mechanics to represent phenomena in the atomic domain. In the Hamiltonian formulation of classical mechanics (for a single particle here), the time evolution of the canonical variables $q(t)$ ('position') and $p(t)$ ('momentum') is governed by Hamilton's equations of motion

$$\dot{q}=\frac{\partial H}{\partial p}, \tag{1.1a}$$

$$\dot{p}=\frac{-\partial H}{\partial q}. \tag{1.1b}$$

Here $H(q,p)$ is the Hamiltonian and is (in our case) just the total mechanical energy of the system. For example, a particle of mass m moving in a conservative force field $F(q)=-\partial V/\partial q$ has the Hamiltonian

$$H(q,p)=\frac{p^2}{2m}+V(q) \tag{1.2}$$

in terms of the potential energy function $V(q)$. For a classical system, $q(t)$ and $p(t)$ are simply ordinary functions of the independent time variable t. They are solutions to the coupled set of differential equations (1.1) subject to the initial conditions $q_0=q(t_0)$, $p_0=p(t_0)$ at some

(arbitrary but definite) initial time $t=t_0$. In this simple example, Eqs. (1.1) are nothing more or less than (equivalent to) Newton's second law of motion

$$m\ddot{q} \equiv ma = \frac{-\partial V}{\partial q} = F(q). \tag{1.3}$$

In the early part of the present century, it became evident that for atomic systems not all of the solutions (or 'orbits' for particle motion) are in fact allowed or realized in nature. For example, only certain orbits, or energy levels, for a bound electron in a hydrogen atom are permitted (as evidenced by the discrete spectrum of the light emitted or absorbed by a hydrogen atom). The program of the old quantum theory (say, 1913–1925) was to find a set of rules that would allow one to select from the (continuous) infinity of classically-allowed solutions (or 'orbits' or energy levels) those actually realized in nature. Bohr's classic 1913 paper gave one such rule in terms of the quantization of the orbital angular momentum l,

$$l = n\hbar, \quad \mathrm{n} = 0, 1, 2, \ldots \tag{1.4}$$

Here $\hbar$ is $h/2\pi$, where h is Planck's constant. The old quantum 'theory' amounted in essence to a set of quantization rules, that were generalizations of Eq. (1.4). It consisted of a set of *ad hoc* guesses guided by Bohr's correspondence principle, which was initially a requirement that certain quantities derived in the (old) quantum theory should pass over into their classical counterparts in a suitable limit.

Heisenberg's 1925 paper laid the foundations of a systematic quantum mechanics by reinterpreting the classical q and p variables as quantities satisfying the commutator relation (in units with $\hbar = 1$)

$$qp - pq \equiv [q, p] = \mathrm{i}. \tag{1.5}$$

(We make *no* claim that Heisenberg, Schrödinger or Dirac originally presented their ideas in the form we represent them here. This is 'Whiggish' history, which we avoid in our study proper.) Hamilton's equations (1.1) and the Hamiltonian (1.2) were to be retained, but the (operators) q and p were now required to satisfy the (commutator) condition of Eq. (1.5). That is, one must seek solutions to the eigenvalue problem

$$\mathrm{H}(q, p)\Psi = E\Psi \tag{1.6}$$

for the allowed eigenvalues E. Here (to make a long story short) Ψ is the (Schrödinger) eigenfunction (or eigenvector). One typically finds a

representation for q and p (e.g., $q \to q$, $p \to -\mathrm{i}\partial/\partial q$) which satisfies Eq. (1.5) and then uses this in the $H(q, p)$ of Eq. (1.2) to re-express Eq. (1.6) as (in units with $2m = 1$)

$$\left[-\frac{\partial^2}{\partial q^2} + V(q)\right]\Psi(\theta) = E\Psi(q). \tag{1.7}$$

This particular representation of Eq. (1.7) is usually referred to as the Schrödinger equation (which Schrödinger in 1925 arrived at independently and by a route different from that indicated here). Dirac in 1925 produced an elegant general set of rules for passing directly from the classical Hamiltonian formulation of a problem to the corresponding quantum-mechanical equations by replacing the classical Poisson brackets $\{q, p\}$ with the commutator $[q, p]$ as

$$\{q, p\} \to \frac{1}{\mathrm{i}}[q, p]. \tag{1.8}$$

Schrödinger also established the formal equivalence of his wave mechanics and of Heisenberg's matrix mechanics. We use the term 'quantum mechanics' to refer to either of these representations without distinction.

Quantum field theory – really, quantum electrodynamics (QED) – was born in 1927 when Dirac applied perturbation theory to an atomic system (such as a hydrogen atom) in a radiation field (such as the electromagnetic field of light). The stationary state problem (e.g., the energy levels in the hydrogen atom) had been solved by quantum mechanics, but the mechanics or details of light emission and absorption remained to be handled. The quantum-mechanical problem considered was

$$\mathrm{i}\frac{\partial \Psi}{\partial \mathrm{t}} = (H_0 + V)\Psi, \tag{1.9}$$

which is just the standard time-dependent Schrödinger equation for an atomic system (whose unperturbed Hamiltonian is H_0) acted upon by an external potential V (here the radiation field due to light). (Similar applications of Schrödinger theory are discussed in more detail in the Appendix which should be consulted if the reader feels at a loss for specifics.) The equations of motion (i.e., Maxwell's equations) for the free electromagnetic field can be transformed into an equivalent set having the form of a denumerably infinite set of uncoupled harmonic oscillators, which can easily be quantized in terms of a set of creation

($a^\dagger$) and annihilation (a) operators satisfying the commutation relation

$$[a, a^\dagger] = 1. \tag{1.10}$$

That is, if Ψ is a state having n quanta (say, photons), then $a^\dagger\Psi$ is a state having $(n+1)$ quanta and $a\Psi$ is a state having $(n-1)$ quanta. Straightforward perturbation theory (think, say, in terms of the strength or intensity of the radiation field as the expansion or 'smallness' parameter) can be applied to compute the transition probability from one atomic (or hydrogen atom) level to another (under the stimulus of the radiation field). Calculations to *lowest* order in this expansion (or perturbation) turned out to be finite and reasonable. However, when this formalism was applied to the problem of the dispersion (or scattering) of light by a collection of atoms, divergent terms were present. During the period from the late 1920s throughout the 1930s, theoretical physicists were occupied with these and other serious consistency problems of quantum field theory, as well as with attempts at formulating a satisfactory relativistic quantum field theory. For our present telescoped 'history' of QFT, we simply pass over this period and turn to the major developments that occurred immediately after the Second World War.

1.5 Renormalized quantum electrodynamics (QED)

Lamb and Retherford in 1947 determined experimentally that two energy levels of hydrogen (the $2s_{\frac{1}{2}}$ and the $2p_{\frac{1}{2}}$) which should have been degenerate in one-electron Dirac theory are in fact separated by a small but finite energy difference. For here and for reference later, it may be worth pointing out that the energy-level notation is nl_j, where n is the principal quantum number of the Bohr theory (that is, the n of Eq. (1.4)) and takes on values (in units of $\hbar$)n $= 1, 2, 3, \ldots$; l is the orbital angular momentum and takes on values $0 \leqslant l \leqslant n-1$ (with the old spectroscopic notation $l=0 \leftrightarrow s$, $l=1 \leftrightarrow p$, $l=2 \leftrightarrow d, \ldots$); and j is the total angular momentum (l plus $s=\frac{1}{2}$, the spin of the electron). Thus, $2s_{\frac{1}{2}}$ stands for $n=2, l=0, j=\frac{1}{2}$. In Dirac's relativistic theory of the hydrogen atom, the energy levels E_{nj}, depend only upon n and j, but not explicitly upon l. For $l=1$ (and $s=\frac{1}{2}$, of course), there are two possible values for j (where $\mathbf{j}=\mathbf{l}+\mathbf{s}$): $j=\frac{1}{2}$ and $j=\frac{3}{2}$. Hence, the $2s_{\frac{1}{2}}$ $(n=2, j=\frac{1}{2})$ and $2p_{\frac{1}{2}}$ $(n=2, j=\frac{1}{2})$ energy levels should be the same (or 'degenerate'). In Figure 1.1 we show the Dirac-theory predictions for the energy level E_{nj} with solid lines and the actually observed $2p_{\frac{1}{2}}$ level with a dotted line. The experimentally observed split between the $2s_{\frac{1}{2}}$ level and the (dotted line)

$2p_{\frac{1}{2}}$ level is the Lamb shift. This difference is observed as a splitting or displacement of the spectral lines emitted when electrons make transitions from one level to a lower one.

In a remarkable calculation that same year, Bethe made a *nonrelativistic* perturbation calculation of the self-energy of an electron in a bound state of a hydrogen atom. The most singular term was a linearly divergent integral that Bethe realized was present for a free electron. (By linearly divergent we simply mean that the calculation yielded an expression involving an integral of the form $\int^\infty \mathrm{d}k$, which diverges linearly (i.e., as a first power of k) as $\lim_{k\to\infty} k \to \infty$.) Arguing that this should be included in the physically observable mass of the electron, he discarded it as a not-separately observable effect. There still remained a logarithmic divergence (i.e., of the form $\int^\infty \mathrm{d}k/k$) of the type already known from Dirac's hole theory which he also discarded in the hope that a relativistic calculation would produce the same logarithmic term plus, possibly, an additional small finite correction. Bethe took the finite remainder of his calculation to be *the* Lamb shift. His calculation produced a result

$$\Delta W/h = 1040 \text{ megacycles } s^{-1} \tag{1.11}$$

to be compared with the early experimental value of 1000 Mc s^{-1}. The extreme accuracy of these measurements becomes clear when we appreciate that $\Delta W_{\text{Lamb}}/W_{\text{energy level}} \approx 10^{-7}$.

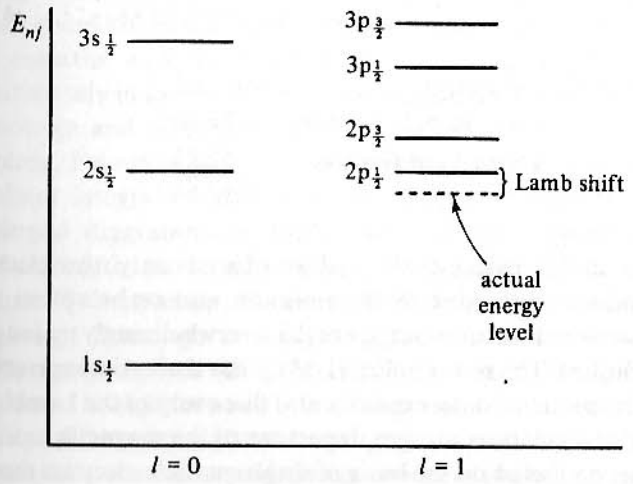


Figure 1.1 Hydrogen atom energy levels in one-electron Dirac theory and the Lamb shift (not drawn to scale).

Kramers pointed out at the 1948 Solvay Conference that observable effects calculated in quantum electrodynamics depend upon e and m only as measured experimentally (that is, upon these *structure-independent* properties of the electron). At this time people were able to do some higher-order corrections to electron scattering cross sections by subtracting out (that is, really, just throwing away) these divergences as they occurred in calculations. However, the answers obtained, while finite, were not unique and the entire procedure was a delicate and ambiguous one. Kroll and Lamb in 1949 made a relativistic quantum-electrodynamics calculation of the level splitting for $2s_{\frac{1}{2}}-2p_{\frac{1}{2}}$ and obtained 1051 Mc s^{-1}. The results were still somewhat ambiguous since subtraction of infinite terms to yield finite ones was not unique because there was no subtraction procedure independent of a specified Lorentz frame. Also, these calculations were carried out to order e^2 in these so-called radiative corrections and were not yet known to remain finite in higher order. 'Radiative correction' is the term used to refer generically to any virtual process in which a photon, say, is radiated (perhaps from an electron) and then reabsorbed (by the same emitter) before it can be observed directly. Virtual processes are essentially those that are allowed by the QED formalism but that cannot be seen directly. We shall give illustrations of several of these below.

Since the Lamb shift remains one of the showpieces of QED, let us give some indication of recent experimental and calculational results.

$$\textit{experimental}:\ \Delta E\equiv\Delta E(2s_{\frac{1}{2}})-\Delta E(2p_{\frac{1}{2}})=1057.77\pm0.01\ \text{Mc s}^{-1}. \tag{1.12}$$

theoretical:	self-energy	1011.45
	vacuum polarization	−27.13
	vertex modification	67.82
		1052.14

(1.13)

Bethe in his initial calculations had estimated only the electron self-energy modification (due to the emission and reabsorption of a virtual photon) which, it turns out, gives the overwhelmingly major part of the contribution. The result in Eq. (1.13) is the theoretical correction to order e^2. Subsequent to the experimental discovery of the Lamb shift and to Bethe's calculation, a slight departure of the magnetic moment from its value, predicted on the basis of single-particle electron theory, was detected experimentally by Foley and Kusch in 1948. That is, according to Dirac (single-electron) theory, the electron should have an

intrinsic magnetic moment of

$$\mu=\frac{e\hbar}{2mc}, \tag{1.14}$$

whereas the experimental value is

$$\mu_{\text{exp}}=\mu+\delta\mu \tag{1.15}$$

with

$$\delta\mu=(0.001\,165\pm0.000\,011)\mu. \tag{1.16}$$

The QED prediction is to order e^2

$$\delta\mu=\frac{\alpha\mu}{2\pi}=0.0\,011\,614\mu \tag{1.17}$$

as shown by Schwinger prior to the experimental measurement.

In his report to the 1948 Solvay Conference, Oppenheimer gave a masterful summary of the state of QED at the time. The first problem was to find a manifestly Lorentz-invariant and gauge-invariant method (or prescription) for subtracting (or neglecting) infinite quantities that arise in corrections made to finite order in $\alpha=e^2/\hbar c(=1/137)$, the fine-structure constant (*beyond* the already-finite terms that are first-order in α). The basis for this was provided by Schwinger and by Tomonaga and applied by Schwinger to calculating, to various orders in α, corrections to the magnetic moment of the electron and to the scattering of electrons by a Coulomb field. Several of these results were simply 'announced' by Schwinger in a 1947 work with no calculational details given. In these formulations a perturbation series expansion in α was essential and the renormalization was proved to remove the infinities only in lowest order. In contrast to the highly formal theory of Tomonaga and of Schwinger, which stressed the wave aspect of the problem, Feynman, whose early papers involved a modification of divergent integrals that ultimately disappeared from the final result, developed diagrammatic techniques that were based on concepts emphasizing the particle-like aspect of the interactions of electrons and photons. Subsequently, Dyson in 1949 showed the equivalence of the Tomonaga–Schwinger and Feynman formalisms. Feynman's diagrammatic expansion has the great virtue that it makes QED relatively simple and straightforward for calculations. Dyson also established the validity of the renormalization program to all orders in α. That is, he proved that, after mass and charge renormalization, every term in the perturbation expansion is *finite*, although the convergence of the series itself has never been established.

The renormalization program in QED consists basically of the following. In doing a perturbation correction to the scattering amplitude to first order in $\alpha = e^2/\hbar c (= 1/137)$, one obtains two types of divergent integrals, one of which can be associated with an electromagnetic mass δm which modifies the 'bare' mass m_0. One simply takes the (finite) physically observed mass m to be

$$m = m_0 + \delta m. \tag{1.18}$$

Only m appears in *calculated quantities* which correspond to physically observable processes, but never m_0 or δm *separately*. Similarly, to first order *corrections* in α another divergent term occurs which again appears only in a suitable linear combination with the 'bare' charge e_0 of the electron. This can be interpreted as a modification of the charge due to (vacuum) polarization effects and the combination is taken to be the physically observed charge e of the electron. Here, too, neither e_0 nor the polarization term appears *separately*, but only the combination making e. Therefore, up through and including all terms of order α and α^2 in the scattering amplitude, charge and mass renormalization may be employed to make the calculated theory finite. It is important to realize that this procedure can be carried out in a completely covariant, and therefore unambiguous, fashion and that these results agree extremely well with experiment (see, for example, Eqs. (1.12)–(1.13) and (1.16)–(1.17) above). The crucial question now becomes what happens when higher-order terms in α are calculated in the perturbation expansion. Divergent terms again appear, but they can *all* be written in terms of those of mass and charge renormalization, and *no* others. In fact, to *every* order α^n, the same procedure of mass and charge renormalization will remove all infinite quantities, leaving finite corrections. This is a remarkable circumstance and is necessary for QED to be useful as a calculational tool. If new types of divergences were to occur in each successive order of perturbation, then, even if we could associate each of these infinities with a renormalization or redefinition of various physically observable quantities, the theory would have *no* predictive power. Any field theory for which a *finite* number of redefinitions is sufficient to remove all the divergences to all orders of perturbation theory is termed *renormalizable*.

The renormalized QED program can be summarized as follows. The time evolution of the state vector $\Phi(t)$ is governed by the interaction Hamiltonian H_I as (in units with $\hbar = 1$)

$$i\frac{\partial \Phi(t)}{\partial t} = H_I \Phi(t) \tag{1.19}$$

which has the formal solution

$$\Phi(t) = \Phi(-\infty) - i\int_{-\infty}^{t} dt_1 H_I(t_1)\Phi(t_1) \tag{1.20}$$

satisfying the initial condition

$$\Phi(t) \xrightarrow[t \to -\infty]{} \Phi(-\infty). \tag{1.21}$$

Here $\Phi(-\infty)$ is the initial state in which the system was prepared. For a scattering process we are interested in the state vector in the remote future $\Phi(+\infty)$ and this is connected to the initial state vector $\Phi(-\infty)$ *via* the S operator (or S matrix) as

$$\Phi(+\infty) = S\Phi(-\infty). \tag{1.22}$$

Once the S matrix (or scattering amplitude) is known, the scattering cross section (or probability) for a given reaction can immediately be calculated (essentially as the square of the modulus of S). If we recursively iterate Eq. (1.20), we obtain the formal expression for S in terms of the Hamiltonian density $H_I(x)$ as

$$S = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int d^4x_1 \ldots \int d^4x_n P\{H_I(x_1) \ldots H_I(x_n)\}. \tag{1.23}$$

Even though we have not specified the precise meaning of all the operations on the right-hand side of Eq. (1.23), we have stated it to indicate both that, once the interaction has been specified in terms of $H_I(x)$, then S can be computed and that the (infinite) series is an expansion in ascending powers of $H_I(x)$ (the perturbation expansion so often referred to in this section). In QED, $H_I(x)$ contains an overall multiplicative constant e, the electron charge.

It is basically an historical accident that the first quantum field theory (QED) compared with experiment turned out to be renormalizable. Salam, in 1952, proved the renormalizability of certain other field theories (in which more than just two renormalization constants appeared), but many (in some loose sense, 'most') quantum field theories are nonrenormalizable. In fact, Fermi's 1933 β-decay theory (the paradigm process of which is the spontaneous decay of a free neutron into a proton with the emission of an electron (or 'β ray')) provides an example of a nonrenormalizable theory. A concise statement of any renormalization program has been given by Matthews and Salam as consisting of three steps:

1. the number of types of infinities must be shown to be finite;
2. a subtraction procedure must be found to remove the infinities;

3. a theoretical justification for this procedure must be found (i.e., it must be shown to hold to all orders in perturbation theory).

Even if a theory is renormalizable, though, this does not necessarily make it useful for calculations since the renormalization program and the calculations themselves can be carried out explicitly only within the framework of perturbation theory. For QED this is fine since the expansion parameter (or coupling constant) $\alpha = 1/137$ is small and the first few terms in the series might reasonably be expected to give a good approximation (even if the series itself should only be an asymptotic one). However, in field theories relevant to nuclear and elementary-particle physics, the coupling constant (which measures the strength or rate of a process) is of the order $g \approx 15$, so that a perturbation expansion is useless to obtain numerical results. These difficulties for strong-interaction theory were amply apparent by the early 1950s. The numerical results were in terrible disagreement with experiment. This was one of the motivations for theorists' turning to other programs, such as dispersion theory and S-matrix theory in the late 1950s and throughout the 1960s. These alternative avenues of research are the subject of the following chapters in this case study.

However, it remained clear that even QED was at base a mathematically inconsistent theory. There had been some hope that the infinities that arose were a result of the approximate nature of the perturbation calculations rather than of the theory itself. As a rather naive analogy, consider the following power series expansion for $x > 0$

$$e^{-1/x} = 1 - \frac{1}{x} + \frac{1}{2!x^2} + \cdots . \tag{1.24}$$

In the limit $x \to 0^+$, every term in the series (except the first) diverges, but the exact result is zero. However, that such is not the case for the renormalization constants of QED was indicated in 1953 by Källen. Independent of any perturbation theory, he argued that in an exact formulation of QED not all of the renormalization constants of QED can be finite. Dyson (1952) and Edwards (1953) have questioned the convergence of the perturbative expansion. Landau (1955) and his coworkers (Landau, Abrikosov and Khalatnikov, 1954a, 1954b, 1954c, 1954d; Landau, 1965) stressed the importance (at high energies) of singularities that are not handled by renormalization. Schwinger has pointed out the fundamental inconsistency of assigning operators for definite-mass fields at localized space-time points (e.g., $\psi(x, t)$) since a precise measurement of these properties would result in arbitrarily large field fluctuations via the uncertainty principle (i.e., the interaction

energy cannot be limited while using an exact space-time description). Also, the actual magnitude of the electron's charge is explainable only once we understand the strong interaction because charge renormalization must, in principle, be linked to *every* charged field with which the electron can interact. Dirac has discussed similar difficulties with QED.

1.6 Feynman diagrams

For reference in later chapters, we now introduce some elementary concepts associated with Feynman diagrams. Before that, though, we define some terminology appropriate to describing scattering phenomena. Some of the best-known and most precise predictions of QED are quantities like the Lamb shift and the anomalous magnetic moment of the electron and these would seem to have little to do with scattering of one particle by another. Nevertheless, most of our basic information about fundamental interactions is gained through scattering processes, as we shall see throughout this book. Figure 1.2 represents a typical scattering event in which an incoming projectile (with initial velocity $\mathbf{v}_0$) is scattered (or deviated) by a target, which will itself recoil (through an angle ϕ with a final velocity $\mathbf{V}$) during the interaction. How the projectile is scattered (as indicated by the angle θ and the final velocity $\mathbf{v}'$ in the figure) is determined by the interaction or force between the target and the projectile. And, conversely, it is through a study of such scattering reactions that we learn something about the details of the forces acting between the particles. In Chapter 2 we discuss scattering formalism in some detail. Feynman diagrams will prove useful for treating scattering reactions.

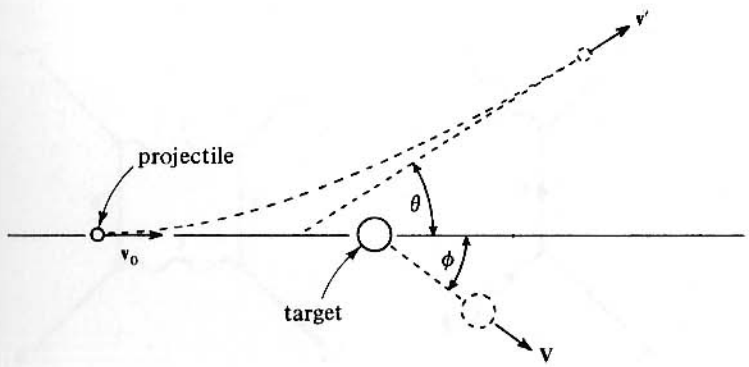


Figure 1.2 The scattering of a projectile by a target.

Basically, Feynman constructed a diagrammatic representation of the perturbation expansion of Eq. (1.23) for the S matrix (or for the scattering amplitude) and a set of rules that allow one to assign a definite mathematical expression to each of these diagrams. That is, there is a one-to-one correspondence between the terms in the perturbation expansion of Eq. (1.23) and the Feynman diagrams (or 'pictures') for any given scattering process. A few simple examples will illustrate the points we shall need for future reference. Figure 1.3 shows the lowest-order Feynman diagram for (free) electron–photon scattering (also known as Compton scattering). This can be pictured as (taking time to flow upward in the figure) the absorption of a photon (γ) by the initial electron (e_i^-), the propagation of a virtual electron (e^-) in the intermediate state and then the subsequent emission of a photon (γ) leaving the final electron (e_f^-). The complete Compton scattering amplitude is obtained by summing *all* possible Feynman diagrams (of which there are infinitely many), allowing for arbitrarily many virtual electrons and photons in the intermediate states. At each $e^-e^-\gamma$ vertex one picks up a factor e in the perturbation expansion. Thus, the Feynman diagram of Figure 1.3 would make a contribution of order e^2 (or α) to the scattering amplitude. We shall not state in any detail the specific Feynman rules for recovering the mathematical form of the contribution to the perturbation expansion from the corresponding diagram. A great virtue of the Feynman diagram technique is that it allows one to 'picture' a scattering process in terms of the exchange of virtual particles.

Figure 1.4 shows two lowest-order Feynman diagrams for electron–electron (or Møller) scattering. If we denote by $p_j, j=1,2,3,4$, the four momenta of the incident and scattered electrons, then (neglecting

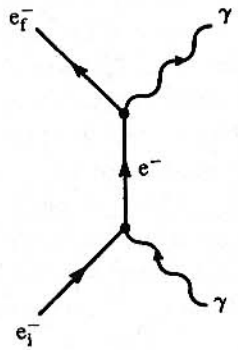


Figure 1.3 Electron–photon (Compton) scattering.

electron spin and photon polarization complications here) the contribution to the scattering amplitude corresponding to Figure 1.4(a) is proportional to

$$\frac{e^2}{(p_1-p_3)^2} \tag{1.25}$$

and that of Figure 1.4(b) to

$$\frac{e^2}{(p_1-p_4)^2}. \tag{1.26}$$

Figure 1.5 is an example of a higher-order contribution to e^-–e^- scattering (of order e^4 or α^2). Finally, Figure 1.6 is a renormalization contribution (a 'vertex' radiative correction) that contributes to charge renormalization. The integral corresponding to this diagram diverges, but the renormalization procedure allows one to assign a *finite* contribution to the scattering amplitude.

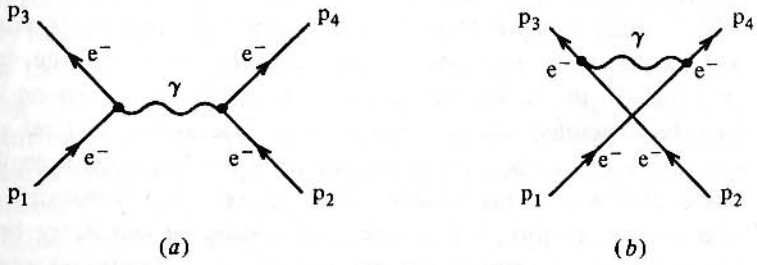


Figure 1.4 Electron–electron (Møller) scattering.

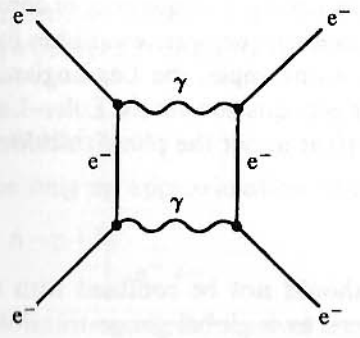


Figure 1.5 A higher-order diagram for electron–electron scattering.

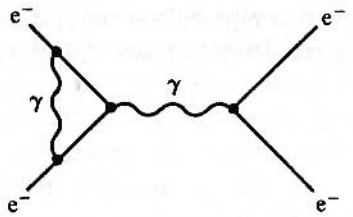


Figure 1.6 A renormalization contribution.

1.7 Gauge field theories

We have already indicated that by the late 1950s even renormalized QFT was unable to cope with strong-interaction phenomena, thus producing interest in the *S*-matrix program. However, QFT did make a strong comeback in the 1970s, in the form of gauge field theories. To complete our brief historical sketch, we now mention some of the key developments in that program. The basic idea used in modern gauge field theories was put forward in 1954 by Yang and Mills. Although their argument concerned a symmetry between protons (p) and neutrons (n) in the strong interactions, the line of reasoning is most directly explained for a simpler case. Just as in classical particle mechanics any invariance or symmetry of the Lagrangian implies the existence of a conserved quantity (or generalized momentum), so in a field theory. (Think, for example, of rotational symmetry implying conservation of angular momentum or of translational symmetry yielding conservation of linear momentum.) Noether's theorem guarantees that, corresponding to any transformation or invariance group that leaves the Lagrangian density (representing the interactions) for the theory invariant, there exists a conserved quantity or constant of the motion (that is, a quantity whose time derivative vanishes by virtue of the equations of motion). As an example, the Lagrangian, which yields the nonrelativistic Schrödinger equation *via* the Euler–Lagrange variational equations, is left invariant under the phase transformation of the wave function $\psi(x)$

$$\psi(x)\rightarrow\psi'(x)=e^{i\alpha}\psi(x) \tag{1.27}$$

where α is a *constant*. (This α should not be confused with the fine structure constant.) This is known as a global gauge transformation since, once α has been chosen, the phase is fixed at every space-time

point (i.e., globally). Essentially, what Yang and Mills argued was that such a 'nonlocal' fixing of the phase (in their case, the fixing of the proton p relative to the neutron n) should be avoided in a local field theory by requiring *local* gauge invariance as

$$\psi(x)\rightarrow\psi'(x)=e^{i\alpha(x)}\psi(x) \tag{1.28}$$

where $\alpha(x)$ is now an arbitrary function of the space-time variable x. However, the Lagrangian for the coupled electromagnetic field ($A_\mu(x)$) and the electron wave function ($\psi(x)$) is no longer invariant under Eq. (1.28) unless the electromagnetic four-potential $A_\mu(x)$ is simultaneously subjected to the transformation

$$A_\mu(x)\rightarrow A'_\mu(x)=A_\mu(x)+\frac{1}{e}\frac{\partial\alpha(x)}{\partial x_\mu}. \tag{1.29}$$

In the context of electromagnetic theory, this is just Weyl's gauge invariance.

Yang and Mills were able to show how to implement a general symmetry group by means of such local gauge fields. The only problem was that each gauge field had associated with it a massless particle, just as $A_\mu(x)$ has the photon. This simply produced too many massless particles that were not found experimentally. For well over a decade, the Yang–Mills theory was considered an interesting but essentially useless curiosity. It is worth noting that their motivation for introducing such local gauge transformations was an abstract or purely theoretical one, not really required by any empirical evidence.

The other problem that provided impetus for the current gauge theories of weak interactions was the fact that the original 1933 Fermi theory of weak interactions is nonrenormalizable and its lowest-order term eventually violates unitarity (or conservation of probability) at high enough energy, a fact realized long ago by Heisenberg in 1936. It was conjectured that Fermi's form of the interaction in which the neutron *n* decays directly into a proton p and electron e and a neutrino ν as

$$n\rightarrow p+e^-+\nu \tag{1.30}$$

might be only an approximation to the actual two-step process

$$\begin{aligned} n&\rightarrow p+W\\ &\quad\hookrightarrow e^-+\nu \end{aligned} \tag{1.31}$$

where W is the so-called (massive) intermediate vector meson. Salam in 1960 had shown that neutral (i.e., uncharged) vector meson theory was

renormalizable. Unfortunately, the (charged) W meson theories are known to be nonrenormalizable unless certain very specific cancellations of divergent terms happen to take place. The problem was to find a renormalizable charged vector meson theory that could describe the weak interactions.

The resolution began when Goldstone in 1961 showed that, even though a Lagrangian (or, equivalently, the equation of motion) for a field theory might possess a certain symmetry, this symmetry could be spontaneously broken so that the solution had a lower degree of symmetry than the Lagrangian. (Spontaneous magnetization of a solid, which does pick out a preferred direction in space, is an example of a solution, or physical situation, having a lower degree of symmetry than the fundamental equations describing the system. The physical solution has only axial symmetry, whereas the equations themselves have complete spherical symmetry (i.e., complete rotational invariance).) Goldstone proved that this can occur whenever the Lagrangian is invariant under a continuous symmetry group but *only* if zero-mass bosons (that is, integer-spin particles) are present. This spontaneous symmetry breaking seemed to offer a plausible mechanism for explaining the broken symmetries (such as SU(3) of Gell-Mann and of Ne'eman) that exist in strong-interaction dynamics, except for the fact that massless particles would again be present, and none were observed experimentally. The proof of this theorem was generalized by Goldstone, Salam, and Weinberg in 1962. In a series of papers Higgs not only established that the proof of the Goldstone theorem breaks down in theories which couple a conserved current to *gauge* fields but also exhibited a particular model field theory in which the Goldstone boson acquires mass and becomes part of a vector meson field (which had initially been massless) so that no massless particles exist in the (broken) solutions to the model.

In the late 1960s Weinberg, and independently Salam, combined the Yang–Mills gauge fields with the Higgs mechanism to produce a unified theory of the weak and electromagnetic phenomena. An analogy between many properties of weak and electromagnetic interactions was used as motivation. In this Weinberg–Salam model the Yang–Mills vector boson field responsible for the weak interactions is put on an equal footing with the photon field (i.e., the electromagnetic four-potential) and both are initially massless, as are the electron and neutrino fields. A Higgs field breaks this symmetry, giving the electron its mass as well as the intermediate vector boson its mass. The massless Yang–Mills vector bosons and the Goldstone bosons conspire so that

the physically observed 'broken' theory has a massive charged spin-1 field W_μ (the charged intermediate vector boson), a neutral massive spin-1 field Z_μ (the neutral intermediate vector boson), and the massless spin-1 photon field A_μ, in addition to a massive electron and a massless neutrino. Initially it was a hope that this theory would be renormalizable. In 1971 't Hooft proved that such is indeed the case. It is precisely the gauge vector–meson theories that are the renormalizable ones. The Higgs meson plays a peculiar role in the Salam–Weinberg model since it is essential to provide the spontaneous symmetry-breaking mechanism and yet, since these Higgs bosons have not to date been observed experimentally, its mass must be made extremely high so that it has essentially no effect on many of the predictions of the theory.

Renormalization assumes a distinctly different status in gauge theories from what it had in previous QFT. In QED, for instance, one *proved* renormalizability. In gauge theories, as is particularly clear in Salam's writing, renormalizability is used as a *criterion* for constructing an acceptable theory. It provides a guide to writing down a suitable Lagrangian just as Lorentz invariance and conservation laws (*via* Noether's theorem) restrict the possible forms of the Lagrangian. This constraint severely limits the choice of Lagrangians.

This fairly complete and well-confirmed Salam–Weinberg model for the unification of the weak and electromagnetic interactions has prompted an attempt to unify these with the strong interactions in a grand unified field theory (or GUT in the trade). One first attempts to build a gauge theory of strong interactions. The basic ingredients in such a theory are quarks (that is, fractionally charged, spin $\frac{1}{2}$ elementary particles), implementation of SU(3) symmetry *via* Yang–Mills gauge fields, the spontaneous breaking of this symmetry by the Higgs mechanism, and the requirement of renormalizability. At first sight one might expect such gauge theories of strong interactions to be of little value for making numerical calculations since ordinary QFT perturbation series diverge for the large coupling constants of the strong interaction. However, a remarkable property of such gauge theories (often referred to as quantum chromodynamics, QCD), known as asymptotic freedom, was proved by Politzer. This means that for certain high-energy scattering processes the lowest-order perturbation calculations provide the major contribution to the cross sections (which are a measure of scattering reactions). The structure, or identifying characteristics, of scattering processes governed by QCD have been calculated using asymptotic freedom and have been confirmed by experiment. One can further attempt to unify these two gauge theories

(Salam–Weinberg and QCD) into one overarching grand unified theory of weak, electro-magnetic and strong interactions (GUT). Discussion of that work is not necessary for our present purposes. There has recently been a merging of ideas from the duality program (from *S*-matrix theory) and from gauge QFT in the superstring theories of current interest. String theories and duality are part of the material of Chapter 8.

1.8 Summary

To give the general reader an overall picture of the problem background out of which the *S*-matrix program emerged, we have presented a sketch of the highlights in the developments of quantum field theory (QFT) in this century. The central theme in that story is one of successive reinterpretations of the formalism (or of the equations) of classical (Hamiltonian and Lagrangian) mechanics: first (Heisenberg in 1925) by promoting the canonical variables q, p of classical mechanics to the noncommuting operators, $[q, p] = \mathrm{i}$, of nonrelativistic quantum mechanics; then (Dirac in 1927) by considering the (Schrödinger) wave function Ψ itself to be a quantized field operator. The long struggle (1927–47) with mathematical inconsistencies in QFT culminated (1947–50) in the highly successful renormalization program of quantum electrodynamics (QED). As we shall see, Heisenberg (in 1943) proposed his initial *S*-matrix theory (SMT) as a possible means of circumventing the (pre-renormalization) divergences of QFT. Subsequently, renormalization procedures proved adequate for the electromagnetic interactions. The Feynman diagram technique (of 1949) provides a useful pictorial representation of the physical scattering processes corresponding to the mathematical expansion of the perturbation series of quantum field theory. When the perturbation-series expansion method of calculation proved inadequate for the strong interactions (mid to late 1950s), theorists turned to dispersion relations and again to the *S*-matrix program. After some initial successes, SMT encountered calculational difficulties at a time (1970s) when QFT made a dramatic comeback in the guise of gauge field theories. Today's (1980s) superstring theories can be seen as a merging of ideas from the duality program of SMT and from gauge field theories.

In this chapter we have defined the scope and context of an historical case study of an episode in modern theoretical physics. The sources of this largely internal history will be the published scientific literature and correspondence from the participants themselves. Cognizance is also taken of the larger social context, such as the structure of the scientific community, within which the theoretical activity took place. The relevance of historical developments to pertinent philosophical issues will be pointed out in the narrative itself. Results of this study will then be used to address questions about the construction, selection and justification of theories in science. The central subject of this book is the *S*-matrix program, from its beginnings in the late 1930s and early 1940s with Wheeler and Heisenberg, as an interim program from 1943 to 1955, as dispersion theory and mass-shell quantum field theory from the mid 1950s to the 1960s, then as the bootstrap and duality models of the 1970s, leading finally to the current superstring theories. We shall analyze the role of the *S*-matrix program played in generating these theories of current interest.

2

Origin of the *S* matrix: Heisenberg's program as a background to dispersion theory[1]

A common perception of Heisenberg's *S*-matrix program of the 1940s is that it encountered difficulties quite early on and then quickly died out. One can easily get the impression that the original Heisenberg program was irrelevant for the theoretical developments that provided the background out of which the dispersion-theory and later *S*-matrix theory program emerged. In this chapter we wish to show that Heisenberg's original program posed a set of questions the criticism of and response to which led to the dispersion-theory program of Goldenberger and Gell-Mann. It is not our purpose to review all of the elementary particle physics of the 1950s, or even all of what today, in retrospect, is judged to have been the 'best' or most important physics of that period.

In correspondence or in direct conversations, Geoffrey Chew[2], Marvin Goldberger[3], and Murray Gell-Mann[4] all recall that Heisenberg's old *S*-matrix program had essentially no direct influence on their own work which led to the dispersion-theory and *S*-matrix theory programs of the late 1950s and of the 1960s. All had known of Heisenberg's general ideas from some lectures given at the University of Chicago by Gregor Wentzel or possibly from reading Heisenberg's papers. It was only later that they became aware of any relevance of their work to Heisenberg's *S*-matrix program.

We begin by reviewing Heisenberg's program (Cushing, 1982, 1986a; Grythe, 1982; Oehme, 1989; Rechenberg, 1989) and the considerable theoretical activity related to it during the period from the mid-1940s to the early 1950s. A summary of the theoretical background available in the mid to late 1930s is given in the Appendix. There the reader can find an elementary discussion of the main results we draw upon, as well as several simple examples that we use later for illustration of subsequent

developments. In the main body of this work there are many references to the equations of the Appendix. Readers with little background in theoretical physics, but with a general knowledge of undergraduate physics, may find it helpful to begin with the Appendix. The ground rules, here and in the rest of this work, for the technical level of the presentation will be roughly the following. Fairly detailed technical arguments will be confined to subjects requiring as background undergraduate physics or what can be gleaned from the Appendix. The discussion of more advanced topics will be less quantitative, although some nodding or 'popular' acquaintance with the 'pictures' (or diagrams) of Feynman will be assumed. The discussion in the latter part of Chapter 1 will suffice for this. Furthermore, several fairly technical developments in the text are set off in smaller type. The essentials of the argument can be followed without studying those sections in detail.

Aside from the difficulties presented for quantum field theory by the confusing status of cosmic ray experiments (in the late 1930s), the early developments, pro and con, in the *S*-matrix program were theory motivated. That is, the divergences present in the QFT formalism and Heisenberg's belief in the need for a fundamental length in such a formalism (coupled with his or anyone else's inability to construct such a theory) led to the formulation of the *S*-matrix program. That program depended in an essential fashion on one's ability to abstract hoped-for general results (as principles of the program) from specific models and incomplete theories. The importance of the early (1940s) *S*-matrix program for subsequent developments in theoretical elementary-particle physics was a series of questions or problems that it gave rise to, such as how causality (essentially as a first-signal principle) was to be incorporated into a scattering formalism, what type of restrictions causality could place on the mathematical form of the scattering amplitudes (which in turn determine the predicted scattering cross sections of experiment) and specifically how the results of scattering experiments would determine the form of the interactions between the scattered particles (at least in model situations governed by the nonrelativistic Schrödinger equation).

An historically important and extremely fruitful interplay between the general theoretical approach to scattering phenomena and experimental practice in the 1930s and 1940's was the study of resonance scattering in nuclear physics. Just as a mechanical system, such as a bridge or even a child's swing, will exhibit exceptionally large responses, vibrations or oscillations when energy is fed into the system at the proper frequency or rate (as by marching over a bridge or by

rhythmically pushing the swing, in our two examples), so atomic and nuclear systems also react particularly strongly to certain stimuli. Hence, the blue color of the sky is due to the preferential (or strong) scattering of light in the blue part of the visible spectrum by atomic energy levels of air molecules. That is, the difference in energy between a pair of bound-state energy levels of an electron in a molecule corresponds to a frequency in the blue part of the spectrum. This produces an especially strong interaction (here, absorption and subsequent reemission of 'blue' light) between the molecule and the light of a certain frequency. In fact, a plot of the scattering cross section (or 'rate' of scattering) versus the frequency of the scattered light would show a peak or 'bump' in the cross section at such a resonance frequency. These resonance phenomena had long been known and understood (both classically and quantum-mechanically) for the interactions between light and atoms. When similar bumps or resonances were observed in the scattering of, say, neutrons from nuclei, it was natural to exploit the analogy with known resonance phenomena. At first the *form* of the mathematical equation used for the atomic resonance formula was essentially just carried over to provide a fit to the nuclear resonance scattering data. This became known as the Breit–Wigner formula. Subsequently, a profound connection was established among causality (as an upper limit on the speed of propagation of a 'cause' to produce an effect at a distant point), analyticity (as a mathematical statement of and constraint upon the smoothness and singularity structure of the function describing the scattering of a projectile by a nucleus) and the resonance or bump structure of these nuclear cross sections. Causality finally provided an underpinning for nuclear resonance formulas. This relation between causality and analyticity would prove a key insight for subsequent work in dispersion theory and S-matrix theory (a theme we develop in later chapters).

The continuity in problems and the chain of workers from the early S-matrix program to other advances in dispersion theory and quantum field theory are major themes of our story. A fertile problem background was set by this early S-matrix theory.

2.1 The S matrix: Wheeler and Heisenberg

Wheeler (1937a, b), in the context of a theoretical description of the scattering of light nuclei, introduced the concept of a scattering matrix.[5] His motivation appears to have been pure nuclear physics and his

S matrix was a tool toward that end. In modern notation, the S matrix consists of elements $S_{\beta\alpha}$, α, $\beta = 1, 2, \ldots, N$, that give the relative strengths of the asymptotic forms of the wave functions for various channels (or types of reactions) as (cf. Eq. (A.33))

$$\psi^{+}_{\alpha\beta}(r_\beta) \xrightarrow[r_\beta \to \infty]{} \frac{i}{2k_\beta r_\beta}[\delta_{\alpha\beta} e^{-ik_\alpha r_\alpha} - S_{\alpha\beta} e^{ik_\beta r_\beta}]. \tag{2.1}$$

Here $e^{-ik_\alpha r_\alpha}$ represents the incident, or incoming, (spherical) wave and $e^{ik_\beta r_\beta}$ the scattered, or outgoing, (spherical) wave. It is the square of such a wave function that is related to the probability of the corresponding reaction taking place (as we show below). In order to keep notational complications to a minimum, we outline the situation only for the $l = 0$ partial wave (or angular momentum state) in Eq. (2.1) and later. There are N possible entrance channels and also N possible exit channels. In Wheeler's paper the scattering matrix is simply an N^2 array of elements connecting the incident wave to possible exit channels. There is no sense of using this S matrix as the central entity in an independent theory. Even though Wheeler presents his scattering matrix essentially as a convenient and powerful calculational tool, he does point out that the elements of this scattering matrix are completely determined by the (vanishing of the) Fredholm determinant of the linear integral equations (which are equivalent to the Schrödinger equation) so that the asymptotic form of these wavefunctions themselves need not be found explicitly.[6] Once these S-matrix elements are known, all the relevant transition probabilities and cross sections can be calculated directly in terms of them. This can be seen as follows.*

By the same type of argument that led from Eqs. (A.20) to Eqs. (A.25) and (A.26), one can show (here for $l = 0$) that, for scattering in the channel $\beta = \alpha$ (i.e., *elastic* scattering), the cross section is

$$\sigma_{sc} \equiv \sigma_{\alpha\alpha} = \frac{\pi}{k_\alpha^2}|S_{\alpha\alpha} - 1|^2, \qquad \beta = \alpha$$

and for transfer from channel α to channel β (i.e., *inelastic* scattering)

$$\sigma_{\alpha\beta} = \frac{\pi}{k_\alpha^2}|S_{\alpha\beta}|^2. \qquad \beta \neq \alpha.$$

* Here and throughout the text, we set off in smaller type technical material that can be passed over without interrupting the continuity of the development.

Both of these can be summarized as

$$\sigma_{\alpha\beta}=\frac{\pi}{k_\alpha^2}|S_{\alpha\beta}-\delta_{\alpha\beta}|^2. \tag{2.2}$$

The (net) reaction cross section (which takes flux out of the entrance channel α) is

$$\sigma_r=\sum_{\beta\neq\alpha}\sigma_{\alpha\beta}=\frac{\pi}{k_\alpha^2}\sum_{\beta\neq\alpha}|S_{\alpha\beta}|^2. \tag{2.3}$$

As we shall discuss in more detail in the next section, conservation of flux (or probability) and the orthogonality of the ψ_α,

$$\langle\psi_\alpha,\psi_\beta\rangle=\delta_{\alpha\beta}, \tag{2.4}$$

imply that

$$\sum_{\gamma=1}^{N}S_{\alpha\gamma}S_{\beta\gamma}^*=\delta_{\alpha\beta} \tag{2.5}$$

or, as a matrix equation,

$$SS^\dagger=1, \tag{2.6}$$

which is a statement of the unitarity of the S matrix. Equations (2.3) and (2.5) together imply that

$$\sigma_r=\frac{\pi}{k_\alpha^2}(1-|S_{\alpha\alpha}|^2), \tag{2.7}$$

which is similar to Eq. (A.26). The total cross section is

$$\sigma_t=\sigma_{sc}+\sigma_r=\frac{2\pi}{k_\alpha^2}(1-\operatorname{Re}S_{\alpha\alpha})=\frac{4\pi}{k_\alpha}\operatorname{Im}f_{\alpha\alpha}(\theta=0), \tag{2.8}$$

since the $l=0$ part of the elastic scattering amplitude $f_{\alpha\alpha}(\theta)$ is just $(S_{\alpha\alpha}-1)/2ik_\alpha$. Here we have argued for the optical theorem (Feenberg, 1932) of Eq. (2.8) only in s-wave approximation. However, the proof goes through in general so that Eq. (A.27) becomes

$$\sigma_t(k_\alpha)=\frac{4\pi}{k_\alpha}\operatorname{Im}f_{\alpha\alpha}(\theta=0). \tag{2.9}$$

This remarkable and important theorem states that the *total* cross section is completely fixed once the (imaginary part of the) forward scattering amplitude is known.

The point to be emphasized is that once the scattering matrix elements $S_{\alpha\beta}$ are given, then all the cross sections, or observables of interest here, can be calculated in terms of them. This gives some indication of the power or importance of the S matrix.

In a series of papers, Heisenberg (1943a, 1943b, 1944, 1946) proposed as an alternative to quantum field theory (QFT) a program whose central entity was a matrix he denoted by S and termed the 'characteristic matrix' of the scattering problem. Cassidy (1981) has shown how Heisenberg interpreted cosmic ray showers or 'explosions' as indicating the existence of a fundamental length, Oehme (1989) has briefly outlined some results contained in Heisenberg's S-matrix papers, and I have previously discussed some of the difficulties with quantum field theories in the 1930s which motivated Heisenberg to seek an alternative to quantum field theory (Cushing, 1982, especially pp. 45–48).

In the period from the late 1930s to the late 1940s, the divergences present in quantum electrodynamics indicated to some that quantum field theory may have been a mistake. Both the theoretical and experimental situations left the future of QED in doubt (Galison, 1983a). Cosmic ray showers (or 'explosions') and the divergence of cross sections beyond a certain energy in a classical (nonlinear) field theory version of Fermi's β-decay formalism were taken by Heisenberg (1936, 1938b) to indicate the existence of a fundamental length and the need for a profound revision of elementary-particle dynamics. Not knowing what that future theory would be, he proposed the S-matrix theory as an interim program. Heisenberg wanted to avoid any reference to a Hamiltonian or to an equation of motion and to base his theory only on observable quantities. The point here is not that Heisenberg felt that the Hamiltonian (in the sense of the total energy) is not a quantum-mechanical observable. Rather, the calculational route from the basic Hamiltonian and the equations of motion of quantum field theory to many of the important experimentally observable quantities (such as bound-state energy levels and scattering cross sections) was an ill-defined one, often producing infinities.[7] Heisenberg wanted a theory based directly on finite, experimentally meaningful quantities. Since Heisenberg wanted to base his theory on general properties that were independent of any particular model, we can see why the S matrix appeared a reasonable place to begin. The formalism outlined in the Appendix indicates that, within the framework of nonrealativistic potential theory, the potential $V(r)$ determines the S matrix and hence the experimentally measurable cross sections. However, if one were given directly the S-matrix elements $S_{\alpha\beta}$, then one would be able to compute the cross sections. In quantum field theory, just as in nonrelativistic potential theory, once the Hamiltonian H for

the system has been given, the scattering matrix is (at least formally) determined. Heisenberg's desire to base a theory on observable quantities only was a return to an idea that had proven useful in his earlier successful formulation of matrix mechanics (Cushing, 1982, p. 19). That is, in his 1925 paper he set out to construct a quantum mechanics wholly in terms of observable quantities. It worked pretty well once – why not try it again!

2.2 Discussion of Heisenberg's *S*-matrix papers

Heisenberg's stated purpose in his seminal paper (1943a), 'The observable quantities in the theory of elementary particles', was to abstract as many general, model-independent features of S as possible. In the abstract and introduction to that paper we read (Heisenberg, 1943a, p. 513):

> The known divergence problems in the theory of elementary particles indicates that the future theory will contain in its foundation a universal constant of the dimension of a length, which in the existing form of the theory cannot be built in in any obvious manner without a contradiction. In consideration of such a later modification of the theory, the present work attempts to extract from the foundation of quantum field theory those concepts which are not likely to be discarded from that future, improved theory and which, therefore, will be contained in such a future theory.
>
> In recent years, the difficulty, which still stands in the way of a theory of elementary particles, has been pointed up in many ways. This difficulty manifests itself surprisingly in the appearance of divergences (infinite self energy of the electron, infinite polarization of the vacuum, and the like), which hinders the development of a mathematically consistent theory and must probably be perceived as an expression of the fact that, in one manner of speaking, a new universal constant of the dimension of a length plays a decisive role, which has not been considered in the existing theory.'

This paper is remarkable for the number of new ideas it introduces, many of which would be put on a firm mathematical basis only years later. We outline below his proof of a key property of the S matrix – namely, unitarity. This property would remain central to the S-matrix program.

Heisenberg proved the unitarity of the S matrix for any system governed by a Hamiltonian. Unitarity, as an independent principle, would remain a key ingredient in the S-matrix program. Some details of his proof are given here. Heisenberg worked in a momentum-space representation outlined by Dirac (1935, pp. 195–200)[8] and defined the S matrix as the matrix coefficient of the outgoing waves.

Let us write this somewhat symbolically[9] (in analogy with Eq. (2.1)) as

$$\psi^+_{i,k} = \phi^-_{i,k} + S_{ik}\phi^+_{i,k}. \tag{2.10}$$

Here the plane wave $\phi_{i,k}$ has been decomposed into its ingoing and outgoing components as

$$\phi_{i,k} = \phi^-_{i,k} + \phi^+_{i,k}. \tag{2.11}$$

In this representation $\phi_{i,k}$ is real and

$$(\phi^+_{i,k})^* = \phi^-_{i,k}. \tag{2.12}$$

Having defined the S matrix in general, Heisenberg then established the unitarity of S within the framework of a Schrödinger equation

$$\sum_j <k|H|j>\psi^+_{i,j} = E_i\psi^+_{i,k}. \tag{2.13}$$

Heisenberg expressed these solutions ψ^+ as a formal operator limit of the solutions to the time-dependent Schrödinger equation (cf. Eq. (A.1)) as

$$\psi^+_{i,k} = \lim_{t\to\infty} e^{iE_it}\langle k|e^{-iHt}|i\rangle. \tag{2.14}$$

This paper contains (in an often symbolic and certainly nonrigorous fashion) the essential elements of formal time-dependent scattering theory, which would later be further developed, for example, by Lippmann and Schwinger (1950), by Gell-Mann and Goldberger (1953) and by Brenig and Haag (1959). Heisenberg's ψ^+ and ψ^- are, respectively, the in and out scattering states of later theory.[10] Since another solution ψ^- to (2.13) can be obtained from (2.14) by letting $t\to-\infty$ and since H is hermitian ($H^\dagger = H$), it follows that

$$\psi^-_{i,k} = (\psi^+_{k,i})^* = \phi^+_{i,k} + S^*_{ki}\phi^-_{i,k}. \tag{2.15}$$

Heisenberg then observed that any linear combination of these ψ^- must also be a solution, in particular, $\sum_j S_{ij}\psi^-_{j,k}$,

whose outgoing part is just $S_{ik}\phi^{+}_{i,k}$. Because the boundary conditions make the solutions to Eq. (2.13) unique, this must be ψ^{+}, from which it follows that

$$SS^{\dagger}=1. \tag{2.16}$$

The essence of this formal proof is that hermicity ($H^{\dagger}=H$) implies unitarity for S. We shall later see that unitarity is also related to conservation of probability.

Heisenberg's development of the scattering matrix concept appears to have been independent of Wheeler's, although Heisenberg does indicate in a footnote (1943a, p. 533) that Wick had mentioned to him a paper by Breit on the use of the scattering matrix. In this footnote, Heisenberg refers to this paper only as 'Breit (Phys. Rev. 1941)'. This is almost certainly a reference to Breit's 'Scattering matrix of radioactive states' that appeared in the December 15, 1940, issue of *The Physical Review*. There Breit applied Wheeler's (1937b) scattering matrix formalism to unstable systems that have a resonance structure. (The 'Breit–Wigner' (1936) formula, Eq. (A.96), is used in this paper by Breit.) Heisenberg thanks Wick for the information but also states that he had not been able to see Breit's paper. (*The Physical Review* did not regularly reach Germany during the war.) He appears to have had no prior or direct knowledge of Wheeler's ideas on the scattering matrix. Dr. Helmut Rechenberg states[11] that, when he asked Heisenberg directly about this, Heisenberg said 'no' to knowing about Wheeler's S-matrix at the time he wrote his own paper on that subject. Professor John Wheeler is also of the opinion[12] that Heisenberg was unaware of Wheeler's work at the time.

But, it does seem implausible that Heisenberg had not even *seen* Wheeler's 1937 *Physical Review* paper on the S matrix. As indicated above, Heisenberg was certainly definite in claiming not to know about (or to have been aware of) Wheeler's work in 1941–42. However, it is not wholly implausible that Heisenberg might not have been aware of Wheeler's scattering matrix concept even if he had come across Wheeler's 1937 paper since the title and context of Wheeler's work indicated a concern with models for light nuclei. Furthermore, Heisenberg's main interests in 1937–40 appear to have been cosmic ray showers and the problems they presented for quantum field theory as then understood. Interestingly, though, Heisenberg (1938a) did publish an article in 1938 in which he referenced a paper by Feenberg in Volume 52 (1937) of *The Physical Review*. This is the same volume that contained Wheeler's S-matrix paper. So, it appears unlikely that

Heisenberg did not at least (literally) 'see' Wheeler's paper, although its content may not have attracted his attention.[13] In any event, Heisenberg certainly brought the concept of the S matrix to the attention of theoretical physicists. It has remained one of the central tools of modern physics.

A reader today looking at Heisenberg's original article (1943a) is amazed at what he 'sees' there (sometimes clearly only in retrospect, though.) There are the removal of the energy-momentum δ-function singularities to yield a true 'function' for S, the derivation of expressions for scattering and production cross sections in terms of the elements of S (cf. Eqs. (2.2) and (2.3) above), the proof of the unitarity of S, the Lorentz invariance of S, the separation of S as[14]

$$S=1+2iT \tag{2.17}$$

with T representing the effects produced by interactions, discussion of the fact that S respects the usual connection between spin and statistics, and expression of S as

$$S=e^{i\eta} \tag{2.18}$$

where η is an hermitian phase matrix. With the decomposition of Eq. (2.17), the unitarity condition of Eq. (2.16) becomes

$$-\frac{i}{2}(T-T^{\dagger})=TT^{\dagger}. \tag{2.19}$$

The diagonal element of this equation is just

$$\operatorname{Im} T_{ii}=\sum_{j}|T_{ij}|^{2}. \tag{2.20}$$

This relation was derived by Heisenberg and is essentially the optical theorem (Eq. (2.9)) relating the total cross section σ_t to the imaginary part of the forward elastic scattering amplitude. It is an elegant and quite 'modern' derivation. The form of Eq. (2.20) makes the meaning of unitarity and the origin of the optical theorem particularly transparent. The quantum-mechanical probability amplitude T_{ij} represents the effects of scattering from channel i to channel j (e.g., particles a and b in channel i going to or scattering into particles c and d in channel j: a + b → c + d; or, more compactly, $i \to j$). So, $|T_{ij}|^2$ is the probability ('rate', 'chance' or cross section) for the reaction $i \to j$. The expression $\sum_j |T_{ij}|^2$, summed over *all possible* outcomes (or final states) j, is the total reaction rate (or cross section). Equation (2.20) relates this total cross section to the imaginary part of T_{ii}, the elastic (e.g., a + b → a + b)

scattering amplitude. That is the optical theorem. Furthermore, Eq. (2.20) shows that a particular amplitude, here T_{ii}, is related to *all* other amplitudes T_{ij}, representing channels (j) that can be connected to channel i. This coupling, or 'entanglement', of reactions with each other will be an important feature of S-matrix dynamics.

The hermitian phase matrix η was the primary quantity to be determined in the theory (Cushing, 1982; Grythe, 1982), essentially by guessing at that time. Its determination would replace equations of motion, such as the Schrödinger equation or the Hamiltonian formalism of quantum field theory. Heisenberg also indicated (1943a) how the matrix η could actually be calculated in those cases where a Hamiltonian H was known, as in quantum mechanics and field theory. In his second paper (1943b) he computed S for particular models of η (i.e., essentially for certain enlightened guesses for the form of η).

An important point for the subsequent development of the S-matrix program is the following. Heisenberg (1944, p. 94, footnote) thanks Kramers for the suggestion that S be considered as an analytic function of the energy variable. (Heisenberg (1946, p. 612) later acknowledged discussing this with Kramers in Leiden in 1942. Grythe (1982) also references some correspondence between Heisenberg and Kramers on this point.) This analyticity suggestion was related to earlier unpublished work by Kramers and Wouthuysen. Here the stationary (or bound) states correspond to the zeros of the S matrix on the (negative) imaginary k axis, where the energy is $E=k^2$. (We shall discuss below poles versus zeros of the S matrix and their connection with bound states.) Wouthuysen, whose association with Kramers began as a graduate student in 1937, recalls:[15]

> In 1940, if I remember well, he [Kramers] proposed to me a subject of research: the study of nuclear resonances, in relation with papers by Breit and Wigner, Kalckar, Oppenheimer and Serber, Kapur and Peierls. A later paper by Seigert (1939) inspired me very much: he pointed out the relations between 'radioactive states' (in the sense of Gamow) and resonance scattering. The Gamow states appeared as singularities for complex energy of the scattering amplitude as well. I proceeded to show the analyticity of the Schrödinger scattering amplitude and in this way found the bound states as singularities with negative energy. These properties I illustrated by simple examples, like square well potentials with square potential barriers. After a seminar I gave on the subject, Kramers' verdict was 'it is interesting and new'. Shortly afterwards, summer 1942, I had to abandon my studies, due to the war circumstances. Later, in 1943, I learned indirectly from Kramers 'that he made good use of my work during a visit of Heisenberg in Leiden'. Clearly (now by hindsight!) Kramers had seen much wider implications of my work than I myself had dreamt of.

Heisenberg (1944) demonstrated this connection explicitly in specific model calculations. A simple example of this correspondence can be seen in the S matrix $S(k)$ for the scattering by a square well (Eqs. (A.39b), (A.42) and the comment following (A.42)). This important reflection by Kramers implied that the scattering matrix not only determined those observables related to cross sections but could also fix the bound-state energies of the system. However, the difficulty was that, since no Hamiltonian was assumed, there were no correspondence rules with classical theory to guide one in constructing (or guessing) S (or η).

2.3 Heisenberg's subsequent role in the program

Heisenberg did not remain a major player in S-matrix theory beyond the late 1940s. The program influenced theoretical physics through the work of others. So, before we detail the history of these developments, let us indicate the path taken by Heisenberg's subsequent research. In 1946 he (Heisenberg, 1946) summarized his S-matrix program and promised a fourth paper in this series. Interestingly enough, this paper appears never to have been published.[16] That manuscript discusses the many-body problem in the S-matrix framework. As we indicate below, the many-body problem posed difficulties for Heisenberg's program. Heisenberg himself soon lost interest in the S-matrix program and turned to the theory of turbulence (1947–48) and then to nonlinear field theory (Heisenberg 1949a, 1950a, 1950b, 1951, 1952, 1953, 1954, 1957, 1958; Heisenberg, Kortel, and Mitter, 1955; Ascoli and Heisenberg, 1957) in which there would be just *one* fundamental field that would underlie all of particle physics. That Heisenberg's interests had shifted back to quantum field theory by the late 1940s is not surprising since the renormalization program had shown how to cope with divergences. One of Heisenberg's earlier motivations for studying a *nonlinear* field theory was that he believed such a theory could account for the 'explosive' cosmic ray events we mentioned in the last chapter (see also Cassidy, 1981, pp. 19–21). That there was a connection between the turbulence and nonlinear quantum field research is already made plausible even by the title of Heisenberg's (1952) paper ('Meson production as a shock wave problem') in which he used previously proposed nonlinear equations to discuss meson production. His view (as summarized much later, Heisenberg, 1966, 1967) seems to have been

that nonlinear problems are overwhelmingly more common in nature than are linear ones. Hence, a nonlinear equation was much more likely to describe a fundamental law of nature than would be a linear one.

Neither Heisenberg's S-matrix program (as a serious, independent program), nor his theory of superconductivity (Heisenberg, 1949c), nor his nonlinear spinor field theory was well received by the majority of theoretical physicists. His attempts at foundationally new theories no longer commanded the attention they once had. For example, Pauli was at first quite enthusiastic about Heisenberg's S-matrix program (as evidenced even in the present chapter by the number of people he described the program to; cf., also, Jost, 1984). But, Pauli was never convinced by the theory and eventually thought the program rather empty. Still, while Pauli remained at Princeton during the war, '... he engaged several of his collaborators at Princeton to explore the S-matrix' (Grythe, 1982, p. 200). Similarly, Pauli was initially interested in Heisenberg's nonlinear spinor theory but soon became an almost derisive critic of it. For example, in his comments on Heisenberg's (1958) paper given at the 1958 CERN Conference, Pauli was very negative (Ferretti, 1958, pp. 122–6).

> I reached the conclusion that [Heisenberg's papers on the spinor model] are mathematically objectionable.
>
> I disbelieve all the more in the possible excuses for such a contradiction.
>
> This I discussed already in April and I wonder that you again repeat it all.
>
> Well, I think that it is superfluous.

Møller also eventually became disenchanted with Heisenberg's S-matrix program (Grythe, 1982, p. 201). In this same vein, Jost states:[17]

> Heisenberg's reputation was of course enormous. He was trusted with an unfailing intuition and with the ability to be able to do almost anything ... His fame began to fade after 1945.
>
> ... Heisenberg's futile attempt at a theory of superconductivity ... dispelled once and for all the spell which he exerted on his friends (and even his enemies).
>
> I was present at H[eisenberg]'s seminar in Princeton (fall 1950) and witness to the violent attack against Heisenberg. Oppenheimer had warned ... [us] ... before the seminar, to treat the speaker kindly. The lion had lost his claws.

Since we do not wish to pursue Heisenberg's nonlinear field theory here, let us indicate briefly his final position in which he had abandoned the S matrix as *the* fundamental entity in a new theory. In a review article on the quantum theory of fields, he stated (1957, p. 270):

> To avoid ... fundamental difficulties ... the efforts of many physicists have in recent years been concentrated on the S matrix. The S matrix is the quantity immediately given by the experiments.
>
> The S-matrix formalism does not by itself guarantee the requirements of relativistic causality.
>
> It is perhaps not exaggerated to say that the study of the S matrix is a very useful method for deriving relevant results for collision processes by going around the fundamental problems. But these problems must be solved some day and one will then have to look for a mathematical formalism that allows one to calculate the masses of the particles and the S matrix at the same time. The S matrix is an important but very complicated mathematical quantity that should be derived from the fundamental field equations; but it can scarcely serve for formulating these equations.

Some small comment is appropriate on the first two paragraphs of this quotation. First, it would be an overstatement to claim that *all* S-matrix elements are given directly by experiment. Some, such as those with two bodies in the initial state, have a reasonably direct relation to experimentally measurable quantities. However, there are many S-matrix elements that one would be extremely hard pressed to find any plausible way to measure. Second, Fierz (1950) had criticized Heisenberg's (1950a, 1950b) attempt to construct a 'convergent field theory' expression for S and had shown it to be not causal. This emphasized the difficulty of implementing causality without having a new Hamiltonian theory.

At the Solvay Conference (Stoops, 1961, pp. 174–5) Heisenberg also reflected upon his reasons for abandoning the S-matrix program:

> When I had worked on the S-matrix for a while in the years 1943 to 1948 I came away from the attempt of construction of a pure S-matrix theory for the following reason: when one constructs a unitary S-matrix from simple assumptions (like a hermitian η-matrix by assuming $S = e^{i\eta}$), such S-matrices always become non analytical at places where they ought to be analytical. But I found it very difficult to construct analytical S-matrices. The only simple way of getting (or guessing) the correct analytical behaviour seemed to be a deduction

> from a Hamiltonian in the old-fashioned manner. ... My criticism [of the original program] comes only from the practical point of view. I cannot see how one could overcome the enormous complications of such a program.

Here, as later, complexity would be a recurrent theme of and problem for the *S*-matrix program.

2.4 Work on the program just after WWII

Let us now return to the historical sequence of developments. Some sense of the status of Heisenberg's early *S*-matrix program can be gotten from Kramers' remarks for a symposium at Utrecht in the spring of 1944 (Kramers, 1944, 139–40; 1956, p. 838):

> III. Heisenberg's recent investigations concerning the possibility of a relativistic description of the interaction that is not based on the use of a Hamiltonian with interaction terms in a Schrödinger equation. Heisenberg considers only free particles and introduces a formalism ('scattering matrix') by means of which the result of a short interaction (scattering) between these particles can be described. Formerly the scattering matrix could be derived from the Hamiltonian, but now we are to consider the scattering matrix as fundamental. We do not care whether a Schrödinger equation for particles in interaction exists; we do care which correspondence requirements exist and how the scattering matrix can obey them. It is interesting that the scattering matrix is also able in principle to answer the question in which stationary states the particles considered can be bound together. These are related to the existence and the position of zeros and poles of the eigenvalues of the scattering matrix, considered as a complex function of its arguments. Heisenberg could already give a (very simple) model of a two-particle system, in which a perfectly sharply relativistically determined stationary state occurs, while there are no divergence difficulties whatsoever.
>
> However promising, this is still only a beginning, and in particular with regard to a correct description of the electromagnetic fields of photons I expect difficulties, which the investigations in this direction will have to overcome. Fortunately, Heisenberg's program is still open in several respects, and one may perhaps expect a great deal from a fortunate combination with further ideas.

In two lengthy papers and in a briefer note, Møller (1945, 1946a, 1946b) studied the properties of Heisenberg's *S* matrix and fleshed out many of Heisenberg's original arguments. (Grythe (1982) has some interesting

comments about the correspondence between Heisenberg and Møller during the period when these papers were written.) In the first paper, he considered the quantum-mechanical dynamics of an arbitrary number of identical interacting particles, gave a careful treatment of the unitarity of *S*, showed that the results of scattering processes (at real positive values of $E=k^2$) are independent of the bound-state energies (1945, p. 18), pointed out the completeness condition for the wave matrix states, proved the Lorentz invariance of *S* from cross-section invariants, and discussed the 'collision constants' (or conserved quantities) arising from the invariance of *S* under sets of transformations (such as the Lorentz transformations). In the second paper, Møller restricted himself mainly to a two-body system and developed extensively the analytic properties of the wave matrix to establish the connection between the zeros of *S* and the bound-state energies. There Møller proved (for a two-body system) that at a zero of *S* in the lower half *k* plane at $k_o=i\kappa$, $\kappa>0$

$$i(-1)^l \left.\frac{dS_l}{dk}\right|_{k_o} > 0 \tag{2.21}$$

for any given partial wave *l* (1946a, p. 29) and showed how resonance energies and half-lives were determined by the *poles* (or singularities) of the analytic *S* matrix in the lower half energy plane at $E-i\Gamma/2$, where Γ is the reciprocal of the lifetime of the state (cf. Eq. (A.101)). He established that for a given matrix *S* there might exist either no Hamiltonian *H* or many Hamiltonians that would yield the same *S*. In other words, even when the usual equations of motion exist, there would not necessarily be a unique correspondence between *S* and *H*. The square-well and Coulomb-potential *S* matrices were given explicitly. Møller was fairly optimistic about the general applicability of results that had been proved in special cases (1946a, p. 45):

> The results obtained for two particle systems in this paper may be supposed to hold also in the general case of a many particle system with possibilities for creation and annihilation processes, the only difference probably being that the number of collision constants ... is then larger than in the case of a simple two particle system.

In 1946 Heisenberg (1946) summarized the status of the *S*-matrix program as set forth in his three papers (1943a, 1943b, 1944) and in Møller's two papers (1945, 1946a).

There Heisenberg used the completeness of the $l=0$ wave functions

$$\int dk\, \psi^{+*}(k,r)\psi^{+}(k,r') + \sum_n \psi_n^*(r)\psi_n(r') = \pi\delta(r-r') \tag{2.22}$$

to express the constants c_n for the bound-state wave functions ($k_n = -i\kappa_n, \kappa_n > 0$)

$$\psi_n(r) \xrightarrow[r\to\infty]{} \frac{c_n}{\sqrt{2}} e^{-ik_n r} \tag{2.23}$$

in terms of the contour integral of the S matrix

$$|c_n|^2 = \oint dk\, S(k) \tag{2.24}$$

where the integral encloses only the pole at $k_n = i\kappa_n$. This was another example of using Hamiltonian-independent, general principles to obtain specific dynamical results about bound states from the S matrix. The point is that the completeness condition of Eq. (2.22) is just a statement that the states $\{\psi(k)\}$ can be used as a basis for expanding 'any' function. It is essentially Fourier's theorem that allows the expansion of an arbitrary function, $f(x)$, in terms of sines and cosines or of exponentials, e^{ikx}. It is a property the ψ's would be expected to have.

Heisenberg (1946, p. 613) acknowledged that Møller had essentially this result in the summer of 1944. Equation (2.24) can also be derived directly by an older method due to Kramers (1938; Ma, 1947). In this paper (1946, p. 609) he also promised the fourth paper (which we referred to above) on many-particle systems. This and his Cambridge lecture at the end of 1947 (Heisenberg, 1949b) appear to have been about the last of Heisenberg's publications on the S matrix. Even at this time, though, Møller (1947, p. 195) remained quite pessimistic about the prospects for a successful conventional quantum field theory.

> [I]t seems that divergencies are intrinsic difficulties of all relativistic quantum field theories of the Hamiltonian form and that the frame offered by the Hamiltonian scheme of quantum mechanics is too narrow.
>
> [I]t even seems doubtful that in any strictly relativistic theory of the future a Hamiltonian and a Schrödinger equation will exist at all in general.

Now that we have indicated what the central concepts of the Heisenberg S-matrix program were, let us consider the impact these ideas had. Since our chief interest in Heisenberg's early S-matrix theory lies in the proposed responses to a series of questions it provoked, we begin by focusing on the work of two theorists, Walter Heitler and Ernst C. G. Stückelberg.

As early as 1941 Heitler (1941) proposed a quantum theory of radiation damping (which refers to the effects caused by the radiation or emission of quanta by a particle) to do for the strong-interaction mesons what classical and quantum theory had done for light. By analogy, the mesons were seen as the quanta of the nuclear force just as the photon was the quantum of the electromagnetic field. The time-dependent theoretical treatment was very much in the spirit of that for the interaction of a quantized atomic system with the electromagnetic field as done by Weisskopf and Wigner (1930) and by Weisskopf (1931). (See Eqs. (A.86)–(A.95) and the discussion following them.) Heitler and Peng (1942) then proposed a scheme for omitting all the divergent terms in the perturbation expansion but still retaining effects of radiation damping. They (1943) used this procedure to make approximate calculations of the matrix elements for meson production in cosmic rays. In the late 1940s Pauli visited the United Kingdom and spoke on the difficulties in quantum field theories (Pauli, 1947) at the 1947 Cambridge Conference. This review included a summary of Heisenberg's S-matrix theory and of Heitler's *ad hoc* recipe for keeping only the finite terms in the perturbation expansion for S. Prior to Pauli's discussions with Heitler about Heisenberg's recent work, the S-matrix program seems to have been unknown at the Dublin Institute for Advanced Studies where Heitler was. Heitler and Hu (1947, p. 124) thank Pauli for having given them an outline of Heisenberg's work. After this exchange, Heitler and Hu (1947) used the approximate matrix elements of Heitler and Peng (1943) to calculate the S matrix and hence the bound states or the (isobar) particle spectrum for the meson–nucleon system. As Heitler and Hu (1947, p. 140) acknowledge, Pauli pointed out that an analytic continuation of an *approximate* expression for S from real k to imaginary values of k need not give reliable values for the zeros (and hence for the isobar masses). However, Heitler and Hu had indicated how bound-state masses could (in principle) be calculated even when a complete Hamiltonian theory is not known.

We can summarize Heitler's procedure as follows (Heitler, 1947; Wentzel, 1947). If the Hamiltonian H is written as $H = H_0 + H'$, where H' (representing the effects of radiation)

causes a perturbation on the eigenstates of H_0, then a reduced matrix element of the operator H' is defined as[18]

$$\langle i|\overline{H'}|j\rangle = \frac{\langle i|H'|j\rangle}{E_j - E_i}. \tag{2.25}$$

The matrix K (sometimes termed the reaction matrix) is defined as

$$K = \sum_{n=0}^{\infty} (\overline{H'})^n H'. \tag{2.26}$$

The equation for the T matrix of Eq. (2.17) takes the form

$$K(1 + iT) = T \tag{2.27}$$

whose (formal) solution is

$$T = (1 - iK)^{-1} K. \tag{2.28}$$

Heisenberg's S matrix (cf. Eq. (2.17)) can be expressed as

$$S = (1 - iK)^{-1}(1 + iK). \tag{2.29}$$

If the interaction Hamiltonian H' is characterized by a coupling constant (or coupling strength) g, then Eq. (2.26) can be seen as an expansion in powers of g. The first non-vanishing term is finite, whereas higher-order ones diverge (i.e., are infinite). Heitler's *ad hoc* recipe consisted in keeping only this lowest-order, finite term for K and dropping (or 'subtracting') all the rest. He (1941) expressed the hope that these equations, which still included radiation damping, could be *exact*, although he later (1947, p. 189) relegated this to a (perhaps interim) working hypothesis:

> The attempt to be described below aims not at a final solution to this difficult problem, but rather at obtaining a preliminary working hypothesis from which can be derived physical results which shall be reasonable and, as far as possible, in agreement with the experiments.

Heitler had generated a relativistically invariant procedure for producing a finite K and hence a finite S matrix whose zeros should give the bound states of the system. Still, this was not a procedure based on some more general principle that eliminated the divergences, but rather a pragmatic rule for simply throwing away all the troublesome terms.

Stückelberg[19] had worked on a consistent classical model for a point electron (1938, 1939, 1941) and later (1942) extended this to a quantized theory. He applied (1944, 1945) Heisenberg's S-matrix formalism to these problems and proposed a series of correspondence rules (i.e., previously-known results to which the quantum expressions must pass over in suitable limits) to guide one in writing down Heisenberg's η matrix, or Heitler's K matrix. These matrices are related as

$$\tan\left(\frac{\eta}{2}\right) = K. \tag{2.30}$$

He also (1946) imposed the constraint that, in a scattering process, a particle cannot appear in the final asymptotic state before the incident particles collide. These correspondence and 'causality' constraints eliminated some of the arbitrariness and the divergences in K (or in η), but still did not determine uniquely the structure of the higher-order terms. This program was an attempt, however, to eliminate the divergences on the basis of independent, generally accepted principles. Stückelberg and his co-workers (Rivier and Stückelberg, 1948; Stückelberg and Rivier, 1950a, 1950b; Stückelberg and Green, 1951) continued in their efforts to develop a quantum field theory free of the divergences present in the Schwinger–Tomonaga–Feynman–Dyson theory, but without ultimate success. Wentzel (1947) gives a nice review of the status of strong-interaction field theory just after the Second World War and, in particular, of the Heisenberg–Stückelberg–Heitler theory. Wentzel's evaluation of Heisenberg's program at that date is that this scheme '... is very incomplete; it is like an empty frame for a picture yet to be painted' (p. 15). He saw Heitler's and Stückelberg's work as attempts to fill this frame.

However, before we continue the historical development of work directly related to Heisenberg's S-matrix program, let us make a few observations about how the S matrix had entered theoretical physics as a practical and important calculational tool by the late 1940s. Most important for theoretical elementary-particle physics was Schwinger's (1948) introduction of the unitary operator $U(t, t_0)$ that gives the time evolution of the state vector $\Psi(t_0)$ at some initial time t_0 to the $\Psi(t)$ at a later time as[20]

$$\Psi(t) = U(t, t_0)\Psi(t_0) \tag{2.31}$$

so that the Lorentz-invariant collision operator S connecting the initial and final states is[21]

$$S = U(\infty, -\infty). \tag{2.32}$$

Schwinger also expressed this S in terms of the hermitian reaction operator K by Eq. (2.29). Dyson (1949a, 1949b) made explicit the relation of Feynman's (1949a, 1949b) method for calculating the

elements of Heisenberg's S matrix and also recalled (1949b, p. 1736) that Stückelberg had anticipated several of Feynman's results. In 1946–47 Dyson had studied under Nicholas Kemmer at Cambridge. Kemmer had Dyson read Heisenberg's S-matrix papers as well as Gregor Wentzel's (1943) *Einführung in die Quantentheorie der Wellenfelder* (published in Vienna during the War). When Dyson came to America in 1947, he was well acquainted both with S-matrix theory and with quantum field theory. Dyson recalls[22]:

> I was well prepared by Kemmer to put this knowledge [of the S matrix and of quantum field theory] to use in the reconstruction of quantum electrodynamics. I well remember the joy of recognition when I suddenly realized that Feynman's rules of calculation were just the practical fulfilment of Heisenberg's S-matrix program. This was in the Fall of 1948.

The fact that Dyson saw Feynman's theory as the *fulfillment* of Heisenberg's program is very clear from a letter to J. R. Oppenheimer that Dyson wrote in 1948[23]:

> I believe it to be probable that the Feynman theory will provide a complete fulfillment of Heisenberg's S-matrix program. The Feynman theory is essentially nothing more than a method of calculating the S-matrix for any physical system from the usual equations of electrodynamics. It appears as an experimental fact (not yet known for certain) that the S-matrix so calculated is always finite; the divergencies only appear in the part of the theory which Heisenberg would in any case reject as meaningless. This seems to me a strong indication that Heisenberg is really right, that the localisation of physical processes is the only cause of inconsistency in present physics, and that so long as all experiments are interpreted by means of the S-matrix the theory is correct.
>
> The Feynman theory exceeds the original Heisenberg program in that it does not involve any new arbitrary hypothesis such as a fundamental length.

Thus, Heisenberg's S matrix as a calculational tool, even if not as the central entity in an independent theory, was very much in the 'consciousness' of theoretical physics before 1950. This renormalization program was so computationally successful that the preponderant majority of theoretical high-energy physicists worked within its tradition beginning in the late 1940s. This success of renormalized quantum electrodynamics undercut, as pointed out earlier, Heisenberg's original motivation (i.e., the inability of quantum field theory to produce finite, unique results) for his S-matrix program. As we shall see later, in the late

1950s and early 1960s, the inability of even renormalized quantum field theory to calculate results for the strong interactions once more led to a new S-matrix program. And, again, once quantum field theory (in the form of gauge field theories) became calculationally adequate, relatively few theorists continued to work on the S-matrix program. In both instances, field theories won the contest because they could be cast into a form that allowed reliable perturbation calculations to be made (since no exact, closed-form solutions are known for realistic situations), while the S-matrix program could not be cast into a form useful for extensive (numerical) calculations. However, once again let us tie off this thread in our story and return to developments more closely related to Heisenberg's program.

A straightforward application of Heisenberg's S-matrix zero $\Leftrightarrow$ bound-state energy rule soon encountered difficulties in the secure testing ground of nonrelativistic quantum mechanics. Ma (1946, 1947) showed by explicit calculation for an exponential potential in Schrödinger theory that there were 'false' or 'redundant' zeros of the S matrix which did not correspond to actual bound-state energies.[24] This discovery, Møller said, made him doubtful about the S-matrix theory and he left the new theory in the summer of 1946 (Grythe, 1982, p. 201). Ter Haar (1946) attempted to use Møller's inequality (2.21) to rule out these false zeros, but Jost (1946, 1947) criticized that argument and concluded that all false zeros could not necessarily be ruled out with that prescription.[25] Jost (1947) considered s-wave scattering and introduced the irregular solution $f(k,r)$ to the radial Schrödinger equation (Eq. (A.19))[26]

$$f(k,r) \underset{r\to\infty}{\longrightarrow} e^{-ikr} \tag{2.33}$$

and an $f(k)$ (today known as 'Jost functions') as[27]

$$f(k)=f(k,r=0). \tag{2.34}$$

The solution regular at the origin is

$$\phi k,r)=\frac{1}{2ik}[f(k)f(-k,r)-f(-k)f(k,r)]. \tag{2.35}$$

$$\underset{r\to\infty}{\longrightarrow} \frac{1}{2ik}[f(k)e^{ikr}-f(-k)e^{-ikr}]. \tag{2.36}$$

From Eq. (A.33) we see that

$$r\psi^{+}(k,r)=\frac{\phi(k,r)}{f(-k)} \tag{2.37}$$

and that

$$S(k)=\frac{f(k)}{f(-k)}=e^{2i\delta(k)}. \tag{2.38}$$

Here $\delta(k)$ is the phase shift (for s-wave scattering). Equation (2.36) makes it clear that $\phi(k,r)$ will be a bound-state wave function for $k=-i\kappa(\kappa>0)$ when $f(-i\kappa)=0$; that is, when $S(k)$ vanishes. (An explicit expression for $f(k)$ for the square well can be read off from Eq. (A.39b).) Jost discussed the analytic properties of these $f(k)$ and laid the foundation for a rigorous treatment of the analytic properties of the scattering amplitude. Unless one confines the discussion to strictly finite-range potentials, there can be redundant zeros in the S matrix found in potential theory. Meixner (1948) also studied the analytic properties of $S(k)$ in Schrödinger theory. Touschek (1948) and Wildermuth (1949a, 1949b) examined many-body effects within the framework of potential theory and Wildermuth (1949c) attempted unsuccessfully (van Kampen, 1951) to resolve the redundant zero problem. These redundant zeros remained a difficulty for the S-matrix program.[28]

Fröberg (1947, 1948a, 1948b), Hylleraas (1948) and Bargmann (1949a, 1949b), all motivated by Heisenberg's program, considered the possible equivalence, in nonrelativistic Schrödinger theory, of the S matrix (or of phase shift $\delta(k)$) and the potential $V(r)$. That is, the question was whether the $V(r)$ could be reconstructed from the $\delta(k)$. In fact, Fröberg (1948a) acknowledged Heisenberg's S-matrix program and thanked Pauli for suggesting the inversion problem to him in Zürich. Hylleraas (1948) also pointed out the importance of the inversion problem for Heisenberg's S-matrix program. Bargmann (1949a, 1949b) cited Møller's (1946a) proof that many Hamiltonians could correspond to one S matrix (cf., comment following out Eq. (2.21) above). Fröberg gave an incomplete, *formal* solution of the inversion problem. Hylleraas somewhat cautiously commented upon and extended Fröberg's work. Then Bargmann gave explicit examples of one phase shift $\delta(k)$ corresponding to several potentials $V(r)$ in Schrödinger theory and thus showed that Fröberg's and Hylleraas' inversion procedure could not be valid in all cases. These results led to Levinson's (1949a, 1949b) interest in the inversion problem, as indicated by

Levinson's (1949a) reference to Fröberg's work. He found conditions on the asymptotic value of the phase shift $(\delta(\infty)-\delta(0))$ which would guarantee that only *one* potential $V(r)$ could have produced this $\delta(k)$. Holmberg (1952) pinpointed the flaw in Fröberg's and Hylleraas' uniqueness agrument for the inversion problem. Gel'fand and Levitan (1951a, 1951b) and Jost and Kohn (1952a, 1925b, 1953) discovered a constructive procedure for finding $V(r)$, given $\delta(k)$ as well as the energies and normalization factors for the bound-state wave functions. Even today, after over three decades of work on the question, a complete solution to the inversion problem is not a simple matter to state. A simplified (and hence incomplete) answer is the following. If $\delta(k)$ is given for all real positive k for $l=0$ and if there are no bound states, then (for suitable restrictions on the potential) there is a unique $V(r)$ for that problem. For each bound state (where the number n of bound states is given by $\delta(0)-\delta(\infty)=n\pi$), there is an arbitrary parameter in the potential (and, hence, no uniqueness). (See Newton, 1966, Chapter 20, and Cushing, 1986a, note 45, for more details.) This general question of the determination of a potential from the experimental phase shifts (which can be gotten rather directly from the observed scattering cross sections) was an important one for the theoretical nuclear physics of the time.

Jost himself characterizes the relevance of the criticism of Heisenberg's S-matrix program to theoretical physics as follows.[29]

> Summing up this special line of development it seems fair to state that the *criticism* of Heisenberg's S-matrix program has stimulated a qualitatively new kind of theoretical research, in which non relativistic wave mechanics serves, not as an analytic tool to calculate some experimental cross-section, but rather as an experimental playground for the discovery of general relationships, which might also be useful elsewhere.

This 'experimental playground' would remain an important tool for research in the S-matrix program, even well into the 1960s.

2.5 The S matrix and nuclear theory

Contact with a long-established line of work in theoretical nuclear physics was made by Hu (1948a). As background, let us say a few words about the relevant nuclear theory. (We return to the compound nucleus model in Chapter 9.) Breit and Wigner (1936, p. 519) had developed (in

analogy with the resonance scattering of light; cf. Eqs. (A.86)–(A.96)) a theory of scattering near a (complex) resonance energy E_c:

> ... [T]heories of ... cross sections ... for the capture of slow neutrons ... explain the anomalously large capture cross sections as a sort of resonance of the s states of the incident particles.

Breit (1940a) summarized this as follows:

> Many nuclear reactions show excitation curves with pronounced peaks which suggest that there is a resonance of the nuclear system to certain energies. (p. 506).
>
> ... [T]he optical case has been cleared up by Weisskopf and Wigner [1930]. ... The band of photons emitted in such a jump is found to have an intensity distribution of the resonance type
>
> $$\frac{\text{const.}}{(E-E_{ji})^2+\Gamma^2/4}$$
>
> with a peak at the emission center E_{ji}. (p. 507)

Breit (1940b) used this approximation to calculate Wheeler's S matrix for the decay of unstable states as

$$S_{\alpha\beta}\approx\delta_{\alpha\beta}+\frac{c_{\alpha\beta}}{E-E_c} \tag{2.39}$$

where the $c_{\alpha\beta}$ are constants. This particular form was based on simplicity (Breit, 1940b, p. 1070):

> The simplest possibility for [poles of the scattering matrix a_{ij}] will be considered
>
> $$a_{ij}\approx\frac{c_{ij}}{E-E_c}+d_{ij}.$$
>
> Here the c_{ij}, d_{ij} are constants and the equation is meant to be only an approximation in the neighbourhood of the complex eigenvalue [E_c].

Very interestingly, Siegert (1939) seems to have been the first to consider the S matrix as an analytic function of k to derive a Breit–Wigner type formula:

> We investigate the singularities of the cross section which occur at certain complex values of the energy. Those singularities which lie near enough to the real axis, cause a sharp resonance maximum on the real axis and we can replace the cross section there by its singular part added to a smooth function of the energy. (p. 750)

Of course, one can 'see' these results today in Siegert's paper, but his motivations and notations were quite different.

An important insight into the origin and generality of the Breit–Wigner resonance formula was obtained by Wigner and his co-workers. We outline below the *formal* derivation of a Breit–Wigner type of formula. Later, we shall discuss a physical or more intuitive explanation of such resonance formulas.

In a series of papers, Wigner (1946a, 1946b) and Wigner and Eisenbud (1947) introduced the derivative matrix $R(E)$ and based a multichannel nuclear scattering theory upon it.[30] If a is the range of the potential, then for a simple single-channel case we have (cf. Eq. (A.37))

$$R(E)=[r\psi_E(r)/d(r\psi_E)/\mathrm{d}r]_{r=a}=\frac{\tan(ka+\delta)}{k} \tag{2.40}$$

where $\psi_E(r)$ is the wave function for the problem. If we set $\psi_E(r)=e^{i\delta}\cos(ka+\delta)\phi_E(r)$, then *outside* the nuclear surface ($r=a$) we can write $\phi_E(r)$ as

$$\phi_E(r)=\frac{1}{r}\left\{\frac{1}{k}\sin[k(r-a)]+R\cos[k(r-a)]\right\} \tag{2.41}$$

with $R(E)=[r\phi_E]|_{r=a}$. By using the complete set of real basis states $\phi_\lambda(r)$ defined by

$$H\phi_\lambda=E_\lambda\phi_\lambda, \tag{2.42a}$$

$$\left.\frac{d(r\phi_\lambda)}{\mathrm{d}r}\right|_{r=a}=0, \tag{2.42b}$$

$$\int_{\substack{\text{nuclear}\\\text{surface}}}\phi_\lambda^2\,\mathrm{d}V=1, \tag{2.42c}$$

one can prove (Wigner and Eisenbud, 1947) that $R(E)$ has the *exact* expansion

$$R(E)=\sum_\lambda\frac{\gamma_\lambda\gamma_\lambda}{(E_\lambda-E)} \tag{2.43}$$

where the γ_λ are independent of E. For example, in the case of the square well (Eq. (A.38)) one can show directly that

$$R(E)\equiv\frac{\tan(Ka)}{K}=\sum_\lambda\frac{2}{a}\frac{1}{(E_\lambda-E)}. \tag{2.44}$$

In the multichannel case Eq. (2.41) becomes

$$\phi_{\alpha\beta}(r_\beta)=\frac{1}{r_\beta}\left\{\frac{\delta_{\alpha\beta}}{k_\alpha}\sin[k_\alpha(r_\alpha-a_\alpha)]+R_{\alpha\beta}\cos[k_\beta(r_\beta-a_\beta)]\right\} \quad (2.45)$$

and Eq. (2.43) generalizes to

$$R_{\alpha\beta}(E)=\sum_\lambda \frac{\gamma_{\lambda_\alpha}\gamma_{\lambda_\beta}}{(E_\lambda-E)} \quad (2.46)$$

where the γ_{λ_α} are again independent of E. Notice that Eq. (2.46) implies that

$$\frac{dR_{\alpha\alpha}}{dE}>0, \quad (2.47)$$

whenever this quantity is defined. The S matrix and this R matrix are related as (cf. Eq. (A.39a))[31] (Wigner and Eisenbud, 1947)

$$S_{\alpha\beta}=e^{-ik_\alpha a_\alpha}\left[\frac{1+i\sqrt{k_\alpha}R_{\alpha\beta}\sqrt{k_\beta}}{1-i\sqrt{k_\alpha}R_{\alpha\beta}\sqrt{k_\beta}}\right]e^{-ik_\beta a_\beta}. \quad (2.48)$$

In the neighborhood of a (real) resonance energy E_λ, $R_{\alpha\beta}$ may be approximated by one term in the sum of Eq. (2.46). Equation (2.48) then gives the Breit–Wigner one-level formula

$$S_{\alpha\beta}\approx e^{-ik_\alpha a_\alpha}\left[\delta_{\alpha\beta}-\frac{2ik_\alpha^{\frac{1}{2}}\gamma_{\lambda_\alpha}k_\beta^{\frac{1}{2}}\gamma_{\lambda_\beta}}{(E-E_\lambda)+i\sum_\nu k_\nu\gamma_\nu^2}\right]e^{-ik_\beta a_\beta}, \quad (2.49)$$

which has the same general *form* as Eq. (2.39) for $\alpha\neq\beta$.

One should not let the formalism obscure the central structure of the argument. It is essentially the completeness of the eigenfunctions ϕ_λ (Eqs. (2.42)) that allows one to obtain the 'resonance' form (Eqs. (2.44) and (2.46)) for $R(E)$ and that, in turn, leads to the Breit–Wigner form (Eqs. (2.48) and (2.49)). The interior of the nucleus remains effectively an unknown black box, since $R(E)$ is defined in terms of quantities at the nuclear surface ($r=a$) *only*.

Motivated by Heisenberg's S-matrix program, Hu (1948a) used directly the analytic properties of S to derive (in the context of potential theory) the Breit–Wigner form in the neighborhood of a complex pole of the S matrix near the positive energy axis. He also clarified the common

distinction that had been made in the literature between the 'potential scattering' that produced the smooth background against which the 'resonance scattering' of Eq. (2.39) stood out (cf. Eqs. (A.112)–(A.114) and the accompanying discussion). Hu showed that the 'potential' and 'resonance' terms both had the same source, the difference in their behavior depending upon the distance of the complex pole from the positive energy axis. He also made the useful observation (which would become the convention eventually adopted) that, since

$$S(k)S^*(k^*)=1, \quad (2.50)$$

a value of k_n on the negative imaginary axis such that $S(k_n)=0$ corresponds to a value $k_n^*=-k_n$ such that $S^*(k_n^*)=\infty$, or $S(-k_n)=\infty$. That is, the *poles* of the S matrix on the positive imaginary axis locate the bound-state energies. From now on we follow this convention illustrated in Figure 2.1. The upper half k-plane maps onto the entire $E=k^2$ physical sheet and the lower half k-plane onto the second (Riemann) sheet in the energy plane.

Hu (1948b) and Eden (1948, 1949a, 1949b) began to study the complicated problem of the analytic continuation of the S matrix when there were production thresholds present. As an aside, let us stress that it is important to appreciate, both here and in subsequent developments discussed throughout the text, that we use the term 'analytic' to include functions that have poles, branch points, and even essential singularities. The key property of such analytic functions is that once they are known exactly in any region however small, then they are (in principle) determined everywhere in their domain of analyticity. We have already

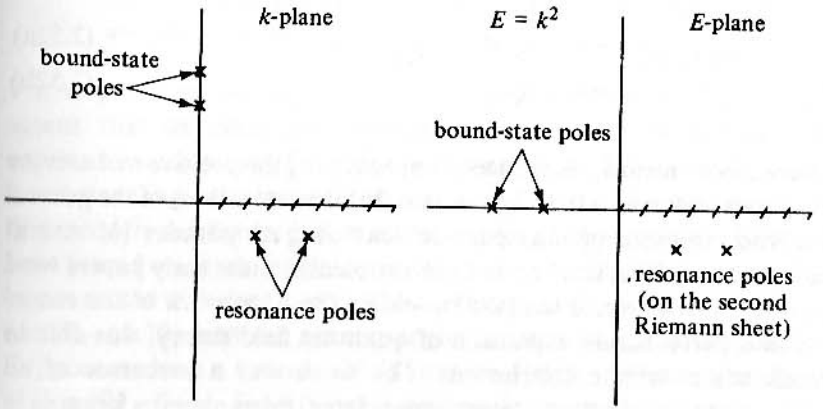


Figure 2.1 The k- and E-planes for the S matrix.

seen that Ning Hu was working with Heitler in Dublin when Pauli informed them of Heisenberg's *S*-matrix program. When Heisenberg visited Cambridge for two months early in 1948, Eden worked with him and again for a month in Göttingen later that year. Eden also recalls[32] that Heisenberg had visited Cambridge briefly a year earlier while Dyson was there.

We can easily see how discontinuities in the S or T matrices come about from Eq. (2.20). Let us take $T_{ii}(E)$ to be a function of the complex energy variable E and appreciate that the symbolic sum in Eq. (2.20) implies conservation of energy and momentum. Then, if E is real and below the two-particle scattering threshold (i.e., E is less than the energy $2m_0c^2$ which follows from the Einstein mass–energy equivalence relation $E=Mc^2$), it follows that $\operatorname{Im} T_{ii}=0$. This simply states that a reaction cannot occur if the total energy available is less than the rest mass of the interacting particles. That is, the optical theorem relates the total cross section σ_t to $\operatorname{Im} T_{ii}$. Below threshold, $\sigma_t=0$ and, hence, $\operatorname{Im} T_{ii}=0$ there. However, once we cross that threshold (by an increase in the total energy E), we have $\operatorname{Im} T_{ii}\neq 0$. Such behavior (of the change in $\operatorname{Im} T_{ii}$) is readily illustrated with the function $F(z)$ of the complex variable $z=|z|\,e^{i\phi}$

$$F(z)=\sqrt{-z}=(e^{-i\pi}z)^{\frac{1}{2}}. \tag{2.51}$$

For $z=x$(real), $F(x)$ is real for $x<0(\phi=\pi)$ and $\operatorname{Im} F(x)=0$ there. For $x>0$, $(\phi=0,2\pi)$ $\operatorname{Im} F(x)\neq 0$ and $\operatorname{Im} F(x)$ undergoes a discontinuity across the positive real axis. In fact, we find

$$\operatorname{Im} F(x)\begin{cases}\overrightarrow{z\to x^+} & -\sqrt{x} \qquad (2.52a)\\ \overrightarrow{z\to x^-} & \sqrt{x} \qquad (2.52b)\end{cases} \qquad x>0$$

where $z\to x^{\pm}$ means, respectively, approaching the positive real axis (or cut) from above ($\phi=0$) or below ($\phi=2\pi$). Investigations of the general analytic structure of multiparticle scattering amplitudes (of several complex variables) is a horrendous problem and the early papers were inconclusive. Later, Eden (1952), within the framework of the renormalized perturbation expansion of quantum field theory, was able to reach some definite conclusions. His work was a precursor of an important area of activity several years later. Eden played a key role in that subsequent development and we return to it in a later chapter.

2.6 Causality and dispersion relations

However, the outgrowth of the Heisenberg *S*-matrix program most fruitful for the eventual use of dispersion relations in elementary particle theory followed from a suggestion made by Kronig (1946). Whereas Kramers (Heisenberg, 1944) had suggested that analyticity be incorporated into the *S*-matrix program of Heisenberg, Kronig now recalled that causality had been a principle useful in deriving dispersion relations for the index of refraction for the propagation of light waves (Kronig, 1926; Kramers, 1927, 1929) (cf. Eqs. (A.80)–(A.85) and the accompanying discussion). In his paper, Kronig (1946) states:

> As is well known, the scattering of monochromatic light by atoms is governed by a relation between the real and imaginary parts of the scattering amplitudes, leading to a familiar connection between the index of refraction and the coefficient of absorption of matter in bulk. ... This relation is a direct consequence of the natural requirement that an electromagnetic field, vanishing at the place of the atom for all times $t<0$ and beginning to act only thereafter, cannot cause the emission of scattered waves before the time $t=0$. In analogy one would expect that a centre of force, influencing the waves associated with material particles in a small region around it, will not give rise to scattered waves before $t=0$ if the primary wave field at the centre is chosen to be equal to zero until this instant.
>
> It hence would seem reasonable to postulate for the scattering of particles a connection between the real and imaginary parts of the scattering amplitudes of the same type as in optics. On the side of the theory it remains to be discussed if this demand is correlated with Heisenberg's condition for S given above [the unitarity relation, $SS^{\dagger}=1$] and what is a suitable extension for the case that new particles are created in the scattering process.

Put very simply, and at a minimum, here the principle of causality[33] means that an effect (or information) cannot instantaneously be communicated across a finite distance in zero time. For example, a light wave traveling at a speed c cannot arrive at an observer located a distance l from a source before a time l/c has elapsed once the light has been turned on. It is enough to indicate here how causality can plausibly be connected with analyticity. (For more details and references, see Eqs. (A.63)–(A.70), and the accompanying discussion, in the Appendix.) Suppose we have a ('reasonable') function $f(t)$ that remains identically zero until $t=0$ (i.e., $f(t)\equiv 0, t\leqslant 0$). This could, for example, represent the response of a system to a light wave that reaches it at $t=0$. We define the

Fourier transform $f(\omega)$ of $f(t)$ as

$$f(\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)\,e^{i\omega t}\,dt.$$

Since $f(t)$ vanishes for $t\leqslant 0$, this becomes

$$f(\omega)=\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} f(t)\,e^{i\omega t}\,dt.$$

In the upper half complex ω-plane ($\mathrm{Im}\,\omega>0$) of Figure 2.2 (where C is the integration contour), the integrand above has an exponential damping factor $e^{-\mathrm{Im}\,\omega t}$. With some additional technical assumptions, this allows one to prove (and we have *not* done this here), that $f(\omega)$ is an analytic function of ω in the upper half complex ω plane. The point for our present purposes is that causality (in the sense that $f(t)$ vanishes identically prior to some time) implies analyticity for $f(\omega)$. We shall encounter variations in this type of argument again.

Kronig's was a qualitative suggestion about the use of causality. The question that now occupied several theorists was just *how* this requirement of causality was to be implemented in the *S*-matrix program and precisely what the connection between causality and analyticity would be. We do not mean to imply that *all* of their efforts were directly connected with Kronig's insight.

Jost, Luttinger and Slotnik (1950) outlined how to calculate higher-order quantum electrodynamics (QED) corrections to the perturbation expansion of the *S* matrix from lower-order terms by using the unitarity

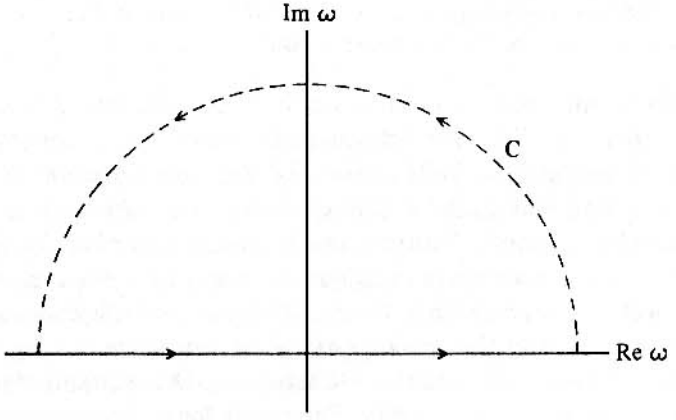


Figure 2.2 The complex ω plane for f(ω).

condition and by also exploiting the analytic properties of these matrix elements to write a dispersion relation for them. In this paper, Wheeler and Toll are thanked for that suggestion. In fact, these authors (Jost *et al.*) had, through Robert Karplus, access to a preliminary version of Toll's thesis (1952) on dispersion relations. Jost 'was ignorant of Kronig's (1946) remark ... and was under the impression that we [Jost *et al.*, 1950] were the first ones to apply dispersion techniques to *S*-matrix elements.'[34] We sketch here this interative procedure for obtaining higher-order corrections from lower-order ones because similar arguments will be important later in the history of the *S*-matrix program. Let us suppress all indices and treat $S(k)$ as a simple function (although the actual problem is more complicated). Then the basic concept is easily illustrated since the S matrix is expanded in a power series in the electric charge as

$$S(k)=1+eS_1(k)+e^2S_2(k)+\ldots \tag{2.53}$$

The unitarity condition, $S^*S=1$, becomes

$$1+e(S_1+S_1^*)+e^2(S_1S_1^*+S_2^*+S_2)+\ldots=1. \tag{2.54}$$

Since the coefficients of each power of e^n must vanish, we find that

$$2\,\mathrm{Re}\,S_2(k)=-|S_1(k)|^2 \tag{2.55}$$

which gives the real part of $S_2(k)$ once $S_1(k)$ has been calculated. But if $S_2(k)$ is an analytic function of k, we can write a dispersion relation (cf. Eqs. (A.67)) as

$$\mathrm{Im}\,S_2(k)=-\frac{1}{\pi}\mathrm{P}\int\frac{\mathrm{Re}\,S_2(k')\,dk'}{(k'-k)}. \tag{2.56}$$

(Here the P symbol on the integral indicates that a particular limiting process – the Cauchy principal value – must be taken to make this singular integral well-defined mathematically.) Rohrlich and Gluckstern (1952) calculated forward Delbrück scattering (essentially the scattering of light by the Coulomb field of a nucleus) directly, using the Feynman–Dyson rules, and then using the dispersion-relation approach of Jost *et al.* (1950). The results agreed. Rohrlich in fact was able to use the dispersion relations (at the suggestion of Joşt) to locate some errors made in the direct, brute-force approach. Toll, who had done the calculation only one way, at first obtained a result that disagreed with Rohrlich's.[35]

Actually, Rohrlich and Gluckstern's (1952) calculation was not just a perturbation-based use of analyticity as suggested in Eqs. (2.53)–(2.56).

In terms of the (complex) elastic scattering amplitude $f(k,\theta)$, the differential scattering cross section is given as

$$\frac{d\sigma}{d\Omega}=|f(k,\theta)|^2. \tag{2.57}$$

As we have already seen, the optical theorem (Feenberg, 1932; Lax, 1950) relates the total cross section to the imaginary part of the *forward* (i.e., $\theta=0$) elastic amplitude as

$$\sigma_t(k)=\frac{4\pi}{k}\,\mathrm{Im}\,f(k,\theta=0). \tag{2.58}$$

In essence, Rohrlich and Gluckstern (1952) were able to use the results of a (relatively simple) approximate calculation of $\sigma_t(k)$, obtain $\mathrm{Im}\,f(k,\theta=0)$ and then use the general dispersion relation for $f(k,\theta=0)$,

$$\mathrm{Re}\,f(k,\theta=0)=\frac{2k}{\pi}\,\mathrm{P}\int\frac{\mathrm{Im}\,f(k',\theta=0)\,dk'}{(k'^2-k^2)} \tag{2.59}$$

to find the complete forward scattering amplitude $f(k,\theta=0)$. From $f(k,\theta=0)$, Eq. (2.57) allows the differential cross section in the *forward* direction to be calculated. Bethe and Rohrlich (1952) extended these methods to an approximate calculation of nonforward Delbrück scattering. These were early and important uses of analyticity to calculate experimentally observable quantities.

Schützer and Tiomno (1951) appear to have been the first to attempt to exploit Kronig's (1946) suggestion that the principle of causality be imposed on Heisenberg's S matrix. They (1951, p. 249) credit Wigner with the surmise that the relation (2.46) for the R matrix might have its basis in that principle. Both Wigner and Bargmann at Princeton are thanked for discussions on this problem. Schützer and Tiomno (1951) considered the s-wave scattering produced by a potential of finite range a and implemented causality by the requirement that there be no scattering of an incident wave until that wave reaches the edge of the scatterer (of radius a). They (1951, p. 251) were able to construct a wave packet that is *rigorously* zero until $t=0$ only because they included negative energies in their Fourier decomposition of $\Psi(r,t)$. As van Kampen points out:[36]

> The absence of negative energies is responsible for most of the complications in my [van Kampen, 1953a] treatment, but also makes it possible for the poles on the imaginary axis that correspond to bound states to exist.

With this assumption of a sharp wave front, Schützer and Tiomno were able to show that the poles of $S(k)$ could only lie on the positive imaginary axis or in the lower half k-plane. They also found that a two-sheeted Riemann surface in the energy plane is required for $S(E)$. They were not able to deduce all of the properties of the R function previously found by Wigner and Eisenbud (1947).

Van Kampen, while working in Denmark, had written a paper (1951) on the analytic continuation of the S matrix, pointing out an equivalence between the bound-state energies E_n and the phase $\eta(k)$. He subsequently went to Princeton where he wrote the paper on the implications of causality for the S matrix. He acknowledges (1953b, p. 1276) Wigner and Bargmann for discussions on this later work. Van Kampen (1953a) applied the causality principle to the Maxwell equations for the scattering of light by a spherically symmetric center of finite size (a). As a step toward establishing the mathematical properties of Heisenberg's S matrix on general grounds, van Kampen was able to show that the S matrix (in this case) was an analytic function of k and that it satisfied a dispersion relation, as well as to derive the analytic properties of Wigner's R function. One undesirable feature of the results was that they depended explicitly upon the size a of the scattering center. Van Kampen (1953b) then considered the implications of causality for the S matrix in nonrelativistic Schrödinger theory, the same problem studied by Schützer and Tiomno. One basic difficulty in a nonrelativistic theory is that there is no maximum velocity of wave propagation. So van Kampen replaced their assumption of a wave packet with a sharp front by a restriction on the probability of finding the particle outside the scattering center. His motivation was the following (1953b, p. 1267):

> The causality condition [for a light wave traveling at the velocity c] was formulated as follows: If at a large distance r_1 from the center of the sphere [of radius a] the ingoing wave packet is zero for all $t<t_1$, then the outgoing packet shall be zero at r_1 for all $t<t_1+2(r_1-a)/c$.
>
> Obviously, for nonrelativistic particles a modification is necessary, since no maximum velocity exists, and one is inclined to postulate: If at any distance r_1 the ingoing wave packet is zero for all $t<t_1$, then the outgoing wave packet must also be zero for $t<t_1$ [as assumed by Schützer and Tiomno (1951)]. ... However, ... there is a serious objection.
>
> The difficulty is that *there are no ingoing or outgoing wave packets that are rigorously zero up to a certain time* ...

For a finite-range potential, he again obtained a dispersion relation for

$S(k)$. He also discussed (1953b, pp. 1275–6) the practical physical indistinguishability between a strictly finite-range potential and a long-range one (such as an exponential potential) and contrasted this with the radical difference in the analytic properties of the S matrix in these two cases. In practice, of course, we consider scattered particles as asymptotically free.

Finally, Toll (1952), who was a student of Wheeler's at Princeton, in a widely-referenced but unpublished Princeton Ph.D. dissertation, conducted an exhaustive study of the connection between causality and dispersion relations for light, including a calculation of Delbrück scattering (Rohrlich and Gluckstern, 1952). Toll and Wheeler (1951) applied dispersion relations to pair production processes. A summary of Toll's thesis was eventually published (Toll, 1956).

Wigner (1955a) used the property (2.47) of Eq. (2.43) for $R(E)$ to derive a famous inequality on the derivative of the phase shift,

$$\frac{\mathrm{d}\delta}{\mathrm{d}k} > -a - \frac{1}{2k}, \tag{2.60}$$

where a is the range of the potential.[37] He (1955b) also gave a beautiful and elementary summary discussion of the implications causality had for the relation (2.46) satisfied by the R matrix. As Wigner (1955a, p. 145) pointed out, a *lower* bound on $\mathrm{d}\delta/\mathrm{d}k$ is required by causality:

> It may be useful ... to derive certain general rules about the energy dependence of phase shifts .... The relations to be derived here are based, fundamentally, on what has come to be called 'the principle of causality'. It states that the scattered wave cannot leave the scatterer before the incident wave has reached it.

To see this qualitatively, consider just the s-wave *outgoing* spherical wave (cf. Eq. (A.33)) and from a wave packet with a momentum distribution $g(k)$ sharply peaked about some momentum k_0 as

$$\Psi_{sc}(r,t) = \frac{-\mathrm{i}}{2r}\int \frac{g(k)}{k}\, \mathrm{e}^{\mathrm{i}[kr + 2\delta(k) - k^2 t]}\,\mathrm{d}k. \tag{2.61}$$

As $t \to +\infty$ the rapid phase oscillations of the exponent in the integrand will produce cancellations that make the integral vanish. (This is essentially the Riemann–Lebesgue lemma familiar from standard functional analysis.) This can be avoided if we simultaneously choose r and t so that the phase

factor remains constant

$$\frac{\mathrm{d}}{\mathrm{d}k}[kr + 2\delta(k) - k^2 t] = 0$$

or

$$t = \frac{r}{2k} + \frac{1}{k}\frac{\mathrm{d}\delta}{\mathrm{d}k}. \tag{2.62}$$

The second term on the right of Eq. (2.62) is often referred to as the delay time. If there were no interaction, the wave would arrive at r at the time $r/2k$. Since the interaction potential $V(r)$ may trap the wave for an arbitrarily long time, there is no fixed upper (positive) bound on $\mathrm{d}\delta/\mathrm{d}k$. However, if the physical force or the wave speed is to remain bounded, the potential $V(r)$ cannot cause the wave to arrive at r arbitrarily *earlier* than it would have in the absence of $V(r)$. This implies that $\mathrm{d}\delta/\mathrm{d}k$ must be bounded from *below* by some negative number. So, the causality requirement gives us a qualitative understanding of Wigner's result (2.60). What Wigner (1955a) actually did can be represented essentially as follows. From the inequality of Eq. (2.47) (for the single-channel case) and Eq. (A.38) giving $R(E)$ in terms of $\delta(k)$, one shows by direct calculation of $\mathrm{d}R/\mathrm{d}E$ that

$$a + \frac{\mathrm{d}\delta}{\mathrm{d}k} > \frac{1}{2k}\sin[2(ka+\delta)] \geqslant -\frac{1}{2k},$$

which is Wigner's inequality, Eq. (2.60). An important implication of this bound is that, at a sharp phase-shift resonance where $\delta(k)$ must pass rapidly through (an odd multiple of) $\pi/2$ as k increases (cf. Eq. (A.122)), $\delta(k)$ must *rise* through $\pi/2$ because $\mathrm{d}\delta/\mathrm{d}k$ can only be large *positive* (but never large negative). Wigner (1955a, p. 146) also pointed out that an expression for $R(E)$ of the form of Eq. (2.43) had previously been shown, by Schützer and Tiomno (1951) and by van Kampen (1953a, 1953b), to be a result of the causality condition. The connection between causality and resonance formulas had become accepted.

2.7 A problem background for the 1950s

In this chapter we have attempted to establish that the extensive interest in dispersion relations and their connection with causality by the early

1950s can be traced directly back to the Heisenberg *S*-matrix program. Essentially no use has been made here of correspondence which Heisenberg himself had with others about his *S*-matrix program. Such is not really relevant for the central thesis of this section: a *prima facie* case, based on the published physics literature and on recollections of physicists active during the time span covered here, that work on questions prompted by Heisenberg's *S*-matrix program generated the background out of which grew the dispersion-theory program of the 1950s and 1960s. (For some correspondence on the *S*-matrix program, see Grythe (1982) and for the interpersonal relations and sociological background to Heisenberg's *S*-matrix program, see Rechenberg (1989).) There had been a large concentration of effort by a sizeable fraction of the community of theoretical physicists in the period 1945–54 on problems growing out of suggestions made in Heisenberg's papers. It will become even more apparent in the following chapters that Heisenberg's *S*-matrix papers raised questions leading to a series of important developments which eventually culminated in the use of dispersion relations in modern high-energy physics. The *S*-matrix and dispersion-theory programs of the 1960s had a major impact upon the theoretical high-energy physics of that decade. Several of the key contributors to that program (e.g., Chew, Gell-Mann and Goldberger) have stated in their recollections that Heisenberg's original *S*-matrix program of the early 1940s provided no direct motivation for or influence upon their own work in the late 1940s and early 1950s. While some of these founders of the later dispersion-theory and *S*-matrix theory program may not have been at the time aware of and involved in the early research on Heisenberg's *S*-matrix program so that they may not have been conscious of any influence of it on their own original work, this network of problems having roots in the Heisenberg program did, nevertheless, provide an essential background for a crucial advance in the field of relativistic dispersion relations, to which we turn in the next chapter.

But, first let us state here a thesis which can be drawn from this chapter. If Heisenberg's *S*-matrix program is remembered at all today, it is usually recalled as a program which encountered difficulties quite early and then quickly died out. From this, one could readily form an opinion that the original Heisenberg program was of little importance for subsequent theoretical developments. While it is true that Heisenberg soon abandoned his *S*-matrix program in favor of his nonlinear field theory of fundamental interactions, Heisenberg's ideas gave rise to a set of questions, such as how the causality requirement, suggested in

the mid 1940s by Kronig as a constraint on the *S* matrix, was to be implemented in the *S*-matrix theory and how the interaction potential of Schrödinger theory could in principle be determined by scattering data. Work on such problems by many theorists produced a series of papers eventually leading to relativistic dispersion relations, a program usually associated with Goldberger and Gell-Mann. The background necessary for Chew's own early investigations on reaction amplitudes was quite independent of any connection with Heisenberg's *S*-matrix theory. However, it was the *confluence* of Chew's and Goldberger's programs that led to the *S*-matrix theory of the 1960s. It is not at all clear that the latter *S*-matrix program could have become a viable candidate for a theory of strong interactions if it had not been for Heisenberg's seminal papers on an earlier program, one quite different from, yet in the same spirit as, its successor. But, we get a bit ahead of our story here.

2.8 Summary

So, we have seen that although Wheeler had already proposed an *S* matrix as a useful calculational tool for theoretical nuclear physics in 1937, Heisenberg in 1943 independently introduced the *S* matrix as the basis for a conceptual framework which might provide an alternative to Hamiltonian quantum field theory. One of his primary motivations for this move was the divergence problems of QFT in the late 1930s, which gave rise to his belief that there might be a fundamental length in nature. Initially, Heisenberg saw the *S*-matrix theory as an interim program based only on observable quantities. Kramers made the suggestion that the *S* matrix be considered an analytic function of its energy variable so that scattering data could (in principle) constrain the values of the bound-state energies. Immediately after the Second World War, Heisenberg's *S*-matrix program looked promising as a general framework within which to construct a coherent theory of interactions among fundamental particles. However, with the great success of the renormalization program in QED, field theory became the main interest among particle theorists. Through the renormalized QED program, the *S* matrix did become a well-known calculational tool. Heisenberg turned to a nonlinear quantum field theory, but neither this nor his other forays into basic theory proved successful.

Nevertheless, criticism of Heisenberg's *S*-matrix program, largely within the framework of nonrelativistic Schrödinger theory as a testing

ground, did stimulate interest in the inversion problem (of determining the potential $V(r)$ from the scattering data or phase shifts, δ_l) and in a theoretical justification for the Breit–Wigner resonance formula used to fit nuclear scattering data. In this latter undertaking, the analyticity of the S matrix suggested by Kramers was of central importance. Kronig in 1946 recalled that causality (essentially in the form of a first-signal principle) had been related to the analyticity properties of the index of refraction in optics to obtain dispersion relations. He suggested that the causality principle might also play a central role for Heisenberg's S matrix. Some progress was made in deriving the Breit–Wigner formula from the causality requirement and dispersion relations were used as an aid in quantum field theory calculations of scattering amplitudes. In 1955 Wigner showed quite explicitly how the causality requirement implied observable consequences for the behavior of a scattering phase shift near a resonance. Interest in problems such as these provided a background out of which emerged in the mid 1950s the dispersion-theory program for the strong interactions.

3

Dispersion relations

There exist some excellent technical reviews of relativistic dispersion relations (Goldberger, 1960, 1961; Jackson, 1961), as well as Goldberger's (1970) own informal recollections of the period from about 1954–69. In addition, Cini (1980) and Pickering (1989a) have written about some of the sociological influences on that program. We shall comment on these later in this chapter and in the concluding chapter. However, let us begin with a few general observations about the mood of theoretical physics in the United States just after the Second World War, at least in one Physics Department, namely the University of Chicago. This is relevant for what follows, since that Department became a center of activity for the dispersion theory program. Wentzel and Fermi were on the faculty then and Goldberger and Chew were graduate students there. One frequent attendee[1] at the theory seminars at Chicago during those years recalls that at that time (around 1948) the general spirit of many of the younger, exceptionally gifted theorists was that physics was something to *do* (never mind studying the works of the great masters) and that nothing significant had been done (in their areas of interest) prior to this. In fact, many other people had worked on these problems (e.g., fixed-source field theory) before, but there was little sense of history among the younger generation. Wentzel's comments (say, from the audience at a talk) on the previous literature were typically received with impatience. None of this, of course, detracts from the significant and original contributions made by this talented and numerically large generation of younger physicists. While Goldberger and Chew were excited by Heisenberg's ideas, which they learned about from Wentzel's lectures and Møller's articles, they did not really pursue this. The S matrix simply was not a hot topic of discussion then.[2] Furthermore, in the work that led to their 1954 paper (with Thirring) on