Lectures in Theoretical Physics

Volume XI-D

Mathematical Methods in Theoretical Physics

LECTURES IN THEORETICAL PHYSICS
University of Colorado, Boulder

Recent titles in this series

1963 Vol. VI. *Lectures in Theoretical Physics,* Wesley E. Brittin and Willard R. Chappell, Editors

1964 Vol. VIIA. *Lorentz Group,* Wesley E. Brittin and Asim O. Barut, Editors
Vol. VIIB. *Elementary Particles,* Wesley E. Brittin and Leona Marshall, Editors
Vol. VIIC. *Statistical Physics, Weak Interactions, Field Theory,* Wesley E. Brittin, Editor

1965 Vol. VIIIA. *Statistical Physics and Solid State Physics,* Wesley E. Brittin, Editor
Vol. VIIIB. *Fundamental Particles and High Energy Physics,* Wesley E. Brittin, Editor
Vol. VIIIC. *Nuclear Structure Physics,* P. D. Kunz, D. A. Lind, and W. E. Brittin, Editors

1966 Vol. IXA. *Mathematical Methods of Theoretical Physics,* W. E. Brittin, A. O. Barut, and M. Guenin, Editors
Vol. IXB. *High Energy and Particle Physics,* W. E. Brittin and A. O. Barut, Editors
Vol. IXC. *Kinetic Theory,* W. E. Brittin, Editor

1967 Vol. XA. *Quantum Theory and Statistical Physics,* Asim O. Barut and Wesley E. Brittin, Editors
Vol. XB. *High Energy Physics and Fundamental Particles,* Asim O. Barut and Wesley E. Brittin, Editors

1968 Vol. XIA. *Elementary Particle Physics,* (Two Volumes) Kalyana T. Mahanthappa, Wesley E. Brittin and Asim O. Barut, Editors
Vol. XIB. *Quantum Fluids and Nuclear Matter,* Kalyana T. Mahanthappa and Wesley E. Brittin, Editors
Vol. XIC. *Atomic Collision Processes,* Sidney Geltman, Kalyana T. Mahanthappa, and Wesley E. Brittin, Editors
Vol. XID. *Mathematical Methods in Theoretical Physics,* Kalyana T. Mahanthappa and Wesley E. Brittin, Editors

LECTURES IN THEORETICAL PHYSICS. 11. Boulder. 1968.

Volume XI-D

Lectures delivered . . . v 11 D.

MATHEMATICAL METHODS IN THEORETICAL PHYSICS

Edited by

KALYANA T. MAHANTHAPPA
and
WESLEY E. BRITTIN

Department of Physics and Astrophysics
University of Colorado

GORDON AND BREACH SCIENCE PUBLISHERS
NEW YORK • LONDON • PARIS

Copyright © 1969 by The Regents of the University of Colorado

Library of Congress Catalog Card Number: 59-13034

Published by GORDON AND BREACH, Science Publishers, Inc.
150 Fifth Avenue, New York, N.Y. 10011

Editorial office for Great Britain:
Gordon and Breach, Science Publishers Ltd.
8 Bloomsbury Way
London W.C.1

Editorial office for France:
Gordon & Breach
7-9 rue Emile Dubois
Paris 14e

Distributed in Canada by:
The Ryerson Press
299 Queen Street West
Toronto 2B, Ontario

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the Publishers.

Printed in the United States of America

CERN C 1621/70
BIBLIOTHEQUE
530.1
LEC

THE LECTURERS

Neil Ashby,
University of Colorado, Boulder, Colorado

Joseph Dreitlein,
University of Colorado, Boulder, Colorado

Albrecht Huber,
University of Birmingham, Birmingham, England

Daniel Iagolnitzer,
Miller Institute for Basic Research in Science and University of California, Berkeley, California

C. Itzykson,
Stanford Linear Accelerator Center, Stanford University, Stanford, California

T. W. B. Kibble,
Imperial College, London, England

William H. Klink,
University of Iowa, Iowa City, Iowa

Elliot H. Lieb,
Massachusetts Institute of Technology, Cambridge, Massachusetts

A. C. Manoharan,
University of Wisconsin, Milwaukee, Wisconsin

Antonio Pagnamenta,
Rutgers University, New Brunswick, New Jersey

M. Resnikoff,
State University of New York, Buffalo, New York

R. Seiler,
University of Pittsburgh, Pittsburgh, Pennsylvania

L. Susskind,
Yeshiva University, New York, New York

J. G. Taylor,
Queen Mary College, London, England

Robert L. Warnock,
Illinois Institute of Technology, Chicago, and Argonne National Laboratory, Argonne, Illinois

Walter Wyss,
Princeton University, Princeton, New Jersey

Dedicated to the memory

of

George Gamow

(1904-1968)

FOREWORD

The Eleventh Boulder Summer Institute for Theoretical Physics was held during June 17-August 23, 1968. The Institute consisted of four parts. The first part (June 17-July 5) was devoted to the Fifth Boulder Conference on Particle Physics. Quantum Fluids and Nuclear Matter were the subject matter of the second part (July 8-July 19). The Symposium on Atomic Collision Processes was held during the third part (July 22-August 2). The last part (August 5-August 23) was devoted to topics of general interest in theoretical physics with an emphasis on mathematical methods. The contributions of all the lecturers, with the exception of two, have been recorded in this eleventh volume of the Lectures in Theoretical Physics which is divided into four parts.

This volume is dedicated to the late George Gamow whose presence contributed much to the growth of the Department of Physics and Astrophyics at the University of Colorado at Boulder.

The Institute was supported in part by a grant from the National Science Foundation and in part by the University of Colorado. The Symposium on Atomic Collision Processes was supported in part by the Advanced Research Projects Agency.

Asim O. Barut contributed to the organization of the Conference in the initial stages. Roy Garstang and Sydney Geltman assumed the responsibilities of organizing the Symposium on Atomic Collision Processes. Consultations with and the moral support of Wesley E. Brittin were extremely helpful. Mrs. Charlotte Walker contributed to the organization of the Institute with her invaluable secretarial help.

Kalayana T. Mahanthappa
Director

PREFACE

The Volume XI-D of the Lectures in Theoretical Physics forms the proceedings of the session on Mathematical Methods in Theoretical Physics which was held during the fourth part of the Eleventh Boulder Summer Institute for Theoretical Physics. The session was devoted mainly to recent developments in S-matrix theory and quantum field theory and applications of group theory to physical problems. The volume contains only the text of lectures and seminars given during the session; no effort has been made to include the discussions.

The editors thanks the lecturers for having sent in their manuscripts in time and proofreading the typed version and Mrs. Marion Higa for her excellent typing of the manuscripts.

Editors,

Kalyana T. Mahanthappa
Wesley E. Brittin

Department of Physics
and Astrophysics
University of Colorado

CONTENTS

1. Group Theoretic Methods

GROUP THEORY AND STOCHASTIC PROCESSES
Joseph Dreitlein 1

INDUCED REPRESENTATION THEORY AND THE POINCARÉ GROUP
William H. Klink 43

THE LORENTZ GROUP AND COMPLEX ANGULAR MOMENTUM
J. G. Taylor 77

A THEOREM ON ANALYTIC PROPERTIES OF UNITARY REPRESENTATIONS
C. Itzykson 127

INFINITE MOMENTUM FRAMES AND PARTICLE DYNAMICS
L. Susskind 135

NUCLEAR QUASI-SPIN MODELS AND THE REPRESENTATIONS OF $U(4) \supset U(2) \otimes U(2)$
M. Resnikoff 183

2. S-Matrix Theory

ON THE COVARIANCE AND ANALYTICITY OF SCATTERING AMPLITUDES
R. Seiler 195

SPACE-TIME PROPERTIES AND PHYSICAL REGION ANALYTICITY IN S-MATRIX THEORY
Daniel Iagolnitzer 221

NONLINEAR ANALYSIS APPLIED TO S-MATRIX THEORY
Robert L. Warnock 273

3. Statistical Physics

TWO-DIMENSIONAL ICE AND FERROELECTRIC MODELS
Elliot H. Lieb 329

SELF-CONSISTENCY AND STABILITY IN THE VARIATIONAL CALCULATION OF THERMODYNAMIC QUANTITIES
Albrecht Huber 355

4. Quantum Field Theory

COHERENT STATES AND INFRARED DIVERGENCES
T. W. B. Kibble 387

SOLUTIONS OF BETHE-SALPETER EQUATIONS
Antonio Pagnamenta 479

5. Axiomatic Quantum Field Theory

ON WIGHTMAN'S THEORY OF QUANTIZED FIELDS
Walter Wyss 533

SOME CONSIDERATIONS ON CONFIGURATION SPACE ANALYTICITY IN GENERAL QUANTUM FIELD THEORY
A. C. Manoharan 577

6. Miscellany

EXACT SOLUTIONS FOR SPIN AMPLITUDES IN EXTERNAL OSCILLATING FIELDS
Neil Ashby 599

HIGH ENERGY SCATTERING IN THE QUARK MODEL
David R. Harrington and Antonio Pagnamenta . . . 621

1. Group Theoretic Methods

GROUP THEORY AND STOCHASTIC PROCESSES†

Joseph Dreitlein‡
Department of Physics and Astrophysics
University of Colorado
Boulder, Colorado

In the last five years, group theory has had extensive applications in particle physics. The purpose of these lectures is to indicate a new area of use for group theoretic techniques, the statistical properties of systems. It will be shown that many physical stochastic processes[1] exhibit sufficient symmetry to permit powerful group theoretic methods to simplify and solve the problems which arise.

In Section I, a simple model of a stochastic process is discussed so that the pertinent ideas and definitions can be fixed. We then go on to develop a theory of stochastic group movements. In Section III, the theory is applied to the solution of a multiple scattering problem. Finally in Section IV, another use of group theory is presented to reduce the problem of Polya walks to a Clebsch-Gordan problem.

The reader who is interested in other approaches to stochastic processes will find a good introduction in Ref. 2.

I. A Mixing Problem

To isolate the relevant ideas involved in Markovian processes, it is apropos to consider a simplified model of a relaxation process. Suppose that in the left hand half of a box are N_1 atoms of a neon, in the other half, N_2 atoms of argon. The mixing problem is simplified so that first an atom in the right side (side 1) crosses to the left (side 2), then an atom from the left side crosses to the right, then right to left, etc. Although the mixing mechanism is in reality more complicated, the simplified description should give a first approximation to the relaxation process.

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.
‡Alfred P. Sloan Fellow, supported in part by Office of Naval Research under contract no. RR-002-05-01, 1-241.

It suffices to fix attention on the history of a single atom. Let $\mathfrak{M}_{ij}'$ be the probability that on step one the atom moves from side j to side i. It is readily seen that

$$\mathfrak{M}' = \left(\begin{array}{c|c} 1 - \frac{1}{N_1} & 0 \\ \hline \frac{1}{N_1} & 1 \end{array}\right) \tag{I.1}$$

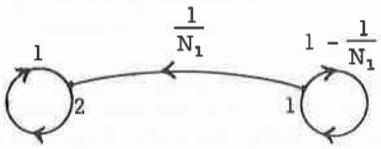


Figure 1.1

Similarly left-right transitions are described by the matrix $\mathfrak{M}''$:

$$\mathfrak{M}'' = \left(\begin{array}{c|c} 1 & \frac{1}{N_2+1} \\ \hline 0 & 1 - \frac{1}{N_2+1} \end{array}\right) \tag{I.2}$$

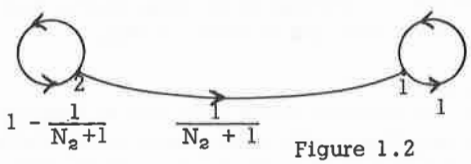


Figure 1.2

Note that after the first step there are N_2+1 atoms on the left side. The accompanying diagrams (Figures 1.1 and 1.2) picture the transitions and the assigned probabilities.[3]

The net result of two steps is summarized by the transition matrix $\mathfrak{M}_{ij}$, again determined by the probability of going from j to i

$$\mathfrak{M} = \mathfrak{M}''\mathfrak{M}' = \begin{pmatrix} 1 - \frac{N_2}{N_1(N_2+1)} & \frac{1}{N_2+1} \\ \frac{N_2}{N_1(N_2+1)} & \frac{N_2}{N_2+1} \end{pmatrix} \tag{I.3}$$

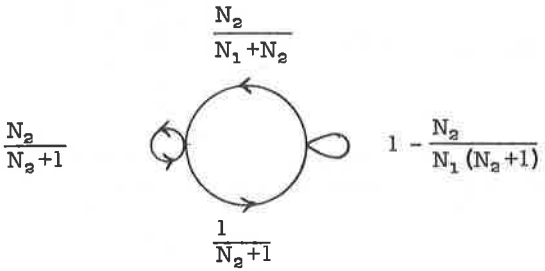


Figure 1.3

These matrices [Eqs. (I.1), (I.2), (I.3)] change the probability of a particle being found in side i (call it p_i) by acting on the state vector w; e.g.,

$$w' = \mathfrak{M} w \quad ; \quad \begin{pmatrix} p_1' \\ p_2' \end{pmatrix} = \mathfrak{M} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \tag{I.4}$$

and yield the new probabilities of being found on the right and left hand sides. It is clear that after an even number of steps N that

$$w(N) = \mathfrak{M}^{\frac{N}{2}} w(0) \tag{I.5}$$

To explicitly evaluate w(N), the matrix $\mathfrak{M}$ is diagonalized. Since it is non-symmetric its right eigenvectors $u^{(i)}$ differ from its left eigenvectors $v^{(i)}$:

$$u^{(1)} = \begin{pmatrix} \frac{N_1}{N_1+N_2} \\ \frac{N_2}{N_1+N_2} \end{pmatrix} \quad u^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$v^{(1)} = (1,1) \qquad v^{(2)} = \left(\frac{N_2}{N_1+N_2} , \frac{-N_1}{N_1+N_2} \right)$$

$$\left(v^{(i)}, u^{(j)} \right) = \delta_{ij} \tag{I.6}$$

Expand w(0) into a sum of eigenvectors:

$$w(0) = (v^{(1)}, w(0))\, u^{(1)} + (v^{(2)}, w(0))\, u^{(2)} \tag{I.7}$$

and note:

$$w(N) = u^{(1)} + (v^{(2)}, w(0)) \left(\frac{1 - \frac{1}{N_1}}{1 + \frac{1}{N_2}} \right)^{\frac{N}{2}} u^{(2)} \qquad (I.8)$$

since the sum $p_1 + p_2$ equals 1 and the eigenvalues are 1 and $[(1 - \frac{1}{N_1})/(1 + \frac{1}{N_2})]$. Since the latter is smaller than unity,

$$\lim_{N \to \infty} w(N) = u^{(1)} = \begin{pmatrix} \frac{N_1}{N_1 + N_2} \\ \frac{N_2}{N_1 + N_2} \end{pmatrix} \qquad (I.9)$$

and the atom being followed ultimately has relative probabilities N_1 and N_2 of being on the right or left hand sides respectively.

The factor $(1 - \frac{1}{N_1})/(1 + \frac{1}{N_2})$ governs the rate of relaxation. Suppose that N_1 and N_2 are very large, and that $\frac{N}{2} = x\left(\frac{N_1 N_2}{N_1 + N_2}\right)$. Then

$$\left(\frac{1 - \frac{1}{N_1}}{1 + \frac{1}{N_2}} \right)^{\frac{N}{2}} \cong \left(1 - \frac{N_1 + N_2}{N_1 N_2} \right)^{x \frac{N_1 N_2}{N_1 + N_2}} \approx e^{-x} \qquad (I.10)$$

which indicates that the transient term dies out exponentially, the lifetime being $2 \frac{N_1 N_2}{N_1 + N_2}$ steps.

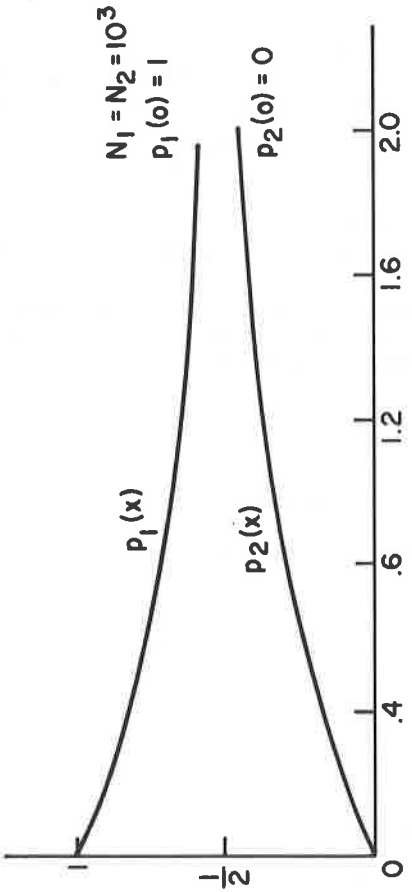


Fig. 1.4. Equilibrization by Mixing

The graph (Fig. 1.4) illustrates the approach to equilibrium under the stated conditions and the table gives $p_1(x)$ and $p_2(x)$:

x	0	0.2	0.4	0.6	0.8	1.0	5.0
$p_1(x)$	1	.0903	.8352	.7744	.7247	.6840	.5034
$p_2(x)$	0	.0907	.1648	.2256	.2753	.3160	.4966

As a physical example, consider the mixing of two isotopes of Ne, Ne^{20} and Ne^{19}. Each isotope initially occupies 1 cm^3. The time taken for a double exchange across the boundary is

$$\tau = \left(\frac{2mkT}{\pi}\right)^{\frac{1}{2}}$$

while the number of each species is $N \approx 2.69 \times 10^{19}$. Hence the relaxation time according to the rough theory at room temperature is about 12 seconds.

There are certain properties of the illustrated Markovian process which can be abstracted from the simple example.

A matrix $\mathfrak{M}$ with the property

$$\sum_i \mathfrak{M}_{ij} = 1; \quad \mathfrak{M}_{ij} \geq 0$$

is called a stochastic matrix and enjoys a number of special properties:

(a) If $\mathfrak{M}$ is stochastic, then $\mathfrak{M}^N$ is stochastic.

Proof:

$$\sum_i (\mathfrak{M}^N)_{ij} = \sum_i \sum_{k_1 k_2 \ldots k_n} \mathfrak{M}_{ik_1} \mathfrak{M}_{k_1 k_2} \ldots \mathfrak{M}_{k_n j}$$

$$= \sum_{k_1 k_2 \ldots k_n} \mathfrak{M}_{k_1 k_2} \ldots \mathfrak{M}_{k_n j} = \sum_{k_n} \mathfrak{M}_{k_n j} = 1$$

Also $(\mathfrak{M}^N)_{ij}$ is clearly non-negative.

(b) One left eigenvector is always $(1,1,1,\ldots,1)$.

This follows from $\sum_i \mathfrak{M}_{ij} = 1$. Therefore $\sum_j u_j^{(i)} = 0$ if $\lambda^{(i)} \neq 1$.

(c) All eigenvalues of $\mathfrak{M}$ are bounded in absolute value by unity.

Proof: Let $\sum_j \mathfrak{M}_{ij} u_j = \lambda u_i$. Then

$$\sum_j (\mathfrak{M}^N)_{ij}\, u_j = \lambda^N u_i \quad .$$

But

$$\Big|\sum_j (\mathfrak{M}^N)_{ij}\, u_j\Big| \le \Big|\sum_j (\mathfrak{M}^N)_{ij}\Big|\, |u_j|$$

so

$$\lambda^N |u_i| \le \sum_j (\mathfrak{M}^{(N)})_{ij} |u_j| \le \sum_j |u_j|$$

Let $|u_i| = \max [|u_1|, |u_2|, \ldots]$ so that

$$|\lambda|^N \le \frac{\sum_j |u_j|}{|u_i|}$$

for all N. This is impossible if $|\lambda| > 1$.
(d) Eigenvalues which are complex occur in pairs, both the eigenvalue and its complex conjugate appear in the spectrum. The property follows from the reality of $\mathfrak{M}_{ij}$.

II. Group Theory and Stochastic Processes

It is often the case that Markovian processes which are of interest in physics exhibit a degree of symmetry, a circumstance intimating the fruitfulness of using group theoretic techniques.[4]

To see how group theoretic considerations arise, consider the problem of a random walk on a line. Let $f(x)$, $f^{(n)}(x)$ be the probabilities of moving a distance x in one step and n-steps respectively. Then $f^{(n)}(x)$ is related to $f(x)$ by convolution operations:

$$f^{(2)}(x) = \int f(x-x')\, f(x')\, dx'$$

$$f^{(3)}(x) = \int f(x-x'')\, f(x''-x')\, f(x')\, dx'dx''$$

$$\cdots$$

$$f^{(n)}(x) = \int f(x-x')\, f^{(n-1)}(x')\, dx' \qquad \text{(II.1)}$$

with normalization condition

$$\int f(x)\, dx = 1 \qquad \text{(II.2)}$$

To evaluate $f^{(n)}(x)$, it is easiest to introduce the Fourier transforms:

$$\hat{f}(k) = \int f(x)\, e^{-ikx} dx$$

$$\hat{f}^{(n)}(k) = \int f^{(n)}(x)\, e^{-ikx} dx \tag{II.3}$$

for then

$$\hat{f}^{(n)}(k) = [f(k)]^n$$

$$f^{(n)}(x) = \frac{1}{2\pi} \int [f(k)]^n\, e^{ikx} dk \tag{II.4}$$

Two examples suffice to show the simplicity of the procedure:

a. $$f(x) = \sqrt{\frac{\alpha}{\pi}}\; e^{-\alpha x^2}$$

Then,

$$\hat{f}(k) = e^{\frac{-k^2}{\alpha^2}} \quad ; \quad [\hat{f}(k)]^n = e^{\frac{-nk^2}{\alpha}} \tag{II.5}$$

So

$$f^{(n)}(x) = \sqrt{\frac{\alpha}{n\pi}}\; e^{\frac{-\alpha}{n} x^2}$$

b. $$f(x) = \tfrac{1}{2}[\delta(x-a) + \delta(x+a)]$$

Then,

$$\hat{f}(k) = \cos ka\,, \quad [\hat{f}(k)]^n = (\cos ka)^n$$

so

$$f^{(n)}(x) = \sum_{m=n, n-2, \ldots, -n} \frac{1}{2^n} \binom{n}{\frac{n-m}{2}} \delta(x-ma) \tag{II.6}$$

A reinterpretation of the procedure will elucidate the symmetry inherent in the problem. The stochastic motion of a particle on a line is prescribed by the probability $f(x)$ of making a step of length x and is independent of the point where the particle is at step 0. With this shift in interpretation goes a new notation: let $\mathfrak{M}(x)$ be the probability of making a movement x. Now x is looked upon as an element of the translation group and, of course,

$$\mathfrak{M}^{(2)}(x) = \int \mathfrak{M}(x-x')\mathfrak{M}(x')\,dx' \tag{II.7}$$

The line itself may be looked upon as the group manifold (and also, trivially, a homogeneous space[5] of the group manifold) of the translation group. The motivation for making the Fourier transformation of $\mathfrak{M}(x)$ becomes clearer since the functions e^{ikx} are the one dimensional representations of the translation group. Since $\mathfrak{M}(x)$ is defined on the group manifold, it can be expanded in a complete set of representations of the group:

$$\mathfrak{M}(x) = \frac{1}{2\pi}\int \hat{\mathfrak{M}}(k)\, e^{i\,kx}\,dk \tag{II.8}$$

The compounding of steps is the convolution operation in the group algebra[6] and is simplest when carried out in irreducible representations of the group.

A related way of looking at the problem is to consider the Markovian matrix $\mathfrak{M}(x_2,x_1)$, the probability of moving from x_1 to x_2 in a single step. It depends only upon x_2-x_1 for a homogeneous process:

$$\mathfrak{M}(x_2,x_1) = \mathfrak{M}(x_2-x_1) \tag{II.9}$$

Consequently, $\mathfrak{M}(x_2,x_1)$ considered as a matrix commutes with the translation group

$$\mathfrak{M}(x_2,x_1) \rightarrow \mathfrak{M}(x_2+a),\ x_1+a) = \mathfrak{M}(x_2,x_1) \tag{II.10}$$

with the consequent simplifying symmetry considerations.

Now, we develop the general theory of stochastic group motion. Consider first a finite group G whose order is $|G|$. Let $g_1, g_2, \ldots, g_{|G|}$ be the elements of G and let $\mathfrak{M}(g)$ be the probability of making a group movement specified by the element g. Ultimately, the group will be allowed to operate on a space, either the group manifold or any one of a number of derived spaces, the homogeneous spaces of the group.

Note that

$$\sum_g \mathfrak{M}(g) = 1 \tag{II.11}$$

and also that $\mathfrak{M}(g,g'^{-1}) \equiv \mathfrak{M}(gg'^{-1})$ may be looked upon as a stochastic matrix since

$$\sum_{g} \mathfrak{M}(g, g'^{-1}) = 1 \tag{II.12}$$

by the group property that gg'^{-1} (for fixed g'^{-1}) runs over all group elements when g runs over all group elements.

The probability of making group movement g in two steps is clearly:

$$\mathfrak{M}^{(2)}(g) = \sum_{g'} \mathfrak{M}(gg'^{-1})\, \mathfrak{M}(g') \tag{II.13}$$

and in n steps:

$$\mathfrak{M}^{(n)}(g) = \sum_{g_{n-1}, g_{n-2}, \ldots g_2, g_1} \mathfrak{M}(g_n g_{n-1})^{-1}\, \mathfrak{M}(g_{n-1} g_{n-2})^{-1} \ldots \quad (g_2 g_1^{-1})(g \tag{II.14}$$

Here $g_1 g_2$ is the element of G obtained from the group multiplication law. To evaluate the convolution of Eq. (II.14), recall that the irreducible representations of a group, $D^{\Gamma}_{ij}(g)$, form a complete set over the group manifold. Therefore

$$\mathfrak{M}(g) = \sum_{\Gamma, i, j} \frac{|\Gamma|}{|G|}\, \mathfrak{M}^{\Gamma}_{ij}\, D^{\Gamma}_{ij}(g)$$

$$\mathfrak{M}^{\Gamma}_{ij} = \sum_{g} \mathfrak{M}(g)\, D^{\Gamma}_{ji}(g^{-1}) \tag{II.15}$$

with the following notations:

D_{ij} = $|\Gamma|$ dimensional irreducible unitary representation of G;

$$i, j = 1, \ldots, |\Gamma|$$

$\mathfrak{M}_{ij}$ = "harmonic" transform of $\mathfrak{M}(g)$

The harmonic transform of $\mathfrak{M}(g)$ plays the central role in the description of the stochastic process. As it turns out, $\mathfrak{M}_{ij}$ control the rate of relaxation to equilibrium.

The central theorem to be displayed is the following. Let

$$\mathfrak{M}^{(n)}(g) = \sum_{\Gamma,ij} \frac{|\Gamma|}{|G|} \mathfrak{M}^{\Gamma}_{ij}(n)\, D_{ij}(g)$$

$$\mathfrak{M}^{\Gamma}_{ij}(1) \equiv \mathfrak{M}^{\Gamma}_{ij} \tag{II.16}$$

then

$$\mathfrak{M}^{\Gamma}_{ij}(n) = \sum_{i,\ldots i_{n-1}} \mathfrak{M}^{\Gamma}_{ii_1}, \mathfrak{M}^{\Gamma}_{i_1 i_2} \ldots \mathfrak{M}^{\Gamma}_{i_{n-1} j}; \tag{II.17}$$

Eq. (II.17) allows the convolution to be carried out by merely multiplying $|\Gamma|$ dimensional matrices together. The main theorem [Eq. (II.17)] follows from the orthogonality and completeness properties of irreducible representations.

The eigenvalues of $\mathfrak{M}^{\Gamma}_{ij}$ will be called roots of the motion. They all lie within or on a unit circle in the complex plane. $\mathfrak{M}^{\Gamma}_{ij}$ itself will be called root or relaxation matrix.

The following are general properties of relaxation matrices:

a. The matrix $\mathfrak{M}^{\Gamma}_{ij}$ has all its entries bounded by unity, a property following from the fact that $\mathfrak{M}(gg'^{-1})$ is a stochastic matrix and from the unitarity of the irreducible representations of G. The eigenvalues are bounded in absolute value by unity; moreover, if none reach this bound, then

$$\lim_{n \to \infty} \mathfrak{M}^{\Gamma}_{ij}(n) \to 0.$$

b. The identity representation of the group, $D(g) = 1$, gives rise to one eigenvalue whose value is unity:

$$\mathfrak{M}'_{i1} = \sum_{g} \mathfrak{M}(g) = 1 .$$

If this is the only eigenvalue whose magnitude is one then

$$\lim_{n \to \infty} \mathfrak{M}^{(n)}(g) = 1/|G|$$

and the asymptotic probabilities of going anywhere on the group manifold are all equal.

c. A sufficient condition for the root matrix to be Hermitian is $\mathfrak{M}(g) = \mathfrak{M}(g^{-1})$. All eigenvalues are, of course, then real.

Before developing the theory of stochastic group motion further, a trivial example should help clarify the manipulations. Suppose the homogeneous random walk on a ring of L sites is to be analyzed. The group inducing the motion is the cyclic group C_L. Its irreducible representations are:

$$D^{(m)}(r) = e^{\frac{2\pi imr}{L}} \qquad n = 0,1,2,\ldots,L-1 \tag{II.18}$$

where r is the group motion of n steps in a clockwise direction and $\Gamma = m$ is the index symbol for an irreducible representation. Now

$$\mathfrak{M}^{(n)}(r) = \frac{1}{L} \sum_{m} \mathfrak{M}^{m}(n)\, e^{\frac{2\pi imr}{L}}$$

$$\mathfrak{M}^{m}(n) = [\mathfrak{M}^{m}(1)]^{n}$$

$$\mathfrak{M}^{(m)}(1) = \sum_{r} \mathfrak{M}(r)\, e^{\frac{-2\pi imr}{L}} \tag{II.19}$$

For the particularly simple case of a probability of $\frac{1}{2}$ of going one step either clockwise or counterclockwise illustrated in Fig. 3.1 one finds

$$\mathfrak{M}^{(m)}(1) = \cos\frac{2\pi M}{L} \tag{II.20}$$

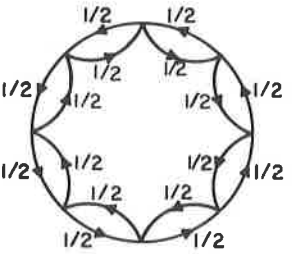


Fig. 3.1. The random walk on a circle.

so that

$$\mathfrak{M}^{(n)}(r) = \frac{1}{L} \sum_{g} \left[\frac{\cos 2\pi m}{L}\right]^{N} e^{\frac{2\pi imr}{L}} \tag{II.21}$$

As $n \to \infty$, $\cos \frac{2\pi m}{L} \to 0$ except for $m = 0$, and $m = \frac{L}{2}$ if L is even. All other motions are transitory and die out eventually. For L even,

$$\lim_{n \to \infty} \mathfrak{M}^{(n)}(r) = \frac{1}{L}\left[1 + (-)^{n+r}\right] \tag{II.22}$$

The particle moving on a cycle of an even number of sites is on odd sites on the odd steps and on even sites on the even steps. There is no "ergodic" behavior in this case but an asymptotic periodic motion with period two.

How many steps must be taken before the asymptotic limit Eq. (II.22) is a good approximation? The transient effects with longest finite relaxation times are obviously generated by the roots.

$$\cos \frac{2\pi}{L}, \quad \cos \frac{2\pi(L-1)}{L}, \quad \cos \frac{2\pi\left(\frac{L}{2}-1\right)}{L}, \quad \cos \frac{2\pi\left(\frac{L}{2}+1\right)}{L}$$

For L large, each of these deviate from magnitude unity by a term of order $\frac{1}{2}\left(\frac{2\pi}{L}\right)^2$ from which it is easy to deduce that only after about L^2 steps do these terms begin dying out. Such behavior is general for diffusion processes where the distance gone is proportional to the square root of the time.

In several cases of physical interest, such as the sphere walk to be treated in the next section, the probability of group motion depends only upon the class to which g belongs. A very considerable simplification results, for then the matrix $\mathfrak{M}_{ij}$ commutes with every matrix $D_{ij}(g)$ and so, by Schur's lemma, must be a multiple of the identity:

$$\mathfrak{M}^{\Gamma}_{ij} \equiv \mathfrak{M}^{\Gamma}\,\delta_{ij}$$

$$\mathfrak{M} = \sum_{g} \frac{\mathfrak{M}(g)}{|\Gamma|}\,\chi^{\Gamma}(g^{-1}) \tag{II.23}$$

where $\chi^{\Gamma}(g)$ is the character of the Γ representation of the group element g. Furthermore, the convolution process becomes:

$$\mathfrak{M}^{\Gamma}(n) = (\mathfrak{M}^{\Gamma})^n \tag{II.24}$$

and so

$$\mathfrak{M}^{(n)}(g) = \frac{1}{|G|} \sum |\Gamma|\,(\mathfrak{M}^{\Gamma})^n \chi^{\Gamma}(g) \tag{II.25}$$

--a substantial simplification!

The roots of the group motion are now

$$\mathfrak{M}^{\Gamma} = \frac{\sum_{i} \mathfrak{M}^{\Gamma}{}_{ii}}{|\Gamma|}$$

Again these are bounded in absolute value by unity and are therefore distribution within or on the unit circle in the complex plane.

If $\mathfrak{M}^{\Gamma} = |\mathfrak{M}^{\Gamma}| e^{i\phi}$, then

$$(\mathfrak{M}^{\Gamma})^{n} = |\mathfrak{M}^{\Gamma}|^{n} e^{in\phi} \tag{II.26}$$

and as a function of n, $\mathfrak{M}^{\Gamma}(n)$ decays along a logarithmic spiral if $|\mathfrak{M}^{\Gamma}| < 1$. See Fig. 2.2.

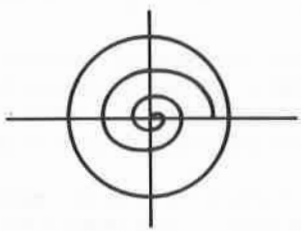


Fig. 2.2. The decay of transients.

(This is the curve along which a moth is purported to fly into a candle.)[7] A relaxation time is given by

$$\tau_{\Gamma} = -\frac{1}{\ell n |\mathfrak{M}^{\Gamma}|} \tag{II.27}$$

If $|\mathfrak{M}^{\Gamma}| = 1$ but $\mathfrak{M}^{\Gamma} \neq 1$, then a periodic asymptotic behavior will develop: if $\varphi = \frac{2\pi}{n}$, the motion repeats itself after n steps. In this scheme, it is quite easy to identify relaxation times and periodic asymptotic behavior. The entire behavior, as advertised, is controlled by the roots of the group motion.

To illustrate a group motion which depends only upon the class structure, a crude approximation to the sphere walk is constructed--the random walk from vertex to vertex of a tetrahedron. The symmetry of the tetrahedron clearly enters if one vertex is as good as any

vertex. The tetrahedron (Fig. 2.3) is converted into itself by the following symmetry operations:

1. Identity mapping -- g_0.

2. Rotations thru $2\pi/3$ radians about an axis from a vertex to the center of the opposite face. There are four of these axes and the operations are $g_{v1}, g_{v2}, g_{v3}, g_{v4}$.

3. Rotations about the same axes but thru $4\pi/3$ radians with operations designated by $g_{v1}', g_{v2}', g_{v3}', g_{v4}'$.

4. Rotations thru π radians about an axis connecting the center of one edge with the center of a diametrically opposite edge. The operations are called g_{E_1}, g_{E_2}, and g_{E_5}.

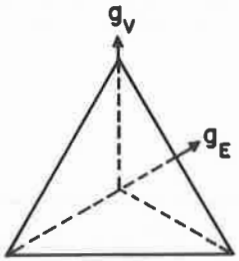


Fig. 3.3. The tetrahedron and symmetry operations.

Note that the four groupings of operations listed specify also the four classes of operations: $\kappa_0, \kappa_V, \kappa_V', \kappa_E$. For example, $g_{E_1}\, g_{v_1}\, g_{E_1}^{-1} = g_{v_2}$.

It is easily visualized that $\mathfrak{M}(g_{E_1}) = \mathfrak{M}(g_{E_2}) = \mathfrak{M}(g_{E_3}) = \frac{1}{3}$ of going from one vertex to any adjacent vertex. The support of $\mathfrak{M}(g)$ is only on the class κ_E. With a knowledge of the character table of the tetrahedral group:

Γ \ κ	κ_0	κ_E	κ_V	$\kappa_{V'}$
Γ_1	1	1	1	1
Γ_2	1	1	ω	ω^*
Γ_3	1	1	ω^*	ω
Γ_4	3	-1	0	0

$\omega = e^{\frac{2\omega i}{3}}$

Table 2.1.

the solution of the problem follows once the roots are determined:

$$\mathfrak{M}^{\Gamma(1)} = \mathfrak{M}^{\Gamma(2)}\,\mathfrak{M}^{\Gamma(3)} = 1;\ \mathfrak{M}^{\Gamma(4)} = -\tfrac{1}{3} \tag{II.28}$$

so that

$$\mathfrak{M}^{(n)}(g_o) = \tfrac{1}{4}\left[1 + 3(-\tfrac{1}{3})^n\right]$$

$$\mathfrak{M}^{(n)}(g_{E_1}) = \mathfrak{M}^{(n)}(g_{E_2}) = \mathfrak{M}^{(n)}(g_{E_3}) = \tfrac{1}{4}\left[1 - (-\tfrac{1}{3})^n\right]$$

$$\mathfrak{M}^{(n)}(g) = 0 \text{ for } g \neq g_o, g_{E_1}, g_{E_2}, g_{E_3} \tag{II.29}$$

The probability distribution on the vertices relaxes quite rapidly to equilibrium as Table 2.2 illustrates. The results are graphed in Fig. 2.4.

n	$\mathfrak{M}^{(n)}(g_o)$	$\mathfrak{M}^{(n)}(g_E)$
1	.000000	.333333
2	.333333	.222222
3	.222222	.259259
4	.259259	.246913
5	.246913	.251028
6	.251028	.249657
7	.249657	.250114
8	.250114	.249961
9	.249961	.250012
10	.250012	.249995

Table 2.2. Relaxation of group motion on tetrahedron

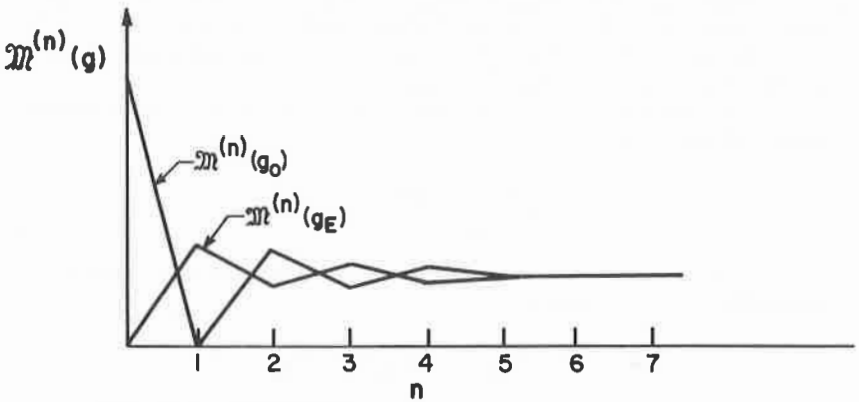


Fig. 2.4. Relaxation of group motion on tetrahedron

A particle starting out at one vertex will after ten steps have $\frac{1}{4}$ probability of being at any vertex to better than one part in ten thousand.

A few words of interpretation go with this example. The four vertices of a tetrahedron form a homogeneous space of the group of the tetrahedron A_4 (see Appendix B) the division subgroup being C_3, the group consisting of g_0, g_{v_1} and $g_{v_1}{}'$:

$$A_4/C_3 = \{\text{set of tetrahedral vertices}\} \qquad \text{(II.30)}$$

In more complicated groups, the expectation that random walks on the homogeneous spaces of the group are treated by the methods developed in this section is born out in practice.

Finally, it might be noted that an invariant subgroup of A_4, the Abelian group D_2 with members g_0, g_{E_1}, g_{E_2}, and g_{E_3} is also an appropriate group to use in analyzing the tetrahedral walk. This group D_2 is part of the group of the cube ("viergruppe") which circumscribes the tetrahedron. Its representations are all one dimensional:

	g_0	g_{E_1}	g_{E_2}	g_{E_3}
Γ_1	1	1	1	1
Γ_2	1	-1	-1	1
Γ_3	1	1	-1	-1
Γ_4	1	-1	1	-1

Table 2.3. Representations of D_2

With $\mathfrak{M}(g_o) = 0$, $\mathfrak{M}(g_{E_1}) = \mathfrak{M}(g_{E_2}) = \mathfrak{M}(g_{E_3}) = \frac{1}{3}$, the solution obtained from Eqs. (II.23) and (II.25) immediately falls out.

To see that the group D_2 is transitive on the tetrahedron (formally), the Burnside lemma[8] is useful.

The number of transitive sets on a space on which a group operates is given by:

$$\frac{1}{|G|} \sum_{g \in G} \Psi(g)$$

$\Psi(g)$ = number of points in space left invariant by g. For D_2, operating on the tetrahedron

$$\Psi(g_o) = 4, \quad \Psi(g_{E_1}) = \Psi(g_{E_2}) = \Psi(g_{E_3}) = 0$$

so that $\sum_{g \in D_2} \Psi(g)/|G| = 1$. Indeed, Burnside's lemma is a useful general condition for exploring transitivity of a group on a space. In general, if the support of $\mathfrak{M}(g)$ encompasses at least the generators of the group,[9] the group will act transitively on the group manifold.

The generalization of the walk on four points is the walk on a simplex to which we now direct attention.

A simplex is a member of the series-point, line, triangle, tetrahedron, etc. It is the simplest polytope σ^N in n dimensions, having N+1 vertices and containing $\binom{N+1}{p} \sigma^p$ subsimplices. The simplex σ^N is obtained as the subset S_{N+1}/S_N where S_N is the symmetric group on n objects.

If the symmetric group is used to study the random walk on σ^N, then for probabilities depending only upon class membership

$$\mathfrak{M}^{(n)}(g) = \sum_{\Gamma} \frac{|\Gamma|}{|G|} \left(\mathfrak{M}^{\Gamma}\right)^n \chi^{\Gamma}(g) \qquad \text{(II.31)}$$

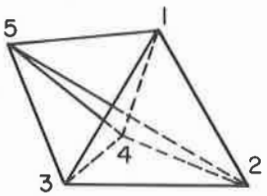


Fig. 2.5. Simplex σ^4 Schlegel diagram.

Since several group operations lead to the same move on the simplex, a quantity of direct interest is really

$$\mathfrak{M}_1 = \sum_{g\in\mathfrak{g}} \mathfrak{M}(g) = \sum_{\Gamma} \frac{|\Gamma|}{|G|} \mathfrak{M}^{\Gamma} \sum_{g\in\mathfrak{g}} \chi^{\Gamma}(g) \qquad \text{(II.32)}$$

where $\mathfrak{g}$ is the little group S_N, if the problem involves starting out from one point. $\mathfrak{M}$ is the probability of returning to the starting vertex. Now observe that

$$\sum_{g} \chi^{(\Gamma, S_{N+1})}(g) = \sum_{g} \sum_{\Gamma'} \chi^{(\Gamma', S_N)}(g) = \sum_{\Gamma'} \delta_{\Gamma',\Gamma_1} \chi^{(\Gamma', S_N)} \qquad \text{(II.33)}$$

Γ_1 = trivial representation of S_N

so that $\sum_{g\in\mathfrak{g}} \chi^{\Gamma}(g)$ is non vanishing if and only if χ^{Γ} when reduced to the subgroup S_N contains the trivial representation of S_N. But this condition is immediately ascertainable since the branching law for the symmetric group is available. If $(1^{\ell_1}, 2^{\ell_2}, 3^{\ell_3}, 3.)$ designates the irreducible representation of the group S_{N+1} with

$$\left(\sum_{n=1}^{N} n\ell_n = n! \right) ,$$

the associated Young pattern

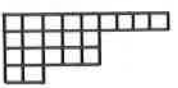


Fig. 2.6.

has ℓ_1 rows of length 1, ℓ_2 rows of length 2, etc. Only the two representations (N+1) and (N,1) can subduce the trivial representation of S_N. Furthermore, each representation subduces the trivial representation of S_n once and therefore

$$\sum_{g\in\mathfrak{g}} \chi^{\Gamma}(g) = (N)! \left[\delta_{\Gamma,(N+1)} + \delta_{\Gamma,(N,1)} \right] \qquad \text{(II.34)}$$

which yields the remarkably simple formula for $\mathfrak{M}_1$:

$$\mathfrak{M}_1 = \frac{|(N+1)|}{(N+1)} \left(\mathfrak{M}^{(N+1)}\right)^n + \frac{|(N,1)|}{N+1} \left(\mathfrak{M}^{(N,1)}\right)^n \qquad \text{(II.35)}$$

To illustrate the calculations, suppose that the probability of moving from one vertex of σ^N to any other in one step is $\frac{1}{N}$. The probability of group movements can be taken to have its support on elements of S_{N+1} which have cycle length N+1. There are N! elements of structure (N+1) of which (N-1)! induce the same motion on the simplex.

Now compute

$$\mathfrak{M}^{\Gamma} = \frac{1}{|\Gamma|} \sum_{g} \mathfrak{M}(g) \left[\chi^{\Gamma}(g)\right]^* \qquad \text{(II.36)}$$

use the theorem that if $g \in \kappa_{(N+1)}$ (the class of elements with a (N+1) cycle, then

$$\chi^{\Gamma}_{(N+1)} = \left\{ \begin{array}{l} (-)^{N_2} \text{ if } \Gamma = \\ 0 \text{ if } \Gamma \neq \end{array} \right\} (N_1, 1^{N_2}) \quad N_2 < N+1 \qquad \text{(II.37)}$$

$$\dim \Gamma^{(N_1, N_2)} = \frac{(N_1+N_2-1)!}{N_2!\,(N_1-1)!}$$

These are representations all of gamma form

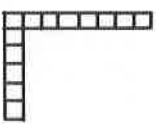


Fig. 2.7.

For the purposes of calculation of $\mathfrak{M}_1$ only the two representations $\Gamma = (N+1)$ and $\Gamma = (N,1)$ are needed:

$$\mathfrak{M}^{(N+1)} = \qquad |(N+1)| = 1$$

$$\mathfrak{M}^{(N,1)} = \frac{(-1)}{N} \qquad |(N,1)| = N \qquad \text{(II.38)}$$

Therefore

$$\mathfrak{M}_1 = \frac{1}{N+1} \left[1 + \frac{(-1)^n N}{N^n}\right] \qquad \text{(II.39)}$$

The remaining probabilities, probability of going from initial vertex to any distinct vertex is clearly

$$\mathfrak{M}_2 = \frac{1 - \mathfrak{M}_1}{N} \tag{II.40}$$

so that the posed problem is solved.

III. The Sphere Walk

In this section, the problem of a random walk on the surface of a sphere in three dimensional space will be posed and solved. The physical motivation for pursuing such a problem has been alluded to in the introduction: the multiple scattering of a particle in a medium composed of scatterers whose recoil is negligible. The tip of the velocity vector of the particle then executes a random walk on a sphere where the probability of moving thru polar angle θ and azimuthal angle φ is prescribed by the differential cross sections.

Group theoretic methods would be immediately applicable if the surface of a sphere were a group manifold. This, however, is impossible because of the confrontation of the two theorems:

a. Every square of a continuous mapping of the sphere surface into itself has a fixed point.[10]

b. The mapping of a group manifold into itself by a group operation other than the identity leaves no point fixed.

The proof of the latter follows immediately from the uniqueness of the identity operation of a group. If we note, however, that the sphere surface is a homogeneous space of the three dimensional rotation group manifold, a procedure is naturally suggested.

The rotation group, or more exactly, the covering group of the rotation group SU(2,C) has a group manifold the surface of a sphere in four dimensions. To see this, consider the transformations defining SU(2,C):

$$U \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Z_1' \\ Z_2' \end{pmatrix} \tag{III.1}$$

These leave invariant the form $|Z_1|^2 + |Z_2|^2$:

$$|Z_1|^2 + |Z_2|^2 = |Z_1'|^2 + |Z_2'|^2 \equiv R^2$$

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 = (x_1')^2 + (y_1')^2 + (x_2')^2 + (y_2')^2$$

$$Z_1 = x_1 + iy_1 \,;\; Z_2 = x_2 + iy_2 \tag{III.2}$$

and so correspond to rotations in four dimensions. Furthermore, the order of the group SU(3,C) is three, just the dimension of the surface of a sphere in four dimensions. All points of the sphere surface are reached only once since:

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \tag{III.3}$$

has the unique solution:

$$U_{11} = Z_1$$

$$U_{21} = Z_2$$

$$|U_{11}|^2 + |U_{21}|^2 = 1 \tag{III.4}$$

To solve the problem of Brownian motion on the surface of a four dimensional sphere, we avail ourselves of the group theoretical methods of the last section. The new feature encountered is a continuous group but, as is known, compact continuous groups have properties completely analogous to finite groups.

Let $\mathfrak{M}(\vec{\varphi})$ be the probability of the group movement labeled by $\vec{\varphi}$. A three dimensional picture of the SU(2,C) group manifold is a three dimensional sphere with the whole surface identified as a single point.

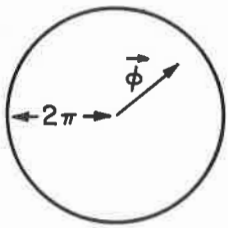


Fig. 3.1. Group manifold for SU(3,C).

The magnitude of $\vec{\varphi}$ and the angular specification of $\vec{\varphi}$ yield three parameters.

How many elements of the group SU(3,C) are there? The answer to this problem is that there is a unique (up to a constant) measure of the group manifold (the Hurwicz invariant measure) and is obtained by treating all group elements democratically.

If α and β are the polar and azimuthal angles of φ then

$$\sum_{g} \to \int d\sigma(y)$$

$$d\sigma(\vec{\varphi}) = d\alpha \sin\beta \, d\beta \sin^2 \frac{\varphi}{2} \, d\varphi \tag{III.5}$$

Integration over the group manifold yields the volume of the group $|G|$ which in this case is:

$$|G| \equiv |SU(2,C)| = \int_0^{\pi} d\alpha \int_0^{2\pi} \sin\beta \, d\beta \int_0^{2\pi} \sin^2 \frac{\varphi}{2} \, d\varphi = 4\pi^2 \tag{III.6}$$

the irreducible representations of $SU(2,C)$ are labeled by j and are $2j+1$ dimensional with $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$.

Then the group motion is analyzed by generalizing Eqs. (II.15) and (II.17):

$$\mathfrak{M}(\vec{\varphi}) = \sum_{jmm'} \frac{|j|}{|G|} \mathfrak{M}^j_{mm'} D^j_{mm'}(\vec{\varphi})$$

$$\mathfrak{M}^j_{mm'} = \int \mathfrak{M}(\vec{\varphi}) D^j_{m'm}(-\vec{\varphi}) \, d\sigma(\vec{\varphi})$$

$$|j| = 2j+1 \tag{III.7}$$

with the convolution of steps being resolved into a matrix multiplication: e.g.:

$$\mathfrak{M}^{(2)}(\varphi) = \int \mathfrak{M}(\varphi, \varphi' - 1) \, \mathfrak{M}(\varphi') d\sigma(\varphi')$$

$$\mathfrak{M}^j_{mm'}(2) = \sum_{m''} \mathfrak{M}^j_{mm''} \mathfrak{M}^j_{m''m'} \tag{III.8}$$

If the group motion depends only upon the class to which $\vec{\varphi}$ belongs then

$$\mathfrak{M}(\vec{\varphi}) = \sum_{j} \frac{|j|}{|G|} \mathfrak{M}^j \chi^j(\vec{\varphi})$$

$$\mathfrak{M}^j = \frac{1}{|j|} \int \mathfrak{M}(\varphi) \chi^*(\varphi) \, d\sigma(\varphi)$$

$$= \frac{4\pi}{|j|} \int_0^{2\pi} \mathfrak{M}(\varphi) \chi^*(\varphi) \sin^2\left(\frac{\varphi}{2}\right) d\varphi \tag{III.9}$$

where the characters of the group SU(2,C) are given by

$$\chi^j(\varphi) = \frac{\sin \frac{2j+1}{2}\varphi}{\sin \frac{\varphi}{2}} \tag{III.10}$$

Given $\mathfrak{M}(\varphi)$, the roots of the group motion $\mathfrak{M}^j$ can be found and the relaxation times computed since

$$\mathfrak{M}^j = \frac{4\pi}{|j|} \int_0^{2\pi} \mathfrak{M}(\varphi) \sin\left(\frac{2j+1}{2}\right)\varphi \sin\left(\frac{\varphi}{2}\right) d\varphi \tag{III.11}$$

which reduces the problem to quadrature.

As a specific example, suppose that $\mathfrak{M}(\varphi)$ is constant between 0 and $\bar{\varphi}$. When a short calculation yields

$$\mathfrak{M}^j = \frac{1}{|j|}\left(\frac{\frac{\sin j\bar{\varphi}}{j} - \frac{\sin(j+1)\bar{\varphi}}{j+1}}{\bar{\varphi} - \sin\bar{\varphi}}\right)$$

$$|j| = 2j+1 \tag{III.12}$$

The accompanying table gives some idea of how fast the $(\mathfrak{M}^j)^n$ go to zero. In the example, $\varphi = \frac{\pi}{6}$

j	$\mathfrak{M}^j$	$(\mathfrak{M}^j)^3$	$(\mathfrak{M}^j)^6$	$(\mathfrak{M}^j)^9$	$(\mathfrak{M}^j)^{12}$
0	1	1	1	1	1
$\frac{1}{2}$	0.979570	.939957	.883519	.830469	.780605
1	.946192	.847107	.717591	.607876	.514936
$\frac{3}{2}$	.900828	.731009	.534374	.390632	.285555
2	.844781	.602882	.363467	.219128	.132108
$\frac{5}{2}$	.779638	.473891	.224573	.106423	.050433
3	.707219	.353722	.125120	.044258	.015655
$\frac{7}{2}$	.629500	.249452	.062226	.015522	.003872
4	.548551	.165063	.027246	.004497	.000742
$\frac{9}{2}$	.466450	.101488	.010300	.001045	.000106
5	.385226	.057167	.003268	.000187	.000011
6	.233829	.012621	.000159	.000002	.000000
7	.104030	.001126	.000001	.000000	.000000
8	.071236	.000361	.000000	.000000	.000000
9	.054659	.000163	.000000	.000000	.000000
10	-0.0830301	-.000572	.000000	.000000	.000000

Table 3.1. Relaxation coefficients $\mathfrak{M}^j$ for the surface of a four-dimensional sphere (only six significant figures have been retained).

From scanning the table, the fast rate of which $(\mathfrak{M}^j)^n$ go to zero can be seen. Only six significant figures have been retained in the calculation. The sign change for $j = 10$ is again reversed at $j = 15$ so the $\mathfrak{M}^j$ do not start an excursion to any significant magnitude of negative values.

How can the properties of the SU(2,C) group walk be used to solve the problem of the random walk on the surface (S_2) of a three dimensional sphere? The answer follows from the theorem[11] that S_2 is the factor space of SO(3) by SO(2). Now SO(3) is a subgroup of SU(2,C) and a complete set on the group manifold is obtained by restricting the representations to integer j values. Furthermore, the integrations over the group manifold encountered in Eqs. (III.7), (III.8), (III.9) and (III.11) now have φ restricted to run from 0 to π.[12]

The example of tetrahedral motion in Section III points the way toward a complete solution of the sphere walk.

Let us assume that the probability of a vector $\hat{e}$ on the unit sphere moving through a polar angle θ is a prescribed function $f(\theta)$. Any azimuthal motion is equally probable.

$$\hat{e} \to \hat{e}' \qquad \hat{e}\cdot\hat{e}' = \cos\theta$$

$$f(\theta) = \text{probability of motion through angle } \theta \tag{III.13}$$

Under the group motion $\vec{\varphi}$, $\hat{e}$ undergoes the following transformation (see Fig. 3.1)

$$\hat{e} \to \hat{e}\cdot\hat{\varphi}\,\hat{\varphi} + (\hat{e} - \hat{e}\cdot\hat{\varphi}\,\hat{\varphi})\cos\varphi + (\hat{\varphi} \times \hat{e})\sin\phi = \hat{e}'$$

$$\hat{\varphi} = \frac{\vec{\varphi}}{|\varphi|} \tag{III.14}$$

which yields the following equation:

$$\cos\theta = \cos^2\alpha + (1 - \cos^2\alpha)\cos\varphi \tag{III.15}$$

If $\mathfrak{M}(\vec{\varphi})$ depends only upon $|\vec{\varphi}|$, it is clear that the probability of $\hat{e}'$ having any azimuthal orientation relative to $\hat{e}$ is independent of the azimuthal orientation, i.e., the probability for finding $\hat{e}'$ depends only upon θ where $\hat{e}\cdot\hat{e}' = \cos\theta$. Each $\mathfrak{M}(\varphi)$ determines an $f(\theta)$ and vice versa. Once $\mathfrak{M}(\varphi)$ is determined the spectrum of relaxation times found from $\mathfrak{M}^j$ are known. All that must be remembered is the restriction of φ to yield the SO(3,R) group manifold and the restriction to integral j values. The function $\mathfrak{M}(\varphi)$ is normalized to unity:

$$\int_0^{\pi} d\alpha \sin\alpha \int_0^{2\pi} d\beta \int_0^{\pi} d\varphi \sin^2\frac{\varphi}{2}\,\mathfrak{M}(\varphi) = 1 \tag{III.16}$$

Now change variables to $\theta[\alpha,\varphi]$ specified by Eq. (III.16):

$$\cos\varphi = \left(\frac{\cos\theta - \cos^2\alpha}{1-\cos^2\alpha}\right)$$

$$\sin^2\frac{\varphi}{2} = \frac{1-\cos\varphi}{2} = \tfrac{1}{2}\left[1 - \frac{\cos\theta - \cos^2\alpha}{1-\cos^2\alpha}\right]$$

$$d\varphi = \frac{\sin\theta\, d\theta}{\left[1 - \left(\frac{\cos\theta - \cos^2\alpha}{1-\cos^2\alpha}\right)^2\right]^{\frac{1}{2}}\left[1-\cos^2\alpha\right]} \tag{III.17}$$

which yields

$$\frac{2\pi}{2}\int_0^{\pi} d\alpha \int_0^{\pi} d\theta \frac{\sin\theta}{\sin\alpha}\left[\overline{\mathfrak{M}}\left(\frac{\cos\theta-\cos^2\alpha}{1-\cos^2\alpha}\right)\right]\left[\frac{1-\frac{\cos\theta-\cos^2\alpha}{1-\cos^2\alpha}}{1+\frac{\cos\theta-\cos^2\alpha}{1-\cos^2\alpha}}\right]^{\frac{1}{2}} = 1$$

$$\mathfrak{M}(\varphi) = \overline{\mathfrak{M}}(\cos\varphi) \tag{III.18}$$

from which we can read off $f(\theta)$:

$$f(\theta) = \pi\int_0^{\pi}\frac{d\alpha}{\sin\alpha}\left[\frac{1-\cos\theta}{1-2\cos^2\alpha+\cos\theta}\right]^{\frac{1}{2}}\overline{\mathfrak{M}}\left(\frac{\cos\theta-\cos^2\alpha}{1-\cos^2\alpha}\right) \tag{III.19}$$

The range of integration in general will not include the points $\alpha = 0$ and $\alpha = \pi$ because $\mathfrak{M}$ vanishes when its argument exceeds unity in magnitude.

$$\cos^2\alpha \le \frac{1+\cos\theta}{2} \tag{III.20}$$

and any singularity in $f(\theta)$ must be integrable since

$$\int f(\theta)\sin\theta\, d\theta = 1 \tag{III.21}$$

by Eq. (III.18). If $f(\theta)$ is specified, then Eq. (III.19) is an integral equation for $\overline{\mathfrak{M}}(\cos\varphi)$.

Suppose the example treated previously [Eq. (III.12)] is considered. It is easy to see geometrically and straight forward to prove analytically that $f(\theta)$ vanishes if $\theta > \bar{\varphi}$. Thus, the particular numerical solution exhibited in Table (3.1) applies for the relaxational Brownean

Motion on a sphere with $f(\theta)$ given by Eq. (III.19). Here $f(\theta)$ is only non-vanishing for $\theta < \bar{\varphi}$, as stated.

The probability density $f(\theta)$ which corresponds to $\mathfrak{M}(\varphi)$ constant up to $\bar{\varphi} = \frac{\pi}{6}$ and zero thereafter can be obtained by performing the integral of Eq. (III.19). The result is:

$$f(\theta) = \begin{cases} \dfrac{1}{\bar{\varphi} - \sin\bar{\varphi}} \tan^{-1}\left(\dfrac{\cos\theta - \cos\bar{\varphi}}{1 - \cos\bar{\varphi}}\right)^{\frac{1}{2}} & \theta \le \bar{\varphi} \\ 0 & \theta > \bar{\varphi} \end{cases} \qquad \text{(III.22)}$$

It is easy to check that $f(\theta)$ is non-negative and obeys Eq. (III.21). It is zero for $\theta = \frac{\pi}{6}$ and maximum for $\theta = 0$, a property which geometric visualization can yield. A nice check is obtained by considering $\mathfrak{M}(\varphi)$ constant over the group manifold of SO(3,R). The Eq. (III.22) yields $f(\theta) = \frac{1}{2}$ which clearly must be the case. A graph of $f(\theta)$ is given in Fig. 3.2 for $\bar{\varphi} = \frac{\pi}{6}$:

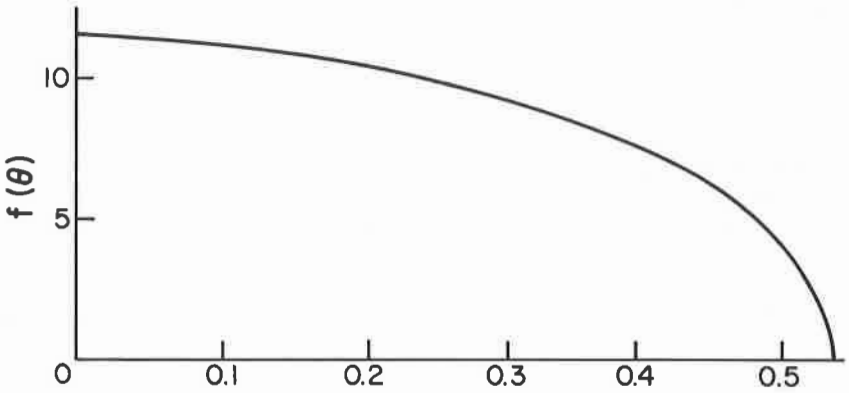


Fig. 3.2. The distribution $f(\theta)$ corresponding to a step function distribution on the group manifold.

The final step in solving the posed problem is the solution of the integral equation (Eq. (III.19)). This equation can be cast into the form

$$f(x) = \pi \int_{-1}^{x} dy \frac{1}{[(1+y)(x-y)]^{\frac{1}{2}}} \overline{\mathfrak{M}}(y)$$

$$x = \cos\theta, \quad y = \cos\varphi \qquad \text{(III.23)}$$

It thus appears as a Volterra integral equation of the first kind with a singular kernel. Fortunately, an exact analytic solution is available. It need only be noted that almost the same integral equation is encountered in the Abel problem.[13] The solution is:

$$\overline{\mathfrak{M}}(y) = \frac{(1+y)^{\frac{1}{2}}}{\pi^2} \frac{d}{dy}\left[\int_{-1}^{y} \frac{f(x)}{(y-x)^{\frac{1}{2}}}\, dx\right] \qquad \text{(III.24)}$$

As a check, one might note that $f(x) = \frac{1}{2}$ yields $\overline{\mathfrak{M}}(y) = \frac{1}{2\pi^2}$.

With the Eq. (III.24), the problem is solved. Given an $f(\cos\theta)$ for the randon step probability on a sphere, we solve for $\overline{\mathfrak{M}}(y) = \overline{\mathfrak{M}}(\cos\varphi) = \mathfrak{M}(\varphi)$ thereby obtaining the relaxation functions from Eq. (III.12). Next, $\mathfrak{M}^{(N)}(\varphi)$ is obtained from the equation.

$$\mathfrak{M}^{(N)}(\varphi) = \sum_j \frac{|j|}{|G|} [\mathfrak{M}^j]^N \chi^j(\varphi) \qquad \text{(III.25)}$$

and finally

$$f^{(N)}(x) = \pi \int_{-1}^{x} dy \frac{1}{[(1+y)(x-y)]^{\frac{1}{2}}} \overline{\mathfrak{M}}^{(N)}(y) \qquad \text{(III.26)}$$

Of course, the quantities of most physical interest are the $\mathfrak{M}^j$ since these control the relaxation rate and enable the infinite sum of Eq. (III.25) to be truncated.

The procedure developed appears to be a useful method in solving the isotropization rate in multiple scattering problems.

IV. Polya Walks and the Clebsch-Gordan Problem

A polya walk is a random walk on a lattice. An example of this type of walk has already been given in Section III--the walk on a line with $f(x) = \frac{1}{2}[\delta(x-a) + \delta(x+a)]$. In this section, a relation between the Polya walk problem and the Clebsch-Gordan problem for groups is exhibited. To some degree, the purpose of this section is more recreational than practical, although the connection between random walks and group coupling problems may serve a heuristic purpose.

To begin, let $\mathfrak{M}(\ell)$ be the probability of making a step of length ℓ, $\ell = 0, \pm1, \pm2, \ldots$. The group is the translation group in a one dimensional lattice, C_∞. The representations of C_∞ are given by

$$D^{(k)}(\ell) = e^{ik\ell} \qquad -\pi < k \leq \pi \qquad \text{(IV.1)}$$

and so

$$\mathfrak{M}(\ell) = \int_{-\pi}^{+\pi} \hat{\mathfrak{M}}(k)\, e^{ik\ell}\, \frac{dk}{2\pi}$$

$$\hat{\mathfrak{M}}(k) = \sum_{n} \mathfrak{M}(\ell)\, e^{-ik\ell} \qquad \text{(IV.2)}$$

Note that the k lie on a circle which can also be taken to be a group manifold--the group manifold of the character group of C_∞ which in this case is $U(1,c)$.[14]

The n step probability $\mathfrak{M}^{(n)}(\ell)$ is easily found.

$$\mathfrak{M}^{(n)}(\ell) = \int_{-\pi}^{+\pi} [\hat{\mathfrak{M}}(k)]^n\, e^{ik\ell}\, \frac{dk}{2\pi} \qquad \text{(IV.3)}$$

The special choice

$$\mathfrak{M}(\ell) = \begin{cases} \frac{1}{2} & \ell = 1, -1 \\ 0 & \ell \neq 1, -1 \end{cases}$$

$$\hat{\mathfrak{M}}(k) = \cos k = \frac{e^{ik} + e^{-ik}}{2} \qquad \text{(IV.4)}$$

suggests that for the symmetric walk $\mathfrak{M}(\ell) = \mathfrak{M}(-\ell)$ the harmonic transforms can be expressed in terms of the characters of $SU(2,C)$

$$\chi^j(\varphi) = \sum_{m=-j}^{m=j} e^{im\varphi}$$

$$j = 0, \pm\tfrac{1}{2}, \pm 1, \ldots \qquad \text{(IV.5)}$$

This is indeed the case

$$\hat{\mathfrak{M}}(k) = \sum_{j=-\infty}^{j=+\infty} [\mathfrak{M}(2j) - \mathfrak{M}(2j+2)]\chi^j(2k) \qquad \text{(IV.6)}$$

The relationship with the Clebsch-Gordan series for $SU(2,C)$ now presents itself, for in the calculation of $\hat{\mathfrak{M}}(k)$, the product $\chi^j(2k)\chi^{j'}(2k)$ is encountered

$$\chi^j(2k)\chi^{j'}(2k) = \sum C^{j''}_{jj'}\chi^{j''}(2k) \qquad \text{(IV.7)}$$

Here the coefficients $C^{j''}_{jj}$, are the coefficients of the Clebsch-Gordan series.

Conversely; if, for example, the result of coupling n spin j particles together is desired, one can solve the random walk problem with $\mathfrak{M}(2j) = \mathfrak{M}(2j-2) = \ldots = \mathfrak{M}(-2j) = \frac{1}{2j+1}$ as the assigned probabilities and then read off the net spins resulting by using

$$C^{j}_{j^n} = (2j+1)^n \left[\mathfrak{M}^{(n)}(2j) - \mathfrak{M}^{(n)}(2j+2)\right] \qquad \text{(IV.8)}$$

As an example, consider the coupling of n spin $\frac{1}{2}$ particles together. The problem is one of a random walk with non-vanishing probabilities resting solely on $\mathfrak{M}(1) = \mathfrak{M}(-1) = \frac{1}{2}$. The n-step probabilities are given by

$$\mathfrak{M}^{(n)}(\ell) = \frac{1}{2^n} \binom{n}{\frac{n-\ell}{2}} : \frac{n-\ell}{2} \text{ integral} \qquad \text{(IV.9)}$$

and so the spectrum of spins resulting have weights

$$C^{J_n}_{\frac{1}{2}} = \binom{n}{\frac{n-2J}{2}} - \binom{n}{\frac{n-2J-2}{2}} \qquad \text{(IV.10)}$$

A result used in the Heisenberg theory of ferromagnetism.[15]

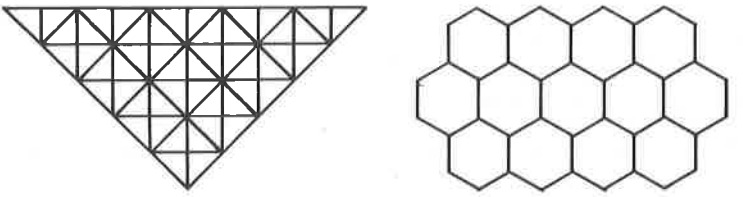
For the coupling of n spin j particles, the answer is obtained upon evaluating the integral

$$\mathfrak{M}^{(n)}(\ell) = \int_{-\pi}^{+\pi} \left[\frac{1}{L+1} \frac{\sin(L+1)k}{\sin k}\right]^n e^{+ik\ell} \frac{dk}{2\pi}$$

$$L = 2j \qquad \text{(IV.11)}$$

and then using Eq. (IV.8)

We turn now from these rather trivial remarks to more interesting problems--Polya walks in a two dimensional lattice. The lattices which will be of interest are of the type which are illustrated in Fig. 4.1.



(a) Triangle lattice (b) Chicken wire lattice

Fig. 4.1. Lattices for Polya Walks

Let the probability of a step $\vec{\ell}$ be $\mathfrak{M}(\vec{\ell})$. Again the now obvious formulae follow:

$$\mathfrak{M}(\ell) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathfrak{M}(\vec{k})\, e^{i\vec{k}\cdot\vec{\ell}}\, \frac{d\vec{k}}{(2\pi)^2}$$

$$\mathfrak{M}(\vec{k}) = \int \mathfrak{M}(\vec{\ell})\, e^{-i\vec{k}\cdot\vec{\ell}}\, d\vec{\ell} \qquad \text{(IV.12)}$$

From here on, the procedure will be carried out informally. Suppose that the probability of a step is $\frac{1}{3}$ for each of three directions $\hat{n}_1$ $\hat{n}_2$ and $\hat{n}_3$. Then $\mathfrak{M}(\vec{\ell})$ has the form

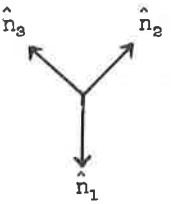


Fig. 4.2. The equiprobable directions of the 3-walk.

$$\mathfrak{M}(\vec{\ell}) = \tfrac{1}{3}[\delta(\vec{\ell}-\hat{n}_1) + \delta(\vec{\ell}-\vec{n}_2) + \delta(\vec{\ell}-\vec{n}_3)] \qquad \text{(IV.13)}$$

and therefore

$$\mathfrak{M}(k) = \frac{\left(e^{i\vec{k}\,\vec{n}_1} + e^{i\vec{k}\,\vec{n}_2} + e^{i\vec{k}\,\vec{n}_3}\right)}{3} \qquad \text{(IV.14)}$$

which is proportional to the character of the three dimensional representation of SU(3,C). The n step walk now involves the coupling problem for the quark representation of SU(3,C), which will yield the Polya walk probabilities in the lattice of Fig. 4.1.a.

To solve the Clebsch-Gordan problem for $[\chi^{(3)}(\vec{k})]^n$, the method of the Young tableau is available. The procedure involves three steps:

i. find the type of irreducible representations in the Kronecker product of the quark representation with itself n times.

ii. find the multiplicity of the representation.

iii. find the contents of the irreducible representations; namely, what lattice vectors $\vec{r}_i$ for each representation (which are found in the character for each representation: $x = \sum_i e^{i\vec{k}\cdot\vec{r}_i}$. Fortunately, the problems involved in these steps have been worked over extensively.[16]

Step (i) is solved by the graphical technique of writing down all Young patterns[17] with n boxes. Each pattern has three rows of boxes with the number of boxes in each row not less than the number of boxes in the following row. Table 4.2 illustrates the Young patterns for ≥ 6.

The pair $(\lambda_1\lambda_2)$ above each pattern indicates the number of columns with one member each and the number of columns with two members each. In terms of $(\lambda_1\lambda_2)$ the dimension of the representation is given by

$$d(\lambda_1\lambda_2) = \frac{(\lambda_1+1)(\lambda_2+1)(\lambda_1+\lambda_2+2)}{2} \qquad \text{(IV.15)}$$

This number is given below each pattern in Table 4.1.

How many different Young patterns are there for a given n? A little reflection shows that this number $\mathfrak{N}(n)$ is equal to the number of tripartite partitions of n, viz, the number of different ways (without regard to order) in which n can be written as the sum of three non-negative integers. A generating function for $\mathfrak{N}(n)$ is G(x):

$$G(x) = \frac{1}{1-x}\,\frac{1}{1-x^2}\,\frac{1}{1-x^3} = \sum_{n=0}^{\infty} \mathfrak{N}(n)\, x^n \qquad \text{(IV.16)}$$

which yields for (n) after a small calculation:

$$\mathfrak{N}(n) = \frac{(n+1)(n+5)}{12} + \tfrac{1}{4}\left(\frac{1+\cos 2\pi}{2}\right) + \tfrac{1}{3}\left(\frac{1+2\cos\frac{2\pi}{3}n}{3}\right) \qquad \text{(IV.17)}$$

This expression can be checked against Table 4.1.

Step (ii) involves counting the multiplicity of the representation which, in this case, is just the dimension of a representation of the symmetric group on n objects associated with a Young pattern. If the multiplicities are denoted by $\mu(\lambda_1\lambda_2\lambda_3)$ (with λ_3 the number of columns with three members each) then

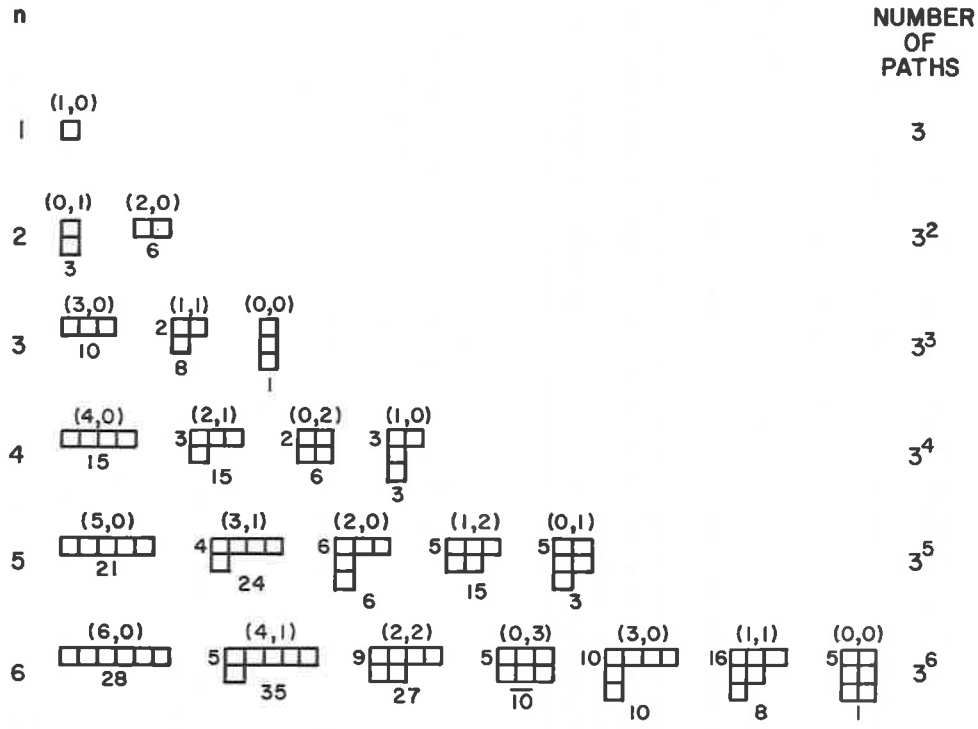


Table 4.1. Young Patterns.

$$\mu(\lambda_1\lambda_2\lambda_3) = \frac{n!(\lambda_1+1)(\lambda_1+\lambda_2+2)(\lambda_2+1)}{(\lambda_1+\lambda_2+\lambda_3+2)!(\lambda_2+\lambda_3+1)\lambda_3!}$$

$$= \frac{n!}{(\lambda_1+\lambda_2+\lambda_3+2)!(\lambda_2+\lambda_3+1)!\lambda_3!}\ d(\lambda_1,\lambda_2) \qquad \text{(IV.18)}$$

As a check of Eq. (IV.18) and Eq. (IV.15), one can verify that

$$3^n = \sum_{\text{patterns}} \mu(\lambda_1\lambda_2\lambda_3)\, d(\lambda_1\lambda_2) \qquad \text{(IV.19)}$$

In Table 4.1, the multiplicities $\mu(\lambda_1\lambda_2\lambda_3)$ are given to the left of each pattern.

Finally, the content of each representation must be found (step (iii)). The solution of this problem is best obtained graphically and is described in the literature.[18]

To illustrate the process of the random walk, the results of visited sites and the relative probability of visit is given in Fig. 4.3.

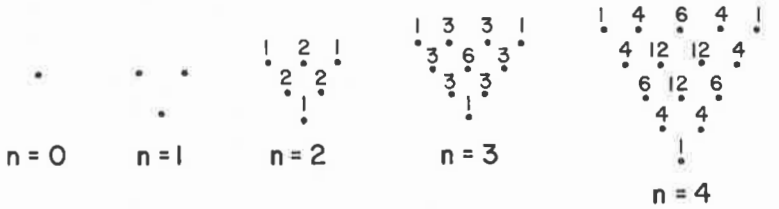


Fig. 4.3.

For this type of Polya walk, three sublattices are distinguished and labeled by the triality quantum number:

$$(\lambda_1+\lambda_2) \equiv t \pmod 3 \qquad \text{(IV.20)}$$

A particle executing a random walk of the type just finished being discussed will visit each sublattice every third move.

The problem of the Polya walk on a chicken wire is interesting because the probability of a step $\vec{\ell}$ is different for even and odd steps. Fig. 4.4 gives the two alternative equiprobable set of directions.

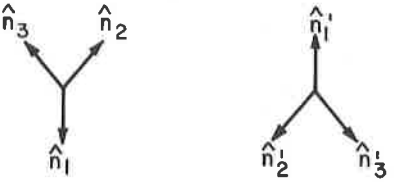


Fig. 4.4. Alternative Equiprobable Sets of Directions.

The solution involves computing $[x^{(3)}]^{[\frac{n+1}{2}]} x^{(\bar{3})[\frac{n}{2}]}$ on step n where $x^{(3)}$ and $x^{(\bar{3})}$ are the characters of the 3 and $\bar{3}$ irreducible representations of SU(3,C) on [x] is the integral part of x. Again three steps are necessary to solve the problem of the chicken wire walk--find the irreducible representations in $D^{(3)[\frac{n+1}{2}]} \otimes D^{(\bar{3})[\frac{n}{2}]}$, their multiplicities and their content. A few manipulations yield the brief results of Table 4.2, which we leave to the pleasure of the reader to complete.

n	YOUNG PATTERN	NUMBER OF PATHS
1	□	3
2	⊓ |	3^2
3	^{2}⊓ ⊞ ⊓	3^3
4	⊓⊓ ^{4}⊞ ⊓⊓ ⊞ ^{2}⊞	3^4

Table 4.2. Young Patterns for Chicken Wire Walk

The number of boxes in the nth step Young patterns is $[\frac{n+1}{2}]+2[\frac{n}{2}]$.

The interplay between the concept of the Polya walk and the coupling problem for compact groups (other groups yield lattices other than the one considered) is heuristic at best. One possible use is to obtain the asymptotic distribution of the number of representations obtained by multiple couplings. For large n, the Polya walk on a lattice can be approximated by the solution of a diffusion problem controlled by the relevant Fokker-Planck equation.

Conclusions and Outlook

The main contents of these notes are the new methods for treating relaxation processes embodied in Sections II and III. It has turned out that the group theoretical methods yield powerful tools for finding the spectrum of relaxation times for a wide variety of processes. It should be obvious that for very diverse problems, group structures are available which reduce the problem to the canonical form presented here. In particular, multiple scattering problems should prove quite tractable once the basic symmetry is recognized.

A topic which was not mentioned but which lies close to the ideas here is the "path integral" formulation of stochastic movements. Feynman's notion of path integral in quantum theory was preceded by such path integrals used in diffusion processes.[19] Some of the group methods used here can be used to treat various path integral problems; e.g., the dipole moment distribution of a polymer. Just how powerful the methods will prove to be remains for the future to determine.

Appendix A: The Group Ring

The treatment of the group movement probability $\mathfrak{M}(g)$ in Section III can be cast into a form of a group ring.

$$\mathfrak{M} \equiv \sum_g \mathfrak{M}(g)g \tag{A.1}$$

Then $\mathfrak{M}$ is a group algebra operation with the multiplication property

$$\mathfrak{M}^2 = \sum_g \mathfrak{M}(g)g \sum_{g'} \mathfrak{M}(g')(g') = \sum_g \left[\sum_{g'} \mathfrak{M}(gg'^{-1})\, \mathfrak{M}(g') \right] g$$

$$= \sum_g \mathfrak{M}^{(2)}(g)g \tag{A.2}$$

Clearly, the generalization holds

$$\mathfrak{M}^n = \sum_g \mathfrak{M}^{(n)}(g)g \tag{A.3}$$

The procedure of taking the harmonic transform of $\mathfrak{M}(g)$ corresponds to forming the right modules of the group ring. A basis for a right module is given by

$$e^{\Gamma}_{ij} = \sum_g D^{\Gamma}_{ji}(g^{-1})g \tag{A.4}$$

The dimension of this right module is $|\Gamma|^2$ and its main property is that

$$\mathfrak{M}(g)e^{\Gamma}_{ij} = \sum_k \left[\sum_{g'} \mathfrak{M}(g')\, D^{\Gamma}_{ki}(g'^{-1}) \right] e^{\Gamma}_{kj} = \sum_k \hat{\mathfrak{M}}^{\Gamma}_{ik}\, e^{\Gamma}_{kj} \tag{A.5}$$

viz, $\mathfrak{M}(g)$ induces an endomorphism of the module M. Furthermore, it is easy to find

$$e^{\Gamma'}_{mn}\, e^{\Gamma}_{ij} = \delta_{\Gamma\Gamma'}\, \delta_{mi}\, \delta_{nj} \tag{A.6}$$

from which follows

$$e^{\Gamma'}\, e^{\Gamma} = |\Gamma|\, \delta_{\Gamma\Gamma'}\, e^{\Gamma}; \sum_i e^{\Gamma}_{ii} \equiv e^{\Gamma} \tag{A.7}$$

and displays theorthogonality and idempotency of the operators

$$I^{\Gamma} \equiv \frac{e^{\Gamma}}{|\Gamma|}$$

$$I^{\Gamma} I^{\Gamma'} = \delta_{\Gamma\Gamma'}\, I^{\Gamma} \tag{A.8}$$

It is these properties which are behind the simplification ensuing from using the harmonic transform of $\mathfrak{M}(g)$.

Many properties of group rings are available in the literature.[6]

Appendix B: The Structure of the Tetrahedral Groups A_4

The tetrahedral group was selected for intensive study to illustrate concretely the basic concepts and operations of "group walks." Abstractly, the tetrahedral group is identical with the alternating group A_4 on four objects, hence the notation.

(α) Presentation

The tetrahedral group is generated by two operations g_{V_1} and g_{E_1} subject to the relations

$$V^3 = 1 \qquad E^2 = 1 \qquad (VE)^3 = 1$$

$$g_{V_1} \equiv V, \quad g_{E_1} \equiv E, \quad g_O \equiv 1 \tag{B.1}$$

In terms of these generators the group operations are:

g_O	g_{V_1}	$g_{V_1}^{\ 1}$	g_{V_2}	$g_{V_2}^{\ 1}$	g_{V_3}	$g_{V_3}^{\ 1}$	g_{V_4}	$g_{V_4}^{\ 1}$	g_{E_1}	g_{E_2}	g_{E_3}
1	V	V^2	VE	EV^2	VE	V^2E	V^2EV^2	VEV	E	V^2EV	VEV^2

Table B.1. Group Elements in Terms of Generators

and the group table is quickly constructed with their use. The Cayley diagram[20] associated with the group presentation is the Schlegel diagram for a truncated tetrahedron Fig. B.1.

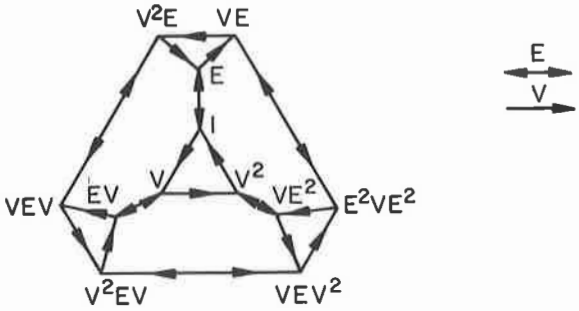


Fig. B.1. Cayley Diagram for A_4

(β) Subgroups

The group consisting of g_O, g_{E_1}, g_{E_2}, g_E is an Abelian invariant subgroup, the group D_2, with $A_4/D_2 =$. Four subgroups of order 3 are generated by rotations about each of the vertices (C_3 groups) while the edge rotations generate three subgroups of order 2. (C_2 groups)

(γ) Classes

Four classes of elements are present: the identity class C_0 and the classes of rotation thru $\frac{2\pi}{3}$, $\frac{4\pi}{3}$ and π, denoted by C_v, $C_{v'}$ and C_E. These latter have a membership of 4, 4 and 3 respectively.

(δ) Representations

Since the group A_4 is itself a finite rotation group, one representation is a three dimensional representation with characters $1 + 2\cos\varphi$ where g is the rotation specified by the group element. It is irreducible and has the characters given in Table B.2.

	$1\ C_0$	$4\ C_v$	$4\ C_{v'}$	$3\ C_E$
Γ_3	3	0	0	-1

Table B.2. Characters of the $_3$ Representation

Besides the trivial one dimensional representation, two other one dimensional representations are present. These can be obtained by inspection using the orthogonality property of group characters and are given in Table B.3.

	$1\ C_0$	$4\ C_v$	$4\ C_{v'}$	$3\ C_E$
Γ_3	3	0	0	-1

Table B.3. Characters of the Γ_1 and Γ_2 Representations

(ε) Induced Representation[21]

There is an intimate connection between induced representations and the walk on the tetrahedron. Let the little group be the group of rotations about one vertex and the "orbiting" operations be g_{E_1}, g_{E_2} and g_{E_3}. Then a four dimensional induced representation is found which decomposes into $D^{(0)}$ and $D^{(3)}$ if the trivial representation of the little group is used. A moment of reflection indicates that the matrices of this induced representation are sufficient to represent completely the stochastic matrix governing the random walk on the tetrahedron. The details of inducing a representation go as follows. Let the little group consist of the operations $\{g_0\ g_{v_1}\ g_{v_1'}\} \equiv L$. Then every element of A_4 can be written uniquely in the form

$$g = o_\beta \ell_\alpha$$

$$o_\beta \in \{g_o, g_{E_1}, g_{E_2}, g_{E_3}\}$$

$$\ell_\alpha \in \{g_o\ g_{v_1}\ g_{v_1}'\} \tag{B.2}$$

Now let β be a vertex index and define

$$o_\beta | 1\,[\Gamma_\ell]\rangle \equiv |\,[\Gamma_\ell]\rangle$$

$$\ell = \text{representation index for little group} \tag{B.3}$$

We also have

$$\ell_\alpha | 1\,[\Gamma_\ell]\rangle = D^{(\Gamma_\ell)}(\ell_\alpha) | 1[\Gamma_\ell]\rangle \tag{B.4}$$

A four dimensional representation is clearly obtained since

$$g|\beta[\Gamma_\ell] = go_\beta | 1[\Gamma_\ell]\rangle = o_{\beta'}\ell_{\alpha'}|[\Gamma_\ell]\rangle$$
$$= D^{(\Gamma_\ell)}(\ell_{\alpha'},)|\beta'[\Gamma_\ell]$$

$$go_\beta = o_{\beta'}\ell_{\alpha'} \tag{B.5}$$

One need only check that $D(g'g) = D(g')D(g)$ to assure that one has a representation.

(ζ) Automorphisms

The set of inner automorphisms in A_4 is isomorphic to the group A_4. There is a wider set of automorphisms Aut A_4 which leads to the full group of the tetrahedron: Aut $A_4 = S_4$ where S_4 is the group of permutations on four objects. The automorphisms of the group indicate the symmetry of the group manifold.

References

1. For a mathematical treatment of stochastic processes, the reader is referred to:
   a. William Feller, An Introduction to Probability Theory and Its Applications (John Wiley, New York, 1967).
   b. Frank Spitzer, Principles of Random Walk (D. Van Nostrand Co., Princeton, New Jersey, 1964).

c. Emanuel Parzen, Stochastic Processes (Holden-Day, San Francisco, 1962).
d. Marck Fisz, Probability Theory and Mathematical Statistics (John Wiley, New York, 1963).
e. E. B. Dynken and V. A. Uspenskii, Random Walks (D. C. Heath, New York, 1963).

2. The theory of Brownian motion can be studied from the following references:
   a. Albert Einstein, Investigations on the Theory of Brownian Motion (Dover, New York, 1956).
   b. Nelson Wax, ed., Selected Papers on Noise and Stochastic Processes (Dover, New York, 1954).
   c. Mark Kac, Probability and Related Topics in Physical Sciences (Interscience Publishers, New York, 1959).
3. These Markov diagrams are used effectively to summarize the properties of Markov processes. See, for example, Ref. 1e.
4. The group theoretic background needed for Sections III and IV can be culled from any of the following references:
   a. Eugene Wigner, Group Theory (Academic Press, New York, 1959).
   b. M. Hamermesh, Group Theory (Addison-Wesley, Reading, Mass., 1962).
5. The concept of homogeneous space is developed in:
   a. Robert Hermann, Lee Groups for Physicists (W. A. Benjamin, New York, 1966).
   b. S. Helgason, Differential Geometry and Symmetric Spaces (Academic Press, New York, 1962).
6. Group algebras have an extensive literature:
   a. Martin Burrow, Representation Theory of Finite Groups (Academic Press, New York, 1965).
   b. M. A. Naimark, Normed Rings (P. Noordloff, Groningen, 1964), Chap. VI.
7. For some interesting properties of logarithmic spirals, see D'Arcy Thompson, On Growth and Form (Cambridge University Press, New York, 1915).
8. The Burnside lemma is presented in Frank Harary, Graph Theory and Theoretical Physics (Academic Press, New York, 1967).
9. The presentation of a group is covered in H. S. M. Coxeter and W. O. Moser, Generators and Relations for Discrete Groups (Springer-Verlag, Berlin, 1965).
10. For the relevant fixed point theorems, the reader is referred to Donald Blackett, Elementary Topology (Academic Press, New York, 1967), p. 151.

11. The homogeneous spaces of the classical group are to be found in Claude Chevalley, Theory of Lie Groups (Princeton University Press, Princeton, New Jersey, 1946).
12. A derivation of the invariant group volume is given in Ref. 12b, p. 325 ff.
13. The solution of this singular integral equation can be found using slight generalizations of methods given by F. G. Tricomi, Integral Equations (Interscience, New York, 1957).
14. The character group has been discussed extensively by L. Pontrjagin, Topological Groups (Princeton Press, Princeton, N. J., 1946).
15. For the Heisenberg theory of ferromagnetism, the reader is referred to J. H. VanVleck, Theory of Electric and Magnetic Susceptibilities (Oxford University Press, Clarendon, 1932), chap. 12.
16. A review of the simple groups with some new methods is given by R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. 34, 1 (1962).
17. The theory of Young tableaux is found in Ref. 4b.
18. The graphical methods have been developed by the author and C. Fronsdal and are reported in C. Fronsdal in Brandeis Lectures, Vol. I (W. A. Benjamin, New York, 1963), p. 510.
19. The use of path integrals in quantum theory is due to Feynman, R. Feynman and A. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill, New York, 1965).
20. An exposition of Cayley diagrams can be found in Ref. 17.
21. The theory of induced representations is found in Ref. 12b and Ref. 13.

INDUCED REPRESENTATION THEORY AND THE POINCARÉ GROUP†

William H. Klink
University of Iowa
Iowa City, Iowa

I. Introduction

Group theory has been found to be a valuable tool for the physicist in his attempt to understand the structure of elementary particles; in the process of using group theory, however, certain typical problems of a mathematical nature arise. Thus, given a group, the physicist is usually interested in knowing what the unitary irreducible representations of the group are, and the spaces--usually Hilbert spaces--on which these representations act; he wishes to know how to carry out the decomposition of an irreducible representation of the group into irreducibles of a subgroup and how to carry out the reduction of a tensor product into irreducible representations of the group. Now for some classes of groups the answers to most of these problems can readily be found. Thus, for the compact Lie groups all the unitary irreducible representations have been classified and there are general --although often complicated--techniques for handling most of the mathematical questions raised by the physicist.[1)]

The purpose of this series of lectures will be to show that for those groups which have irreducible representations which can be written as induced representations, that most questions of interest to the physicist can be answered using techniques developed by G. W. Mackey.[2)] It will be shown that the decomposition of an irreducible induced representation of a group into irreducibles of a subgroup, the reduction of n-fold tensor products, the explicit evaluation of Clebsch Gordan coefficients, possibly the evaluation of "D" functions of a group, all can be handled by the theory of induced representations developed by Mackey.

But the question immediately arises as to how large the class of groups is which has irreducible representations which can be written as induced representations. Even restricting the irreducible representations to unitary representations acting on Hilbert spaces,

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

the number of such groups is large. There are, however, two classes of groups of particular interest in physics, namely those groups having invariant subgroups and the noncompact simple groups. It should be emphasized that, as the word "induced" suggests, induced representations of a group are built up from certain supposedly known irreducible representations of subgroups; but these subgroups are very often either the compact Lie groups or Abelian groups whose irreducible representations are known.

For that class of groups having invariant subgroups, Mackey has shown that all the unitary irreducible representations of such groups can be written as induced representations.[2] Two subclasses of these groups have been used extensively in physics, namely the semidirect product groups (having Abelian invariant subgroups), and the disconnected groups. Probably the best known example of a semidirect product group, a group which will be dealt with in great detail in Sec. IV, is the Poincaré group; other examples include the Galilean group and the Euclidean group. An example of a disconnected group is the infinite dihedral group, consisting of all rotations about an axis plus a 180° rotation about the perpendicular axis.

While Mackey has shown that all irreducible unitary representations of groups having invariant subgroups can be written as induced representations, the same is not true of the noncompact simple groups. It may be that all irreducible unitary representations of the noncompact simple groups can be generated via induced representations acting on suitably defined spaces, but there are certainly classes of irreducible representations such as the discrete series which cannot be written as induced representations. In these lectures only the so-called principal series of irreducible representations of noncompact simple groups will be discussed, since the Hilbert spaces on which the principal series act are quite simply defined. For other classes of irreducible representations of the noncompact simple groups which can be written as induced representations, see Refs. 3 and 4.

Probably the best known example of a noncompact simple group in physics is the group SL(2,C), the covering group of the Lorentz group. It will be discussed in some detail in Sec. II--which summarizes the relevant parts of Mackey's induced representation theory--where it will be used to illustrate some of the ideas of induced representation theory. Sec. III will discuss the connection between the abstract spaces on which the unitary operators of induced representation theory act, and the spaces of interest to the physicist. Finally, Sec. IV will study the Poincaré group in detail from the point of view of induced representation theory.

II. Mathematical Preliminaries

To define an induced representation, consider a (separable, locally compact)[5] group G and a (closed) subgroup H. Let L(h) be a representation of H acting on a Hilbert space $\mathcal{H}(L)$. Consider functions f which carry elements g of G into the space $\mathcal{H}(L)$ with the property that f(h g) = L(h) f(g). This forms the vector space $\mathcal{H}(U^L)$:

$$\mathcal{H}(U^L) = \{ f \mid f(g) \in \mathcal{H}(L),\ f(hg) = L(h)f(g)\ \forall h \in H, g \in G\} \tag{II.1}$$

which is made into a Hilbert space by restricting f(g) so that

$$||f||^2 = \int_{G/H} d\mu(g)\ ||f(g)||^2_{\mathcal{H}(L)} < \infty \tag{II.2}$$

where $d\mu(g)$ is the measure over cosets G/H.

Since f is an element of $\mathcal{H}(L)$, the notation $||f(g)||^2_{\mathcal{H}(L)}$ designates the norm of f taken as an element of $\mathcal{H}(L)$.

The induced representation U^L then is defined by its action on elements of the Hilbert space $\mathcal{H}(U^L)$:

$$U^L(g_o)f(g) = \sqrt{\rho(g, g_o)}\ f(gg_o) \tag{II.3}$$

where g_o is an arbitrary but fixed element of G, and f is in $\mathcal{H}(U^L)$. The notation U^L is meant to indicate the unitary representation of G induced by the representation L of H. The function $\rho(g, g_o)$ is chosen so that U^L will in fact be unitary. This means that

$$\begin{aligned} ||U^L(g_o)f||^2 &= \int_{G/H} d\mu(g)\ ||U^L(g_o)\ f(g)||^2_{\mathcal{H}(L)} \\ &= \int_{G/H} d\mu(g)\ \rho(g, g_o)||f(gg_o)||^2_{\mathcal{H}(L)} \\ &= \int_{G/H} d\mu(gg_o^{-1})\ \rho(gg_o^{-1}, g_o)||f(g)||^2_{\mathcal{H}(L)} \end{aligned} \tag{II.4}$$

If the coset measure $d\mu(g)$ were invariant under G, so that $d\mu(g\ g_o) = d\mu(g)$, then the function $\rho(g, g_o)$ would be set equal to one; however, in general $d\mu(g)$ is not invariant under G, but only quasi-invariant (that is, sends sets of measure zero into sets of measure zero[2]) and the function $\rho(g, g_o)$ is so chosen to compensate for the lack of invariance of the coset measure under the action of G. Thus, $d\mu(gg_o^{-1})\ \rho(gg_o^{-1}\ g_o) = d\mu(g)$ and

$$||U^L(g_o)f||^2 = \int_{G/H} d\mu(g)\ ||f(g)||^2_{\mathcal{H}(L)}$$

$$= ||f||^2 \qquad \text{(II.5)}$$

The induced representation U^L can therefore be made unitary by an appropriate choice of $\rho(g, g_o)$.

The question then arises--under what conditions can the induced representations defined by (II.3) be made irreducible? The answer depends on the class of groups being considered.

Consider first the class of groups having invariant subgroups. Mackey has shown that if one extends the notion of representation to include projective representations, then all unitary irreducible representations of this class of groups can be written as induced representations. Since only "true" or vector representations will be considered in these lectures, it is sufficient to consider a subclass which includes semidirect product groups with invariant Abelian subgroups, and disconnected groups. Semidirect product groups are defined as groups whose elements can be written uniquely as

$$g = k\, g_c \qquad \text{(II.6)}$$

with g in G, k in K, the invariant subgroup, and g_c an element of G/K. By a theorem of Pontriagin,[6] a disconnected group G always has as an invariant subgroup the connected part of G--called K again--while it is assumed that G/K is isomorphic to a finite group.

For these two classes of groups, the inducing subgroups are obtained using little group theory.[7] Thus, let M be a representation of the above mentioned invariant subgroup K. Then the inducing subgroup H of G is given by

$$H = \{g \mid g \in G,\ M(gkg^{-1}) \cong M(k)\ \forall\ k \in K\} \qquad \text{(II.7)}$$

[The symbol $\cong$ means "is equivalent to."] Some further conditions must generally be put on the representations of H; these conditions can be found in Refs. 2 and 7.

The inducing subgroups for the second class of groups having unitary irreducible representations which can be written as induced representations, namely the noncompact simple groups, are obtained from the Iwasawa decomposition,[8] written as

$$G = N\, A\, K_m \qquad \text{(II.8)}$$

where K_m is the maximal compact subgroup of G, A is the maximal noncompact Abelian subgroup of G, and N is a nilpotent subgroup of G. Define the centralizer C(A) of G to be that set of elements of K_m which commute with A:

$$C(A) = \{k_m \mid k_m \in K_m,\ k_m A = A k_m\} \tag{II.9}$$

Then the principal nondegenerate series of G is induced by the subgroup

$$H = N\,A\,C(A) \tag{II.10}$$

while the principal degenerate series of G are induced by an even larger subgroup than (II.10), obtained by putting restrictions on A so that C(A) includes more elements of K_m. For a more complete survey of how these various inducing subgroups are obtained, see Refs. 3 and 4.

As an example, consider the group SL(2,C), the covering group of the homogeneous Lorentz group. SL(2,C), the set of 2 x 2 complex matrices with determinant one, can be written

$$SL(2,C) = \left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right\},\ a_{11}a_{22} - a_{12}a_{21} = 1$$

$$= N\,A\,(SU(2)) \tag{II.11}$$

where SU(2) is the maximal compact subgroup, A is the set of matrices of the form $\begin{pmatrix} |a_{22}|^{-1} & 0 \\ 0 & |a_{22}| \end{pmatrix}$ and N is the set of matrices of the form $\begin{pmatrix} 1 & a_{12} \\ 0 & 1 \end{pmatrix}$. The centralizer C(A) is seen to be that set of elements of SU(2) of the form $\begin{pmatrix} \frac{a_{22}^*}{|a_{22}|} & 0 \\ 0 & \frac{a_{22}}{|a_{22}|} \end{pmatrix}$, that is, $\begin{pmatrix} e^{-i\varphi} & 0 \\ 0 & e^{i\varphi} \end{pmatrix}$, with $e^{i\varphi} = \frac{a_{22}}{|a_{22}|}$. Then the subgroup which induces the principal nondegenerate series of SL(2,C) is

$$H = \left\{\begin{pmatrix} a_{22}^{-1} & a_{12} \\ 0 & a_{22} \end{pmatrix}\right\} \tag{II.12}$$

which has representations

$$h \to L^{[m\rho]}(a_{22}, a_{12}) = |a_{22}|^{i\rho} \left(\frac{a_{22}}{|a_{22}|}\right)^m \tag{II.13}$$

with ρ real and m an integer.

The induced representation $U^{[m\rho]}(SL(2,C))$ is then obtained from (II.1) and (II.3). Naimark[9] has chosen a particularly simple form for the coset representatives G/H, choosing elements $\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$. Then the induced representation $U^{[m\rho]}$ acts on functions $f(z)$, with norm

$$||f||^2 = \int dz\, |f(z)|^2 < \infty \tag{II.14}$$

The measure over coset elements, $dz \equiv dxdy$ is readily obtained since the set of elements $\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$ form an Abelian subgroup of $SL(2,C)$. However, it is also readily seen that under the action of an arbitrary group element of $SL(2,C)$, the measure dz does not remain invariant. In fact the coset representative $\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$ gets sent into the representative $\begin{pmatrix} 1 & 0 \\ z' & 1 \end{pmatrix}$, where $z' = \dfrac{a_{11}z+a_{21}}{a_{12}z+a_{22}}$. Under such a transformation Naimark[9] shows that dz is transformed into

$$dz' \equiv dx'dy' = |a_{12}z+a_{22}|^{-4}\, dz \tag{II.15}$$

Therefore the function $\rho(g,g_0)$ of (II.3) must be chosen to be $|a_{12}z+a_{22}|^4$ in order that the induced representation $U^{[m\rho]}$ be unitary.

Given the definition of an induced representation of the group G acting on the Hilbert space $\mathcal{H}(U^L)$, it is possible to state the key theorem which makes it possible to carry out the tensor product decomposition of induced representations and hence provide a means of obtaining Clebsch-Gordan coefficients. This theorem is called the subgroup theorem by Mackey, and, as has been done in the preceding parts of this section, only a sketch of the relevant ideas will be presented. The person interested in the measure theoretic complications is urged to read the original work by Mackey.

Thus, consider two (closed) subgroups H and H′ of G with L a representation of H on the space $\mathcal{H}(L)$. Ignoring any measure theoretic difficulties (such as the regular relatedness of the groups H and H′), and setting $\rho(g,g_0)$ equal to one for simplicity (in order to focus on the algebraic aspects of the theorem) consider the induced reducible representation $U^L(H')$ to see how it reduces into invariant subspaces. To this end consider the double coset decomposition

$$G = \bigcup_D H\, g_D\, H' \tag{II.16}$$

where g_D is a set of elements of G labeling the double cosets $H\, g_D\, H'$ and define the functions

$$f_D(g) = f(g) \qquad g \in H\, g_D\, H'$$
$$= 0 \qquad g \notin H\, g_D\, H' \tag{II.17}$$

Now the set $\{f_D(g)\}$ forms a subspace $\mathcal{K}_D(U^L)$ of $\mathcal{K}(U^L)$ invariant under $U^L(H')$ since, if g is in the double coset $H\, g_D\, H'$ so is gh' and therefore $U^L(H')\, f_D(g)$ is in $\mathcal{K}_D(U^L)$. The Hilbert space $\mathcal{K}(U^L)$ is thus seen to break up into a direct sum (integral) of subspaces $\mathcal{K}_D(U^L)$ labeled by double cosets.

Mackey then shows that on a subspace $\mathcal{K}_D(U^L)$, fixed by the double coset label D, that the representation $U^L(H')$ is equivalent to an induced representation, induced by a subgroup of H'. To see this define a mapping $\tilde{f}_D$ carrying elements of H' into $\mathcal{K}(L)$ with the property that $\tilde{f}_D(h') = f_D(g_D h')$. Then

$$f_D(hg_D h') = L(h)\, f_D(g_D h')$$
$$= L(h)\, \tilde{f}_D(h') \tag{II.18}$$

shows that $\tilde{f}_D$ and f_D define one another uniquely.

Now consider the subgroup

$$H_D = g_D^{-1}\, H\, g_D \cap H' \tag{II.19}$$

It is not hard to see that H_D induces representations of H' on the subspace $\mathcal{K}_D(U^L)$. Thus let h_D be an element of H_D. Then

$$\begin{aligned}
\tilde{f}_D(h_D h') &= \tilde{f}_D(g_D^{-1} h g_D h') \\
&= f_D(hg_D h') \\
&= L(h)\, f_D(g_D h') \\
&= L(h)\, \tilde{f}_D(h') \\
&= L(g_D h_D g_D^{-1})\, \tilde{f}_D(h') \\
&= L_D(h_D)\, \tilde{f}_D(h')
\end{aligned} \tag{II.20}$$

where $L_D(h_D) \equiv L(g_D h_D g_D^{-1})$ is a representation of H_D inherited from H via (II.19). Thus, the functions $\tilde{f}_D(h')$ are of the form (II.1), with norm $||\tilde{f}_D||^2 = \int_{H'/H_D} d\mu(h')\, ||\tilde{f}_D(h')||^2_{\mathcal{K}(L_D)}$, and the induced representation $U^L(H')$ on $\mathcal{K}_D(U^L)$ is seen to be equivalent to the induced

representation $U^{L_D}(H')$, induced by the representation L_D of H_D.

The subgroup theorem shows how an induced representation of G breaks up into induced representations of H', a subgroup of G. In Sec. III it will be shown how the subgroup theorem is used to find the coefficients which actually reduce out the reducible representation $U^{L_D}(H')$ for generally $U^{L_D}(H')$ is not irreducible so to arrive at irreducible representations of H', a further reduction is generally needed.

In their reduction of irreducible representations of the group SL(2,C) into representations of SU(1,1), Sciarrino and Toller[10] have made use of the subgroup theorem. The group H of the subgroup theorem is chosen to be the subgroup inducing the principal series of SL(2,C) namely the group (II.12). H' is chosen to be SU(1,1) so that it is necessary to calculate the double cosets

$$SL(2,C) = \bigcup_D H\, g_D\, SU(1,1) \tag{II.21}$$

or

$$\left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right\} = \left\{\begin{pmatrix} p^{-1} & q \\ 0 & p \end{pmatrix}\right\} g_D \left\{\begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}\right\} \tag{II.22}$$

where $p \neq 0$, p, q, α and β are complex, and $|\alpha|^2 - |\beta|^2 = 1$ (the notation is that of Sciarrino and Toller). One double coset representative is conveniently chosen to be the identity element $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and upon multiplying elements of H and SU(1,1) together in (II.22), it is clear that $|p| = \sqrt{|a_{22}|^2 - |a_{21}|^2}$ which means that the identity double coset will not cover elements of SL(2,C) of the type $|a_{21}| > |a_{22}|$. It is not hard to show that elements with $|a_{21}| > |a_{22}|$ are covered by a second double coset which has as a convenient representative the element $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

After checking that these two double cosets indeed cover all elements of SL(2,C) it is necessary to find what the group H_D, Eq. (II.19), is. A simple calculation shows that

$$H_D = U(1) = \left\{\begin{pmatrix} \frac{\alpha}{|\alpha|} & 0 \\ 0 & \frac{\alpha^*}{|\alpha|} \end{pmatrix}\right\}$$

which has representations $\left(\frac{\alpha}{|\alpha|}\right)^m$ for the identity double coset and $\left(\frac{\alpha}{|\alpha|}\right)^{-m}$ for the other double coset, inherited from the representation of the inducing subgroup of SL(2,C), Eq. (II.13). The subgroup H_D does not induce irreducible representations of SU(1,1) and a further reduction is needed which will be discussed in Sec. III.

The subgroup theorem can be used to decompose the tensor products of induced representations. To show this, consider the outer product group $\{(g_1, g_2, \ldots, g_n)\}$ of ordered n-tuples of elements of G. If $H_1, H_2, \ldots H_n$ are inducing subgroups of G, then the outer product group has representations induced by the outer product inducing group $(H_1, H_2, \ldots, H_n)$, this outer product inducing group having representations $L_1 \otimes L_2 \otimes \ldots \otimes L_n$ acting on the tensor product space $\mathcal{H}(L_1) \otimes \mathcal{H}(L_2) \otimes \ldots \otimes \mathcal{H}(L_n)$. The induced representations of the outer product group act on functions $F(g_1, g_2, \ldots, g_n)$ which form the Hilbert space $\mathcal{H}(U^{L_1 \otimes L_2 \otimes \ldots \otimes L_n})$.

If the induced representation $U^{L_1 \otimes L_2 \ldots \otimes L_n}(g_1, g_2, \ldots g_n)$ is restricted to elements of the diagonal subgroup $G = \{(g, g, \ldots, g)\}$ for g in G, the subgroup theorem is directly applicable. Thus, the functions $F(g_1, g_2, \ldots, g_n)$ will be decomposed into functions labeled by double cosets:

$$(g_1, g_2, \ldots, g_n) \in (H_1, H_2, \ldots, H_n)(g_{D_1}, g_{D_2}, \ldots g_{D_n})(G, G, \ldots, G) \tag{II.23}$$

with $g_{D_1}, g_{D_2}, \ldots g_{D_n}$ elements of G. Then according to the subgroup theorem the group inducing representations of the diagonal group $(G, G, \ldots, G)$ is

$$\begin{aligned} H_D &= (g_{D_1}, g_{D_2}, \ldots, g_{D_n})^{-1} (H_1, H_2, \ldots, H_n)(g_{D_1}, g_{D_2}, \ldots, g_{D_n}) \\ &\qquad \cap (G, G, \ldots, G) \\ &= (g_{D_1}^{-1} H_1 g_{D_1}, g_{D_2}^{-1} H_2 g_{D_2}, \ldots, g_{D_n}^{-1} H_n g_{D_n}) \cap (G, G, \ldots, G) \\ &= (g_{D_1}^{-1} H_1 g_{D_1} \cap G, g_{D_2}^{-1} H_2 g_{D_2} \cap G, \ldots, g_{D_n}^{-1} H_n g_{D_n} \cap G) \end{aligned} \tag{II.24}$$

which has representations inherited from $L_1 \otimes L_2 \otimes \ldots \otimes L_n$. The n-fold tensor product has then been reduced to an induced representation of G given by the group (II.24). Again it will be shown in Sec. III how the use of "D" functions on the group G enables one to actually obtain the Clebsch-Gordan coefficients which carry out reduction.

As an example of a tensor product decomposition, consider again the group SL(2,C); as was shown in Eq. (II.12), the principal series of representations of SL(2,C) are induced by the subgroup $H = \left\{\begin{pmatrix} a_{22}^{-1} & a_{12} \\ 0 & a_{22} \end{pmatrix}\right\}$ with H having representations $|a_{22}|^{ip}\left(\frac{a_{22}}{|a_{22}|}\right)^m$. The induced representation is defined on the Hilbert space of functions

f(z), Eq. (II.14). Then the tensor product of two irreducible representations $[m_1\rho_1]$ and $[m_2\rho_2]$ of SL(2,C), $U^{[m_1\rho_1]\otimes[m_2\rho_2]}$ acts on the functions $F(z_1, z_2)$ of the Hilbert space

$$\mathcal{H}(U^{[m_1\rho_1]\otimes[m_2\rho_2]}) = \{F(z_1, z_2) \mid ||F||^2 = \int dz_1 dz_2 |F(z_1, z_2)|^2 < \infty\} \tag{II.25}$$

and the problem is to reduce this Hilbert space into a direct integral of Hilbert spaces invariant under SL(2,C).

To this end consider the double coset decomposition of the outer product group

$$(SL(2,C)_1, SL(2,C)_2) = \bigcup_D (H_1, H_2)(g_D, e)(SL(2,C), SL(2,C)) \tag{II.26}$$

with H_1 and H_2 both of the form (II.12). It is not hard to see that there are two double cosets whose representatives can be chosen to be $g_D = e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $g_D = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. When $g_D = e$ is chosen, the inducing subgroup of the diagonal group, Eq. (II.24), is seen to be the subgroup of Eq. (II.12), and thus, for this choice of double coset labels an irreducible induced representation is obtained. However, for for the choice $g_D = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ a simple calculation using (II.24) shows that the inducing subgroup is $\left\{\begin{pmatrix} |a_{22}|^{-1} & 0 \\ 0 & |a_{22}| \end{pmatrix}\right\}$, so that the induced representation of SL(2,C) generated from this double coset is reducible. In fact because it is reducible the representation induced when $g_D = e$ can be ignored since it occurs only on a set of measure zero in the decomposition of the reducible representation generated by the other double coset. This further decomposition has been carried out by Naimark.[11]

It is to be noted that the subgroup H_D of H', Eq. (II.19), can be no smaller than the identity subgroup. But the identity subgroup induces the so called regular representation, where $U^1(g_o)$ acts on functions f(g) which have norm

$$||f||^2 = \int_G d\mu(g) |f(g)|^2 < \infty$$

and $d\mu(g)$ is right invariant Haar measure. The "1" on $U^1(g_o)$ indicates the 1 dimensional representation of the identity subgroup.

For semidirect product groups with Abelian invariant subgroups, the reduction of the regular representation into irreducibles can always be carried out. Therefore, for such groups and other groups for which the reduction of the regular representation is known, the subgroup theorem is decisive.

III. Relation of Induced Representation Theory to Quantities of Interest in Physics; the D Functions

The connection between induced representation theory --in particular the abstract spaces on which the induced representations act-- and the spaces of physical interest can be made via the D functions. Before defining the D functions it is useful to observe that the functions f(g), elements of the space $\mathcal{K}(U^L)$, Eq. (II.1), can be thought of as functions over cosets because of the condition f(hg) = L(h)f(g); for if G is decomposed into right cosets with respect to the inducing subgroup H,

$$G = \bigcup_c H\, g_c \tag{III.1}$$

with g_c a set of elements of G labeling the right cosets, then in particular, $g = hg_c$ for g in G and h in H, so that

$$f(g) = L(h)f(g_c)\ . \tag{III.2}$$

The induced representation then acts on $f(g_c)$ in the following way:

$$\begin{aligned} U^L(g_o)f(g_c) &= \sqrt{\rho(g_c, g_o)}\, f(g_c g_o) \\ &= \sqrt{\rho(g_c, g_o)}\, f(hg_{c'}) \\ &= \sqrt{\rho(g_c, g_o)}\, L(h)f(g_{c'}) \end{aligned} \tag{III.3}$$

where use has been made of (III.1) to write that $g_c g_o = hg_{c'}$. Thus, the element g_o of G moves the coset representative g_c to the representative $g_{c'}$. Eq. (III.3) is a basic equation showing how functions over cosets transform under the action of an arbitrary group element.

For example, in the group SL(2,C) Naimark[9] chooses the elements $\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}$, z complex, as coset representatives and considers the the action of a group element on functions over cosets f(z); the action $z \to z' = \frac{a_{11}z+a_{21}}{a_{12}z+a_{22}}$ is obtained from the equation

$$\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = h \begin{pmatrix} 1 & 0 \\ z' & 1 \end{pmatrix} \tag{III.4}$$

where h is an element of the inducing subgroup H, Eq. (II.12).

Given the transformation properties of functions over cosets, the connection between induced representation theory and quantities of physical interest is made through the use of D functions, defined as

$$U(g_o)|[\chi]x\rangle = \int_{x'} D^{[\chi]}_{x'x}(g_o)|[\chi]\ x'\rangle \ . \tag{III.5}$$

More generally

$$D^{[\chi]}_{x_o x}(g) \equiv \langle[\chi]\ x_o|U(g)|[\chi]\ x\rangle \tag{III.6}$$

for a general element g in G. $[\chi]$ denotes a set of labels specifying irreducible representations of G, while x and x′ are "eigenvalues" of a complete set of commuting elements of the Lie algebra of G. x_o is also a set of "eigenvalues," but in general of a different set of commuting elements of the Lie algebra of G. $U(g_o)$ is a unitary operator of G acting on the (in general improper) basis states $|[\chi]x\rangle$. The symbol $\int$ means an integration or summation over the variables x′, depending on whether they are discrete or continuous (or both).

Consider the operator

$$U(g_o)\ D^{[\chi]}_{x_o x}(g) \equiv D^{[\chi]}_{x_o x}(gg_o) \tag{III.7}$$

In particular restrict the elements of G in the D functions to coset representatives $D^{[\chi]}_{x_o x}(g_c)$. Then $D^{[\chi]}_{x_o x}(g_c)$ can be thought of, on the one hand, as a concrete realization of the basis state $|[\chi]x\rangle$, for

$$\begin{aligned} U(g_o)\ D^{[\chi]}_{x_o x}(g_c) &= D^{[\chi]}_{x_o x}(g_c g_o) \\ &= \int_{x'} D^{[\chi]}_{x_o x'}(g_c)\ D^{[\chi]}_{x'x}(g_o) \\ &= \int_{x'} D^{[\chi]}_{x'x}(g_o)\ D^{[\chi]}_{x_o x'}(g_c) \end{aligned} \tag{III.8}$$

which agrees with the transformation properties of the state $|[\chi]x\rangle$, Eq. (III.5).

On the other hand, for specified values of x_o, $D^{[\chi]}_{x_o x}(g_c)$ may also be thought of as a function over cosets (although $D^{[\chi]}_{x_o x}(g_c)$ will not, in general, be an element of $\mathcal{H}(U^L)$, but rather will be in the rigged Hilbert space[12] of $\mathcal{H}(U^L)$), with transformation properties

$$U(g_o)\, D^{[\chi]}_{x_o x}(g_c) = D^{[\chi]}_{x_o x}(g_c g_o)$$

$$= \sum_{x_o'} D^{[\chi]}_{x_o x_o'}(h)\, D^{[\chi]}_{x_o' x}(g_c') \,, \qquad \text{(III.9)}$$

where $g_c g_o = h g_c'$.

The question is whether $D^{[\chi]}_{x_o x_o'}(h)$ can be chosen so that it is equal to the matrix element of the operator $L(h)$. To see that this is possible, choose the labels x_o so that they include in the complete set of commuting elements of the Lie algebra of G, the elements of the inducing subgroup H. Label the irreducible representations of H by χ_H and the diagonal elements of H by x_H. Then x_o includes χ_H and x_H. Any other elements of x_o will be called $\tilde{x}_o$. Then x_o and x_o' can be chosen so that

$$U(g_o) D^{[\chi]}_{x_o x}(g_c) = \sum_{x_o'} D^{[\chi]}_{x_o x_o'}(h)\, D^{[\chi]}_{x_o' x}(g_c')$$

$$= \sum_{\tilde{x}_o' \chi_H' x_H'} \langle [\chi] \tilde{x}_o \chi_H x_H | U(h) | [\chi] \tilde{x}_o' \chi_H' x_H' \rangle D^{[\chi]}_{x_o' x}(g_c')$$

$$= \sum_{x_H'} \langle [\chi] \tilde{x}_o \chi_H x_H | U(h) | [\chi] \tilde{x}_o \chi_H x_H' \rangle\, D^{[\chi]}_{x_o' x}(g_c')$$

$$= \sum_{x_H'} D^{[\chi_H]}_{x_H x_H'}(h)\, D^{[\chi]}_{(\tilde{x}_o \chi_H x_H') x}(g_c') \,. \qquad \text{(III.10)}$$

It only remains to verify that the irreducible representation $[\chi_H]$ appears in the irreducible representation space of G labeled by $[\chi]$. To see whether this is the case use the subgroup theorem, letting $H = H'$. Consider any double coset other than the identity double coset; then $H_D = g_D^{-1} H g_D \cap H$ and the Frobenius reciprocity theorem[13] can be used to see whether the representation of H induced by H_D contains the irreducible representation χ_H.

With the labels x_o fixed by (III.10) so that $D^{[\chi]}_{x_o x}(g_c)$ transforms properly, (III.10) then also fixes the normalization of the states $|[\chi] x\rangle$. For the norm of functions over cosets, Eq. (II.2) is known and thus,

$$\langle[\chi]x'|[\chi]x\rangle = \left(D^{[\chi]}_{x_o x'}, , D^{[\chi]}_{x_o x}\right)$$
$$= \int_{G/H} d\mu(g_c)\left(D^{[\chi]}_{x_o x'}(g_c), D^{[\chi]}_{x_o x}(g_c)\right)_{\mathcal{K}(L)} \tag{III.11}$$

Thus, it is seen that the D functions play a dual role, acting as both a concrete realization of $|[\chi]x\rangle$ and as a function over cosets which transforms properly.

It is to be noted that often the D functions can be obtained from knowledge of the transformation properties of the induced representation. Consider, for example, the group SL(2,C); how does a state $|[m\rho]JM\rangle$, where the labels J,M come from the SU(2) subgroup of SL(2,C), transform under the action of an element of SL(2,C)? To answer this question consider a right coset decomposition of SL(2,C) with respect to its inducing subgroup, Eq. (II.12), choosing not $\left|\begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}\right|$ as right coset representatives, but elements from $\frac{SU(2)}{U(1)}$. Now SU(2) can be written as $\begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix}$, $|u|^2+|v|^2=1$, so that representatives of the cosets $\frac{SU(2)}{U(1)}$ can be written $\begin{pmatrix} |u| & v \\ -v^* & |u| \end{pmatrix}$ and thus, the right coset decomposition of SL(2,C) is now written

$$SL(2,C) = \bigcup_v H \begin{pmatrix} |u| & v \\ -v^* & |u| \end{pmatrix} \tag{III.12}$$

with every element of SL(2,C) in a unique right coset specified by v. Further, the square integrable functions over cosets f(v) have norm

$$||f||^2 = \int_{\frac{SU(2)}{U(1)}} d\mu(v)\ |f(v)|^2 < \infty \tag{III.13}$$

and the action of an arbitrary group element of SL(2,C) is computed from (III.3). (See Naimark, Ref. 9, p. 156.)

Define the kernel transformation

$$\varphi(JM) = \int_{\frac{SU(2)}{U(1)}} d\mu(v)\ D^{J\,*}_{mM}(v)\ f(v) \tag{III.14}$$

Then since the action of a group element of SL(2,C) on f(v) is known, using the orthonormality of the SU(2) D functions allows one to calculate the action of the group on the elements $\varphi(J,M)$. But this gives precisely the SL(2,C) D functions, for

$$U(g_o)\varphi(J,M) = \sum_{J'M'} D^{[m\rho]}_{JM\,J'M'}(g_o)\,\varphi(J',M') \qquad \text{(III.15)}$$

where g_o is an element of SL(2,C).

Quite generally if an appropriate kernel as in (III.13) can be found, then the D functions for the group can be found. Note that it is actually necessary only to compute the action of an element of the form $\begin{pmatrix} D^{-1} & 0 \\ 0 & D \end{pmatrix}$ of SL(2,C) on the "wave functions" $\varphi(J,M)$; for it is possible to decompose SL(2,C) into double cosets with respect to SU(2),

$$SL(2,C) = \bigcup_D SU(2)\, g_D\, SU(2)$$

and it is not hard to see that g_D can be chosen to be of the form $\begin{pmatrix} D^{-1} & 0 \\ 0 & D \end{pmatrix}$. But

$$D^{[m\rho]}_{JM\,J'M'}(R_1 g_D R_2) = \langle [m\rho]JM|\, U(R_1)U(g_D)U(R_2)\,|[m\rho]J'M'\rangle \qquad \text{(III.16)}$$

where R_1 and R_2 are elements of SU(2), and the action of R_1 and R_2 on J,M and J',M' is known and results in SU(2) D functions.

It is clear that if the D functions of the group G have transformation properties as functions over cosets (III.3) then all of the results of Sec. II can be utilized. For example, the subgroup theorem may be used to decompose an induced irreducible representation into a reducible representation of a subgroup. Using the supposedly known D functions of both the group and its subgroup, it is possible to calculate the coefficients which carry out the reduction. Thus, consider the D function of the group G, written as $D^{[\chi]}_{x_o x}(g_D h_c)$ where g_D is a double coset label and h_c is a coset representative of H'/H_D, where H_D is defined by Eq. (II.19):

$$H' = \bigcup_c H_D\, h_c \qquad \text{(III.17)}$$

Choose $\tilde{f}_D(h_c)$, defined in Eq. (II.18) to be the (in general improper) D function $D^{[\chi']}_{y_o y}(h_c)$, where

$$D^{[\chi']}_{y_o y}(h_c) = \langle [\chi']y_o|\, U(h_c)\,|[\chi']\, y\rangle \qquad \text{(III.18)}$$

is assumed to have the correct transformation properties, Eq. (II.20) and $[\chi']$ labels irreducible representations of H'. Then the coefficients which carry out the reduction of G into H' are the overlap between the irreducible function $D^{[\chi']}_{y_o y}(h_c)$ and the reducible function $D^{[\chi]}_{x_o x}(g_D h_c)$:

$$\langle [\chi']y;y_o x_o D|[\chi]x\rangle = \left(D^{[\chi']}_{y_o y}\,,\, D^{[\chi]}_{x_o x}\right)$$

$$= \int_{H'/H_D} d\mu(h_c)\left(D^{[\chi']}_{y_o y}(h_c),\, D^{[\chi]}_{x_o x}(g_D h_c)\right)_{\mathcal{H}(L_D)}; \tag{III.19}$$

the labels y_o and x_o are of course fixed, but also function as degeneracy labels in the event that the reduction of an irreducible representation of G into irreducibles of H' is not multiplicity free.

Consider again the example of the reduction of the irreducible representation of SL(2,C) into irreducibles of SU(1,1), discussed in Sec. II. There are two relevant classes of irreducible representations of SU(1,1) called the discrete and principal series; both of these representations will be discussed in Sec. IV in connection with the spacelike representations of the Poincaré group. Suffice it to say that the D functions of SU(1,1) can be written $D^{[s\pm]}_{i_o i}(SU(1,1))$, with s discrete or continuous, and i_o and i both discrete. Then assuming that the D functions of SU(1,1) transform properly as functions over cosets $\frac{SU(1,1)}{U(1)}$, labeled by β, Eq. (II.22), it is possible to write, for the identity double coset,

$$\langle [s\pm]i;i_o J_o M_o|[m\rho]JM\rangle$$

$$= \int_{\frac{SU(1,1)}{U(1)}} d\mu(\beta) D^{[s\pm]*}_{i_o i}(\beta)\, D^{[m\rho]}_{J_o M_o\, JM}(\beta) \tag{III.20}$$

In the analysis used by Toller, no multiplicity arises, so that the labels i_o, J_o and M_o are irrelevant. However, several examples will be given in Sec. IV where multiplicity does occur, and then such labels will be relevant.

Finally, to obtain Clebsch-Gordan coefficients arising in the reduction of an n-fold tensor product one uses n D functions and goes through the procedure outlined in Sec. II, Eq. (II.23)ff. After choosing double coset representatives, a reducible induced representation usually results, induced by the subgroup

$$H_D = (g_{D_1}^{-1} H_1 g_{D_1} \cap G, g_{D_2}^{-1} H_2 g_{D_2} \cap G, \ldots, g_{D_n}^{-1} H_n g_{D_n} \cap G) \qquad \text{(II.24)}$$

with representations inherited from $L_1 \otimes L_2 \otimes \ldots \otimes L_n$. Choosing the function $F(g_{c_1}, g_{c_2}, \ldots, g_{c_n})$ to be the n products

$$D^{[\chi_1]}_{x_{o_1} x_1}(g_{c_1})\, D^{[\chi_2]}_{x_{o_2} x_2}(g_{c_2}) \ldots D^{[\chi_n]}_{x_{o_n} x_n}(g_{c_n})$$

and the function $\tilde{f}_D(\tilde{g}_c)$ to be the D function $D^{[\chi]}_{x_o x}(\tilde{g}_c)$, where $\tilde{g}_c$ labels right cosets of G with respect to H_D and x_o is chosen so that $D^{[\chi]}_{x_o x}(\tilde{g}_c)$ transforms in the same manner as $\tilde{f}_D(\tilde{g}_c)$, gives Clebsch-Gordan coefficients

$$\langle [\chi] x; x_o, x_{o_1}, x_{o_2}, \ldots, x_{o_n}, \{D\} \,|\, [\chi_1] x_1; [\chi_2] x_2; \ldots, [\chi_n] x_n \rangle$$

$$= \int_{G/H_D} d\mu(\tilde{g}_c) \left(D^{[\chi]}_{x_o x}(\tilde{g}_c), D^{[\chi_1]}_{x_{o_1} x_1}(g_{D_1} \tilde{g}_c) \ldots D^{[\chi_n]}_{x_{o_n} x_n}(g_{D_n} \tilde{g}_c) \right)_{\mathfrak{K}(L_1 \otimes \ldots \otimes L_n)}$$

(III.21)

where again $x_o, x_{o_1}, x_{o_2} \ldots, x_{o_n}$, and $\{D\}$ label possible degeneracies arising in the n-fold reduction.[14] An example of formula (III.21) will be given in Sec. IV. Suffice it to say that (III.21) is sufficiently general to handle even those reductions which are not multiplicity free. This is in contrast to methods such as that of Anderson, et al,[14] where projection operators are used to obtain the Clebsch-Gordan coefficients.

IV. Analysis of the Poincaré Group Using Induced Representation Theory.

Moussa and Stora[15] have shown how one can use Mackey's induced representation theory to analyze the Poincaré group. In this section the covering group of the Poincaré group will be analyzed and it is thus useful to have a matrix form of the covering group of the Poincaré group.

Now, the covering group of the Lorentz subgroup of the Poincaré group is SL(2,C). Further, the action of an element Λ of SL(2,C) on a space-time translation a is known:[9]

$$H(a') = \Lambda\, H(a)\, \Lambda^{+} \qquad \text{(IV.1)}$$

where H(a) is a Hermitian matrix of the form

$$H(a) = \begin{pmatrix} a_t + a_z & a_x - ia_y \\ a_x + ia_y & a_t - a_z \end{pmatrix} \tag{IV.2}$$

Quite obviously H(a′) is also Hermitian and linearly related to H(a).

How then are these Poincaré group elements to be written as a matrix? It is not hard to see that a suitable choice is

$$P = \left\{ \begin{pmatrix} \Lambda & H(a)\,\Lambda^{-1+} \\ 0 & \Lambda^{-1+} \end{pmatrix} \right\} \tag{IV.3}$$

and that elements of P combine in the correct way.

In order to write out the induced representations of P it is first necessary to find the inducing subgroups. It is well known from little group theory[7] (Eq. (II.7)) that the inducing subgroups consist of all translations {H(a)} plus those Lorentz transformations satisfying

$$\Lambda\, H(p)\, \Lambda^{+} = H(p) \quad . \tag{IV.4}$$

p is a momentum vector which arises as an irreducible representation label in $e^{ip\cdot a}$ of the invariant Abelian translations ($p\cdot a = Ea_t - \vec{p}\cdot\vec{a}$).

There are essentially three classes (not including $p \equiv 0$) of energy momentum vectors p satisfying (IV.4):

1. $p^2 > 0$ or $|H(p)| > 0$; then a standard vector for these timelike representations can be chosen to be $p_o = (M, \vec{O})$, or $H(p_o) = M \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = M\, I_2$. The little group is $\Lambda\Lambda^{+} = I_2$ which is the definition of SU(2), and the inducing subgroup of the timelike representations is thus

$$H = \begin{pmatrix} SU(2) & \{H(a)\}SU(2) \\ 0 & SU(2) \end{pmatrix} \tag{IV.5}$$

which has representations

$$H \longrightarrow e^{ip_o\cdot a}\, D^{[s]}_{i\,i'}(SU(2)) \tag{IV.6}$$

where $D^{[s]}_{i\,i'}(SU(2))$ are the irreducible matrix elements of SU(2) (Wigner functions).

2. $p^2 = 0$, choose $p_o = (\frac{1}{2}, 0, 0, \frac{1}{2})$ as a standard energy momentum vector for this lightlike class of representations. Then $H(p_o) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\Lambda H(p_o)\Lambda^{+} = H(p_o)$ gives, since elements of SL(2,C) can be written out explicitly:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a_{11}^* & a_{21}^* \\ a_{12}^* & a_{22}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} |a_{11}|^2 & a_{11}a_{21}^* \\ (a_{11}a_{21}^*)^* & |a_{22}|^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{(IV.7)}$$

so that $a_{21} = 0$, $|a_{11}| = 1$, and thus, the set of elements of SL(2,C) leaving $H(p_o)$ invariant is of the form $\begin{pmatrix} \frac{a_{11}}{|a_{11}|} & a_{12} \\ 0 & \frac{a_{11}^*}{|a_{11}|} \end{pmatrix}$ which can also be written

$$\overline{E(2)} \equiv \left\{ \begin{pmatrix} e^{i\varphi/2} & \beta e^{-i\varphi/2} \\ 0 & e^{-i\varphi/2} \end{pmatrix} \right\} \qquad \text{(IV.8)}$$

with $e^{i\varphi/2} = \frac{a_{11}}{|a_{11}|}$ and $\beta e^{-i\varphi/2} = a_{12}$. In the form (IV.8) it is not hard to see that $\overline{E(2)}$ is the covering group of the two dimensional Euclidean group $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \beta_x \\ \beta_y \end{pmatrix}$ with $\beta = \beta_x + i\beta_y$.
The inducing subgroup for the lightlike representations is thus

$$H = \begin{pmatrix} \overline{E(2)} & \{H(a)\}\overline{E(2)}^{-1+} \\ 0 & \overline{E(2)}^{-1+} \end{pmatrix} \qquad \text{(IV.9)}$$

The representations of $\overline{E(2)}$ are themselves induced representations, and the D functions of $\overline{E(2)}$ are labeled by a single parameter s, which is either integer or half integer (the finite spin mass zero representations) or a positive number (the continuous spin-mass zero representations).[16] Thus,

$$H \longrightarrow e^{ip_o \cdot a} D_{ii'}^{[s]}(\overline{E(2)}) \qquad \text{(IV.10)}$$

3. $p^2 < 0$, choose $p_o = (000M)$, then $H(p_o) = M\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\Lambda \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \Lambda^+ = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ shows that the relevant subgroup here is SU(1,1), resulting in the inducing subgroup

$$H = \begin{pmatrix} SU(1,1) & \{H(a)\}\, SU(1,1)^{-1+} \\ 0 & SU(1,1)^{-1+} \end{pmatrix} \qquad \text{(IV.11)}$$

which has two classes of unitary irreducible (and therefore infinite dimensional) representations, called the discrete series and the principal series. There are other classes of unitary irreducible representations of SU(1,1), such as the supplementary series, but these other classes do not appear in the reduction of the regular representation of SU(1,1).

Just as with SU(2) and $\overline{E(2)}$, the eigenvalues of the generator of the U(1) subgroup of SU(1,1) is chosen to be in the complete set of commuting elements of the Lie algebra of SU(1,1), so that the D functions of SU(1,1) will be written $D^{[s\pm]}_{ii'}(SU(1,1))$, where i and i′ are integers or half integers. [s±] will label irreducible representations of SU(1,1).

Consider first the discrete series of SU(1,1). Then s is a negative integer or half integer, and i (or i′) ranges from -s, ascending in integer steps to infinity--this corresponds to the + sign--or starts from +s and ranges to minus infinity--which corresponds to the - label.

The principal series of representations of SU(1,1) can be written as induced representations, induced, according to Eq. (II.10), by NA of the Iwasawa decomposition which is of the form

$$\begin{pmatrix} 1+in & -in \\ in & 1-in \end{pmatrix} \begin{pmatrix} \cosh\beta & \sinh\beta \\ \sinh\beta & \cosh\beta \end{pmatrix}, \quad n,\beta \text{ real} \tag{IV.12}$$

and has irreducible representations, labeled by a continuous parameter $s > 0$. For any choice of s, i ranges from minus infinity to plus infinity. However a distinction must be made between i a half integer--which will be labeled by a minus sign--and i an integer--labeled by a plus sign.

Thus, all the relevant representations of SU(1,1) are labeled by [s±] so that representations of the inducing subgroup are

$$H \longrightarrow e^{ip_0 \cdot a} \; D^{[s\pm]}_{ii'}(SU(1,1)) \tag{IV.13}$$

It should be noted that whereas the sign of the energy is not an invariant for the spacelike representations, it is an invariant for the timelike and lightlike representations. Thus, a sign ± will be adjoined to the label s on the D functions of SU(2) and $\overline{E(2)}$, indicating the sign of the energy. Then it is clear that all the inducing subgroups discussed have representations which can be labeled [s±], although the ± takes on different meaning for the various inducing subgroups.

The notation [s±] has been chosen so that all the inducing subgroups can be handled on an equal footing. Thus, let $\hat{\Lambda}$ be an element

of SL(2,C) satisfying $\hat{\Lambda} p_o = p_o$, that is,

$$\hat{\Lambda}\, H(p_o)\, \hat{\Lambda}^{+} = H(p_o) \tag{IV.14}$$

for any of three choices of p_o. Then an element of any inducing subgroup can be written

$$h = \begin{pmatrix} \hat{\Lambda} & H(a)\, \hat{\Lambda}^{-1+} \\ 0 & \hat{\Lambda}^{-1+} \end{pmatrix} \tag{IV.15}$$

with representation

$$h \longrightarrow e^{ip_o \cdot a}\, D_{ii'}^{[s\pm]}(\hat{\Lambda}) \tag{IV.16}$$

The Hilbert spaces $\mathcal{H}(U^{L})$ on which the induced representations of the Poincaré group act are written $\mathcal{H}(U^{p_o s\pm}) \equiv \mathcal{H}(U^{M^2 s\pm})$ and the norm of functions in these Hilbert spaces is

$$\begin{aligned} \|f_i\|^2 &= \int d\mu(\Lambda_c)\, \|f_c(\Lambda_c)\|^2_{\mathcal{H}(p_o s\pm)} \\ &= \sum_i \int d\mu(\Lambda_c)\, |f_i(\Lambda_c)|^2 < \infty \end{aligned} \tag{IV.17}$$

where the range of i depends on the representation $[s\pm]$ which is chosen. Λ_c is a set of elements of SL(2,C) labeling the right cosets; that is

$$SL(2,C) = \bigcup_c \{\hat{\Lambda}\}\, \Lambda_c$$

$$\Lambda = \hat{\Lambda}\, \Lambda_c \tag{IV.18}$$

The induced representations of the Poincaré group act on the spaces $\mathcal{H}(U^{M^2 s\pm})$ as

$$\begin{aligned} U^{[M^2 s\pm]}(\Lambda_o a_o)\, f_i(\Lambda_c) &= f_i[(\Lambda_c, 0)(\Lambda_o a_o)] \\ &= e^{ip_o \cdot \Lambda_c a_o} \sum_{i'} D_{ii'}^{[s\pm]}(\hat{\Lambda}) f_{i'}(\Lambda_{c'}) \end{aligned} \tag{IV.19}$$

where $\hat{\Lambda}$ and $\Lambda_{c'}$ are defined by $\Lambda_c \Lambda_o = \hat{\Lambda}\, \Lambda_{c'}$.

In order to define spaces which are more physically transparent than $\mathcal{H}(U^{[M^2 s\pm]})$, consider "wave functions" $\varphi_i(\vec{p})$ defined by

$$\varphi_i(\vec{p}) = f_i(\Lambda_c) \ , \quad p = \Lambda_c^{-1} p_o \tag{IV.20}$$

Then

$$\begin{aligned} U(\Lambda_o, a_o)\varphi_i(\vec{p}) &\equiv U(\Lambda_o, a_o)\, f_i(\Lambda_c) \\ &= e^{ip\cdot a_o} \sum_{i'} D_{ii'}^{[s\pm]}(\hat{\Lambda})\, \varphi_{i'}(\vec{p}') \end{aligned} \tag{IV.21}$$

where $p' = \Lambda_{c'}^{-1} p_o$; since $\Lambda_c \Lambda_o = \hat{\Lambda}\, \Lambda_{c'}$, it follows that $\Lambda_o \Lambda_{c'}^{-1} = \Lambda_c^{-1} \hat{\Lambda}$ and applying p_o to both sides of this equation gives

$$\begin{aligned} \Lambda_o \Lambda_{c'}^{-1} p_o &= \Lambda_c^{-1} \hat{\Lambda}\, p_o \\ \Lambda_o\, p' &= p \end{aligned} \tag{IV.22}$$

Define (p, Λ_o) as[18]

$$(p, \Lambda_o) = \Lambda_c(\Lambda_o p)\, \Lambda_o\, \Lambda_c^{-1}(p) \tag{IV.23}$$

where $\Lambda_c(p)$ means that coset element which carries p to p_o. Then $\hat{\Lambda} = \Lambda_c \Lambda_o \Lambda_{c'}^{-1}$ means, because of (IV.22), that

$$\begin{aligned} \hat{\Lambda} &= \Lambda_c(p \to p_o)\, \Lambda_o \Lambda_{c'}^{-1}(p' \to p_o) \\ &= \Lambda_c(p)\, \Lambda_o\, \Lambda_c^{-1}(\Lambda_o^{-1} p) \\ &= (\Lambda_o^{-1} p,\ \Lambda_o) \end{aligned} \tag{IV.24}$$

Therefore

$$U(\Lambda_o, a_o)\varphi_i(\vec{p}) = e^{ip\cdot a_o} \sum_{i'} D_{ii'}^{[s\pm]}(\Lambda_o^{-1} p, \Lambda_o)\varphi_{i'}(\Lambda_o^{-1} p) \tag{IV.25}$$

and it is only a short step to seeing how a (non-normalizable) momentum basis vector transforms. Denote the momentum basis vector by $|[M^2 s\pm]\vec{p}\ \sigma\rangle$, where M^2 can range over all of the reals; the choice of M^2 then dictates the meaning of $s\pm$. $\vec{p}$ is the vector momentum, while σ is a "spin" component. Now a general vector can be written

$$|\psi\rangle = \sum_{\sigma} \frac{d^3\vec{p}}{\sqrt{\vec{p}^2+M^2}} \varphi_\sigma(\vec{p}) |[M^2 s\pm]\ \vec{p}\sigma\rangle \tag{IV.26}$$

so that

$$\begin{aligned} U(\Lambda_o,a_o)|\psi\rangle &= \sum_{\sigma} \int \frac{d^3\vec{p}}{\sqrt{\vec{p}^2+M^2}} [U(\Lambda_o,a_o)\varphi_\sigma(\vec{p})] |[M^2 s\pm]\ \vec{p}\ \sigma\rangle \\ &= \sum_{\sigma\sigma'} \int \frac{d^3\vec{p}}{\sqrt{\vec{p}^2+M^2}} e^{ip\cdot a_o} D^{[s\pm]}_{\sigma\sigma'}(\Lambda_o^{-1}p,\Lambda_o)\varphi_{\sigma'}(\Lambda_o^{-1}p) |[M^2 s\pm]\vec{p}\sigma\rangle \\ &= \sum_{\sigma\sigma'} \int \frac{d^3\vec{p}'}{\sqrt{\vec{p}'^2+M^2}} e^{i\Lambda_o p'\cdot a_o} D^{[s\pm]}_{\sigma\sigma'}(p',\Lambda_o)\varphi_{\sigma'}(p') |[M^2 s\pm]\Lambda_o p'\sigma\rangle \end{aligned} \tag{IV.27}$$

and thus,

$$U(\Lambda_o,a_o)|[M^2 s\pm]\vec{p}\sigma\rangle = e^{i\Lambda_o p\cdot a_o} \sum_{\sigma'} D^{[s\pm]}_{\sigma'\sigma}(p,\Lambda_o) |[M^2 s\pm]\overrightarrow{\Lambda_o p}\ \sigma'\rangle \tag{IV.28}$$

so that the transformation properties of basis vectors are completely specified.

Up to this point it has not been necessary to make an actual choice of coset representatives. Eqs. (IV.23) and (IV.28) show that what usually is termed an orbiting operation or boost can equivalently be called a choice of coset representatives. There are many choices possible, but three can be singled out as being particularly useful.

The first choice corresponds to rotationless Lorentz transformations, which in SL(2,C) correspond to Hermitian matrices. Thus, $\Lambda_c = \Lambda_c^+$ and a convenient way of writing such Hermitian matrices is

$$\Lambda_c^{-1} = R(\hat{p})\, L_z(|p|)\, R^{-1}(\hat{p}) \tag{IV.29}$$

where $L_z(|p|)$ is a pure Lorentz transformation along a fixed z axis, of magnitude given by the magnitude of the momentum p (remember that $\Lambda_c^{-1}(p)$ orbits the particle from p_o to p, Eq. (IV.23)) and $R(\hat{p})$ is an element of SU(2) specified by the polar angles of $\vec{p}$. For the Hermitian choice of coset representatives, the index σ of (IV.28) is then interpreted as being the spin component along the fixed z axis.

A second choice generates so called helicity states. In this case the coset representatives are chosen to be[18]

$$\Lambda_c^{-1} = R(\hat{p})\, L_z(|p|) \tag{IV.30}$$

which is non-Hermitian. The usefulness of such coset representatives --emphasized by Jacob and Wick, and others[19]--lies in the fact that helicity states are rotationally invariant so that if a helicity state is defined to be

$$|[M^2 s\pm]\vec{p}\lambda\rangle \equiv U(\Lambda_c^{-1}(p))|[M^2 s\pm]\vec{0}\lambda\rangle \tag{IV.31}$$

then under an arbitrary rotation R

$$\begin{aligned} R|[M^2 s\pm]\vec{p}\lambda\rangle &= R\, R(\hat{p})L_z(|p|)|[M^2 s\pm]\vec{0}\lambda\rangle \\ &= R(\hat{p})\, R^{-1}(\hat{p})\, R\, R(\hat{p})L_z(|p|)|[M^2 s\pm]\vec{0}\lambda\rangle \\ &= R(\hat{p})\, R_z(\varphi)L_z(|p|)|[M^2 s\pm]\vec{0}\lambda\rangle \\ &= R(\hat{p})\, L_z(|p|)\, e^{i\varphi\lambda}|[M^2 s\pm]\vec{0}\lambda\rangle \quad . \end{aligned} \tag{IV.32}$$

A third choice of coset representatives--less well known than the first two choices--will be called Rideau cosets.[20] These coset representatives are defined by

$$\Lambda_c = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\epsilon} \begin{pmatrix} \lambda & 0 \\ \zeta & \lambda^{-1} \end{pmatrix}, \quad \epsilon = 0,1, \quad \lambda > 0, \ \zeta \text{ complex} \tag{IV.33}$$

No confusion should result from the symbol λ in the Rideau cosets (which is Rideau's notation) and the λ labeling helicity in(IV.31). It should be noted that the coset representatives (IV.29) and (IV.30) are generally used for timelike or lightlike representations. For the Rideau cosets ϵ is zero for the timelike or lightlike representation. For the spacelike representations, ϵ can be either zero or one. The explicit dependence of λ and ζ on the momentum p is obtained from the "orbiting" equation

$$H(p) = \Lambda_c^{-1}(p)\, H(p_o)\, \Lambda_c^{-1+}(p)$$

$$\begin{pmatrix} E+p_z & p_x-ip_y \\ p_x+ip_y & E-p_z \end{pmatrix} = \begin{pmatrix} \lambda^{-1} & 0 \\ -\zeta & \lambda \end{pmatrix} H(p_o) \begin{pmatrix} \lambda^{-1} & -\zeta^* \\ 0 & \lambda \end{pmatrix}; \tag{IV.34}$$

thus

$$\lambda = \sqrt{\frac{M}{E+p_z}}\,,$$

$$\zeta = -\frac{p_x+ip_y}{\sqrt{M(E+pz)}} \qquad \text{(IV.35)}$$

for the timelike case. Similar solutions can be obtained for p_o lightlike and spacelike.

The great advantage of the Rideau cosets is that for ε zero, the set of Λ_c forms a group. Therefore the analogue of the Wigner rotation $(p,\Lambda_c) = \Lambda_c(\Lambda_o p)\Lambda_c\Lambda_c^{-1}(p)$ is the identity rotation and no D functions appear when a particle is carried from one frame to another by Rideau cosets. The disadvantage of the Rideau cosets is that they seem not to correspond to obviously physical transformations.

Once the transformation properties of states, Eq. (IV.28), are known, it is possible to compute the D functions. Thus,

$$D^{[M^2 s\pm]}_{\vec{p}'i\,\vec{p}\sigma}(\Lambda_o,a_o) = \langle[M^2 s\pm]\vec{p}'i|\,U(\Lambda_o,a_o)\,|[M^2 s\pm]\,\vec{p}\sigma\rangle$$
$$= e^{i\Lambda_o p\cdot a_o} D^{[s\pm]}_{i\sigma}(p,\Lambda_o)\sqrt{\vec{p}^2+M^2}\;\delta^3(p'-\Lambda_o p) \qquad \text{(IV.36)}$$

and, as was shown in Sec. III, in order to obtain Clebsch-Gordan coefficients, it is necessary to check that the D functions $D^{[X]}_{x_o x}(g_c)$ transform properly as induced representations. For the Poincaré group this is readily done, since

$$U(\Lambda_o,a_o)D^{[M^2 s\pm]}_{\vec{p}'i\,\vec{p}\sigma}(\Lambda_c) = D^{[M^2 s\pm]}_{\vec{p}'i\,\vec{p}\sigma}[(\Lambda_c,0)(\Lambda_o a_o)]$$
$$= D^{[M^2 s\pm]}_{\vec{p}'i\,\vec{p}\sigma}[(\hat{\Lambda},\Lambda_c a_o)(\Lambda_{c'},0)]$$
$$= \sum_{i'} \langle[M^2 s\pm]\vec{p}\sigma|\,U^{-1}(\Lambda_{c'})\,|[M^2 s\pm]\hat{\Lambda}^{-1}p'i'\rangle^*$$
$$\times\, e^{ip'\cdot\hat{\Lambda}^{-1}\Lambda_c a_o} D^{[s\pm]*}_{i'i}(p',\hat{\Lambda}^{-1})$$
$$= \sum_{i'} e^{ip'\cdot\hat{\Lambda}^{-1}\Lambda_c a_o} D^{[s\pm]*}_{i'i}(p',\hat{\Lambda}^{-1}) D^{[M^2 s\pm]}_{\hat{\Lambda}^{-1}p'i',\vec{p}\sigma}(\Lambda_{c'})$$
$$\text{(IV.37)}$$

where, as before, $\Lambda_c\Lambda_o=\hat{\Lambda}\Lambda_{c'}$. But it is clear from (IV.37) that the transformation properties of the Poincaré D functions will agree with (IV.19) only if p' is chosen to be p_o. For then

$$U(\Lambda_o,a_o)D^{[M^2 s\pm]}_{p_o i\,\vec{p}\sigma}(\Lambda_c) = \sum_{i'} e^{ip_o\cdot\Lambda_c a_o} D^{[s\pm]*}_{i'i}(p_o,\hat{\Lambda}^{-1})\, D^{[M^2 s\pm]}_{p_o i'\,\vec{p}\sigma}(\Lambda_{c'})$$

$$= e^{ip_o\cdot\Lambda_c a_o}\sum_{i'} D^{[s\pm]*}_{i'i}(\hat{\Lambda}^{-1})\, D^{[M^2 s\pm]}_{p_o i'\,\vec{p}\sigma}(\Lambda_{c'})$$

$$= e^{ip_o\cdot\Lambda_c a_o}\sum_{i'} D^{[s\pm]}_{ii'}(\hat{\Lambda})\, D^{[M^2 s\pm]}_{p_o i'\,\vec{p}\sigma}(\Lambda_{c'}) \qquad \text{(IV.38)}$$

which agrees with (IV.19). The label σ corresponds to something like a spin component. Its interpretation depends on the choice made of coset representatives. The index i is not specified in (IV.38) and will be shown to act as a degeneracy parameter in the reduction of tensor products. It is to be noted that the indices p_o and i on the Poincaré D function correspond to labels arising from the inducing subgroup of P and therefore correspond to the "mixed basis" D functions, Eq. (III.10).

With the transformation properties of the D functions fixed, it is possible to compute the Clebsch-Gordan coefficients arising in the reduction of n-fold tensor products of the Poincaré group. It will be seen that a natural division occurs between the 2-fold reduction[21] and the $n > 2$-fold reduction.[22]

Consider first the 2-fold reduction. From the point of view of induced representation theory there are now tensor product functions $F_{i_1 i_2}(\Lambda_{c_1},\Lambda_{c_2})$ with norm

$$||F_{i_1 i_2}||^2 = \int d\mu(\Lambda_{c_1})d\mu(\Lambda_{c_2})||F_{i_1 i_2}(\Lambda_{c_1},\Lambda_{c_2})||^2_{\mathfrak{K}[M_1^2 s_1\pm]\otimes[M_2^2 s_2\pm]}$$

$$= \sum_{i_1 i_2} d\mu(\Lambda_{c_1})d\mu(\Lambda_{c_2})|F_{i_1 i_2}(\Lambda_{c_1},\Lambda_{c_2})|^2 < \infty \qquad \text{(IV.39)}$$

which transform under the action of a group element, like

$$U(\Lambda_o,a_o)F_{i_1 i_2}(\Lambda_{c_1},\Lambda_{c_2}) = e^{ip_{o_1}\cdot\Lambda_{c_1}a_o}e^{ip_{o_2}\cdot\Lambda_{c_2}a_o}$$

$$\times\sum_{i_1' i_2'} D^{[s_1\pm]}_{i_1 i_1'}(\hat{\Lambda}_1)D^{[s_2\pm]}_{i_2 i_2'}(\hat{\Lambda}_2)$$

$$\times F_{i_1' i_2'}(\Lambda_{c_1'},\Lambda_{c_2'}) \qquad \text{(IV.40)}$$

with $\Lambda_{c_1}\Lambda_o = \hat{\Lambda}_1\Lambda_{c_1'}$ and $\Lambda_{c_2}\Lambda_o = \hat{\Lambda}_2\Lambda_{c_2'}$.

To find the double cosets relevant to the reduction of the 2-fold tensor product, it is sufficient to consider the SL(2,C) subgroup of the Poincaré group, for the translations, being an invariant subgroup, will always occur in the inducing subgroups. Thus, consider elements $\hat{\Lambda}_1$, $\hat{\Lambda}_2$ of the two inducing subgroups, so that

$$(SL(2,C)_1, SL(2,C)_2) = \bigcup_D (\{\hat{\Lambda}_1\}, \{\hat{\Lambda}_2\})(e,\Lambda_D)(SL(2,C), SL(2,C)) \quad \text{(IV.41)}$$

For most cases of physical interest the elements $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ will be elements of SU(2). Even if this is not the case it is not hard to see that a suitable choice for double coset representatives is $\Lambda_D = \begin{pmatrix} D & 0 \\ 0 & D^{-1} \end{pmatrix}$, $D > 0$ so that the inducing subgroup H_D of P is (not including the translations)

$$\begin{aligned} H_D &= (e,\Lambda_D)^{-1}(\{\hat{\Lambda}_1\}, \{\hat{\Lambda}_2\})(e,\Lambda_D) \cap (SL(2,C), SL(2,C)) \\ &= U(1) \end{aligned} \quad \text{(IV.42)}$$

Define the function

$$\tilde{f}_{i_1 i_2 D}(\tilde{\Lambda}_c) \equiv \tilde{f}_{i_1 i_2 D}(\tilde{\Lambda}_c, \Lambda_D \tilde{\Lambda}_c) = F_{i_1 i_2}(\Lambda_{c_1}, \Lambda_{c_2}) \quad \text{(IV.43)}$$

where the notation $\tilde{f}_D$ comes from (II.18), $\tilde{\Lambda}_c$ is a right coset representative of SL(2,C) with respect to the subgroup U(1), and the subscripts i_1, i_2, and D are labels which are unchanged by the action of the group--and hence label invariant subspaces of the tensor product space:

$$\begin{aligned} U(\Lambda_o, a_o)\tilde{f}_{i_1 i_2 D}(\tilde{\Lambda}_c) &= \tilde{f}_{i_1 i_2 D}[(\tilde{\Lambda}_c, 0)(\Lambda_o, a_o)] \\ &= e^{ip_{o_1}\cdot\tilde{\Lambda}_c a_o} e^{ip_{o_2}\cdot\tilde{\Lambda}_c a_o} \sum_{i_1', i_2'} D^{[s_1\pm]}_{i_1 i_1'}(u) \\ &\quad \times D^{[s_2\pm]}_{i_2 i_2'}(u)\, \tilde{f}_{i_1', i_2', D}(\tilde{\Lambda}_{c'}) \\ &= e^{ip_D\cdot\tilde{\Lambda}_c a_o}(u)^{i_1+i_2}\, \tilde{f}_{i_1 i_2 D}(\tilde{\Lambda}_{c'}) \quad ; \end{aligned} \quad \text{(IV.44)}$$

here u is an element of U(1), the rotation subgroup of the inducing subgroups of P, $p_D = p_{o_1} + \Lambda_D^{-1} p_{o_2}$ and $\tilde{\Lambda}_c$ and u are defined as usual by $\tilde{\Lambda}_c \Lambda_o = u \tilde{\Lambda}_{c'}$.

It is now a short step to computing the Clebsch-Gordan coefficients from Eq. (III.21). Choose $F_{i_1 i_2 D}(\Lambda_{c_1}, \Lambda_{c_2})$ to be

$$D^{[M_1^2 s_1 \pm]}_{p_{o_1} i_1 \vec{p}_1 \sigma_1}(\Lambda_{c_1}) \; D^{[M_2^2 s_2 \pm]}_{p_{o_2} i_2 \vec{p}_2 \sigma_2}(\Lambda_{c_2})$$

which can be written

$$D^{[M_1^2 s_1 \pm]}_{p_{1o} i_1 \vec{p}_1 \sigma_1}(\tilde{\Lambda}_c) \; D^{[M_2^2 s_2 \pm]}_{p_{o_2} i_2 \vec{p}_2 \sigma_2}(\Lambda_D \tilde{\Lambda}_c) \; .$$

Then the Clebsch-Gordan coefficients will be the overlap between these D functions and the D function $D^{[M^2 s \pm]}_{p_D (i_1 + i_2) \vec{p} \sigma}(\tilde{\Lambda}_c)$; note that the four vector p_D is essential in order that $D^{[M^2 s \pm]}_{p_D (i_1 + i_2) \vec{p} \sigma}(\tilde{\Lambda}_c)$ transform like $\tilde{f}_{i_1 i_2 D}(\tilde{\Lambda}_c)$, Eq. (IV.44).

Thus, the Clebsch-Gordan coefficients are

$$\langle [M^2 s \pm] \vec{p} \sigma ; i_1 i_2 D \,|\, [M_1^2 s_1 \pm] \vec{p}_1 \sigma_1 ; [M_2^2 s_2 \pm] \vec{p}_2 \sigma_2 \rangle$$

$$= \int_{\frac{SL(2,C)}{U(1)}} d\mu(\tilde{\Lambda}_c) \left(D^{[M^2 s \pm]}_{p_D (i_1 + i_2) \vec{p} \sigma}(\tilde{\Lambda}_c) , D^{[M_1^2 s_1 \pm]}_{p_{o_1} i_1 \vec{p}_1 \sigma_1}(\tilde{\Lambda}_c) D^{[M_2^2 s_2 \pm]}_{p_{o_2} i_2 \vec{p}_2 \sigma_2}(\Lambda_D \tilde{\Lambda}_c) \right)_{\mathcal{H}(1 \otimes 2)_D}$$

$$= \int_{\frac{SL(2,C)}{U(1)}} d\mu(\tilde{\Lambda}_c) D^{[M^2 s \pm]^*}_{p_D (i_1 + i_2) \vec{p} \sigma}(\tilde{\Lambda}_c) \; D^{[M_1^2 s_1 \pm]}_{p_{o_1} i_1 \vec{p}_1 \sigma_1}(\tilde{\Lambda}_c) \; D^{[M_2^2 s_2 \pm]}_{p_{o_2} i_2 \vec{p}_2 \sigma_2}(\Lambda_D \tilde{\Lambda}_c)$$

$$= \int_{\frac{SL(2,C)}{U(1)}} d\mu(\tilde{\Lambda}_c) \; D^{[s \pm]^*}_{i_1 + i_2, \sigma}(p, \tilde{\Lambda}_c) \; D^{[s_1 \pm]}_{i_1 \sigma_1}(p_1, \tilde{\Lambda}_c) D^{[s_2 \pm]}_{i_2 \sigma_2}(p_2, \Lambda_D \tilde{\Lambda}_c)$$

$$\times \delta^3 (p_D - \tilde{\Lambda}_c p) \; \delta^3 (p_{o_1} - \tilde{\Lambda}_c p_1) \; \delta^3 (p_{o_2} - \Lambda_D \tilde{\Lambda}_c p_2) \qquad \text{(IV.45)}$$

There are 9 delta functions and 5 integration variables, leaving 4 delta functions which express energy-momentum conservation. Furthermore there are 3 D functions of the little groups $\{\hat{\Lambda}_1\}, \{\hat{\Lambda}_2\}$, and $\{\hat{\Lambda}\}$.

Consider the case when $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are both elements of SU(2). Then if the sign of the energy for both 1 and 2 is positive, $\hat{\Lambda}$ will also

be an element of SU(2), and the Clebsch-Gordan coefficients obtained can be used in the usual partial wave expansion of the scattering amplitude. The choice of coset representations will of course dictate the interpretation of σ_1 and σ_2, as for example helicities.

Consider the rotation $(p_1, \tilde{\Lambda}_c)$. Now $(p_1, \tilde{\Lambda}_c) = \Lambda_c(\tilde{\Lambda}_c p_1)\tilde{\Lambda}_c \Lambda_c^{-1}(p_1) = \tilde{\Lambda}_c \Lambda_c^{-1}(p_1)$. Define $p_1' = \Lambda_c(p_1)p = \Lambda_c(p_1)(p_1+p_2)$, which is a vector in the rest frame of particle 1. But $(p_1, \tilde{\Lambda}_c)\Lambda_c(p_1) = \tilde{\Lambda}_c$ and applying p to both sides gives

$$(p_1, \tilde{\Lambda}_c)\Lambda_c(p_1)p = \tilde{\Lambda}_c p$$

$$(p_1, \tilde{\Lambda}_c)p_1' = p_D \qquad \text{(IV.46)}$$

where use has been made of the delta functions in (IV.46). But p_D is a vector with only a z component, as can be seen from its definition. Therefore $(p_1, \tilde{\Lambda}_c)$ is the rotation in the rest frame of particle 1 carrying the vector p' into the z axis. A similar construction is easily carried out for particle 2.

If the signs of the energy are both positive then it is not difficult to show[21] that the rotation appearing in the D function $D^{[s\pm]*}_{i_1+i_2\,\sigma}(p, \tilde{\Lambda}_c)$ carries the vector $\Lambda_c^{-1}(p)p_1$ (which is p_1 transformed into the center of mass) into the z axis in the center of mass. In the event that one of the signs of the energy is negative, however, the invariant mass

$$M^2 = p_D^{\,2} = |H(p_D)|$$

$$= M_1^2 + M_2^2 - 2M_1 M_2 \left(\frac{D^2+D^{-2}}{2}\right) \qquad \text{(IV.47)}$$

ranges from minus infinity to $(M_1 - M_2)^2$. In this case $(p, \tilde{\Lambda}_c)$ can be an element of either SU(1,1) $\overline{E(2)}$ or SU(2).

It will be shown that $(p, \tilde{\Lambda}_c)$ is the transformation which carries $\Lambda_c^{-1}(p)p_1 = \Lambda_c^{-1}(p_1+p_2)p_1$ into the z axis in the standard frame defined by p_o. To see this define $\tilde{p} = \Lambda_c^{-1}(p_D)p_{o_1}$. Then $\tilde{p}$ is aligned along the z axis in the standard frame since p_D has no x or y components. Therefore

$$\tilde{p} = \Lambda_c^{-1}(p_D)p_{o_1} = \Lambda_c^{-1}(p_D)\,\tilde{\Lambda}_c\, p_1$$

$$= [\Lambda_c^{-1}(p_D)\tilde{\Lambda}_c\,\Lambda_c(p)]\Lambda_c^{-1}(p)p_1 = (p, \tilde{\Lambda}_c)\Lambda_c^{-1}(p)p_1 \qquad \text{(IV.48)}$$

so, once the vectors p_1 and p_2 are chosen in some arbitrary frame, the transformation (IV.48) is fixed.

The Clebsch-Gordan coefficients (IV.45), for particle 1 timelike and of positive energy and particle 2 timelike and of negative energy, are of use in the expansion of the scattering amplitude in terms of the momentum transfer variable $(p_1-p_2)^2$, for then p_2 is the 4-momentum of an antiparticle with negative energy.[23]

Consider finally the case of the reduction of n-fold tensor products of the Poincaré group, for $n > 2$. Again it is only necessary to find double cosets with respect to the outer product group $(SL(2,C)_1, SL(2,C)_2, \ldots, SL(2,C)_n)$:

$$(SL(2,C)_1, SL(2,C)_2, \ldots, SL(2,C)_n) = \bigcup_{D_1} (SU(2)_1, SU(2)_2, \ldots SU(2)_n) \times (\Lambda_{D_1}, \Lambda_{D_2}, \ldots, \Lambda_{D_n})(SL(2,C), SL(2,C), \ldots, SL(2,C)), \tag{IV.49}$$

where it has been assumed that the inducing subgroups all are timelike, and of positive energy, this being the case of most physical interest. A concrete choice of double coset representatives is given in Ref. 22. Suffice it to say that with any choice of double coset representatives, the inducing subgroup of the diagonal group $(SL(2,C), SL(2,C), \ldots, SL(2,C))$ is the identity subgroup, so that the function $\tilde{f}_{i_1 i_2 \ldots i_n D}(\Lambda)$ is a function over $SL(2,C)$ such that

$$\tilde{f}_{i_1 i_2 \ldots i_n D}(\Lambda) \equiv \tilde{f}_{i_1 i_2 \ldots i_n}(\Lambda_{D_1}\Lambda, \Lambda_{D_2}\Lambda, \ldots, \Lambda_{D_n}\Lambda) = F_{i_1 i_2 \ldots i_n}(\Lambda_{C_1}, \Lambda_{C_2}, \ldots, \Lambda_{C_n}) \tag{IV.50}$$

Here D refers to all of the relevant double coset labels of (IV.49) and remains unchanged, as do the spin labels $i_1, i_2, \ldots, i_n$, under the action of an arbitrary Poincaré transformation:

$$\begin{aligned} U(\Lambda_o, a_o)\tilde{f}_{i_1 i_2 \ldots i_n D}(\Lambda) &= \tilde{f}_{i_1 i_2 \ldots i_n D}[(\Lambda,0)(\Lambda_o, a_o)] \\ &= e^{ip_{o_1}\cdot\Lambda_{D_1}\Lambda a_o} \ldots e^{ip_{o_n}\cdot\Lambda_{D_n}\Lambda a_o} \\ &\quad \times \tilde{f}_{i_1 i_2 \ldots i_n D}(\Lambda\Lambda_o) \\ &= e^{ip_D\cdot\Lambda a_o}\tilde{f}_{i_1, i_2 \ldots i_n D}(\Lambda\Lambda_o) \end{aligned} \tag{IV.51}$$

where p_D is now defined to be $\Lambda_{D_1}^{-1} p_{o_1} + \Lambda_{D_2}^{-1} p_{o_2} + \ldots + \Lambda_{D_n}^{-1} p_{o_n}$.

The Clebsch-Gordan coefficients are thus the overlap between n D functions $D^{[M_1{}^2 s_1 \pm]}_{p_{o_1} i_1 \vec{p}_1 \sigma_1}(\Lambda_{c_1}) \ldots D^{[M_n{}^2 s_n \pm]}_{p_{o_n} i_n \vec{p}_n \sigma_n}(\Lambda_{c_n})$ and the D function $D^{[M^2 s \pm]}_{p_D i \vec{p} \sigma}(\Lambda)$:

$$\langle [M^2 s\pm]\vec{p}\sigma; i, i_1, i_2, \ldots, i_n D \,|\, [M_1{}^2 s_1 \pm]\vec{p}_1\sigma_1; [M_2{}^2 s_2 \pm]\vec{p}_2\sigma_2; \ldots;$$

$$\times [M_n{}^2 s_n \pm]\vec{p}_n\sigma_n\rangle$$

$$= \int_{SL(2,C)} d\mu(\Lambda)\, D^{[M^2 s\pm]*}_{p_D i \vec{p}\sigma}(\Lambda)\, D^{[M_1{}^2 s_1 \pm]}_{p_{o_1} i_1 \vec{p}_1 \sigma_1}(\Lambda_{D_1}\Lambda) \ldots D^{[M_n{}^2 s_n \pm]}_{p_{o_n} i_n \vec{p}_n \sigma_n}(\Lambda_{D_n}\Lambda)$$

$$= \int_{SL(2,C)} d\mu(\Lambda)\, D^{[s\pm]*}_{i\sigma}(p, \Lambda)\delta^3(p_D - \Lambda p)\, D^{[s_1 \pm]}_{i_1 \sigma_1}(p_1, \Lambda_{D_1}\Lambda)\, \delta^3(p_{o_1} - \Lambda_{D_1}\Lambda)$$

$$\ldots D^{[s_n \pm]}_{i_n \sigma_n}(p_n, \Lambda_{D_n}\Lambda)\, \delta^3(p_{o_n} - \Lambda_{D_n}\Lambda) \qquad \text{(IV.52)}$$

When these 3 (n+1) delta functions are evaluated, it is found that 3 result in momentum conservation, 6 compensate for the integration over SL(2,C) and the remaining 3n-6 can be solved in terms of invariants $p_i \cdot p_j$ and $\epsilon_{\alpha\beta\gamma\delta} p_n^{\alpha} p_{n-1}^{\beta} p_{n-2}^{\gamma} p_1^{\delta}$. Energy conservation is contained in the 3n-6 delta functions. It does not seem to be possible to interpret the rotations appearing in the n + 1 D functions in (IV.52) in any physically obvious way. The rotation (p,Λ) behaves like a rigid body rotation, rotating all n momentum vectors through the same angle.[22] The n remaining rotations then correspondingly change the spin direction of the n particles.

Conclusion

Mackey's theory of induced representations has been shown to provide a valuable tool for the physicist, providing a rather complete analysis of those groups whose irreducible unitary representations can be written as induced representations.

The author wishes to thank Dr. E. Thieleker of Argonne National Laboratory for many helpful discussions on mathematical aspects of induced representation theory and Drs. G. Fuchs and C. Itzykson who provided much valuable criticism during the course of these lectures.

References

1. See, for example, R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. 34, 1 (1962); J. P. Antoine

and D. Speiser, in Group Theoretical Concepts and Methods in Elementary Particle Physics, ed. by F. Gürsey (Gordon and Breach, New York, 1962).
2. G. W. Mackey, The Theory of Group Representations (Dept. of Mathematics, The University of Chicago, Chicago, 1955); and references cited therein.
3. I. M. Gelfand, M. A. Naimark, Unitäre Darstellungen der Klassischen Gruppen (Akademie Verlag, Berlin, 1957), and W. H. Klink, J. Math. Phys. 9, 1669 (1968).
4. E. M. Stein, High Energy Physics and Elementary Particles (IAEA, Vienna, 1965), ed. by A. Salam.
5. The measure theoretic properties of locally compact groups will not be discussed in these lectures; see, however, Refs. 2 and 4.
6. L. S. Pontriagin, Topological Groups, 2nd ed. (Gordon and Breach, New York, 1966).
7. J. S. Lomont, Applications of Finite Groups (Academic Press, New York, 1959); see also Ref. 2.
8. R. Hermann, Lie Groups for Physicists (W. A. Benjamin, Inc., New York, 1966); S. Helgason, Differential Geometry and Symmetric Spaces (Academic Press, New York, 1962).
9. M. A. Naimark, Linear Representations of the Lorentz Group (The MacMillan Co., New York, 1964).
10. A. Sciarrino and M. Toller, J. Math. Phys. 8, 1252 (1967).
11. M. A. Naimark, American Math. Soc. Trans., series 2, 36, 101 (1964).
12. I. M. Gelfand, and G. E. Shilov, Generalized Functions (Academic Press, New York, 1964); A. Böhm, in Lectures in Theoretical Physics, Vol. IXA, ed. by W. E. Brittin et al (Gordon and Breach, New York, 1967).
13. See Ref. 2, p. 129, ff.
14. R. L. Anderson, R. Raczka, M. A. Rashid, and P. Winternitz, "Clebsch-Gordan Coefficients for the Lorentz Group--I," preprint IAEA, IC/67/50, Trieste, Italy.
15. P. Moussa and R. Stora, Lectures in Theoretical Physics, Vol. VIIA, ed. by W. E. Brittin and A. O. Barut (University of Colorado Press, Boulder, Colorado, 1965).
16. J. F. Boyce, R. Delbourgo, Abdus Salam, and J. Strathdee, preprint ICTP/67/9, Trieste, Italy.
17. V. Bargmann, Ann. Math. 48, 568 (1947); see also Refs. 10 and 16.
18. It is to be noted that this definition is not the usual one given, as for example, the article of J. Dreitlein in Ref. 15. Quite generally the notation of Mackey has been followed, which means

that an induced representation is defined as $U(g_o)\, f(g) = f(gg_o)$ (that is, group multiplication acts to the right) rather than the definition given by Moussa and Stora in Ref. 15, namely $U(g_o)f(g) = f(g_o^{-1}g)$. As a consequence $D^{[\chi]}_{x_o x}(g)$ functions have the "standard states" x_o to the left, as bra vectors, rather than $D^{[\chi]}_{x x_o}(g)$, where they would be ket vectors. The advantage of Mackey's choice is that all results finally are in terms of ket vectors $|[\chi]x\rangle$. The disadvantage is that coset representatives will be different than conventional definitions. This fact has been taken into account in defining the various coset representatives, Eqs. (IV.29) and (IV.30).

19. M. Jacob and G. C. Wick, Ann. Phys. 7, 404 (1959); G. C. Wick, Ann. Phys. 18, 67 (1962); L. Durand III, Lectures in Theoretical Physics, Vol. IV, ed. by W. E. Brittin, B. W. Downs, and J. Downs (John Wiley and Sons, New York, 1962).
20. G. Rideau, Commun. Math. Phys. 3, 218 (1966).
21. For a list of references concerning the reduction of two timelike positive energy representations, see Ref. 15; M. Kummer, J. Math. Phys. 7, 997 (1966); W. H. Klink, University of Iowa preprint 67-59 (to be published in J. Math. Phys.)
22. W. H. Klink and G. J. Smith, Commun. Math. Phys. 10, 231 (1968) and references cited therein.
23. F. T. Hadjioannou, Nuovo Cimento 44, 185 (1966); see also the work of H. Joos in Ref. 15, and Ref. 16.

THE LORENTZ GROUP AND COMPLEX ANGULAR MOMENTUM†

J. G. Taylor
Department of Physics
Queen Mary College
London, England

I. A Review of Group Theoretic Essentials

A. Introduction

In these nine lectures I want to consider the problem of Regge poles from the point of view of their embedding in the Lorentz group in contradistinction to the 3-dimensional rotation group. The non-covariant formulation of Regge pole theory is by means of the partial wave expansion of a two-body scattering amplitude. In order to perform this in a covariant and crossing symmetric fashion it is necessary to consider the covariant angular momentum separation for both positive or negative values of the invariant $P_\mu{}^2$, where P_μ is the 4-vector $P_\mu = p_\mu{}^{(1)} + p_\mu{}^{(2)}$ and $p_\mu{}^{(i)}$ is the energy-momentum 4-vector of the ith particle. Such a discussion itself requires an understanding of the various representations of the little group of P_μ, which are SO(3), SO(2,1), E(2) (the two dimensional Euclidean group) and finally SO(3,1). This discussion is also based on harmonic analysis and its extension analogous to the Laplace transform extension of the Fourier transform on these groups. To be able to give this discussion let me begin by describing these ideas of harmonic analysis on groups and so on; I will do this in a descriptive fashion, not trying to prove theorems I quote, but trying to describe the main ideas and the main structures that are involved.

Let me begin by going briefly over the fundamental concepts I will use in groups and functional analysis. It will help to jog your memories as well as set the notation I will use.

B. Group Theory[1)]

An abstract group G is a set of elements a, b, ... which is closed under multiplication:

$$a, b \in G \Rightarrow ab \in G$$

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

and which is associative: $a(b) = (ab)c$, and for which each element a has an inverse a^{-1} with $aa^{-1} = a^{-1}a = e$, where e is the identity of the group. A group G_1 is homomorphic to a group G_2 if to each $a_1 \in G_1$ there is assigned a unique element $a_2 \in G_2$, so that if b_1 and b_2 are similarly related then $b_1 b_2$ is assigned to $a_1 a_2$. We can denote the assignment $a \to b$ as a mapping h: $h(a) = b$, so we require for a homomorphism that $h(a_1 a_2) = h(a_1) \cdot h(a_2)$. Two groups G_1 and G_2 are isomorphic if there is a 1:1 mapping h which preserves multiplication.

An automorphism of G is an isomorphism of G into itself; this is an inner automorphism if $\exists\, x \in G$ so that $h(a) = xax^{-1}$, $\forall\, a \in G$; it is an outer automorphism otherwise. Two elements a, b commute if ab = ba; two subgroups G_1, G_2 commute if any element of G_1 commutes with any element of G_2. A group is Abelian of any pair if its elements commute. Two elements a, b, are conjugate if $\exists\, x \in G$: $a = xbx^{-1}$. A subgroup G_1 is self-conjugate (or invariant or normal) if $xgx^{-1} \in G_1$ for any $g \in G_1$, any $x \in G$. A subgroup of a group G is a subset of G which is itself closed under multiplication (and so is a group in its own right). If every element of a group G can be written uniquely as the product of an element of a subgroup G_1 with an element of a subgroup G_2 which commutes with G_1, for which $G_1 \cap G_2 = \{e\}$ then G is the direct product of G_1 and G_2, denoted by

$$G = G_1 \times G_2 \quad .$$

Let G_1 be an Abelian invariant subgroup of the group G. Then G is the semi-direct product of G_1 and the subgroup G_2 if every element of G can be written uniquely as $g = g_1 g_2$, $g_i \in G_i$, $G_1 \cap G_2 = \{e\}$. We see that the product of g with $f = f_1 f_2$ in G can be written as

$$fg = f_1 f_2 g_1 g_2 = (f_2 f_1 g_1 f_2^{-1})\, f_2 g_2 \text{, where } f_1 f_2 g_1 f_2^{-1} \in G_1 \text{, } f_2 g_2 \in G_2 \text{ ;}$$

we may write $f_2 g_1 f_2^{-1} = \alpha_{f_2}(g_1)$, where α_{f_2} is the inner automorphism of G defined by f_2. We write $G = G_1 \wedge G_2$. For any invariant subgroup G_1 we may define the quotient group G/G_1 by means of the left cosets $gG = \{gg_1, g_1 \in G_1\}$. Such left cosets may be multiplied together $(gg_1 g' g_1' = gg'(g'^{-1} g_1 g') g_1'$ and since $g'^{-1} g_1 g' \in G_1$, and $g_1' \in G_1$ then $gg_1 g' g_1 \in$ coset defined by gg'). This set of left cosets together with this multiplication law is denoted by G/G_1.

A group G with no non-trivial invariant subgroup is termed simple; if it has no non-trivial Abelian invariant subgroup it is termed semi-simple.

C. Example: The Inhomogeneous Lorentz Group

Let x be a real 4-vector with components x_μ $(\mu = 0,1,2,3)$, with Lorentz length $x^2 = g^{\mu\nu} x_\mu x_\nu$, where $g^{\mu\nu} = 0$ $(\mu \neq \nu)$ $g^{00} = +1$, $g^{11} = -1$.

Let the linear transformation

$$x \to x' = \Lambda x + a$$

with

$$x_\mu{}' = \Lambda_\mu{}^\nu x_\nu + a_\mu$$

be denoted (a,Λ). If Λ is a real 4×4 matrix which preserves the Lorentz length of any vector and a is a real 4-vector, then (a,Λ) is called an inhomogeneous Lorentz transformation; when $a = 0$, $(0,\Lambda)$ is a homogeneous Lorentz transformation. The set of such transformations forms a group, $\mathcal{P}$ (sometimes called the Poincaré group), with group multiplication law

$$(a',\Lambda')(a\ \Lambda) = (a' + \Lambda' a, \Lambda'\Lambda) \quad .$$

It will be true that $\mathcal{P}$ is a group if we can show that the homogeneous Lorentz group, $\mathcal{L}$, the subgroup of elements $(0,\Lambda)$, is actually a group. This follows from the condition

$$x'^2 = x^2$$

i.e.,

$$x_\mu g^{\mu\nu} x_\nu = \Lambda_\mu{}^{\mu'} x_{\mu'} g^{\mu\nu} \Lambda_\nu{}^{\nu'} x_{\nu'},$$

or

$$\Lambda_\mu{}^{\mu'} g^{\mu\nu} \Lambda_\nu{}^{\nu'} = g^{\mu'\nu'}$$

or

$$\Lambda^T g \Lambda = g \quad .$$

If this is true for Λ_1, Λ_2 then it will also be true for $(\Lambda_1\Lambda_2)$, as is seen from

$$(\Lambda_1\Lambda_2)^T g (\Lambda_1\Lambda_2) = \Lambda_2{}^T \Lambda_1{}^T g \Lambda_1 \Lambda_2 = \Lambda_2{}^T g \Lambda_2 = g \quad .$$

Evidently the identity matrix $I \in \mathcal{L}$, and so is Λ^{-1} for any $\Lambda \in \mathcal{L}$ (since $\Lambda^T g \Lambda = g \Rightarrow g = (\Lambda^{-1})^T g \Lambda^{-1}$, and $(\det \Lambda)^2 = 1$, $\det \Lambda = \pm 1$, so Λ^{-1} exists.) There is another important subgroup of $\mathcal{P}$ besides $\mathcal{L}$; it is T_4, the translation subgroup, composed of (a,I). Since we may write

$$(a,\Lambda) = (a,I)(0,\Lambda)$$

and evidently $\mathfrak{L} \cap T_4 = \{e\}$, where $e = (0,I)$, and also T_4 is an invariant subgroup of $\mathcal{P}$:

$$(a,\Lambda)(b,I)(a,\Lambda)^{-1} = (a,\Lambda)(b,I)(-\Lambda^{-1}a,\Lambda^{-1}) = (a,\Lambda)(b-\Lambda^{-1}a,\Lambda^{-1})$$

$$= (a + \Lambda b - a, I) = (\Lambda b, I) \in T_4 \ ,$$

so $\mathcal{P} = T_4 \Lambda \mathfrak{L}$. Since T_4 is also Abelian we see that $\mathcal{P}$ is not even semi-simple. Evidently T_4 is isomorphic to the whole 4-dimensional real Euclidean space $\underline{R}_4$; the structure of $\mathfrak{L}$ is more interesting. Indeed since $\det \Lambda = \pm 1$ we may split $\mathfrak{L}$ into two parts, $\mathfrak{L}_+$ and $\mathfrak{L}_-$ with $\det \Lambda = \pm 1$ respectively. Since also if we take the 00 component of $\Lambda_\mu{}^{\mu'} g^{\mu\nu} \Lambda_\nu{}^{\nu'} = g^{\mu'\nu'}$ we obtain

$$(\Lambda_0{}^0)^2 - (\Lambda_1{}^1)^2 - (\Lambda_2{}^2)^2 - (\Lambda_3{}^3)^2 = 1$$

so $\Lambda_0{}^0 \geqslant +1$ or $\leqslant -1$.

In the first case, $\Lambda_0{}^0 \geqslant +1$ we see that if $x_0 > |\underline{x}|$, $\underline{x} = (x_1, x_2, x_3)$ then $(\Lambda x)_0 \geqslant |\underline{\Lambda x}|$; if $\Lambda_0{}^0 \leqslant -1$ then $(\Lambda x)_0 \leqslant |\underline{\Lambda x}|$. We may divide $\underline{R}_4$ into three regions: $x^2 > 0$ (timelike x), $x^2 = 0$ (lightlike x) and $x^2 < 0$ (spacelike x). These regions may be subdivided into

(i) $x_0 > |\underline{x}|$ (interior of the forward cone, $\mathcal{V}_+$)
(ii) $x_0 < -|\underline{x}|$ (interior of the backward cone, $\mathcal{V}_-$)
} $x^2 > 0$

(iii) $x \neq 0$ { $x_0 > 0$; $x_0 < 0$ }
$x = 0$
} $x^2 = 0$. (the light cone).

Then we may split $\mathfrak{L}$ into four parts $\mathfrak{L}_{\pm}^{\uparrow\downarrow}$, with $\Lambda \in \mathfrak{L}^\uparrow$ if $\Lambda_0{}^0 > ((\Lambda_1{}^1)^2 + (\Lambda_2{}^2)^2 + (\Lambda_3{}^3)^2)^{\frac{1}{2}}$, and $\Lambda \mathcal{V}_+ \in \mathcal{V}_+$. Similarly for $\mathfrak{L}_\downarrow$. If we denote by π the space reflection

$$\pi = \begin{pmatrix} 1 & & & \\ & -1 & \underline{0} & \\ & \underline{0} & -1 & \\ & & & -1 \end{pmatrix} ,$$

θ the time reflection

$$\theta = \begin{pmatrix} -1 & & \underline{0} & \\ & 1 & & \\ \underline{0} & & 1 & \\ & & & 1 \end{pmatrix}$$

then $\pi \mathfrak{L}_+^\uparrow = \mathfrak{L}_-^\uparrow$, $\theta \mathfrak{L}_+^\uparrow = \mathfrak{L}_-^\downarrow$, $\theta\pi \mathfrak{L}_+^\uparrow = \mathfrak{L}_+^\downarrow$, and $\mathfrak{L}_+^\uparrow$ is an invariant subgroup of $\mathfrak{L}$ with $\mathfrak{L}/\mathfrak{L}_+^\uparrow = (1, \theta, \pi, \theta\pi)$. We call $\mathfrak{L}_+^\uparrow$ the proper Lorentz group.

D. Topology[1)]

In order to be able to discuss continuity and connectedness we set up the notion of open sets in a general set A (elements a,b,c,...). The open sets $\underline{O}$ are subsets of A which are closed under arbitrary reunions ($O_\alpha \in \underline{O} \Rightarrow \bigcup_\alpha O_\alpha \in O$) and finite intersection ($O_1, O_2 \in O \Rightarrow O_1 \cap O_2 \in \underline{O}$); also the null-set φ and the whole set A are open.

The closed sets are the complements of open sets.

The neighborhoods of a point $a \in A$ are the sets N(a), each of which contains an open set containing a. The set A with the collection $\underline{O}$ is called a topological space, and the collection $\underline{O}$ characterizes a topology τ (which is equally well characterized by the collection of closed sets or the collection of neighborhoods of each of the points of A). A function f (single valued) on a topological space A_1 with values in a topological A_2 is called continuous if the inverse image of every open set in A_2 is open in A_1.

A particularly important example of a topological space is called a Hilbert space H. This is a linear vector space over the complex number field $\underline{C}$ which is equipped with a scalar or inner product (f,g) for any pair of elements $f,g \in H$, with the properties:

(i) (f,g) is a complex number with $(f,g)^* = (g,f)$

(ii) it is linear in g, antilinear in f

(iii) $(f,g) \geq 0 = 0 \Leftrightarrow f = 0$.
Then $||f|| = (f,f)^{\frac{1}{2}}$ satisfies the triangular inequality

$$||f + g|| \leq ||f|| + ||g||$$

and the Hilbert space topology is defined by means of the open sets which are the spheres $\{||f - f_0|| < \epsilon, \forall f_0 \in H, \epsilon > 0$.

(iv) A Hilbert space is also required to be complete:

$$||f_n - f_m|| \to 0 \text{ as } n,m \to \infty \Rightarrow \exists f \in H \ni ||f_n - f|| \to 0 \text{ as } n \to \infty .$$

(v) We can also add the separability condition: $\exists \{f_n\}$ dense in H, i.e. the smallest closed subset containing the linear subspace generated by $\{f_n\}$ is the whole of H. The set $\{\varphi_n\}$ is termed an orthonormal set if $(\varphi_n, \varphi_m) = \delta_{nm}$ and is termed complete if $(f,\varphi_n) = 0, \forall_n \Rightarrow f = 0$. Any $f \in H$ can be expanded uniquely as $f = \Sigma a_n \varphi_n$, with $a_n = (f,\varphi_n)$. A linear operator A is a mapping of a subset D_A (the domain of A) of H into H which is linear on A: $A(\alpha\varphi + \beta\psi) = \alpha A(\varphi) + \beta A(\psi)$, if $\alpha\varphi + \beta\psi, \varphi, \psi \in D_A$, any $\alpha,\beta \in \underline{C}$. If D_A is dense in H (the closure of D_A, the smallest closed subset

containing D_A, is equal to H), we may define a unique operator A^+, the adjoint of A, by $(Af,g) = (f,A^+g)$, $\forall f \in D_A$. A is self-adjoint (Hermitian) if $A^+ = A$; it is unitary if $A^+A = AA^+ = I$. A linear operator A is bounded if

$$\exists \lambda > 0 \ni ||A\varphi|| \leqslant \lambda ||\varphi||, \forall \varphi \in H.$$

Then $||A|| = \inf ||A\varphi||/||\varphi||$. Finally a projection P is a linear operator $\ni P^+ = P, P^2 = P$, P is bounded, and $H = PH \oplus (I - P)H$ defines a direct sum decomposition of any $\varphi \in H$ into $P\varphi \in PH$, $(I - P)\varphi \in (I - P)H$, with $(P\varphi, (I-P)\varphi) = 0$. In general subspaces M, N of H are orthogonal, $M \perp N$ if $(m,n) = 0$, $\forall m \in M$, $n \in N$.

A compact set is one for which every open covering is finite. A locally compact space is one for which every point has a compact neighborhood.

E. Topological Groups[1)]

A topological group is a set G equipped with the structure of a group and of a topological space which are compatible, so that $(x,y) \to xy^{-1}$ is a continuous function of $G \times G \to G$. If every point of G has a neighborhood which is homeomorphic to a bounded open set in R^n then G is called an n-parameter group (a homeomorphism is a 1:1 bi-continuous mapping). Then every n-parameter group is locally compact. If a point is parametrized by the real numbers $a_1 \ldots a_n$ in some neighborhood and by $(b_1 \ldots b_n)$ in an overlapping neighborhood we can consider $b_1 \ldots b_n$ as n functions of the parameters $a_1 \ldots a_n$. If these functions are indefinitely differentiable (C^∞) then the group is called a C^∞ manifold. If moreover the group operation $xy \to xy^{-1}$ gives a differentiable function for the parameters of xy^{-1} when regarded as functions of the parameters of x and y then G is a Lie group.

Evidently $\mathcal{P}$, T_4 and $\mathcal{L}$ are respectively 10-, 4-, and 6-parameter Lie groups which are locally compact but not compact in the obvious topology on the parameters.

There are various properties of n-parameter groups which are important. Firstly there are various notions of connectedness. A topological space G is connected if it cannot be expressed as the union of 2 disjoint open sets. If x(t) is a continuous function of some interval I of the real line into G then x(t) traces out a path as t moves through I. G is arcwise connected if any two points x, y may be joined by a path lying in G. Taking $I = [0,1]$ then we require a path x(t) in G so that $x(0) = x$, $x(1) = y$. Two paths x(t) and y(t) are called homotopic if they can be continuously deformed into each other, i.e., if there exists a continuous function x(s,t) on $I \times I$ into G so that $x(0,t) = x(t)$, $x(1,t) = y(t)$.

Consider a given point $P_0 \in G$, and let π denote the sets of homotopically equivalent curves which begin and end at P_0. Each element of π is thus composed of a set of closed curves homotopically equivalent to a given closed curve beginning and ending at P_0. We may equip π with a multiplication law by defining, for any closed curves $x(t)$, $y(t)$, the product curve $xy(t) = x(2t)$ $(0 \leq t \leq \frac{1}{2})$, $= y(2t-1)$, $(\frac{1}{2} \leq t \leq 1)$. Also $x^{-1}(t) = x(1-t)$. π now forms a group, called the fundamental group of G. It is, in general, a discrete Abelian group; if it has m elements then G is called m-ply connected (m = 1, G is simply connected). Then given any group G, which may in general be multiply connected, we may always choose a unique simply connected group $\bar{G}$ which is locally isomorphic to G, and determined by π; $\bar{G}$ is called the covering group of G, and in fact $G = \bar{G}/\pi$. We may piece together the homeomorphic images in R^n of the neighborhood of the elements to form the parameter or group space (everywhere locally isomorphic with R_n). If this parameter space is itself compact then the Lie group will be compact; if not it will be locally compact.

A second very important property for any separable locally compact group is the existence of an invariant measure. To begin with we define the Borel subsets of any topological space G as the smallest family of sets containing the open sets and closed under the two operators of taking complements and countable intersections. A measure μ on G is then a positive-valued function on the Borel subsets of G which is countably additive: $\mu(\bigcup_i E_i) = \sum_i \mu(E_i)$ if $E_i \cap E_j = \varphi$, $i \neq j$. A right invariant Haar measure has the additional property of $\mu(Ex) = \mu(E)$, $\forall x = G$, $\forall$ Borel set $E \subset G$. It can be proved that every locally compact topological group has a right invariant Haar measure (which is non-trivial); it also has a left invariant Haar measure. Further the right (or left) invariant Haar measure is unique in the sense that if μ_1 and μ_2 are both such measures then there is a positive real number λ so that $\mu_1(E) = \lambda\mu_2(E)$, $\forall E \in C$. If G is compact it is always possible to take the Haar measure so that $\mu(G) = 1$. If G is not compact $\mu(G)$ is never defined. Further if G is compact we may take μ to be both left and right invariant. If $(t_1 \ldots t_n)$ are the group parameters then the Haar measure defines an invariant integral $\int d\mu(t_1 \ldots t_n)$ over the group manifold.

F. The Groups SO(3), SU(2), SL(2,C), Etc.

The group of rotations about the origin in 3-dimensions is given by the group of 3 x 3 real orthogonal matrices, with determinant plus one (so excluding transformations from a right- to a left-handed set of axes). This group may be described by points inside the sphere of radius π. Any rotation may be described by a vector $\underline{n}$ along a direction about which the rotation occurs and of length equal to the angle of rotation. Since rotations of $+\pi$ and $-\pi$ give equal results then

opposite ends of a diameter on the surface of the sphere must be identified. This implies that the group is doubly connected, since a closed curve which jumps across such a diameter once can never be reduced to a point. The related simply connected group has two identical spheres, where now one doesn't identify opposite ends of a diameter but equivalent points on the surfaces of the two spheres. The corresponding group is SU(2), the set of 2 x 2 unitary matrices of determinant plus one. Instead of giving the details of the relationship of SO(3) and SU(2) we will consider the relation of $\mathfrak{L}_+^\uparrow$ and S(2,C), since the earlier relation will follow from it.

The group SL(2,C) is the set of 2 x 2 complex matrices with determinant plus one, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, with $a_{11}a_{22} - a_{12}a_{21} = 0$. In terms of the Pauli matrices $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ we may represent a real 4-vector x by the Hermitian matrix

$$X = x_0\sigma + \underline{x}\cdot\underline{\sigma} = \begin{pmatrix} x_0+x_3 & x_1 - ix_2 \\ x_1+ix_2 & x_0-x_3 \end{pmatrix}, \quad \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Consider the mapping $X \to X' = A \times A^+$ and conversely any Hermitian matrix X defines a real 4-vector x_μ, with $x_\mu = \frac{1}{2}\mathrm{Tr}(X\sigma_\mu)$. Since X' is also a Hermitian 4 x 4 matrix it defines a new point in $\underline{R}^4$, x', and the transformation $x \to x'$ is linear, $x_\mu' = \Lambda_\mu{}^\nu(a)x_\nu$, $\Lambda_\mu{}^\nu = \frac{1}{2}\mathrm{Tr}[A\sigma_\lambda A^+\sigma_\mu]g^{\nu\lambda}$. We have that $\det X = x^2 \Rightarrow \det X' = x'^2$, so $\Lambda(a) \in \mathfrak{L}$. Thus we have a mapping of SL(2,C) into $\mathfrak{L}$ defined by $\Lambda(a)$. Since this mapping is continuous and SL(2,C) is connected it can only map SL(2,C) into the connected component of the identity 1 in $\mathfrak{L}$; this is seen to be $\mathfrak{L}_+^\uparrow$, since neither θ nor π can be obtained continuously (requiring a change of sign of determinant) from 1 in $\mathfrak{L}$. We also see that SL(2,C) has 6 real parameters, as does $\mathfrak{L}_+^\uparrow$, so that we expect SL(2,C) to be equal to the whole of $\mathfrak{L}_+^\uparrow$. This is indeed the case, and the explicit form of the relation between SL(2,C) and $\mathfrak{L}_+^\uparrow$ is given by the following set of relations (which I will not prove):

(a) For given $a \in SL(2,C)$, $\Lambda(a)$ is given by:

$$\Lambda_o{}^o = |a_o|^2 + \sum_{k=1}^{3} |a_k|^2$$

$$\Lambda_o{}^k = a_o\bar{a}_k + \bar{a}_o a_k + i\varepsilon_{k\ell m} a_\ell \bar{a}_m$$

$$\Lambda_k{}^o = a_o\bar{a}_k + \bar{a}_o a_k - i\varepsilon_{k\ell m} a_\ell \bar{a}_m$$

$$\Lambda_\ell{}^k = \delta_{k\ell}\left(|a_0|^2 - \sum_{k=1}^{3}|a_k|^2\right) + a_k\bar{a}_\ell + \bar{a}_k a_\ell$$

$$+ i\epsilon_{k\ell m}(\bar{a}_0 a_m - a_0\bar{a}_m)$$

(b) For given $\Lambda \in \mathcal{L}_+^\uparrow$, a is given by:

(i) $a_0 + \underline{a}\cdot\underline{\sigma} = N^{-1}\left\{\Lambda_\nu{}^\nu + \sum_{k=1}^{3}(\Lambda_0{}^k + \Lambda_k{}^0 + i\epsilon_{ok\rho\tau}\Lambda_\tau{}^\rho)\,\sigma_k\right\}$

with $N^2 = 4 - \operatorname{tr}\Lambda^2 + (\operatorname{tr}\Lambda)^2 + i\epsilon_{\mu\rho\lambda\tau}\Lambda_\mu{}^\rho\Lambda_\lambda{}^\tau$ if $\Lambda_\nu{}^\nu \neq 0$

(ii) $a_0 = 0$, $a_k = \pm\{[\tfrac{1}{2}(\Lambda_0{}^0 - 1)]^{\frac{1}{2}}(\underline{e}_1 \times \underline{e}_0)_k$

$$+ i[\tfrac{1}{2}(\Lambda_0{}^0 + 1)]^{\frac{1}{2}}(\underline{e}_1)_k\}$$

with $(\underline{e}_0)_k = \left\{\sum_{\ell=1}^{3}(\Lambda_0{}^\ell)^2\right\}^{-\frac{1}{2}}\Lambda_0{}^k$, $\underline{e}_1$ the normalized real eigenvector belonging to the eigenvalue $(\Lambda_0{}^0 + 1)$ of the matrix $M_k{}^\ell = \Lambda_0{}^0\delta_{k\ell} + \Lambda_k{}^\ell$ if $\Lambda_\nu{}^\nu = 0$.

In both cases we see that the sign ambiguity of $a(\Lambda)$ means that the identity of $\mathcal{L}_+^\uparrow$ is the image of $+\sigma_0$ and $-\sigma_0$ in SL(2,C), and that $\mathcal{L}_+^\uparrow$ is isomorphic to $SL(2,C)/\{+\sigma_0, -\sigma_0\}$. The relation between SO(3) and SU(2) can be deduced from the above relation, and as we expect, SO(3) is isomorphic to $SU(2)/\{+\sigma_0, -\sigma_0\}$. Also we see that $\mathcal{L}_+^\uparrow$ and so $\mathcal{L}$ and $\mathcal{P}$ are doubly connected. Indeed we may easily see that to the rotation $\underline{n}$ in SO(3) correspond the two rotations $\pm\exp\{\frac{1}{2}\underline{\sigma}\cdot\underline{n}\}$ in SU(2) (the factor $\frac{1}{2}$ in the exponent essentially causes the double-valuedness since, for example, $\exp[\frac{i}{2}\sigma_1(\theta+2\pi)] = \exp[\frac{i}{2}\sigma_1\theta]$). We may also describe SL(2,C) by means of complex rotations. We make the ansatz $a = \exp\{\frac{1}{2}\underline{\sigma}\cdot\underline{b}\}$, where $\underline{b}$ is a complex rotation vector. Let $\sum_{k=1}^{3} b_k^2 = -(\eta + i\varphi)^2 \neq 0$. Then using the usual power series expansion for the exponential function and the properties of the Pauli matrices,

$$a = \operatorname{ch}\tfrac{1}{2}(\eta + i\varphi) + i\,\underline{\sigma}\cdot\underline{b}\,\frac{\operatorname{sh}\frac{1}{2}(\eta + i\varphi)}{(\eta + i\varphi)}$$

whereas if $\eta + i\varphi = 0$ we have $a = 1 + \frac{1}{2}\underline{\sigma}\cdot\underline{b}$. We now solve for $\underline{b}$ in terms of a and find

$$\frac{i}{2}b_k = \frac{a_k(\eta+i\varphi)}{2\,sh\frac{1}{2}(\eta+i\varphi)} = \frac{a_k \log[\,ch(\eta+i\varphi) + sh\frac{1}{2}(\eta+i\varphi)]}{sh\frac{1}{2}(\eta+i\varphi)}$$

$$= \frac{a_k \log(a_0 + \sqrt{a_0^2 - 1}}{\sqrt{a_0^2 - 1}}$$

As a function of a_0 this has a cut from $-\infty$ to -1; the interval $|a_0| \leq 1$ gives the real b_k, so the subgroup SU(2). Other than the cut we can obtain the whole of SL(2,C) by exponentiation.

The important subgroups of SL(2,C) are: SU(2), SU(1,1), and SL(2,R). We have already defined SU(2) as the set of unitary matrices in SL(2,C): $U^+U = 1$. If $U = u_0 + i\,\underline{u}\cdot\underline{\sigma}$ with $u_0^2 + \underline{u}^2 = 1$ then u_0, u_1, u_2, u_3 are real.

SU(1,1) contains elements V of SL(2,C) with $V^+\sigma_3 = \sigma_3 V^{-1}$, so if $V = u_0 + v_1\sigma_1 + v_2\sigma_2 + iv_3\sigma_3$, $v_0 - v_1^2 - v_2^2 + v_3^2 = 1$ then $v_0, \underline{v}$ are real. SL(2,R) has the restriction $a^+\sigma_2 = \sigma_2 a^{-1}$, so we can take $a = a_0 + a_1\sigma_1 + ia_2\sigma_2 + a_3\sigma_3$ with real a_0, a_1. We may relate SU(1,1) and SL(2,R) by the 1:1 mapping $a_0 = v_0$, $a_1 = v_1$, $a_2 = -v_3$, $a_3 = v_2$.

We may write the invariant integral over the various subgroups as follows. We parametrize $a \in SL(2,C)$, $a = a_0 + \underline{\sigma}\cdot\underline{a}$, $d\mu(a) = C_0^2 \prod_{i=0}^{3} Da_i\, \delta(a_0^2 - \sum_{i=1}^{3} a_i^2 - 1)$ where $Dz = dxdy$, $\delta(z) = \delta(x)\delta(y)$. This is invariant, since $d\mu(ab) = d\mu(a)\frac{\partial(ab)}{\partial a} = d\mu(a)\det b = d\mu(a)$.

$a \in SL(2,R)$, $a = a_0 + a_1\sigma_1 + ia_2\sigma_2 + a_3\sigma_3$; $d\mu(a) = C_0 \prod_{\lambda=0}^{3} da_\lambda\, \delta(a_0^2 + a_2^2 - a_1^2 - a_3^2 - 1)$

$u \in SU(2)$, $u = u_0 + i\underline{\sigma}\cdot\underline{u}$; $d\mu(u) = C_0 \prod_{\lambda=0}^{3} du_\lambda\, \delta\left(\sum_{\lambda=3} u_\lambda^2 - 1\right)$

$v \in SU(1,1)$, $v = v_0 + v_1\sigma_1 + v_2\sigma_2 + iv_3\sigma_3$; $d\mu(v) = C_0 \prod_{\lambda=0}^{3} dv_\lambda\, \delta(v_0^2 + v_3^2 - v_1^2 - v_2^2 - 1)$

where the last three measures involve only real differentials and δ-functions. SU(2) is compact, with measure $c_0\pi^2$ so we take $c_0\pi^2 = 1$, and use the same c_0 for the other measures. In terms of the parametrization given earlier for SU(2), $u \in \exp\{\frac{1}{2}\underline{\sigma}\cdot\underline{b}\}$, with $u_0 = \cos\frac{1}{2}|\underline{b}|$, $u_k = -\frac{b_k}{|\underline{b}|}\sin\frac{1}{2}|\underline{b}|$, we have

$$d\mu(u) = c(1 - \cos|\underline{b}|)\frac{d^3b}{|\underline{b}|}, \quad c = (8\pi^2) - 1 .$$

G. Lie Algebras[2)]

Let me restrict my discussion to groups whose elements are finite-dimensional matrices. We can obtain a great deal of the important properties of a Lie group by restricting our attention to the infinitesimal elements of G, which define the Lie algebra L_G of G. We consider the differentiable one-parameter subsets of G. Each of these can be regarded as defining an infinitesimal element A, by $\frac{d}{dt}\varphi(t)\Big|_{t=0} = A$, $\varphi(t)$ being the general element of the subset. Since there are at most n parameters there can be at most n linearly independent infinitesimal elements $A_1, \ldots, A_n$. For if we parametrize G by $(t_1 \ldots t_n)$, and choose the origin of $t_1 \ldots t_n$ so that for the one-parameter subset $\varphi(t) = \varphi(t_1(t), \ldots, t_n(t))$ then $t_i(0) = 0$ $(1 \leq \tau \leq n)$. Hence

$$\frac{d\varphi}{dt}\Big|_{t=0} = A = \sum \frac{\partial\varphi}{\partial t_i}\Big|_{t_i=0} \left(\frac{dt_i}{dt}\right)_{t=0} = \sum a_i A_i$$

where

$$a_i = \left(\frac{dt_i}{dt}\right)_{t=0}, \quad A_i = \frac{\partial\varphi}{\partial t_i}\Big|_{t_i=0} .$$

Hence the set of n elements $A_1, \ldots, A_n$ form a basis of L_G. The linear space spanned by these elements is the Lie algebra of G; it is closed under the commutator bracket $[A,B] = AB - BA$. For if $\varphi(t), \theta(t)$ are differentiable one-parameter subsets, with $\varphi(0) = \theta(0) = 1$, if $A = A = \frac{d}{dt}\varphi(t)\Big|_{t=0}$, then $\theta(t)A\theta^{-1}(t)$ is a differentiable one-parameter subset so its derivative at $t = 0 \in L_G$; but

$$\frac{d}{dt}(\theta(t)A\theta^{-1}(t))\Big|_{t=0} = BA - AB$$

if $B = \frac{d\theta}{dt}\Big|_{b=0}$ so $[A,B]_- \in L_G$.

Then $[A_i, A_j] = C_{ijk}A_k$ where the C_{ijk} are the structure constants of the group.

In general, the Lie algebra of G, L_G, is the set of differential operators X acting on functions on the group by

$$Xf(g) = \frac{d}{dt} f(g\varphi(t))\Big|_{t=0}$$

where $\varphi(t)$ is any analytic one-parameter subgroup (i.e., in the parameter space if $t_i(t)$ is the ith parameter describing $\varphi(t)$ then $t_i(t)$ is an

analytic function of t, for $1 \leq l \leq n$). (L_G is the tangent space to G at the identity.)

A number of Lie groups may have isomorphic Lie algebras, such as SU(2) and SO(3). We see that we cannot always exponentiate to obtain a general group element in such a case.

Let us consider the infinitesimal elements of SO(3), $\mathfrak{L}_+^\uparrow$, SL(2C), $\mathcal{P}$ and SU(2). For $\mathfrak{L}_+^\uparrow$, we consider $\Lambda = g + \varepsilon$, so $\Lambda^T g \Lambda = g \Rightarrow \varepsilon^T = -\varepsilon$. Then ε may be written as

$$\varepsilon = \sum_{\substack{\mu,\nu=0 \\ \mu<\nu}}^{3} i\,\varepsilon_{\mu\nu} J_{\mu\nu} \quad ,$$

with $J_{\mu\nu} = -J_{\nu\mu}$, so that the 6 matrices $J_{\mu\nu}$ are the infinitesimal elements of $\mathfrak{L}_+^\uparrow$ (where $(iJ_{\mu\nu})_{\lambda\sigma} = g_{\mu\lambda}\, g_{\nu\sigma} - g_{\mu\sigma}\, g_{\nu\lambda}$). From the explicit form of $J_{\mu\nu}$ we may obtain

$$i\left[J_{\mu\nu}, J_{\lambda\sigma}\right] = g_{\mu\lambda} J_{\nu\sigma} + g_{\nu\sigma} J_{\mu\lambda} - g_{\mu\sigma} J_{\nu\lambda} - g_{\nu\lambda} J_{\mu\sigma} \qquad \text{(I.1)}$$

For T_4, we write

$$a = \sum_{\mu=0}^{3} a_\mu P_\mu$$

and thus $(a,1) = \sum a_\mu (P_\mu, 1)$. We regard the P_μ's as operators on some space, and take the $J_{\mu\nu}$'s as operators on the same space, still with $(a,1) = \sum a_\mu (P_\mu, 1)$. Evidently since T_4 is Abelian,

$$\left[P_\mu, P_\nu\right]_- = 0 \qquad \text{(I.2)}$$

whilst we may easily see that P_μ transforms as a vector under $\mathfrak{L}$:

$$\left[J_{\mu\nu}, P_\lambda\right]_- = g_{\mu\lambda} P_\nu - g_{\nu\lambda} P_\mu \qquad \text{(I.3)}$$

Thus a general infinitesimal element is of form

$$(a,\Lambda) = 1 + i\,\varepsilon_\mu P_\mu - \frac{i}{2}\,\varepsilon_{\mu\nu} J_{\mu\nu}$$

where the right hand side is regarded as an operator on some space. Then we can derive the commutation relation between P_λ and $J_{\mu\nu}$ by using the equation

$$(0,\Lambda)^{-1}(P,1)(0,\Lambda) = (\Lambda^{-1}P,1) \tag{I.4}$$

with

$$(P,1) = 1 + i\epsilon_\mu P_\mu,\quad (0,\Lambda) = 1 + \frac{i}{2}\epsilon_{\mu\nu}J_{\mu\nu},\quad (\Lambda^{-1}P,1) = 1 + i(\Lambda^{-1})_\nu P_\nu$$

so (I.4) becomes

$$\left(1 + \frac{i}{2}\epsilon_{\mu\nu}J_{\mu\nu}\right)\left(1 + i\epsilon_\lambda P_\lambda\right)\left(1 - \frac{i}{2}\epsilon_{\mu\nu}J_{\mu\nu}\right) = 1 + i(\Lambda^{-1}\epsilon)_\nu P_\nu$$

or

$$\frac{i}{2}\epsilon_{\mu\nu}\epsilon_\lambda\left[J_{\mu\nu},P_\lambda\right]_- = i\left[(\Lambda^{-1}\epsilon)_\lambda - \epsilon_\lambda\right]P_\lambda \tag{I.5}$$

But $\Lambda = 1 - \frac{i}{2}\epsilon_{\mu\nu}J_{\mu\nu}$ (where $J_{\mu\nu}$ is now the 4 x 4 matrix described above), so that $\Lambda^{-1} = 1 + \frac{i}{2}\epsilon_{\mu\nu}J_{\mu\nu}$. Thus

$$(\Lambda^{-1}\epsilon)_\lambda - \epsilon_\lambda = \frac{i}{2}\epsilon_{ab}(J_{ab})_{\lambda\tau}\epsilon_\tau\ ,$$

and from (I.5),

$$[J_{\mu\nu}P_\lambda]_- = +i(J_{\mu\nu})_{\tau\lambda}P_\tau = i(g_{\nu\lambda}P_\mu - g_{\mu\lambda}P_\nu) \tag{I.6}$$

A simple representation of P_μ and $J_{\nu\lambda}$ is given by differentiable operators acting on differentiable functions of 4-vectors:

$$P^\mu = i\frac{\partial}{\partial X_\mu}\ ,\quad J_\mu{}^\nu = i\left(X_\mu\frac{\partial}{\partial X_\nu} - X_\nu\frac{\partial}{\partial X_\mu}\right)$$

Thus the infinitesimal elements of T_4 are the P_μ, and of $\mathcal{P}$ are the set of 10 P_μ and $J_{\mu\nu}$. Further, by restricting elements of $\mathcal{L}$ to have $\epsilon_{oi} = 0$ we see that the infinitesimal elements of SO(3) are J_{23}, J_{31}, J_{12}. If we denote this set by J_1, J_2, J_3 respectively then we obtain the well-known angular momentum commutation relations $[J_i,J_k] = i\,\epsilon_{ik\ell}J_\ell$. If we define $K_i = J_{io}$ then we find from the above commutation relations that

$$\left[J_\ell,K_j\right]_- = i\epsilon_{ijk}K_k$$

$$\left[K_i,K_j\right]_- = i\epsilon_{ijk}J_k$$

If we introduce $\underline{M} = \frac{1}{2}(\underline{J} + i\underline{K})$, $\underline{N} = \frac{1}{2}(\underline{J} - i\underline{K})$ then we easily see

$$\left[M_i, M_j\right]_- = i\epsilon_{ijk}M_k\,,\ \left[N_i, N_j\right]_- = i\epsilon_{ijk}N_k\,,\ \left[M_i N_j\right] = 0\ .$$

If we were using $\underline{K}$ instead of $i\underline{K}$ (so have ix_0 instead of x_0) we would be working with $SO(4) = SO(3) \times SO(3)$; it is the extra i which causes the non-compactness of $SO(3,1)$.

H. Representation Theory[3]

A preresentation L of a group G is a homomorphism of G into a set of linear operators on a normed vector space H(L) so that $L_{g_1} L_{g_2} = L_{g_1 g_2}$, for any $g_1, g_2 \in G$. There are evident difficulties of defining a product of unbounded operators, so we restrict the representations to the case when L_g is bounded and everywhere defined for all $g \in G$. We will also require that if G is a topological group then $L_g\varphi$ is continuous in g, $\forall \varphi \in H(L)$. The dimensionsof the representation is the dimension of the space H(L).

A unitary representation is one for which L_g is unitary, $\forall_g \in G$, and H(L) a Hilbert space. Two representations L, L′ are equivalent if there exists a 1:1 linear mapping A of H(L) to H(L′) so that $AL_gA^{-1} = L_g'$, $\forall_g \in G$. It may be shown to any representation of a compact group is equivalent to a unitary representation; for a non-compact group a unitary representation can never be finite dimensional.

A subspace M of H(L) is invariant if $L_gM \subset M$, $\forall_g \in G$. L is termed irreducible if it has no non-trivial invariant subspace. If M is invariant H is a Hilbert space and L unitary then $M^\perp$ is invariant

$$[(M^\perp = (g:(g,g') = 0)\ ,\ g' \in M)]\ .$$

For if $\varphi \in M^\perp$ then for any $\psi \in M$, $(\psi, L\varphi) = (L^{-1}\psi\varphi)$. But $L^{-1}\psi \in M$, so $(L^{-1}\psi, \varphi) = 0 \Rightarrow L\varphi \in M^\perp \Rightarrow LM^\perp \subset M^\perp$ (note that for any $g \in G$, $(L_g)^{-1} = L_{g^{-1}}$). If L is reducible and M is an invariant subspace then L may be completely reduced into a direct sum of a representation L_1 on M and L_2 on $M^\perp$:

$$L = L_1 \oplus L_2 \quad \text{on} \quad M \oplus M^\perp = H(L)\ .$$

(Such a complete reduction will not be true for non-unitary representations, in general.)

The important representation is the regular representation, defined in $L_2(G,\mu) = [$complex-valued functions f on G for which $\int |f(g)|^2\, d\mu(g) < \infty$ where μ is the right-invariant Haar measure on G$]$. The regular representation T_g is defined by $T_{g_1} f(g) = f(gg_1)$, and is evidently unitary:

$$\int |(T_{g_1}\varphi)(g)|^2 \, d\mu(g) = \int |f(gg_1)|^2 \, d\mu(g) = \int |f(gg_1)|^2 \, d\mu(gg_1)$$
$$= \int |f(g)|^2 \, d\mu(g)$$

One of the basic problems of the theory of group representations is to decompose an arbitrary representation into irreducible representations. It is necessary to consider a direct integral of unitary representations. This is defined as follows. Let S be a Borel space (with a set of Borel subsets given on it) and μ a measure on S. For each $s \in S$ let L^s be a unitary representation of a group G in a Hilbert space H_s. $L_2(S,\mu,H)$ is defined as the set of cross-sections φ_s (where $\varphi_s \in H_s$, $\forall_s \in S$) for which $\int_s (\varphi_s,\varphi_s)\, d\mu(s) < \infty$. The direct integral representation $L = \int L^s d\mu(s)$ is defined by

$$(L_g\varphi)(s) = L^s{}_g \varphi_s, \quad \forall \varphi_s \in H_s,$$

and is evidently unitary. Then the first step of the decomposition of an arbitrary unitary representation of an arbitrary separable locally compact group G gives a direct integral of mutually disjoint primary representations (where two representations are mutually disjoint if no subrepresentation of one is equivalent to a subrepresentation of the other; a representation is primary if it cannot be written as the direct sum of two mutually disjoint representations). This decomposition is obtained by decomposing by means of a suitable set of projections.

We now have to analyze the primary representations. They come in three types:

Type I: equal to a direct sum of equivalent irreducible representations of form $\Sigma \oplus L_o$, where L_o is an irreducible representation.

Type II: the equivalence classes within the quasi-equivalence class to which the representation belongs form an ordered semi-group isomorphic to all the positive real numbers (the quasi-equivalence class is composed of primaries which are not disjoint; the semi-group operator is of taking direct sums, the ordering that of being equivalent to a subrepresentation). Essentially type II can always be cut in half, i.e. $2^{-n}L$ is defined, all n, so λL is, for any real number λ with $0 < \lambda < 1$.

Type III: The above ordered semi-group has just one member, ∞.

Type I contains all representations of the following groups: every compact group, every semi-simple Lie group, every locally compact commutative group. We note that we have only described the

reduction to irreps (irreducible unitary representations). We have now to consider how to construct all irreps. For a large class of groups, including the ones we are interested in, all irreps may be obtained by Frobenius' method of induction. We consider the case that $G = T \wedge \mathcal{L}$ where T is an Abelian invariant subgroup. Each irrep of T is defined by a character $\chi \in \hat{T}$ (i.e. a complex-valued homomorphism T); if T is an n-parameter group with parameters $(t_1 \ldots, t_n)$ then the characters are of the form

$$\chi(t_1 \ldots t_n) = \exp i \left(\sum_{\alpha=1}^{n} t_\alpha p^\alpha \right) ,$$

where $p^1 \ldots p^n$ are any real numbers. Let the inner automorphism defining the product in $T \wedge \mathcal{L}$ be a linear operator acting on the parameter space. We will denote the action of $\Lambda \in \mathcal{L}$ on $t \in T$ by the symbol $\wedge t$; when $n = 4$ and $G = \mathcal{P}$, $T = T_4$, then $\wedge t$ is the actual form of this linear automorphism. Let us fix the character p_o, say, and let $\mathcal{L}_{p_o}$ be the little group of p_o, that is the set of $\Lambda \in \mathcal{L}$ $\Lambda p_o = p_o$. We may identify $G/T \wedge \mathcal{L}_{p_o}$ with $\mathcal{L}/\mathcal{L}_{p_o}$. If we define the orbit through p_o as the set of all Λp_o, as $\Lambda \in \mathcal{L}$, then each point p on an orbit will identify the element Λ so that $p = \Lambda p_o$. Since Λ is unique to within multiplication by an element of $\mathcal{L}_{p_o}$ we may identify $G/T \wedge \mathcal{L}_{p_o}$ with the orbit of p_o. We now wish to set up a representation of G. Let us suppose that we have a unitary representation U of $\mathcal{L}_{p_o}$ on the Hilbert space H(U). We now consider functions φ on G with values in H(U) which satisfy the subsidiary condition

$$\varphi(\ell_{p_o} g) = e^{i\Sigma t_\alpha p_o^\alpha} U_\ell \varphi(g)$$

if $\ell \in \mathcal{L}_{p_o}$ and $\ell_{p_o} = (t, \ell) \in T \wedge \mathcal{L}_p$. We define the representation U of G on this set of functions by $U_{g_o}\varphi(g) = \varphi(gg_o)$. This will only be unitary if we can introduce a suitable invariant measure μ on $G/T \wedge \mathcal{L}_{p_o}$, so on the orbit of p_o. Provided this can be done then we consider the representation U on the functions in $L_2(G/T \wedge \mathcal{L}_{p_o}, H(U), \mu)$ which satisfy the above consistency condition, the norm in this space being $\int ||\varphi(g)||^2 d\mu(p)$ (where the norm is indepedent of the element g chosen in a particular coset of $G/T \wedge \mathcal{L}_{p_o}$ by the consistency condition). Let $\varphi(g)$ be such a function. We may regard the functions in $L_2(G/\mathcal{L}_{p_o} \wedge T, H(U), \mu)$ as functions of p, the point on the orbit of p_o which have been identified with the particular element of $\mathcal{L}/\mathcal{L}_{p_o}$. To obtain the complete operation of the representation on $\varphi(p)$ take any element $g = (t, \Lambda) \in G$. Let L_p denote a suitable representative of $G/T \wedge \mathcal{L}_{p_o}$ corresponding to p on the orbit of p_o, so that $L_p p_o = p$; t_p is termed a

boost, and belongs to $\mathcal{L}$. Then we need to calculate $U_g\varphi(p)$.

We take $\varphi(p)$ as defined by $L_p^{-1}\varphi(p) = \varphi(L_p^{-1})$ so that

$$U_g\varphi(p) = \varphi(L_p^{-1}g) = \varphi(L_p^{-1}gL_{\Lambda^{-1}p}L_{\Lambda^{-1}p}^{-1})$$

But

$$L_p^{-1}gL_{\Lambda^{-1}p} = L_p^{-1}(t,\Lambda)L_{\Lambda^{-1}p} = (L_p^{-1}t, R_{p_o})$$

where $R_{p_o} = L_p^{-1}\Lambda L_{\Lambda^{-1}p} \in \mathcal{L}_{p_o}$, since $L_p^{-1}\Lambda L_{\Lambda^{-1}p_o} = L_p^{-1}\Lambda(\Lambda^{-1}p) = L_p^{-1}p = p_o$, (and is called the Wigner rotation). Thus by the consistency conditions on the functions in $L_2(G/T \mathcal{L}_{p_o}, H(U), \mu)$ we have

$$U_g\varphi(p) = e^{i\Sigma_\alpha (L_p^{-1}t)_\alpha p_o^\alpha} \cdot U(R_{p_o})\varphi(\Lambda^{-1}p)$$

where

$$\sum (L_p^{-1}t)_\alpha p_o^\alpha = \sum (L_pL_p^{-1}t)_\alpha (L_pp_o)^\alpha = \sum t_\alpha (L_pp_o)^\alpha = \sum t_\alpha p^\alpha$$

so we obtain finally

$$U_g\varphi(p) = e^{i\Sigma_\alpha t_\alpha p^\alpha} U(R_{p_o})\varphi(\Lambda^{-1}p) \qquad (I.7)$$

A number of important properties concerning induced representations may be proved. In particular the induced representation obtained from an irrep of a little group of type I is also an irrep, and that if the orbits are suitable smooth then all possible irreps of G are obtained by induction from all possible irreps of the little groups $\mathcal{L}_{p_o}$ for varying p_o (the smoothness is that the set of points p_o, one from each orbit, may be chosen as to be an analytic subset of $\hat{T} = \underline{R}^n$. [An analytic subset is the map of a Borel function on the unit interval, and is isomorphic to the unit interval to within a set of measure zero.] Finally if all the irreps of the little group are of type I, so are those of G.

We may make (I.7) more transparent, or at least more familiar, if we spell out the action of U_g on a suitable set of basis vectors. These basis vectors will be labeled by p, and by the irrep U of $\mathcal{L}_{p_o}$. Let j denote this latter labeling; a final labeling will be λ, denoting the various other quantum numbers needed to specify states in the representation.

We may take the matrix elements of (I.7) with $\langle p,j,\lambda|$, obtaining

$$\langle p,j,\lambda|U(a,\Lambda)|\varphi\rangle = \sum_{\mu} e^{i\Sigma a_\alpha p^\alpha} D^j_{\lambda\mu}(R_{p_o}) \langle \Lambda^{-1}p,j,\mu|\varphi\rangle$$

where

$$D^j_{\lambda\mu}(R_{p_o}) = \langle p,j,\lambda|U(R_{p_o})|\Lambda^{-1}p,j,\mu\rangle$$

Since U is unitary then removing φ and replacing p by Λp, we have

$$U(a,\Lambda)|p,j,\lambda\rangle = e^{i\Sigma a_\alpha (\Lambda p)^\alpha} \sum_{\mu} D^j_{\mu\lambda}(S_{p_o})|\Lambda p,j,\mu\rangle$$

where $S_{p_o} = L_{\Lambda p}^{-1} \Lambda L_p$ is an alternative form of the Wigner rotation, again with $S_{p_o} \in \mathcal{L}_{p_o}$.

I. Representations of the Poincaré Group[4] (without parity or time inversion)

These are classified by irreps of little groups. Note that $P_\mu P^\mu$ and $W_\mu W^\mu$ are Casimir operators which commute with all the generators P_μ, $J_{\mu\nu}$, of $\mathcal{P}$; the irreps of $\mathcal{P}$ specified by values of $P^2 = P_\mu P^\mu$ and $W^2 = W_\mu W^\mu$ (W_μ is the Pauli-Lubanski vector $W_\mu = -\frac{1}{2}\epsilon_{\mu\nu\lambda\sigma} J^{\nu\lambda} P^\sigma$ with $W_\mu P^\mu = 0$, $[W_\mu, P_\nu]_- = 0$). We diagonalize P_μ also in states of any irrep, so have states $|P^2, W^2, P_\mu, ?\rangle$. We will see that all the induced reps of $\mathcal{P}$ have $W^2 = -p^2 j(j+1)$ for $p^2 \neq 0$, where j is a suitable label for the corresponding irrep of the little group of p; if $p^2 = 0$, $W^2 = +\rho^2$, where ρ is a similar label. Thus our states will be $|p_\mu, j, \lambda\rangle$, where, however, λ will not in general be invariant, e.g., it is usually taken to be the eigenvalue of J_{12}.

To show the above statement about the relation between W^2 and j will specify the possible types of little group. These are specified in terms of the standard momentum, which we will denote by p. These are of four different kinds, with correspondingly different little groups:

(1) Timelike: $\hat{p}^2 > 0$.

We may take $\hat{p} = (\pm\sqrt{\hat{p}^2}, \underline{0})$; acting in states with $P_\mu = \hat{p}_\mu$ gives $W = \sqrt{\pm\hat{p}^2}\,(0, J_{23}, J_{31}, J_{12})$, and the little group $\mathcal{L}_{\hat{p}}$ is generated by the three angular momenta, so is the rotation group SO(3), as seen by inspection. Then $-W^2|\hat{p},j,\lambda\rangle = +p^2(J_{12}^2 + J_{31}^2 + J_{23}^2)|\hat{p},j,\lambda\rangle = \hat{p}^2 j(j+1)|\hat{p},j,\lambda\rangle$ where $j = 0, \frac{1}{2}, 1, \ldots$ correspond to the known irreps D^j of SO(3).

(2) Spacelike: $\hat{p}^2 < 0$.

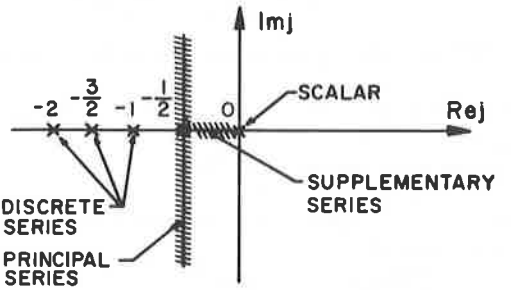
Take $\hat{p} = \sqrt{-p^2}\ (0,0,0,1)$, and W_μ reduces to $\sqrt{-p^2}\,(J_{12}, J_{20}, J_{01}, 0)$, the generators of $R_{\hat{p}}$, which satisfy the commutation rules $[J_{12}, J_{20}]_- = iJ_{01}$, $[J_{01}, J_{12}]_- = iJ_{20}$, and $[J_{20}, J_{01}]_- = iJ_{12}$. These become the standard angular momentum commutation rules for SO(3) in terms of J_{12}, iJ_{01}, iJ_{20}, so we have for these generators the invariant metric $-(x_3^2 - x_1^2 - x_2^2)$, and they generate the group SO(2,1). Again $-W^2 = \hat{p}^2 j(j+1)$ for each representation D^j of SO(2,1). These unitary irreps of SO(2,1) are of four kinds:

(a) principal series, $\mathrm{Re}\, j = -\frac{1}{2}$, $-\infty < \mathrm{Im}\, j < \infty$, all infinite dimensional, with λ taking all integer or half integer values between $-\infty$ and $+\infty$ (there are also many valued irreps with fractional λ, but these will not be considered).

(b) supplementary series, $-\frac{1}{2} < \mathrm{Re}\, j < 0$, $\mathrm{Im}\, j = 0$; again infinite dimensional, $\lambda = 0, \pm 1, \pm 2, \ldots$.

(c) discrete series, $j = -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$, of two kinds, depending on the sign of λ: D^{j+}, with $\lambda = -j, -j+1, -j+2, \ldots$ or D^{j-}, with $\lambda = j, j-1, j-2, \ldots$. They are all semi-infinite dimensional.

(d) scalar representation, $j = 0$, $\lambda = 0$, the only finite dimensional irrep of SO(2,1). We may show these irreps graphically in the j-plane:



We note that the subgroup of SL(2,C) corresponding to SO(2,1) is SL(2,R) as we see by requiring $AXA^+ = X$ with $X = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, corresponding to $\hat{p} = (0,0,1,0)$. For then $X^{-1}AXA^+ = 1 = (A^T)^{-1}A^+$, so $A^T = A^+$ or A is real. However SL(2,R) is infinitely connected with universal covering group $\underline{R}^3$, which covers it infinitely often. Also $SL(2,R) \approx SU(1,1)$.

(3) Lightlike case: $\hat{p} = 0$.

Take $\hat{p} = (\omega, 0, 0, \omega)$, so $W_\mu = (J_{12}, -\pi_2, \pi_1, J_{12})$, where $\pi_1 = J_{10} - J_{13}$, $\pi_2 = J_{20} - J_{23}$) so $[J_{12}, \pi_1] = i\pi_2$, $[J_{12}, \pi_2] = -i\pi_1$, $[\pi_1, \pi_2] = 0$, and $\mathfrak{L}_{p_0} = SO(2)$ $T(2) = E(2)$, the two-dimensional Euclidean group. The irreps of E(2) are given by $W^2 = \omega^2 (\pi_1^2 + \pi_2^2) = \rho^2$, where the rep of $SO(2) \wedge T(2)$ is D^ρ. Of course ω is not invariant. There are two kinds of irrep of E(2):

(a) Principal series, D^ρ, $0 < \rho < \infty$, Im $\rho = 0$, infinite dimensional, with λ taking all integer or half integer values between $-\infty$ and $+\infty$.

(b) Discrete series, which are all one-dimensional, $\rho = 0$, so $\pi_1 = \pi_2 = 0$ and we have only J_{12} left. Then $W_\mu |\hat{p}, 0, \lambda\rangle = \hat{p}_\mu J_{12} |\hat{p}, 0, \lambda\rangle = \lambda P_\mu |\hat{p}, 0, \lambda\rangle$, so $W_\mu = \lambda P_\mu$, and λ is a Poincaré invariant; its values are $0, \pm\frac{1}{2}, \pm 1, \ldots$, and the irrep is $D^{0\lambda}$. (These representations are those used in the theory of massless particles.)

(4) Null case: $\hat{p} = 0$.

Then $\mathfrak{L}_0 = \mathfrak{L}$. There are two Casimir operators, $\frac{1}{2} J_{\mu\nu} J_{\mu\nu} = j_0^2 - \sigma^2 + 1$, and $\frac{1}{4} \epsilon_{\mu\nu\lambda\sigma} J_{\mu\nu} J_{\lambda\sigma} = 2ij_0\sigma$. We can specify also $(J_{12}^2 + J_{23}^2 + J_{31}^2)$ and J_{12} as given when acting on the basis vectors:

$$(J_{12}^2 + J_{23}^2 + J_{31}^2) |j_0 \sigma j \lambda\rangle = j(j+1) |j_0 \sigma j \lambda\rangle$$

$$J_{12} |j_0 \sigma j \lambda\rangle = \lambda |j_0 \sigma j \lambda\rangle$$

These irreps $D^{j_0\sigma}$ of SO(3,1) are of two kinds (both infinite dimensional):

(a) Principal series: Re $\sigma = 0$, $-\infty <$ Im $\sigma < \infty$, $j_0 = 0, \frac{1}{2}, 1, \ldots$ with $j = j_0, j_0 + 1, j_0 + 2 \ldots, \lambda = -j, -j+1, \ldots, j$.

(b) Supplementary series: $0 <$ Re $\sigma < 1$, Im $\sigma = 0$, $j_0 = 0$, and $j = 0, 1, 2, \ldots, \lambda = -j, -j+1, \ldots, +j$.

We return to calculating matrix elements of the generators in these cases in Section II.

J. Harmonic Analysis on Groups[3),4)]

Let me consider the simplest case of Abelian locally compact G; then $\hat{G}$ is also an Abelian locally compact group in the topology of uniform convergence on compact subsets of G, and there are Haar measures μ, $\hat{\mu}$ for G, $\hat{G}$ respectively. Let $f(g) \in \mathfrak{L}_2(G, \mu)$, and define the Fourier transform of f as $\hat{f}(x) = \int_G f(g) x(g) d\mu(g)$. Then the Fourier inversion theorem is $f(g) = \int_{\hat{G}} f(x) \bar{x}(g) d\hat{\mu}(x)$, and the Plancherel theorem is $\int_{\hat{G}} |\hat{f}(x)|^2 d\hat{\mu}(x) = \int_G |f(g)|^2 d\mu(g)$. The measure $\hat{\mu}$ is called the Plancherel measure. We note that the Fourier transformation gives a direct integral decomposition of the regular representation of G into one-dimensional representations. For the regular representation L_g is defined by

$$L_{g_1} f(g) = f(gg_1) \quad ,$$

so

$$\widehat{L_{g_1} f} = \int f(gg_1)x(g)d\mu(g) = \int f(gg_1)x(gg_1)\bar{x}(g_1)d\mu = \bar{x}(g_1)\hat{f}(x) .$$

Thus

$$L_{g_1} f(g) = \int_{\hat{G}} \hat{f}(x)\bar{x}(g_1)x(g)d\hat{\mu}(x) \qquad (I.8)$$

where in the term in the integral with a given value of x, L_{g_1} acts through $\bar{x}(g_1)$, so is one-dimensional. Since the irreps of G are given by the one-dimensional characters then we may regard (I.8) as a decomposition of the regular representation into a direct integral of the unitary irreducible representations.

The situation is more complicated when we have a non-Abelian group, since then the unitary irreducible representations are no longer one-dimensional; if the group is not compact then none of them are even finite dimensional. It turns out that for groups of type I we may still write the regular representation as a direct integral over a suitable set of the irreducible unitary representations. It is necessary to replace the one-dimensional character $x(g)$ by an irreducible operator $T_x(g)$ satisfying $T_x(g_1 g_2) = T_x(g_1)T_x(g_2)$. Then we may define the Fourier transform now by $\hat{f}(x) = \int f(g)T_x(g)d\mu(g)$, with inverse

$$f(g) = \int \text{Trace}\left[\hat{f}(x)T_x^{+}(g)\right] d\hat{\mu}(x)$$

where

$$\int |f(g)|^2 \, d\mu(g) = \int \text{Trace}\left[\hat{f}^{+}(x)\hat{f}(x)\right]d\mu(x) \quad .$$

We can see this by the "formal" statement

$$\int d\hat{\mu}(x) \, \text{Trace}(T_x^{+}(g)T_x(g') = \delta(g - g') \quad .$$

The Plancherel measure is again determined uniquely, and the regular representation is now, using $\widehat{L_{g_1} f}(x) = T_x^{+}(g_1)\hat{f}(x)$, that

$$L_{g_1} f(g) = \int \text{Trace}\left[T_x^{+}(g_1)\hat{f}(x)T_x(g)\right]d\hat{\mu}(x)$$

and is represented by the action of $T_x^{+}(g_1)$, for each value of x (though of course it cannot be the sum of one-dimensional irreps).

It turns out that for SL(2,C) only the principal series has non-zero Plancherel measure; the supplementary series does not contribute. There is a general result that only the continuous series of irreps contribute in the Plancherel formula for non-compact groups.[3]

K. Harmonic Analysis in Special Groups[4]

(a) SU(2). We use the Euler decomposition for any rotation (so any element of SU(2)), so have

$$g = e^{i\frac{\sigma_3}{2}x}\, e^{i\frac{\sigma_1}{2}\theta}\, e^{i\frac{\sigma_3}{2}\varphi}$$

(where the corresponding SO(3) element is $e^{iJ_{12}x}\, e^{iJ_{23}\theta}\, e^{iJ_{12}\varphi}$). The harmonic analysis depends on the operators T_j (with x replaced by $j = 0, \frac{1}{2}, 1, \ldots$). We need to obtain the matrix elements of T_j in a suitable basis in H_j, the space of the representation. Let this basis be $|j,\lambda\rangle$. Then we may solve the equation $T_j(g_1 g_2) = T_j(g_1)T_j(g_2)$, by taking

$$T_{j,\lambda\mu}(g) = \langle j\lambda | g | j\mu \rangle = D_{\lambda\mu}{}^{j}(\underline{\varphi})$$

where $\underline{\varphi} = (x,\theta,\varphi)$. Then we define the Fourier coefficient

$$\hat{f}_{\lambda\mu}{}^{j} = \int f(\underline{\varphi})\, D_{\lambda\mu}{}^{j}(\underline{\varphi})^{*}\, d\mu(\underline{\varphi})$$

where the invariant measure is $[dx\,d(\cos\theta)\,d\psi]/8\pi^2$. The inverse Fourier transform is

$$f(\underline{\varphi}) = \sum_{j,\lambda,\mu} (2j+1)\, \hat{f}_{\lambda\mu}{}^{j}\, D_{\lambda\mu}{}^{j}(\underline{\varphi})$$

and the Plancherel equation is

$$\sum_{j,\lambda,\mu} (2+1)\, |\hat{f}_{\lambda\mu}{}^{j}| = \int d\mu(\underline{\varphi})\, |f(\underline{\varphi})|^2 \quad .$$

We see that the irrep of dimension (2j+1) appears (2j+1) times in the regular representation, as is known in general.

The corresponding results for $\mathcal{L}_+^{\uparrow}$ are as follows. We may again take an Euler decomposition for the group element:

$$\Lambda = \Lambda(\underbrace{\varphi,\theta,\psi}_{\text{rotation}};\ \underbrace{J;\ 0,\theta',\psi'}_{\text{boost rotation}}) = e^{-i\varphi J_{12}}\, e^{-i\theta J_{23}}\, e^{-i\psi J_{12}}\, e^{-i\zeta J_{03}} \times e^{-i\theta' J_{23}}\, e^{-i\psi' J_{12}}$$

We can see how such a parametrization occurs as follows. For any Λ we denote by $L_j(\theta)$ the boost in the j direction, $L_j(\theta) = e^{-i\theta J_{oj}}$, with velocity $|\underline{v}| = \sinh\theta$ along the j-axis. Then $\Lambda L_1(\theta)$ has the (01)-component $\Lambda_o{}^o \sh\theta + \Lambda_o{}^1 \ch\theta$. For $\Lambda \in \mathcal{L}_+^\uparrow$, $\Lambda_o{}^o > \Lambda_o{}^1$, so we may choose real θ, say θ_1, so that $\Lambda_o{}^o \sh\theta_1 + \Lambda_o{}^1 \ch\theta_1 = 0$. We may similarly choose real θ_2 and θ_3 so that $L_1(\theta_1)L_2(\theta_2)L_3(\theta_3)$ has the form of rotation matrix (if $\Lambda_{o\,3} = 0$, $1 \le j \le 3$ so that $\Lambda_o{}^o = 1$ then $\Lambda_j{}^o = 0$, $1 \le j \le 3$, since $\Lambda_o{}^{o2} = 1 + \sum_{j=1}^3 (\Lambda_j{}^o)^2$). We may reduce the boosts to one boost along the 3 axis together with a rotation about the 3 axis and one about the 1 axis, so giving the above generalized Euler decomposition of Λ.

We take the basis states of the representation $D^{j_o\sigma}$ to be $|j_o\sigma\, j\, m\rangle$, where

$$J_{12}|j_o\sigma jm\rangle = m|j_o\sigma jm\rangle,\ (J_{12}{}^2 + J_{23}{}^2 + J_{31}{}^2)|j_o\sigma jm\rangle = j(j+1)|j_o\sigma j_m\rangle\ .$$

Then

$$\begin{aligned} D^{j_o\sigma}_{j'm'jm}(\Lambda) &= \langle j_o\sigma j'm'|\Lambda(\varphi\theta\psi;\zeta;0,\theta',\psi')|j_o\sigma jm\rangle \\ &= \sum_{a,b} \langle j_o\sigma j'm'|R(\varphi\theta\psi)|j_o\sigma j'a\rangle\langle j_o\sigma j'a|e^{-i\zeta J_{o3}}|j_o\sigma jb\rangle \\ &\quad \times \langle j_o\sigma jb|R(0,\theta',\psi')|j_o\sigma jm\rangle \end{aligned}$$

Now $[J_{o3}, J_{12}] = 0$ so

$$\langle j_o\sigma j'a|e^{-i\zeta J_{o3}}|j_o\sigma jb\rangle = \delta_{ab}\, d^{j_o\sigma}_{j'aj}(\zeta)$$

and

$$D^{j_o\sigma}_{j'm'jm} = \sum_a D^{j'}_{m'a}(\varphi\theta\psi)\, d^{j_o\sigma}_{j'aj}(\zeta)\, D^{j}_{am}(0\ \theta'\ \psi')$$

where $-\min(j,j') \le a \le \min(j,j')$. (We will return to the d-functions later.)

We may now perform the Fourier transform and its inversion, using that the invariant measure on $\mathcal{L}_+^\uparrow$ is

$$d\mu = \left(\frac{1}{8\pi^3}\right) d\varphi d(\cos\theta)\, d\psi \cdot \tfrac{1}{2} \cdot \sh\zeta d(\ch\zeta) \left(\frac{1}{4\pi}\right) d(\cos\theta') d\psi',$$

where the first and last factors are the rotation measures, the middle factor the measure for boosts in the 3-direction. Then

$$\hat{f}^{j_o\sigma}_{j'm'jm} = \int d\mu(\varphi,\theta,\psi,\zeta,\theta',\psi')\, f(\varphi,\theta,\psi,\zeta,\theta',\psi') D^{j_o\sigma}_{j'm'jm}(\varphi\theta\psi\zeta\theta'\psi')$$

where $f(\varphi,\theta,\psi,\zeta,\theta',\psi') = f(\Lambda(\varphi,\theta,\psi;\zeta;0,\theta',\psi'))$. The inverse Fourier transform is

$$f(\Lambda(\varphi,\ldots)) = \int_{-i\infty}^{+i\infty} (-i)d\sigma \sum_{j_o=0,\frac{1}{2},1,\ldots} \sum_{j,m,j',m'} (j_o^2-\sigma^2)(2j+1)(2j'+1) \times \hat{f}^{j_o\sigma}_{j'm'jm} D^{*j_o\sigma}_{j'm'jm}(\varphi,\ldots)$$

We see that the Plancherel measure is concentrated on the principal series only; the supplementary series has Plancherel measure zero. Finally

$$\int |f(\Lambda(\varphi,\ldots))|^2 d\mu = \sum_{j_o=0,\frac{1}{2},\ldots} \sum_{j,m,j',m'} \int_{-i\infty}^{+i\infty} (-i)d\sigma |\hat{f}^{j_o\sigma}_{j'm'jm}|^2$$

We state the corresponding Plancherel measure for SO(2,1). This measure is concentrated on the principal series and the discrete series with $j \neq -\frac{1}{2}$; we have

$$f^{j}_{\lambda\mu} = \frac{1}{8\pi^2} \int_0^{2\pi} d\varphi \int_1^{\infty} d(\mathrm{ch}\,\beta) \int_0^{2\pi} d\psi\, f(\varphi,\beta,\psi)\, D^{j*}_{\lambda\mu}(\varphi,\beta,\psi)$$

with

$$D^{j}_{\lambda\mu}(\varphi,\beta,\psi) = \langle j,\lambda | e^{-i\varphi J_{12}} e^{-i\beta J_{01}} e^{-i\psi J_{12}} | j,\mu\rangle$$
$$= e^{-i(\psi\mu-\varphi\lambda)} \langle j,\lambda | e^{-i\beta J_{01}} | j,\mu\rangle \quad ,$$

$$f^{j\pm}_{\lambda\mu} = \frac{1}{8\pi^2} \int_0^{\pi} d\varphi \int_1^{\infty} d(\mathrm{ch}\,\beta) \int_0^{2\pi} d\psi\, f(\varphi,\beta,\psi)\, D^{j\pm*}_{\lambda\mu}(\varphi,\beta,\psi)$$

with

$$f(\varphi,\beta,\psi) = \sum_{j^{\pm},\lambda,\mu} (2j+1)\, f^{j\pm}_{\lambda\mu} D^{j\pm}_{\lambda\mu}(\varphi,\beta,\psi) + i\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \sum_{\lambda,\mu} \frac{dj(2j+1)}{2\tan\pi j} \times f^{j}_{\lambda\mu} D^{j}_{\lambda\mu}(\varphi,\beta,\psi)$$

provided that

$$\int_0^{\pi} d\varphi \int_1^{\infty} d(\mathrm{ch}\,\beta) \int_0^{\pi} d\psi \; |f(\varphi,\beta,\psi)|^2 < \infty$$

Thus the supplementary series and the discrete representations $D^{\pm}_{-\frac{1}{2}}$ are excluded, since they have measure zero (and indeed would not give square integrable functions). The principal series has $\tan \pi j$ replaced by $\cot \pi j$ when μ is half an odd integer.

The Plancherel measure for E(2) is concentrated on the principal series and equal to $\theta(\rho)d\rho^2$.

It is possible to consider indefinitely differentiable functions on the above groups, and so their duals i.e. to consider distributions on $\hat{G}$. This is important in discussions of character; we will indeed discuss generalized functions on $\hat{G}$ to define Regge poles in the next section.

II. Covariant Angular Momentum Decomposition[5)]

A. One-Particle States

We consider again the representations of the Poincaré group, but now from the point of view of one-particle states, two-particle states, and so on. The generators of $\mathcal{P}$ were given in Sec.I.G, being the operators $P_\mu, J_{\mu\nu}$, and satisfying the commutation rules:

$$[P_\mu, P_\nu]_- = 0$$

$$[P_\lambda, J_{\mu\nu}]_- = i(g_{\lambda\mu}P_\nu - g_{\lambda\nu}P_\mu)$$

$$[J_{\mu\nu}, J_{\rho\pi}]_- = i(g_{\mu\rho}J_{\nu\lambda} + g_{\nu\pi}J_{\mu\rho} - g_{\mu\pi}J_{\nu\rho} - g_{\nu\rho}J_{\mu\pi}) .$$

We also discussed the Pauli-Lubanski vector $W_\mu = -\frac{1}{2}\epsilon_{\mu\nu\lambda\rho}J^{\nu\lambda}P^\rho$ for which $P_\mu W^\mu = 0$, $[P_\mu, W_\nu]_- = 0$, $[W_\mu, W_\nu]_- = i\epsilon_{\mu\nu\lambda\rho}W^\lambda P^\rho$. In terms of the operators $K_i = J_{io}$, $J_i = \frac{1}{2}\epsilon_{oijk}J^{jk}$ we have $W_o = \underset{\sim}{J}\cdot\underset{\sim}{P}$ and $W_i = \epsilon_{ijk}K_jP_k + J_iP_o$.

The representations of $\mathcal{P}$ are labeled by the values of the Casimir operators $P_\mu P^\mu = P^2$ and $W_\mu W^\mu = W^2$, as we described in Sec. I.I. A single particle state will be labeled by the value m^2 of P^2, by the value of W^2, and certain other numbers. These other numbers will be the components p_μ of the particle's momenta and the components of the particle's spin. We have already considered the representation of $\mathcal{P}$ for $m^2 > 0$ in Sec. I.I. and have that $-W^2 = m^2 j(j+1)$, where j is the spin of the particle; also $W_\mu = m(0, J_{23}, J_{31}, J_{12})$, which are the generators of the little group SO(3) of p_μ, so that a complete commuting set

of operators describing the one-particle states will be P^2, p_μ, W^2 and one component of W_μ. Thus we have states $|P^2, p_\mu, W^2, W_a\rangle$, and we take the normalization

$$\langle m^2, p_\mu, W^2, W_a | m^2, p_\mu', W^2, W_a' \rangle \cdot \Delta^+(p) = (2\pi)^4 \delta^4(p_\mu - p_\mu')\delta(W_a - W_a')$$

where $\Delta^+(p) = 2\pi\theta(p_o)\delta(p^2 - m^2)$. Apart from m^2, W^2 (assumed the same for all particles of a given type) there will be a total of four parameters labeling any state.

B. Two-Particle States

Let the momenta be $p_\mu(1)$, $p_\mu(2)$, and let $P_\mu = p_\mu(1) + p_\mu(2)$, $Q_\mu = p_\mu^{(1)} - p_\mu^{(2)}$. The normalization of the two-particle states $|P_\mu, Q_\mu, W_{1a}, W_{2a}\rangle$ is

$$\langle P_\mu' Q_\mu' W_{1a}' W_{2a}' | P_\mu Q_\mu W_{1a} W_{2a} \rangle \Delta^+(p^{(1)})\Delta^+(p^{(2)}) = (2\pi)^8 \delta^4(P - P') \times \delta^4(Q - Q')\delta_{W_{1a}, W_{1a}'} \cdot \delta_{W_{2a}, W_{2a}'} \cdot$$

In order to have a covariant labeling we introduce (following Feldman and Matthews[6]) the notion of felicity for each particle. The value of the felicity of the ith particle is

$$\lambda_i = \frac{2 W_\mu^{(1)} p^\mu}{\Delta(t,1,2)} \cdot \frac{t}{|t|} \qquad (t = P^2)$$

where the factor $t/|t|$ is included to enable a unified discussion of the case for negative t to be given and $\Delta(t,1,2) = [t^2 + m_1^4 + m_2^4 - 2tm_1^2 - 2tm_2^2 - 2m_1^2 m_2^2]^{\frac{1}{2}}$, with m_1 and m_2 the masses of the two particles respectively. We remark that $\Delta(t,1,2)/2\sqrt{|t|}$ is the center of mass momentum in the (1,2) channel, since in the center of mass $p^{(1)} = (\sqrt{\underline{p}^2 + m_1^2}, \underline{p})$, $p^{(2)} = (\sqrt{\underline{p}^2 + m_2^2}, -\underline{p})$, so $t^{\frac{1}{2}} = \sqrt{\underline{p}^2 + m_1^2} + \sqrt{\underline{p}^2 + m_2^2}$, whose solution is $|\underline{p}| = \frac{\Delta(t,1,2)}{2\sqrt{|t|}}$. Thus in the c. of m. frame, for $t > 0$,

$$\lambda_1 = \frac{2 W_o^{(1)} \sqrt{t}}{2 |\underline{p}| \sqrt{t}} = \frac{W_o^{(1)}}{|\underline{p}|} = \frac{\underline{J}^{(1)} \cdot \underline{p}}{|\underline{p}|} \text{ (from Sec. II.A), and}$$

$$\lambda_2 = -\frac{\underline{J}^{(2)} \cdot \underline{p}}{|\underline{p}|}$$

which is precisely the usual helicity for the two particles.

We write the S-matrix element for the two-particle to two-particle scattering process $1 + 2 \to 3 + 4$ as

$$\langle P_\mu', Q_\mu', \lambda_3, \lambda_4 | S | P_\mu, Q_\mu, \lambda_1, \lambda_2 \rangle$$

with $P_\mu' = p_\mu^{(3)} + p_\mu^{(4)}$, $Q_\mu' = p_\mu^{(3)} - p_\mu^{(4)}$. By translational invariance we can extract a common δ^4-function, to give

$$\langle \ldots | S | \ldots \rangle = i(2\pi)^4 \delta^4 (P - P') \langle P', Q', \lambda_3, \lambda_4 | T | P\ Q\ \lambda_1\ \lambda_2 \rangle \quad .$$

Since P and P′ are identical and remain unchanged we will not write them in explicitly. Further only two components of Q or Q′ are needed.

Let us first see how to rotate the direction of Q′ to Q. This will be achieved by an element of the little group of P_μ. We define the orthogonal system of unit vectors

$$n^\mu = P^\mu \sqrt{|t|}/t; \quad \gamma^\mu = -(P^2 Q^\mu - (P\cdot Q) P^\mu)(\Delta(t,1,2)\sqrt{|t|})$$

$$\beta^\mu = c^{-1} \epsilon^{\mu\nu\alpha\beta} Q_\nu Q_\alpha' P_\beta, \quad \alpha^\mu = \epsilon^{\mu\nu\alpha\beta} \beta_\nu \gamma_\alpha n_\beta, \quad -c^2 = \det \begin{pmatrix} Q^2 & Q\cdot Q' & Q\cdot P \\ Q'\cdot Q & Q'^2 & Q'\cdot P \\ P\cdot Q & P\cdot Q' & P^2 \end{pmatrix}$$

Then $n^2 = -\gamma^2 = t/|t|$, so one or other of n_μ, γ_μ are time-like. Since α_μ and β_μ are orthogonal to both n_μ and t_μ then they must both be space-like. Then the generators of the little group of P_μ are $(M_\alpha, M_\beta, M_\gamma) = (W_\alpha/\sqrt{t}, W_\beta/\sqrt{t}, W_\gamma/\sqrt{|t|})$ where $W_\alpha = W_\mu \alpha^\mu$, etc. This follows since it is easy to see that $[M_\alpha, M_\beta]_- = iM_\gamma$, and the other two commutation relations involving cyclic permutations of α, β and γ follow for the orthogonal unit vectors n, γ, β, α chosen above. For $t > 0$ these generators are like J_{12}, J_{23}, J_{31}; for $t < 0$ like J_{12}, iJ_{01}, iJ_{02}, as we expect from the discussion in Sec. I.H. Now evidently β_μ is orthogonal to Q_μ and Q_μ' so that we can go from the direction of Q_μ to that of Q_μ' by a rotation about β_μ, so

$$|Q'\lambda_3\lambda_4\rangle = e^{iM_\beta \chi} |Q\,\lambda_3\lambda_4\rangle \quad .$$

χ is the angle between Q and Q′. If we work in the center of mass frame in the channel (1,2) then $\cos\chi = \underline{p}\cdot\underline{p}'/|\underline{p}|\,|\underline{p}'|$ where

$$|\underline{p}| = \Delta(t,1,2)/2\sqrt{|t|}, \qquad |\underline{p}'| = \Delta(t,3,4)/2\sqrt{|t|} \quad ,$$

and

$$s = (p^{(1)}-p^{(3)})^2 = (\sqrt{|p|^2+m_1^2} - \sqrt{|p'|^2+m_3^2})^2$$
$$- (|p|^2+|p'|^2 - 2|p||p'| \cos\chi)$$

Then

$$\cos\chi = \left[s - (\sqrt{|p|^2+m_1^2} - \sqrt{|p'|^2+m_3^2}+|p|^2+|p'|^2 \right]/2|p||p'|$$

and an evaluation of this gives

$$\cos\chi = \frac{2st+t^2 - (m_1^2+m_2^2+m_3^2+m_4^2)t + (m_1^2-m_2^2)(m_3^2-m_4^2)}{\Delta(t,1,2)\Delta(t,3,4)} \qquad (II.1)$$

Thus we wish to evaluate $\langle Q,\lambda_3\lambda_4|e^{-iM_\beta\chi T}|Q\lambda_1\lambda_2\rangle$. In order to develop a covariant angular momentum expansion we need to expand this matrix element in terms of covariant angular momentum representations. The most natural variable to choose for this is W^2, not J^2 (which is not covariant). We expand the state $|Q,\lambda_i,\lambda_j\rangle$, regarded as a function of the element L_Q of the little group of P_μ which transforms some standard vector Q_o into Q, $L_QQ_o = Q$. Thus

$$|Q,\lambda_i,\lambda_j\rangle = \Lambda(L_Q)\,|Q_o\lambda_i\lambda_j\rangle$$

We now use the Plancherel measure $\hat{\mu}$ on the little group $\mathfrak{L}_p$ of P_μ to write

$$|Q,\lambda_i,\lambda_j\rangle = \int d\hat{\mu}\,(W^2,M)|W^2M\lambda_i\lambda_j\rangle\langle W^2M\lambda_i\lambda_j|\Lambda(L_Q)|Q_o\lambda_i\lambda_j\rangle \qquad (II.2)$$

The labels for the complete set of states $|W^2M\lambda_i\lambda_j\rangle$ are eight in all (when P_μ is also taken into account); M is one of M_α, M_β or M_γ. It is convenient to take M_γ for M, since $|Q,\lambda_3\lambda_4\rangle$ is an eigenstate of M_γ with eigenvalue $(\lambda_3-\lambda_4)$. We see this from the fact that

$$W_\mu = -\tfrac{1}{2}\epsilon_{\mu\nu\lambda\sigma}(J_3+J_4)^{\nu\lambda}(p_3+p_4)^\sigma$$
$$= W_\mu^{(3)} + W_\mu^{(4)} - \tfrac{1}{2}\epsilon_{\mu\nu\lambda\sigma}(J_3{}^{\nu\lambda}p_4{}^\sigma + J_4{}^{\nu\lambda}p_3{}^\sigma)$$

Now $|Q\lambda_3\lambda_4\rangle = e^{-iM_\beta\chi}|Q'\lambda_3\lambda_4\rangle$ so that

$$M_\gamma|Q\lambda_3\lambda_4\rangle = e^{-iM_\beta\chi}(e^{+iM_\beta\chi}M_\gamma e^{-iM_\beta\chi})|Q'\lambda_3\lambda_4$$

But

$$e^{+iM_\beta\chi} W_\mu \gamma_\mu e^{-iM_\beta\chi} = W_\mu \gamma_\mu' \quad ,$$

where γ_μ' is obtained by rotating the direction of Q' from that of Q back to that of Q', so

$$\gamma_\mu' = -(P^3 Q_\mu' - (P\cdot Q')P_\mu)/(\Delta(t,3,4)\sqrt{|t|})$$

(this can be very easily seen in the center of mass system when $t > 0$, for then $\gamma_\mu = (0, \frac{p}{|p|})$, $\gamma_\mu' = (0, \frac{p'}{|p'|})$).

Then

$$W_\mu Q'^\mu |Q'\lambda_3\lambda_4\rangle = 2(-W_\mu^{(3)} + W_\mu^{(4)})|Q'\lambda_3\lambda_4\rangle \quad .$$

Then

$$M_\gamma |Q\lambda^3\lambda^4 = \frac{2t(W^3 - W^4)}{|t|\Delta(t,3,4)} P^\mu |Q\lambda_3\lambda_4\rangle = (\lambda_3 - \lambda_4)|Q\lambda_3\lambda_4\rangle$$

In a similar fashion we may prove that

$$M_\gamma |Q\lambda_1\lambda_2\rangle = (\lambda_1 - \lambda_2)|Q\lambda_1\lambda_2\rangle \quad .$$

Then

$$\langle Q'\lambda_3\lambda_4|T|Q\lambda_1\lambda_2\rangle = \int\int \langle Q\lambda_3\lambda_4|e^{-iM_\beta\chi}|W^2 M\lambda_3\lambda_4\rangle d\hat{\mu}(W^2, M\lambda_3\lambda_4)$$

$$\times \langle W^2 M\lambda_3\lambda_4|T(t)|W'^2 M'\lambda_1\lambda_2\rangle d\hat{\mu}(W'^2 M'\lambda_1\lambda_2)\langle W'^2 M'\lambda_1\lambda_2|Q\lambda_1\lambda_2\rangle \tag{II.3}$$

In order to treat the cases $t > 0$ and $t < 0$ simultaneously we replace the integration symbol by a summation symbol:

$$\int d\hat{\mu}(W^2 M\lambda_1\lambda_2\rangle = \sum_{W^2 M\lambda_1\lambda_2} f(W^2\lambda_{12})$$

where $\lambda_{12} = \lambda_1 - \lambda_2$ and where for

$t > 0$ $(SO(3))$, $f(W^2, M\lambda_1, \lambda_2) = (2j+1)\theta(j - |\lambda_{12}|)$

$t < 0$ $(SO(2,1))$, $f(W^2, M, \lambda_1, \lambda_2) = \frac{i(2j+1)}{2\tan \pi j}$ (principal series)

$= (2j+1)\theta(j + |\lambda_{12}|)\theta(-1-j)$ (discrete series)

We note that $\theta(j-|\lambda_{12}|)$ arises in the SO(3) formula for f since the state $|W^2, M, \lambda_1, \lambda_2\rangle$ is defined only for $j \geq |\lambda_{12}|$, as one can see easily in the center of mass. A similar argument holds for the θ-function in the discrete series contribution of SO(2,1), this time going into the brick-wall frame, $P_\mu = (0,0,0,\sqrt{|t|})$. We must replace the $\tan \pi j$ by $\cot \pi j$ in the principal series contribution of SO(2,1) when $|\lambda_{12}|$ is half an odd integer. The discrete series D_j^+ contributes for $\lambda_{12} > 0$; D_j^- for $\lambda_{12} < 0$.

Then we have the state normalization

$$\langle W^2, M, \lambda_1, \lambda_2 | W'^2, M', \lambda_1', \lambda_2', \rangle = \delta_{MM'} \delta_{W^2 W^2} \delta_{\lambda_1 \lambda_1'} \delta_{\lambda_2 \lambda_2'} f^{-1}(W^2, \lambda_{12})$$

where the Kronecker delta has to be taken as a Dirac delta in the principal series of SO(2,1). Then we define reduced amplitudes and functions:

$$\langle W^2 M \lambda_3 \lambda_4 | T | W'^2 M' \lambda_1 \lambda_2 \rangle = \delta_{W^2 W'^2} \delta_{MM'} f^{-1}(W^2, \lambda_{max}) \times \langle \lambda_3 \lambda_4 | T(W^2, t) | \lambda_1 \lambda_2 \rangle$$

(with $\lambda_{max} = \max(|\lambda_{12}|, |\lambda_{34}|)$, and $\delta_{MM'}$ arises since T is invariant, and we use the Wigner-Eckart theorem), and

$$\langle W^2 M \lambda_3 \lambda_4 | e^{-iM_\beta \chi} | W'^2 M' \lambda_3' \lambda_4' \rangle = \delta_{\lambda_3 \lambda_3'} \delta_{\lambda_4 \lambda_4'} \delta_{W^2, W'^2} \times f^{-1}(W^2 \lambda_{34}) d^{W^2}_{MM'}(\chi)$$

(since the felicities are Lorentz covariant $d^{W^2}_{MM'}$ will not depend on λ_3 or λ_4, since λ_3 and λ_4 only enumerate the multiplicity of the representation W^2). Evidently $d^{W^2}_{MM'}(0) = \delta_{MM'}$.

We have also to evaluate $\langle W^2, \lambda_{34}, \lambda_3 \lambda_4 | Q \lambda_3 \lambda_4 \rangle$ which we denote by $C_{34}(t)$. This can be determined from the state normalization:

$$\int \langle W^2 M, \lambda_3, \lambda_4 | Q'', \lambda_3, \lambda_4 \rangle \frac{\Delta(t,3,4)}{2(4\pi)^2 t} d\Omega(Q'') \langle Q'', \lambda_3 \lambda_4 | W'^2, M', \lambda_3, \lambda_4 \rangle = \delta_{W^2 W'^2} \delta_{MM'} f^{-1}(W^2, \lambda_{34})$$

where

$$\frac{d^4P}{(2\pi)^4} \frac{d^4Q''}{(2\pi)^4} \Delta^+(p_1) \Delta^+(p_2) = \frac{d^4P}{(2\pi)^4} \frac{\Delta(t,1,2)}{(4\pi)^2 2t} \cdot d\Omega(Q'') \quad ,$$

and we drop the d^4P integration.

Further

$$\langle Q\lambda_3\lambda_4 | e^{-iM_\beta\chi} | W^2 M\lambda_3\lambda_4\rangle = \sum_{W_1^2} \langle Q\lambda_3\lambda_4 | W_1^2\lambda_{34}\lambda_3\lambda_4\rangle f(W_1^2,\lambda_{34})$$

$$\times \langle W_1^2\lambda_{34}\lambda_3\lambda_4 | e^{-iM_\beta\chi} | W^2 M\lambda_3\lambda_4\rangle$$

$$= \sum_{W_1^2} \langle Q|\lambda_3\lambda_4 | W_1^2\lambda_{34}\lambda_3\lambda_4\rangle f(W_1^2,\lambda_{34})\delta_{W_1^2 W^2} f^{-1}(W^2,\lambda_{34})$$

$$\times d^{W^2}_{\lambda_{34},M}(\chi)$$

$$= C^*_{34}(t)\, d^{W^2}_{\lambda_{34},M}(\chi) \quad .$$

Then

$$\sum_{\lambda_1,\lambda_2} \int \langle W^2 M\lambda_1\lambda_2 | Q''\lambda_1\lambda_2\rangle \frac{d\Omega(Q'')}{2t(4\pi)^2} \Delta(t,1,2) \langle Q''\lambda_1\lambda_2 | W'^2 M'\lambda_1\lambda_2\rangle$$

$$= \sum_{\lambda_{12}} \int C^*_{12}(t)\, d^{W^2}_{M^{12}\lambda_{12}}(\chi) d\Omega(\chi) \frac{\Delta(t,1,2)}{2t(4\pi)^2}\, C_{12}(t)\, d^{W'^2}_{M'\lambda_{12}}(\chi)$$

If we take

$$C_{12}(t) = \left(\frac{8\pi t}{\Delta(t,1,2)}\right)^2$$

then this becomes

$$\frac{1}{2\pi} \int \sum_{\lambda_{12}} d^{W^2 *}_{M\lambda_{12}}(\chi)\, d^{W'^2}_{M'\lambda_{12}}(\chi)$$

which has then the normalization

$$\delta_{W^2 W'^2} \delta_{MM'} f^{-1}(W^2,\lambda_{12}) \quad ,$$

as we expect.

Thus we finally get in (II.3):

$$\langle Q'\lambda_3\lambda_4 | T | Q\lambda_1\lambda_2\rangle = C^*_{34}(t)\, C_{12}(t) \sum_{W^2} f(W^2,\lambda_{max}) d^{W^2}_{\lambda_{34},\lambda_{12}}(\chi)$$

$$\times \langle \lambda_3\lambda_4 | T(W^2,t) | \lambda_1\lambda_2\rangle \quad (\text{where } \lambda_{max} = \max(|\lambda_{12}|, |\lambda_{34}|)) \qquad \text{(II.4)}$$

Equation (II.4) was derived by Feldman and Matthews,[6] using the Sommerfeld-Watson trick. Here we have derived it without any analyticity assumptions on the generalized partial wave amplitudes $\langle\lambda_3\lambda_4|T(W^2,t)|\lambda_1\lambda_2\rangle$.

In the case of $t > 0$, (II.4) becomes the usual partial wave expansion. For $t < 0$ we recognize that it becomes the Regge background integral, along the line $\mathrm{Re}j = -\frac{1}{2}$, together with a discrete set of contributions for $\mathrm{Re}j \leq -1$. The derivation of (II.4) in this case made certain assumptions about the square integrability of a certain function on the group SO(2,1). To see what the assumption is let us consider (dropping the helicity suffices) the T-matrix element $\langle Q'P|T|QP\rangle$. Let Q_0, P_0 be standard values of Q,P obtained by a suitable Lorentz transformation L_Q^{-1}:

$$L_Q Q_0 = Q, \quad L_Q P_0 = P .$$

We also require

$$L_{Q'} Q_0' = Q', \quad L_{Q'} P_0 = P$$

(so conserving total energy and momentum). Then

$$\langle Q'P|T|QP\rangle = \langle Q_0' P_0|U(L_{Q'}^{-1},)\, TU(L_Q)|Q_0 P_0\rangle$$

where U is some unitary representation of $\mathcal{L}_+^\uparrow$. Using the Lorentz invariance of T we obtain

$$\langle Q'P|T|QP\rangle = \langle Q_0 P_0|U(L_{Q'}^{-1} L_Q)\, T|Q_0 P_0\rangle = f(g)$$

where the Lorentz transformation $g = L_{Q'}^{-1} L_Q$ is such that $gP_0 = P_0$, so g belongs to the little group of P_0. What we have done is expanded the function f(g) over the irreps of this little group of nonzero Plancherel measure. This can be done provided f(g) is square integrable over the group manifold itself. In the case $P^2 = t < 0$, when this little group is SO(2,1), the noncompact parameter of SO(2,1) is the parameter χ; the measure on SO(2,1) is $d(\cos\chi)$, where $\cos\chi$ is given by Eq. (II.1), in other words is proportional to ds. Then f(g), regarded as a function s, must decrease for large s as $s^{-\frac{1}{2}-\varepsilon}$ ($\varepsilon > 0$) for it to be square integrable over SO(2,1). It is known that unitarity places bounds on the increase of F(s,t) for large s with negative t:

$|F(s,o)| < s(\ell n\, s)^2$, $|F(s,t)| < s^{\frac{3}{4}}(\ell n\, s)^{\frac{2}{3}}$, $(t<0)$ (Froissart bound)[7]

$|F(s,o)| < s^2(\ell n\, s)^2$, $|F(s,t)| < s^{\frac{3}{2}}(\ell n\, s)^{\frac{3}{2}}$, $t<0$ (Greenberg-Low bound)[8]

which certainly allows a decrease of $F(s,t)$ faster than $s^{-\frac{1}{2}}$. However it also allows a slower decrease than this, so that our use of Plancherel measure is not justified. Indeed we expect to see Regge poles in the complex j-plane. These will correspond to nonunitary representations of $SO(2,1)$[10] and will give a contribution to the summation in (II.4) of the form

$$\sum_n d^{W_n^2}_{\lambda_3\lambda_4,\lambda_1\lambda_2}(\chi)\frac{\langle\lambda_3\lambda_4|T(W_n^2,t)|\lambda_1\lambda_2\rangle}{\tan\pi j_n}(2j_n+1) \qquad (II.5)$$

where $-W_n^2 = tj_n(j_n+1)$ is a given function of t. Such a "Regge pole" representation of $SO(2,1)$ may be of the form described by Bargmann:[9] For $g \in SU(1,1)$, $g = \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}$ with $|\alpha|^2 - |\beta|^2 = 1$ we let $f(z)$ be a function on the unit circle with norm $||f||^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(e^{i\theta})|^2 d\theta$ and obvious inner product with similar functions. The representation $\mathfrak{D}^\ell(g)$ is defined by

$$\mathfrak{D}^\ell(g)\, f(z) = (\alpha + \beta z)^{2\ell+2} f\left(\frac{\alpha z - \bar\beta}{\bar\alpha - \beta z}\right)$$

This coincides with the principal series when $\ell = -\frac{1}{2} + is$; in general it is a continuous bounded representation of g, with operator norm

$$||\mathfrak{D}^\ell(g)|| = \exp(|2\,\mathrm{Re}\,\ell + 1|\zeta)$$

where

$$g = \begin{pmatrix} e^{-i\mu} & 0 \\ 0 & e^{i\mu} \end{pmatrix}\begin{pmatrix} \mathrm{ch}\,\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix}\begin{pmatrix} e^{-i\nu} & 0 \\ 0 & e^{i\nu} \end{pmatrix}.$$

The previously mentioned unitarity bounds on the amplitudes require $\mathrm{Re}\, j_n(t) < 1$ for $t < 0$. We may also include Regge cuts as well as the supplementary series[11] but will not do this explicitly. There may also be nonunitary finite dimensional representations, corresponding to elementary particles, though only with spin $0, \frac{1}{2}$ for $t < 0$. Which of these moving or non-moving poles can actually occur evidently

depends sensitively on the large s behaviour of $F(s,t)$; this is not obtainable by symmetry considerations alone but will require properties of of a dynamical nature.

Let us now turn to the $t = 0$ situation. Here there are two possibilities: $P_\mu \neq 0$, with little group $E(2)$, or $P_\mu = 0$, with little group $SO(3,1)$. Let us take $P_\mu \neq 0$ first, since this is related to the $t \neq 0$ cases (since there is a 3-parameter little group, so the same number of quantum numbers are required for each state). Exactly the same expression (II.4) is valid in the form

$$\langle Q'\lambda_3\lambda_4|T|Q\lambda_1\lambda_2\rangle = C^*_{34}(0)\, C_{12}(0) \int_0^\infty d\rho\; d^\rho_{\lambda_{34},\lambda_{12}}(\zeta)\langle\lambda_3\lambda_4|T(\rho)|\lambda_1\lambda_2\rangle$$

The angle χ is now determined by the fact that

$$d^\rho_{\lambda_{34},\lambda_{12}}(\zeta) = \langle\rho,\lambda_{34}|e^{-i\pi_1\zeta}|\rho\lambda_{12}\rangle$$

$$= \lim_{t\to 0}\langle\rho,\lambda_{34}|e^{-iM_\beta\chi}|\rho,\lambda_{12}\rangle$$

where $M_\beta = \frac{\pi_1}{\sqrt{|t|}}$, and we take the generators of $E(2)$ to be $\omega(J_{10} - J_{13})$, $\omega(J_{20} = J_{23})$ and J_{12}. Also for $t \to 0$, $\chi \sim -a\sqrt{t}$, where

$$a = -\frac{4(s - s_t)}{(m_1^2 - m_2^2)(m_3^2 - m_4^2)}$$

so that $\zeta = -a$, where one can show that[6]

$$s_t = (m_1^2 m_4^2 - m_2^2 m_3^2)(m_1^2 + m_4^2 - m_2^2 - m_3^2)\Big(\sum_1^4 m_i\Big) X^{-1}$$

where

$$X = (m_1^2 m_4^2 - m_2^2 m_3^2)(m_1^2 + m_4^2 - m_2^2 - m_3^2) - (m_1^2 m_2^2 - m_3^2 - m_4^2)(m_1^2 + m_2^2 - m_3^2 - m_4^2)$$

A universal form of the three cases (but with $P_\mu \neq 0$) has been suggested by Feldman and Matthews.[6] They show that one may take a universal contour C in the $-W^2$-plane, assuming suitable analyticity of the functions $\langle\lambda_3,\lambda_4|T(W^2,t)|\lambda_1\lambda_2\rangle$ in $-W^2$. The contour C starts at $+\infty+i\varepsilon$, runs down about the real axis to a point between $j = |\lambda|$ and $|\lambda| - 1$ and then goes back to $+\infty-i\varepsilon$ below the real axis, where $|\lambda| = \max(|\lambda_{12}|, |\lambda_{34}|)$. Then $-W^2$ cuts the real axis between $|\lambda|t(|\lambda|-1)$ and $|\lambda|t(|\lambda|+1)$. The function $f(W^2,\lambda) = [2i\tan\pi(j-|\lambda|)]^{-1}$ for all these cases. Since $-W^2 = tj(j+1)$, when $t > 0$ the contour encircles the poles of f at the points $j = |\lambda| + n$ (n is a positive integer or zero)

and gives the usual partial wave expansion. When $t < 0$ the contour C becomes the principal series plus the terms of the discrete series with $j = -1, -\frac{3}{2}, \ldots, -|\lambda|$ which we already described (these terms arise for nonsense values, i.e. for values of j which are excluded in the case $t > 0$). Regge poles will again arise due to the fact that the scattering amplitude does not have a suitably fast asymptotic decrease for $s \to \infty$.

I want to finish this discussion by remarkling briefly on the d-functions which arise in the three different cases of SO(3), SO(2,1) and E(2).

From the nature of the commutation relations between M_α, M_β, M_γ we have

$$e^{iM_\beta\chi} M_\gamma e^{-iM_\beta\chi} = (\cos\chi M_\gamma - \sin\chi M_\alpha)$$
$$e^{iM_\beta\chi} M_\alpha e^{-iM_\beta\chi} = (\sin\chi M_\gamma + \cos\chi M_\alpha)$$

Let $U = e^{-iM_\beta\chi}$, and consider

$$-\frac{d^2U}{d\chi^2} - \cot\chi\frac{dU}{d\chi} + \frac{UM_\gamma^2 - 2\cos\chi\, M_\gamma U M_\gamma + M_\gamma^2 U}{\sin^2\chi} = M_\beta^2 U + i\cot\chi M_\beta U$$

$$+\frac{1}{\sin^2\chi}\,(UM_\gamma U^{-1}UM_\gamma - 2\cos\chi M_\gamma UM_\gamma + M_\gamma UU^{-1}M_\gamma U)$$

$$= M_\beta^2 U + i\cot\chi M_\beta U + \frac{1}{\sin^2\chi}[(\cancel{\cos\chi M_\gamma} + \sin\chi M_\alpha)UM_\gamma$$

$$- 2\cos\chi\cancel{M_\gamma UM_\gamma} + M_\gamma U(\cancel{\cos\chi M_\gamma} - \sin\chi M_\alpha)]$$

$$= M_\beta^2 U + i\cot\chi M_\beta U + \frac{1}{\sin\chi}(M_\alpha UM_\gamma - M_\gamma UM_\alpha)$$

Now

$$M_\alpha UM_\gamma - M_\gamma UM_\alpha = (M_\alpha UM_\gamma U^{-1} - M_\gamma UM_\alpha U^{-1})U$$
$$= M_\alpha(\cos\chi M_\gamma + \sin\chi M_\alpha)U - M_\gamma(-\sin\chi M_\gamma + \cos\chi M_\alpha)U$$
$$= \sin\chi(M_\alpha^2 + M_\gamma^2)U + \cos\chi\,[M_\alpha, M_\gamma]_- U$$
$$= \sin\chi(M_\alpha^2 + M_\gamma^2)U - i\cos\chi M_\beta U\,.$$

So

$$-\frac{d^2U}{d\chi^2} - \cot\chi\frac{dU}{d\chi} + \frac{UM_\gamma^2 - 2\cos\chi M_\gamma UM_\gamma + M_\gamma^2 U}{\sin^2\chi} = M^2 U \quad \text{(II.5)}$$

where $M^2 = M_\alpha^2 + M_\beta^2 + M_\gamma^2$.

But

$$W^2 = W_\mu W^\mu = W_\mu(\alpha^\mu+\beta^\mu+\gamma^\mu+n^\mu)(\alpha_\nu+\beta_\nu+\gamma_\nu+n_\nu)W^\nu$$
$$= (W_\alpha+W_\beta+W_\gamma)(W^\alpha+W^\beta+W^\gamma)$$

since

$$(\alpha^\mu+\beta^\mu+\gamma^\mu+n^\mu)(\alpha_\nu+\beta_\nu+\gamma_\nu+n_\nu) = \delta^\mu_\nu$$

and

$$W^\nu n_\nu = 0 .$$

Hence

$$W^2 = W_\alpha W^\alpha + W_\beta W^\beta + W_\gamma W^\gamma = -tM^2 .$$

Taking matrix elements of (II.5) between $|W^2, M\rangle$ and $|W^2 M'\rangle$ we obtain

$$\left[\frac{d^2}{d\chi^2} + \cot\chi\frac{d}{d\chi} - \frac{W^2}{t} - \frac{M^2 - 2MM'\cos\chi + M'^2}{\sin^2\chi}\right] d^{W^2}_{MM'}(\chi) = 0 . \quad \text{(II.6)}$$

For SO(3), $\chi = \theta$, $-W^2 = tj(j+1)$, we have

$$\left(\frac{d^2}{d\theta^2} + \cot\theta\frac{d}{d\theta} + j(j+1) - \frac{\mu^2 - 2\mu\lambda\cos\theta + \lambda^2}{\sin^2\theta}\right) d^j_{\mu\lambda}(0) = 0 \quad \text{(II.7)}$$

for SO(2,1), $\chi = i\beta$, $-W^2 = tj(j+1)$ with $t < 0$, we have

$$\left(\frac{d^2}{d\beta^2} + \coth\beta\frac{d}{d\beta} - j(j+1) - \frac{\mu^2 - 2\mu\lambda\cosh\beta + \lambda^2}{\sinh^2\beta}\right) d^j_{\mu\lambda}(\beta) = 0 \quad \text{(II.8)}$$

for $SO(2) \wedge T(2)$, $\chi = \xi\sqrt{|t|}$, $j(j+1) \to \rho^2$, so multiplying by t and letting $t \to 0$ in (II.6) gives

$$\left[\frac{d^2}{d\xi^2} + \frac{1}{\xi}\frac{d}{d\xi} + \rho^2 - \frac{(\mu-\lambda)^2}{\xi^2}\right] d^\rho_{\mu\lambda}(\xi) = 0 .$$

A similar discussion can be given for the whole SO(3,1) group, though it gives coupled differential equations for $d^{j\sigma}_{j\mu\pm1 j}$, and $d^{j\sigma}_{j\mu j'}$; these are more fully discussed in Salam et al.[6]

The solutions to (II.6) and (II.7) can be given as hypergeometric functions, using that $d^{j}_{\lambda\mu}(1) = \delta_{\lambda\mu}$:

$$d^{j}_{\mu\lambda}(z) = \left[\frac{\Gamma(j+\mu+1)\Gamma(j-\lambda+1)}{\Gamma(j+\lambda+1)\Gamma(j-\mu+1)}\right]^{\frac{1}{2}} \left(\frac{1+z}{2}\right)^{\frac{1}{2}(\mu+\lambda)} \left(\frac{1-z}{2}\right)^{\frac{1}{2}(\mu-\lambda)}$$

$$\times F\left(-j+\mu,\ j+\mu+1,\ \mu-\lambda+1,\ \frac{1-z}{2}\right) \Gamma(\mu-\lambda+1)^{-1} \quad .$$

where

$$F(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} \quad ,$$

where $(a_n) = a(a+1)\ldots(a+n-1)$.

The above are correct for the SO(3) functions and the principal series of SO(2,1); the discrete series can be obtained by analytic continuation $j,\mu,\lambda \to -j,-\lambda-1,-\mu-1$.

The $SO(2) \wedge T(2)$ functions are $d^{\rho}_{\lambda\mu}(\xi) = J_{\lambda-\mu}(\rho\xi)$ which are Bessel functions.

The SO(3,1) functions can be written down as a finite sum of suitable hypergeometric functions.[6]

Let me now turn to the case $P_\mu = 0$ $(t = 0)$, for which the little group is SO(3,1). I will not try to give an explicitly covariant analysis in this case. I consider again the T-matrix element

$$\langle p_1, W_1, W_{1a}; p_2, W_2, W_{2a} | T | p_3, W_3, W_{3a}; p_4, W_4, W_{4a} \rangle$$

where

$$p_1+p_2 = p_3+p_4 = P = 0;\ Q = p_1-p_2 = 2p_1\ ,\ Q' = p_3 - p_4 = 2p_3 \quad .$$

To be precise the state vector component describing the second particle belongs to the conjugate space to the first particle component. (It is a "bra" instead of the "ket" as written in the Dirac notation.) We first combine the particle one and two spin representations (of the rotation group) and reduce into irreducible representations, so have

$$\langle p_1 W_1, W_{1a}; p_2, W_2, W_{2a} | T | p_3, W_3 W_{3a}; p_4, W_4, W_{4a} \rangle$$

$$= \sum_{J,J'} \langle W_1, W_{1a} | JA W_2 W_{2a} \rangle \langle QJA | T | Q'J'A' \rangle \langle W_4, W_{4a} J'A' | W_3, W_{3a} \rangle$$

where we will take W_a to be J_{12}, and so $A = W_{1a} - W_{2a}$, $A' = W_{3a} - W_{4a}$.

We now consider the Lorentz transformation which takes the vector Q' into one with the direction of Q, denoted by L. We note that since $p_1 = -p_2$ and $p_1^2 = m_1^2$, $p_2^2 = m_2^2$ then $m_1 = m_2 = M$; similarly $m_3 = m_4 = \mu_3 = \mu$; the process we are considering is the forward scattering of a particle of mass M with one of mass μ. In particular we work in the frame in which $Q = (2M, 0)$, $Q' = 2(\sqrt{p^2+\mu^2}, \underline{p})$, with $\underline{p}$ along the z-axis. Then $L = e^{iJ_{03}\zeta}$, where $\zeta = \sinh^{-1}(p/\mu)$. Hence

$$\langle QJA|T|Q'J'A'\rangle = \langle QJA|TU(e^{iJ_{03}\zeta})|QJ'A'\rangle$$

We now perform an expansion of this matrix element, regarded as a function of ζ, over the little group of P_μ, i.e. over $SL(2,C)$ in the same manner as we did above for the case $P_\mu \neq 0$. Thus we have

$$\langle QJA|T|Q'J'A'\rangle = \sum_{j_o} \int_{-i\infty}^{+i\infty} (-i)d\sigma(j_o^2-\sigma^2)(2J+1)(2J'+1)D^{j_o\sigma}_{JAJ'A'}(\zeta) \times \langle JA|T^{j_o\sigma}|J'A'\rangle$$

where the reduced matrix element depends on Q and Q' only on their lengths i.e. on $Q^2 = 4M^2$, $Q'^2 = 4\mu^2$, which are constants. Also

$$D^{j_o\sigma}_{JAJ'A'}(\zeta) = \langle JAj_o\sigma|e^{iJ_{o3}J}|J'A'j_o\sigma\rangle = \delta_{AA'}\, d^{j_o\sigma}_{JAJ'}(\zeta)$$

so

$$\langle QJA|T|Q'J'A'\rangle = \delta_{AA'} \sum_{j_o=} \int_{-i\infty}^{+i\infty} (-i)d\sigma(j_o^2-\sigma^2)(2J+1)(2J'+1)d^{j_o\sigma}_{JAJ'}(\zeta) \times \langle J|T^{j_o\sigma}|J'\rangle \quad . \tag{II.9}$$

The covariant form of ζ is $\cosh\zeta = Q\cdot Q'/\sqrt{Q^2Q'^2}$.

The expression (II.9) has been extended to $P_\mu \neq 0$ by Salam et al;[6] they did this by expanding $\langle QJA|T(P)|Q'J'A'\rangle$ as a set of functions of P^2 multiplied by suitable polynomials in the components of P; each such function of P^2 will then have an expansion similar to (II.9), though with reduced matrix element $\langle J|T(P^2)^{j_o\sigma}|J'\rangle$; the expression reduces to (II.9) at $P_\mu = 0$. Again we have the problem of convergence related to (II.9); this is particularly bad in this case since for forward scattering we expect $F(s,o) \sim s$ for large s, where $s = M^2+\mu^2-2M\mu\cosh\zeta$. There is the possibility that for large enough

negative t we may have that $F(s,t)$ goes to zero faster than s^{-1} as $s \to \infty$ (if Regge behavior $s^{\alpha(t)}$ holds, with $\alpha(t)<-1$ for negative enough t), for then $\int |F(s,t)|^2 \, sds < \infty$. Before we discuss this and other problems related to Regge pole theory proper let us discuss the implications of the expansion (II.9), and in particular its extension to non-unitary representations of SL(2,C) which give a contribution of the form

$$\left(j_o^2 - \sigma^2(t)\right) d^{j_o\sigma(t)}_{JAJ'}(\zeta) \langle J| T(t)^{j_o\sigma(t)} |J'\rangle$$

where $\sigma(t)$ need not be purely imaginary (such a contribution is called a Lorentz pole, in analogy to the Regge pole-type contributions in (II.5)). The effect of such a Lorentz pole is very interesting: it generates a whole series of Regge poles $W_n^2 = W_n^2(t)$, of the form $W_n^2(t) = \sigma(t) - n - 1$, $n = 0,1,2,\ldots$. These new Regge trajectories are called daughter trajectories, and are also being discussed by Pagnamenta in his talk on the Bethe-Salpeter equation.[12] To see how the daughter trajectories arise we have to expand the representation function $d^{j_o\sigma}_{JAJ'}$ of SL(2,C) into the representation function $d^{j_o\sigma}_{\lambda\mu}$ of SU(1,1). We will assume the factorisability of the residue at the Lorentz pole, so the Lorentz pole contribution is $(j_o^2 - \sigma^2(t))\, b_J(t) b_{J'}(t) d^{j_o\sigma(t)}_{JAJ'}(\zeta)$. We are trying to expand in eigenstates of $J_{12}^2 - J_{02}^2 - J_{02}^2$, employed in the SO(3,1) representation space. Since

$$e^{-i\zeta J_{03}} = e^{i\frac{\pi}{2}J_{31}} e^{-i\zeta J_{01}} e^{-i\frac{\pi}{2}J_{31}}$$

we have

$$d^{j_o\sigma}_{j'Aj}(\zeta) = \sum_{\nu,\nu'} d^{j'}_{A\nu'}\left(-\tfrac{\pi}{2}\right) \langle j'\nu'| e^{-i\zeta J_{01}} |j\nu\rangle \, d^{j}_{\nu A}\left(\tfrac{\pi}{2}\right)$$

(where the SO(3) subgroup used in the j',j,A labeling is generated by J_{12}, J_{23}, J_{31} and has J_{12} diagonal). If we use a complete set of SO(2,1) states we have

$$d^{j_o\sigma}_{j'Aj}(\zeta) = \sum_{\nu,\nu'} \sum_{J} d^{j'}_{A\nu'}\left(-\tfrac{\pi}{2}\right) K^{j_o\sigma}_{\nu'}(J,j')^* \, d^{J}_{\nu',\nu}(\zeta) \, K^{j_o\sigma}_{\nu}(J,j) \, d^{j}_{\nu A}\left(\tfrac{\pi}{2}\right) f^{-1}(J,\nu) \qquad \text{(II.10)}$$

where $K^{j_o\sigma}_{\nu}(J,j) = \langle J,\nu| j\nu\rangle$ where the state $|J,\nu\rangle$ is an SO(2,1) state; the summation over J is over the discrete and principal series of

SO(2,1). The function (K^*K) has poles in J at $J = \sigma - n - 1$, $n = 0,1,2\ldots$ so that the contributions from such poles are obtained by writing down the residue in (II.10) (having displaced the J contour to encircle such poles). The resulting contribution then becomes the sum of factorisable contributions of Regge poles at $j_n(t) = \sigma(t) - n - 1$. These are the daughter poles. An alternative approach to daughters is to use O(4) in place of O(3,1). This symmetry group arises by "Wick rotation" from O(3,1) i.e. by replacement of x_0 by ix_0. Such a replacement is valid in perturbation theory. This has been investigated considerably in the Bethe-Salpeter equation. In the ladder approximation this equation is

$$M(Q,Q',P) = \frac{1}{\frac{1}{4}(Q-Q)^2+m^2} + \lambda \int \frac{d^4Q''\ M(Q'',Q',P)}{[\frac{1}{4}(Q-Q'')^2+m^2][\frac{1}{4}(P+Q'')^2-M^2][\frac{1}{4}(P-Q'')^2-\mu^2]} \tag{II.11}$$

where m is the mass of the exchanged particle, and we have assumed equal masses for the target and projectile in the t-channel, $m_1 = m_3 = M$, $m_2 = m_4 = \mu$. We may perform the Wick rotation for $P^2 < (M+\mu)^2$, both in Q, Q' and the internal variable Q''. If we take $P = (\varepsilon, 0)$ then (II.11) becomes

$$M(Q,Q',\varepsilon) = \frac{1}{[\frac{1}{4}(Q_0-Q_0')^2+\frac{1}{4}(\underline{Q}-\underline{Q}')^2+m^2]}$$

$$+\lambda \int \frac{d^4Q''\ M(Q'',Q',\varepsilon)}{[\frac{1}{4}(Q-Q'')^2+m^2][\frac{1}{4}(i\varepsilon+Q_0'')^2+\underline{Q}''^2+M^2][\frac{1}{4}(i\varepsilon-Q_0'')^2+\underline{Q}''^2+M^2]} \tag{II.12}$$

where $\frac{1}{4}(Q-Q'')^2$ is evaluated in the Euclidean metric. If we take $\varepsilon = 0$ we see the explicit O(4) invariance of $M(Q,Q',0)$, since the inhomogeneous term and the kernel of the integral equation (II.12) both have this invariance. In such a case the solution of (II.12) may be obtained by expanding in terms of 4-dimensional harmonic functions, which involve replacing the ordinary Legendre functions P_ℓ by Gegenbauer polynomials $C_{\alpha-\ell}^{\ell-1}$. The resulting solution is found to have daughter trajectories as for the SO(3,1) case.

To see how the expansion in Gegenbauer polynomials goes, let me reconsider the partial wave B.S. equation after its Wick rotation (as given in Pagnamenta's notes[12]):

$$M_\ell(Q_4, |\underline{Q}|; Q_4', |\underline{Q}'|) = V_\ell(Q_4, |\underline{Q}|; Q_4', |\underline{Q}'|)$$

$$+\frac{1}{2\pi^2} \int \frac{dQ_4''|\underline{Q}''|^2 d|\underline{Q}''| V_\ell(Q_4|\underline{Q}|; Q_4''|\underline{Q}''|) M_\ell(Q_4'', |\underline{Q}''|, Q_4', |\underline{Q}'|}{[\frac{1}{4}(i\sqrt{t}-Q_4'')^2+\underline{Q}''^2+m^2][\frac{1}{4}(i\sqrt{t}+Q_4'')^2+\underline{Q}''^2+M^2]}$$

where

$$V_\ell(Q_4,|\underline{Q}|,Q_4',|\underline{Q}'|) = \frac{2}{\pi|\underline{Q}||\underline{Q}'|}\, Q_\ell\, \frac{(Q_4+Q_4')^2+\underline{Q}^2+\underline{Q}'^2+4\mu^2}{8|\underline{Q}||\underline{Q}'|}$$

The Gegenbauer polynomials $C_n^\lambda(x)$ are orthonormal polynomials on the interval $[-1,1]$ with respect to the function $(1-x^2)^{\lambda-\frac{1}{2}}$, and so

$$\int_{-1}^{+1} C_n^\lambda(x)\, C_m^\lambda(x)(1-x^2)^{\lambda-\frac{1}{2}}\, dn\ \delta_{nm}$$

and satisfy the differential equation

$$(1-x^2)y'' - (2\lambda+1)xy' + n(n+2\lambda)y = 0$$

for $y = C_n^\lambda(x)$. It is convenient to use the functions

$$D_n^{\ell+1}(x) = 2^\ell\Gamma(1+\ell)\left[\frac{2\Gamma(1+n)(\ell+n+1)(1-x^2)^\ell}{\pi\Gamma(n+2\ell+2)}\right]^{\frac{1}{2}} C_n^{\ell+1}(x)$$

which are orthonormal in n on $(-1,+1)$ with respect to $(1-x^2)^{\frac{1}{2}}$. Using the addition theorem for the Gegenbauer polynomials:[15]

$$C_n^1(\bar{\omega}) = \frac{\pi}{2(n+1)} \sum_{\ell=0}^{n} D_{n-1}^{\ell+1}(\omega')(2\ell+1)P_\ell(z)$$

where $\bar{\omega} = \omega\omega' + z\sqrt{+\omega^2}\sqrt{1-\omega'^2}$, we can expand the potential V_ℓ and the function M_ℓ in Gegenbauer polynomials in the angles between Q or Q′ and the axis: $\omega = Q\cdot P/\sqrt{Q^2}\sqrt{P^2}$, $\omega' = Q\cdot P'/\sqrt{Q'^2}\sqrt{P^2}$, $\bar{\omega} = Q\cdot Q'/\sqrt{Q^2}\sqrt{Q'^2}$, and $z = \frac{Q-Q'}{|\underline{Q}||\underline{Q}'|}$.

Then

$$M_\ell(Q_4,|\underline{Q}|;Q_4',|\underline{Q}'|) = \pi \sum_{\lambda,\mu=0}^{\infty} M_{\ell,\lambda\mu}(Q,Q',t)\frac{D_\lambda^{\ell+1}(\omega)D_\mu^{\ell+1}(\omega')}{2(\ell+\lambda+1)} \qquad \text{(II.13)}$$

where $Q = \sqrt{Q^2}$, $Q' = \sqrt{Q'^2}$.

$$I_\ell(Q_4,|\underline{Q}|;Q_4',|\underline{Q}'|) = \pi \sum_{\lambda=0}^{\infty} \frac{I_{\ell\lambda}(Q,Q',t)}{2(\ell+\lambda+1)}\, D_\lambda^{\ell+1}(\omega)\, D_\lambda^{\ell+1}(\omega') \qquad \text{(II.14)}$$

where now the O(4) symmetry of I requires $\lambda = \mu$ in comparison with Eq. (II.13)

$$I_{\ell,\lambda}(Q,Q',t) = \frac{1}{2\pi QQ'} \cdot E_{\ell+\lambda}\left(\frac{Q^2+Q'^2+\mu^2}{2QQ'}\right)$$

where E_n is the function of the second kind related to C_n^1 in a similar way that Q_ℓ is related to P_ℓ. This arises from the expansion

$$[\mu^2+(Q-Q')^2]^{-1} = \frac{1}{2QQ'} \sum_{n=0}^{\infty} E_n\left(\frac{Q^2+Q'^2+\mu^2}{2QQ'}\right) C_n^{(1)}(\bar{\omega})(-1)^n \qquad \text{(II.15)}$$

and the addition theorem for $C_n^{(1)}(\bar{\omega})$ is used, so that

$$[\mu^2+(Q-Q')^2]^{-1} = \frac{\pi}{2QQ'} \sum_{n=0}^{\infty} E_n\left(\frac{Q^2+Q'^2+\mu^2}{2QQ'}\right) \frac{1}{2(n+1)} \sum_{\ell=0}^{n} D_{n-\ell}^{\ell+1}(\omega)$$

$$D_{n-\ell}^{\ell+1}(\omega')(2\ell+1)P_\ell(z)(-1)^n$$

so that the ℓth partial wave projection of the exchange term is as given in (II.14), taking $n-\ell = \lambda$. The resulting equation for $T_{\ell,\lambda\mu}$ is

$$T_{\ell,\lambda\mu}(Q,Q',t) = I_{\ell,\lambda}(Q,Q')\,\delta_{\lambda\mu} + \frac{1}{4\pi} \sum_{\nu=0}^{\infty} \frac{1}{(\ell+\nu+1)}$$

$$\times \int dQ''\,(Q'')^3 I_{\ell,\lambda}(Q,Q'') M_{\ell,\nu\mu}(Q'',Q',t)\, G_{\lambda\nu}^{\ell}(Q'',t) \qquad \text{(II.16)}$$

with

$$G_{\lambda\nu}^{\ell}(Q,t) = \int_{-1}^{+1} \frac{d\omega\,(1-\omega^2)^{\frac{1}{2}} D_\lambda^{\ell+1}(\omega) D_\nu^{\ell+1}(\omega)}{[Q^2+m^2-\frac{1}{4}t-i\sqrt{t}\,Q\omega][\frac{1}{4}Q^2+M^2-\frac{1}{4}t+i\sqrt{t}\,Q\omega]}$$

Evidently when $t = 0$, $G_{\lambda\nu}^{\ell}(Q) = \delta_{\lambda\nu}[Q^2+M^2]^{-1}[Q^2+M^2]^{-1}$ and we have $M_{\ell,\lambda\mu} = \delta_{\lambda\mu} M_{\ell,\lambda}$ and the Eq. (II.16) decouples. Let us take this case from now on; (II.16) becomes

$$M_{\ell,\lambda}(Q,Q') = I_{\ell,\lambda}(Q,Q') + \frac{1}{4\pi(\ell+\lambda+1)} \int \frac{dQ''(Q'')^3 I_{\ell,\lambda}(Q,Q'') M_{\ell,\lambda}(Q'',Q)}{[Q''^2+M^2][Q''^2+m^2]}$$

If we consider this equation as a Fredholm equation then the Fredholm determinant can have zeros in the variable $(\lambda+\ell)$ which now replaces

ℓ in the partial wave B.S. equation arising from (II.12); such a zero is known to generate a Regge trajectory, so we have

$$M_{\ell,\lambda}(Q,Q') \sim b(Q,Q')/[\ell+\lambda-\alpha(0)] \quad .$$

Then the partial wave projection M_ℓ will have poles at $\ell = \alpha(0), \alpha(0)-1, \ldots$ etc. (the poles at $\alpha(0)-$(odd integer) turn out to decouple if the $(-1)^n$ in the expansion (II.15) is used correctly; a further set arise at $\alpha(0)-1$, $\alpha(0)-3$, etc. These are the daughters that we saw from O(3,1). They can be followed a little way from $t = 0$ though this has not been done analytically beyond the first daughter;[13] the trajectories are then quite far to the left in the ℓ-plane. A numerical analysis of the daughters to t near threshold is discussed in Dr. Pagnamenta's lectures;[12] their integer spacing is approximately preserved.[16]

We see that the family of trajectories can only be reached on the mass shell by taking the momenta of the particles to be infinite, for the particles 1 and 2 have momenta $\frac{1}{2}(P \pm Q)$ and masses M and μ; this is different from the O(3,1) analysis which gave Toller poles with particles 1 and 2 having the same mass. However the Salam et al[14] extension away from $P_\mu = 0$ will give Toller poles for unequal masses for all four external particles. Hence the limitation of the result for the B.S. equation to equal mass scattering can be removed.

Indeed the simplest case of daughter trajectories is in non-relativistic scattering by a Coulomb potential, where the complete symmetry group of the system is O(4), not O(3) (giving rise to the degeneracy of the levels of the hydrogen atom for differing ℓ but the same n). We have for this case $E = -\alpha/(\ell+n_r+1)^2$ are the energy levels, where n_r is the radial quantum number ($n_r = 0,1,2,\ldots$). Then the Regge poles lie at $\ell(E) = -1-n_r+\sqrt{\frac{\alpha}{-E}}$; these are always in a family of trajectories, one member of the family occurring for each value of n_r. The relativistic B.S. equation with zero mass exchange also has this family behaviour; the equation possesses O(4) symmetry[17] even for non-zero energy, and so have the same degeneracy as the hydrogen atom. Then for this case

$$\ell(E) = -1-n_r+f^{-1}(E)$$

where $E=f(1+n_r+\ell) = f(n)$, where n is the principle quantum number. We expect that $E \to 0$ as $n \to \infty$ so that $\ell(E) \to +\infty$ as $E \to 0-$.

There is another reason for introducing daughters: to prevent a singularity at $t = 0$ which arises from a single Regge pole contribution in the t-channel. Such a singularity can occur in the scattering of equal mass particles with spin (the scalar case has no such singularity) or in the scattering of unequal mass particles. In this case

the Regge pole contribution has a threshold factor $(|Q||Q'|, \alpha(b)$, and as $t \to 0$, for unequal mass particles, we have

$$|Q| = Q^2 + (P\cdot Q)^2/t \to \infty, |Q'| \to \infty$$

The cancellation of this singularity is achieved by the first daughter, with pole position $\alpha_1(t) = \alpha(t)-1$. Similarly this daughter has to have its singularity (which is not completely canceled by the parent trajectory) canceled by a further trajectory at $\alpha_1(t)-1$, and so it goes.[18]

We note finally that the generalisation of Lorentz poles to the case $P^2 \neq 0$ indicates that the daughter trajectories which are born from a given parent should remain at integral spacing for all P^2; Dr. Pagnamenta's[12] trajectories (for $M/\mu = 10$) have this property (to within the limits of computational errors); the results of Chung and Snider[16] (for $M/\mu = 1$) seem to disagree with this, at least for the second daughter trajectory, though again computing errors may bring this result into question. It would seem that this situation is still open.

Let me now turn to the problem of extending the limits of validity of the group theoretic expansion over the little group of P_μ so as to be able to incorporate Regge poles.[19] In order to do this let me now use normalizable states. We have (for spinless particles, for convenience)

$$\iint |p_1 p_2\rangle \frac{d^4p_1}{(2w)^4}\frac{d^4p_2}{(2w)^4}\delta(p_1^2-m_1^2)\delta(p_2^2-m_2^2)\langle p_1 p_2| = \iint |PQ\rangle d^4p \times \frac{\Delta(t,1,2)}{(2\pi)^4 2t} d\Omega(Q)\langle PQ| = I$$

so the 2 particle space of normalisable wave functions $f(p_1, p_2)$ with norm

$$||f||^2 = \int\int d^4p_1\, d^4p_2\, \delta(p_1^2-m_1^2)\,\delta(p_2^2-m_2^2)\,|f(p_1 p_2)|^2$$

can be expressed as functions of P, Q with norm

$$||f||^2 = \int\int d^4p\, d\Omega(Q)\frac{\Delta(t,1,2)}{(2\pi)^4 2t}\,|f(P,Q)|^2$$

where $P = p_1+p_2$, $Q = p_1-p_2$. If we introduce the function $\tilde{f}(P,Q) = \left(\frac{\Delta(t,1,2)}{2t}\right)^{\frac{1}{2}} f(P,Q)$ the norm becomes

$$||f||^2 = \int dt \int d^4P\, \delta(P^2-t) \int d\Omega(Q)\,|\tilde{f}(P,Q)|^2$$

We now wish to finish the reduction of the direct product of the one-particle states. Let Q_o be a standard position of $Q\{(m,0)$ if $Q^2>0$, $(0,0,0,1)$ if $Q^2<0\}$. Then we can project out the Q dependence of $\tilde{f}(P,Q)$ by integrating over the little group of P, with

$$\tilde{f}_{MM'}(P,W^2) = \int \tilde{f}(P,Q)\, D^{W^2}_{MM'}(L_Q)\, d\Omega(Q)$$

where $L_Q Q_o = Q$, $L_Q P = P$; we will have

$$\int |\tilde{f}(P,Q)|^2\, d\Omega(Q) = \sum_{W^2,MM'} |\tilde{f}_{MM'}(P,W^2)|^2$$

Let us now consider the T-matrix element heuristically, neglecting problems of convergence, so that

$$(f_{12},Tf_{34}) = \int f_{12}^*(p_1,p_2) f_{34}(p_3 p^4) \prod_{j=1}^{4} \delta(p_i^2 - m_i^2)\, d^4 p_i\, T(p_1 p_2 p_3 p_4)$$
$$\times \delta^4\left(\sum_{i=1}^{4} p_i\right)$$

Using the group property of $D^{W^2}_{MM'}(L_Q)$ we can reduce this expression using

$$T(p_1 p_2 p_3 p_4) = T(P,Q,Q') = \sum_{W^2,M,M'} d^{W^2}_{MM'}(\chi_{QQ'})\, T_{M'M}(W^2,t) \quad .$$

where $\chi_{QQ'}$ is the angle χ defined in Eq. (I.8). Then

$$(f_{12},Tf_{34}) = \int d^4P\, \delta(P^2-t) f_{12}(P,Q) f_{34}(P,Q') \sum_{W^2,M,M'} d^{W^2}_{MM'}(\chi_{QQ'})$$
$$\times T_{M'M}(W^2,t)\, d\Omega(\Omega)\, d\Omega(Q')dt$$
$$= \sum_{M,M',W^2} \int \tilde{f}_{(12,34)MM'}(W^2,t) T_{M'M}(W^2,t)\, dt$$

where

$$\tilde{f}_{(12,34)MM'}(W^2,t) = \int d^4P\, \delta(P^2-t) f_{12}(P,Q) f_{34}(P,Q') d^{W^2}_{MM'}(\chi_{QQ'})$$
$$\times\, d\Omega(Q)\, d\Omega(Q')$$
$$= \sum_{M''} \int d^4P\, \delta(P^2-t)\, \tilde{f}_{(12)MM''}(P,W^2)\, \tilde{f}_{(34)M''M'}(P,W^2)$$

Thus we can represent T as a bilinear functional on the space H_{1234} of functions of four momenta, with norm

$$||\tilde{f}_{12,34}||^2 = \int \sum_{W^2,M,M''} |\tilde{f}_{(12),M,M''}(P,W^2)|\, |\tilde{f}_{(34)M,M''}(P,W^2)|^2\, d^4P$$

We note that H_{1234} with this norm is the normed space formed from H_{12}, H_{34}, each with norm

$$||f_{(12)}||^2 = \int \sum_{W,M,M'} |f_{(12)MM'}(P,W^2)|^2\, d^4P \quad .$$

The norm in H_{1234} is obtained as the

$$\inf_i \sum_i ||f_{12}^{(i)}|| \otimes ||f_{34}^{(i)}|| \quad ,$$

the infimum being taken over all possible expressions of $f_{12} \otimes f_{34}$ as

$$\sum_i f_{12}^{(i)} \otimes f_{34}^{(i)} \quad .$$

We may deduce from unitarity that

$$|\langle f_{12}, Tf_{34}\rangle| = |\langle \tilde{f}_{1234}, T\rangle| = |\langle f_{12}, (s-1)f_{34}\rangle|$$
$$\leq 2||f_{12}||\, ||f_{34}|| \leq 2||\tilde{f}_{1234}||$$

(by the property of the norm on H_{1234}, sometimes called the greatest cross norm, since it is greater than all other cross norms, such as $||f_{12}||\, ||f_{34}||$). Hence T is a bounded linear functional on H_{1234}. Thus we have to obtain the expression of the most general bounded linear functional on H_{1234}, which is also invariant. Now this can have more general expressions than

$$\sum_{MM',W^2} \tilde{f}_{(1234)MM'}(W^2,t)\, T_{M'M}(W^2,t)dt$$

with the summation over W^2 along the principal and suitable parts of the discrete series. It has been shown by Toller[19] that we may allow functionals of the above form with the summation over W^2 taken as an integral along a contour with $\mathrm{Re}\ell \leq 1$ (for the range of $t < 0$); the restriction $\mathrm{Re}\ell \leq 1$ corresponds closely to the unitarity restriction I discussed earlier. The restriction arises from the fact that

$$|d^{\ell}_{MM'}(\zeta)| < P_{\mathrm{Re}\ell}(\cosh 2\zeta) < |\cosh 2\zeta|$$

for complex ℓ restricted to $\mathrm{Re}\ell \leq 1$; for ℓ in this region it is possible to show that any function $f_{1234} \in H_{1234}$ gives a function $\tilde{f}_{(1234)MM'}(P,W^2)$ which is analytic in W^2 (or ℓ) in this region. Thus the Regge pole or Regge cut contributions can be allowed. The final problem involved with this approach is to show that the Regge type of contribution of Eq. (II.5) is actually a Regge pole contribution i.e. that it is the continuation of the crossed channel SO(3) partial wave expansion. In other words that the expression for the scattering amplitude, for $t > 0$ and $s < 0$ given by expanding in partial waves and then performing the Sommerfeld-Watson trick to give a background contribution plus Regge poles:

$$M(s,t) = \sum_n \frac{\beta(t)(2\alpha_n(t)+1)\, Q_{\alpha_n(t)}(\cos\theta_t)}{\sin\pi\,\alpha_n(t)} + \int_{-\frac{1}{2}-i\infty}^{+\frac{1}{2}+i\infty} \frac{(2\ell+1)d\ell Q_{\ell}(\cos\theta_t)}{\sin\pi\ell}$$

will continue to the region $t < 0$ and $s > 0$ preserving its form. This has been shown by Boyce[20] under the condition that the functions involved are square integrable over the group manifold and that there are no Regge poles to the right of $\mathrm{Re}\ell = -\frac{1}{2}$ for $t > 0$. I will not give this proof; it evidently has to be extended within the functional framework I gave above; this has not been done yet.

Let me finish this discussion of the theoretical aspects of Regge poles and group theory by briefly describing three alternative approaches to the subject. The first involves the complex inhomogeneous Lorentz group (CILG). It is assumed[21] that the S-matrix is invariant under CILG (an assumption which may not be valid if the S-matrix is defined by means of time ordered products of some field operators). Non-unitary representations which are unitary on RILG (the real inhomogeneous Lorentz group) are considered which are irreducible; the states of such a representation are classified by a complex 4-vector

$P_\mu + iQ_\mu = \bar{P}_\mu$ with complex mass $(P_\mu + iQ_\mu)^2 = M$. For $M \neq 0$ the little group of $P_\mu + iQ_\mu$ is the complex rotation group (isomorphic to $SL(2,C)$), with irreducible representations labeled by $\chi = (\rho, n)$, where ρ is a complex number and n an integer. The single particle states transform as

$$U(a,L)|P_\mu + iQ_\mu\rangle = e^{iP_\mu a^\mu} |L_\mu{}^\nu P_\nu + iL_\mu{}^\nu Q_\nu\rangle \quad .$$

Then the S-matrix for 2-particle scattering can be reduced in the usual way:

$$\langle \bar{P}_\mu^{(1)} \bar{P}_\mu^{(2)} | S | \bar{P}_\mu^{(3)} \bar{P}_\mu^{(4)} \rangle = \int_\Omega d\chi \, \langle \bar{P}_\mu, g^{12} | S(2) | \bar{P}_\mu g^{34} \rangle$$

where g^{12}, g^{34} are suitable complex rotations. Then as $s \to \infty$ it can be shown that the S-matrix element behaves as $t^{\frac{1}{2}(|\mathrm{Re}\rho| - 2)}$ from a given point χ; by taking a set of point-contributions in the integral i.e. δ-functions in χ, asymptotic Regge pole behaviour is obtained, with χ playing the role of ℓ. This approach evidently suffers from the difficulty that S-matrix elements are not in general invariant under CILG; that Regge poles can occur will depend on certain dynamical properties (for example the use of unitarity and analyticity, etc.).

An apparently alternative approach[22] has been to set up local representations of SO(3), that is representations of only a portion of the whole SO(3) group manifold. Such local representations can be associated with complex angular momentum. These representations ar are non-unitary; I have already discussed how to deal with non-unitary representations of SO(2,1) and SO(3), and such representations are indeed the ones we required. Thus this approach is contained in the functional discussion we gave, at least as far as non-unitary representations defined over the whole group manifold. It is not at all clear that local non-global representations are still required; further discussion has to be given of this.

Finally a far more general approach has been set up by Lurcat,[23] by his theory of fields on the Lorentz group. Here the mass and spin are treated in an equivalent fashion so that a particle may have both a mass and a spin spectrum. However the theory is too sketchy to be said to exist dynamically yet.

Acknowledgments

I would like to thank Professor Mahanthappa for the stimulating hospitality extended during my stay at the Institute for Theoretical Physics, W. Klink and A. Pagnamenta for very useful discussions, and P. Matthews, G. Feldman, A. Salam, R. Delbourgo, S. F. Boyce and J. Strathdee for communicating their results prior to publication.

References

1. See, for example, L. S. Pontryagin, Topological Groups (Gordon and Breach, New York, 1966).
2. See, for example, P. M. Cohn, Lie Groups (Cambridge University Tracts in Mathematics, Cambridge University Press, 1957).
3. On the general properties of group representations:
   a) G. W. Mackey, Bull. Am. Math. Soc. 69, 628 (1963): "Infinite Dimensional Group Representations," and references therein.
   b) G. W. Mackey, "Group Representations," lecture notes, Mathematical Institute, Oxford (1967) (unpublished).
   c) G. W. Mackey, Induced Representations of Groups and Quantum Mechanics (W. A. Benjamin, Inc., New York, 1968).
   d) I. E. Segal, Bull. Amer. Math. Soc. 71, 419 (1965) and references therein.
4. For more detailed discussions of representations of $\mathcal{P}$, $\mathcal{L}$, SL(2,C) etc., see:
   a) M. A. Naimark, Linear Representations of the Lorentz Group (Pergamon Press, Oxford, 1964).
   b) I. M. Gelfand, M. I. Graev and N. Vilenkin, Generalized Functions, Vol. 5 (Academic Press, N. Y., 1966).
   c) W. Rühl, "Lectures on the Lorentz Group and Harmonic Analysis," CERN lecture notes, 1967-68, (unpublished), and references therein.
5. The original suggestion of using a group theoretic approach to complex angular momentum and partial wave expansions was given in H. Joos, Lectures in Theoretical Physics, Vol. VIIA, ed. by Wesley E. Brittin and Asim O. Barut (University of Colorado Press, Boulder, 1965), p. 132.
6. G. Feldman and P. Matthews, Phys. Rev. 168, 1587 (1968). See also J. Strathdee, R. Delbourgo and A. Salam, Phys. Rev. 164, 1981 (1967).
7. M. Froissart, Phys. Rev. 123, 1053 (1961).
8. W. Greenberg and F. E. Low, Phys. Rev. 124, 2047 (1961).
9. V. Bargmann, Annals of Mathematics 48, 568 (1947).
10. M. Toller, Nuovo Cimento 37, 631 (1965).
11. F. Hadjioannou, Nuovo Cimento 44, 185 (1966).
12. A. Pagnamenta, this volume.
13. See, for example, D. Z. Freedman and J. M. Wang, Phys. Rev. 153, 1596 (1967). V. Chung and D. R. Snider, Phys. Rev. 162, 1639 (1967). A. Pagnamenta, Ref. 12, and references quoted therein.
14. R. Delbourgo, A. Salam, and J. Strathdee, "Partial Wave Analysis in Terms of the Homogeneous Lorentz Group," Trieste preprint (1968) (unpublished).

15. The Gegenbauer polynomials are discussed, for example, in the Bateman Manuscript Project, California Institute of Technology, Higher Transcendental Functions, Vol. 1 (McGraw-Hill, Inc., New York, 1953).
16. See also Chung and Snider, Ref. 13.
17. R. E. Cutkosky, Phys. Rev. 96, 1135 (1954).
18. See Freedman and Wang, Ref. 13. J. C. Taylor, Oxford University preprint (1967) (unpublished).
19. See M. Toller, Nuovo Cimento 54, 295 (1968). We follow the ideas presented in this reference, though the detailed working out of these ideas is presented in a very different fashion.
20. J. F. Boyce, J. Math. Phys. 8, 675 (1967).
21. E. H. Roffman, Phys. Rev. Letters 16, 210 (1966), and Brandeis University thesis (unpublished).
22. See, for example:
    a) M. Andrews and J. Gunson, J. Math. Phys. 5, 1391 (1964).
    b) G. Domokos, Nucl. Phys. 47, 124 (1963).
23. F. Lurcat, Physics 1, 95 (1964).

A THEOREM ON ANALYTIC PROPERTIES OF UNITARY REPRESENTATIONS†

C. Itzykson‡*
Stanford Linear Accelerator Center
Stanford University
Stanford, California

Unitary representations of non-compact Lie groups have been investigated recently with the aim of applying them to various problems of particle physics. Much of the basic facts about these representations have been worked out by mathematicians and physicists in the last twenty years. However new developments tend to concentrate on certain particular aspects such as analyticity properties on which less seems to be known. Consider for instance the compact groups. It is known in this case that all their (irreducible) representations are finite dimensional, equivalent to unitary ones, and furthermore obtainable through the decomposition of tensor products of some fundamental one. As a result the matrix elements of these representations are polynomials in the elements of the fundamental representation. This corresponds to a very simple analytic structure. One may ask whether there exists analogous properties in the case of non-compact groups. This problem is a formidable one if one wants complete results. However it is possible at least to make some steps in this direction.

To get some orientation let us consider a specific example. Let us take for instance the group SL(2R) of two by two real matrices of determinant one:

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad ad - bc = 1$$

Consider the following unitary representation. Its vector space is the set of analytic functions $f(z)$ analytic for $\mathrm{Im}\, z > 0$ and such that the norm:

---

† Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

‡ Present address: Institute of Advanced Study, Princeton, New Jersey.

* Permanent address: Service de Physique Theorique, CEN Saclay, BP n° 2, 91, Gif sur Yvette, France.

$$||f||^2 = \frac{1}{\pi}\int dx\ |f(x)|^2 \tag{1}$$

is finite. The group acts as:

$$[U(g)f](z) = \frac{1}{bz+d}\ f\left(\frac{az+c}{bz+d}\right) \quad . \tag{2}$$

It is easily verified that this representation is unitary. In fact it is one of the discrete series mentioned in Prof. Taylor's lectures. Now to compute the matrix elements of the representation we need some basis states. For this purpose it is customary to diagonalize the compact subgroup of rotations of the form

$$r = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \quad . \tag{3}$$

It is readily verified that a complete orthonormal basis is defined by the functions:

$$f_n(z) = \frac{(z-i)^n}{(z+i)^{n+1}} \quad , \quad n = 0,1,\ldots \quad . \tag{4}$$

These functions satisfy:

$$\text{(i)}\quad (f_m, f_n) = \delta_{m,n}$$

$$\text{(ii)}\quad [U(n)\ f_n](z) = \frac{1}{-\sin\varphi z + \cos\varphi}\ \frac{\left(\frac{\cos\varphi z + \sin\varphi}{-\sin\varphi z + \cos\varphi} - i\right)^n}{\left(\frac{\cos\varphi z + \sin\varphi}{-\sin\varphi z + \cos\varphi} + i\right)^{n+1}}$$

$$= e^{-i(2n+1)\varphi}\ \frac{(z-i)^n}{(z+1)^{n+1}} = e^{-i(2n+1)\varphi}\ f_n(z) \quad . \tag{5}$$

Let us now compute some simple matrix element of the non-compact subgroup of elements of the form:

$$\ell = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \tag{6}$$

where λ is a real positive number. For instance

$$(f_o, U(\ell) f_o) = \frac{1}{\pi}\int_{-\infty}^{+\infty} dx\ \frac{1}{x-i}\cdot\frac{\lambda}{\lambda^2 x + i} \quad .$$

The integral is readily obtained by closing the contour in the upper half plane with the result that:

$$(f_o, U(\ell)f_o) = \frac{2\lambda}{\lambda^2 + 1} = \frac{1}{ch\theta} \tag{7}$$

if we set $\lambda = e^{\theta}$.

This simple expression exhibits two interesting facts. The first one is that the given matrix element is meromorphic in the complex λ plane (which implies in particular that it is a periodic function of θ of period $2i\pi$). Furthermore $(f_o, U(\ell)f_o)$ has singularities at $\lambda = \pm i$. The crucial fact is that these singularities are located on the unit circle. What we are going to prove in the following is that this is indeed a very general property.

The subgroup of matrices ℓ of the form (6) is isomorphic to the simplest non-compact Lie group the multiplicative group of positive real numbers, which we shall denote by R_+. Its "natural complexification" is C_o the multiplicative group of complex numbers (i.e. the complex plane minus the origin). Notice that C_o has a compact real form, namely U(1) the multiplicative group of complex numbers with modulus one. Now R_+ is also isomorphic with the additive group of real numbers whose complexification--the additive group of complex numbers--is a simply connected Lie group which is in fact an infinite covering of C_o. (This, by the way, illustrates the property that in a general sense a complexification of a Lie group need not be unique.)

We shall concentrate our attention on the pair R_+, C_o as it arises in several applications. Suppose now that we are given a Hilbert space $\mathcal{H}$ and a unitary representation $U(R_+)$ in $\mathcal{H}$. This representation can be quite general. Special cases arise when R_+ is a subgroup of some larger group and $U(R_+)$ is the restriction to R_+ of some unitary irreducible representation of this larger group. Such was the situation in the previous example. Frequent applications also occur when dealing with the Lorentz group, when R_+ is isomorphic to the set of "boosts" along a certain direction. Now select some vector $|f)$ in the Hilbert space $\mathcal{H}$, and consider the diagonal matrix element:

$$(f|\,U(\lambda)\,|f) = K(\lambda) \tag{8}$$

with λ a real positive number, i.e. an element of R_+. We may now ask what analytic properties, if any, the functions $K(\lambda)$ may have if λ is regarded as a complex parameter (different from zero). We shall prove the following rather striking property.

Theorem: If $K(\lambda)$ is the restriction to the real positive axis of some function (denoted by the same symbol) analytic in a neighborhood of

the unit circle, then the vector $|f)$ is an invariant vector of the representation $U(R^+)$; in other words $U(\lambda)|f) = |f)$ for all positive λ.

We have already seen one example of this phenomenon above. It is clear that in general the domain of analyticity can be quite small. The theorem states that it is impossible for $K(\lambda)$ to be extended to an analytic function on the whole of C_o. That does not mean that $K(\lambda)$ cannot be analytic on a covering of C_o, i.e. a multisheeted function. For instance the irreducible representations of R_+ such that $U(\lambda)|f) = \lambda^{i\alpha}|f)$, with α real, are examples of this situation, for the only case then where $K(\lambda) = \lambda^{i\alpha}(f|f)$ is analytic in a neighborhood of the unit circle is when it is uniform; i.e. $\alpha = 0$ and $|f)$ is indeed invariant. So except for the trivial case of an invariant $|f)$ no domain of analyticity of $K(\lambda)$ imbedded in C_o can contain the whole unit circle.

Let us turn to the proof. The first fact to notice is that, it is equivalent to say that $U(\lambda)|f) = |f)$ or $(f|U(\lambda)|f) = (f|f)$. Indeed $||\,|f) - U(\lambda)|f)||^2 = 2(f|f) - 2\,\mathrm{Re}\,(f|U(\lambda)f)$ so that if $(f|U(\lambda)|f) = (f|f)$, a real quantity, the norm of $|f) - U(\lambda)|f)$ vanishes and $|f) = U(\lambda)|f)$. In the following we shall normalize $|f)$ by the condition $(f|f) = 1$, clearly an unimportant factor. We want to make best use of the unitarity of $U(\lambda)$. For instance one has as a consequence Schwartz's inequality:

$$|K(\lambda)|^2 = |(f|U(\lambda)|f)|^2 < 1 \tag{9}$$

due to the normalization $(f|f) = 1$. Actually (9) is a special case of a more general inequality which may be obtained as follows. Let $\lambda_k = e^{\theta_k}$, $k = 0,1,\ldots n$ be a sequence of real positive numbers (θ_k real) and g_k a corresponding sequence of arbitrary complex numbers. Then the vector $\sum_o^n g_k\, U(e^{\theta_k})|f)$, has a non-negative norm. Using the unitarity of $U(\lambda)$ we deduce that the following quadratic form is non-negative:

$$\sum_{p,q=0}^{n} \bar{g}_q \quad K(e^{\theta_p - \theta_q})\, g_p \geqslant 0. \tag{10}$$

Furthermore we have, again from unitarity,

$$K(\lambda) = (f|U(\lambda)f) = (U(\lambda^{-1})f|f) = \overline{(f|U(\lambda^{-1})f)} = \overline{K(\lambda^{-1})}$$

for λ real positive. Now assume that, as implied by the hypothesis of the theorem, $K(\lambda)$ can be analytically continued on a neighborhood of the unit circle. We can take this neighborhood small enough if we wish to make it invariant under the transformation $\lambda \to \overline{\lambda^{-1}}$. Then the

two analytic functions $K(\lambda)$ and $\overline{K(\lambda^{-1})}$, both analytic in the same neighborhood of the unit circle, coincide on its intersection with the real positive axis hence are identical:

$$K(\lambda) = \overline{K(\lambda^{-1})} \quad . \tag{11}$$

This says in particular that $K(\lambda)$ is real on the unit circle (recall that $K(1) = 1$). For simplicity of notation we shall denote:

$$F(\theta) = K(e^{\theta}) \quad , \tag{12}$$

and we can transcribe the conditions (10) and (11) in terms of the function $F(\theta)$. Since the sets $\{\theta_k\}$ and $\{g_k\}$ appearing in the inequality (10) are arbitrary we are allowed to make any convenient choice. So let us assume that:

$$\theta_k = 2k\theta \quad , \quad g_k = \frac{(-1)^k}{(2\theta)^{2n}} \binom{n}{k} \quad , \tag{13}$$

where θ is an arbitrary positive quantity and $\binom{n}{k}$ stands for the usual binomial coefficient $\binom{n}{k} = \frac{n!}{k!n-k!}$. Condition (10) now reads:

$$\frac{1}{(2\theta)^{2n}} \sum_{S=-n}^{+n} (-1)^s \, F(2s\theta) \sum_{\substack{p,q=0 \\ p-q=s}}^{n} \binom{n}{p} \binom{n}{q} \geqslant 0$$

From the identity $(1+x)^n(1+x)^n = (1+x)^{2n}$ we derive that for $-n \leq s \leq n$:

$$\sum_{\substack{p,q=0 \\ p-q=s}}^{n} \binom{n}{p} \binom{n}{q} = \binom{2n}{n+s} = \frac{2n!}{n+s\,!\,n-s\,!}$$

hence the above condition may be rewritten as:

$$\frac{1}{(2\theta)^{2n}} \sum_{s=-n}^{+n} (-1)^s \frac{2n!}{(n-s)!\,(n+s)!} \, F(2s\theta) \geqslant 0 \tag{14}$$

The left hand side of this expression is nothing but a $2n$-th finite difference. To see what this means let us recall that for an arbitrary function $F(x)$ the successive finite differences are defined through:

$$(\Delta_\theta F)(x) = \frac{F(x+\theta) - F(x-\theta)}{2\theta}$$

$$\Delta_\theta{}^m F = \Delta_\theta(\Delta^{m-1} F) \tag{15}$$

By induction, one easily derives the following formula:

$$(\Delta_\theta{}^m F)(x) = \frac{1}{(2\theta)^m} \sum_{\ell=0}^{m} (-1)^\ell \frac{m!}{\ell!\, m-\ell!} F(x+(m-2\ell)\theta) \quad , \tag{16}$$

which for $m = 2n$ and $x = 0$ gives:

$$(\Delta_\theta{}^{2n} F)(0) = \frac{(-1)^n}{(2\theta)^{2n}} \sum_{s=-n}^{+n} (-1)^s \frac{2n!}{(n-s)!\,(n+s)!} F(2s\theta) \quad .$$

Hence we have:

$$(-1)^n (\Delta_\theta{}^{2n} F)(0) \geqslant 0 \tag{17}$$

Now we can write $F(\theta) = F_+(\theta) + F_-(\theta)$ by breaking it into its even and odd parts

$$F_\pm(\theta) = \frac{F(\theta) \pm F(-\theta)}{2} \quad . \tag{18}$$

Translated in terms of F our hypothesis states that $F(\theta)$ is analytic in a neighborhood of the imaginary axis and periodic: $F(\theta+2i\pi) = F(\theta)$, so that we can make the neighborhood a strip (by reducing it eventually) of the form $|\mathrm{Re}\ \theta| \leq 2\alpha$ for instance. Furthermore $F(\theta)$ is real on the imaginary axis as implied by (11). Clearly the even and odd parts enjoy the same properties. Again condition (11) implies that on the real axis F_+ is real while F_- is purely imaginary. Now F_+ in the neighborhood of $\theta = 0$ is analytic hence infinitely differentiable. Furthermore the inequality (17) only involves the even part F_+. It is well known (and easily verified) that for an infinitely differentiable function the finite differences tend in the limit $\theta \to 0$ to the corresponding derivatives. Hence the even derivatives of $F_+(\theta)$ alternate in sign at $\theta = 0$ (the odd derivatives are obviously zero). That means that along the imaginary axis $\theta = i\psi$, $F_+(\theta) = F_+(i\psi)$ which is again a real (even) function of ψ has all its derivatives with respect to ψ non-negative at $\psi = 0$. But since the odd derivatives vanish anyhow all derivatives with respect to ψ of $F_+(i\psi)$ are non-negative at $\psi = 0$. Now considered as a function of ψ $F_+(i\psi)$ is analytic in $|\mathrm{Im}\ \psi| \leq 2\alpha$. We have at the origin a Taylor series for $F_+(i\psi)$ convergent in a

circle of radius 2α with all coefficients non-negative. As a result for $0 \leq \psi < 2\alpha$ $F_+(i\psi)$ is non-negative together with all its derivatives. The process can be repeated starting from this point $\psi = \alpha$ say until we reach in a finite number of steps the point $\theta = 2i\pi$, implying that all along the segment $\theta \leq \psi \leq 2\pi$ $F_+(i\psi)$, together with all its derivatives is non-negative. Now from the periodicity we have:

$$0 = F_+(2i\pi) - F_+(0) = \int_0^{2\pi} \frac{d}{d\psi} F(i\psi)\, d\psi \quad .$$

Since $\frac{d}{dx} F_+(i\psi)$ is non-negative (and continuous) this implies that:

$$\frac{d}{d\psi} F_+(i\psi) = 0 \quad , \quad 0 \leq \psi \leq 2\pi \quad ,$$

or that F_+ is constant on the imaginary axis and hence in all the neighborhood of that axis where it is assumed analytic. This constant value, by the way, is its value at $\theta = 0$ i.e. one. Hence on the intersection of the neighborhood with the real axis $F_+(\theta)$ is equal to one. Since $F_-(\theta)$ is purely imaginary along the real axis, one has for θ real and close to zero:

$$|F(\theta)|^2 = |F_+(\theta)|^2 + |F_-(\theta)|^2 = 1 + |F_-(\theta)|^2 \quad .$$

Schwartz's inequality on the other hand, Eq. (9), requires $|F(\theta)|^2$ not to be larger than one. The conclusion is then that $F_-(\theta)$ has to vanish for θ close to zero, or $F(\theta) = 1$ for $-2\alpha < \theta < 2\alpha$. Returning to the original notation we have $K(\lambda) = 1$ for λ close to 1. From our initial remark this implies that $U(\lambda)|f) = |f)$ for λ close to one. By the very definition of representations this cannot hold in the neighborhood of the identity unless it holds for all values of λ real positive. Indeed for any such λ there always exists an integer t and a λ_0 lying in the given neighborhood of the identity such that $\lambda = \lambda_0^t$ and $U(\lambda)|f) = U(\lambda_0)^t|f) = U(\lambda_0)^{t-1}U(\lambda_0)|f) = U(\lambda_0)^{t-1}|f)$ etc. until we get $U(\lambda)|f) = |f)$. The theorem is thus proved.

Finally let us make a few comments. The "positivity" property that we have used is in fact the starting point of a classical theorem in Fourier analysis due to Bochner which claims essentially that there exists a (positive) measure $d\mu(\alpha)$ such that:

$$K(\lambda) = \int_{-\infty}^{+\infty} \lambda^{i\alpha}\, d\mu(\alpha) \quad . \tag{19}$$

Conversely given a positive measure $d\mu(\alpha)$ we can define a unitary representation of R_+ on the Hilbert space of square integrable functions with respect to $d\mu(\alpha)$ through:

$$[U(\lambda)f](\alpha) = \lambda^{i\alpha} f(\alpha) \tag{20}$$

If $\int d\mu(\alpha) < \infty$ then $K(\lambda)$ as defined in (19) is essentially $(f_o | U(\lambda) | f_o)$ where f_o is a constant function. This construction can serve to build interesting examples. Taking for instance

$$d\mu(\alpha) = \frac{\operatorname{ch} b\alpha + \cos a\alpha}{\operatorname{ch} \pi \frac{\alpha}{2}} \, d\frac{\alpha}{2} \tag{21}$$

with $0 \leq b < \pi/2$, we have obviously a positive measure with $\int d\mu(\alpha) < \infty$.

Then $K(\theta)$ given by (19) turns out to be equal to:

$$\frac{2\operatorname{ch}\theta \cos b}{\operatorname{ch} 2\theta + \cos 2b} + \frac{2\operatorname{ch}\theta \operatorname{ch} a}{\operatorname{ch} 2\theta + \operatorname{ch} 2a}$$

again a meromorphic function in the complex plane of the variable $\lambda = e^{\theta}$ with singularities on (and off) the unit circle. Moreover the singularities on the unit circle has an arbitrary location depending on the parameter b.

In conclusion let us point that the study of analytic properties in the general case of a non-compact group seems to remain a largely unexplored field.

The author takes this opportunity to thank Prof. W. Panofsky for his hospitality at the Stanford Linear Accelerator Center where this work was completed.

INFINITE MOMENTUM FRAMES AND PARTICLE DYNAMICS†‡

Leonard Susskind
Belfer Graduate School of Science
Yeshiva University
New York, New York

I. Infinite Momentum Kinematics

A. Introduction

It has become fashionable among theorists in recent years to criticize the Schrödinger picture with its associated instantaneous state vector satisfying $i\frac{\partial|\psi\rangle}{\partial t} = H|\psi\rangle$. Chew[1] and many others have maintained that only the complete scattering event has operational meaning when long range forces are absent and have therefore advocated a pure S-Matrix Theory of Strong Interactions. Even among axiomatic field theorists, the attitude is sometimes expressed that the observables of relativisitic field theory are the S-matrix elements and that local interpolating fields are auxiliary quantities which are useful for insuring reasonable locality properties for the S-matrix.

One rarely encounters the notions of instantaneous state or of instantaneous measurements of operator observables as in the original Dirac-Von-Neumann[2] formulations of quantum mechanics.

Of course, in a formal sense a Schrödinger Picture necessarily exists if the S-matrix view is correct. A complete set of states $|\psi_{in}\rangle$ may be postulated with the property that superpositions such as

$$\int A(\psi)\, e^{iE_\psi t}\, |\psi\rangle$$

approach states

$$\int A(\psi)\, e^{iE_\psi t}\, |\psi_o\rangle$$

with $|\psi_o\rangle$ being a set of well separated, freely moving, incoming particles in the limit $t \to -\infty$. The quantities E_ψ are the energy of the state $|\psi_{in}\rangle$ which equals the sum of the energies of the asymptotic free particles.

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.
‡Supported in part by AF-AFOSR Grant No. 1282-67.

Likewise, a set of out-states $|\psi_{out}\rangle$ may be postulated which span the same space and the transformation matrix $\langle\psi_{in}|\psi'_{out}\rangle$ is the S-matrix. Any state may be expanded in the $|\psi_{in}\rangle$ or $|\psi_{out}\rangle$ and will evolve in time according to

$$e^{iHt}\int A(\psi)|\psi_{in}\rangle = \int A(\psi)\, e^{iE_\psi t}|\psi_{in}\rangle \quad .$$
$$\text{out} \qquad\qquad\qquad\qquad\qquad\qquad \text{out}$$

For example, the expectation value of any operator Q as a function of time may be computed if we know $\langle\psi_{in}|Q|\psi_{out}\rangle$.

$$\langle Q\rangle_t = \int e^{i(E_\psi - E_{\psi'})t}\langle\psi_{in}|Q|\psi'_{in}\rangle\, A(\psi)\, A^*(\psi')$$
$$= \int e^{i(E_\psi - E_{\psi'})t}\langle\psi_{in}|Q|\psi''_{out}\rangle\, S^\dagger_{\psi'';\psi'}\, A(\psi)\, A^*(\psi')$$

where $S^\dagger_{\psi'';\psi'} = \langle\psi'_{out}|\psi'_{in}\rangle$.

Therefore the question is not whether the Schrödinger picture exists in a mathematical sense, but whether operational meaning can be given to instantaneous values of the variables describing the Schrödinger picture. That is, can we do an experiment at a given time, certain to within an error smaller than the typical duration of interaction between hadrons?

Most theorists at present seem to be of the opinion that the answer is no. The reasoning goes as follows: In order to measure a variable associated with a system at a given time t_0, an apparatus or external system must interact with the system for an interval Δt centered about t_0 with Δt being small in comparison with the typical time scale for variations of the state of the system. For hadronic systems this time is $\sim 10^{-23}$ sec. However, one can scarcely expect to get an external probe into and out of a system in a time shorter than the time needed for light to traverse the system which again is a time $\sim 10^{-23}$ sec. Thus it does not seem possible to localize the time of measurement much more than to be able to say that it coincided with the scattering event. In this view the events of the world are quantized in discrete scattering events. Goldberger and Watson have even attempted to argue that the usual continuous time evolution of the Schrödinger equation is a long time average over many collision processes.[3]

The set of lectures I am going to give are reactionary in that I will propose a return to the good old "instant" picture. The way around the argument given above is very simple. The time scale of internal motions of strongly interacting systems cannot really be said to be 10^{-23} sec. It is in fact a function of the velocity of the center

of mass of the system as seen by the laboratory observer. For systems moving with $v \approx c$ the internal motions are so time dilated that the duration of collisions can be made as long as we like. By inserting a probe along the direction of motion of the system at controllable macroscopic distances along its motion we can control the time of probing to any desired accuracy relative to the time scale of collisions.[4] Thus we are led to ask how an observer describes a system of hadrons whose center of mass frame moves relative to him with a velocity approaching that of light.

B. Free Particles

We shall consider the description of any system from two reference frames.[5] The first, called the ordinary reference frame (O.R.F.) is a frame in which all momenta are finite. The coordinate axes for the O.R.F. are labelled x', y', z', t'.

The second frame called the limiting reference frame (L.R.F.) is a frame moving in the -z direction with respect to the O.R.F. with a velocity which will be allowed to approach the velocity of light. (We take $c = 1$.) The L.R.F. has coordinate axes x, y, z, t related to x', y', z', t' by

$$x = x'$$

$$y = y'$$

$$z = z' \cosh \omega + t' \sinh \omega$$

$$t = t' \cosh \omega + z' \sinh \omega \quad .$$

In the limit that the L.R.F. has velocity 1, $\cosh \omega = \sinh \omega = e^{\omega}/2$. Hence

$$x = x'$$

$$y = y'$$

$$z = (z' + t')e^{\omega}/2$$

$$t = (z' + t')e^{\omega}/2$$

The space-like surface $t = 0$ which is the "space" hypersurface for the L.R.F. degenerates to a surface tangent to the light cone from the point of view of O.R.F.

$$t = 0 \equiv z' + t' = 0$$

In the L.R.F. all momenta have been boosted up to very large values in the +z direction. For a particle with momentum k'_x, k'_y, k'_z and energy ω'_k in the O.R.F., the momentum in the L.R.F. is

$$\begin{aligned} k_x &= k'_x \\ k_y &= k'_y \\ k_z &= k'_z \cosh\omega + \omega'_k \sinh\omega \\ \omega_k &= \omega'_k \cosh\omega + k'_z \sinh\omega \end{aligned} \tag{I.1}$$

which tends to

$$\begin{aligned} k_x &= k'_x \\ k_y &= k'_y \\ k_z &= \omega_k = e^{\omega}(k'_z + \omega'_k) \end{aligned} \tag{I.2}$$

as the hyperbolic angle ω becomes infinite.

The finite quantity $k_z e^{-\omega}$ will be denoted by α and equals $k'_z + \omega'_k$. α will be called the longitudinal ratio.

In the L.R.F. where k_z becomes infinite we shall use α to describe the z component of motion. The transverse motion can be described by the usual transverse momentum $(k_x, k_y) = (k'_x, k'_y)$ which will be denoted by a 2-vector K. In general, a capitalized momentum means the transverse component of the 3-momentum labeled by the same lower case letter.

More generally we shall write $|(\alpha_1 K_1, \ldots, (\alpha_n K_n)\rangle$ for the state of an n particle system. For bra vectors we will use $\langle(\beta_1 L_1), \ldots, (\beta_m L_m)|$.

The physical significance of replacing k_z by α is interesting. It takes into account the fact that for all systems moving with $v \approx 1$ all dimensions along the z axis are Lorentz-contracted by a factor $e^{-\omega}$. Hence, in order to describe the structure in the z direction, the z axis must be "stretched" by a factor e^{ω}. The stretching of z must be accompanied by an opposite rescaling of the conjugate momentum variable. Hence α is the conjugate to the stretched z axis.

We next define the notion of a "kinematic subgroup of the Poincare group" for a given observer.

By definition it is the subgroup which leaves invariant the space like hypersurface of initial conditions for that observer. For the O.R.F. this hypersurface is $t' = 0$ and the kinematic subgroup consists

of spatial translations and rotations. The reason for picking out this subgroup is that in a dynamical theory such as field theory, the generators of the kinematic subgroup do not depend on the Hamiltonian.

Their action is independent of the form of the interaction between particles. The other generators which form the dynamical part of the group displace the surface of initial conditions and therefore, in order to determine the state in the transformed frame, the system must propagate from one space-like surface to another. Therefore, the dynamics influence the form of these generators. For the O.R.F. the dynamical generators are the energy and boost generators.

The space-like surface of the L.R.F. degnerates to a surface tangent to the light cone. The subgroup which carries such a surface into itself is larger than the corresponding subgroup for the O.R.F. It consists of seven generators instead of six. The kinematic subgroup for L.R.F. can be described either from the point of view of O.R.F. or L.R.F.

The first kinematic generator for L.R.F. is that of z-boosts which act on z' and t' as follows

$$z' \to z' \cosh \nu + t' \sinh \nu$$

$$t' \to t' \cosh \nu + z' \sinh \nu$$

Therefore $z' + t' \to (z' + t')e^{\nu}$ and the surface $z' + t' = 0$ is preserved.

Acting on the momentum variables, the z boost gives

$$k'_z \to k'_z \cosh \nu + \omega'_k \sinh \nu$$

$$\omega'_k \to \omega'_k \cosh \nu + k'_z \sinh \nu$$

$$k'_x \to k'_x$$

$$k'_y \to k'_y \qquad \text{(I.3)}$$

In the L.R.F. this is

$$k_z \to k_z \cosh \nu + \omega_k \sinh \nu$$

$$\omega_k \to \omega_k \cosh \nu + k_z \sinh \nu \qquad \text{(I.4)}$$

or in terms of α

$$\alpha \to e^{\nu}\alpha$$

$$K \to K \qquad \text{(I.5)}$$

more generally

$$|(\alpha_1 K_1), \ldots, (\alpha_n K_n)\rangle \to |(e^{\nu}\alpha_1 K_1) \ldots (e^{\nu}\alpha_n K_n)\rangle . \tag{I.6}$$

Thus z-boosts are scale transformations of the α. L.R.F. quantum-mechanics must be invariant under rescaling of the α's.

A second, third and fourth element of the kinematic subgroup must be those transformations which are merely spatial rotations in the L.R.F.

For rotations about x we have

$$\begin{aligned} k_x &\to k_x \\ k_y &\to k_y + \epsilon\, k_z \\ k_z &\to k_z - \epsilon\, k_y \end{aligned} \tag{I.7}$$

for infinitesimal angle ϵ.

In terms of α and K this is

$$\begin{aligned} K_x &\to K_x \\ K_y &\to K_y + \epsilon\, \alpha\, e^{\omega} \\ e^{\omega}\alpha &\to e^{\omega}\alpha - \epsilon\, K_y \end{aligned} \tag{I.8}$$

If we now choose ϵ such that in the limit $e^{\omega} \to \infty$ the product $\epsilon\, e^{\omega}$ is finite and equal to Q_y we have

$$\begin{aligned} K_x &\to K_x \\ K_y &\to K_y + \alpha\, Q_y \\ \alpha &\to \alpha \end{aligned} \tag{I.9}$$

Similarly a rotation about y gives

$$\begin{aligned} K_y &\to K_y \\ K_x &\to K_x + \alpha\, Q_x \\ \alpha &\to \alpha \end{aligned} \tag{I.10}$$

Generally

$$|(\alpha_1 K_1)\ldots(\alpha_n K_n)\rangle \to |(\alpha_1, K_1+\alpha_1 Q),\ldots(\alpha_n, K_n+\alpha_n Q)\rangle \qquad (I.11)$$

Rotations about z are generated in the obvious manner, α being invariant and K transforming as a 2-vector.

Translations along the x or y axes are obvious members of the kinematic subgroup and are generated by K_x and K_y. A translation by 2-vector R multiplies the state vector by $e^{iR\cdot\Sigma K_i}$ where ΣK_i is the total transverse momentum.

Likewise z-translations (in the rescaled z variable) are generated by α and multiply the state vector by $e^{a\sum_i \alpha_i}$ where a is the distance of translation (measured in dilated units).

The kinematic subgroup then contains the three translations, three rotations and the z boosts in the L.R.F.

Our next task is to determine how the state vector propagates from one value of t to another. The generator of time translations is the Hamiltonian which in the L.R.F. is given by

$$\sum_i \omega_i$$

where $\omega_i = \sqrt{k_i^2 + M_i^2}$, the sum being over all particles in the system. Since ω_i tends to $\alpha_i e^{\omega}$, the total energy tends to k_z, the total z momentum. In this limit the energy differences between states with the same value of α becomes lost unless we expand ω_i to next order in $e^{-\omega}$.

$$\sum_i \omega_i = \sum \sqrt{e^{2\omega}\alpha_i^2 + K_i^2 + M_i^2}$$

$$\sim e^{\omega}\sum_i \alpha_i + e^{-\omega}\sum_i (K_i^2 + M_i^2)/2\alpha_i + 0(e^{-2\omega}) \qquad (I.12)$$

The term $e^{\omega}\sum \alpha_i$ is the total z momentum which becomes infinite. It is conserved and for an eigenstate of the z momentum has no effect on the state amplitude other than to multiply it by a phase factor. We shall make it a rule to consider only eigenvectors of α_{total} in which case we may simply ignore the term $e^{\omega}\sum \alpha_i$ altogether since it is the same for all states which can couple together. We are left with the term

$$e^{-\omega}\sum_i \frac{K_i^2 + M_i^2}{2\alpha_i}$$

which goes to zero as $e^{\omega} \to \infty$. This is a manifestation of the Einstein time dilation. Since relative and transverse motions take place on a time scale dilated by the factor e^{ω} it is reasonable that the part of the Hamiltonian which generates such motions goes as $e^{-\omega}$. As with the z axis, we shall rescale the t axis defining a time "time" $\mathfrak{T} = e^{-\omega}t$. The conjugate variable will be called H and is given by

$$H = \sum_i \frac{K_i^2 + M_i^2}{2\alpha_i} \tag{I.13}$$

We can gain some insight into the meaning of H by considering its action in terms of the O.R.F.

From (I.13) and the identifications of α with $k'_z + \omega'_k$ and K as k'_x, k'_y we have

$$H = \sum_i \frac{k'^2_x + k'^2_y + M^2}{2(k'_z + \omega'_k)}$$

Now using $M^2 = (\omega'_k)^2 - (k')^2$ gives

$$H = \tfrac{1}{2} \sum_i \frac{(\omega'_k)^2 - k'^2_z}{k'_z + \omega'_k} = \tfrac{1}{2} \sum_i (\omega'_k - k'_z) \tag{I.14}$$

This is clearly half the sum of the t′ and z′ translation generators and therefore generates translations along the light-like direction

$$y = x = z - t = 0$$

Hence L.R.F. quantum mechanics may be thought of as quantum mechanics in which the state vector is defined for an observer whose space is tangent to the light cone. The "Hamiltonian" generates changes in the state from one surface to the next along light-like directions perpendicular to the tangent line.

One very striking feature of the kinematic structure is its resemblance to two-dimensional galilean quantum mechanics in which each particle has a 2-vector momentum q_x, q_y, a mass μ and an energy, $E = \frac{q^2}{2\mu} + B$ where B is the internal energy of the particle. The theory is to have invariance under the following transformations:

a) Velocity boosts, $q \to q + \mu v$
b) translations generated by q
c) rotations which rotate the 2-vector q.

d) gauge transformations generated by the sum of the masses.
e) time translations generated by $\frac{q^2}{2\mu} + B$.
Compare a - e with the following L.R.F. transformations
a') x,y rotations. $K \to K + \alpha Q$
b') x,y translations generated by K
c') rotations about z which rotate K
d') Longitudinal translations generated by α.
e') τ translations generated by $H = \frac{K^2}{2\alpha} + \frac{M^2}{2\alpha}$
The structure is identical to a - e if the identifications are made,

$$\alpha \leftrightarrow \mu$$

$$K \leftrightarrow q$$

$$Q \leftrightarrow v$$

$$\frac{m^2}{2\alpha} \leftrightarrow B \qquad \text{(I.15)}$$

The analogy runs much deeper than just the theory of free particles. It is due to an isomorphism between a subgroup of the Poincare group and the two dimensional galilean group.

C. Bound States and Wave Functions

Another concept which has come into disfavor is the Schrödinger wave function for deeply bound composites.[6] It has been widely claimed that the Schrödinger wave function description, even with relativistic kinematics is inadequate to describe the charge distribution of a bound state such as the deuteron when the binding energy becomes large.[7] The generally accepted picture[8] is that for $M < \sqrt{2}m$ (M is the deuteron mass, m the nucleon mass) the range of the charge distribution is given by

$$r^{-2} = 2/m^2 \qquad \text{(I.16)}$$

Eq. (I.16) it is claimed represents ultra-relativistic effects such as pair effects and the like.

As the binding lessens and M becomes greater than $\sqrt{2}m$ but less than 2m the effect of an anomalous threshold is felt and the range satisfies

$$r^{-2} = 4M^2 - \frac{M^4}{m^2} \qquad \text{(I.17)}$$

which begins to look like the wave function expression one might expect. In the limit $M \to 2m$ we recover the usual non-relativistic expression.

Now in all previous attempts to explain deeply bound states through wave functions, the bound state is studied from its rest frame where it is assumed our non-relativistic intuitions are most applicable. However, we have seen a remarkable analogy between non-relativistic quantum mechanics and the L.R.F. description of particle systems. This prompted me to return to the abandoned wave function description but not as applied to the bound state at rest but as it applies to the bound state with velocity near c. Our method will be to write the expression for the range of the charge distribution of a two dimensional bound state and directly apply the correspondences of Eq. (I.15). For a bound state of internal energy B_d composed out of two particles of mass μ_p and μ_n and internal energy B_p and B_n the charge range is given by

$$r^{-1} = \frac{2\sqrt{2\mu(-B_d + B_p + B_n)}}{\mu_n}\ (\mu_p + \mu_n) \qquad \text{(I.18)}$$

where μ is the reduced mass $\mu_p\mu_n/(\mu_p + \mu_n)$.

In the correspondence of N.R. physics with L.R.F. physics the analogue of μ is not m, the true mass, but rather α, the longitudinal ratio. Applying correspondences (I.15) gives

$$r^{-1} = 2\sqrt{2\,\frac{\alpha_n\alpha_p}{\alpha_p + \alpha_n}\left(\frac{m^2}{2\alpha_p} + \frac{m^2}{2\alpha_n} - \frac{M^2}{2(\alpha_p + \alpha_n)}\right)}\ \frac{\alpha_p + \alpha_n}{\alpha_n} \qquad \text{(I.19)}$$

Note that r^{-1} is invariant under a change of scale of the α's as required by invariance under longitudinal boosts. We shall simplify our calculation by scaling the α's so that $\alpha_p + \alpha_n = 1$. This gives

$$r^{-2} = \frac{4\alpha_n(1-\alpha_n)}{\alpha_n^2}\left(\frac{m^2}{\alpha_n} + \frac{m^2}{1-\alpha_n} - M^2\right) = \frac{4}{\alpha_n^2}\left[-M^2\alpha_n(1-\alpha_n) + m^2\right] \qquad \text{(I.20)}$$

The neutron in the deuteron does not have a definite value of α_n. In general we can expect a superposition of proton-neutron states with all values of α_p and α_n subject to $\alpha_p + \alpha_n = 1$.

Furthermore, the L.R.F. is a frame which is boosted to such a large velocity that all momenta are in the +z direction. Hence we have the further constraint that $\alpha_p > 0$ and $\alpha_n > 0$. This allows all α_n in the range $0 < \alpha_n < 1$.

The actual charge range corresponds to that value of α_n for which r is maximum. Hence let us find the minimum of r^{-2} from Eq. (I.20)

$$\frac{d}{d\alpha_n} r^{-2} = 4\left(-\frac{2m^2}{\alpha_n^3} + \frac{M^2}{\alpha_n^2}\right) = 0$$

or

$$\alpha_n = \frac{2m^2}{M^2} \qquad (I.21)$$

There are two possible cases. Either $\frac{2m^2}{M^2} > 1$ or $\frac{2m^2}{M^2} < 1$.

In the first case the extremum of r occurs at a value of α_n which would require $\alpha_p < 0$. The minimum of r^{-2} in the range $0 < \alpha_n < 1$ occurs at the end point $\alpha_n = 1$ in this case. The range is given by $r^{-2} = 4m^2$ for $M < \sqrt{2}m$. On the other hand if case two prevails, the minimum of r^{-2} is in the range 0 - 1 so that r^{-2} is given by combining Eqs. (I.20) and (I.21)

$$r^{-2} = 4M^2 - \frac{M^4}{m^2} \qquad \text{for } M > \sqrt{2}m \ .$$

Thus with a single wave function picture of the bound state we reproduce the deep binding normal threshold result of Eq. (I.16) and the loose binding anomalous threshold result of (I.17).

It appears that all the arguments which militate against the use of both the Schrödinger picture and the wave function are not relevant in the L.R.F.

In the next lecture we shall see that the arguments against another reactionary viewpoint, namely that of the Fock space representation of the canonical commutation relations of field theory, are also not relevant in the L.R.F.

II. Fock Space and Perturbation Theory at Infinite Momentum

A. Weinberg's Rules

Another old-fashioned idea which has gone out of style along with the Schrödinger Picture and wave function is the Fock-space description of the Schrödinger picture state vector. The reason concerns the existence of vacuum effects in any relativistic field theory.[9] For example, in quantum electrodynamics the interaction has matrix elements connecting the bare vacuum to a state with a photon and a pair. Hence it follows that the bare vacuum is not an eigenvector of the total Hamiltonian. This is serious because no state in Fock space other than the bare vacuum is both normalizable and carries exactly zero 3-momentum.[9] Therefore the true vacuum which has these properties cannot be in Fock space.

Let me at this point remind you of the technical meaning of Fock space. Fock space is the Hilbert space spanned by the countable

basis of vectors formed by operating a finite number of times with creation operators on the bare vacuum. States can be "out of Fock space" if they have too many bare particles. For example if we enumerate a basis of single particle states with the integers th n a state such as n particles in the nth mode (n goes from 1 to ∞) would be out of Fock space. In general the super-Fock space consisting of all vectors spanned by the uncountable basis of any number of particles in any mode will contain many mutually orthogonal countable Hilbert spaces and part of the problem of field theory is to determine which such subspace is the space in which physics actually takes place. Since it cannot be Fock space we can begin to make guesses as to its nature. The first such guess may be based on the observation that it must contain a vacuum vector with zero momentum which is normalizable, not the bare vacuum, and hence not in Fock space. We may conjecture that the actual space of eigenstates of finite energy are the states spanned by finite numbers of creation operators on this physical vacuum. As far as I know this does not violate any known theorems such as Haag's theorem. We shall show that if this is the case for some simple field theories that they may be reformulated in terms of the degrees of freedom of the L.R.F. particle variables and that in the reformulated version, the physical states are in Fock space.

We shall derive the form of L.R.F. field theory by means of taking the L.R.F. limit of perturbation theory. The resulting rules which were first obtained by S. Weinberg will be referred to as Weinberg's Rules.[10] Weinberg's rules will be compared with rules obtained from a Fock space quantum theory formulated directly in terms of L.R.F. particle degrees of freedom. The resulting equivalence is not dependent on the validity of perturbation theory although we shall not prove that here.

We begin with a restatement of the old-fashioned pre-Feynmanized perturbation theory.

The space of states is the space generated by repeated application of creation operators $\varphi^+(k)$ on the bare vacuum, the bare vacuum being the unique state annihilated by all $\varphi^-(k)$. The $\varphi^+(k)$ and $\varphi^-(k)$ are Hermitian conjugates satisfying

$$[\varphi^-(k), \varphi^+(k)] = \delta(k-k')\,\omega_k \tag{II.1}$$

with $\omega_k = +(k^2 + m^2)^{\frac{1}{2}}$.

The Hamiltonian is $H_o + V$ with H_o being

$$\int \frac{dk}{\omega_k}\, \varphi^+(k)\, \varphi^-(k)\omega_k = \int dk\, N(k)\, \omega_k \tag{II.2}$$

The interaction we shall take to be the simplest of all, namely $\lambda \int d^3x\, \varphi^3(x)$ which is expressed in terms of $\varphi^{\pm}(k)$ as

$$V = \lambda \int \frac{dk}{\omega_k} \frac{d\ell}{\omega_\ell} \frac{dm}{\omega_m} [\varphi^+(k)\,\varphi^+(\ell)\,\varphi^+(m) + 3\,\varphi^+(k)\,\varphi^+(\ell)\,\varphi^-(-m) + 3\,\varphi^+(k)\,\varphi^-(-\ell)\,\varphi^-(-m) + \varphi^-(k)\,\varphi^-(\ell)\,\varphi^-(m)]\,\delta(k+\ell+m) \qquad (II.2')$$

According to the Lippman-Schwinger formalism[11] the transition matrix element for an energy conserving process from A to B is given by

$$T(A;B) = \lim_{z \to \omega_A + i\varepsilon} \langle A | T(z) | B \rangle \qquad (II.3)$$

where ω_A is the sum of all particle energies in the state A or B and T(z) is given by

$$T(z) = V + \frac{1}{z - H} V \quad . \qquad (II.4)$$

If (II.4) is expanded in powers of λ we obtain the explicit series solution

$$T(z) = V + V\frac{1}{z - H_o} V + V \frac{1}{z - H_o} V \frac{1}{z - H_o} V + \ldots \qquad (II.5)$$

The "old fashioned rules" for T follow from Eq. (II.5) by inserting a complete set of Fock Space states between the V's. The particular form of V in Eq. (II.2) causes transitions of four types:

a) Any particle of momentum k can break up into a pair (ℓ, m) with amplitude $\lambda\delta(k-\ell-m)$.

b) The inverse process, $(\ell, m) \to k$ with the same amplitude.

c) Three particles (k, ℓ, m) can be emitted with amplitude $\lambda\delta(k+\ell+m)$.

d) Three particles, (k, ℓ, m) can be absorbed with amplitude $\lambda\delta(k+\ell+m)$.

The factors $\frac{1}{z-H_o}$ contribute factors like $\frac{1}{\omega - \sum_i \omega_i + i\varepsilon}$ where the sum is over all particles in the intermediate state in question.

One final rule in the evaluation of T is to include a phase space factor ω^{-1} for each internal particle integrated over.

The rules for the nth order contribution to T are summarized in the following 8 steps:

1. Draw all possible nth order Feynman graphs for the process under consideration. (For well studied reasons it is unnecessary to draw any graphs with disconnected bubbles leading from no particles to no particles. Such graphs when summed contribute an overall insignificant phase factor to the scattering matrix.)

2. Break each diagram into the n! possible time ordered graphs in which different time orderings of the n vertices are treated as distinct. This defines for each time ordered graph n - 1 intermediate states.

3. For each vertex include the factor $\lambda\delta(k_1-k_2-k_3)$ or $\lambda\delta(k_1+k_2+k_3)$ depending on whether it is a transition of type (a,b) or (c,d).

4. For each intermediate state include a factor $(\omega-\sum_i \omega_i+i\epsilon)^{-1}$.

5. For each internal line include the factor $\frac{dk}{\omega_k}$.

6. Integrate over all internal momenta.

7. Sum over time orderings.

8. Sum over graphs.

The contribution of a single time ordering of the vertices is not Lorentz invariant and will depend on the frame in which the calculation is made. We shall now consider the contribution of such time ordered "old fashioned" graphs when the calculations are made in the L.R.F.

According to the considerations of Sec. I the external lines of a process should be labeled by α's and K's instead of (k_x,k_y,k_z). The internal lines can also be labeled with values of α and K. For the external lines the α are all greater than zero. For the internal lines we cannot be sure that as the momentum of the entire process becomes large the sums over states (Integrals over momenta) are saturated by states in which all particles carry z-momentum greater than zero. Although we shall show this to be the case, the possibility of negative α's in internal lines will be allowed for the present.

The rules 1 - 8 are replaced by 1′ - 8′ which are derived by straightforward substitution of αe^{ω} for k_z and K for (k_x,k_y). Also the free energy of a particle is given by $\alpha e^{\omega}+\frac{K^2+m^2}{2\alpha e^{\omega}}$. Rules 1′ and 2′ are identical to 1 and 2.

3′. For each vertex include

$$\lambda\delta(\alpha_1-\alpha_2-\alpha_3)\,e^{-\omega}\,\delta(K_1-K_2-K_3)$$

or

$$\lambda\delta(\alpha_1+\alpha_2+\alpha_3)\,e^{-\omega}\,\delta(K_1+K_2+K_3)$$

the second choice applying to vacuum processes.

4′. For each intermediate state include a factor

$$\left[e^{\omega}\left(\sum_{\text{ext}}\alpha_i-\sum_{\text{int}}|\alpha_j|\right)+\frac{e^{-\omega}}{2}\left(\sum_{\text{ext}}\frac{K_i^2+m_i^2}{\alpha_i}-\sum_{\text{int}}\frac{K_j^2+m_j^2}{\alpha_j}+i\varepsilon\right)\right]^{-1}$$

If all internal α's are greater than zero then the sum over the α_j in the intermediate state gives

$$\sum_{\text{int}}|\alpha_j|=\sum_{\text{int}}\alpha_j$$

which by momentum conservation equals $\sum_{\text{ext}}\alpha_i$. Hence in that case the entire factor is order e^{ω}. If any α_j is less than zero $\sum_{\text{int}}|\alpha|\neq\sum_{\text{ext}}\alpha$ and the factor is order $e^{-\omega}$. Therefore as $\omega\to\infty$ those processes which have internal lines carrying negative α will go to zero in comparison with the processes with all $\alpha_j>0$. For $\lambda\varphi^3$ theory in the L.R.F. we may say that intermediate as well as initial and final values of α are positive.

5′. For each internal line include the factor $dK\frac{d\alpha}{\alpha}$.

Counting powers of e^{ω} we have $e^{\omega(N_1-N_2)}=e^{-\omega}$ where N_1 = number of intermediate states, N_2 = number of vertices. The $e^{-\omega}$ combines with the overall energy delta function $\delta(E_{in}-E_{out})$ to give

$$\delta(e^{\omega}E_{in}-e^{\omega}E_{out})=\delta\left[\sum_{in}\frac{K^2+m^2}{2\alpha}-\sum_{out}\frac{K^2+m^2}{2\alpha}\right]$$

which is the δ-function of effective L.R.F. energy.

Rules 6′, 7′ and 8′ are identical to 6, 7 and 8.

Note that because all internal α's must be greater than zero for a process to survive in the limit $e^{\omega}\to\infty$, those processes involving emission of three particles from the vacuum must vanish in the limit since momentum conservation demands that at least one particle have negative z-momentum.

As an exercise the reader can compute the second order contribution to two particle scattering from the various possible time orderings of the Feynman graphs in Fig. 1.

For example, the time ordered graph a_1 gives a contribution to the amplitude (αK), $(\alpha' K')\to(\beta L)$, $(\beta' L')$ which is given by

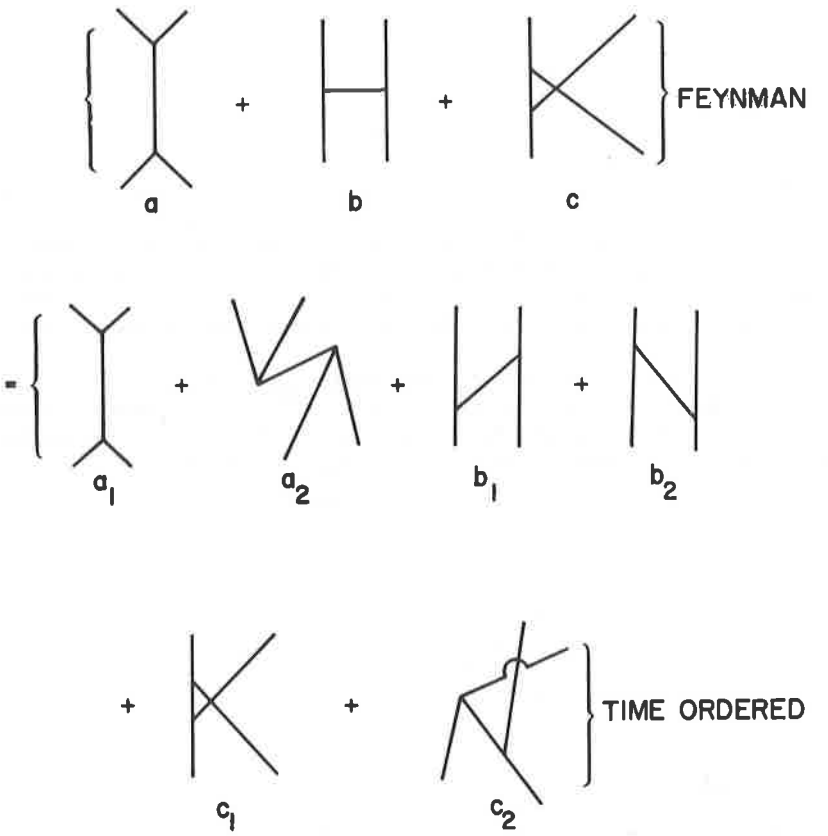


Fig. 1

$$\delta\left(\frac{K^2+m^2}{\alpha}+\frac{(K')^2+m^2}{\alpha'}-\frac{L^2+m^2}{\beta}-\frac{(L')^2+m^2}{\beta'}\right)$$
$$\times\frac{\lambda^2}{(\beta+\beta')}\left[\frac{m^2+(L')^2}{\beta'}+\frac{m^2+L^2}{\beta}-\frac{m^2+(L+L')^2}{\beta+\beta'}\right]^{-1}. \quad \text{(II.6)}$$

This can be expressed in terms of the Mandelstam invariants s, t, u.

$$s=(\omega_\ell+\omega_{\ell'})^2-(\ell_z+\ell_z')^2-(L+L')^2$$
$$=\left[(\beta+\beta')e^{\omega}+\frac{L^2+m^2}{2\beta e^{\omega}}+\frac{(L')^2+m^2}{2\beta' e^{\omega}}\right]^2-(\beta+\beta')^2e^{2\omega}-(L+L')^2$$
$$=\left(\frac{L^2+m^2}{\beta}+\frac{(L')^2+m^2}{\beta'}\right)(\beta+\beta')-(L+L')^2 \quad \text{(II.7)}$$

Combining (II.6) and (II.7) gives the usual covariant expression

$$\frac{\lambda^2}{s-m^2}$$

for the process. Notice that in an ordinary frame the graph a_1 and the vacuum process a_2 would combine to give $\frac{\lambda^2}{s-m^2}$ whereas in the L.R.F. the process a_1 alone gives the covariant result.

B. L.R. F. Formulation

The main conclusion about rules 1′ through 8′ is that they permit a formulation of L.R.F. quantum mechanics directly in terms of the L.R.F. variables without reference to any limiting procedure.

Define the vacuum vector $|0\rangle$ such that $\langle 0|0\rangle = 1$. The rest of the space of states is generated by Fock space creation operators $\varphi^+(\alpha K)$ which create particles in the L.R.F. with longitudinal ratio α and transverse momentum K. Let φ^+ and φ^- satisfy canonical commutation relations

$$[\varphi^+(\alpha K)\varphi^-(\beta L)] = \delta(K-L)\,\delta\left(1-\frac{\alpha}{\beta}\right) \quad \text{(II.8)}$$

We also introduce a generator H which propagates the state from one surface to the next in the manner discussed in Sec. I.

$$H = H_o + V \quad \text{(II.9)}$$

$$H_o = \int dK\,\frac{d\alpha}{\alpha}\,\varphi^+(\alpha K)\,\varphi^-(\alpha K)\,\frac{K^2+m^2}{2\alpha} \quad \text{(II.10)}$$

$$V = \lambda \int dK dL dM \frac{d\alpha}{\alpha} \frac{d\beta}{\beta} \frac{d\gamma}{\gamma} [3\varphi^{+}(\alpha K)\varphi^{-}(\beta L)\varphi^{-}(\gamma M) + 3\varphi^{+}(\beta L)\varphi^{+}(\gamma M)\varphi^{-}(\alpha K)] \delta(\alpha-\beta-\gamma)\delta(L+M-K) \qquad (II.11)$$

Because Weinberg's rule forbid vacuum processes there is no term in V with products of 3 creation or annihilation operators. Because these terms are absent, the bare vacuum $|0\rangle$ is an eigenvector of H and is the physical vacuum. Hence the Hamiltonian of (II.9) will not cause the eigenvectors to be out of Fock space because of vacuum effects. The theory is free of the only proven objection to Fock Space representations of canonical commutation relations.

Actually the lack of vacuum structure is another manifestation of the "Galilean" structure of L.R.F. physics.

The Hamiltonian in Eq. (II.9) gives rise to a perturbation series which is term by term identical to Weinberg's rules. We leave it as an exercise to prove this.

C. Fermions

Most field theories require modifications of the above procedures in order to obtain a L.R.F. Fock space theory which is equivalent. We shall demonstrate the necessary modifications for a theory with fermions and bosons coupled through the interaction Lagrangian $\bar{\psi}\psi\varphi$.

In place of rules 1 - 8 the correct old fashioned rules are 1″ - 8″. For a detailed exposition of these rules we refer the reader to Schweber's book.[12]

1″. Draw all Feynman graphs and consider each fermion line as a directed line flowing forward for particles and backward for antiparticles.

2″. Divide the diagram into the n! possible time ordered processes as in rule 2. Particle lines flow from past to future and antiparticle lines from future to past.

3″. For each vertex include a delta of incominb less outgoing momentum and a factor of the fermion boson coupling constant.

4″. For the initial and final states include the Dirac spinors u and $\bar{u}$ for particles and u^{c} and $\bar{u}^{c}$ for antiparticles.

For the internal states include $\frac{1}{\omega - \sum_i \omega_i + i\varepsilon}$ as in rule 4. The sum is over fermion and boson lines.

For each internal boson line include the factor $\frac{dk}{\omega}$. For each internal fermion line of momentum $\vec{p}$ include $\frac{d^3p}{\omega_p}(\omega_p\gamma_o - \vec{p}\cdot\gamma + m)$. For each internal antifermion line of momentum $\vec{q}$ include $\frac{dq}{\omega_q}(\omega_q\gamma_o - q\cdot\gamma - m)$.

6″. Integrate over internal momenta and sum over internal spins as would be done for the Feynman graph.

7″ and 8″ are the same as 7 and 8.

Let us apply these rules to the Feynman graph in Fig. 2. The figure shows the two possible time orderings of the S-channel fermion pole graph.

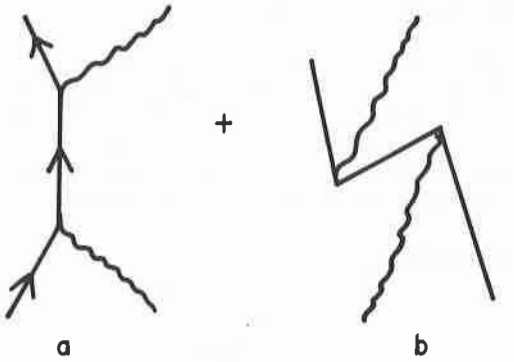


Fig. 2

The first ordering labeled (a) is similar to the corresponding boson scattering graph calculated in Sec. II.A. If the theory involving fermions worked just like the $\lambda\varphi^3$ model we might expect this first time ordering to give the entire covariant graph and the second time ordering to vanish as $e^{-2\omega}$ in the infinite momentum limit. We let μ = mass of boson, am = fermion mass, $\omega_k = (k^2 + \mu^2)^{\frac{1}{2}}$, $\nu_p = (p^2 + m^2)^{\frac{1}{2}}$ and $\vec{p},\vec{k},\vec{p}',\vec{k}'$ are the initial fermion, initial boson, finial fermion and final boson momenta. Graph a then gives

$$\bar{u}_p \frac{[\gamma_o \nu_{p+k} - \gamma\cdot(p+k) + m]}{(\nu_p + \omega_k - \nu_{p+k})\,\nu_{p+k}} \; u_p' \qquad \text{(II.12)}$$

Time ordering (b) gives

$$\bar{u}_p \frac{[\gamma_o \nu_{p+k} + \gamma\cdot(p+k) + m]}{(\nu_p + \omega_k + \nu_{p+k})\,\nu_{p+k}} \; u_p' \qquad \text{(II.13)}$$

Adding these contributions gives

$$\frac{\gamma_o(\nu_p + \omega_k) - \gamma\cdot(p+k) + m}{(\nu_p + \omega_k)^2 - \nu_{p+k}^2} \qquad \text{(II.14)}$$

which is identical to the usual Feynman graph.

The expression in Eq. (II.13) refers to a vacuum process in which the internal fermion carries negative α in the limit that the external lines have positive infinite momentum. Such processes always vanished in $\lambda\varphi^3$ theory. Let us see if they vanish in the present theory.

The term in (II.13) is expressed as

$$-\frac{\bar{u}_p[(\nu_p+\omega_k)\gamma_o - \gamma\cdot(p+k) + m]u_p'}{(\nu_p+\omega_k+\nu_{p+k})\nu_{p+k}} + \frac{\bar{u}_p\gamma_o(\nu_{p+k}+\nu_p+\omega_k)u_p'}{(\nu_p+\omega_k+\nu_{p+k})\nu_{p+k}} \qquad \text{(II.15)}$$

The first contribution goes to zero as $e^{-2\omega}$ since the numerator is a scalar and therefore is order 0 in e^{ω}. The denominator goes as $e^{2\omega}$.

The second term

$$\frac{\bar{u}_p\gamma_o u_p'}{\nu_{p+k}}$$

survives because both numerator and denominator go as e^{ω}. The numerator goes as e^{ω} because the spinors u_p and u_p' are normalized to $\bar{u}u = 1$.

Thus the vacuum effects and negative α effects do not vanish in this theory. One might have serious doubts as to whether a self-contained L.R.F. formulation of fermion theories exists in view of this fact. In particular the presence of vacuum terms would seem to leave us just where we started with respect to the question of Fock space representations of the canonical commutation relations. The situation is no better for either pseudo-scalar or pseudo-vector coupling.

The remarkable fact that will occupy our attention for the remainder of this lecture is that the entire effect of all graphs in which negative α fermions appear can be absorbed into a very simple term in the L.R.F. effective Hamiltonian in such a way as to entirely replace the graphs with negative α by a new kind of 4-particle vertex involving only positive α lines.

The persistence of negative α graphs in the infinite momentum limit is a special case of a wide class of phenomena in which in some limit a subspace of states should become unimportant but which still manifests itself in the scattering amplitude. For example, the exchange of a very heavy meson can contribute to low energy scattering although the probability for producing such a meson, even virtually, vanishes. Similarly in the non-relativistic limit all energies are assumed much less than mc^2. Thus the energy needed to produce a third particle from two incident particles becomes much greater than the

available energy. It should be expected that the portion of the space of states with more than 2 particles is simply unimportant when the energies studied are non-relativistic. Nevertheless, the process in which a virtual third particle is exchanged is not unimportant and accounts for the non-relativistic potential.

In both cases the possible excitation of some energetically distant mode is replaced by a contribution to the Hamiltonian. For example it is not experimentally possible at present to distinguish the difference between a four point interaction and a heavy intermediate vector meson in the weak interaction.

Non-relativistically we can replace an exchange graph by an addition to the potential.

I maintain that the processes involving negative α fermions are of this class of energetically distant processes which, if they contribute at all, do so in a manner indistinguishable from an additional term in the Hamiltonian. Therefore let me digress and derive the conditions for this to be so.

Divide the space of states into two orthogonal subspaces labeled P and Q. P labels that part of the space of states which is expected to survive when the limit in question is taken. In our present case P is all states with no negative z-momentum fermions. Q is the subspace in which one or more fermions has negative z-momentum.

We assume the Hamiltonian is decomposable into a free part, H_o, and interaction part, V, such that the Lippmann-Schwinger formalism for the T-matrix is satisfied.

$$T(z) = V + V\frac{1}{z - H_o}T(z) = V + V\frac{1}{z-H_o}V + V\frac{1}{z-H_o}V\frac{1}{z-H_o}V + \ldots \tag{II.16}$$

Let us define $T_Q(z)$ as the sum of all contributions to (II.16) with every intermediate state being a Q state. If $\hat{Q}$ is the projection operator for Q then

$$T_Q(z) = V\frac{\hat{Q}}{z-H_o}V + V\frac{\hat{Q}}{z-H_o}V\frac{\hat{Q}}{z-H_o}V + \ldots$$

$$= V\frac{\hat{Q}}{z-H_o}V + V\frac{\hat{Q}}{z-H_o}T_Q(z) \tag{II.17}$$

We shall only be interested in matrix elements of T_Q between P states. If we know these then we can eliminate the Q type states altogether. If we define $\tilde{V}(z) = V + T_Q(z)$ then

$$T(z) = \tilde{V}(z) + \tilde{V}(z)\frac{\hat{P}}{z-H_o}\tilde{V}(z) + \\ + \tilde{V}(z)\frac{\hat{P}}{z-H_o}\tilde{V}(z)\frac{\hat{P}}{z-H_o}\tilde{V}(z) + \ldots \qquad \text{(II.18)}$$

Eq. (II.18) can be verified by substituting Eq. (II.17) and observing that each term of (II.16) appears exactly once. Eq. (II.18) can be summed to give

$$T(z) = \tilde{V}(z) + \tilde{V}(z)\frac{\hat{P}}{z-H_o}T(z) \qquad \text{(II.19)}$$

or simply

$$T(z) = \tilde{V}(z) + \tilde{V}(z)\frac{1}{z-H_o}T(z) \qquad \text{(II.20)}$$

where T and $\tilde{V}$ are matrices in the P subspace.

The condition that the entire effect of Q be replaced by an additional term in the Hamiltonian in the P subspace is that $\tilde{V}(z)$ and therefore $T_Q(z)$ be independent of z.

In that case the equation is identical to the L.S. equation with an interaction $\tilde{V}$.

When I say that T_Q is independent of z I mean that it be independent of z when z is allowed to run over the values of energy typical to P states. Naturally we do not expect T to be independent of z when z approaches values typical of Q state energies. We must keep this in mind when applying these ideas.

Now let us rreturn to the graph studied in Eq. (II.12) but considered as a contribution to $T_Q(z)$. In order to study the z dependence in a manner appropriate to the L.R.F. we note that all L.R.F. energies have the form

$$E = e^{\omega}\sum\alpha + e^{-\omega}H$$

Hence the range of L.R.F. energies is reasonably parametrized by writing z as $e^{\omega}\Sigma\alpha + e^{-\omega}\zeta$. As ζ varies all L.R.F. energies are covered by finite ζ. If z itself were used all L.R.F. energies would be covered by a variation of z of order $e^{-\omega}$.

The graph gives a contribution to $V\frac{\hat{Q}}{z-H_o}V$ of the form

$$\frac{\bar{\mu}_p[\gamma_o\,\nu_{p+k} + \gamma\cdot(p+k) - m]\,\mu_p'}{(z - \nu_p - \omega_k - \nu_{p+k} - \nu_p' - \omega_k')\nu_{p+k}} \qquad \text{(II.21)}$$

which can be divided into two terms as in (II.15). The surviving contribution as $\omega \to \infty$ is

$$\frac{\bar{\mu}_p \gamma_o (\nu_{p+k} + \nu_p + \omega_k) \mu_p'}{(z - \nu_p - \omega_k - \nu_{p+k} - \nu_{p'} - \omega_k')\nu_{p+k}}$$

$$\frac{\bar{\mu}_p \gamma_o [2\alpha_p + 2\alpha_k] \mu_p' e^{\omega}}{(e^{\omega}(\alpha_p + \alpha_k) + \frac{z}{e^{\omega}} - 3e^{\omega}\alpha_p - 3e^{\omega}\alpha_k) e^{\omega}(\alpha_p + \alpha_k)}$$

$$= \frac{\bar{\mu}_p \gamma_o \mu_p'}{(\alpha_p + \alpha_k)} e^{-\omega} \qquad \text{(II.22)}$$

Thus, as advertised, this contribution to T_Q is z independent and its matrix elements can be absorbed into H.

A very similar contribution arises from the "u-channel" fermion exchange graph if the exchanged fermion carries $\alpha < 0$. I have investigated a great many other possible contributions to T_Q and have so far not found any that survive the limit $e^{\omega} \to \infty$. It is therefore my conjecture that the only necessary term in H needed to counter negative α fermions is a 4-pronged interaction whose matrix elements are given in (II.22) and the corresponding "u-channel" term.

Hence, although the contributions of vacuum graphs do not vanish in the L.R.F., the Q type states do disappear from the space of finite energy states. Their effect is absorbed into a simple contribution to H which is itself not a vacuum term since it has matrix elements only between P states. The bare vacuum is still the physical vacuum and Haag's theorem is not a difficulty.

The arguments given above are of course strictly heuristic. In order to demonstrate the existance of a self-contained L.R.F. formulation we must show that a sensible approach to the removal of divergences exists. This problem is presently under investigation by D. Flory.

The Fock space approach may not be left intact when ultra-violet divergences are considered. The state can leave Fock space not only because of the large available volume of space in which to produce vacuum particles but also because too many bare particles may be produced with high energy. These divergences may be removable by substituting the complete physical particle for the bare particle. Hopefully all divergences can be absorbed into the structure of a physical particle and there are no divergences in the transition from certain physical particle vectors to actual scattering states. This is the subject of the next section.

III. Physical Particle Representation

A. Physical Particle Basis

Because the transition from bare to physical particles may be ultra-violet divergent and may therefore lead the state vector out of Fock space, I shall rewrite the equations of L.R.F. field theory in terms of a physical particle basis, abstracted from Fock space. The basis states are not yet energy eigenstates but contain what is expected to be the divergent effects of interaction, namely the construction of the virtual "clouds" of each particle. We may hope (and in fact, perturbation theory bears out this hope in several examples) that the final effects of the interaction in causing real scattering among these physical particle states is small (in the sense of T. Kato) and convergent.

The ideas will be presented in terms of a set of scalar particles described by a set of fields φ_μ with interaction $\lambda_{\mu\nu\sigma}\varphi_\mu\varphi_\nu\varphi_\sigma$. Unless explicitly needed the index μ will be suppressed and we shall write $V = \lambda\varphi^3$. The masses m_μ need not be equal.

We consider the L.R.F. formulation given in Sec. II. The following notations are convenient:

1) An arbitrary set of bare particles $(\alpha_1 K_1, \ldots, \alpha_n K_n)$ will be labeled by an upper case letter from the early part of the alphabet.

$$A \equiv (\alpha_1 K_1, \ldots, \alpha_n K_n)$$

Sums over A will mean summations over all bare particle configurations. In particular A can stand for (0), the bare vacuum. Thus

$$\sum_A |A\rangle\langle A| = I \qquad \text{(III.1)}$$

2) The operator A^+ is the product $\varphi^+(\alpha_1 K_1)\ldots\varphi^+(\alpha_n K_n)$. If A represents (0), the no particle state, then A^+ is the unit operator. A^- is by definition the Hermitian conjugate of A^+ and equals $\varphi^-(\alpha_1 K_1)\ldots\varphi^-(\alpha_n K_n)$.

3) The vectors $|ABC\ldots\rangle$ are defined by

$$|ABC\ldots\rangle = A^+B^+C^+\ldots|0\rangle \ . \qquad \text{(III.2)}$$

They are independent of the order of A,B,C,... and contain all the bare particles of A,B,C... juxtaposed.

If A is the null set (0) then

$$|A,B,C,\ldots\rangle = |B,C,\ldots\rangle \qquad \text{(III.3)}$$

4) An arbitrary polynomial of creation operators is denoted by $\mathcal{P}^+$, $\mathcal{D}^+$, $\mathcal{R}^+$, or $\mathcal{S}^+$.

$$\mathcal{P}^+ \equiv \sum_A f(A)\, A^+ \tag{III.4}$$

$$\mathcal{P}^- \equiv \sum_A f(A)^*\, A^- \tag{III.5}$$

Also

$$|\mathcal{P},\mathcal{D},\mathcal{R}\ldots\rangle = \mathcal{P}^+\mathcal{D}^+\mathcal{R}^+\ldots|0\rangle \tag{III.6}$$

5) That unique polynomial of creation operators which creates one single <u>physical</u> particle of type μ and momentum α, K is given by $a^+_{\alpha K\mu}$

$$a^+_{\alpha K\mu} = \sum_{A=1}^{\infty} \psi_{\alpha K\mu}(A)\, A^+ \tag{III.7}$$

where $\sum_{A=1}^{\infty}$ means a sum over bare particles excluding the (0) configuration.

The notation $|\alpha_1 K_1 \mu_1, \ldots \alpha_n K_n \mu_n\rangle$ shall be reserved for physical particle states defined by $a^+_{\alpha_1 K_1 \mu_1}\ldots a^+_{\alpha_n K_n \mu_n}|0\rangle$. More often than not we shall suppress the index μ and write

$$|\alpha_1 K_1, \ldots, \alpha_n K_n\rangle = a^+_{\alpha_1 K_1}\ldots a^+_{\alpha_n K_n}|0\rangle\ .$$

The Hermitian conjugate of $a^+(\alpha K)$ is $a^-(\alpha K)$ but a^+ and a^- unlike φ^+ and φ^- do not satisfy canonical commutation relations.

$$[a^+(\alpha K),\, a^-(\alpha' K')] \neq \delta(K-K')\,\delta(1-\tfrac{\alpha}{\alpha'})$$

The reason that $|\alpha_1 K_1, \ldots \alpha_n K_n\rangle$ have not been used in field theory is because they are not orthonormal although they may span the space of states. In order to use the $|\alpha_1 K_1, \ldots, \alpha_n K_n\rangle$ in any sort of useful manner, the inner products $\langle \beta_1 L_1, \ldots, \beta_m L_m|\alpha_1 K_1, \ldots, \alpha_n K_n\rangle$ must be known as well as the matrix elements $\langle \beta_1 L_1, \ldots \beta_m L_m|H|\alpha_1 K_1, \ldots, \alpha_n K_n\rangle$.

A useful representation for $\langle \beta_1 L_1, \ldots | \alpha_1 K_1, \ldots \rangle$ can be obtained with the aid of Lemma 1.

Lemma 1. For all A, B, C, D

$$\langle AB | CD \rangle = \sum_{E_1, E_2, E_3, E_4} \langle A | E_1 E_2 \rangle \langle B | E_3 E_4 \rangle \times \langle E_1 E_3 | C \rangle \langle E_2 E_4 | D \rangle \qquad \text{(III.8)}$$

The sums over E_1, E_2, E_3, E_4 are over all bare particle configurations for each E. Included in the E is the vacuum set. A special case is when B = D = (0). Then E_3 and E_4 must both be (0) for $\langle B | E_3 E_4 \rangle$ to be non-zero and E_2 must be (0) for $\langle E_2 E_4 | D \rangle$ to be non-zero. Hence in this case we have

$$\langle A,0 | C,0 \rangle = \sum_1 \langle A | E_1 0 \rangle \langle 0 | 00 \rangle \langle 00 | 0 \rangle \langle E_1 0 | C \rangle$$

$$= \langle A | E_1 \rangle \langle E_1 | C \rangle$$

By definition $\langle A,0| = \langle A|$ and $|C0\rangle = |C\rangle$ so that (III.8) expresses the trivial identity in (III.1).

For B = (0) but $D \neq (0)$ we prove (III.8) as follows.

$$\langle A0 | CD \rangle \equiv \langle A | CD \rangle$$

$$= \langle A | D^+ | C \rangle$$

$$= \sum_{E_1} \langle A | D^+ | E_1 \rangle \langle E_1 | C \rangle$$

$$= \sum_{E_1} \langle A | E_1{}^+ | D \rangle \langle E_1 | C \rangle$$

$$= \sum_{E_1 E_2} \langle A | E_1{}^+ | E_2 \rangle \langle E_2 | D \rangle \langle E_1 | C \rangle$$

$$= \sum_{E_1 E_2} \langle A | E_1 E_2 \rangle \langle E_1 | C \rangle \langle E_2 | D \rangle$$

The last line is equivalent to (III.8) when E_2 and E_3 are (0) which is the only non-zero term in the E_2, E_3 sum if B = (0).

In general when A, B, C and D are not (0) Eq. (III.8) expresses the fact that in order to compute $\langle AB|CD\rangle$ we must link up all possible pairs of bare particles from bra and ket with delta functions and form a symmetrized product. Those links from A to C are included in E_1, those from A to D in E_2 and so on.

Eq. (III.8) can be used to prove a generalization for inner products such as $\langle A_1A_2\ldots A_m|C_1\ldots C_n\rangle$.

$$\langle A_1A_2\ldots A_m|C_1\ldots C_n\rangle$$

$$= \sum_{E_{ij}} \langle A_1|E_{11}E_{12}\ldots E_{in}\rangle\langle A_2|E_{21}\ldots E_{2n}\rangle\ldots\langle A_m|E_{m1}\ldots E_{mn}\rangle$$

$$\times \langle E_{11}E_{21}\ldots E_{m1}|C_1\rangle\ldots\langle E_{1n}E_{2n}\ldots E_{mn}|C_n\rangle \tag{III.9}$$

which is abbreviated by

$$\langle A_n\ldots|C_m\ldots\rangle = \sum_E \langle A_m|E\ldots\rangle\otimes\langle E\ldots|C_n\rangle \ . \tag{III.10}$$

Consider for example $\langle A_1A_2A_3|C_1C_2\rangle$. If we define $A_1^+A_2^+ = A^+$ then

$$\langle A_1A_2A_3|C_1C_2\rangle = \langle AA_3|C_1C_2\rangle \tag{III.11}$$

which by Eq. (III.8) is

$$\sum_{F_1F_2E_5E_6} \langle A|F_1F_2\rangle\langle A_3|E_5E_6\rangle\langle F_1E_5|C_1\rangle\langle F_2E_6|C_2\rangle$$

$$= \sum_{FE} \langle A_1A_2|F_1F_2\rangle\langle A_3|E_5E_6\rangle\langle F_1E_5|C_1\rangle\langle F_2E_6|C_2\rangle \tag{III.12}$$

We now use (III.8) to evaluate $\langle A_1A_2|F_1F_2\rangle$

$$\langle A_1A_2|F_1F_2\rangle = \sum_E \langle A_1|E_1E_2\rangle\langle A_2|E_3E_4\rangle\langle E_1E_3|F_1\rangle\langle E_2E_4|F_2\rangle \tag{III.13}$$

The quantities $\sum_{F_1} \langle E_1 E_3 | F_1 \rangle \langle E_5 F_1 | C_1 \rangle$ and $\sum_{F_2} \langle E_2 E_4 | F_2 \rangle \langle F_2 E_6 | C_2 \rangle$ which appear when Eq. (III.13) is plugged into Eq. (III.12) are equal to $\langle E_1 E_3 E_5 | C_1 \rangle$ and $\langle E_2 E_4 E_6 | C_2 \rangle$ respectively. We prove this as follows:

$$\langle E_1 E_3 | F_1 \rangle \langle E_5 F_1 | C_1 \rangle = \langle E_1 E_3 | F_1 \rangle \langle F_1 | E_5^- | C_1 \rangle$$

$$= \langle E_1 E_3 | E_5^- | C_1 \rangle = \langle E_1 E_3 E_5 | C_1 \rangle \quad .$$

The result when the sum on F is made is

$$\sum_E \langle A_1 | E_1 E_2 \rangle \langle A_2 | E_3 E_4 \rangle \langle A_3 | E_5 E_6 \rangle$$

$$\times \langle E_1 E_3 E_5 | C_1 \rangle \langle E_2 E_4 E_6 | C_2 \rangle$$

$$= \langle A_3 | E \ldots \rangle \otimes \langle E \ldots | C_2 \rangle \quad . \qquad \text{QED}$$

The process may be repeated to prove (III.9) for all n and m.

Since Eq. (III.9) is linear in each A_i and C_j it may be applied to polynomials $\wp$.

$$\langle \wp_1 \ldots \wp_m | \mathfrak{D}_1 \ldots \mathfrak{D}_n \rangle$$

$$= \sum \langle \wp_1 | E_{11} E_{12} \ldots E_{1n} \rangle \ldots \langle \wp_m | E_{m1} \ldots E_{mn} \rangle$$

$$\langle E_{11} E_{21} \ldots E_{m1} | \mathfrak{D}_1 \rangle \ldots \langle E_{1n} \ldots E_{mn} | \mathfrak{D}_n \rangle$$

$$= \langle \wp_m | E \ldots \rangle \otimes \langle E \ldots | \mathfrak{D}_n \rangle \qquad \text{(III.14)}$$

Substituting the $a^+(K)$ for the $\wp^+$ we get

$$\langle \beta_1 L_1, \ldots, \beta_m L_m | \alpha_1 K_1, \ldots, \alpha_n K_n \rangle = \langle \beta L_m | E \ldots \rangle \otimes \langle E \ldots | \alpha K_n \rangle \qquad \text{(III.15)}$$

The final step in the elimination is the replacement of the states $|E\ldots\rangle$ in (III.15) by some set of states which are not as far removed from the eigenvectors of energy momentum as are the bare states. We could use the $|\alpha K \ldots\rangle$ themselves but because of their non-orthonormality this is complicated. Instead we shall use "in" (or "out") energy momentum eigenstates labeled by $|E^{in}\rangle$. We assume that $|E^{in}\rangle$ is expandable in the $|A\rangle$

$$|0^{in}\rangle = |0\rangle$$

$$|\alpha K^{in}\rangle = |\alpha K\rangle = \sum_{1}^{\infty} \psi_{\alpha K}(A)|A\rangle$$

For E not equal to the vacuum or single particle state we assume

$$|E^{in}\rangle = \sum_{1}^{\infty} C_{EA}|A\rangle \ .$$

We further define

$$E^{+}in = I \qquad \text{for } E = (0)$$

$$E^{+}in = a^{+}(\alpha K) \qquad \text{for } E = \text{single particle.}$$

For all other E

$$E^{+}in = \sum_{A=1}^{\infty} C_{EA} A^{+}$$

Then for all E

$$E^{+}_{m}|0\rangle = |E_{m}\rangle$$

We can also define

$$|E^{in} F^{in} \ldots\rangle$$

by

$$E^{+}in\, F^{+}in \ldots |0\rangle$$

It is easy to show that in Eqs. (III.14) and (III.15) we can replace all E by E^{in} to obtain

$$\begin{aligned}
&\langle \mathcal{P}_1 \ldots \mathcal{P}_m | \mathcal{Q}_1 \ldots \mathcal{Q}_m \rangle \\
&= \sum_{E} \langle \mathcal{P}_1 | E_{11}{}^{in} \ldots E_{1n}{}^{in} \rangle \ldots \langle \mathcal{P}_m | E_{m1}{}^{in} \ldots E_{mn}{}^{in} \rangle \langle E_{11}{}^{in} \ldots E_{m1}{}^{in} | \mathcal{Q}_1 \rangle \\
&\qquad \times \langle E_{1n}{}^{in} \ldots E_{mn}{}^{in} | \mathcal{Q}_n \rangle \\
&\equiv \langle \mathcal{P}_m | E^{in} \ldots \rangle \otimes \langle E^{in} \ldots | \mathcal{Q}_n \rangle \qquad \text{(III.16)}
\end{aligned}$$

and

$$\langle \beta_1 L_1, \ldots \beta_m L_m | \alpha_1 K_1 \ldots \alpha_n K_n \rangle = \langle \beta L_m | E^{in} \rangle \otimes \langle E^{in} | \alpha K_n \rangle \qquad \text{(III.17)}$$

In order to use (III.17) an approximation will be used in which the sums over E^{in} are truncated, retaining only vacuum and one physical particle configurations. The inner product will then involve only amplitudes $\langle \alpha K | E_1^{in} \ldots E_n^{in} \rangle$ where the E^{in} are either vacuum or single particles. The rule for a projection $\langle \alpha K | 0 \ldots 0 \gamma_1 J_1 \cdots \gamma_n J_n \rangle$ with r vacuum configurations is just $\langle \alpha K | \gamma_1 J_1 \cdots \gamma_n J_n \rangle$.

Hence in this approximation the inner products among physical particle vectors are known functionals of the single particle projections which we denote by

$$\begin{aligned} &\langle \beta_1 L_1 \ldots \beta_m L_m | \alpha_1 K_1 \ldots \alpha_n K_n \rangle \\ &\approx \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle \end{aligned} \qquad \text{(III.18)}$$

Since the single particle state is orthogonal to $|0\rangle$ there is no contributions in which $\langle \alpha K | 0 \rangle$ appear.

As an example the inner product $\langle \beta_1 L_1, \beta_2 L_2 | \alpha_1 K_1, \alpha_2 K_2 \rangle$ is given by seven terms, two with the form

$$\begin{aligned} \int \frac{d\gamma_{11} d\gamma_{22}}{\gamma_{11}\gamma_{22}} dJ_{11} dJ_{22} &\langle \beta_1 L_1 | \gamma_{11} J_{11} \rangle \langle \beta_2 L_2 | \gamma_{22} J_{22} \rangle \langle \gamma_{11} J_{11} | \alpha_1 K_1 \rangle \\ &\times \langle \gamma_{22} J_{22} | \alpha_2 K_2 \rangle \\ = \langle \beta_1 L_1 | \alpha_1 K_1 \rangle &\langle \beta_2 L_2 | \alpha_2 K_2 \rangle \\ = \delta\left(1 - \frac{\beta_1}{\alpha_1}\right) \delta\left(L_1 - K_1\right) &\delta\left(1 - \frac{\beta_2}{\alpha_2}\right) \delta\left(L_2 - K_2\right) \end{aligned} \qquad \text{(III.19)}$$

four with the form

$$\begin{aligned} \int \frac{d\gamma}{\gamma} dJ \, &\langle \beta_1 L_1 | \gamma_{11} J_{11}, \gamma_{12} J_{12} \rangle \langle \beta_2 L_2 | \gamma_{22} J_{22} \rangle \\ &\times \langle \gamma_{11} J_{11} | \alpha_1 K_1 \rangle \langle \gamma_{12} J_{12}, \gamma_{22} J_{22} | \alpha_2 K_2 \rangle \\ &= \int \langle \beta_1 L_1 | \alpha_1 K_1, \gamma J \rangle \langle \gamma J, \beta_2 L_2 | \alpha_2 K_2 \rangle \frac{d\gamma}{\gamma} dJ \end{aligned} \qquad \text{(III.20)}$$

and one more term given by

$$\int \frac{d\gamma_{11}\cdots d\gamma_{22}}{\gamma_{11}\cdots\gamma_{22}}\, dJ_{11}\cdots dJ_{22}$$

$$\times \langle \beta_1 L_1 | \gamma_{11} J_{11}, \gamma_{12} J_{12}\rangle \langle \beta_2 L_2 | \gamma_{21} J_{21}, \gamma_{22} J_{22}\rangle$$

$$\times \langle \gamma_{11} J_{11}, \gamma_{21} J_{21} | \alpha_1 K_1\rangle \langle \gamma_{12} J_{12}, \gamma_{22} J_{22} | \alpha_2 K_2\rangle \qquad \text{(III.21)}$$

B. The Hamiltonian

Although the Hamiltonian is a very complicated function of the a^+ and a^- if indeed it can be expressed as a function of the a^+ and a^-, the matrix elements $\langle \beta_1 L_1 \ldots \beta_m L_m | H | \alpha_1 K_1 \ldots \alpha_n K_n\rangle$ are simple. In the L.R.F. the Hamiltonian has the form

$$H = \int C(\alpha K, \beta L)\, \varphi^+_{\alpha K} \varphi^-_{\beta L} + \int V(\alpha K, \beta L, \gamma J)\, \varphi^+_{\alpha K} \varphi^-_{\beta L} \varphi^-_{\gamma J} + \text{h.c.} \qquad \text{(III.22)}$$

The first term is the sum of the bare particle unperturbed energies and the second is the contribution from $\lambda\varphi^3$ which contains no vacuum terms in the L.R.F. We now introduce Lemma II.

Lemma II. For all A,B,C,D and operator H of the form in Eq. (III.22)

$$\begin{aligned}\langle AB|H|CD\rangle = &\sum_E \langle A|H|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_1E_3|C\rangle\langle E_2E_4|D\rangle \\ &+ \langle A|E_1E_2\rangle\langle B|H|E_3E_4\rangle\langle E_1E_3|C\rangle\langle E_2E_4|D\rangle \\ &+ \langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_1E_3|H|C\rangle\langle E_2E_4|D\rangle \\ &+ \langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_1E_3|C\rangle\langle E_2E_4|H|D\rangle \\ -\sum_E &\langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_1|H|E_1'\rangle\langle E_1'E_3|C\rangle\langle E_2E_4|D\rangle \\ &- \langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_2|H|E_2'\rangle\langle E_1E_3|C\rangle\langle E_2'E_4|D\rangle \\ &- \langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_3|H|E_3'\rangle\langle E_1E_3'|C\rangle\langle E_2E_4|D\rangle \\ &- \langle A|E_1E_2\rangle\langle B|E_3E_4\rangle\langle E_4|H|E_4'\rangle\langle E_1E_3|C\rangle\langle E_2E_4'|D\rangle \end{aligned} \qquad \text{(III.23)}$$

The structure consists of the sum of four "outside terms" which might be written as

$$\langle A_2|H|E\ldots\rangle \otimes \langle E\ldots|B_2\rangle + \langle A_2|E\ldots\rangle \otimes \langle E\ldots|H|B_2\rangle$$

less the sum of four inside terms written symbolically as

$$- \langle A_2|E\ldots\rangle \otimes \langle E|H|E'\rangle\langle E'\ldots|B_2\rangle$$

If we generalize Eq. (III.23) to polynomials in A,B,C,D and in particular to the polynomials $a^+(\alpha K)$ and replace the sum on E by a sum on E^{in} the structure simplifies greatly.

This is due to the simplification

$$\langle \beta L|H|E^{in}\ldots\rangle = L^2+m^2/2B \langle \beta L|E_{in}\ldots\rangle$$

and

$$\langle E^{in}\ldots|\alpha K\rangle = K^2 + m^2/2\alpha \langle E^{in}\ldots|\alpha K\rangle$$

on the outside and

$$\langle E^{in}|H|E^{in\prime}\rangle = \omega_E \langle E^{in}|E^{in\prime}\rangle$$

on the inside terms. ω_E is the energy of $|E\rangle$. Hence

$$\begin{aligned}&\langle \beta_1 L_1, \beta_2 L_2|H|\alpha_1 K_1, \alpha_2 K_2\rangle \\ &= \langle \beta_1 L_1|E_1{}^{in}E_2{}^{in}\rangle\langle \beta_2 L_2|E_3{}^{in}E_4{}^{in}\rangle\langle E_1{}^{in}E_3{}^{in}|\alpha_1 K_1\rangle\langle E_2{}^{in}E_4{}^{in}|\alpha_2 K_2\rangle \\ &\quad \times [\,(L_1{}^2+m^2)/2\beta_1+(L_2{}^2+m^2)/2\beta_2+(K_1{}^2+m^2)/2\alpha_1+(K_2{}^2+m^2)/2\alpha_2 \\ &\qquad - \omega_{E_1} - \omega_{E_2} - \omega_{E_3} - \omega_{E_4}\,] \qquad \text{(III.24)}\end{aligned}$$

It is a simple matter to generalize these equations to an arbitrary matrix element

$$\langle \beta_1 L_1\ldots\beta_m L_m|H|\alpha_1 K_1\ldots\alpha_n K_n\rangle$$

There are m + n outside terms and n·m inside terms.

$$\begin{aligned}&\langle \beta_1 L_1\ldots\beta_m L_m|H|\alpha_1 K_1\ldots\alpha_n K_n\rangle \\ &\quad = \langle \beta L_m|E^{in}\ldots\rangle\otimes\langle E^{in}\ldots|\alpha K_n\rangle \\ &\qquad \times\Big[-\sum_1^{m\cdot n}\omega_E + \sum_1^{m}\frac{L^2+m^2}{2\beta} + \sum_1^{n}\frac{K^2+m^2}{2\alpha}\Big] \qquad \text{(III.25)}\end{aligned}$$

The introduction of the vacuum and single particle saturation for the sums further simplifies $\langle H \rangle$ and produces a functional of the single particle amplitudes $\langle \beta L | \alpha_1 K_1 \ldots \alpha_n K_n \rangle$ and the masses m^2. The internal particles γJ have energy $\frac{J^2 + m^2}{2\gamma}$ and the vacuum energy is zero. The functional is denoted

$$\langle \beta L | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K \rangle \left(\sum \frac{L^2 + m^2}{2\beta} + \sum \frac{K^2 + m^2}{2\alpha} - \sum \frac{J^2 + m^2}{2\gamma} \right)$$

For example

$$\langle \beta_1 L_1 , \beta_2 L_2 | H | \alpha_1 K_1 , \alpha_2 K_2 \rangle$$

is a sum of seven terms, two with the form

$$\delta \left(1 - \frac{\beta_1}{\alpha_1} \right) \delta (K_1 - L_1) \, \delta \left(1 - \frac{\beta_2}{\alpha_2} \right) \delta (K_2 - L_2)$$

$$\times \left[\frac{K_1^2 + m^2}{2\alpha_1} + \frac{K_2^2 + m^2}{2\alpha_2} \right] \qquad \text{(III.26)}$$

four with the form

$$\int \langle \beta_1 L_1 | \alpha_1 K_1 , \gamma J \rangle \langle \gamma J , \beta_2 L_2 | \alpha_2 K_2 \rangle$$

$$\times \left[\frac{L_1^2 + m^2}{2\beta_1} + \frac{K_2^2 + m^2}{2\alpha_2} - \frac{J^2 + m^2}{\gamma} \right] \frac{d\gamma}{\gamma} \, dJ \qquad \text{(III.27)}$$

and one with the form

$$\int \frac{d\gamma_{11}}{\gamma_{11}} \cdots \frac{d\gamma_{22}}{\gamma_{22}} \, dJ_{11} \ldots dJ_{22}$$

$$\times \langle \beta_1 L_1 | \gamma_{11} J_{11} , \gamma_{12} J_{12} \rangle \langle \beta_2 | \gamma_{21} J_{21} , \gamma_{22} J_{22} \rangle$$

$$\times \langle \gamma_{11} J_{11} , \gamma_{21} J_{21} | \alpha_1 K_1 \rangle \langle \gamma_{12} J_{12} , \gamma_{22} J_{22} | \alpha_2 K_2 \rangle$$

$$\times \left[\frac{L_1^2 + m^2}{\beta_1} + \frac{L_2^2 + m^2}{\beta_2} + \frac{K_1^2 + m^2}{\alpha_1} + \frac{K_2^2 + m^2}{\alpha_2} - \frac{J_{11}^2 + m^2}{\gamma_{11}} - \frac{J_{21}^2 + m^2}{\gamma_{21}} \right.$$

$$\left. - \frac{J_{12}^2 + m^2}{\gamma_{12}} - \frac{J_{22}^2 + m^2}{\gamma_{22}} \right] \qquad \text{(III.28)}$$

C. Composite Nature of Hadrons

G. Frye and I have tried to use the structure defined in the previous sections to implement the bootstrap theory of hadrons. Each single particle state can be decomposed into a part lying in the space spanned by $|\alpha_1 K_1 \ldots \alpha_n K_n\rangle$ with $n \geq 2$ and a part orthogonal to this subspace. It is our assumption that this orthogonal part is very small or perhaps not present at all. Thus we write

$$|\alpha K \mu\rangle = \sum_{n=2}^{\infty} \int e_{\alpha K \mu}(\alpha_1 K_1 \mu_1 \ldots \alpha_n K_n \mu_n) |\alpha_1 K_1 \mu_1 \ldots\rangle d^n K \frac{d^n \alpha}{\alpha_1 \ldots \alpha_n} \tag{III.29}$$

The theory then contains as its dynamical equations two types of equations

1. Projection equations

$$\langle \beta L | \alpha K \rangle = \delta\left(1 - \frac{\beta}{\alpha}\right) \delta(L - K)$$

$$= \sum \int \langle \beta L | \alpha_1 K_1 \ldots \alpha_n K_n \rangle \, e_{\alpha K}(\alpha_1 K_1 \ldots) \times d^n K \, d^n \alpha / \alpha_1 \ldots \alpha_n$$

$$\langle \beta_1 L_1 \ldots \beta_m L_m | \alpha K \rangle = \sum \int \langle \beta_1 L_1 \ldots \beta_m L_m | \alpha_1 K_1 \ldots \alpha_n K_n \rangle \times e_{\alpha K}(\alpha_1 K_1 \ldots)$$

$$\langle \beta_1 L_1 \ldots \beta_m L_m | \alpha_1 K_1 \ldots \alpha_n K_n \rangle = \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle$$

2. Energy equations

$$\sum \int \langle \beta_1 L_1 \ldots \beta_m L_m | H | \alpha_1 K_1 \ldots \alpha_n K_n \rangle \, e_{\alpha K}(\alpha_1 K_1 \ldots) \frac{d^n K \, d^n \alpha}{\alpha_1 \ldots \alpha_n}$$

$$= \langle \beta_1 L_1 \ldots | H | \alpha K \rangle = (K^2 + m^2)/2\alpha \, \langle \beta_1 L_1 \ldots \beta_m L_m | \alpha K \rangle$$

$$\langle \beta_1 L_1 \ldots \beta_m L_m | H | \alpha_1 K_1 \ldots \alpha_n K_n \rangle = \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle \times \left(\sum \frac{L^2 + m^2}{2\beta} + \sum \frac{K^2 + m^2}{2\alpha} - \sum \frac{J^2 + m^2}{2\gamma} \right)$$

I shall digress here in order to point out how the derivation of Eqs. (III.29)-(III.33) would break down in an O.R.F.

We would not want to define the physical particle states in terms of the operators a^+_k which act on the bare vacuum to give a single particle. Such operators contain in addition to the creation operators associated with the particle, those creation operators which create the physical vacuum $|v\rangle$. Acting with products such as $a^+_k a^+_\ell$ on $|0\rangle$ would create the vacuum "stuff" twice and would not produce anything like a two body state.

If we define a^+_k as the polynomial in φ^+_k which gives $a^+_k|v\rangle$ as the single particle state then Lemma I is no longer applicable to inner products.

Secondly the Hamiltonian does not have the form given in Eq. (III.22) but contains products of three creation operators and three annihilation operators. Lemma II is not applicable to such operators. Hence Eqs. (III.29)-(III.33) rely heavily on the non-relativistic character of the L.R.F. description.

The L.R.F. has another advantage. The symmetry under the 8 parameter subgroup of the Poincare group including the "kinematic subgroup" and H can be maintained manifestly even in truncated approximations to the theory. This is sufficient to guarantee that the energy spectrum for single particle states has the form $\frac{K^2+m^2}{2\alpha}$. In the O.R.F., truncation and mutilation of a theory generally reduces the symmetry from Poincare invariance to translational and rotational symmetry. The result is that the energy spectrum cannot be guaranteed to have the form $(k^2+m^2)^{\frac{1}{2}}$. The difference lies in the fact that the kinematic subgroup for the L.R.F. is a seven parameter group while for the O.R.F. it is only a six parameter group.

The required invariance can be imposed on the functions $\langle \beta_1 L_1 \ldots | \alpha K\rangle$. They shall be invariant under transformations (I.6), (I.9), (I.10), transverse and longitudinal translations and rotations about $\mathcal{Y}$. Translational invariance requires $\langle \beta_1 L_1 \ldots | \alpha K\rangle$ to be diagonal in transverse and longitudinal momentum. Thus it has the form

$$\delta\left(1 - \frac{\beta_1+\ldots+\beta_m}{\alpha}\right)\delta(K - L_1 - \ldots - L_m)$$
$$\times P(\alpha, \beta_i, K, L_i)$$

Invariance under (I.6), (I.9) and (I.10) require P to be a function of ratios β_i/α and invariant transverse variables $L_i - \frac{\beta_i}{\alpha}K$.

$$P = P\left(\frac{\beta_i}{\alpha}, L_i - \frac{\beta_i}{\alpha}K\right)$$

Finally, invariance under γ-rotations requires P to depend on scalar products of the $\left(L_i - \frac{\beta_i}{\alpha} K\right)$ in the two dimensional x,y space.

With this in mind let us return to Eqs. (III.29) - (III.33). The invariance of the single particle projections and of the phase space $dJ\, d\gamma/\gamma$ and the region of γ and J integration insure that the multiparticle projections

$$\langle \beta_1 L_1 \ldots | \alpha_1 K_1 \ldots \rangle$$

defined by Eq. (III.31) are invariant. Therefore in order that the right hand side of (III.30) be guaranteed invariant we need only require

$$e_{\alpha K}(\alpha_1 K_1 \ldots \alpha_n K_n)$$

to be invariant.

$$e_{\alpha K}(\alpha_1 K_1 \ldots \alpha_n K_n) = \delta\left(1 - \frac{\alpha_1 + \ldots + \alpha_n}{\alpha}\right) \delta(K - K_1 - \ldots - K_n)\, E\left(\frac{\alpha_i}{\alpha}, K_i - \frac{\alpha_i}{\alpha} K\right)$$

Again the invariance of the volume element $\frac{d\alpha\, dK}{\alpha}$ and the region of integration guarantee the right hand side of Eq. (III.29) to be an invariant function of βL and αK. The only such function is a constant multiple of $\delta\left(1 - \frac{\beta}{\alpha}\right) \delta(K - L)$. Hence Eq. (III.29) is a single normalization condition which need be applied for only one value of αK.

The energy equations obtained by substituting (III.33) into (III.32) have the form

$$\sum \int \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle \left(\sum \frac{L^2 + m^2}{2\beta} + \frac{K^2 + m^2}{2\alpha} - \frac{J^2 + m^2}{2\gamma} \right) \times e_{\alpha K}(\alpha_1 K_1 \ldots) \frac{d^n K\, d^n \alpha}{\alpha_1 \ldots \alpha_n}$$

$$= \frac{K^2 + m^2}{2\alpha} \sum \int \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle\, e_{\alpha K}(\alpha_1 K_1 \ldots) \frac{d^n K\, d^n \alpha}{\alpha_1 \ldots \alpha_n}$$

or

$$\sum \int \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle \, \alpha \left(\sum_1^m \frac{L^2+m^2}{\beta} + \sum_1^n \frac{K^2+m^2}{\alpha} - \sum \frac{J^2+m^2}{\gamma} - \frac{K^2}{\alpha} \right)$$

$$\times e_{\alpha K}(\alpha_1 K_1 \ldots \alpha_n K_n) \frac{d^n K \, d^n \alpha}{\alpha_1 \ldots \alpha_n}$$

$$= m^2 \sum \int \langle \beta L_m | \gamma J \ldots \rangle \otimes \langle \gamma J \ldots | \alpha K_n \rangle \, e_{\alpha K}(\alpha_1 K_1 \ldots) \frac{d^n K \, d^n \alpha}{\alpha_1 \ldots \alpha_n}$$

It is easy to check that both sides of (III.37) are invariant and hence may be evaluated for a single value of αK. Eq. (III.37) is an eigenvalue, eigenvector problem for the eigenvalue M^2 and eigenvector E. The kernel is a given functional of the functions P and the mass spectrum.

The structure of the equations of the theory are

$$P = A(P, m^2)$$

$$m^2 = a(P, m^2)$$

where a and A are nonlinear functionals, P is the set of P functions and m^2 the set of masses. The set of operations implied by A and a can be made more explicit. Let us imagine a parametrized form of P and a set of masses which we will call P^{in} and M^{2in}. The subscript in means input. In terms of P^{in} and M^{2in} the kernel for Eq. (III.37) can be obtained and a solution for the eigenvectors E and eigenvalues M^2 called E^{out} and M^{2out} can be found. Eq. (III.29) is then used to normalize E and (III.30) then defines P^{out}. Our equations are

$$P^{out} = P^{in}, \quad m^{2\,out} = m^{2\,in}$$

A single particle exchange model similar to the single particle exchange model of N/D bootstrap theory may be set up by truncating the expansion of one particle states at n = 2 and then evaluating the two body matrix elements $\langle \beta_1 L_1, \beta_2 L_2 | \alpha_1 K_1, \alpha_2 K_2 \rangle$ and $\langle \beta_1 L_1, \beta_2 L_2 | H | \alpha_1 K_1, \alpha_2 K_2 \rangle$ retaining the terms in (III.19), (III.20), (III.26) and (III.27).

Having obtained the projections and masses, the multiparticle matrix elements of H can be constructed and the energy equations set up for values of energy in the scattering region. The scattering amplitude and particle masses and couplings have been obtained in this manner for pi-nucleon systems by G. Frye and he finds them to be in agreement with the Chew reciprocal bootstrap theory when the mass of

the nucleon is taken to be much larger than that of π. The method can also be applied to π-π problems and a bootstrap calculation of ρ may be made which is again similar to N/D calculations.

IV. Application to High Energy Collisions

A. General Theory

Graham Frye and I are considering the application of the L.R.F. description of hadronic systems to the description of high energy systems incident on targets at rest or near at rest in the laboratory frame. We wish to know what the postulated L.R.F. description of the system with high momentum implies about high energy collisions.

Our idea, somewhat oversimplified, can be expressed as follows:

Let us suppose that the degrees of freedom of the incident high energy system and the target remain independent during the whole course of the collision. By this we mean that the state of the system is spanned by vectors in a product space formed from the cartesian product of the space of an infinite momentum system and the space of an ordinary finite system.

$$|\psi\rangle = \sum_{ij} |\psi_i\rangle \, |\varphi_j\rangle$$

where ψ_j refers to the L.R.F. description of the incident system and φ_j to the O.R.F. description of the target. The independence of the degrees of freedom does not mean that there is no interaction between target and system. It just says that the scattering matrix acts in this product space. This is a very non-trivial constraint on the form of the s-matrix at asymptotic energies. The products of collision in the final state must be decomposable into two distinct subsets, the first subset associated with the target must have finite momentum (meaning much less than the incident system) in the lab frame. The second subset must contain only particles of finite transverse momentum, and a finite fraction of the incoming longitudinal momentum. Furthermore the invariant mass of each of the final subsets must be small in comparison with the incoming momentum. Any particle in the final state with large transverse momentum would be neither an L.R.F. particle nor an O.R.F. particle.

Now these constraints are in fact the most prominent feature of many high energy collision events as seen empirically. It is therefore of interest to pursue these ideas further. The first observation I shall make is that the above constraints require α for the L.R.F. system to be conserved.

Consider an event in which a system of momentum $(\alpha e^{\omega}, K)$ hits a target of momentum $\vec{q}$. The final state involves a system of momentum $(\alpha' e^{\omega}, K')$ and a system of momentum $\vec{q}'$ with $|\vec{q} - \vec{q}'| \ll e^{\omega}$. Momentum and energy conservation give

$$(\alpha - \alpha')e^{\omega} = q_z' - q_z \tag{IV.1}$$

$$K - K' = q'_{(x,y)} - q_{(x,y)} \tag{IV.2}$$

$$(\alpha - \alpha')e^{\omega} + 0(e^{-\omega}) = \omega_{q'} - \omega_q \tag{IV.3}$$

From (IV.1), (IV.2) and (IV.3) we find that t, the invariant momentum transfer squared, is

$$-t = (K - K')^2 + 0(e^{-\omega}) \tag{IV.4}$$

and the square of the center of mass energy is

$$s = 2\alpha e^{\omega}(q_o - q_z) + \text{finite} \tag{IV.4'}$$

From (IV.1) α is conserved to within terms of order $e^{-\omega}$. (IV.3) constrains the initial and, final target momenta to satisfy $(\omega' - q_z') = (\omega - q_z)$ to within terms of order $e^{-\omega}$.

The conservation of α will be assumed to be more general that just for initial and final states. We shall assume α conserved for the course of the collision. For a collision at fixed α the space of states is then $|\psi_\alpha\rangle|\varphi\rangle$. We postulate a Hamiltonian $H = H_o^i + H_o^T + V$ where H_o^i acts on the L.R.F. incident degrees of freedom, H_o^T on the target degrees and V couples the two systems. We know that $H_o^i = \alpha e^{\omega} + 0(e^{-\omega})$. Since α is conserved in the collision the term αe^{ω} may be dropped.

To do this consistently we must make a slight modification of the Schrödinger equation. Let $\hat{P}_z$ be the z component of momentum operator for the target degrees of freedom. The total z-momentum is $\alpha e^{\omega} + \hat{P}_z$. Instead of writing $i\frac{\partial}{\partial t} = H$ we write

$$\left(i\frac{\partial}{\partial t} + i\frac{\partial}{\partial z}\right)|\psi\rangle = (H - \hat{P}_z - \alpha e^{\omega})|\psi\rangle$$

$$= (H_o^T - \hat{P}_z)|\psi\rangle + V|\psi\rangle$$

where $H_o^T - \hat{P}_z$ is an operator which acts on the target degrees of freedom. It is a kind of Hamiltonian which propagates in the $z + t$ direction instead of the t direction. We shall denote it by h^T and use it as though it were the target Hamiltonian remembering that its eigenvalues

are $\omega = P\gamma$ instead of just ω.

Defining $t + \gamma = u$

$$i\frac{\partial}{\partial \mu}|\psi\rangle = (h^T + V)|\psi\rangle$$

Also in order to get a scattering amplitude which does not go as $e^{-\omega}$, V must be of order 1 and not $e^{-\omega}$. That is the collision processes take place on the scale of ordinary time and not on the dilated internal time of the incoming system. Hence we may ignore the terms in H which go as $e^{-\omega}$ since they represent processes which take a very long time. Actually we may ignore them for the duration of the interaction between the two sets of degrees of freedom. After target and system have separated there is an infinite amount of time available for the $e^{-\omega}$ terms to act as final state interaction allowing the products of collision to become final asymptotically separated particles. The point may be summarized by saying that since the free motion of the incoming system is time dilated while the collision process is not, the impulse approximation becomes exact as $e^{\omega} \to \infty$. The Hamiltonian is then $V(\tau, i) + h^T$ where τ and i denote target and incident degrees of freedom. As stated $V(\tau, i)$ must not have matrix elements connecting different values of α.

$$[V, \alpha] = 0 \tag{IV.5}$$

Thus the Schrödinger equation becomes

$$i\frac{\partial|\psi\rangle}{\partial u} = V(i,\tau)|\psi\rangle + h^T|\psi\rangle \tag{IV.6}$$

We define a kind of interaction representation by factoring out the free time development of the target

$$i\frac{\partial|\tilde{\psi}\rangle}{\partial u} = \tilde{V}(i,\tau)|\tilde{\psi}\rangle \tag{IV.7}$$

where

$$|\tilde{\psi}\rangle = e^{-ih^T\mu}|\psi\rangle \tag{IV.8}$$

and

$$\tilde{V}(\mu) = e^{ih^T u}\, v\, e^{-ih^T u} \tag{IV.9}$$

The solution is given by

$$|\tilde{\psi}(u_2)\rangle = U\, e^{i\int_{u_1}^{u_2} \tilde{V}(u)\,du}\, |\tilde{\psi}(u_1)\rangle \qquad \text{(IV.10)}$$

Where U means u ordered product defined in analogy with the usual time ordered product. The scattering amplitude is given by

$$S_{\psi_i\varphi_i \to \psi_f\varphi_f} = \langle \psi_i^{in}| \langle \varphi_i^{in}| U\, e^{i\int_{-\infty}^{\infty} \tilde{V}(u)du} |\varphi_f^{out}\rangle |\psi_f^{out}\rangle \qquad \text{(IV.11)}$$

where ψ_i is the initial configuration of the incoming system, ψ_f its final configuration and φ_i and φ_f are initial and final target states. The vectors $|\psi_i^{in}\rangle$, $|\psi_f^{out}\rangle$, $|\varphi_i^{in}\rangle$, $|\varphi_f^{out}\rangle$ are the eigenvectors of H_o^i and H_o^T corresponding to the given asymptotic boundary condition when V is absent.

Actually Eq. (IV.11) is a very elaborate generalization of the impulse and eikonal approximations. To see the impulsive character of the formula we note that the only dependence that the formula has on the part of the Hamiltonian which generates the motions of the incident system by itself is through the dependence on the initial and final vectors $|\psi_i^{in}\rangle$ and $|\psi_f^{out}\rangle$. These will be complicated superpositions of physical particle or bare particle states whose wave functions will depend on H_o^i. That is to say H_o^i affects the motion before and after collision but not during.

We can see the eikonal aspects if we make a simple model of the target in which it is like a rigid potential. Its excited states involve so much energy that they are not important at all. The state of the target is simply called $|\varphi\rangle$ and $\langle\varphi|\tilde{V}(i,\tau,u)|\varphi\rangle \equiv V(i,u)$. Then Eq. (IV.10) becomes

$$|\psi\rangle_{\mu_2} = U\, e^{i\int_{u_1}^{u_2} V(i,u)du}\, |\psi\rangle_{\mu_1} \qquad \text{(IV.12)}$$

which expresses the scattering as a product of successive scatterings at different u-values. Eq. (IV.11) becomes

$$S_{i\to f} = \langle \psi_i^{in}|\, u\, e^{i\int V(iu)du}\, |\psi_f^{out}\rangle \qquad \text{(IV.13)}$$

which can be compared with the relativistic eikonal approximation for a point particle.

$$S = \langle i|e^{i\int V(u)du}|f\rangle \qquad \text{(IV.14)}$$

In all equations the operator V is of course diagonal in α but possibly dependent on α. Hence the scattering operator should really be

$$U\, e^{i\int \tilde{V}(i,\tau,u,\alpha)du} \tag{IV.15}$$

For the rest of this lecture I will use the model of the rigid potential of Eq. (IV.12) and (IV.13) only to save myself the tedious insertion of the target variables.

We must now determine the properties of $V(i,\alpha)$.

$V(i,\alpha)$ is an operator in the space of the L.R.F. degrees of freedom. Hence its matrix elements in the physical particle basis must be expandable. At this point, in order to illustrate the ideas Frye and I are proposing I shall make some approximations. The approximations are by no means crucial to the arguments but to go beyond them requires some further developments of the physical particle representation. The first approximation is to truncate the expansion of matrix elements of V at two particle contributions.

$$\langle \beta L|V|\alpha K\rangle \approx \int dL_1\, dL_2\, dK_1\, dK_2\, \frac{d\beta_1}{\beta_1}\,\frac{d\beta_2}{\beta_2}\,\frac{d\alpha_1}{\alpha_1}\,\frac{d\alpha_2}{\alpha_2}$$

$$\times e_{\beta L}(\beta_1 L_1, \beta_2 L_2)^{*}\langle \beta_1 L_1, \beta_2 L_2|V|\alpha_1 K_1, \alpha_2 K_2\rangle$$

$$\times e_{\alpha K}(\alpha_1 K_1, \alpha_2 K_2)\ . \tag{IV.16}$$

The second approximation is to assume that matrix elements of V between two body states can be expressed as linear sums of matrix elements between the constituent particles.

$$\langle \beta_1 L_1, \beta_2 L_2|V|\alpha_1 K_1, \alpha_2 K_2\rangle = \langle \beta_1 L_1|V|\alpha_1 K_1\rangle \delta(L_2 - K_2)$$

$$\times \delta\left(1 - \frac{\alpha_2}{\beta_2}\right) + \text{symmetrizing terms}. \tag{IV.17}$$

Combining (IV.17) with (IV.16) gives a homogeneous equation for the single particle matrix elements of V. Let

$$\langle \beta L|V|\alpha K\rangle = V(\alpha, L, K)\,\delta(\alpha - \beta) \tag{IV.18}$$

and

$$e_{\alpha K}(\alpha_1 K_1, \alpha_2 K_2) = E\left(\frac{\alpha_1}{\alpha},\ K_1 - \frac{\alpha_1}{\alpha}K\right)\delta\left(1 - \frac{\alpha_1+\alpha_2}{\alpha}\right)$$

$$\times \delta(K - K_1 - K_2)\ . \tag{IV.19}$$

In Eq. (IV.18) we make use of the assumed α-conservation and in Eq. (IV.19) we use invariance of the single particle wave functions under longitudinal boosts and transverse "galilean boosts." It is the assumed validity of the invariance of the description of the incident system under the L.R.F. transformations described in Sec. I which is the important point. From Eqs. (IV.16), (IV.17), (IV.18) and (IV.19) we obtain

$$V(\alpha,K,L) = \int E\left(\frac{\alpha}{\gamma},L'-\frac{\gamma}{\alpha}L\right)^* E\left(\frac{\alpha}{\gamma},K'-\frac{\gamma}{\alpha}K\right)\frac{\alpha^2}{\gamma^2(\alpha-\gamma)}$$

$$\times\,\delta(L-K-L'-K')\,V(\gamma,L',K')\,d\gamma dK'dL' \qquad \text{(IV.20)}$$

It is easy to prove that V can only depend on the combination L - K and not L + K. Imagine making a very small rotation of the entire incident system relative to the target. The rotation is about the x axis and takes every Ky into Ky + αQ. The scattering amplitude should not be altered by such a small rotation although the description of the incoming system changes by a finite amount. Therefore the scattering amplitude must be invariant under a change of every transverse momentum corresponding to a "galilean boost." We guarantee this by demanding that V be invariant. Thus $V(\alpha,K,L)$ must be invariant under $K \to K + \alpha Q$, $L \to L + \alpha Q$. The combination K - L is invariant and K + L is not. Hence $V = V(\alpha, K - L)$. Define $\Delta = K - L$. Eq. (IV.20) then gives

$$V(\alpha\Delta) = \int E\left(\frac{\alpha}{\gamma},L'-\frac{\gamma}{\alpha}L\right)^* E\left(\frac{\alpha}{\gamma},K'-\frac{\gamma}{\alpha}K\right)\frac{\alpha^2}{\gamma^2(\alpha-\gamma)}$$

$$\delta(K'-L'-\Delta)\,V(\gamma,\Delta)\,d\gamma dL'dK' \qquad \text{(IV.21)}$$

which is a homogeneous integral equation in one variable for each value of Δ. Let $U = L' - \frac{\gamma}{\alpha}L$ and $V = K' - \frac{\gamma}{\alpha}K$

$$V(\alpha\Delta) = \int E\left(\frac{\alpha}{\gamma},U\right)^* E\left(\frac{\alpha}{\gamma}V\right)\frac{\alpha^2}{\gamma^2(\alpha-\gamma)}$$

$$\delta(U - V - \Delta)\,V(\gamma\Delta)\,dUdVd\gamma \qquad \text{(IV.22)}$$

Define $R(\alpha,\gamma,\Delta)$ by

$$R\left(\frac{\alpha}{\gamma}\Delta\right) = \int E\left(\frac{\alpha}{\gamma}U\right)^* E\left(\frac{\alpha}{\gamma}V\right)\frac{\alpha^2}{\gamma(\alpha-\gamma)}\,\delta(U-V-\Delta)\,dUdV \qquad \text{(IV.23)}$$

Then

$$V(\alpha\Delta) = \int_0^\alpha \frac{d\gamma}{\gamma} \; R\left(\frac{\alpha}{\gamma},\Delta\right) V(\gamma,\Delta) \; . \qquad \text{(IV.23)}$$

Eq. (IV.23) and all that precedes it can be generalized if more than one type of particle exist. The matrix of single particle matrix elements of V denoted by $V_{ab}(\alpha\Delta)$ satisfies a matrix generalization of (IV.23)

$$V_{ab}(\alpha\Delta) = \int_0^\alpha \frac{d\gamma}{\gamma} \; R_{ab}^{\;cd}\left(\frac{\alpha}{\gamma},\Delta\right) \; V_{cd}(\gamma\Delta) \qquad \text{(IV.24)}$$

In (IV.24) the indices a, b, c, and d label the discrete variables which determine particle type.

Eq. (IV.24) is a linear integral eigenvalue equation for an eigenvector $V(\alpha\Delta)$ with eigenvalue 1. The variable Δ enters only as a parameter. Let us begin investing the solutions of (IV.24) supposing for the moment that the eigenvalue 1 is non-degenerate. Then there is a single linearly independent solution and all solutions are obtained from this solution by multiplication by a function of Δ.

Let x be a real number and consider $Vab(\alpha x,\Delta)$.

$$Vab(\alpha x,\Delta) = \int_0^{\alpha x} \frac{d\gamma}{\gamma} \; R_{ab}^{\;cd}\left(\frac{\alpha x}{\gamma},\Delta\right) V_{cd}(\gamma\Delta) \quad . \qquad \text{(IV.26)}$$

Now make the change of variables $\gamma \to x\gamma$.

$$Vab(\alpha x,\Delta) = \int_0^\alpha \frac{d\gamma}{\gamma} \; R_{ab}^{\;cd}\left(\frac{\alpha}{\gamma},\Delta\right) V_{cd}(\gamma x,\Delta) \qquad \text{(IV.27)}$$

Hence $V\,ab(\alpha x,\Delta)$ satisfies Eq. (IV.25) if $V\,ab(\alpha\Delta)$ does. Therefore $V\,ab(\alpha x,\Delta)$ must be a multiple of $V\,ab(\alpha\Delta)$

$$V\,ab(\alpha x,\Delta) = D(x,\Delta) \qquad \text{(IV.28)}$$

Let $x = 1 + \varepsilon$ and use $D(1,\Delta) = 1$.

$$V\,ab(\alpha\Delta) + \alpha\varepsilon \frac{\partial V}{\partial \alpha} = V\,ab(\alpha\Delta) + \varepsilon \frac{\partial D}{\partial x} V\,ab$$

$$\alpha \frac{\partial V}{\partial \alpha} = d(\Delta)\, V \qquad \text{(IV.28)}$$

In (IV.28), $d(\Delta)$ is $\frac{\partial D}{\partial x}$ at $x = 1$. The solution of (IV.28) is

$$V\,ab(\alpha\Delta) = C\,ab(\Delta)\, \alpha^{d(\Delta)} \qquad \text{(IV.29)}$$

In order to understand the significance of Eq. (IV.29) we go back to Eq. (IV.13) and take the first term in the expansion of the exponential.

$$T \approx \int V \, du$$

which from (IV.29) gives

$$T\,ab(\alpha,\Delta) \approx C\,ab(\Delta)\,\alpha^{d(\Delta)} \qquad \text{(IV.30)}$$

For circular symmetry of the target $T\,ab(\alpha\Delta)$ will depend only on α and $|\Delta|$ so that Δ may be replaced by the invariant t. Using Eq. (IV.4) and (IV.4′) gives an asymptotic form for the scattering amplitude

$$T\,ab(st) \approx C\,ab(t)\,s^{d(t)} \qquad \text{(IV.31)}$$

for $S \sim e^{(w)}$.

Eq. (IV.31) is identical to the form of the high energy limit of Regge Pole theory. The $Cab(t)$ are the residue functions and $d(t)$ is the Regge trajectory. If more than one linearly independent solution of (IV.24) exists then the general solution will be a sum of terms like (IV.31) with a form analogous to a sum of Regge Poles.

The equation for the Regge trajectory is found by using the solution in (IV.30) in (IV.24).

$$Cab(\Delta)\,\alpha^{d}(\Delta) = \int_{o}^{\alpha} \frac{d\gamma}{\gamma} R_{ab}^{cd}\left(\frac{\alpha}{\gamma}, \Delta\right) Ccd(\Delta)\,\gamma^{d(\Delta)}$$

or

$$Cab(\Delta) = \int_{o}^{\alpha} R_{ab}^{cd}\left(\frac{\alpha}{\gamma},\Delta\right)\left(\frac{\gamma}{\alpha}\right)^{d(\Delta)} \frac{d\gamma}{\gamma} Ccd(\Delta)$$

Let $\frac{\gamma}{\alpha} = \mu$

$$Cab(\Delta) = \int_{0}^{1} R_{ab}^{cd}\left(\frac{1}{\mu},\Delta\right)\mu^{d(\Delta)} \frac{d\mu}{\mu} Ccd(\Delta)$$

The condition for a solution is

$$\mathrm{Det}\left[\int_{0}^{1} R_{ab}^{cd}\left(\frac{1}{\mu},\Delta\right)\mu^{d(\Delta)} \frac{d\mu}{\mu} - \delta ac\,\delta db\right] = 0 \qquad \text{(IV.32)}$$

With R_{ab}^{cd} being thought of as a matrix in a space for which $\delta ac \delta db$ is the unit matrix.

B. Consequences

We conjecture that V and not the transition matrix T is the object with pure Regge Pole behavior. We also assume the Regge Poles are just those postulated by S-matrix workers. The first order of $Ue^{i\int Vdu}$ will have the usual Regge structure. The higher orders will certainly induce non-Regge terms. I shall not attempt to compute the non-Regge contributions. Some simple plausible although non-rigorous arguments will be used to show that the non-Regge terms have reasonable properties and help improve the agreement with experiment. We assume that individual matrix elements of V are very small and comparable to typical inelastic scattering processes at large energy. The first correction equals

$$-\int du \int du'\ U(V(\mu)\ V(\mu'))$$

$$= i \int du\, du'\ V(\mu)\ V(\mu') + i \int d\mu du' [V(u), V(u')]$$

$$= [i \int du V(u)]^2 + i \int du du' [V(u), V(u')] \qquad \text{(IV.33)}$$

The second commutator term I don't know how to evaluate. I shall assume that [V,V] is an operator of similar order of magnitude to V itself and produces effects of the same general order of magnitude.

Elastic matrix elements of $(\int V)^2$ are different because they consist of sums of positive definite terms summed over all available channels. At high energies there are a great many channels to consider and this term may represent effects an order of magnitude larger than those contained in V. Inelastic matrix elements of $(\int V)^2$ are of course not sums of positive definite terms and will tend to be smaller than elastic. The elastic amplitude

$$\sum_{\psi} \int \langle \alpha L | V | \psi_{\alpha K'} \rangle \langle \psi_{\alpha K'} | V | \alpha K \rangle\ dK'$$

is an integral over t and t' where $-|K'-L|^2 = t$, $|K'-K|^2 = t'$. Assuming $\langle \alpha L | V | \alpha' K' \rangle$ behaves like $\alpha^{d(t)}$ the entire contribution is an integral over $\alpha^{d(t)+d(t')-1}$ and the entire amplitude will go as $\alpha^{C(t'')}$ where $C(t'')$ is the maximum value of $d(t) + d(t') - 1$ which appears in the integral. The -1 is present because in the intermediate state summation an α^{-1} phase space factor appears. This prescription gives rise to a leading behavior which is identical to the Regge-Cut phenomena discussed by Amati, Fubini and Stanghellini and by Mandelstam.

Further iterations produce an infinite sequence of Regge-cuts, the branch points following trajectories which becomes flatter and flatter.

For example, let us consider the effect of iterating Pomeranchuk. Let Pomeranchuk be given by

$$d'(t) = d'(u) + c'(t)$$

$$= 1 + c'(t)$$

The leading power generated by iteration is $d^2(t'')$ given by

$$d^2(t'') = \text{Max}\,[1+c'(t)+c'(t')] = \text{Max}\,[1+c'(t+t')]$$

The generalization to n iterations is given by $d^n(\tau)$

$$d^n(\tau) = \text{Max}\,[1+c'(t+t'+\ldots+t^n)]$$

The maximum occurs at

$$t = t' = \ldots = t^n = \tau/n^2$$

and

$$d^n(t) = 1 + nc'\frac{t}{n^2} = 1 + c'\frac{t}{n}$$

For large n the trajectory is flat and gives rise to an amplitude which goes as a single power of s for all t. For elastic scattering this appears to be correct at high energies.

For inelastic scattering the sums over states corresponding to an iteration of V are not sums over positive definite quantities. Therefore it is not clear that the first term of the expansion in V does not dominate some region of energy. At sufficiently large energy, the amplitude should be dominated by iteration of Pomeranchuck trajectories together with a single quantum number exchange trajectory which leads a flattening of the effective quantum number exchange trajectory. At lower energies however, where the single power of V may dominate, we may see a t-dependent power law $s^{d(t)}$ for inelastic phenomena. This is also consistent with experimental charge exchange data.

Actually the theory proposed here cannot be correct. The reason is that the hermiticity of V conflicts with crossing symmetry. Consider for example the matrix element between chargeless self charge conjugate mesons of the Pomeranchuk part of V. Hermiticity requires it to be real while crossing symmetry requires it to be complex and in fact pure imaginary at t = 0. Similar observations can be made about charge exchange trajectories. My own guess is that the potential V should not be Hermitian, reflecting the possibility of excitation of states outside the product space $|\psi_i\rangle\,|\varphi_j\rangle$. Probability may leak into

states with particles which cannot be uniquely associated with target or incident system, possibly through multiperipheral mechanisms, the complex parts of V implied by a demand of manifest crossing symmetry reflecting the leakage of probability into configurations far from the originally proposed product configurations.

References

1. W. Heisenberg, Physik 120, 513, 673 (1943); H. Stapp, Phys. Rev. 125, 2139 (1962); and G. F. Chew, S-Matrix Theory of Strong Interactions (W. A. Benjamin, New York, 1961).
2. P. A. M. Dirac, Principles of Quantum Mechanics (Oxford Press, 1958); and J. Von Neumann, Mathematical Foundations of Quantum Mechanics (Princeton University Press, Princeton, N.J., 1955).
3. M. Goldberger and K. Watson, Phys. Rev. 127, 2284 (1962).
4. L. Susskind, "Gedanken Experiments Involving Hadrons," Phys. Rev. Letters (to be published).
5. L. Susskind, Phys. Rev. 165, 1535 (1968).
6. L. Susskind and G. Frye, Phys. Rev. 165, 1553 (1968).
7. R. Blankenbecler and L. F. Cook, Jr., Phys. Rev. 119, 1745 (1960).
8. F. Gross, Phys. Rev. 134B, 406 (1964).
9. R. Haag, Det. Kong. Danske Vid. Selsk. Mat.-fys. Medd. 29, No. 12 (1955).
10. S. Weinberg, Phys. Rev. 150, 1313 (1966).
11. E. Lippmann and J. Schwinger, Phys. Rev. 79, 469 (1950).
12. S. Schweber, An Introduction to Relativistic Quantum Field Theory (Row Peterson and Co., 1961).
13. D. Amati, S. Fubini, A. Stanghellini, Phys. Letters 1, 29 (1962) and S. Mandelstam, Nuovo Cimento 30, 1113, 1117, 1148 (1963).

NUCLEAR QUASI-SPIN MODELS AND THE REPRESENTATIONS OF $U(4) \supset U(2) \otimes U(2)$†

M. Resnikoff
Department of Physics and Astronomy
State University of New York
Buffalo, New York

I would like to talk today about an interesting group theoretical problem, the solution of which has application to quasi-spin models of nuclear physics. The mathematical method itself has further mathematical applications which we can discuss at the end of the lecture.

To provide physical motivation, we will first discuss the quasi-spin model of Kerman[1] and its generalization by Hecht[2] and Pang.[3] We will go into some detail on the quasi-spin model, but most is on slides so we can go through it quickly. The essential reason for the use of the decomposition $U(4) \supset U(2) \otimes U(2)$ in quasi-spin models is the following--the model may be formulated in group theoretical terms involving the rotation group in five dimensions $O(5)$ for the J coupling scheme and $O(8)$ for the LS coupling scheme. The physical quantum numbers then appear in the decomposition $O(5) \supset O(3) \otimes O(3)$, or $O(8) \supset O(6) \supset O(3) \otimes O(3)$. In this latter chain, $O(6)$ is isomorphic to $SU(4)$ and $O(3)$ is isomorphic to $SU(2)$, hence the problem of reducing $SU(4) \supset SU(2) \otimes SU(2)$.

Kerman[1] was responsible for the phrase "quasi-spin." The quasi-spin model incorporates two essential requirements: first, independent particle motion in a deformed field (the Nilsson model), and second, explicit two particle forces or pairing forces. The Nilsson model has been very successful in describing single particle properties of rotating nuclei, and pairing forces are known to be important because of the even-odd mass differences and the absence of single particle excitations in the low energy spectrum of even-even nuclei.[1] The equilibrium shape of the system is determined by the competition between the single particle tendency for deformations with the two particle force to give spherical systems.

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

The pairing part of the Hamiltonian for N identical particles of configuration j^N is

$$H = - G \sum_{m,m'>0} (-1)^{2j-m-m'} a^+_{jm} a^+_{j-m} a_{j-m'} a_{jm'}$$

The state function which diagonalizes the total Hamiltonian, single particle plus pairing force, is

$$A_+ A_+ \ldots A_+ a^+_{jm_1} a^+_{jm_2} \ldots a^+_{jm_v} |0\rangle$$

A_+ is the operator which creates a pair coupled to $J = 0$, and v is the seniority number, the number of particles remaining after removal of $J = 0$ pairs. It is only necessary to require a pair coupled to $J = 0$ because of the seniority force, an assumption that all pairing force energies E_J are equal except the lowest E_0. The seniority force assumption allows us to express the Hamiltonian in terms of N, the number of particles, and the operators $A_\pm$, A_0. We won't go into this further.

The pairing operators may be written

$$A_+ = \sum_{m>0} (-1)^{j-m} a^+_{jm} a^+_{j-m}$$

$$A_- = \sum_{m>0} (-1)^{j-m} a_{j-m} a_{jm}$$

$$A_0 = \tfrac{1}{2} \sum_{m>0} (a^+_{jm} a_{jm} + a^+_{j-m} a_{j-m} - 1)$$

The operators $A_\pm$ create (annihilate) pairs of particles coupled to $J = 0$, and satisfy the commutation relations of the generators of O(3), hence the designation quasi-spin. The Hamiltonian H may be written

$$H = - G A_+ A_-$$

and the energies levels are

$$E = - \tfrac{1}{4}(N - v)(2j + 3 - N - v) \cdot G$$

when the Hamiltonian acts on the above state function.

We next wish to generalize this model to spin $\frac{1}{2}$ particles--the neutron and proton: $m_+ = -\frac{1}{2}$, neutron and $m_+ = +\frac{1}{2}$ proton--with spin s, m_s and orbital angular momentum ℓ, m_ℓ. We construct pairs coupled to total orbital angular momentum $L = 0$.[3] There are 28 such operators, forming the generators of O(8). I remind you from the rotation group O(3), there exists $\frac{1}{2}\cdot 3\cdot 2 = 3$ planes, an angle in each, hence three parameters and generators. For O(8), there exists $\frac{1}{2}\cdot 8\cdot 7 = 28$ generators. Just as for $A_\pm, A_o$, we can separate the generators into pair creation, annihilation, and number conserving operators:

(a) Pair creation operators: (i) spin up proton, spin down proton

$$Q_+^{p,-p} = \sum_m (-)^{\ell-m} a^+_{\ell m\uparrow p} a^+_{\ell -m\downarrow p}$$

(ii) spin up proton, spin up neutron

$$Q_+^{pn} = \sum_m (-)^{\ell-m} a^+_{\ell m\uparrow p} a^+_{\ell -m\uparrow n}$$

(iii) spin up proton, spin down neutron
(iv) spin up neutron, spin down neutron
(v) spin down proton, spin up neutron
(vi) spin down proton, spin down neutron

The Pauli principle eliminates the other choices.

(b) Pair annihilation operators (similar to above)
(c) Number conserving operators:
(i) Number operators

$$Q_o^{pp}, Q_o^{nn}, Q_o^{-p-p}, Q_o^{-n-n}$$

(ii) Spin exchange operators

$$Q_o^{p-p}, Q_o^{n-n}, Q_o^{-pp}, Q_o^{-nn}$$

(iii) Charge exchange operators

$$Q_o^{pn}, Q_o^{np}, Q_o^{-p-n}, Q_o^{-n-p}$$

(iv) Space exchange operators

$$Q_o^{p-n}, Q_o^{n-p}, Q_o^{-pn}, Q_o^{-np}$$

To more clearly see that these 28 operators represent the generators of O(8), we may couple the pairs to total spin S = 1, total isospin T = 0 or total spin S = 0, total isospin T = 1, i.e., we obtain the tensor character of these operators. With the appropriate Clebsch-Gordanary, this is

$$^{2S+1,2T+1}A^{+}(M_S,M_T) = \sqrt{\ell+\tfrac{1}{2}} \sum_{m,m_t,m_s} \langle \ell m\, \ell -m | oo\rangle \langle \tfrac{1}{2} m_t \tfrac{1}{2} m_t' | TM_T\rangle$$

$$\times \langle \tfrac{1}{2} m_s \tfrac{1}{2} m_s' | SM_S\rangle\ a^{+}_{\ell m\, m_t\, m_s}\ a^{+}_{\ell -m\, m_t' m_s'}$$

Example:

$$^{13}A^{+}(0,1) = Q_{+}^{n-n} \qquad ^{31}A^{+}(1,0) = Q_{+}^{np}$$

$$^{13}A^{+}(0,0) = 1/\sqrt{2}(Q_{+}^{p-n}+Q_{+}^{n-p}) \qquad ^{31}A^{+}(0,0) = 1/\sqrt{2}(Q_{+}^{n-p}+Q_{+}^{-np})$$

$$^{13}A^{+}(0,-1) = Q_{+}^{p-p} \qquad ^{31}A^{+}(-1,0) = Q_{+}^{-n-p}$$

Number operators:

$$Q_o = -\tfrac{1}{2}(Q_o^{nn}+Q_o^{-n-n}+Q_o^{pp}+Q_o^{-p-p}) = -\tfrac{1}{2}N_{op}+\Omega$$

$$S_o = \tfrac{1}{2}(Q_o^{nn}+Q_o^{pp}-Q_o^{-n-n}-Q_o^{-p-p})$$

$$T_o = \tfrac{1}{2}(Q_o^{nn}+Q_o^{-n-n}-Q_o^{pp}-Q_o^{-p-p})$$

$$E_{oo} = \tfrac{1}{2}(Q_o^{nn}+Q_o^{-p-p}-Q_o^{pp}-Q_o^{-n-n})$$

These are the four linear commuting operators of this rank four group. N_{op} = total number of nucleons, and Ω is the spatial degeneracy, e.g., $= \ell+1$ if no spin, isospin. Finally, the exchange operators may be written:

$$S_{+}=1/\sqrt{2}(Q_o^{n-n}+Q_o^{p-p}) \qquad S_{-}=1/\sqrt{2}(Q_o^{-nn}+Q_o^{-pp}) \qquad E_{1-1}=Q_o^{p-n}$$

$$T_{+}=1/\sqrt{2}(Q_o^{np}+Q_o^{-n-p}) \qquad T_{-}=1/\sqrt{2}(Q_o^{pn}+Q_o^{-p-n}) \qquad E_{11}=Q_o^{n-p}$$

$$E_{10}=1/\sqrt{2}(Q_o^{n-n}-Q_o^{p-p}) \qquad E_{-10}=1/\sqrt{2}(Q_o^{-nn}-Q_o^{-pp}) \qquad E_{-11}=Q_o^{-np}$$

$$E_{01}=1/\sqrt{2}(Q_o^{np}-Q_o^{-n-p}) \qquad E_{0-1}=1/\sqrt{2}(Q_o^{pn}-Q_o^{-p-n}) \qquad E_{1-1}=Q_o^{-pn}$$

In this form it can be seen that $\underset{\sim}{S}, \underset{\sim}{T}$ and $E_{\alpha\beta}$ form the subgroup O(6) within O(8), that $\underset{\sim}{S}, \underset{\sim}{T}$ commute, hence the physically meaningful decomposition $O(8) \supset O(6) \supset O(3) \otimes O(3)$. The quantum numbers of O(8) are the following. Subject to a given seniority v, there can be μ_1 neutrons with spin up, μ_2 with spin down, μ_3 protons with spin up, and μ_4 with spin down. The quantum numbers of O(8) are then the nucleon number N(or v) and linear combinations of the μ_i, as given by $Q_0, \ldots, E_{00}$.

We can then examine the O(6) representations contained in O(8). This, in general, is not unique, but if v is small, say 0,1 (near closed shell), the maximum degeneracy is two. This degeneracy can be handled by the properties of the wave functions under conjugation. Then, given an O(6) representation, we must find the values of $\underset{\sim}{S}, \underset{\sim}{T}$ which occur. Again, if v = 0 or 1, the maximum degeneracy is two. However, if v = 2, if we are away from a closed shell, the degeneracy is larger and other means must be found to construct wave functions. We address ourselves to this problem. We wish to ask the question --what are the $U(2) \otimes U(2)$ representations contained in U(4)? With what degeneracy do they occur? What are the explicit wave functions?

To answer these specific questions, let us first look at the more general problem of $U(nm) \supset U(n) \otimes U(m)$. We may construct a state of U(n) from the n dimensional vectors $\underset{\sim}{z}_i = (z_i^{(1)}, \ldots, z_i^{(n)})$ where i = 1,...,m for an m-rowed Young tableau of U(n). From these vectors, we may construct the differential operators acting on the Hilbert space

$$C^{\alpha\beta} = \sum_{i=1}^{m} z_i^{(\alpha)} \frac{\partial}{\partial z_i^{(\beta)}} \qquad \alpha, \beta = 1, \ldots, n$$

$$C_{ij} = \sum_{\alpha=1}^{n} z_i^{(\alpha)} \frac{\partial}{\partial z_j^{(\alpha)}} \qquad i, j = 1, \ldots, m$$

The n^2 operators $C^{\alpha\beta}$ are the generators of U(n), and commute with the operators C_{ij}. The generators satisfy the commutation relations

$$[C^{\alpha\beta}, C^{\alpha'\beta'}] = \delta_{\alpha'\beta} C^{\alpha\beta'} - \delta_{\alpha\beta'} C^{\alpha'\beta}$$

To construct a state $|f,\alpha\rangle$ associated with an irreducible representation of U(n), we construct Casimir operators C^p, p = 1,...,m,...,n

$$C^p = C^{\mu_1\mu_2} C^{\mu_2\mu_3} \ldots C^{\mu_p\mu_1}$$

and require the state $|f,\alpha\rangle$ to be diagonal

$$C^p|f,\alpha\rangle = \omega^p|f,\alpha\rangle$$

This is simply Schur's lemma. With an m-rowed Young tableau, the C^p for $p > m$ are dependent on the C^p with $p < m$.

On the other hand, since the operators C_{ij} commute with the generators $C^{\alpha\beta}$ and thus the Casimir operators C^p, we may label an irreducible representation with the operators C_{ij}. In fact, there is a close correspondence between the operators C_{ij} and a Young tableau. Let f_i be the number of boxes in the ith row of a Young tableau, $i = 1,\ldots,m$ where $m < n$, and let $\lambda_i = f_i - f_{i-1}$, $\lambda_m = f_m$ be the overhangs between the ith and (i-1)st rows. Then Moshinsky[4] has shown that a sufficient condition that a state $|f,\alpha\rangle$ be associated with an irreducible representation of U(n) is

$$C_{ii}|f,\alpha\rangle = f_i|f,\alpha\rangle$$

$$C_{ij}|f,\alpha\rangle = 0 \qquad i < j \qquad ij = 1,\ldots,m$$

The explicit wave function is

$$|f,\alpha\rangle = \sum_{h_{ij}} C_{h_{ij}} (z_1^{(\mu)})^{\lambda_1} (\Delta_{12}^{(\mu\nu)})^{\lambda_2} \ldots (\Delta_{12\ldots m}^{(\mu\lambda\ldots\gamma)})^{\lambda_m}$$

where $(z_1^{(\mu)})^{\lambda_1}$ is a shorthand notation for $(z_1^{(1)})^{h_{11}} (z_1^{(2)})^{h_{12}} \ldots (z_1^{(n)})^{h_{1n}} \equiv (z_1^{(\mu)})^{\lambda_1}$, $h_{11}+h_{12}+\ldots+h_{1n}=\lambda_1$, $\Delta_{12}^{(\mu\nu)}$ are the 2 x 2 anti-symmetric forms $(z_1^{(\mu)} z_2^{(\nu)} - z_1^{(\nu)} z_2^{(\mu)})$, Δ_{123} are 3 x 3 anti-symmetric forms, and so on. It is important to note that the operation $C_{ij}|\rangle = 0$ for $i < j$ orders the lower indices. To obtain the maximum state $|f,\alpha_M\rangle$, we require

$$C^{\alpha\beta}|f,\alpha_M\rangle = 0$$

The explicit state is

$$|f,\alpha_M\rangle = C\,(z_1^{(1)})^{\lambda_1} (\Delta_{12}^{(12)})^{\lambda_2} \ldots (\Delta_{12\ldots m}^{(12\ldots m)})^{\lambda_m}$$

Note that the upper indices are also now ordered. It is easy to show that this state is diagonal with respect to C^p. But now, what happens if we operate up or down with C_{ij}, i.e., <u>not</u> consider only the states

$C_{ij} \mid \rangle = 0$, $i < j$? Because of the commutativity, it will still be diagonal with respect to C^p, hence associated with an irreducible representation of U(n). The crucial point is, the state is also diagonal with respect to C_p,

$$C_p = C_{i_1 i_2} C_{i_2 i_3} \cdots C_{i_p i_1} \qquad \text{for } p = 1, \ldots, m$$

and will also be associated with an irreducible representation of U(m)! Thus, the operations with C_{ij} lower us from the maximum state to an arbitrary state of U(m), and the operators C_{ij} are generators of U(m). This freedom to operate with C_{ij} arises because the Moshinsky condition, while sufficient, is not necessary; the only necessary condition is that C_p, C^p be diagonal. In general, we may operate up and down with both $C_{ij}, C^{\alpha\beta}$, and obtain a state of U(m) $\otimes$ U(n) contained in a symmetric representation of the group U(mn). If we relabel indices

$$z_i^\alpha \equiv z^{(i-1)n+\alpha} = z^\mu \qquad \mu = 1, \ldots, mn$$

and the state vector is associated with the symmetric representation, the number of boxes $h = \lambda_1 + 2\lambda_2 + \ldots + m\lambda_m$ of the one row tableau equal to the degree of the polynomial in z^μ. Each representation of U(m) $\otimes$ U(n) occurs once in the symmetric representation of U(mn). However, we wish to consider more general representations than the symmetric representations of U(mn). Let us go to the specific case of U(4), and generalize the results to other than symmetric representations.

In U(2), the maximum state is

$$\left| \begin{matrix} j = m_2 \\ j\ \lambda \\ j = m_1 \end{matrix} \right\rangle = C (z_1^1)^{2j} (z_1^1 z_2^2 - z_2^1 z_1^2)^\lambda ,$$

$$J^o = \tfrac{1}{2} \left(z_1^1 \frac{\partial}{\partial z_1^1} - z_1^2 \frac{\partial}{\partial z_1^2} + z_2^1 \frac{\partial}{\partial z_2^1} - z_2^2 \frac{\partial}{\partial z_2^2} \right)$$

$$J^o \left| \begin{matrix} j \\ j\ \lambda \\ j \end{matrix} \right\rangle = j \mid \rangle \qquad .$$

We may lower with $C^- = C^{21}$ and obtain the state

$$\left| \begin{matrix} m_2 \\ j\ \lambda \\ j = m_1 \end{matrix} \right\rangle = C' (z_1^1)^{j-m_1} (z_1^2)^{j+m_1} (\Delta_{12}^{(12)})^{\lambda}$$

Then $J^o \mid \rangle = m_2 \mid \rangle$ and $\underset{\sim}{J}^2 \mid \rangle = j(j+1) \mid \rangle$. Similarly, we may operate down with $C_- = C_{21}$, and obtain

$$\left| \begin{matrix} j \\ j\ \lambda \\ m_1 \end{matrix} \right\rangle = C'' (z_1^1)^{j-m_2} (z_2^1)^{j+m_2} (\Delta_{12}^{(12)})^{\lambda}$$

where $J_o \mid \rangle = m_2 \mid \rangle$ and $\underset{\sim}{J}^2 \mid \rangle = j(j+1) \mid \rangle$. Or, we may operate down with both C^- and C_-. If we relabel indices $z_1^1 = z^1$, $z_1^2 = z^2$, $z_2^1 = z^3$ and $z_2^2 = z^4$, we obtain the symmetric representation, with one vector $\underset{\sim}{z} = (z^1, \ldots, z^4)$. Let us consider more than one row $\underset{\sim}{z}_i$, $i = 1, \ldots, 4$. The $U(4)$ operators in the $U(2) \otimes U(2)$ decomposition become

$$J_o = \tfrac{1}{2}(C^{11} + C^{33} - C^{22} - C^{44}) \equiv J_1^o \qquad J^o = \tfrac{1}{2}(C^{11} + C^{22} - C^{33} - C^{44}) \equiv J_2^o$$

$$C_+ = z_\alpha^1 \frac{\partial}{\partial z_\alpha^2} + z_\alpha^3 \frac{\partial}{\partial z_\alpha^4} \qquad C^+ = z_\alpha^1 \frac{\partial}{\partial z_\alpha^3} + z_\alpha^2 \frac{\partial}{\partial z_\alpha^4}$$

$$C_- = z_\alpha^2 \frac{\partial}{\partial z_\alpha^1} + z_\alpha^4 \frac{\partial}{\partial z_\alpha^3} \qquad C^- = z_\alpha^3 \frac{\partial}{\partial z_\alpha^1} + z_\alpha^4 \frac{\partial}{\partial z_\alpha^2}$$

where α is summed from 1 to 4 and C^{ii} are the generators of $U(4)$. The $U(4)$ conditions on the state vector are

$$C_{ii} \mid f, \alpha\rangle = f_i \mid f, \alpha\rangle \qquad i = 1, \ldots, 4$$

$$C_{ij} \mid f, \alpha\rangle = 0 \quad .$$

The $U(2) \otimes U(2)$ conditions are

$$C_+, C^+ \mid f, \alpha_M\rangle = 0$$

and

$$J_1^o (J_2^o) \mid f, \alpha_M\rangle = j_1 (j_2) \mid f, \alpha_M\rangle \quad .$$

If we apply these conditions to the state vector, we get the solution:

$$|\lambda;j_1\mu_1 j_1, j_2\mu_2 j_2\rangle = (\Delta_{1234}^{(1234)})^{\lambda_4} \sum_{n_i} A_{n_i} S_{n_i}$$

$$S_{n_i} = (z_i)^{n_1}(\Delta)^{n_2}(\Delta^{12})^{n_3}(\Delta^{13})^{n_4}(S_1)^{n_5}(S_2)^{n_6}$$

$$\times (\Delta_{123}^{123})^{n_7}(T_1)^{n_8}(T_2)^{n_9}(T_3)^{n_{10}}(T_4)^{n_{11}}(U_1)^{n_{12}}(U_2)^{n_{13}}$$

where

$$\Delta = z_1^1 z_1^4 - z_1^2 z_1^3 \quad , \quad S_1 = (\Delta^{23})^2 + (\Delta^{14})^2 - 2\Delta^{12}\Delta^{34} - 2\Delta^{13}\Delta^{24}$$

$$S_2 = z_1^1\Delta^{14} - z_1^1\Delta^{23} - 2z_1^3\Delta^{12} \quad , \quad T_1 = \Delta^{123}\Delta^{234} - \Delta^{124}\Delta^{134}$$

$$T_2 = z_1^1\Delta^{134} + z_1^3\Delta^{123} \quad , \quad T_3 = z_1^1\Delta^{124} - z_1^2\Delta^{123}$$

$$T_4 = \Delta^{14}\Delta^{123} + \Delta^{23}\Delta^{123} - 2\Delta^{13}\Delta^{124}$$

$$U_1 = z^1\Delta^{14}\Delta^{124} - z^1\Delta^{24}\Delta^{123} + z^2\Delta^{23}\Delta^{123} + z^1\Delta^{12}\Delta^{234} - z^2\Delta^{13}\Delta^{124}$$

$$U_2 = z^1\Delta^{14}\Delta^{134} + zz^1\Delta^{34}\Delta^{123} - z^2\Delta^{13}\Delta^{134} - z^3\Delta^{13}\Delta^{124} + z^3\Delta^{23}\Delta^{123}$$

The $U(4) \supset U(2) \otimes U(2)$ constraints on the exponents n_i are:

$$n_1+2n_3+n_6+n_7+2n_{10}+n_{11}+2n_{12} = 2j_1$$

$$n_1+2n_4+n_6+n_7+2n_9+n_{11}+2n_{13} = 2j_2$$

$$n_1+2n_2+n_6+n_9+n_{10}+n_{12}+n_{13} = \lambda_1$$

$$n_3+n_4+2n_5+n_6+n_{11}+n_{12}+n_{13} = \lambda_2$$

$$n_7+n_8+n_9+n_{10}+n_{11}+n_{12}+n_{13} = \lambda_3$$

Finally, each of the terms S_{n_i} is not independent because the following relations hold:

$$z^1\Delta^{23} + z^3\Delta^{12} = z^2\Delta^{13} \qquad \Delta^{14}\Delta^{123} + \Delta^{12}\Delta^{134} = \Delta^{13}\Delta^{124}$$

$$z^1\Delta^{24} + z^4\Delta^{12} = z^2\Delta^{14} \qquad \Delta^{24}\Delta^{123} + \Delta^{12}\Delta^{234} = \Delta^{23}\Delta^{124}$$

$$z^4\Delta^{23} + z^2\Delta^{34} = z^3\Delta^{24} \qquad \Delta^{12}\Delta^{34} + \Delta^{14}\Delta^{23} = \Delta^{13}\Delta^{24}$$

$$z^1\Delta^{234} - z^2\Delta^{134} + z^3\Delta^{124} - z^4\Delta^{123} = 0 \ .$$

We get the relations

$$(S_2)^2 = (z^1)^2 S_1 - 4\Delta\,\Delta^{12}\Delta^{13} \qquad (U_1)^2 = S_1 (T_3)^2 + 4\Delta T_1 (\Delta^{12})^2$$

$$(T_4)^2 = (\Delta^{123})^2 S_1 - 4\Delta^{12}\Delta^{13} T_1 \qquad (U_2)^2 = S_1 (T_2)^2 + 4\Delta T_1 (\Delta^{13})^2$$

and this implies the conditions $n_6, n_{11}, n_{12}, n_{13} = 0, 1$. Also,

$$S_2 T_4 = (\Delta^{13} U_1 + \Delta^{12} U_2), \quad U_1 U_2 = 4\Delta\Delta^{12}\Delta^{13} T_1 - ((z^1)^2 T_1 + (\Delta^{123})^2 \Delta) S_1$$

$$z^1 T_4 = \Delta^{13} T_3 - \Delta^{12} T_2 , \qquad \Delta^{123} S_2 = -(\Delta^{13} T_3 + \Delta^{12} T_2)$$

$$T_3 T_2 = (\Delta^{123})^2 \Delta - (z^1)^2 T_1 , \qquad z^1 \Delta^{123} S_1 = \Delta^{13} U_1 - \Delta^{12} U_2$$

$$T_2 U_1 = -(z^1 S_2 T_1 + \Delta^{123} T_4 \Delta)$$

$$\Delta^{123} U_2 = T_4 T_2 - 2 z^1 T_1 \Delta^{13} , \qquad T_4 U_2 = 2 S_2 T_1 \Delta^{13} + \Delta^{123} S_1 T_2$$

$$\Delta^{123} U_1 = -T_4 T_3 + 2 z^1 T_1 \Delta^{12} , \qquad T_4 U_1 = 2 S_2 T_1 \Delta^{12} - \Delta^{123} S_1 T_3$$

$$T_3 U_2 = -z^1 S_2 T_1 + \Delta^{123} T_4 \Delta$$

and this implies that whenever all the integers n_i in each set of integers below is non-zero, the product of polynomials may be re-expressed,

$$(n_6, n_{11}),\ (n_{12}, n_{13}),\ (n_1, n_{11}),\ (n_1, n_5, n_7)\ (n_{10}, n_{11}),\ (n_6, n_7)$$
$$(n_1, n_{12}),\ (n_6, n_{12}),\ (n_1, n_{13}),\ (n_6, n_{13}),\ (n_9, n_{12})$$
$$(n_2, n_{12}),\ (n_{11}, n_{12}),\ (n_2, n_{13}),\ (n_{11}, n_{13}),\ (n_{10}, n_{13})$$

These conditions completely specify the terms S_{n_i}, and provide an implicit expression of the representations j_1, j_2 contained in an irreducible representation of U(4) with the degeneracy with which they occur.

Of course, one could ask whether we have all the $U(2) \otimes U(2)$ representations contained in a representation of U(4). This is a counting operation, to show that

$$\sum g\,(2j_1+1)(2j_2+1) = N$$

where g is the degeneracy of a representation j_1, j_2 and N is the dimension of the U(4) representation. This summation is performed in a preprint,[5] where, incidentally, more details of the discussed research are included. Additional operators may be constructed which break this degeneracy--this is also included in the above preprint.

With the results of this calculation, it is now possible to return to the nuclear problem and solve for seniority number $v > 1$.

References

1. A. K. Kerman, Ann. Phys. 12, 300 (1961).
2. K. T. Hecht, Phys. Rev. 139, B794 (1965) and Nucl. Phys. 102, 11 (1967).
3. S. C. Pang, Univ. of Michigan technical report, November, 1967.
4. M. Moshinsky, J. Math. Phys. 4, 1123 (1963).
5. M. Resnikoff, "The Representations of $U(4) \supset U(2) \otimes U(2)$," SUNY at Buffalo preprint, June, 1967.

ON THE COVARIANCE AND ANALYTICITY OF SCATTERING AMPLITUDES†‡*

R. Seiler
University of Pittsburgh
Pittsburgh, Pennsylvania

Introduction

The S-matrix amplitudes are of outstanding importance in each theory for scattering processes. In various attempts, there have been derived covariance and analyticity properties from basic assumptions.[16] They are fundamental for the derivation of dispersion relations.

In a pure S-matrix approach[17),11),13)] the aim is to build up a theory starting from covariant and analytic amplitudes defined in a physical region (real domain in the mass shell). The discussion of crossing and CPT symmetries involves the problem of analytic extensions of real analytic functions into a complex domain, invariant under the action of complex Lorentz transformations with determinant one. The analysis of this problem is the main object of these lectures.

In the first lecture, I will speak on geometries in a linear vector space and the corresponding symmetry groups. The technique developed--due to E. Artin[1)]--provides the instrument for a topological classification of the symmetry groups as well as the little groups (Lemma 1).[2)]

The second lecture contains a brief survey of the representation theory of the real Lorentz group.[4)] Restrictions and extensions of spinor representations are considered (II.C).

In the last lecture, we start with an exposition of the CPT theorem in field theory[10)] (III.A). Only the parts necessary for the comparison with the analog problem in analytic S-matrix theory (III.B) are

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

‡This work was partially supported by ARPA (SD-67).

*Supported in part by the U. S. Atomic Energy Commission under Contract AT-30-1-3829.

mentioned. Finally, the analytic extensions of real analytic scattering amplitudes, especially the problem of uniqueness is analyzed (III.C).

In most of the work done in S-matrix theory, the amplitude is decomposed into a sum of standard covariants with analytic coefficients. We are not going into this problem,[18),19)] although it is intimately connected to the question of analytic extensions of real analytic covariant functions.

I. Geometries and Isometries[1)]

A. Notations and Basic Facts

1. Let V be an n dimensional vectorspace over the field k of the real or complex numbers and (.,.) a bilinear mapping of the set $V \otimes V$ of all pairs $[x,y]$, $x \in V$, $y \in V$, into k

$$\begin{aligned} (x+y,z) &= (x,z) + (y,z) \\ (x,y+z) &= (x,y) + (x,z) \end{aligned} \qquad x \in V,\ y \in V$$

$$(\alpha x, y) = \alpha (x,y) \qquad \alpha \in k$$

$$(x, \alpha y) = \alpha (x,y)$$

Such a bilinear mapping is called a metrical structure over V. A basis in V associates to a metrical structure over V a metrical tensor g. A metrical structure is nonsingular if $(x,y) = 0$ for all $y \in V$ implies $x = 0$. This is equivalent to $\det g = 0$.

2. An element $x \in V$ is orthogonal to $y \in V$,

$$x \perp y$$

if $(x,y) = 0$. There is the following theorem which gives an interesting classification of a wide class of metrical structures:

Theorem 1: If the orthogonality relation is reflexive, i.e. if $x \perp y$ implies $y \perp x$, the metrical structure is either pseudo-orthogonal (symmetric) or symplectic (antisymmetric)

$$(u,v) = \pm (v,u) \qquad u \in V,\ v \in V\ . \tag{I.1}$$

Since this theorem is not very well known, I will give you a simplified version of Artin's proof:[1)]

a. Start from the identity

$$(x,(x,y)z) - (x,(x,z)y) = 0$$

By assumption, we get the main equation

$$(x,y)(z,x) - (y,x)(x,z) = 0 \ . \qquad (I.2)$$

(I.2) provides for $x = z$ the implementation (the second relation by interchanging x and y):

$$(x,y) \neq (y,x) \rightarrow (x,x) = 0, \ (y,y) = 0 \qquad (I.3)$$

b. Suppose there exists elements $x,y \in V$ such that

$$(x,y) \neq (y,x) \qquad (I.4)$$

We show that for all $z \in V$, $(z,z) = 0$. By (I.3) it is sufficient to assume $(x,z) = (z,x)$. But this is only compatible with (I.2), (I.4) if $(x,z) = (z,x) = 0$ and $(y,z) = (z,y) = 0$ (the later relation again by interchanging x and y). But since $((x+z),y) = (x,y) \neq (y,x) = (y,(x+z))$ we get from (I.3) $(x+z,x+z) = 0 = (z,z)$.

c. Either for all $x,y \in V$, $(x,y) = (y,x)$ or $(x+y,x+y) = (x,y) + (cy,x) = 0$.

3. The metrical structure allows the definition of an orthogonal complement $U^{\perp}$ to a subspace $U \subset V$

$$U^{\perp} = \{x \mid (x,y) = 0, \ y \in U\}$$

and a radical

$$\text{rad } U = U \cap U^{\perp} \ .$$

Radicals contain only light vectors

$$x \in \text{rad } U \rightarrow (x,x) = 0 \ .$$

We already mentioned that a vectorspace V with rad $V = 0$ is called nonsingular.
A hyperbolic plane H is a two-dimensional subspace in V spanned by the vectors p and q such that $(p,p) = 0 = (q,q)$, $(p,q) = 1$.
A straight line is a one-dimensional subspace of V spanned by the element $g \in V$, $\langle g \rangle$ $(g,g) \neq 0$.
The subspace $U \subset V$ is the orthogonal sum of the subspaces $R \subset V$, $S \subset V$

$$U = R \perp S$$

if U is the direct sum of R and S and if R is orthogonal to S. The following theorem gives the standard decomposition of a subspace $U \subset V$ into a nonsingular part W and a purely singular one rad U. The nonsingular part can be further decomposed into standard types, namely into straight lines in the case of a pseudoorthogonal metrical structure or into hyperbolic planes in the case of a symplectic metrical structure.

Theorem 2: Each subspace $U \subset V$ of a nonsingular vectorspace[1] V has the decomposition

$$U = W \perp \operatorname{rad} U.$$

If the metrical structure is pseudoorthogonal U can again be decomposed into straight lines

$$U = \langle g_1 \rangle \perp \ldots \perp \langle g_n \rangle \quad .$$

If the metrical structure is symplectic, U is the direct sum of hyperbolic planes $H_i = \langle p_i, q_i \rangle$

$$U = H_1 \perp \ldots \perp H_n \ .$$

4. Let V and W be two linear vectorspaces over k with metrical structures. And let σ be an isomorphism of V onto V such that

$$(x,y)_V = (\sigma x, \sigma y)_W \ , \ x,y \in V.$$

Then σ is called an isometry. For $V = W$ the isometries form a group $L(n,k)$, $n = \dim V$. The norm $|A| = \sum_{i,k} |A_{i,k}|$, $A \in L$, provides a topology on $L(n,k)$[2]. Such a group provides a decomposition of

$$V^m = \underbrace{V \oplus \ldots \oplus V}_{m}$$

into equivalents classes $V^m(k)$:

$$k \sim k' (L(n,k)) \leftrightarrow k' = \Lambda k, \ \Lambda \in L(n,k) \ .$$

The topological structures of the isometry groups $L(n,k)$ and the little groups

$$L(n,k;\underline{W}) = \{\Lambda \in L^1(n,k) \,|\, \Lambda \underline{W} = \underline{W}, \ \underline{W} = (W_1 \ldots W_k)\}$$

($L^1(n,k)$ denotes the one component of $L(n,k)$) are characterized in the following:

Lemma 1[2]: a. The group $L(n,k)$, $k = R$ or C, is connected[3] if the metrical structure is symplectic.

b. $L(n,k)$ has two connected components ($N = L(n,k)/L^1(n,k)$, $N = Z_2$[4]) if the metrical structure is pseudo-orthogonal and either $k = C$ or there is now light vector in V.

c. $L(n,k)$ has four connected components ($N = Z_2 \otimes Z_2$[5]) if the metrical structure is pseudo-orthogonal $k = R$ and there is at least one light vector in V.

d. The little group $L(n,k;\underline{W})$ is connected.

Remarks: a. The Lorentz group has already the most complicated topological structure of all the pseudo-orthogonal and symplectic groups.

b. We will use the notation $L^1(n,k)$ and $L_+(n,k)$ for the one components of the isometry groups $L(n,k)$.

c. A part of the proof will be given in the next section (I.B).

B. Examples

1. Consider the four-dimensional vectorspace V over the field of the real numbers R. Let the metrical tensor g in the basis e_0, e_1, e_2, e_3 be

$$g = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}$$

This is an orthogonal metrical structure and, since $\det g \neq 0$, non-singular. V is the orthogonal sum of the four straight lines $\langle e_i \rangle$, $i = 0,1,2,3$

$$V = \langle e_0 \rangle \perp \ldots \perp \langle e_3 \rangle$$

but it can also be written as the orthogonal sum of a hyperbolic plane

$$H = \left\langle \frac{1}{\sqrt{2}} (e_0 + e_1), \frac{1}{\sqrt{2}} (e_0 - e_1) \right\rangle$$

and the subspace

$$U = \langle e_2 \rangle \perp \langle e_3 \rangle$$

$$V = H \perp U. \tag{I.5}$$

Consider the subspace $W = \langle e_0 + e_1, e_2, e_3 \rangle$. It is easy to see that

$$\text{rad } W = \langle e_0 + e_1 \rangle$$

and

$$W = U \perp \text{rad } W \quad .$$

The group of this metrical structure is the Lorentz group $L(4,R)$. There is an important subgroup N of all elements reducing according to (I.5), which is the direct sum of the two-dimensional Lorentz group $L(2,R)$ and the two-dimensional rotation group $O(2,R)$,

$$N = L(2,R) \oplus O(2,R).$$

If we use the basis vectors $\frac{1}{\sqrt{2}}(e_0 \pm e_1)$ in H, the group $L(2,R)$ is realized by the matrix group

$$\left\{ \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{\frac{1}{2}(1-\alpha)} \Bigg| \; \lambda \text{ real}, \; \alpha = \pm 1 \right\} \quad .$$

It is obvious that $L(2,R)$ has four components of connectivity characterized by $(\alpha, \operatorname{sign}\lambda)$.[6]

The elements of $O(2,R)$ are rotations in U and realized by the matrices

$$\left\{ \begin{pmatrix} \cos g & \sin g \\ -\sin g & \cos g \end{pmatrix}, \; \begin{pmatrix} \sin g & \cos g \\ \cos g & -\sin g \end{pmatrix} \Bigg| \; -\pi < g \leq \pi \right\}$$

This group has two components of connectivity.

The number of the components of connectivity of $L(4,R)$ can be found by the following argument. The determinant and sign of Λ_{00} are continuous functions on the group $L(4,R)$. The identity 1, the space inversion I_s, the time inversion I_t and the space time inversion $I_{st} = I_s I_t$ are, therefore, in different components of connectivity. Since for each $\Lambda \in L(4,R)$ with $\det\Lambda = 1, \Lambda_{00} > 0$ there is a frame of coordinates such that Λ has one of the two normal forms[4]

$$\begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & 1/\lambda & 0 & 0 \\ 0 & 0 & \cos g & \sin g \\ 0 & 0 & -\sin g & \cos g \end{pmatrix} \qquad \begin{pmatrix} 1+\frac{1}{2}\tau^2 & -\frac{1}{2}\tau^2 & -\tau & 0 \\ \frac{1}{2}\tau^2 & 1-\frac{1}{2}\tau^2 & -\tau & 0 \\ -\tau & \tau & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{matrix} \lambda > 0 \\ \\ \tau \neq 0, \text{ real} \end{matrix}$$

Λ is in the one component $L_+^\uparrow$ of $L(4,R)$.

Each element of $L(4,R)$ can be multiplied by $1, I_s, I_t, I_{st}$ such that it is in the one component of $L(4,R)$. Therefore, the number of components of connectivity is four.

2. Consider now the two-dimensional vectorspace V over the complex numbers. Let the metrical tensor ε be in the basis e_1, e_2

$$(\varepsilon^{ik}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad .$$

This is a symplectic structure and since $\det\varepsilon \neq 0$ nonsingular. V is a hyperbolic plane since

$$(e_1, e_1) = (e_2, e_2) = 0$$

$$(e_1, e_2) = 1 \quad .$$

The group $Sp(2,C)$ of this metrical structure is isomorphic to the group of matrices with determinant 1 as it is easily seen by the following argument. Let A be an element of $Sp(2,C)$. A is uniquely determined by the pictures e_1' and e_2' of the basis vectors

$$e_1' = Ae_1$$

$$e_2' = Ae_2 \quad . \tag{I.7}$$

A simple calculation shows that $\det A = 1$, since A is isometric. A very useful tool in handling symplectic groups is the

Theorem 3[1]: Each symplectic transformation $A \in Sp(2,C)$ is a product of symplectic transvections $\tau_\lambda(x)$, $x \in V$. Definition

$$\tau_\lambda(x)\cdot y = y + \lambda^{-1}(y,x)x, \quad y \in V, \ \lambda \in C, \ \lambda \neq 0 \ .$$

Proof: a. From the definition follows that $\tau_\lambda(x) \in Sp(2,C)$.

b. Again, an arbitrary transformation $A \in Sp(2,C)$ is uniquely defined by (I.7). The transformation $\tau_\lambda(e_1' - e_1)$, $\lambda = (e_1' e_1)$, maps e_1 on e_1'.

c. It remains to show that two hyperbolic planes

$$H = \langle m, n \rangle$$

$$H' = \langle m, n' \rangle$$

can be mapped one onto the other by a product of symplectic transvections such that

$$m \to m,\ n \to n' \quad .$$

i. Let $(n,n) \neq 0$. Then $\tau_\mu(n'-n)$, $\mu = (n',n)$, maps n on n' and has m as a fixed point.

ii. $(n,n') = 0$. This can only be true if $n = n'$.

It is an immediately consequence of the above lemma that the group $Sp(2,C)$ or equally $SL(2,C)$ has only one component of connectivity, and that $SL(2,C)$ is even simply connected.

II. Covariant S-Matrix Amplitudes

In this section, we are going to review covariance properties of the S-matrix amplitudes in field theory. Such properties are basic elements of the Analytic S-matrix framework.[5]

Let us start from the following assumption: There exist two Fockspaces H_{in} and H_{out} of many particle scattering states on which a representation $U(a,\Lambda)$ of the inhomogeneous quantum mechanical Lorentz group i $SL(2,C)$ acts as in a free field theory. (For a discussion of this point in the general theory of quantized fields we refer to Ref. 6.) H_{ex}, ex ε {out, in} is spanned by products of free particle states $\psi_{\alpha_1..\alpha_n}(p_1..p_n)$ with mass m and spin s[7], momentum p and spin in the z-direction α. They are formal elements of H_{ex}.

The S-matrix is defined by

$$\left(\psi^{out}_{\alpha_1..\alpha_n}(p_1..p_n), S\,\psi^{out}_{\alpha_1'..\alpha_n'}(p_1'..p_n')\right) = \left(\psi^{out}_{\alpha_1..\alpha_n}(p_1..p_n),\ \psi^{in}_{\alpha_1'..\alpha_n'}(p_1'..p_n')\right)$$

which is the basic quantity for the calculation of transition probabilities.

The S-operator commutes with $U(a,\Lambda)$. The S-matrix elements are covariant functions of the momenta and spins though the explicit form of this covariance depends on the special choice of the basis in H_{out}. It will turn out that by choosing an especially appropriate basis, the S-matrix elements are tensor valued covariant functions. By a nonsingular transformation in the vector space of the tensor valued covariant functions, we will get the most simple covariant amplitudes.

In the first four subsections, we prepare the necessary material from the theory of the unitary continuous representations of the inhomogeneous quantum mechanical Lorentz group i $SL(2,C)$. The last one contains the amplitudes in the different bases.

A. The Spinor Representations of SL(2,C)[7]

Consider again the second example (I.B.2), namely the 2-dimensional vectorspace C^2 over the complex numbers with the metrical tensor

$$(\varepsilon^{ik}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

The simplest nontrivial representation of the group of isometries Sp(2,C) is

$$Sp(2,C)\ A \longrightarrow \bar{A} \in Sp(2,C)\ .$$

It is a standard result[8] that all finite dimensional representations of SL(2,C) are equivalent to a representation $D^{[r,s]}$ in the linear vectorspace of the tensors $\psi_{\alpha_1 \ldots \alpha_r \dot{\beta}_1 \ldots \dot{\beta}_s}$ which are symmetric in the dotted and undotted indices respectively and transform according to

$$\psi_{\alpha_1 \ldots \alpha_r \dot{\beta}_1 \ldots \dot{\beta}_s} \to A_{\alpha_1}{}^{\alpha_1'} \ldots A_{\alpha_r}{}^{\alpha_r'} \bar{A}_{\dot{\beta}_1}{}^{\dot{\beta}_1'} \ldots \bar{A}_{\dot{\beta}_s}{}^{\dot{\beta}_s'} \psi_{\alpha_1' \ldots \alpha_r' \dot{\beta}_1' \ldots \dot{\beta}_s'} \qquad \text{(II.1)}$$

Such a representation is called a spinor representation of the group SL(2,C).

The following remarks make the result plausible:

1. The representation $A \to \bar{A}$ is not equivalent to the identity. Consider the matrix

$$\begin{pmatrix} i & \sqrt{2} \\ -\sqrt{2} & i \end{pmatrix} \in SL(2,C)$$

which is mapped on

$$\begin{pmatrix} -i & \sqrt{2} \\ -\sqrt{2} & -i \end{pmatrix}$$

The corresponding characters are $2i$ and $-2i$.

2. If the tensor $\psi_{\alpha_1 \alpha_2}$ is not symmetric in α_1 and α_2, we can split ψ into a symmetric ψ^s and an antisymmetric ψ^a part. Since the transformation (II.1) preserves the symmetry of the tensor, the representation would be reducible.

B. The Connection with the Representations of L↑(4,R)

1. The Paulimatrices together with the identity

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

span a four-dimensional vectorspace over C,

$$V = \{\underset{\sim}{x} | \underset{\sim}{x} = \sum x^i \sigma_i \ , \ x^i \in C\} \tag{II.2}$$

V is isomorphic to C^4 by the mapping τ

$$\tau : x \to x^i = \tfrac{1}{2} \text{ Trace } \underset{\sim}{x}\, \sigma_i \qquad i = 0,1,2,3 \tag{II.3}$$

The Hermitian matrices are the real elements in V with respect to the involution

$$\underset{\sim}{x} = \sum x^i \sigma_i \longrightarrow \underset{\sim}{x}^* = \sum \bar{x}^i \sigma_i \quad .$$

Obviously, they form a subspace in V isomorphic to R^4 by the mapping τ.

The elements of SL(2,C) in V are characterized by

$$(x,x) = x^{0\,2} - \vec{x}^2 = 1$$

since

$$(x,x) = \det \underset{\sim}{x} \quad . \tag{II.4}$$

The subset of all unitary matrices in SL(2,C) are characterized by the reality condition

$$x^0 = \bar{x}^0$$

$$\vec{x} = - \vec{\bar{x}} \quad .$$

This is a consequence of the relation

$$A^{*-1} = \sigma_2 A \sigma_2 \qquad A \in SL(2,C) \tag{II.5}$$

The last equation is the algebraic expression for the geometrical fact, that SL(2,C) is the group of isometries of the two-dimensional linear vectorspace over C with the metrical structure

$$\epsilon = i\sigma_2$$

In general, one writes an equivalent formula

$$A^T \epsilon A = \epsilon \tag{II.6}$$

2. The transformation in V

$$\underset{\sim}{x}' = A \underset{\sim}{x} A^{*}, \qquad A \in SL(2,C) \tag{II.7}$$

induces via the isomorphism τ a real Lorentz transformation $\Lambda(A)$ in R^4. The mapping $A \to \Lambda(A)$ is a representation of $SL(2,C)$ by elements of $L(4,R)$ equivalent to $D^{[\frac{1}{2},\frac{1}{2}]}$ (II.1) and it can be shown to be a local homeomorphism. Since $SL(2,C)$ is simple connected,[18] it has to be the universal covering group of the one component $L_+^{\uparrow}(4,R)$ of $L(4,R)$.

Using the formalism of the spinor representation (II.7) reads

$$\underset{\sim}{x}'_{\alpha\dot{\beta}} = A_{\alpha}^{\ \alpha'} \bar{A}_{\dot{\beta}}^{\ \dot{\beta}'} \underset{\sim}{x}_{\alpha'\dot{\beta}'} \tag{II.7'}$$

By multiplying an arbitrary tensor with the metrical tensor ϵ and $\underset{\sim}{x}_{\alpha\dot{\beta}}$ the number of dotted and undotted indices can be changed at convenience. Of especial importance is the corresponding procedure in momentum space where one has to multiply by ϵ and

$$\begin{aligned} \underset{\sim}{p}_{\alpha\dot{\beta}} &= \frac{1}{m} (p^{o}\sigma_{o} + \vec{p}\vec{\sigma})_{\alpha\beta} \\ \underset{\sim}{\bar{p}}^{\dot{\alpha}\beta} &= \frac{1}{m} (p^{o}\sigma_{o} - \vec{p}\vec{\sigma})_{\alpha\beta} \end{aligned} \tag{II.8}$$

(Here $(p^{o}, \vec{p})$ is a vector in R^4 transforming covariant to $(x^{o}, \vec{x})$.) For the last expression, one usually writes

$$\underset{\sim}{p}^{\dot{\alpha}\beta} = \underset{\sim}{\bar{p}}^{\dot{\alpha}\beta} \quad .$$

C. Restrictions and Extensions of the Spinor Representation $D^{[r,s]}$

1. The irreducible representation $D^{[r,s]}$ of $SL(2,C)$ induces a reducible representation of the subgroup $SU(2)$, since the representation

$$A \to A \ , \ A \in SU(2)$$

is equivalent to the identity representation (II.5). The irreducible parts are equivalent to the standard representation D^{s} in the vector space of the totally symmetric tensors $\psi_{\alpha_1 \ldots \alpha_{2s}}$

$$\psi_{\alpha_1 \ldots \alpha_s} \longrightarrow A_{\alpha_1}^{\ \alpha'_1} \cdots A_{\alpha_{2s}}^{\ \alpha'_{2s}} \psi_{\alpha'_1 \ldots \alpha'_{2s}} \tag{II.9}$$

$D^{[r,s]}$ as a representation of SU(2) is unitary in the vectorspace of the tensors $\psi_{\alpha_1 \ldots \alpha_{2r}, \dot\beta_1 \ldots \dot\beta_{2s}}$ if one defines the scalar product

$$(\varphi,\psi) = \sum_{\alpha,\beta} \overline{\varphi_{\alpha_1 \ldots \alpha_{2r}\, \dot\beta_1 \ldots \dot\beta_{2s}}}\, \psi_{\alpha_1 \ldots \alpha_{2r}\, \dot\beta_1 \ldots \dot\beta_{2s}} \quad . \tag{II.10}$$

2. The representation $D^{[r,s]}$ of SL(2,C) has an analytic extension to a representation of SL(2,C) $\otimes$ SL(2,C) by the following procedure: The group SL(2,C) is isomorphic to the subgroup F in SL(2,C) $\otimes$ SL(2,C) consisting of the real elements with respect to the involution

$$A \otimes B \longrightarrow B^* \otimes A^* \quad .$$

The elements of F can also be characterized by

$$A \otimes B \in F \leftrightarrow A = \sum a^i \sigma_i \quad , \; B = \sum b^i \sigma_i \in SL(2,C)$$

$$\bar a = b$$

Consider the representation $D_C^{[r,s]}$ of SL(2,C) $\otimes$ SL(2,C) acting in the same vectorspace as the representation $D^{[r,s]}$ and defined by

$$\psi_{\alpha_1 \ldots \alpha_{2r}\, \dot\beta_1 \ldots \dot\beta_{2s}} \to A_{\alpha_1}^{\alpha_1'} \ldots A_{\alpha_{2r}}^{\alpha_{2r}'} B_{\dot\beta_1}^{\dot\beta_1'} \ldots B_{\dot\beta_{2s}}^{\dot\beta_{2s}'} \psi_{\alpha_1' \ldots \alpha_{2r}' \dot\beta_1' \ldots \dot\beta_{2s}'} \tag{II.11}$$

This representation, analytic in (a,b), is uniquely defined by the restriction to the subgroup F. This is a consequence of the following:

Lemma 2: Suppose $f(z,w)$ is analytic in a neighborhood of $z = w = 0$ and zero on the intersection of U with the plane $z = \bar w$. Then f vanishes identically in U.

Proof: The statement is obvious after the substitution

$$z' = \tfrac{1}{2}(z+w) \qquad w' = \frac{1}{2i}(z-w) \quad .$$

Remarks: a. Consider the special case of the representation of SL(2,C) by $L_+^\uparrow(4,R)$ or $D^{[\frac{1}{2}\frac{1}{2}]}$.

$$A = \sum_i a^i \sigma_i \to \Lambda^\mu{}_\nu = \sum_{i,k} a^{i-k}\, t_{ik}{}^\mu{}_\nu$$

$$2t_{i,k}{}^\mu{}_\nu = \operatorname{trace} \sigma_i \sigma_\nu \sigma_k \sigma_\mu \quad .$$

The extension is explicitly given by

$$A \otimes B \longrightarrow \Lambda^{\mu}{}_{\nu} = \sum_{i,k} a^{i} b^{k} t_{ik}{}^{\mu}{}_{\nu} \quad .$$

$SL(2,C) \otimes SL(2,C)$ is mapped on $L_{+}(4,C)$.

b. Because $\Lambda^{\mu}{}_{\nu}(a,\bar{a})$ is harmonic

$$\sum \frac{\delta^{2} \Lambda^{\mu}{}_{\nu}}{\delta a^{i} \delta \bar{a}^{-i}} = 0,$$

the uniqueness of the analytic extension is also a consequence of a general theorem.[8]

D. Results of the Theory of the Unitary Representations of the Quantum Mechanical Lorentz Group[4]

This part will be extensively covered in the lectures of J. G. Taylor. Therefore, I quote only the result we need later.

An irreducible unitary representation of $iSL(2,C)$ which allows a physical interpretation as a particle with mass $m > 0$ is unitary equivalent to a canonical representation $[m,s]$, $m > 0$, s halfinteger,

$$[m,s]: \ (a,A) \longrightarrow U(a,A)$$

where $U(a,A)$ is a unitary operator in the Hilbert space $H[m,s]$,

$$H[m,s] = \{ (f_{t}(p)) \,|\, t=-s,-s+1,..s, p \in R^{4}, \ f_{t}(p) \in L^{2}_{\mu}, \quad d\mu = \theta(p^{0}) \, \delta(p^{2}-m^{2}) d^{4}p \}$$

with the scalar product

$$(f,f) = \sum_{t} (f_{t}, f_{t})_{L^{2}_{\mu}}$$

The explicit form of $U(a,A)$ is given by

$$(U(a,A)f)_{t}(p) = e^{i(a,p)} \sum_{s} D_{t}{}^{s}(B^{-1}_{p \leftarrow k} \, A \, B_{q \leftarrow k}) \, f_{s}(\Lambda^{-1}(A)p)$$

$$k = (m,0)$$

$$p = \Lambda(A)q \ .$$

The $B_{p \leftarrow k}$ is an arbitrary element of $SL(2,C)$ with the only restriction

$$B_{p\leftarrow k} \underset{k}{\sim} B^*_{p\leftarrow k} = \underline{p} \tag{II.14}$$

D is the image of the rotation $B^{-1}_{p\leftarrow k} A B_{q\leftarrow k}$ in the irreducible representation D^s of SU(2) (C.1). It is convenient to use the following formal basis $|p,\alpha\rangle$

$$f = \sum_\alpha \int d\mu(p) |p,\alpha\rangle \quad f_\alpha(p)$$

the $|p,\alpha\rangle$ transform contragradiently to $f_\alpha(p)$ and are covariantly normalized

$$U(a,A)|p,\alpha\rangle = e^{i(a,\Lambda p)} |\Lambda,\beta\rangle \; D_\beta{}^\alpha (D^{-1}_{\Lambda p\leftarrow k} A B_{p\leftarrow k})$$

$$\langle p,\alpha|p',\alpha'\rangle = 2p_0 \; \delta(\vec{p}-\vec{p}') \; \delta_{\alpha\alpha'}, \tag{II.13'}$$

Since we are not going to prove the above statement, some remarks might be appropriate:

1. It is clear that the translations are represented unitarily. Since $B^{-1}_{\Lambda p\leftarrow k} AB_{p\leftarrow k}$ is an element of SU(2), $D(B^{-1}_{\Lambda p\leftarrow k} \Lambda B_{p\leftarrow k})$ makes sense and is a unitary matrix.

2. By (II.14) $B_{p\leftarrow k}$ is not uniquely defined. It is easy to see that each $B_{p\leftarrow k}$ can be multiplied from the right by an arbitrary element of SU(2) which corresponds to an ordinary rotation. This can be used to choose the direction of quantization.

3. Besides the canonical basis of the representation which is characterized by a rotation free boost and which implies

$$U(0,B_{p\leftarrow k})|k,\alpha\rangle = |p,\alpha\rangle$$

there are several others in common use.

The helicity basis uses a direction of quantization in the rest-frame along the momentum p. The relation to the canonical basis is given by

$$|p,\alpha\rangle^H = |p,s_3\rangle \; D_{s_3}{}^\alpha(U) \tag{II.15}$$

where U is a unitary matrix characterized by

$$U \, \underline{p} \, U^* = |\underline{p}| \sigma_3$$

The canonical and the helicity basis have two principal disadvantages. First, the covariance equation (II.13') contains a transformation

matrix D depending on p. This is very inconvenient (for instance, for the Fourier transformation). Second, $|p_1 s_3\rangle$ and $|p_1 \alpha\rangle^H$ do not transform as tensor quantities. These difficulties are overcome if one introduces the spinor bases[9]

$$|p\rangle^{\alpha} = |p,s_3\rangle\, D_{s_3}{}^{\alpha}(B^{-1}_{p\leftarrow k})$$

$$|p\rangle_{\dot\beta} = |p,s_3\rangle\, \bar D_{\beta}{}^{s_3}(B_{p\leftarrow k}) \qquad \text{(II.16)}$$

with the transformation law[8]

$$U(a,A)|p\rangle^{\alpha} = e^{i(a,\Lambda p)}|\Lambda p\rangle^{\delta}\, D_{\delta}{}^{\alpha}(A)$$

$$U(a,A)|p\rangle^{\dot\beta} = e^{i(a,\Lambda p)}|\Lambda p\rangle^{\dot\delta}\, \bar D_{\dot\delta}{}^{\dot\beta}(A) \qquad \text{(II.13'')}$$

Between the dotted and undotted bases holds the relation[9]

$$|p\rangle^{\alpha}\, \underset{\sim}{p}_{\alpha\dot\beta} = |p\rangle_{\dot\beta} \qquad \text{(II.17)}$$

The normalization of the spinor basis turns out to be

$${}^{\alpha}\langle p|p'\rangle^{\alpha} = \tilde p^{\dot\alpha\alpha'}\, 2po\, \delta(\vec p-\vec p')$$

$${}_{\beta}\langle p|p'\rangle_{\beta} = \underset{\sim}{\bar p}_{\dot\beta\beta},\, 2po\, \delta(\vec p-\vec p') \qquad \text{(II.18)}$$

E. The Amplitudes

Let us now consider the S-matrix elements for states built up by free particle wave functions or rather by products of the basis elements $|p,\alpha\rangle$ (to get a real S-matrix element, one would have to smear this quantity by wave functions $f_{\alpha}(p)$),

$$\langle k_1\ k_n|S|k_1'\ k_m'\rangle, k_i = (p_i,\alpha_i)\ .$$

If S commutes with $U(a,\Lambda)$ the matrix elements have to satisfy the relations

$$\langle k_1|U^{-1}(a,A)\, SU(a,)|..k_m\rangle = \langle k_1..|S|..k_m\rangle \qquad \text{(II.19)}$$

or explicitly in the formal canonical basis (II.13')[10]

$$\langle \underline{p},\underline{\alpha}|S|\underline{p}',\underline{\alpha}'\rangle = \exp -i\sum_{e}(a,\Lambda p_e)\exp\sum_{k}(a,\Lambda p_k)\langle \underline{\Lambda p},\underline{\beta}|\underline{\Lambda p}'\underline{\beta}'\rangle \bar{D}_{\beta}^{\alpha}(R)D_{\beta'}^{\alpha'}(R)$$

$$R = B^{-1}_{\Lambda p \leftarrow k}\, A\, B_{p\leftarrow k} \qquad \text{(II.19')}$$

The matrix elements are almost tensor valued functions, but not in the strict sense. The next step will be to change the formal basis in H_{ex} such that the matrix elements will be tensor valued functions. This is done by introducing the spinor basis (II.16). The equation (II.19) has now the form

$$\overset{\alpha}{\underline{p}}\langle \underline{p}|S|\underline{p}'\rangle^{\alpha'} = \exp -i\sum_{e}(a,\Lambda p_e)\exp i\sum_{k}(a,\Lambda p_k)^{\beta}\langle \underline{\Lambda p}|S|\underline{\Lambda p}'\rangle^{\beta'} D_{\beta}^{\alpha}(A)\bar{D}_{\dot\beta}^{\dot\alpha'}(A) \qquad \text{(II.19'')}$$

In the same manner transforms the amplitude $A_{\alpha_1 \ldots \dot\alpha_m}(p_1|p_m)$ defined by the equation (the δ-function absorbs the translational symmetry)

$$A_{\alpha_1 \ldots \dot\alpha_m}(p_1 .|.p_m)\,\delta(p_1+\ldots-p_m) = {}^{\alpha_1}\langle p_1 .|S-1|.p_m\rangle^{\dot\alpha_m} \qquad \text{(II.20)}$$

III. Realanalytic Covariant Functions and Their Continuation into the Complex Space[2)]

The problem of a possible analytic extension of an analytic covariant function and their uniqueness appears in the context of the CPT theorem in quantum field theory (Section A)[10)] as well as in analytic S-matrix theory (Section B).[11),5)] In both approaches, the existence of an analytic extension is trivial. The nontrivial part is to show uniqueness. It is one of the most striking features of quantum field theory that the uniqueness follows from the Wightman axioms alone whereas in analytic S-matrix theory one has to introduce a special and physically unmotivated assumption to get the desired result.

A. The CPT Operator in Quantum Field Theory

Consider the two asymplotic free fields $A^{ex}_{\alpha}(x)$ (connected by an interpolating field $A_{\alpha}(x)$)

$$A^{ex}_{\alpha}(x) = \left(\frac{1}{2\pi}\right)^{\frac{3}{2}} \int d\mu(p)\,(e^{-ipx}\, a^{ex}_{\alpha}(p) + e^{ipx}\, b^{ex}_{\dot\alpha}(p)^{+})$$

$$\text{with} \qquad [a^{ex}_{\alpha}(p),\, a^{ex}_{\beta}(q)^{+}] = 2p_o\, \underset{\sim}{p}_{\alpha\dot\beta}\, \delta(\vec{p}-\vec{q}) \qquad \text{(III.1a)}$$

$$\left[b_{\alpha}^{ex}(p),\ b_{\dot{\beta}}^{ex}(q)^{+} \right] = 2p_o\ \bar{\tilde{p}}_{\dot{\alpha}\beta}\ \delta(\vec{p}-\vec{q}) \tag{III.1b}$$

$$\left[a_{\alpha}^{ex}(p),\ b_{\beta}^{ex}(q)^{(+)} \right] = 0 \tag{III.1c}$$

$$a_{\alpha}(p)\,\Omega = b_{\dot{\alpha}}(p)\,\Omega = 0 \tag{III.2}$$

Both fields have the following transformation property

$$U(a,B)\,A_{\alpha}^{ex}(x)\,U^{-1}(a,B) = D_{\alpha}^{s\ \ \sigma}(B^{-1})\,A_{\sigma}^{ex}(\Lambda(B)x + a)$$

$$(a,B) \in iSL(2,C) \quad . \tag{III.3}$$

The commutation relations (III.1) together with (III.2) allow us to reconstruct two Fock spaces H_{ex}, isomorphic to Fock spaces over the one particle space

$$H = H[m,s] \oplus H[m,s] \quad .$$

$H[m,s]$ carries a unitary irreducible representation of $iSL(2,C)$. There are three CPT operators

$$\theta^{ex}\,A_{\alpha}^{ex}(x)\,\theta^{ex} = \eta\,A_{\alpha}^{ex}(-x)^{+}$$

$$\theta\ A_{\alpha}^{in}(x)\ \theta\ \ = \eta\,A_{\alpha}^{out}(-x)^{+} \qquad \eta = \begin{cases} 1 & 2s \text{ even} \\ i & 2s \text{ odd} \end{cases}.$$

They are connected to the S-matrix by the relation

$$\theta^{-1}\,\theta^{out} = S = \theta^{in}\,\theta^{-1} \quad .$$

The existence of the CPT symmetry has the following implication for the S-matrix elements (II.19″)

$$\langle a_{\alpha_1}^{out}(p_1)\cdot b_{\alpha_n}^{out}(p_n)\,S\,a_{\alpha_{n+1}}^{out}(p_{n+1})^{+}\cdot b_{\alpha_m}^{out}(p_m)^{+}\rangle$$

$$= \xi\,\langle a_{\alpha_m}^{out}(p_m)\cdot b_{\alpha_{n+1}}(p_{n+1})\,S\,a_{\alpha_n}^{out}(p_n)^{+}\cdot b_{\alpha_1}^{out}(p_1)^{+}\rangle \tag{III.4}$$

ξ is a phase factor depending on the number and the type of the particles. A similar relation holds for the amplitude A (II.20)

$$A_{\alpha_1\,\dot{\alpha}_m}(p_1\,..\,|\,..\,p_m) = A_{\alpha_m\,\dot{\alpha}_1}(\bar{p}_m\,..\,|\,..\,\bar{p}_1) \tag{III.5}$$

The main contribution to the existence proof of the CPT symmetry in Axiomatic Field Theory is one part of the theorem of Bargmann, Hall and Wightman which we state without proof:

Lemma 3 (1 of B.H.W.): Let $f(z_1 \ldots z_n)$ be a covariant and analytic function in the tube

$$T_n = \{ z = (z_1 \ldots z_n) \mid z_i = \xi_i - i\eta_i,\ \eta_i \in V_+,\ \xi_i \in R^4 \}$$

and $S(\Lambda)$ a tensor representation of $L_+^\uparrow(4,R)$:

$$f_\alpha(\Lambda z_1 \ldots \Lambda z_n) = S_\alpha^{\ \beta}(\Lambda)\, f_\beta(z_1 \ldots z_n),\ \Lambda \in L_+^\uparrow(4,R) \ .$$

Then there exists a unique analytique extension of f to the extended tube

$$T_n' = \bigcup_A A\, T_n$$

$$A \in L_+(4,C) \ .$$

The statement can easily be reduced to a lemma of V. Bargmann:[12] The set $B = \{A \in L_+(4,C) \mid A\, T_n \cap T_n \neq 0\}$ is connected (formulation of R. Jost).

B. The CPT Operator in Analytic S-Matrix Theory

The amplitudes (II.20) $A_{\alpha_1 \ldots \mid \ldots \dot\alpha_m}(p_1 \ldots \mid \ldots p_m)$, (p_i, α_i) momentum and spin of the ith particle or antiparticle, are supposed to be defined in a L_+-invariant domain D_R[11] in the real manifold

$$M_R = \{ (p_1 \ldots p_m) \mid p_i \in R^4,\ p_i^2 = m_i^2,\ \sum p_i = 0 \}$$

and real analytic[12] in D_R. (We assume $\prod_i m_i \neq 0$.)

Further, they transform according to a spinor representation $D^{[k,m-k]}$ of $SL(2,C)$

$$A_{\alpha_1 \dot\alpha_m}(p_1 . \mid . p_m) = \prod_{j,e} B_{\alpha_j}^{\ \alpha_j'}\, \bar B_{\dot\alpha_e}^{\ \dot\alpha_e'}\, A_{\alpha_1' \ldots \dot\alpha_m'}(\Lambda^{-1}(B)p_1 \ldots \Lambda^{-1}(B)p_m) \qquad \text{(III.7)}$$

A CPT operation defined in analogy to quantum field theory has to connect two amplitudes, by now unrelated

$$A_{\alpha_1 \ldots \dot\alpha_m}(p_1 \mid p_m) = \chi\, A^\theta_{\alpha_m \ldots \dot\alpha_1}(\bar p_m \mid \bar p_1),\quad \chi = (-1)^k \qquad \text{(III.8)}$$

($(\bar{p}_i, \alpha_i)$ momentum and spin of the antiparticle). With the "substitution rule"[13] (III.8) reads

$$A_{\alpha_1 \ldots \dot{\alpha}_m}(p_1 \mid p_m) = \chi\, A^{\theta}_{\alpha_1 \ldots \dot{\alpha}_m}(-p_1 \mid -p_m) \qquad \text{(III.9)}$$

By assumption[13] there shall be only one amplitude $\tilde{A}$ for a physical process and his CPT transformed and $\tilde{A}$ shall be analytic in a domain containing the physical region of the process and the CPT transformed one. For this assumption to make sense, one has to have a unique prescription how to get this universal amplitude $\tilde{A}$ from the physical amplitude A. The situation would be especially simple in the case of a unique analytic extension of the physical amplitude into the CPT transformed region, since the covariance property (III.7) assures the existence of an analytic extension: There is an analytically parametrized path $\omega(t)$ in SL(2,C) SL(2,C) such that $\Lambda(\omega(0)) = 1$ and $\Lambda(\omega(1)) = -1$.

In the remaining part of these lectures, we discuss the question of the uniqueness of analytic continuations for the amplitudes

$$A_{1 \ldots m}(p_1 \ldots \mid \ldots p_m) \quad .$$

C. The Analytic Extension

Let $A_\alpha(p)$ be a covariant analytic function defined in a real domain $D_R \subset M_R$. We are going to investigate three questions. Is there an analytic continuation of $A_\alpha(p)$ being $L_+(4,C)$ covariant on:

1. The set $L(4,C)\, D_R$?
2. A complex neighborhood of each point z in the domain of holomorphy[14] of A_α?
3. A complex neighborhood D of the real domain D_R?

1. The first question can almost trivially be answered by yes. There exist by the same arguments as in B an analytic continuation of A_α to $L_+(C)\, D_R$. It is sufficient to show $L_+(C)$-covariance in D_R. But this is easy to see since

$$\underline{p} \in D_R,\ \underline{p}' = \Lambda \underline{p} \in D_R,\ \Lambda \in L_+(4,C)$$

implies the existence[13] of $\Lambda' \in L_+$ such that

$$\underline{p}' = \Lambda' \underline{p} \quad .$$

Remark: A corresponding argument in the case of the Bargmann-Hall-Wightman theorem is impossible, since only the imaginary part of the points z and $z' = \Lambda z$ are known to be in the forward tube. It is just at this point where the lemma of Bargmann saves the situation.

2. The second question is answered by the following theorem, which is valid not only for the Lorentz group but for each isometry group of a pseudo-orthogonal or symplectic geometry (Section I). It is a generalization of a theorem by H. P. Stapp.[11] Because of the strong analyticity structure on M_R (Remark 12) it is enough to consider domains in $V^m(R^n)$.

Theorem 4: Let $A_\alpha(\underline{p})$ be a real analytic, tensor valued function defined in the real domain $D_R \in V^m(R^n)$ and let S be a tensor representation of a pseudo-orthogonal or symplectic group $L(n,R)$ such that for all elements Λ of the one component $L_+^\dagger(n,R)$ and $\underline{p} \in D_R$, $p' = \Lambda\underline{p} \in D_R$

$$A(\underline{p}) = S(\Lambda)\, A(\Lambda^{-1}\underline{p}) \ . \tag{III.10}$$

Then each point of the Riemann manifold[14] (R,π,D),[14] $D \subset V^m(C)$ is contained in a schlicht neighborhood U such that (III.10) holds for $\underline{w} \in U$, $\underline{w}' = \sum \underline{w} \in U$, $\sum \in L_+(n,C)$.[15]

We are not going to review the reduction of this theorem to lemma 1 and lemma 4--this was done by P. Minkowski and D. N. Williams[15]--but to develop some ideas for the proof in the following remarks:

a. $A_\alpha(\Lambda(\lambda_1 \ldots \lambda_\alpha)\underline{p})$ is a real analytic function of p for $p \in D_R$ and real analytic in the parameters λ_i of the Lie group $L(n,R)$. Since A is continuous in all variables together the theorem of Hartog's[8] assures the analyticity in all variables together. Each point $(\lambda = 0, \underline{p})$, $\underline{p} \in D_R$, is contained in a complex schlicht neighborhood such that (III.10) holds.

b. The same holds for each point $(\lambda = 0, \underline{p})$, where $\underline{p}$ is the projection of a point $r \in R$,

$$\pi(r) = \underline{p}$$

and the neighborhood intersected with the plane $\lambda = 0$ contains r.

c. This is still not enough for the proof because of the following geometrical property of the vectorspace V, typical for geometries with light vectors.[16] There is possibly a point $\underline{w} \in D_R$ such that for each fixed positive number M, there exists in each neighborhood of $\underline{w}$ a pair of points $\underline{z}$, $\underline{z}' = \Lambda\underline{z}$, $\Lambda \in L_+(n,C)$, such that

$$\min_{\omega \in P} \ \max_{0 \leq t \leq 1} |\omega(t) - w| > M$$

where P is the set of continuous paths

$$P = \{\omega \mid \omega(t) = \Lambda(t)\underline{z},\ \omega(1) = \underline{z}',\ \omega(0) = \underline{z},\ \Lambda(t) \in L_+(n,C)$$

d. The counter statement of c for points $\underline{w}$ with the "full dimensionality" is a consequence of the following:

Lemma 4: Let the dimension d of the space

$$\langle \underline{w} \rangle = \langle w_1 .. w_m \rangle \subset V$$

be maximal, i.e. $d = \min(m,n)$.

Then there exist for each neighborhood N of the identity in $L(n,C)$ a neighborhood U of $\underline{w}$ such that for $\underline{k} \in U$, $\underline{k}' = \Lambda\underline{k} \in U$, $\Lambda \in L_+(n,C)$ there is a $\sum \in N$ such that $\underline{k}' = \sum \underline{k}$.

For the purely algebraic proof of lemma 4 and the full proof of theorem 2, we refer to Ref. 2.

3. Before we go into the discussion of the last question, we have to analyze a particular geometrical property of real domains in $V^m(R^n)$. To do this, we need the following two

Definitions: Two elements $(\underline{k},\underline{k}')$ in $D_R \subset V^m(R^n)$ are called limit points if there exists two sequences in $V^m(C^n)$ $\{\underline{k}_i\}$, $\{\underline{k}_i'\}$ converging to $(\underline{k},\underline{k}')$ (in the Eucledian norm) and $\underline{k}_i \sim \underline{k}_i'$ $(L_+(n,C))$. For each $\underline{k} \in V^m(C)$ exists a normal form $\underline{k}T$ (Theorem 2)

$$\underline{k}T = (k_i T_1^i \ldots k_i T_m^i) = (e_1 .. e_s,\ p_i .. p_t,\ 0..0)$$

where T is a nonsingular linear transformation and

$$\text{rad}\ \langle k \rangle = \langle p_i .. p_t \rangle,\ \langle \underline{k} \rangle = \text{rad}\ \langle k \rangle \perp \langle e_1 .. e_s \rangle \quad .$$

For the limit points there is still another equivalent definition depending on the geometry in V:

Lemma 5: The points $k \in D_R$, $k' \in D_R$ are limit points, if

a. The geometry is symplectic and

$$(k_r, k_s) = (k_r', k_s') \quad r,s = 1..m \quad .$$

b. The geometry is pseudo-orthogonal and either

$$\dim \langle \underline{k} \rangle \neq \dim \langle \underline{k}' \rangle, \quad \text{or}$$

$$\dim \langle k \rangle = \dim \langle k' \rangle \ \langle n - 2 ,$$

$$\text{or } \dim \langle \underline{k} \rangle = \dim \langle \underline{k}' \rangle = n-1 \quad \text{and } \underline{k} \sim \underline{k}' \ (L(n,C)),$$

or dim $\langle \underline{k} \rangle$ = dim $\langle \underline{k}' \rangle$ = n-1, $\underline{k} \not\sim \underline{k}'$ ($L(n,C)$) and

$$\det \sum \cdot \Lambda = +1$$ (for the definition of $\sum$ and Λ see below).

or dim $\langle \underline{k} \rangle$ = dim $\langle \underline{k}' \rangle$ = dim $\langle \underline{k}' \rangle$ = n, $\underline{k} \sim \underline{k}'$ ($L_+(n,C)$).

Now, I have to give you the definition of $\sum$ and Λ. There is a non-singular matrix T transforming $\underline{k}$ and $\underline{k}'$ into a normal form (II.11):

$$\underline{k}T = (e_1 \ldots e_{n-2}),\ p,\ 0)$$

$$\underline{k}'T = (e_i' \ldots e_{n-2}'),\ 0,\ p')$$

Further, there are two elements q, q' in V such that

$$(q,q) = 0 \qquad (p,q) = 1 \qquad (q',q') = 0 \qquad (p',q') = 1 \ .$$

$\sum$ and Λ are fixed by the requirements

$$e_i' = \sum e_i \qquad e_i' = \Lambda e_i' \qquad i = 1 \ldots n-2 \ .$$

$$p' = \sum p \qquad q' = \Lambda p'$$

One can show that although $\sum$ and Λ are not uniquely defined det $\sum \Lambda$ is unique. I will not give the lengthy proof of lemma 5.[2] Now we are ready to characterize the domains $D_R \subset V^m(R^n)$ allowing a positive answer to question three (for an arbitrary $A_\alpha(p)$):

Theorem 5 (H. P. Stapp): Suppose the assumptions of theorem 2 together with the geometrical properties of D_R hold:

1. $D_R = L_+(n,R)\, D_R'$, D_R' is a bounded domain in $V^m(R^n)$.
2. Limit points in D_R' are L‡(n,R)-equivalent.

Then there exists a schlicht, complex, $L_+(n,C)$-invariant domain $U \subset D_R$ with an analytic extension of $A_\alpha(p)$ onto U such that (III.10) holds for $\Lambda \in L_+(n,C)$, $p \in U$.

Proof: 1. There certainly exists a schlicht complex domain $U_o' \subset D_R'$.

2. If $L_+(n,C)U_o'$ does not already have the desired property, there are points (p_o, p_o') in U_o, $L_+(n,C)$-equivalent, not having property (III.10).

3. Consider $U_{i+1}' \subset U_i' \subset U_o'$, i = 1,2,... . Either there is a $U_i = L_+(n,C)\, U_i'$ with the properties claimed in theorem 5 or not. The

latter is inconsistent with the second assumption.[2)]

The following example[17] gives an illustration of the necessity of the second assumption. Consider the vectorspace V(I.B.2)

$$V = \langle p,q \rangle \qquad (p,p) = 0, \quad (q,q) = 0, \quad (p,q) = (q,p) = 1 \quad .$$

The function

$$f(k,k') = \alpha \sqrt{\beta^2 + \beta'^2}$$

is analytic in

$$D_R = \{ (k,k') \mid k = \alpha p + \beta q, \; k' = \alpha' p + \beta' q, \; \alpha,\beta,\alpha',\beta' \text{ real}, \beta^2 + \beta'^2 \neq 0 \} \; .$$

Yet, f is not even $L_+(2,R)$ invariant in D_R since

$$f(k,k') = -f(-k,-k') \quad .$$

Remark: For two pseudo-orthogonal geometries, lemma 5 provides a simpler formulation of the assumptions in theorem 5 (for a complete discussion, see Ref. 2).

1. If rad V = 0 assumption 2 is superfluous.
2. If dim rad V = 1 (and only in this case) assumption 2 can be replaced by the following: two $L_+(n,R)$-equivalent points in D_R are $L\uparrow(n,R)$ equivalent.
3. Theorem 5 is a generalization of theorem 4.

In conclusion of this section, let me apply the general result (Theorem 5) to the following physically motivated,[11)] situation: Let $A_\alpha(p)$ be a real analytic covariant function defined in the schlicht domain D_R contained in the mass shell (now massless particles),

$$A_\alpha(p) = S_\alpha{}^\beta(\Lambda) \, A_\beta(\Lambda^{-1} p), \; \Lambda \in L\uparrow(4,R) \; .$$

Let further D_R have the geometrical property,

$$D_R = U(\Lambda) D'_R \qquad , \quad D'_R \text{ bounded.}$$

$$\Lambda \in L_+(4,R)$$

Then there exists a schlicht complex neighborhood D

$$D_R \subset D$$

which is $L_+(4,C)$ invariant, such that $A_\alpha(p)$ is $L_+(4,C)$ covariant on D.

Remarks

1] Recall V is a vectorspace over the field C or R.
2] For convenience we identify the abstract group and the matrix group.
3] Path connected.[3)]
4] Cyclic group of order 2.
5] Group of Klein.
6] $\{(\alpha, \mathrm{sign}\lambda)\}$ is a representation of $L(2,R)$ if one defines the multiplication $(\alpha, \mathrm{sign}\lambda)(\alpha', \mathrm{sign}\lambda') = (\alpha\alpha', \mathrm{sign}\lambda\lambda')$ and a realization of $Z_2 \otimes Z_2$.
7] We neglect the possiblity of different masses.
8] This follows from (II.12″) and (II.16).
9] $p_{\alpha\beta}$ is the straightforward generalization of the matrix defined by Eq. (II.8).
10] $\Lambda p = (\Lambda p_1 \ldots, \Lambda p_m)$.
11] Connected, open set, not the vector part of Λp.
12] We use real analytic in the sense, A_α is the trace of an analytic function in a complex neighborhood $U(p) \subset V^m(C^4)$ of each point $p \in D_R$.
13] Recall D_R is a connected open set and $\Pi m_i \neq 0$.
14] R manifold over D, projection π.
15] Since U is schlicht, we identify points in D with the corresponding points in U.
16] This was pointed out by R. Jost.[15)]
17] Private communication of P. Minkowski.
18] Compare with the remark following Theorem 3.

References

1. E. Artin, Geometrical Algebra (Wiley, Interscience Tracts 1957).
2. R. Seiler, Helv. Phys. Acta 39, 641 (1966).
3. L. Pontrjagin, Topological Groups (Princeton University Press, Princeton, N. J., 1946).
4. E. Wigner, Ann. of Math. 40, 149 (1939).
5. R. J. Eden, P. V. Landskoff, D. I. Olive, J. C. Polkinghorne, The Analytic S-Matrix (Cambridge 1966).
6. K. Hepp in Axiomatic Field Theory, Brandeis University 1965.
7. B. L. van derWaerden, Die Gruppentheoretische Methode in der Quanten mechanik (J-Springer, Berlin 1932).
8. H. Cartan, Theorie elementaire des functions analytiques; Edition Scientifique, Hermann, 1961.
9. H. Joos, Fortschr. Phys. 10, 65, 1962.
10. R. Jost, Helv. Phys. Acta 30, 409 (1957).
11. H. P. Stapp, Lawrence Radiation Laboratory Report UCRL-10843.
12. R. Jost, General Theory of Quantized Fields, AMS 1965.

13. G. Chew, Analytic S-Matrix Theory (Benjamin, 1961)
14. B. A. Fuks, Analytic Functions of Several Complex Variables, AMS.
15. P. Minkowski, D. N. Williams and R. Seiler in Lectures in Theoretical Physics (Univ. of Colo. Press, Boulder, 1964).
16. H. Epstein in Axiomatic Field Theory, Brandeis University 1965 (Gordon and Breach).
17. W. Heisenberg, Z. Physik 120, 513 (1943).
18. K. Hepp, Helv. Phys. Acta 37, 55 (1964).
19. D. N. Williams, UCRL-11113 preprint.

SPACE-TIME PROPERTIES AND PHYSICAL REGION ANALYTICITY IN S-MATRIX THEORY†‡

Daniel Iagolnitzer*
Miller Institute for Basic Research in Science
and
Department of Physics
University of California
Berkeley, California

I. Introduction

Space-time causality properties are intimately connected to momentum space analyticity properties. This well-known fact is the basis of dispersion theory, and of certain forms of field theory. The purpose of this paper is to describe some of the recent work on one particular aspect of this connection, namely the connection between macroscopic space-time causality properties and physical-region analyticity properties of the S-matrix. This aspect has special significance because it deals with the quantities that are most closely related to actual physical measurements: it does not, on the one hand, refer to conceptions of causality that attribute a fundamental physical significance to the idealization of the space-time point; and it does not, on the other hand, refer to extensions of the S-matrix, either off the mass shell or away from the physical region.

The connection between macroscopic space-time properties and physical region analyticity properties is now well understood, as a result of a number of independent studies. The aim of this article is to review these works together with the closely connected works on space-time description in S-matrix theory.

A. Physical-Region Analytic Structure

In Sec. II we give a precise description[1] of some analyticity properties that the physical-region S-matrix is believed to enjoy.

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

‡A large part of this work has been done in collaboration with Dr. H. P. Stapp.

*Permanent address: Service de Physique Theorique, CEN SACLAY BP No. 2 Gif-sur-Yvette.

These properties were originally suggested by perturbation theory. In brief, they are that the S-matrix decomposes into a sum of connected parts each having only one energy momentum conservation law delta function (cluster decomposition property), that each connected part divided by this delta function--called a scattering function--is furthermore analytic except on positive α Landau surfaces, and that near these surfaces they are "plus $i\epsilon$" limits of certain analytic functions (plus $i\epsilon$ rule).

These three properties constitute what we shall call the normal analytic structure.*

A further important property is the rule that gives the discontinuity around a physical-region singularity surface as a function of scattering functions (Cutkovsky-type formulas).

B. Macroscopic Causality and Normal Analytic Structure

In Sec. III an equivalence is established between the normal analytic structure and a macroscopic space-time condition, called macrocausality.[2),3)]

In Sec. IIIA a weak concept of macroscopic space-time localization is introduced; it makes precise the notion that, for appropriate forms of the momentum space wavefunctions, particles are asymptotically localized in the direction of their energy momentum.

In Sec. IIIB, space-time displacements of the particles involved in a process are considered and macrocausality is formulated, by means of a macroscopic limit involving space-time dilation, in a way that refers only to this asymptotic notion of localization and such that in a dilated coordinate system a classical description becomes relevant.

The condition expresses the requirement that transfers of energy momentum not ascribable to (stable) physical particles be damped exponentially under dilation. Specifically, it says this: suppose the initial plus final particles of a process can be separated in groups k such that in the macroscopic limit there is no classical multiple scattering process connecting the initial and final particles except possibly for those that involve connections within the various groups k; there is no connection between the different groups k. Then a factorization of the transition probability W into a product of the partial transition probabilities W_k should hold: namely the remainder $R = W - \prod_k W_k$ should fall off in the macroscopic limit. Moreover, the rate of fall off should be that suggested by classical arguments and in particular should be exponential under certain conditions on the form of the (initial plus final) wavefunctions.

* A more precise definition is given later.

This factorization property is shown in Appendix II to be equivalent* to a corresponding fall-off property of the connected amplitudes, unless there is a connected multiple scattering process connecting the initial and final particles.

This form of macrocausality involving directly the connected amplitudes is shown in Sec. IIID, E to be equivalent ** to the normal analytic structure of the S-matrix. The derivation of the normal structure from macrocausality is given in Sec. IIID. The easier converse part is outlined in Sec. IIIE.

The methods used in Appendix II are adapted from Refs. 4,5. In those references the first part of the normal structure, namely the cluster decomposition property, was derived from the space-time factorization property, that the remainder $R = W - \prod_k W_k$ fall to zero when the various groups k become infinitely displaced from each other. In contrast to the macrocausality condition formulated in Sec. III, no assumption on the precise rate of fall off was needed and no concept of space-time localization of particles was explicitly introduced. On the other hand, fall off was assumed for all directions of displacements. (See the end of Appendix II.)

A basic idea used in Sec. IIIA, namely the use of Gaussian type wave packets that shrink with a displacement parameter τ was adapted from Ref. 6. In that work the analyticity in momentum transfer, inside a complex ellipse (around the physical region), of two body functions (the absorptive parts due to different channels associated with two initial particles and the scattering function of the two-body channel) was shown to be equivalent to a condition of exponential fall off for the probability of any process generated by an appropriate initial state of the two particles, when they are displaced from each other along a direction normal to the relative velocity in the center of mass system.

These results are stronger in one sense than those described in Sec.III, because they give a domain of analyticity in the complex plane (around the physical region). They are, however, restricted to two-body functions and to analyticity in momentum transfer alone. The work of Ref. 6 has been reviewed in Ref. 7 and extended in Ref. 8.

C. Discontinuity Formulas and Factorization Properties for Causal Configurations

The macrocausality property of Sec. III was stated for certain configurations of groups k of initial and final particles that we will call noncausal.

---

* The precise mathematical results are given later. A strict equivalence is obtained if a few exceptional points are excluded.

** See footnote above.

In Sec. IV we first describe certain factorization properties of transition amplitudes for causal configurations--namely those for which there is a classical connected multiple-scattering process G connecting the initial and final particles--and we show the equivalence between, on the one hand, macrocausality plus these factorization properties, and, on the other hand, the normal analytic structure plus the discontinuity formulas.[9]

The factorization property associated with a connected network G of intermediate particles that has no multiple lines* is this:[10),11)] the transition amplitude of the process is equal--up to a term that falls off exponentially in the appropriate macroscopic limit--to the integral I(G) of the product of transition amplitudes associated with the various vertices of G over all possible real, positive-energy, on mass-shell energy momenta of the intermediate particles.

If the network G has no closed loop, a slightly different formulation can be given.[10)] If G is, for instance, associated with the classical process $A + B \to A_1 + M$, $M + C \to B_1 + C_1$, then the integral I(G) is equal to the transition amplitude of the process $M + C \to B_1 + C_1$, where the (initial) wavefunction of particle M is computed as a (mass shell) "outgoing wave" in the process $A + B \to A_1 + M$ from the given states of A, B, A_1.

The physical interpretation of the properties is clear: in the macroscopic limit, the vertices of G are associated with independent subprocesses.

The described factorization properties for causal configurations are expressed in terms of transition amplitudes. In contrast to the case of macrocausality in Sec. III, no equivalent condition in terms of transition probabilities has been obtained. [The results described in Sec. VB show that these factorization properties imply classical-type formulas for transition probabilities. However, various approximations are involved, and the converse does not seem possible.]

The derivation of the factorization properties from the discontinuity formulas is given in Sec. IVA. The converse is outlined in Sec. IVB. We also briefly describe at the end of Sec. IVB the related results of Ref. 12 in which a pole-type behavior of the scattering function (but no analyticity) was derived from a weaker form of the described factorization property for the special case of one intermediate particle.

D. S-matrix Theory and Physical Space-Time Description of Processes

In Sec. V we describe some closely related works on space-time

* Multiple lines are sets of more than one line that run between the same vertices. The case when multiple lines are involved is discussed in Sec. IV.

description in S-matrix theory, which start from the described physical region analytic structure of the S-matrix. In contrast to the results of Secs. III and IV that refer only to a certain asymptotic limit and to a corresponding asymptotic notion of localization, the results of Sec. V give an approximate description valid also nonasymptotically of the localization of the various subprocesses and of the space-time propagation of the (external and intermediate) particles.

In Sec. VA certain formulas for the transition amplitudes of processes of single and multiple scattering are obtained in terms of space-time wavefunctions.[2),10)] These are defined as the formal Fourier transforms of the mass shell momentum space wavefunctions and the results obtained give the basis for a characterization of space-time localization by means of these quantities.* (The latter is completely consistent with the asymptotic notion of localization of Sec. IIIA, but says more.)

In Sec. VB similar formulas are obtained for the transition probabilities in terms of à la Wigner density and efficiency functions of energy momentum and space time.[13)] The latter are defined in terms of the density and efficiency operators, which are the general quantum description of the initial, resp. final particles when information is not necessarily complete, resp. when the measurement apparatus select ensembles of final states.

The formulas then derived are of a classical type, with the density and efficiency functions replacing the corresponding classical quantities, namely the probability densities and measurement efficiencies.

The density and efficiency functions are, moreover, shown to enjoy the properties of the classical quantities, after appropriate smoothing, or even without smoothing, under conditions that enforce $\hbar$ as well as the Compton wavelength $\hbar/mc$ to be negligible.

These results show how S-matrix theory can lead to the classical space-time description of processes, which is the only ground where experience and theory meet, and they give the basis for physical interpretations of the density and efficiency functions.

The factorization properties of Sec. IV and the related results of Sec. V are crucial for an S-matrix theory of measurements. This is briefly discussed at the end of Sec. V.

E. Conclusion

In the Conclusion (Sec. VI) of this paper, we discuss the general implications of the foregoing results in S-matrix theory, with

* The location of a particle, as defined by the region of intersection of its space-time wavefunction with that of another particle is never spatially confined in its rest frame to less than its Compton wavelength.

emphasis on the discussion of the status of space time in that theory. We also describe briefly and discuss at the end related results of local field theory.

II. Analyticity Properties

A. Scattering Functions*

We consider for simplicity a theory with no superselection rules and with only spinless particles.

The basic observable in a scattering experiment can be considered to be the function $S_{mn}(\varphi_1 \ldots \varphi_m; \psi_1 \ldots \psi_n)$ the square of which is the transition probability** from an initial system of m particles represented by unit norm wavefunctions $\varphi_1 \ldots \varphi_m$ to a final system of n particles represented by unit norm wavefunctions $\psi_1 \ldots \psi_n$.

The norm of a wavefunction φ_i (or ψ_i) is the L^2 norm in the on-mass-shell momentum space of particle i:

$$|\varphi_i| = [(2\pi)^{-3} \int |\varphi_i(p_i)|^2 \delta(p_i^2 - \mu_i^2)\theta(p_{io}) d^4 p_i]^{\frac{1}{2}}, \qquad (II.1)$$

where μ_i is the mass of particle i.

The momentum space wavefunctions can be assumed to span either the space L^2 of square integrable functions or a dense subspace for instance the Schwartz[17] space $\mathcal{D}$ of infinitely differentiable functions with compact support.

The functional S_{mn} is assumed to be linear in the wavefunctions of the initial particles and antilinear in the wavefunctions of the final particles.

This linearity, together with the boundedness of the S-matrix (probability interpretation):

$$|S_{mn}(\varphi_1 \ldots \varphi_m, \psi_1 \ldots \psi_n)| \leq \prod_i |\varphi_i| \prod_j |\psi_j| \qquad (II.1')$$

implies the existence of the momentum space kernels $S_{mn}(p_1, \ldots, p_m,$

---

* The introduction to the S-matrix formalism given below is that of Ref. 1a. An introduction that is closer to physics is given in Ref. 15. On the other hand, a comparison to the more usual presentation which starts with the operator S that maps an Hilbert space $\mathcal{H}$ of free particle states onto itself is given in Appendix I.

** This strict probablity interpretation is not crucial in this paper and could be replaced by a weaker condition.

$q_1, \ldots, q_n$) as tempered distributions* in the space of all on-mass-shell momenta.

Conservation of energy momentum requires this distribution to be concentrated on the set M, defined as the intersection of the hyperplane $\Sigma p_i - \Sigma q_j = 0$ with the above space of all on-mass-shell momenta.

This distribution--when restricted to the set of points at which at least two of the initial and final 4-vectors p_i, q_j are not collinear--is furthermore[19] the product of a δ function $\delta(\Sigma p_i - \Sigma q_j)$ with a corresponding kernel.**

The connected kernels S^c_{mn} are uniquely defined in terms of the kernels S_{mn} by the iterative solution to the system of equations:

$$S_{mn}(P_m, Q_n) = \sum_{\{k\}} \beta_{\{k\}} \prod_{k \in \{k\}} S^c_{m_k n_k}(P_{m_k}, Q_{n_k}) \quad ,$$

$$S^c_{11}(P, Q) = S_{11}(P, Q) \quad , \qquad \text{(II.2)}$$

where the $\{k\}$ are all subdivisions of the set (m,n) into subgroups (m_k, n_k), P_m and Q_n denote the sets of the initial and final momenta and the coefficients $\beta_{\{k\}}$ are phase factors that are equal to plus one (or minus one).†

The kernel $S^c_{mn}(P_m, Q_n)$ is a linear combination of S_{rt} and is likewise a tempered distribution that is the product of a δ function of energy momentum conservation with a kernel $T_{mn}(P_m, Q_n)$ defined on M.

The kernel $T_{mn}(P_m, Q_n)$ is called a scattering function.

The analyticity properties described in this chapter refer to these scattering functions.

---

* The continuity in L^2 norm with respect to each variable implies the continuity in the topology of the Schwartz[17] space S. This in turn implies that the functional S_{mn} is a tempered distribution (nuclear theorem). (Even if the physical wave packets are restricted to D, there exists a (unique) continuation of S_{mn} to function in L^2 for which (II.1′) still holds.)

From the continuity in L^2-norm, precise bounds on the order of these distributions can also be computed.[18]

** The derivatives of $\delta(\Sigma p_i - \Sigma q_j)$ are excluded[19] because of the boundedness (II.1′) of the S-matrix.

† They are all equal to plus one if no fermions are involved. If fermions are involved, then these phases are plus or minus one, depending on the relative order of the subdivisions $\{m_k, n_k\}$ and $\{m,n\}$ (see Appendix IB).

B. Landau Surfaces

Consider a multiple scattering connected graph G associated with the initial and final particles of a process, the set I of its internal lines and a set of corresponding 4-vectors k_i. Each line has an orientation (sense) that specifies the direction of flow of the energy momentum k_i.

To each cycle (closed loop) of G that can be constructed on the internal lines, associate the function z(i): z(i) = 0 if z does not contain the line i, z(i) = +1 (resp. z(i) = -1) if it contains it with the correct (resp. opposite) orientation.

The Landau equations of the graph G are the loop equations

$$\sum_{i \in I} z(i)\, \alpha_i k_i = 0 \qquad (\text{II}.3)$$

for each cycle z, plus the equations of energy momentum conservation at each vertex and the (real, positive energy) mass shell constraint for each internal and external line.

To each internal line of G, let us now associate (in addition to the sense defined above) a sign σ_i, and let α_i in the Landau equations associated with G be restricted by the condition $\sigma_i \alpha_i > 0$.

The Landau surface L(G) is obtained by eliminating the coefficients α_i and the internal k_i. (It is confined to the space M of external momenta.) The L(G) of the trivial one vertex graph is defined to be the empty set.*

A graph with all $\sigma_i > 0$ is denoted G^+ and the set $L(G^+)$ is called a positive α Landau surface. A basic remark that has been emphasized in Ref. 20 and will be used in Sec. III is this: the Landau equations of G^+ are precisely equations of classical kinematics, the (positive) coefficient $\alpha_i k_{io}$ representing the time of propagation of particle i between its creation and annihilation.

Certain definitions will prove useful: M_o is the subset of M where two (or more) initial or two (or more) final energy momentum 4-vectors are collinear.

$L_o(G)$ is the subset of L(G) - M_o that lies on no L(G') where G' is a nontrivial contraction** of G. (G is a trivial contraction of itself.)

---

* Graphs G that arise from insertion of vertices on sets of lines that all begin at one vertex and all end at one vertex of some other diagram will be excluded, since these G do not give new Landau surfaces.

** A contraction is a graph obtained by equating the two end points of certain lines and removing all lines that begin at their own end points.

The physical side of $L_0(G^+)$ is the (well-defined) side to which it would be shifted by a formal scale transformation that increases by a common factor the masses associated with the internal lines of G.

The physical region of G is the subset of M defined by the Landau equations for G without the loop equations (i.e. by the mass shell and conservation law constraints alone).

Some properties of Landau surfaces are these: (i) The union of all positive α Landau surfaces is the union of all $L_0(G^+)$, plus M_0. (This is trivial.)

(ii) Each surface $L_0(G^+)$ is a real analytic submanifold of of codimension 1.[2)]

(iii) The number of graphs G^+ that give positive α Landau surfaces $L_0(G^+)$ that enter bounded portions of the physical region is finite.[21)]

(iv) $L_0(G^+)$ lies on the boundary of the physical region of G^+. Furthermore, in some neighborhood of any point of $L_0(G^+)$ the physical region of G^+ either coincides with $L_0(G^+)$ or lies on its physical side.[2),11)]

(v) If two $L_0(G^+)$ coincide in a neighborhood of a point, then their physical sides coincide at this point.[2)]

By virtue of (v), the physical side of $L_0(G^+)$ is an intrinsic characteristic of the surface that does not depend on the particular G^+ used to define it.

Property (ii) asserts that for any point P of $L_0(G^+)$ there is a real neighborhood U and a real analytic function ℓ (of real analytic local coordinates* of $M-M_0$ near P) such that grad $\ell \neq 0$ in U and $L_0(G^+)$ is defined by $\ell = 0$ in U. The sign of ℓ is conventionally chosen to positive on the physical side of $L_0(G^+)$.

The power series for ℓ at P defines a function $\underset{\sim}{\ell}$ analytic in a certain complex neighborhood of P.

The function $\underset{\sim}{\ell}$ should always be understood as an abbreviation of $\underset{\sim}{\ell}[z;P,L_0(G^+)]$, where the z are analytic local coordinates of $M-M_0$ near P of $L_0(G^+)$. Two different functions $\underset{\sim}{\ell}$ could be used to define the same surface $L_0(G^+)$ at P, but it is not hard[11)] to show that the two corresponding regions $\text{Im}\ \underset{\sim}{\ell} > 0$ will then coincide in some complex neighborhood $\underset{\sim}{U}$ of P.

C. Analyticity Properties

The normal analytic structure is defined by the following properties (1) through (3):

Property 1. "Each scattering function is analytic** at P unless

* Real analytic local coordinates are such that all energy momentum 4-vectors are real analytic functions of them.

** A function is analytic at P if it is an analytic function of analytic local coordinates.

P lies on the union L_c^+ of positive α Landau surfaces of the (connected) graphs G associated with the initial and final particles of the process."

By virtue of properties (i) through (iii) of Sec. IIB, the set L_c^+ has dimension one less than that of M.

Property 2. "If P lies on the Landau surface $L_0(G^+)$ of a given G^+ and if any other $L_0(\hat{G}^+)$ containing P coincides with $L_0(G^+)$ near P, then P has a real neighborhood U and a complex neighborhood $\underset{\sim}{U}$ that contains U, such that the scattering function in U is the boundary value of a function analytic in a domain that lies in $\text{Im}\, \underset{\sim}{z} > 0$. This domain is the intersection of $\underset{\sim}{U}$ with an open half cone $\underset{\sim}{C}$ that can be made arbitrarily close to the open half space $\text{Im}\, \underset{\sim}{z} > 0$ by taking $\underset{\sim}{U}$ sufficiently small."

By virtue of the edge of the wedge theorem,[22] properties (1) and (2) ensure the existence of a "plus iε" analytic continuation of the scattering function past the surface $L_0(G^+)$.

The points which satisfy the conditions of property (2) are called quasi-simple. The points P in a neighborhood of which there is furthermore no other surface $L_0(\hat{G}^+)$ are called simple. The class of quasi-simple points thus includes, in addition to simple points, the points P that lie on $L_0(\hat{G}^+)$ that coincide with $L_0(G^+)$, or that are end points of some other $L_0(G'^+)$ (to which they do not belong). In the latter case, G (or one of the $\hat{G}$ such that $L_0(\hat{G}^+)$ coincides with $L_0(G^+)$) is a contraction of G′ and the consistency of the property (2) at P and at points that lie on $L_0(G'^+)$ in the neighborhood of P is a consequence of the following property: (vi)[2] $L_0(G^+)$ and $L(G'^+)$ are tangent at P and their physical sides (in the neighborhood of P) lie on a common side of the tangent hyperplane.

Property (2) is called the plus iε rule for quasi-simple points. By virtue of properties (i) through (v) of Sec. IIB, almost every point lying on any positive α Landau surface is quasi-simple. The only remaining points are those of M_0--which are confined to a set of dimension 2 less than the dimension of the complete set L_c^+--and the points that lie on two distinct surfaces $L_0(G^+)$--which are confined to a set of dimension 1 less than the dimension of L_c^+.

Most of the latter points are covered by a third property,[3] which is the generalization of property (2) to points P such that for some fixed G^+, any $L_0(\hat{G}^+)$ that contains P coincides near P with the $L_0(G^+)$ of some contraction G_ϵ^+ of G^+. These points are called semi-simple. It is proved in Ref. 2 that the domains $\text{Im}\, \underset{\sim}{z}_i > 0$--where the index i labels the (necessarily finite) number of distinct surfaces $L_0(G^+)$ that contain P--have a nonempty intersection that contains a real limit point.

Property 3. This is the plus iε rule for semi-simple points. It says that "the property analogous to (2) holds at semi-simple points, with the intersection of the $\text{Im}\,\underset{\sim}{\ell}_i > 0$ replacing the domain $\text{Im}\,\underset{\sim}{\ell} > 0$."*

Properties (1) through (3) define the normal analytic structure. They say nothing, however, about points of M and about points that lie on distinct surfaces $L_o(G^+)$ for two or more G^+ that are contractions of no common parent.

The latter points are covered by

Property 4. (Independence property). "In the neighborhood of any point P of $L_c^+ - M_o$, the scattering function is equal to a sum of terms, each corresponding to one independent parent and having the associated analyticity properties."

A plus iε rule is not expected at the points of M_o. A continuity property has been assumed for them in Ref. 14.

A final property is the rule that gives the general discontinuity of S around any physical-region singularity. It is derived in Ref. 14 from properties (1) through (4) plus the continuity property assumed there for M_o points, together with the analogous properties of the connected parts of S^{-1}.** [Property (4) and the continuity at M_o points are needed even though the discontinuity is evaluated at a quasi-simple point.]

Property 5. "If P lies on $L_o(G^+)$ and on no $L(\hat{G}^+)$ for $\hat{G}^+$ different from G^+, and if G^+ has no multiple lines, then the discontinuity of S is given by the Cutkovsky formula obtained by replacing each vertex v of G^+ by the corresponding kernel S_v: namely it is equal to the integral D of the product of the S_v's, over all (real, positive energy) on-mass-shell internal momenta of G.

If G has multiple lines, one must furthermore insert in the corresponding lines α the operator S_α^{-1} defined by $S_\alpha^{-1} S_\alpha = I_{\alpha'}$ where S_α is the restriction of S to the subspace corresponding to the set of lines α, and I_α is the corresponding restriction of unity."

At points where $S = S_c$, the discontinuity is defined as a product of the energy-momentum conservation law delta function of the process times $T - T^{(L)}$, where $T^{(L)}$ denotes the boundary value along a domain of $\text{Im}\,\underset{\sim}{\ell} < 0$ of a counterclockwise analytic continuation of the scattering function T around the surface $\underset{\sim}{\ell} = 0$. We consider as part of the

---

* The consistency of the plus iε rules at P and at the points that lie on other surfaces L_o^+ in the neighborhood of P is a consequence of a generalization of property (vi).

** An assumption that singularities associated with solutions of the Landau equations that require α's of both signs do not systematically cancel the positive-α singularities is also used, but it is believed that this assumption is unnecessary.[16]

property (5) the fact that the function $T^{(L)}$ satisfies the minus iε rule at P. That is, it has analyticity properties similar to properties (1) and (2), but with the domain Im $\underset{\sim}{\ell} < 0$ replacing the domain Im $\underset{\sim}{\ell} > 0$. (These properties of $T^{(L)}$ were proved in Ref. 14.)

The integral D defined in property (5) vanishes on the unphysical side of $L_o(G^+)$ by virtue of property (iv) of Sec. B. This was used in Ref. 14, where S was written as a sum of D plus a term S′ that was proved to have a minus iε continuation around P. The vanishing of D on the unphysical side of $L_o(G^+)$ then implies that S′ is the minus iε continuation of S. This property of D will also be used in Sec. IVB.

On the other hand, D is a certain definite function of the S_v's (and S_α^{-1}'s). Various properties of D can thus be derived from the properties of these latter functions. These are kernels of bounded operators (in the sense of (II.1′) and D can itself be shown[23] to be the kernel of a bounded operator, and thus a tempered distribution. Moreover, certain analyticity properties of D can be derived from the analyticity properties of the S_v's (and S_α^{-1}'s). Various results that give arguments of internal consistency to the analytic structure defined by properties (1) through (5) can thus be obtained.[24]

In the remainder of this section we consider for simplicity a simple point that lies on a surface $L_o(G^+)$ associated with a graph G with no multiple lines (and such that the S_v's reduce to their connected parts) and we briefly describe two analyticity properties of D that only depend on property (1) of the S_v's and that will be used in Sec. IVB.

The various scattering functions T_v's are easily seen to be analytic at the points determined by the various constraints. (The requirement that P be simple assures that these points can lie on no positive α Landau surface of the process v.) Thus the kernel D_r (i.e. D divided by the energy momentum conservation law delta function of the process) can first be shown to be <u>analytic</u> on the physical side of $L_o(G^+)$ (in the neighborhood of P).

Consider next[16] at least one phase factor $\rho_i = \delta(k_{io} - (\underset{\sim}{k}_i^2 + m_i^2)^{\frac{1}{2}})$ associated to an internal line i and write it as

$$\rho_i = \frac{1}{2i\pi}\left[(k_o - (\underset{\sim}{k}_i^2 + m_i^2)^{\frac{1}{2}} + i\varepsilon)^{-1} - (k_o - (\underset{\sim}{k}_i^2 + m_i^2)^{\frac{1}{2}} - i\varepsilon)^{-1}\right].$$

By introducing analytic continuations of the corresponding T_v's slightly off mass shell, D_r can be shown to be itself equal to a corresponding difference $D_r^+ - D_r^-$ of terms that have in the neighborhood of P analyticity properties analogous to properties (1) and (2); they obey plus and minus iε rules respectively.

This result shows in particular that T is equal to D_r^+ up to an analytic function[22] in the neighborhood of P. It can be similarly shown to be equal (up to an analytic function) to a Feynman-type integral where "plus iε poles" replace for each line i the corresponding mass shell δ-functions.

III. Macroscopic Causality and Normal Analytic Structure

A. Space-Time Localization of Particles

To formulate a causality condition, some notion of space-time position is needed. There is no difficulty with space-time displacements: from group theoretic considerations (Appendix IA), one is led to interpret the four-vector a in the equation

$$\varphi^a(p) = \varphi(p)\, e^{ipa} \qquad \text{(III.1)}$$

as a space-time displacement.

On the other hand, the position of a relativistic particle cannot be determined precisely. The sharpest definition allowable either in practice or mathematically is of the order of the particle Compton wavelength.

For the purposes of the present chapter, it is sufficient to introduce a weak notion of a macroscopic space-time localization that roughly expresses the idea that the asymptotic space-time localization of a particle represented by $\varphi(p)$ lies in its velocity cone V, which is defined as the closure of the set of all space-time points that can be reached by traveling from the origin along any direction that is parallel to any on-mass-shell four-vector p such that $\varphi(p)$ is nonvanishing.

To make this idea precise, consider a given arbitrarily small open region R, the closure of which does not intersect the velocity cone V and a dilation that takes each point u′ to the point u′τ. This dilation takes the region R to the dilated region R_τ.

The region R_τ becomes infinitely large as $\tau \to \infty$. Our basic assumption is that it makes sense, at least asymptotically, to speak of the probability $P(R_\tau)$ that the particle represented by $\varphi(p)$ can be considered to be in R_τ. Moreover, this probability should become small when the space-time wavefunction corresponding to $\varphi(p)$ becomes uniformly small in a region R_τ', where the open R′ contains the closure of R. (Notice that R_τ' contain all points that lie at distance $\alpha\tau$ from points R_τ, where α is positive and τ goes to infinity.)

Different space-time functions can be defined. Usual forms are

$$f(x) = \int \varphi(p)\, e^{ipx}\, \frac{dp}{2p_o} \quad , \qquad \text{(III.2)}$$

and

$$\tilde{\varphi}(x) = \int \varphi(\underline{p})\, e^{ipx}\, \frac{d\underline{p}}{(2p_o)^{\frac{1}{2}}} \quad . \quad * \qquad \text{(III.3)}$$

It is not possible in general to attach a physical meaning to the values of either $f(x)$ or $\tilde{\varphi}(x)$ at a given point x. However, both $\max_{x \in R_\tau} |f(x)|$ and $\max_{x \in R_\tau} |\tilde{\varphi}(x)|$ have a rapid (i.e. faster than any inverse power of τ) fall off under certain conditions on $\varphi_\tau(\underline{p})$. Under other conditions, the fall off is exponential. Thus, although the actual values of these different space-time functions are different, the nature of the rate of fall off is the same. And this nature is the same also for the various other quantities that have been proposed as measures of the space-time densities (cf. Sec. V). Thus it seems reasonable to take the nature of the rate of fall off given by these functions to be a reliable indication of the nature of the rate of fall off of $P(R_\tau)$.

For example, suppose the wave function $\varphi(\underline{p})$ belongs to the Schwartz space $\mathcal{D}$. If the closure of R does not intersect V_-, then Ruelle's lemma[25] implies that both $\max_{x \in R_\tau} |f(x)|$ and $\max_{x \in R_\tau} |\tilde{\varphi}(x)|$ have a rapid fall off with τ.

If we consider a set of similarly shaped regions R_u centered at different space-time points u, then this fall off is also uniform on compact subset of u's, as long as the closure of the set of R_u does not intersect V.

Consider next Gaussian-type wavefunctions that shrink with the displacement parameter τ. In particular, consider[3] the class of wavefunctions of the form

$$\varphi_\tau(\underline{p}) = \chi(\underline{p})\, e^{-(\underline{p}-\underline{P})^2 \gamma \tau} \quad , \qquad \text{(III.4)}$$

where $\underline{P}$ is a given 3-vector, γ is a positive constant and the $\chi(\underline{p})$ are confined to the class of functions of $\mathcal{D}$ that have a fixed compact support, are analytic in some fixed complex neighborhood N of $\underline{P}$ and satisfy $|\chi(\underline{p})| \leq 1$ if $\underline{p}$ is real or in N.

Theorem. "Both $\max_{x \in R_\tau} |\tilde{\varphi}(x)|$ and $\max_{x \in R_\tau} |f(x)|$ have an exponential fall off if the intersection of the closure of R with the line parallel to the 4-vector $P(P_o = (\underline{P}^2 + m^2)^{\frac{1}{2}}, \underline{P})$ drawn from the origin is void. This fall off is, moreover, uniform on any compact subset of u's over which this intersection remains void and, furthermore, the constant of exponential fall off decreases no faster than linearly in the $\gamma \to 0$ limit.

---

* $p_o = (\underline{p}^2 + m^2)^{\frac{1}{2}}$

"In particular, for all γ smaller than some fixed γ_0, the following inequalities hold:

$$\max_{x \in R_\tau} |f_{\chi,\gamma}(x)| < Ce^{-\alpha\gamma\tau},$$

$$\max_{x \in R_\tau} |\tilde{\varphi}_{\chi,\gamma}(x)| < De^{-\beta\gamma\tau}, \qquad \text{(III.5)}$$

where C and α (resp. D and β) do not depend on χ, γ, or τ."

An outline of the proof is given in Appendix III.

The fall off indicated by the symbol $\Rightarrow 0$ will mean a fall off of the type specified by (III.5) when the $\varphi_\tau(\underline{p})$ under consideration are of the form (III.4). The $\varphi(\underline{p})$ in $\mathcal{D}$ of interest will be the $\gamma = 0$ limit of the $\varphi_\tau(\underline{p})$ of (III 4). The meaning of $\Rightarrow 0$ switches to rapid fall off in this case.

Our general assumption about space-time localization is that for any $\varphi_\tau(\underline{p})$ of the form (III.4) (including $\gamma = 0$), $P(R_{u\tau}) \Rightarrow 0$ uniformly in u on compact sets such that the intersection of the closure of the union of the R_u with V is void. The set V is the velocity cone of $\varphi_\tau(\underline{p}) = \chi(\underline{p})$ for $\gamma = 0$, but for $\gamma \neq 0$, it is the line through the origin with direction P. Thus V is defined by the "effective support" of $\varphi_\tau(\underline{p})$, which for the case $\gamma \neq 0$ is simply the point $\underline{p} = \underline{P}$.

B. Macroscopic Causality

Let $\varphi_{i\tau}$ be a set of initial and final wavefunctions of the form (III.4) and let $\varphi_{i\tau}{}^{u_i\tau}$ be a corresponding set of displaced wavefunctions:

$$\varphi_{i\tau}{}^{u_i\tau}(\underline{p}_i) = \varphi_{i\tau}(\underline{p}_i)\, e^{i(p_i u_i)\tau} \qquad \text{(III.6)}$$

It is convenient to visualize the limit $\tau \to \infty$ in a space-time coordinate system scaled to τ. Then the displacements $u_i\tau$ become fixed displacements u_i and the regions R_τ of the preceding section become fixed regions R.

In this system, the particle i is localized in its displaced velocity cone $V_i{}^{u_i}$ with tip at u_i (that is, the probability $P(R) \Rightarrow 0$ in the $\tau \to \infty$ limit of the closure of R does not intersect the corresponding $V_i{}^{u_i}$).

The requirement that the transfer of energy momentum from the initial particles to the final particles can be ascribed to (stable) physical particles is the requirement that one can draw a network of lines representing a classical multiple scattering process that is compatible with both the momentum energy requirements on the various initial and final particles and also the space-time requirements entailed by the

positions of the various initial and final velocity cones $V_i^{u_i}$. Specifically, the energy momentum of each external line of the classical process must satisfy the appropriate mass shell constraint, and its momentum must lie in the effective support of $\varphi_{i\tau}(p)$. Moreover, the vertex of the multiple scattering diagram that represents the interaction involving the external particle i must lie in the corresponding $V_i^{u_i}$.

For a fixed set of $\varphi_{i\tau}$, there may be certain sets $u \equiv \{u_i\}$ of displacements such that it is possible to find a classical multiple scattering process that satisfies the conditions just described. These u are called causal with respect to $\varphi_\tau \equiv \{\varphi_{i\tau}\}$. The set consisting of the u's that are noncausal is called $A(\{\varphi_{i\tau}\})$. The set consisting of the u's such that one can find no connected classical process is called $A_c(\{\varphi_{i\tau}\})$. In the case $\gamma > 0$, the A and A_c only depend on $P \equiv \{P_i\}$ and they will also be denoted $A(P)$ and $A_c(P)$ respectively.

Macrocausality is the requirement that

$$|S^c(\{\varphi_{i\tau}^{u_i\tau}\})| \Rightarrow 0 \qquad \text{(III.7)}$$

uniformly on compact subsets of $A_c(\{\varphi_{i\tau}\})$.*

For $\gamma = 0$, the symbol $\Rightarrow 0$ denotes a rapid fall off and for $\gamma > 0$, it denotes a bound of the type (III.5):

$$|S^c(\{\varphi_{i\tau}^{u_i\tau}\})| < Ce^{-\alpha\gamma\tau} \qquad \text{(III.7')}$$

for γ smaller than some fixed γ_0 and where C and α are independent of γ, τ and the χ_i's.

This form of the macrocausality condition will be derived in Appendix II (except at exceptional values of P) from the factorization property $W - \prod_k W_k \Rightarrow 0$ described in Sec. I. (The converse is straightforward.) It is justified in the framework of a classical model. All transfers of energy momentum not ascribable to stable physical particles in accordance with the classical ideas are assumed to give only effects that fall off at some arbitrary but fixed exponential rate. For sufficiently small γ, these terms are masked by effects $\sim e^{-\alpha\gamma\tau}$ that are classically ascribable to stable physical particles: these latter effects come from classical multiple scattering processes in which some external particles are not localized in their effective

* The physical justification of macrocausality (see below) requires that the velocity cones used in the construction of A or A_c be "slightly larger" than the ones defined by the effective supports of the $\varphi_{i\tau}$. This modification does not significantly alter any arguments, as long as P is kept away from M_0.

velocity cones, or the momenta are not in the effective supports. Similar classical effects occur in the $\gamma = 0$ limit, where they give contributions having rapid fall off. Thus, the dominant effect for small enough γ has the same type of fall off as $P(R_T)$ itself. This is the assertion of (III.7).

A final remark: Suppose $Z(\underline{p})$ is some function of all the initial plus final variables $\underline{p}_i$ that is analytic with an appropriate choice of units in each (complex) component $p_{i\mu}$ of each vector $\underline{p}_i$ for $|p_{i\mu} - P_{i\mu}| \leq 1$. (The corresponding domain of analyticity in $\underline{p} \equiv \{\underline{p}_i\}$ is N.) Then consider the function $\hat{\chi}(\underline{p}) = Z(\underline{p})\chi(\underline{p})$, where χ is a product $\prod_i \chi_i$ of functions of the form (III.4) that has its compact support in N.

Macrocausality says a priori nothing unless Z has itself a product form.* However, it is easily seen[3]--by using a multidimensional Cauchy expansion for Z around the point $P \equiv \{P_i\}$--that (III.7′) still holds for this nonproduct $\hat{\chi}(\underline{p})$ (with eventually different C and γ_0). This fact is used in Sec. D.

C. Causal Displacements[2]

In this section, we describe some basic results, which will be used later, about the set A_c of noncausal displacements. By causal displacements, we will only refer here to those that are associated with connected diagrams (i.e. do not belong to A_c).

(i) "If the (effective) support of $\varphi_T \equiv \prod_i \varphi_{i_T}$ contains no positive α Landau point and no point of M_0, then all nontrivial displacements belong to A_c."

The trivial displacements consist of common displacements of all particles plus displacements of particles along lines in their own velocity cones.

Property (i) is easily derived from the remark that positive α Landau equations are precisely equations of classical kinematics[20] (Sec. II).

Let P be a point of $M-M_0$. Let Γ be a set of u's** that are linear combinations of $\ell = [3(n + m)-4]$ u_k such that the corresponding $q_k = u_k(p - P)$ are real local analytic coordinates of $M-M_0$ at P. (A specific set of u_k and q_k is exhibited at the beginning of Sec. D.)

The following properties are proved in Ref. 2:

(ii) "If the support of $\chi = \prod_i \chi_i$ is sufficiently small, then the intersection of Γ with the set of displacements that are trivial with

---

* We have specifically restricted the statement of macrocausality to the case of product wavefunctions $\varphi_{i_T} u_{i_T}$, because the classical arguments are more compelling in this case. In any case, the physical arguments are difficult to extend to "wavefunctions" that mix initial and final variables.

** These u are sets $\{u_i\}$ of displacements of each particle (see Sec.B).

respect to χ is empty."

Properties (i) and (ii) ensure that if P does not lie on L_c^+, then all the displacements in the set Γ belong to $A_c(P)$ and also belong to $A_c(\chi)$ if the support of χ is sufficiently small.

(iii) "If P is a simple or quasi-simple point of $L_c^+ - M_o$, there is only one causal direction u in the set Γ at P (i.e. which does not belong to $A_c(P)$) and this direction is determined by the direction of the gradient at P to the Landau surface* L_o that contains P.

"For any open cone C that contains that direction, the support of χ can, moreover, be chosen sufficiently small so that all u's different from zero in the complement C' of C in Γ lie in $A_c(\chi)$."

The set of all causal displacements is obtained by adding trivial displacements to those that lie in Γ.

A similar property has been obtained for semi-simple points. It says that:

(iv) "At a semi-simple point P, the analogous property holds, with the direction of the gradient to L_o at P being replaced by the closed convex cone C_P that is formed of positive linear combinations ov vectors $u_c(P)$ that are gradients at P to the various L_o that contain P."

These properties will be used in Sec. D to derive from the macrocausality condition the normal analytic structure of the S-matrix [properties (1) through (3) of Sec. II].

At a non-semi-simple point P, a general plus iε rule can be derived if the union of the causal u's in the set Γ at P is confined to the union of a finite number of disjoint closed convex cones. It says that the scattering function is equal in the neighborhood of P to a sum of terms, one corresponding to each cone and having the associated analyticity properties. (This property becomes property (4) of Sec. II if each cone is associated with an independent parent.) The points P that satisfy this condition are called type I points.

There are (exceptional) points that are not type I points. These are called type II points. (They include all points of M_o.) For type II points, no analyticity property has been derived so far from macrocausality. However, the cluster decomposition property can still be derived even though some of the points that lie on several surfaces of partial energy momentum conservation are type II points. This is easily seen by considering common displacements d_k of the particles of each subgroup k such that $d \equiv \{d_k\}$ belongs to $A_c(\chi)$. (This is always possible if the support of χ is sufficiently small.)

---

* This is the gradient at P of any real analytic $\ell(q)$ with grad $\ell \neq 0$ such that $\ell(q) = 0$ defines L_o locally and such that ℓ is positive on the physical side of L_o (see Sec. IIB).

D. Derivation of the Normal Analytic Structure from Macrocausality

Consider the $3(n + m)$ components of the $(n + m)$ vectors $p_i - P_i$. For simplicity, we may choose from among them (as is always possible) a set of $\ell = 3(n + m) - 4$ analytic local coordinates q_k of $M\text{-}M_o$ at P. The set Γ of Sec. C is then the set of $u \equiv \{u_i\}$ such that each u_i is space like and such that the four components of the set $\{u_i\}$ that correspond to the four components of the set of vectors $p_i - P_i$ that are not in the set $q = (q_1 \ldots q_\ell)$, are fixed at zero. The displacement u is thus represented by a vector $u \equiv (u_1 \ldots u_\ell)$. (This particular choice of q_k and u_k slightly simplifies the notations, but is not important in the proofs.)

On the other hand, the initial and final wavefunctions will be restricted to a class of functions $\{\varphi_{i\tau}{}^{u_i\tau}\}$ of the form (III.4), where the support of $\chi = \prod_i \chi_i$ lies in the neighborhood N defined at the end of Sec. B. Moreover, the intersection of N with the complex mass shell $\tilde{M}$ should lie in the region covered by the above-described local analytic coordinate system.

The connected amplitude $S_c(\{\varphi_{i\tau}{}^{u_i\tau}\})$ of the process will then be written

$$e^{iPv} S_c(\{\varphi_{i\tau}{}^{u_i\tau}\}) = \int T(q)\chi(q)J(q)e^{-iqv}e^{-\gamma|v|\mu(q)}dq, \quad \text{(III.8)}$$

where $v = \tau u$, $T(q)$ denotes the scattering function, $J(q)$ is the Jacobian that comes from the elimination of the δ-functions of energy momentum conservation and of mass shell constraint, and $\mu = \sum_i (p_i - P_i)^2$, where i ranges over all the initial and final particles.

The function $\mu(q)$ is positive for real q and vanishes for $q = 0$. The functions $J(q)$ and $\mu(q)$ are analytic in complex neighborhoods of the origin and N is taken small enough to lie in both of these. Finally, $J(0)$ is nonzero.

We introduce

$$\hat{T}(v,t) = \int F(q)\, e^{-t\mu(q)}\, e^{-iqv}\, dq, \quad \text{(III.9)}$$

with

$$F(q) = T(q)\,\chi(q)\,J(q)\,. \quad \text{(III.10)}$$

Then (III.8) gives

$$\hat{T}(v,\gamma|v|) = S_c(\{\varphi_{i\tau}{}^{u_i\tau}\})\, e^{iPv}\,. \quad \text{(III.11)}$$

In subsection (a) below, the analyticity of $F(q)$ at $q = 0$ is derived from macrocausality, in case P of $M\text{-}M_0$ lies on no positive-α Landau surface. In subsection (b), the plus iε rule for $F(q)$ at $q = 0$ is obtained for semi-simple points P. The corresponding properties of T itself, considered as a distribution, follow[2),3)] from these properties of $F(q)$, together with the facts that $J(q)$ is analytic and nonzero at $q = 0$, and that $\chi(q)$ is analytic and can be taken nonzero at $q = 0$.

(a) Analyticity at Non-Landau Points

Consider a point $P \equiv \{P_i\}$ of $M\ M_0$ that does not lie on L_c^+. Then choose the support of χ sufficiently small so that it contains no point of L_c^+ plus M_0, and so that the set of all unit norms u's of the form $(u_1 . \ldots . u_\ell)$ is a compact subset of $A_c(\{\varphi_{i\tau}\})$ for $\gamma \geq 0$ (see Sec. C).

The special case $\gamma = 0$ gives the rapid fall off condition:

$$\hat{T}(v,0) \Rightarrow 0 \qquad \text{(III.12)}$$

for $|v| \to \infty$, uniformly in all directions of v.

Then (III.9) shows that $F(q)$ is the Fourier transform of $\hat{T}(v,0)$:

$$F(q) = (2\pi)^{-\ell} \int \hat{T}(v,0)\, e^{iqv}\, dv\ , \qquad \text{(III.13)}$$

and is thus infinitely differentiable.

To show that it is also analytic at $q = 0$, write:

$$(2\pi)^{\ell} F(q) = \int e^{iqv}\, dv \left\{ \hat{T}(v, \gamma_0 |v|)\, e^{\gamma_0 |v| \mu(q)} - \int_0^{\gamma_0 |v|} dt\, \frac{\partial}{\partial t} \left[\hat{T}(v,t)\, e^{t\mu(q)} \right] \right\} . \qquad \text{(III.14)}$$

Equations (III.7′) and (III.11) immediately entail that the term $F_1(q)$ coming from the first term under the integral is convergent and, moreoyer, analytic in the domain E defined as the intersection of N with

$$\operatorname{Re} \mu(q) + \frac{|\operatorname{Im} q|}{\gamma_0} < \alpha\ , \qquad \text{(III.15)}$$

where α is the fall off parameter in (III.7′).

The domain E includes the origin since $\mu(0) = 0$.

The integral $F_2(q)$ associated with the second term under the integral of Eq. (III.14) is a well-defined function for q real in E since F and F_1 are both well defined. We shall now show that it is also analytic in E.

Introducing (III.9) into the term of the integrand of (III.14) that corresponds to F_2, one obtains

$$e^{iqv}\frac{\partial}{\partial t}[\hat{T}(v,t)\,e^{t\mu(q)}] = \int F(q')\,e^{i(q-q')v}\,e^{t(\mu(q)-\mu(q'))}[\mu(q)-\mu(q')]dq'. \tag{III.16}$$

(The transference of $\partial/\partial t$ to the integrand is allowed because $F(q)$ belongs to $\mathcal{D}$.)

Hefer's Theorem[26] allows one to write

$$\mu(q) - \mu(q') = \underset{\sim}{\rho}(q,q') \cdot (q - q') \quad , \tag{III.17}$$

where the components ρ_j $(j = 1,2,\ldots,\ell)$ of $\underset{\sim}{\rho}$ are analytic in the product of the domains N in q and q' space. Eq. (III.16) then becomes

$$e^{iqv}\frac{\partial}{\partial t}[\hat{T}(v,t)\,e^{t\mu(q)}] = \nabla_v[e^{iqv}e^{t\mu(q)}\underset{\sim}{H}(q,v,t)] \quad , \tag{III.18}$$

where

$$\underset{\sim}{H}(q,v,t) = \int F(q')e^{-iq'v}\,e^{-t\mu(q')}\underset{\sim}{\rho}(q,q')\,dq' \quad . \tag{III.19}$$

We may thus write

$$(2\pi)^{\ell}F_2(q) = -\lim_{R\to\infty}\Big[\int_{|v|<R} dv \int_0^{\gamma_0|v|} dt\,\nabla_v[e^{iqv}e^{t\mu(q)}\underset{\sim}{H}(q,v,t)]\Big] . \tag{III.20}$$

For fixed R, we may change the order of integration and perform a first integration over v in the domain $t/\gamma_0 < |v| < R$. Then Gauss's theorem gives this integral as a difference of integrals over the boundary surfaces at $|v| = R$ and $|v| = t/\gamma_0$.

The contribution $F_2^R(q)$ to $F_2(q)$ associated with the surface $|v| = R$ vanishes in the $R \to \infty$ limit because of the following two fall off properties of the components of $\underset{\sim}{H}(q,v,t)$.

(i) "The function $F(q')\rho_j(q,q')e^{-t\mu(q')}$ belongs to $\mathcal{D}$ as a function of q'. This yields a uniform rapid fall off of the components H_j of $\underset{\sim}{H}$ in the $|v| \to \infty$ limit, for fixed t.

The following bound is also obtained[3] for any integer $n > 0$:

$$|H_j(q,v,t)| < \frac{C_0^{(n)} + C_n t^n}{|v|^n} \quad , \tag{III.21}$$

where $C_0^{(n)}$ and C_n are independent of j, v, t and q for q in E."

(ii) "Equation (III.7′) and the remark at the end of Sec. IIIB yield the bound:

$$|H_j(q,v,t)| < C' e^{-\alpha t} \qquad \text{(III.22)}$$

for $0 < t \leqq \gamma_0 |v|$, where C' is also independent of j, v, t and q for q in E

If we divide the region of integration over t into the two parts $0 < t < t(R)$ and $t(R) < t < \gamma_0 R$, where $t(R) = \ell\beta^{-1} \ln R$ and $\beta = \alpha - \mathrm{Re}\,\mu(q)$ (β is positive for any q in E), the two corresponding contributions to $F_2^R(q)$ vanish in the $R \to \infty$ limit because of Eqs. (III.21) and (III.22) respectively. (Details are given in Ref. 3.)

The function $F_2(q)$ is thus equal to the contribution of the surface t/γ_0:

$$(2\pi)^{\ell} F_2(q) = \gamma_0 \int e^{iq\gamma_0} e^{\gamma_0 |v| \mu(q)} \hat{v} \cdot \underline{H}(q, v, \gamma_0 |v|)\, dv, \qquad \text{(III.23)}$$

where $\hat{v} = |v|^{-1} v$.

Since each component H_j is an analytic function of q in N, $F_2(q)$ is analytic in E by virtue of (III.7′) (and the remark at the end of Sec. IIIB).

This proves the analyticity of F(q) and thus of T(q) at $q = 0$ and completes the proof of property (1) of Sec. II.

(b) The Plus iε Rule

Consider the cone C_P of causal u at a semi-simple point P and a slightly larger cone C that contains C_P. We choose the support of χ sufficiently small so that all $v \neq 0$ in the complement C′ of C lie in the acausal set $A_C(\{\varphi_{i\gamma}\})$ for $\gamma \geqq 0$.

The cone C has[3] a nonempty open polar cone C^+, which is the set of all vectors $y = (y_1 \ldots y_\ell)$ such that $y \cdot v > 0$ for all $v \neq 0$ in the closure of C.

The proof of the plus iε rule proceeds as follows. The $\hat{T}(v,t)$ of (III.9) is written as:

$$\hat{T}(v,t) = \theta(C)\hat{T}(v,t) + \theta(C')\hat{T}(v,t), \qquad \text{(III.24)}$$

where $\theta(C)$ is one for v in C and zero otherwise.

The function F(q) divides accordingly. The first term,

$$F_C(q) = \int \theta(C)\hat{T}(v,0) e^{iqv}\, dv\, (2\pi)^{-\ell},$$

is (as a distribution[3]) the boundary value of a function analytic for Im q in C^+ due to the exponential fall off of e^{iqv} as $|v| \to \infty$. [$\hat{T}(v,0)$ is bounded by a constant (see (II.1′).]

The second term is

$$F_{C'}(q) = \int \theta(C')\hat{T}(v,0)e^{iqv}\, dv\, (2\pi)^{-\ell} \quad .$$

The intersection of $|u| = 1$ with C' is a compact subset of $A_c(\chi)$. Thus the arguments of subsection (a) can be used again.

The only difference is that there is an extra surface term at the boundary of C'. This term is confined to the boundary of C. With methods similar to those described in subsection (a), it can be shown[3] to be the boundary value of a function analytic in the intersection of Im q in C^+ with a neighborhood of the origin.

The sum $F(q)$ is, therefore, analytic in the intersection of Im q in C^+ with some neighborhood of $q = 0$.

This proves property (3) of Sec. II and property (2) is a special case.

E. The Converse

In this section, we give an outline of the proof[3] that properties (1) through (4) of Sec. II imply conversely the causality property (III.7) for any compact subset of A_c.** For simplicity, we consider only functions χ with sufficiently small compact support, so that in particular the conditions of Sec. IIID are satisfied, but this is not important in the case $\gamma > 0$.

If the support of χ contains no positive α Landau point, then all nontrivial u's lie in A_c for $\gamma \geqq 0$ (see Sec. C). On the other hand, property (1) ensures that $T(q)$ is analytic in the support of χ and in some complex neighborhood N. Then a distortion of contour analogous to that of Appendix III yields (III.7) for $\gamma > 0$. For $\gamma = 0$, the proof is a simple consequence of Ruelle's lemma.

If P is a simple, quasi-simple, or semi-simple point of $L_c^+ - M_o$ the set A_c is described in Sec. IIIC. For any u of A_c, there is a set of coordinates $(q_1 \ldots q_\ell)$ such that u belongs to the corresponding set Γ (see Sec. IIIC*). By virtue of property (3) of Sec. II, there is an Im q in the domain of analyticity of $T(q)$ such that Im $q \cdot u < -\epsilon$, where ϵ is positive and can be made independent of the particular u in the compact subset of A_c considered. This property ensures that one can find suitable distorted contours in the two cases $\gamma > 0$ and $\gamma = 0$ discussed below.

For $\gamma > 0$, the method is analogous to that of Appendix III. Some slight modifications are needed because T is no longer analytic every-

---

* For a general set $\{u_k\}$ and $\{q_k = u_k(p - P)\}$, a displacement in Γ is a combination $\sum_1^\ell \lambda_k u_k$. Then $u(p - P) = \Sigma \lambda_k q_k$. We will still denote this expression $q \cdot u$.

** So far, a complete proof has been given in Ref. 3 only for properties (1) and (2). (A weaker form of (III.7) has on the other hand been proved in the general case.)

where in the real support of χ. With a suitable separation of the domain of integration into an internal part around $q = 0$ and an external part, the contribution of the external part can be evaluated by using the boundedness of the S-matrix [Eq. (II.1')]. The contribution of the internal part is evaluated by using a distorted contour that avoids the (codimension 1) singularity.

For $\gamma = 0$, the proof is analogous to that given in special cases in Refs. 10 and 11. Consider, for simplicity, a given u that lies in the corresponding Γ. We may write (see Sec. IIID):

$$\hat{T}(v,0) = \int T(q)\chi(q)J(q)e^{-iqv}\,dq \quad , \qquad \text{(III.25)}$$

where $v = \tau u$.

Consider a function $\xi(q)$ that has a compact support that contains the support of χ, is analytic inside its own support and has on the boundary of that support a behavior analogous to the function $e^{-[1/(x^2-1)]^2}$. The support of ξ is chosen so that u lies also in $A_c(\xi)$.

By using a suitable distortion of the support of ξ into the complex q-space, the Fourier transform $\hat{T}_\xi(v')$,

$$\hat{T}_\xi(v') = \int T(q)\xi(q)e^{-iqv'}\,dq, \qquad \text{(III.25)}$$

can be shown to have a uniform rapid fall off as $|v'| \to \infty$ for all v' that lie in $A_c(\xi)$. On the other hand, the Fourier transform $\hat{\chi}_\xi(v')$ of $\chi(q)J(q)/\xi(q)$ falls off rapidly in all directions.

Since $\hat{T}(v,0)$ can be written as a convolution product of $\hat{T}_\xi(v)$ and $\hat{\chi}_\xi(v)$, it is then easy to show* that it falls off rapidly if v lies in $A_c(\xi)$. The details of these converse proofs are given in Ref. 3.

IV. Discontinuity Formulas and Factorization Properties for Causal Configurations

A. Derivation of the Factorization Properties From the Discontinuity Formulas

Consider a set of initial and final $\varphi_{i\tau}^{u_i\tau}$ of the form (III.4) (with $\gamma \geq 0$). Suppose, for simplicity, that P lies on $L_o(G^+)$ and on no $L(\hat{G}^+)$ for $\hat{G}^+$ different from G^+, and that $S = S^c$ at P.

Property (5) of Sec. II entails that

$$S^c(\{\varphi_{i\tau}^{u_i\tau}\}) - D(\{\varphi_{i\tau}^{u_i\tau}\}) = S^{c(L)}(\{\varphi_{i\tau}^{u_i\tau}\}) \quad . \qquad \text{(IV.1)}$$

But $T^{(L)}$ obeys a minus iε rule analogous to property (2) of Sec. II (see Sec. II). The results of Sec. IIIE can then be used to prove

* We also use the slow growth (i.e. no faster than a polynomial) of $\hat{T}_\xi(v')$ for the causal directions of v'. This is a consequence of the fact that T is a tempered distribution.

that (for a sufficiently small compact support of χ)

$$S^{c}{}^{(L)}(\{\varphi_{i\tau}{}^{u_i\tau}\}) \neq 0 \qquad \text{(IV.2)}$$

on any compact subset of u's that belong to a set $A_c^-(\{\varphi_{i\tau}\})$, which is defined similarly to A_c, but the "anticausal" u's are excluded rather than the causal u's. (The anticausal u's have directions opposite to those of the causal u's.)

The set A_c^- thus contains in particular (for a sufficiently small support of χ) the causal u's.

In the case of a graph G with no multiple lines, (IV.1) and (IV.2) then yield immediately the factorization property of transition amplitudes described in Sec. I, for $D(\{\varphi_{i\tau}{}^{u_i\tau}\})$ is the integral over the (real, positive energy) on-mass-shell momenta of G of the product of transition "amplitudes" associated with each vertex of G. They are amplitudes between states that are either wavefunctions $\varphi_{i\tau}{}^{u_i\tau}$ in the case of an external line or pure momentum space eigenstates in the case of an internal line. (They can always be defined as kernels of operators as in Sec. II.)

The physical interpretation of this property is that in the $\tau \to \infty$ limit the vertices of G are associated with independent subprocesses.

If G is associated with the classical process $A + B \to A_1 + M$, $M + C \to B_1 + C_1$ (see Sec. I), then $D\{\varphi_{i\tau}{}^{u_i\tau}\}$ is equal to the transition amplitude of the process $M + C \to B_1 + C_1$, where the initial state of M is the "outgoing wave" of the process $A + B \to A_1 + M$:

$$\varphi_M(\underline{k}) = \int \varphi_A^*(\underline{p}_A)\varphi_B^*(\underline{p}_B)\varphi_{A_1}(\underline{p}_{A_1})\, S\left(\underline{p}_{A_1}, \underline{k}, \underline{p}_A, \underline{p}_B\right) \frac{d\underline{p}_A}{2p_{Ao}} \frac{d\underline{p}_B}{2p_{Bo}} \frac{d\underline{p}_{A_1}}{2p_{A_1o}} \,. \qquad \text{(IV.3)}$$

This result extends to all G not having closed loops and is important in the S-matrix theory of measurement, which is described at the end of Sec. V.

If the graph G has multiple lines, a physical interpretations similar to that described above, although not so limpid, can still be given. The $S_\alpha{}^{-1}$'s can be considered to compensate for the fact that the interaction between the particles of the multiple line α would otherwise be counted twice: once in each of the two S_v's that involve this multiple line.

To see this, one may consider the $S_v{}^{(\alpha)}$ obtained by joining together one of the two S_v's with $S_\alpha{}^{-1}$. The decompositions of S_v and $S_v{}^{(\alpha)}$ into "bubble diagrams" considered in Ref. 14 exhibit the fact that the decomposition of $S_v{}^{(\alpha)}$ contains precisely the terms of the

decomposition of S_v that involve no interaction between the particles of line α.

(The discontinuity formula is obtained in Ref. 14 by using precisely this property of $S_v^{(\alpha)}$.)

One final remark. In the case of wave packets in D ($\gamma = 0$), the term $D(\{\varphi_{i\tau}^{u_i\tau}\})$--which falls off rapidly for u in the intersection of A_c and A_c---can be shown to fall off as an inverse power of $\tau^{\frac{3}{2}}$ for a causal (or anticausal) u. The rate is understandable physically as an effect of the "spreading of the wave packets" associated with the propagation of the intermediate particles.

B. The Converse

We have proved in Sec. A that

$$S^c(\{\varphi_{i\tau}^{u_i\tau}\}) - D(\{\varphi_{i\tau}^{u_i\tau}\}) \Rightarrow 0 \qquad \text{(IV.5)}$$

on any compact subset of A_c.

If, conversely, we start from this assumption, then the methods of Sec. IIIC show that $T - D_r$, which is the kernel $S - D$ divided by the energy momentum conservation law delta function, is the boundary value of a function analytic in a domain of $\text{Im}\, \underset{\sim}{\ell} < 0$. Since D vanishes on the nonphysical side of $L_0(G^+)$, this function is precisely the counterclockwise continuation $T^{(\mathcal{L})}$ of T and property (5) is obtained.

From a physical point of view, one may wish to demand that Eq. (IV.5) hold a priori, but only for essentially causal u's.

Consider a cone C of u's of the form $(u_1 \ldots u_\ell)$ that contains the cone of causal u's and is arbitrarily close to it. We assume (i) that (IV.5) holds uniformly on the set of unit norm u's that lie in the closure of C, and (ii) the macrocausality condition of Sec. III, namely,

$$S^c(\{\varphi_{i\tau}^{u_i\tau}\}) \Rightarrow 0 \qquad \text{(IV.6)}$$

on the closure of the complement of C.

The decomposition of D_r into a difference $D_r^+ - D_r^-$ (see end of Sec. II) implies that $D(\{\varphi_{i\tau}^{u_i\tau}\}) \Rightarrow 0$ on any compact subset of u's that lie in the intersection of A_c and A_c^-. The assumptions (i) and (ii) thus imply (IV.5) for the set of u's that lie in the closure of the complement of the cone C^- opposite to C. The earlier arguments of this section then lead back to property (5). That is, the discontinuity can be derived directly from stronger physical requirements, instead of the procedures of Ref. 14.

Description of Ref. 12. In Ref. 12, a weaker factorization property was assumed in the case of the graph G with one intermediate particle M described previously (see Sec. I and Sec. A).

Particles C, B_1, C_1 were supposed to be displaced by a common 4-vector ρ and the quantity

$$H(\rho) = S^C(\rho) - \theta(\rho_o)D(\rho) \quad , \tag{IV.7}$$

(where the wave packets have been left implicit) was assumed to satisfy the condition

$$\int |H(\rho)|^2 \, d^4\rho < +\infty \quad . \tag{IV.8}$$

Eq. (IV.8) expresses the idea that $H(\rho)$ is small for large $|\rho_o|$. This is analogous to macrocausality for $\rho_o < 0$ and is a factorization property for causal configuration for $\rho_o > 0$; it is in effect opposed to the property $\int |D(\rho)|^2 \, d^4\rho = +\infty$.

Eq. (IV 8) is then shown to imply a pole-type behavior of the scattering function:

$$T \cong \frac{1}{2\pi} \frac{1}{2(\underset{\sim}{k}^2 + m^2)^{\frac{1}{2}}} \frac{T_1(\underset{\sim}{p}_{A_1}, \underset{\sim}{k}, \underset{\sim}{p}_A, \underset{\sim}{p}_B) T_2(\underset{\sim}{p}_{B_1}, \underset{\sim}{p}_{C_1}, \underset{\sim}{k}, \underset{\sim}{p}_C)}{k_o - (\underset{\sim}{k}^2 + m^2)^{\frac{1}{2}} + i\varepsilon} + T' , \tag{IV.9}$$

where k is the sum of the (on-mass-shell) 4-vectors p_A, p_B minus p_{A_1}, T_1 and T_2 denote the scattering functions of the processes $A + B \to A_1 + M$, $M + C \to B_1 + C_1$ and the possible singularities of T' in k are at most "square integrable."

V. S-Matrix Theory and Physical Space-Time Description

In Secs. IIID and IVA, we have seen how the physical region analytic structure of the S-matrix, including the discontinuity formula, implies corresponding fall off properties of the connected amplitudes for noncausal configurations and factorization properties for causal configurations.

In the present section, it is only some sort of smoothness rather than analyticity itself that is required, and corresponding approximations are used. In contrast to the foregoing work, we shall not specify the precise degree of smoothness or the precise bound on the background terms, but refer the reader to the original articles for these details.

A. Space-Time Wavefunctions in S-Matrix Theory

In this section, we consider smooth wave packets $\varphi_i(\underset{\sim}{p}_i)$ with compact supports in spheres of radius d centered at the points $\underset{\sim}{P}_i$.

Suppose $P \equiv \{P_i\}$ is not a positive α Landau point, and that d is sufficiently small. Then if the scattering function is sufficiently smooth in points $\{p_i\}$ determined by the supports of the φ_i, it can in first approximation be replaced by its value at P.

Consider the formal Fourier transform of the mass shell wave-function defined by [Eq. (III.2)]

$$f(x) = \int \varphi(\underset{\sim}{p})\, e^{ipx}\, \frac{d\underset{\sim}{p}}{2p_o} \quad .$$

A power series expansion about $\underset{\sim}{p} = \underset{\sim}{P}$ gives

$$p_o = P_o + \underset{\sim}{V}(\underset{\sim}{p} - \underset{\sim}{P}) + \mathcal{O}(d^2) \quad , \tag{V.1}$$

where $\underset{\sim}{V} = P_o^{-1}\underset{\sim}{P}$.

If $P_o^{-1}\underset{\sim}{x}_o\underset{\sim}{d}^2$ is small compared to one, $f(x)$ becomes a function of $\underset{\sim}{x} - \underset{\sim}{V}x_o$:

$$f(x) \approx \alpha \int \varphi(\underset{\sim}{p})\, e^{-i\underset{\sim}{p}(\underset{\sim}{x}-\underset{\sim}{V}x_o)}\, \frac{d\underset{\sim}{p}}{2p_o} \quad , \tag{V.2}$$

where α is a phase. On the other hand, for large x_o, the function $f(x)$ is mainly confined to the velocity cone defined by $\varphi(\underset{\sim}{p})$. Thus the quantity $P_o|f(x)|^2$ has roughly the behavior characteristic of a classical probability function corresponding to a momentum space distribution $|\varphi(\underset{\sim}{p})|^2$. This interpretation will be confirmed by the following results and those of Sec. B.

(a) Single Scattering[2)]

Consider initial wave packets $\varphi_i(\underset{\sim}{p}_i)$ and final wave packets $\psi_j(\underset{\sim}{q}_j)$. The connected amplitude of the process is

$$S^c(\{\varphi_i\},\{\psi_j\}) = \int \prod_i \varphi_i^*(\underset{\sim}{p}_i) \prod_j \psi_j(\underset{\sim}{q}_j) T(\ldots\underset{\sim}{q}_j\ldots,\ldots\underset{\sim}{p}_i\ldots)$$

$$\times\, \delta(\Sigma\, p_i - \Sigma\, q_j) \prod_i \frac{d\underset{\sim}{p}_i}{2p_{io}} \prod_j \frac{d\underset{\sim}{q}_j}{2q_{jo}} . \tag{V.3}$$

By using the equality,

$$\delta(\Sigma\, p_i - \Sigma\, q_j) = (2\pi)^4 \int \exp[i(\Sigma\, p_i - \Sigma\, q_j)x]\, d^4x \quad , \tag{V.4}$$

we obtain for small d and a nonpositive α Landau point $P \equiv \{P_i, Q_j\}$ the approximation

$$S^c(\{\varphi_i\},\{\psi_j\}) \approx (2\pi)^4\, T(\{P_i\},\{Q_j\}) \int \prod_{i,j} f_i(x) f_j^*(x)\, d^4x \ . \tag{V.5}$$

Thus, apart from a constant factor, the connected amplitude is given as the integral over space time of the product of the space-time wavefunctions f (or f*) at the same point x. In particular, this

amplitude is negligibly small unless all the regions where the various space-time wavefunctions are not negligible overlap.

The implication is that the function defined by Fourier transformation can be given a space-time interpretation. This interpretation is here derived from the smoothness of the momentum space scattering function, rather than being independently postulated. It is supported by the analysis given below of the pole contribution.

On the other hand, it seems impossible to give a space-time interpretation to the functions defined by Fourier transformation unless the scattering functions have some smoothness property.

(b) Incoming and Outgoing Waves

Consider the scattering $A + B \to A_1 + M$ and the "outgoing wave" of particle M defined in terms of the states of particles A, B, A_1 in (IV.3):

$$\varphi_M(\underset{\sim}{k}) = \int \varphi_A^*(\underset{\sim}{p}_A)\varphi_B^*(\underset{\sim}{p}_B)\varphi_{A_1}(\underset{\sim}{p}_{A_1})T(\underset{\sim}{p}_{A_1}, \underset{\sim}{k}, \underset{\sim}{p}_A, \underset{\sim}{p}_B)\delta(p_A + p_B - p_{A_1} - k) \times \frac{d\underset{\sim}{p}_A}{2p_{Ao}} \frac{d\underset{\sim}{p}_B}{2p_{Bo}} \frac{d\underset{\sim}{p}_{A_1}}{2p_{A_1 o}} \quad ,$$

where $k_o = (\underset{\sim}{k}^2 + m^2)^{\frac{1}{2}}$. Incoming waves can be defined analogously.

With the same smoothness assumptions as in the previous case, one can write[10]

$$\varphi_M(\underset{\sim}{k}) \approx (2\pi)^4 T \int f_A(x) f_B(x) f_A^*(x) e^{ikx}\, d^4x \quad , \qquad (V.7)$$

where $k_o = (\underset{\sim}{k}^2 + m^2)^{\frac{1}{2}}$ and T is evaluated at the point determined by the vectors $\underset{\sim}{P}_A$, $\underset{\sim}{P}_B$, $\underset{\sim}{P}_{A_1}$.

The corresponding space-time function is

$$f_M(y) = (2\pi)^4 T \int f_A(x) f_B(x) f_{A_1}^*(x) D(y - x)\, d^4x \quad , \qquad (V.8)$$

where

$$D(x) = \int e^{ikx}\, \delta(k^2 - m^2)\, \theta(k_o)\, g(\underset{\sim}{k})\, d^4k \quad , \qquad (V.9)$$

and $g(\underset{\sim}{k})$ is any function equal to one for all values of $\underset{\sim}{k}$ determined by the supports of the momentum space wavefunctions of A, B, A_1.

In the case where $K_o^{-1} d^2 |x_o - y_o|$ and $K_o^{-1} d |\underset{\sim}{x} - \underset{\sim}{y}|$ are small compared to one, the function $D(y - x)$ can be replaced by

$$D'(y - x) = \frac{(2\pi)^2}{2K_o}\, \delta\left(\underset{\sim}{y} - \underset{\sim}{x} - \frac{\underset{\sim}{K}}{K_o}(y_o - x_o)\right) \quad , \qquad (V.10)$$

and for large $(y_o - x_o)$, the function $f_M(y)$ is confined to the appropriate velocity cone.

(c) Multiple Scattering

Consider the scattering $A + B + C \to A_1 + B_1 + C_1$ previously described, where the particles C, B_1, C_1 are displaced by a 4-vector ρ and where the classical process $A + B \to A_1 + M$, $M + C \to B_1 + C_1$ is possible.

Then, near the one-particle pole, the scattering function of the process can be written

$$T \approx \frac{1}{2\pi} \frac{1}{2(\underline{k}^2 + m^2)^{\frac{1}{2}}} \frac{T_1(\underline{p}_{A_1}, \underline{k}, \underline{p}_A, \underline{p}_B)\, T_2(\underline{p}_{B_1}, \underline{p}_{C_1}, \underline{k}, \underline{p}_C)}{k_o - (\underline{k}^2 + m^2)^{\frac{1}{2}} + i\varepsilon} \qquad (V.11)$$

up to a smooth background term, where $k = p_A + p_B - p_{A_1}$, and T_1 and T_2 are the scattering functions--that will be supposed smooth--for the processes $A + B \to A_1 + M$, $M + C \to B_1 + C_1$.

The connected amplitude $S^C(\rho)$ of the process can now be written

$$S^C(\rho) \approx (2\pi)^4 T_1 T_2 \int f_A(x) f_B(x) f_{A_1}{}^*(x) f_C(y) f_{B_1}{}^*(y) f_{C_1}{}^*(y) D^C(y-x+\rho) \times d^4x\, d^4y \quad , \qquad (V.12)$$

where

$$D^C(x) = \int e^{ikx} \frac{1}{2(\underline{k}^2 + m^2)^{\frac{1}{2}}} \frac{1}{k_o - (\underline{k}^2 + m^2)^{\frac{1}{2}} + i\varepsilon} g(\underline{k})\, d^4k = \theta(x_o)\, D(x). \qquad (V.13)$$

If the products $f_A(x)\, f_B(x) f_{A_1}{}^*(x)$ and $f_C(y) f_{B_1}{}^*(y) f_{C_1}{}^*(y)$ have negligible values when $|x|$ and $|y|$ are larger than some length ℓ (where $|x| = (x_o{}^2 + x^2)^{\frac{1}{2}}$, then, for ρ_o positive and large compared to ℓ, the factor $\theta(y_o - x_o + \rho_o)$ can be replaced by one, which corresponds to replacing the pole in (V.11) by a mass shell δ function.

This result is analogous to that of Sec. IVA. It says that $S^C(\rho)$ can be evaluated as the transition amplitude of the process $M + C \to B_1 + C_1$, where the state of M is determined by the mass shell "outgoing wave" of M in the process $A + B \to A_1 + M$, or by the corresponding space-time function $f_M(y)$ determined in subsection (b). The latter describes the localization of the intermediate propagating particle M.

These results extend to the case when there are several poles.

It is worth emphasizing that they depend on the large (causal) separation of the overlap regions of the space-time wavefunctions, but not on the separation of the individual regions where these functions have nonnegligible values.

B. S-Matrix Theory and Classical Description of Interactions[13]

(a) Density and Efficiency Functions

When the information on an initial particle is not necessarily complete, it is represented by a density operator ζ that is Hermitian, positive definite as is $1 - \zeta$ and has trace one:

$$0 \leqq \zeta \leqq 1, \tag{V.14a}$$

$$\mathrm{Tr}\,\zeta = 1. \tag{V.14b}$$

The momentum space kernel $\zeta(\underset{\sim}{p},\underset{\sim}{p}')$ reduces to the product $\varphi^*(\underset{\sim}{p}')\varphi(\underset{\sim}{p})$ in the case of a pure state.

A one-particle measurement apparatus that selects an ensemble of final states is similarly represented by an efficiency operator F that satisfies (V.14a).

In nonrelativistic quantum mechanics, the Wigner's density function $g_W(\underset{\sim}{P},\underset{\sim}{x})$ is defined as

$$g_W(\underset{\sim}{P},\underset{\sim}{x}) = \int \zeta(\underset{\sim}{p},\underset{\sim}{p}')e^{i(\underset{\sim}{p}-\underset{\sim}{p}')\underset{\sim}{x}}\,\delta\left(\frac{\underset{\sim}{p}+\underset{\sim}{p}'}{2} - \underset{\sim}{P}\right)\frac{d\underset{\sim}{p}}{(2p_o)^{\frac{1}{2}}}\,\frac{d\underset{\sim}{p}'}{(2p_o')^{\frac{1}{2}}}. \tag{V.15}$$

Efficiency functions are defined similarly.

The Wigner's density function is not in general a positive function in contrast to a classical probability density. However, the following results hold:[13]

(i) When smeared over a phase-space region of at least unit volume ($\hbar = c = 1$), the Wigner's function is positive (and no greater than one):

$$0 \leqq g_W(\underset{\sim}{P},\underset{\sim}{x})^*\,\tau_{B_x}(\underset{\sim}{x})\,\tau_{B_P}(\underset{\sim}{P}) \leqq 1 \tag{V.16}$$

for any set (B_x, B_P) such that $B_x B_P \geqq \frac{1}{2}$, where * denotes a convolution and the τ's are Gaussians of widths B_x and B_P respectively:

$$\tau_{B_x}(\underset{\sim}{x}) = (2\pi B_x)^{-\frac{3}{2}}\exp\left[-\tfrac{1}{2}\frac{\underset{\sim}{x}^2}{B_x^2}\right].$$

(ii) Eq. (V.16) can be used to show that the function g_W is, in a good approximation, positive (and no greater than one) without smearing, if it is slowly varying in phase-space regions of volume large compared to one.

(iii) If the information on a particle is a classical distribution $\mu(\underset{\sim}{P},\underset{\sim}{x})$ up to quadratic errors A_x and A_P in position and momentum

respectively ($A_x A_B \geq \frac{1}{2}$) in the quantum case), then the function g_W has the form of the corresponding classical probability density:*

$$g_W(\underline{P},\underline{x}) = \mu(\underline{P},\underline{x}) * \tau_{A_x}(\underline{x})\, \tau_{A_P}(\underline{P}) \quad , \tag{V.17}$$

and is, in particular, positive.

If $\mu(\underline{P},\underline{x}) = \delta(\underline{P} - \underline{P}_o)\delta(\underline{x} - \underline{x}_o)$, and if $A_x A_P = \frac{1}{2}$, g_W is the Wigner's function of the pure state:

$$\varphi(\underline{p}) = (2\pi A_P)^{-\frac{3}{2}} \exp[i\underline{p}\underline{x}_o] \exp\left[-\frac{(\underline{p} - \underline{P}_o)^2}{2A_P^2}\right] \quad . \tag{V.18}$$

Although the Gaussian-type wavefunctions of Sec. III may contain a function χ, their Wigner's function also becomes positive in the $\gamma\tau \to \infty$ limit.

In Ref. 13, a density function of energy momentum and space-time that is better adapted to the relativistic case was defined by

$$j(P,x) = \int \zeta(\underline{p},\underline{p}') \exp[i(p - p')x]\, \delta\left(\frac{p + p'}{2} - P\right) \frac{d\underline{p}}{2p_o} \frac{d\underline{p}'}{2p_o'} \tag{V.19}$$

where $p_o = (\underline{p}^2 + m^2)^{\frac{1}{2}}$, $p_o' = (\underline{p}^2 + m^2)^{\frac{1}{2}}$. Efficiency functions were defined similarly.

From this definition, it is easy to show that j can be written (without approximation) as a function of the variables P and $\underline{x} - \underline{V}x_o$.

We shall consider kernels $\zeta(\underline{p},\underline{p}')$ that have compact supports in a sphere** of radius d in the variable $\underline{u} = \underline{p} - \underline{p}'$ (or contain a damping factor $\exp[-u^2/d^2]$).

If d^{-1} is large compared to the Compton wavelength $1/m$ of the particle, $j(P,x)$ is equal in good approximation to

$$\begin{aligned} j_F(P,x) &= [m^2 \int j(\lambda P,x)\, d\lambda]\, \delta(P^2 - m^2)\, \theta(P_o) \\ &= g_F(\underline{P},\underline{x} - \underline{V}x_o)\, \delta(P^2 - m^2)\, \theta(P_o) \quad , \end{aligned} \tag{V.20}$$

where λ is a scalar.

The function j_F--which is covariant, as is j itself--contains a mass shell δ function.

---

* According to information theory, the density operator ζ [resp. the probability $g(\underline{P},\underline{x})$ in the classical case] is determined by extremizing Tr ζ ln ζ (resp. the integral $\int g \ln g$).

** Correspondingly, the Wigner function is slowly varying in intervals small compared to d^{-1}.

The function $g_F(\underline{P},\underline{x})$ satisfies itself (when d^{-1} is large compared to $1/m$) the properties previously described for the Wigner's function.

Properties (i) and (ii) then ensure that the density and efficiency functions satisfy the properties of the classical probability densities and measurement efficiencies respectively, after appropriate smearing or, without smearing, under conditions that enforce that $\hbar$ as well as $\hbar/mc$ (= $1/m$) are negligible.

Property (iii) ensures that the Gaussian-type wave packets of Sec. III enjoy classical properties in the limit $\gamma\tau \to \infty$ ($\gamma\tau$ then becomes large compared to $1/m$).

One final remark: the quantity formally associated in this framework with a classical probability in space time is

$$F(x) = \int 2P_o j(P,x)\, d^4P \quad , \qquad (V.21)$$

the current associated with j being $j_\mu = 2P_\mu j$ ($\mu = 0,1,2,3$). In the approximation where $\hbar/mc$ is negligible, $F(x)$ is positive.

In the case of a pure state represented by a wavefunction $\varphi(\underline{p})$ with support in a sphere of radius d centered at $\underline{P}$, $F(x)$ is also equal (in the above approximation) to the functions $P_o|f(x)|^2$ and $|\tilde{\varphi}(x)|^2$, where f and $\tilde{\varphi}$ are defined in (III.2).

(b) Single Scattering*

Consider, for simplicity, a scattering $A + B \to A_1 + B_1$. The transition probability W of the process can be written (without approximation)

$$W = \int \zeta_A(\underline{p}_A,\underline{p}_A')\zeta_B(\underline{p}_B,\underline{p}_B')F_{A_1}^*(\underline{p}_{A_1},\underline{p}_{A_1}')F_{B_1}^*(\underline{p}_{B_1},\underline{p}_{B_1}')$$

$$\times\, T(\underline{p}_{A_1},\underline{p}_{B_1},\underline{p}_A,\underline{p}_B)T(\underline{p}_{A_1}',\underline{p}_{B_1}',\underline{p}_A',\underline{p}_B')\delta(p_A+p_B-p_{A_1}-p_{B_1})$$

$$\times\delta(p_A'+p_B'-p_{A_1}'-p_{B_1}')\prod_i \frac{d\underline{p}_i}{2p_{io}}\,\frac{d\underline{p}_i'}{2p_{io}} \quad , \qquad (V.22)$$

where the index i ranges over A, B, A_1, B_1.

Consider density and efficiency operators having smooth kernels with compact support in a sphere of radius d in the variable $\underline{u} = \underline{p} - \underline{p}'$ (or with damping factors $\exp[-\underline{u}^2/d^2]$ and suppose that the scattering function is smooth in the domain determined by these kernels.

* The formulas obtained in subsections b) and c) below,[13] under appropriate approximations, are particular cases of physical interest of the general formula[16] (without approximation):

$$W = \int \prod_{i=1}^{m+n} dP_i\, dx_i [\Pi j_i(P_i,x_i)]\hat{S}(P_1,\ldots P_{m+n},x_1\ldots x_{m+n})$$

where $\hat{S}$ is defined from the scattering function in a way similar to (V.19).

If we approximate the scattering functions involved in (V.22) by their values at the points $(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B)$, where $\underline{P}=(\underline{p}+\underline{p}')/2$, we have

$$W \approx \int j_A(P_A,x) j_B(P_B,x) E_{A_1}(P_{A_1},x) E_{B_1}(P_{B_1},x) (2\pi)^{-4} |T(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B)|^2 \times d^4x \prod_i d^4P_i \quad , \tag{V.23}$$

where j_A, j_B, E_{A_1}, E_{B_1}, are the density and efficiency functions of the the particles A, B, A_1, B_1.

This formula is precisely a classical type formula for a classical local scattering.

Suppose now that instead of replacing the scattering functions by their values at the points $\{\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B\}$, we keep also the first-order terms in the variables $\underline{u}_{A_1}, \underline{u}_{B_1}, \underline{u}_A, \underline{u}_B$ by writing

$$T^*(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B) T(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B) \approx |T(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, \underline{P}_B)|^2 \times \exp[i(\underline{u}\nabla_P\theta + \ldots)] \quad , \tag{V.24}$$

where θ is the phase of T (the derivatives of the modulus of T do not appear in first order).

Then the product of the density and efficiency functions in (V.23) are no longer taken at the same point x: relative displacements (that depend on the derivatives of θ) appear. This corresponds not to a local interaction, but to an interaction that takes place in a (small) space-time grain, and the corresponding formula is said to be of a coarse-grained type.[27]

(c) Multiple Scattering

The formula analogous to (V.7) is

$$\zeta_M(\underline{k},\underline{k}') \approx \int (2\pi)^4 |T(\underline{P}_{A_1}, \underline{P}_{B_1}, \underline{P}_A, -)|^2 \delta(P_A+P_B-Q_1-K) e^{iux} j_A(P_A,x) \times j_B(P_B,x) E_{A_1}(P_{A_1},x)\, dP_A\, dP_B\, dP_{A_1}\, dx \quad , \tag{V.25}$$

where $K = (k + k')/2$, $u = k - k'$, $k_o = (\underline{k}^2 + m^2)^{\frac{1}{2}}$, $k_o' = (\underline{k}'^2 + m^2)^{\frac{1}{2}}$.

It yields the formula analogous to (V.8) for the density function of the "outgoing" particle M:

$$j_M(K,y) = \int (2\pi)^4 |T|^2 \delta(P_A + P_B - P_{A_1} - K) j_A(P_A,x) j_B(P_B,x) E_{A_1}(P_{A_1},x) \times \frac{1}{2K_o} \hat{D}(K, \underline{y} - \underline{x} - \frac{\underline{K}}{K_o}(y_o - x_o))\, dP_A\, dP_{A_1}\, dx \quad , \tag{V.26}$$

where

$$\hat{D}(K,\underset{\sim}{x}) = \int e^{i\underset{\sim}{u}\underset{\sim}{x}}\, \delta\left(K^2 - m^2 + \tfrac{1}{2}\left(\left(\frac{\underset{\sim}{K}}{K_o}\underset{\sim}{u}\right)^2 - \underset{\sim}{u}^2\right)\right) \theta([\underset{\sim}{K} + \underset{\sim}{u})^2 + m^2]^{\frac{1}{2}}$$

$$\times\, \theta([\underset{\sim}{K} - \underset{\sim}{u})^2 + m^2]^{\frac{1}{2}})\, d\underset{\sim}{u}, \qquad (V.27)$$

and, if $1/m$ is small compared to d^{-1}, it can be replaced in Eq. (V.26) by

$$\hat{D}'(K, y - x) = \delta(K^2 - m^2)\, \theta(K_o)\delta\left(\underset{\sim}{y} - \underset{\sim}{x} - \frac{\underset{\sim}{K}}{K_o}(y_o - x_o)\right). \qquad (V.28)$$

In the case of the scattering $A + B + C \to A_1 + B_1 + C_1$, previously described, the formula obtained in Ref. 13 then says that for large ρ_o positive the transition probability of the process is equal to the transition probability of the process $M + C \to B_1 + C_1$, where the momentum space density kernel of M is determined by (V.25), and its density function is determined by (V.26) and (V.28).

The formula obtained in the latter case is of a classical type, associated with the description of the process $A + B + C \to A_1 + B_1 + C_1$ in terms of the two successive interactions $A + B \to A_1 + M$, $M + C \to B_1 + C_1$.

C. S-Matrix Theory of Measurements[10]

The factorization properties for causal configurations in the case of a graph G with no closed loop, and the related results of this section are crucial for an S-matrix theory of measurements: by virtue of these results, the initial and final wavefunctions* of some process can be interpreted as the outgoing waves of earlier processes and the incoming waves of later processes respespectively. These earlier and later processes are interpreted as the processes whereby the initial and final wavefunctions are measured.

The pole contributions are for practical purposes the only contribution when the distances between the locations of the various processes are macroscopic; one need not make the measurements of the initial and final particles in the "infinite past" or "infinite future."

The pole factorization property is thus a key element of the S-matrix theory of measurements. The significance of the mass shell constraint emerges clearly; if one performs measurements in such a way that the pole contribution is, for practical purposes, the only contribution, then the processes of measurement are effectively disengaged from the process being measured, by virtue of the factorization property of the residue.

---

* Respectively density or efficiency operators.

VI. Conclusion

We have described in this work a variety of relationships between physical region analyticity properties suggested by perturbation theory and macroscopic space-time properties suggested by classical ideas. An effective equivalence has been established between these physical region analyticity properties and the asymptotic behavior of transition amplitudes in a certain macroscopic limit. Various approximate but not only asymptotic space-time properties have also been derived from these same analyticity properties.

These results can be used in either direction. If the analyticity properties of the S-matrix are considered the fundamental assumptions of S-matrix theory, then the results actually provide for a derivation of the (macroscopic) space-time structure of physical phenomena from principles that do not explicitly contain the notions of space or time: macroscopic space-time can thus be understood without introducing any a priori space-time substructure.

In the other direction, if one assumes that certain classical ideas about the space-time structure of phenomena are valid in the appropriate asymptotic limit [and specifically that dynamical effects can be considered to be transmitted over macroscopic distances only by (stable) physical particles, apart from effects of finite range], then one can derive the corresponding physical region analyticity properties of the (mass shell) S-matrix. Analyticity at all points that do not lie on positive α Landau surfaces and the plus $i\varepsilon$ rule at the positive α points can be derived from conditions on the transition probabilities themselves for noncausal configurations of groups of particles (macrocausality). The discontinuity formulas can also be derived from conditions on transition amplitudes for causal configurations, but these have no equivalent formulation in terms of transition probabilities. It thus seems difficult to consider that they represent a basic physical principle* and one may instead derive[14] the general discontinuities around the physical region singularities from the above analyticity properties derived from macrocausality and analogous properties of the connected parts of S^{-1}.** (See end of Sec. II.)

---

* This conclusion is confirmed by the consideration of graphs with multiple lines.

** The properties of S^{-1} can be considered as consequences either of unitarity (if the latter is admitted) or of an extended version of macrocausality.

Regularity properties at the exceptional type II points are also needed. They have not been so far derived from macrocausality. In the following viewpoints, they can be considered as consequences of "maximal analyticity."

In the general framework of analytic S-matrix theory, the analyticity thus derived in the physical region can be considered the primitive analytic structure that is extended by the principle of maximal analyticity away from the physical region.

The analytic structure derived from the principle of maximal analyticity alone in the physical region itself is consistent with that derived from macrocausality. From the requirement that all singularities be generated by the phase-space singularities, one immediately derives[28] that all singularities lie on Landau surfaces. It is not yet known whether the only way to satisfy this condition plus unitarity (or $SS^{-1} = I$) is with all singularities confined to positive α surfaces and satisfying the $i\varepsilon$ rules.* Thus, at present, one needs to invoke macrocausality to justify these constraints. However, the macrocausality condition need not be independently postulated in that approach; it can be derived at the non-Landau points from analyticity. This provides a basis for the validity of the corresponding classical ideas and thus for a general statement of the condition.

A. Local Field Theory and the S-Matrix

We end this conclusion with a discussion of related results of local field (or local procedures) theory. Microscopic causality (and ad hoc spectrum conditions plus technical assumptions (e.g. distributions)) imply the existence of asymptotic states in the $t \to \pm\infty$ limit and of the operator S between the corresponding spaces $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$ (Haag-Ruelle scattering theory).

The question arises to what extent the microscopic space-time structure of local field theory implies the macroscopic space-time structure described in this paper, in particular macrocausality, or the equivalent physical region analyticity properties of the (mass shell) S-matrix.

On the one hand, various factorization properties (for noncausal and causal configurations) of the type described in this paper have been obtained in Ref. 30. (These results generalize those of Ref. 31.) The conclusion of Ref. 30 seems, however, to be that to obtain general results, it is necessary to use (besides the standard Wightman axioms)

* In Ref. 29, there has been an attempt to derive not only the discontinuity formulas but also the analyticity on the mixed α surfaces. This attempt is, however, vitiated by a technical difficulty. As in Ref. 14, the $i\varepsilon$ rule at the positive α Landau points is part of the assumptions. (If one assumes that there is a plus or minus $i\varepsilon$ rule at all points, then it might be possible to derive from arguments of internal consistency (see Sec. II) that the sign must be identical at all points.)

some assumptions of regularity or "permanence of smoothness" of some field functions. (Partial results, such as the macrocausality condition of Sec. III, in the case when the velocity cones of the initial particles all intersect at times later than those of the final particles, can be derived within the framework of the Wightman axioms.)

On the other hand, substantial results in the direct derivation of analyticity properties have only been obtained so far for $2 \to 2$ processes. In the linear program, the scattering function $T(s,t)$ (where s is the energy and t the momentum transfer) is shown to be analytic in domains that go much beyond the physical region (and are sufficient to prove crossing, etc.).

However, although the domain of analyticity in s admits the real axis as a boundary, no analyticity has been obtained for s real (in the physical region). (The analyticity obtained in t has been shown in Ref. 8 to be equivalent to a macroscopic locality condition that has been described in Sec. I.)

In Ref. 32, analyticity in the elastic physical region of a $2 \to 2$ process has been derived in the nonlinear program. However, besides the Wightman axioms, it seems necessary to use also here some assumptions of regularity or permanence of smoothness of some field functions.

Appendix I. Fock Space of States and the S-Matrix

A. Basic Assumptions and General Properties of the S-Matrix

In this appendix we describe the formulation of S-matrix theory that starts with the operator S that maps a Fock space $\mathcal{H}$ of free particle states onto itself. As in Sec. II, we consider for simplicity a theory with no superselections rules and with only spinless particles.

Initial and final systems are assumed to be represented by rays (that is unit vectors up to a phase) in an Hilbert space $\mathcal{H}$ of "free particle states." Moreover, each initial or final system is assumed to determine a unique ray. Conversely, every ray in $\mathcal{H}$ is assumed to represent a unique physically realizable initial (or alternatively final) system.

Each initial (respectively final) system is considered to be the initial (respectively final) state of a physical system "sub-specie aeternitatis" (that is "with all its evolution"), that has also a unique final (respectively initial) state. This defines a one to one correspondence between rays of $\mathcal{H}$, which we denote $\hat{S}$, that takes the ray φ_{in} associated with the initial state of a system "sub-specie aeternitatis" to the ray φ_{out} associated with its final state. Finally, initial (respectively final) probabilities $Pr(\varphi,\psi)$ that an initial (respectively final) state φ occurs in the initial (respectively final) state ψ are defined and

assumed to be equal to the ray product $\varphi \cdot \psi$ (i.e. the unitary inner product $|\langle\varphi|\psi\rangle|^2$ of unit norm representative vectors).

The transition probability $W(\varphi,\psi)$ from an initial state φ to a final state ψ is then equal to $Pr(\hat{S}\varphi,\psi) = \psi \cdot \hat{S}\varphi$.

1. Hilbert Space of Free Particle States

The explicit construction of the Hilbert space $\mathcal{H}$ of states may be outlined as follows:[33] the Poincaré group--or at least its proper orthochronous part $P_+^\uparrow$ is asserted to be a symmetry group:

$$Pr(\varphi,\psi) = Pr({}^g\varphi, {}^g\psi) \tag{AI.1}$$

or equivalently

$$\varphi \cdot \psi = {}^g\varphi \cdot {}^g\psi$$

(for all g that belong to $P_+^\uparrow$ and all φ,ψ) where ${}^g\varphi, {}^g\psi$ are obtained from φ,ψ by the transformation g.

Eq. (AI.1) implies the existence of a unitary* representation of the covering group $\bar{P}_+^\uparrow$ of $P_+^\uparrow$ in $\mathcal{H}$.

One-particle states are then assumed to span spaces $\mathcal{H}_{m,s}$ of irreducible unitary representations. (These are labeled by two numbers the mass m--that we restrict to be strictly positive--and the spin s (which is integer or half integer).)

For spinless particles, $\mathcal{H}_{m,s}$ is the space of square integrable functions $\varphi(\vec{p})$ with respect to the measure $\frac{dp}{2(p^2+m^2)^{\frac{1}{2}}}$ and the transformation law is:

$$(a,\Lambda)\varphi(p) = e^{ipa}\,\varphi(\Lambda^{-1}p) \tag{AI.2}$$

where a is a space-time translation and Λ a Lorentz transformation.

For a theory with at most a countably infinite number of types of (stable) particles, labeled by the integers $t = 1,2\ldots$, the Hilbert space $\mathcal{H}$ of free particle states is then asserted to be the Fock space associated with these particles.

$$\mathcal{H} = \oplus\, \mathcal{H}_{\underline{n}} \quad (\underline{n} = n_1 \ldots n_t \ldots) \tag{AI.3}$$

where $\oplus$ denotes a direct sum of subspaces $\mathcal{H}_{\underline{n}}$, the latter being tensor products of one particle spaces,** that are symmetrized or

---

* Antiunitarity is included by an assumption of continuity of U(g) with respect to g.

** A well defined order of the types of particles is adopted.

antisymmetrized with respect to the interchange of identical particles.*

At this stage a final assumption on the correspondence between physical (initial or final) systems and rays in $\mathcal{H}$ may be added: a composite system of $\underset{\sim}{n}$ particles that are represented by rays φ_i, or equivalently wavefunctions $\varphi_i(\underline{p}_i)$ (up to a phase), is asserted to be represented (up to a phase) by the (symmetrized or antisymmetrized) product of the individual wavefunctions:

$$\varphi_{\underset{\sim}{n}}(P_{\underset{\sim}{n}}) = \mathfrak{S}\left(\prod_i \varphi_i(\underline{p}_i)\right)$$

(This assumption will be discussed at the end of Sec. B.)

2. The S-Matrix-General Properties

Unitarity (including linearity) and energy momentum conservation can be derived[33] from two physical requirements; the conservation by $\hat{S}$ of probabilities ($\Pr(\varphi,\psi) = \Pr(\hat{S}\varphi,\hat{S}\psi)$), and Poincaré invariance of transition probabilities:

$$\varphi\cdot\psi = \hat{S}\varphi\cdot\hat{S}\psi \tag{AI.5}$$

$$\psi\cdot\hat{S}\varphi = {}^{g}\psi\cdot\hat{S}\,{}^{g}\varphi \tag{AI.6}$$

Eq. (AI.5) implies that $\hat{S}$ can be represented by a unitary or antiunitary operator S of $\mathcal{H}$.

Eq. (AI.6) implies that

$$U^{-1}(g)\, SU(g) = \omega(g)\, S \tag{AI.7}$$

where $\omega(g)$ is a one-dimensional representation of $\bar{P}^{\uparrow}_{+}$. The one-dimensional representations of $\bar{P}^{\uparrow}_{+}$ being constant, Eq. (AI.7) yields the commutation of $U(g)$ with S.

This property, when restricted to small space-time translations a, for which $U(a) \approx 1 + i P^{\mu}a_{\mu}$ yields

$$SPS^{+} = P \qquad \text{if S is unitary}$$

$$SPS^{+} = -P \qquad \text{if S is antiunitary}$$

If S were antiunitary then multiplication by complex conjugation would give a unitarity S satisfying the first of these two relations, which is necessary if we want to identify P^{o} with physical energy.

* Even though we only consider spinless particles, we will not assume here the spin statistics connection.

3. Comparison with the Formalism of Sec. II

In the framework of the formalism described in this Appendix, the functional S_{mn} of Sec. II appears as a derived quantity, namely the restriction of S to initial systems of m particles and final systems of n particles.

The unitarity of the S-matrix and Eq. (AI.4) imply the linearity properties of the functional S_{mn} with respect to each one particle wave packet, described in Sec. II, and the commutation of U(a) and S implies the energy momentum conservation property there stated. (Energy momentum conservation implies conversely translational invariance.)[19] Unitarity can also be derived[1] from linearity and a weak requirement of conservation of probabilities ($|\varphi| = |S\varphi|$).

In the method of this appendix both the linearity and unitarity of S are derived. But it is assumed that the initial (respectively final) transition probabilities $Pr(\varphi,\psi)$ are given by the unitary inner product $|\langle\varphi|\psi\rangle|^2$. And in place of the weak conservation of probabilities ($|\varphi| = |S\varphi|$) one needs the conservation of transition probabilities (AI.5).

Finally, the fact that the scattering functions are either symmetric or antisymmetric, which is a consequence of analyticity,[1] is built into the Hilbert space formalism described here.

It is easy to show that conversely an operator S of a Fock space $\mathfrak{H}$ can be formally reconstructed from the functionals S_{mn}, by using the boundedness property (II.1') (extended to arbitrary wave packets φ_m and ψ_n of m initial and n final variables respectively). See also Ref. 34.

These formal constructions do not require, however, in the framework of Sec. IIA, that arbitrary vectors in $\mathfrak{H}$ (that can have components in various $\mathfrak{H}_n$) represent actual physical systems.*

If arbitrary vectors in $\mathfrak{H}$ do represent physical systems, certain interesting consequences can be derived. Specifically, in Sec. B below, the ray in $\mathfrak{H}$ that represents a composite system will be determined from weak physical assumptions in the case when the individual states have components in various $\mathfrak{H}_n$ (the result obtained is precisely that given by the usual creation operator formalism).[36]

From this result, the phase factors occurring in the cluster decomposition of S (Eq. II.2) can be determined either from the space-time condition of Ref. 5 (see Sec. I) or from a slightly extended formulation of macrocausality. This is discussed in Appendix IIA.

---

* It may also be noted that whereas in Sec. IIA physical state vectors could be assumed to span only dense subspaces, such as $\mathfrak{D}$ of $\mathfrak{H}_{m,o}$, this is hardly possible in this Appendix because the operator S does not always conserve smoothness. (It can be proved that S conserves smoothness, but only away from the Landau singularities.)[35]

B. States of Composite Systems

In this section, we briefly describe with some slight modifications the results of Ref. 36 (a comparison with Ref. 36 itself is given at the end).

a) We introduce first the physical requirement that initial (respectively final) states of particle that have nonoverlapping velocities do not "interfere" (this idea is contained in particular in the assumption on space-time localization of Sec. IIIA).

Consider states φ,φ'; and ψ,ψ' that are represented (up to a phase) by sets of wavefunctions $\{\varphi_m(P_m)\}$, $\{\varphi'_{m'}(P_{m'})\}$, and $\{\psi_m(P_m)\}$, $\{\psi'_{m'}(P_{m'})\}$ such that the velocities of the φ's do not overlap those of the ψ's.

In this case we assume a factorization of the probability that the composite system ξ of φ and ψ occurs in the composite system ξ' of φ' and ψ':

$$\xi\cdot\xi' = (\varphi\cdot\varphi')(\psi\cdot\psi') \tag{AI.8}$$

By the methods of Ref. 36 it can then be shown that the ξ's can be represented (up to a phase) by a bilinear mapping* of representative vectors $|\varphi\rangle$ and $|\psi\rangle$.

This mapping can then be shown[36] to have a unique bilinear extension for states with overlapping velocities (in any pair $\mathcal{H}_m$, $\mathcal{H}_n$).

It likewise follows from the physical requirement (AI.8) that any composite system can generally be represented by a vector obtained by a multilinear mapping from the set of vectors that represent the individual systems.

b) In the case when the individual component states have fixed numbers of particles of each type (i.e. are represented by unique wavefunctions $\varphi_m(P_m)$ and $\psi_n(P_n)$) a slight extension of Eq. (AI.4) says that ξ can be represented up to a phase by the wavefunction

$$\xi_{\underset{\sim}{m}+\underset{\sim}{n}}(P_{\underset{\sim}{m}+\underset{\sim}{n}}) = \mathcal{S}(\varphi_{\underset{\sim}{m}}(P_{\underset{\sim}{m}})\psi_{\underset{\sim}{n}}(P_{\underset{\sim}{n}})) \tag{AI.9}$$

The phase could a priori depend on the states $\varphi_{\underset{\sim}{m}}$ and $\psi_{\underset{\sim}{n}}$. The linearity property of a) however, only allows a dependence on the numbers $\underset{\sim}{m}$ and $\underset{\sim}{n}$ of particles of each type.

This in turn implies, by using again the linearity property, that for arbitrary φ and ψ, ξ can be represented by the set of wavefunctions:

---

* This mapping is in fact also biunitary for nonoverlapping states:

$$\langle\xi|\xi'\rangle = \langle\varphi|\varphi'\rangle\langle\psi|\psi'\rangle$$

$$\xi_{\underset{\sim}{m}}(P_{\underset{\sim}{m}}) = \sum_{\underset{\sim}{n}} \alpha(\underset{\sim}{n},\underset{\sim}{m}-\underset{\sim}{n})\ (\varphi_{\underset{\sim}{n}}(P_{\underset{\sim}{n}})\psi_{\underset{\sim}{m}-\underset{\sim}{n}}(P_{\underset{\sim}{m}-\underset{\sim}{n}})) \qquad \text{(AI.10)}$$

with unknown phases α.

c) The phase factors α are determined in Ref. 36 by using the physical identity of the composite states $\{\varphi$ and $\psi\}$ and $\{\psi$ and $\varphi\}$; and also $\{\varphi$ and $\{\psi$ and $\chi\}\}$ and $\{\{\varphi$ and $\psi\}$ and $\chi\}$, the corresponding rays being then identical.

The phases α are then shown to be reducible to

$$\alpha(\underset{\sim}{m},\underset{\sim}{n}) = \sum_{\underset{\sim}{n}} (-1)^{\sum_{s>t}^{f} n_s(m_t-n_t)} \qquad \text{(AI.11)}$$

where $\sum^{f}$ denotes summation over fermions only (that is the particles for which antisymmetrization is used). This is precisely the result obtained from the creation annihilation formalism with the stipulation that all operators commute, except that fermion operators all anti-commute with one another.

1. Comparison with Ref. 36

In Ref. 36 itself, Eq. (AI.8) was assumed to hold if at some time t_o, space-time wavefunctions of the φ's and ψ's have disjoint compact supports.

Eq. (AI.9) is then itself derived from the requirement that, at least for such nonoverlapping states, the probability of observing given momenta and positions must be the same for the composite state ξ of φ and ψ and for φ and ψ separately.

If for instance φ and ψ are one particle states, it is required that that:

$$|\xi(\underset{\sim}{p}_1,\underset{\sim}{p}_2)|^2 = |\varphi(\underset{\sim}{p})|^2\,|\psi|\underset{\sim}{p}_2|^2 \qquad \text{(non identical particles)}$$

$$|\xi(\underset{\sim}{p}_1,\underset{\sim}{p}_2)|^2 = |\varphi(\underset{\sim}{p}_1)|^2\,|\psi(\underset{\sim}{p}_2)|^2 + |\varphi)\underset{\sim}{p}_2)|^2\,|\psi(\underset{\sim}{p}_1)|^2$$

(identical particles) (AI.12)

and similar equations for the space-time wavefunctions.

This approach, though satisfactory in non-relativistic quantum mechanics, presents difficulties in relativistic S-matrix theory, where space-time wavefunctions cannot be strictly associated with probabilities. (Note also that space-time wavefunctions that have disjoint supports at time t_o, do not have this property for $t_o' \neq t_o$ arbitrarily close to t_o.)

Appendix II. Connectedness Structure of the S-Matrix

In this appendix we derive the fall off property (III.7) of the connected amplitudes, on compact subsets of A_c, from the factorization property:

$$\hat{R} = W - \prod_k W_k \Rightarrow 0 \qquad \text{(AII.1)}$$

uniformly on compact subsets of displacements that contain no causal displacement associated with a network of particles involving a connection between the specified groups k. The wavefunctions $\varphi_{i\tau}$ have the form (III.4) with $\gamma \geqq 0$.

We will also need a slight generalization to cover cases when $\varphi_{i\tau}$ is, instead, a sum $\lambda\psi_{i\tau}+\mu\psi'_{i\tau}$ of functions of the form (III.4) that may be centered at different $\underline{P}_i$ and have different values of γ. This generalization also follows from the physical ideas of Sec. III, provided the groups k are still non-causal with respect to each other for an arbitrary combination of $\psi_{i\tau}$ and $\psi'_{i\tau}$. Moreover the fall off should be uniform on the set λ, μ such that $|\lambda|^2 + |\mu|^2 = 1$. In Sec. A below, the ψ and ψ' will be chosen such that the above condition is satisfied.

Eq. (AII.1) implies the existence of a phase factor β which may depend on the wavefunctions $\varphi_{i\tau}{}^{u_i\tau}$ and is not uniquely determined in general such that:

$$|R| < |\hat{R}|^{\frac{1}{2}}$$

where

$$R = S - \beta \prod_k S_k \qquad \text{(AII.2)}$$

and S and S_k denote the transition amplitudes of the process and of the various subprocesses k (the wavefunctions $\varphi_{i\tau}{}^{u_i\tau}$ are left implicit).

Eqs. (AII.1) and (AII.2) imply that $R \Rightarrow 0$. We will show in Sec. A below that a fall off of this type can still be obtained with a phase β independent of the $\varphi_{i\tau}{}^{u_i\tau}$. This in turn will imply (Sec. B) the condition (III.7) if we exclude some exceptional points $P \equiv \{P_i\}$. The latter are those for which there exists a displacement $u \equiv \{u_i\}$ and two different corresponding causal diagrams the connected parts of which link together different groups of particles. These points belong in particular to the class of type II points of Sec. IIIC.

A. Factorization Properties of Transition Amplitudes

Consider a process in which the wavefunction $\varphi_{i\tau}$ of particle i is a sum $\lambda\psi_{i\tau} + \mu\psi'_{i\tau}$ with:

$$\lambda = \prod_k S_k(\psi'_{i\tau})$$

$$\mu = -\prod_k S_k(\psi_{i\tau}) \qquad \text{(AII.3)}$$

where the dependence of the S_k on the other wavefunction and on the displacements is left implicit.

The linearity of the S-matrix implies that:

$$\lambda\mu[\beta(\psi_{i\tau})-\beta(\psi'_{i\tau})] = -\lambda R(\psi_{i\tau})-\mu R(\psi'_{i\tau})+R(\lambda\psi_{i\tau}+\mu\psi'_{i\tau}) \qquad \text{(AII.4)}$$

We want to compare the values of $\beta(\psi_{i\tau})$ and $\beta(\psi'_{i\tau})$. If $|\lambda|$ for instance is smaller than $|\hat{R}(\psi'_{i\tau})|^{\frac{1}{4}}$, then $|R(\psi'_{i\tau})|$ is itself smaller than $3|\hat{R}(\psi'_{i\tau})|^{\frac{1}{4}}$ for any value of β. Since $|\hat{R}|^{\frac{1}{4}} \Rightarrow 0$, we will restrict our attention to the case when $|\lambda|$ and $|\mu|$ are both larger than the corresponding $|\hat{R}|^{\frac{1}{4}}$.

In this case we obtain:

$$|\beta(\psi_{i\tau}) - \beta(\psi'_{i\tau})| \leqq 3|\hat{R}|^{\frac{1}{4}} \qquad \text{(AII.5)}$$

where $|\hat{R}|$ is the maximum value of $|\hat{R}(\lambda\psi_{i\tau}+\mu\psi'_{i\tau})|$ for $|\lambda|^2+|\mu|^2 = 1$. The slight generalization of (AII.1) (discussed below it) entails that $|\hat{R}|^{\frac{1}{4}} \Rightarrow 0$. The the phase difference $\Rightarrow 0$ for $\psi_{i\tau}$ and $\psi'_{i\tau}$ in the class considered.

By using a chain of relations of this type, we may moreover show that a similar result holds if for all i the wavefunctions $\psi'_{i\tau}$ are chosen different from the $\psi_{i\tau}$.

This result will be used later to prove that β can be chosen independent of the wavefunctions $\psi_{i\tau}$. However we must first show that β can be chosen independent of the displacements $\{\tau u_i\}$.

Consider the case where $\psi'_{i\tau} = \psi_{i\tau}{}^{\rho u_i}$ with $0 \leqq \rho \leqq 1$. Eq. (AII.6) yields:

$$\beta(\{\psi_{i\tau}{}^{u_i\tau}\}) - \beta(\{\psi_{i\tau}{}^{u_i(\tau+\rho)}\}) \quad 0 \qquad \text{(AII.7)}$$

(This fall off is uniform for $0 \leqq \rho \leqq 1$.)

Eq. (AII.7) in turn implies[37] that, for a fixed set $\{u_i\}$, β has a

limit β_0 in the $\tau \to \infty$ limit.* More precisely:

$$\beta(\tau) - \beta_0 \to 0 \tag{AII.8}$$

where $\to$ denotes a rapid fall off in the case $\gamma = 0$ and a bound of the type:

$$|\beta(\tau) - \beta_0| < \frac{1}{\gamma} C e^{-\alpha\gamma\tau} \tag{AII.9}$$

for $\gamma > 0$.

In Ref. 37 it is moreover proved that (AII.8) still holds uniformly on sufficiently small compact subsets of u's around a given $u_0 \equiv \{(u_0)_i\}$.

By using a chain of such compact subsets, it is then possible to prove that the limit β_0 is independent of the displacements $\{\tau u_i\}$.

Then Eq. (AII.6) can be used again to show that β_0 is independent of the wavefunctions, and we obtain the factorization property:

$$S - \beta_0 \prod_k S_k \to 0 \tag{AII.10}$$

where β_0 can only depend on the numbers of particles of each type in each group k. If we consider different subdivisions $\{k\}$ into groups k, we will denote the corresponding phases by $\beta_{\{k\}}$.

The various $\beta_{\{k\}}$ must satisfy certain constraints. Consider a subdivision $\{k\}$ and a subdivision $\{\ell\}$ into groups ℓ each of which is composed of some of the groups k. Since there is still no causal connection between the groups ℓ or between the groups k that compose a given ℓ, the corresponding factorization properties imply that

$$\beta_{\{k\}} = \beta_{\{\ell\}} \prod_\ell \beta_{\{k\}_{k \in \ell}} \tag{AII.11}$$

The $\beta_{\{k\}}$ just defined are the ones that must be used in Eq. (II.2) to define the connected parts of the S-matrix.

If we admit the results of Appendix IB, and a new slight extension of the macrocausality condition to cover cases involving arbitrary

* This result is due to the inequality

$$|\beta(\tau')-\beta(\tau)| \leq |\beta(\tau')-\beta(\tau'-\rho)|+\ldots+|\beta(\tau'-(n-1)\rho)-\beta(\tau)|$$

$$\leq \int_{\tau_0}^{\infty} |\beta(\tau')-\beta(\tau'-\rho)|\,d\tau'$$

where $\tau' \geq \tau \geq \tau_0$, $\tau' - \tau = n\rho$.

initial and final states (with components in various $H_{\underline{n}}$) in each subprocess k, then the linearity of the S-matrix and of the product that generally represents composite states implies moreover that all the $\beta_{\{k\}}$ are equal to a single constant β_o (up to a sign*), which may be chosen equal to one.[5]

B. Fall-off Properties of the Connected Amplitudes

A proof by induction will be used: Eq. (III.7) holds for $S^c_{11} = S_{11}$. We will assume that it holds for all $S^c_{m',n'}$, where $m' \leq m, n' < n$ or $m' < m, n' \leq n$ and will prove it for S^c_{mn} (we leave implicit the dependence of the S's or S_c's on the (displaced) wavefunctions).

We consider points P that do not belong to the class of exceptional points that have been defined above and functions χ with sufficiently small supports so that they contain no such point.

Consider a compact subset Σ of A_c. It may contain causal u's corresponding to (non connected) networks of particles. We then divide Σ into a finite number of compact subsets $\Sigma_{\{\hat{k}\}}$ in each of which either $S \Rightarrow 0$ or Eq. (AII.10) holds for specified groups $\hat{k}$. We will prove that $S^c_{mn} \to 0$ uniformly on each $\Sigma_{\{\hat{k}\}}$. Since the number of $\Sigma_{\{\hat{k}\}}$ is finite, it will ensure that $S^c_{mn} \to 0$ uniformly on Σ.

By definition of the connected parts (Eq. (II.2)), S_{mn} is a sum of terms that we classify into three categories: 1) S^c_{mn} itself, 2) all the terms associated with subdivisions $\{k\}$ that are either identical to, or subdivisions of the subdivision $\{\hat{k}\}$ considered and 3) remaining terms.

Due to (AII.11) the terms in 2) combine to give precisely the term $\beta_{\{\hat{k}\}} \prod_{\hat{k}} S_{\hat{k}}$. The various terms in 3) are products of $S^c_{m_k n_k}$. By virtue of the exclusion of the exceptional points defined above, at least one of the $\{u_i\}_{i \in k}$ belongs to the corresponding $A_c(\{\varphi_{i\tau}\}_{i \in k})$ and the corresponding $S^c_{m_k n_k} \Rightarrow 0$. (The other terms are bounded by one by virtue of (II.1').) A comparison with Eq. (AII.10) then yields that $S^c_{mn} \to 0$.

One final remark: the symbol $\to$ (see (AII.9)) replaces the symbol $\Rightarrow 0$ that is expected. This is due to the method used in Sec. A. However this slight weakening of the fall off property does not affect the derivation of the analyticity properties**given in Sec. III. Thus the stronger fall off $\Rightarrow$ of (III.7) can be obtained using the results of Sec. IIIE.

---

* If fermions are involved, $\beta_{\{k\}}$ is equal to one or minus one. This latter case arises because of Eq. (AI.11) and because a general subdivision $\{m_k, n_k\}$ of $\{m,n\}$ cannot be written in the same order as $\{m,n\}$.

** Some slight but non essential modifications are needed in the proof.

1. Comparison with Ref. 5

In Ref. 5 common displacements d_k of the particles involved in each group k were considered and $\hat{R} = W - \prod_k W_k$ was assumed to fall off in the $\tau \to \infty$ limit. There was no assumption on the fall off rate.

In the present work the particles in each k are no longer displaced by a common d_k. There is no assumption on $\hat{R}$ in the case of causal configurations, and there is an assumption of strong fall off in the case of non-causal configurations of the groups k.

In Ref. 5, it was proved that S^c_{mn} vanishes when $\tau \to \infty$ if all the d_k are not identical. This in turn implies that the momentum space kernel of S^c_{mn} contains no partial energy momentum conservation law δ-function. But the fall off condition is not sufficient to obtain analyticity.

Appendix III. Fall Off Properties of Space Time Wavefunctions

Consider the wavefunctions (III.4): $\varphi_{(\gamma\tau)}(\underset{\sim}{p}) = \chi(\underset{\sim}{p})e^{-(\underset{\sim}{p}-\underset{\sim}{P})^2\gamma\tau}$ where χ is analytic in a neighborhood of $\underset{\sim}{P}$ and γ is positive:

We study in this appendix the fall off properties of the space time wavefunctions defined in (III.2).

We will adopt the notation $\hat{f}(x)$ for either one of the functions f and $\tilde{\varphi}$ and the notation $\hat{\chi}(\underset{\sim}{p})$ for the functions $\chi(\underset{\sim}{p})/2p_o$ and $\chi(\underset{\sim}{p})/(2p_o)^{\frac{1}{2}}$ respectively.

We will prove the exponential fall off of $\hat{f}(u\tau)$ in the $\tau \to \infty$ limit for a given 4-vector u that is not parallel to P $((\underset{\sim}{P}^2+m^2)^{\frac{1}{2}}, \underset{\sim}{P})$, and the precise bound of Eq. (III.5).

If we use the variable $\underset{\sim}{q} = \underset{\sim}{p} - \underset{\sim}{P}$ and denote from now on by $\chi(\underset{\sim}{q})$ the function $\hat{\chi}(\underset{\sim}{P}+\underset{\sim}{q})$ ($\chi(\underset{\sim}{q})$ is analytic at $\underset{\sim}{q} = 0$ in either case), we can write:

$$\hat{f}(u\tau) = e^{i(Pu)\tau} \int \chi(\underset{\sim}{q}) \exp(-\tau[\underset{\sim}{q}^2\gamma - i(\underset{\sim}{q}u - u_o[(\underset{\sim}{P}+\underset{\sim}{q})^2+m^2)^{\frac{1}{2}} - P_o]])\, d\underset{\sim}{q} \qquad \text{(AIII.1)}$$

Consider for any positive α, the set of surfaces Σ_α in the space of complex $\underset{\sim}{q}$:

$$\mathrm{Re}[q^2\gamma - i(qu - u_o[((P+q)^2+m^2)^{\frac{1}{2}} - P_o])] = \alpha\gamma \qquad \text{(AIII.2)}$$

The intersection of Σ_α with the space of real $\underset{\sim}{q}$ is the sphere $S_\alpha: \underset{\sim}{q}^2 = \alpha$.

If it is possible to find on Σ_α a 3-dimensional surface L_α bounded by S_α, continuous in α, closed and bounded and lying in the domain of analyticity of χ (and of $((\underset{\sim}{P}+\underset{\sim}{q})^2+m^2)^{\frac{1}{2}}$), then $\hat{f}(u\tau)$ is easily seen to have an exponential fall off $e^{-\alpha\gamma\tau}$ in the $\tau \to \infty$ limit: to see this divide the region of integration over real $\underset{\sim}{q}$ into the two parts $|\underset{\sim}{q}^2| > \alpha$ and

$|\underset{\sim}{q}|^2 < \alpha$. The contribution of the part $|\underset{\sim}{q}|^2 > \alpha$ has the fall off $e^{-\alpha\gamma\tau}$ because of the presence of the Gaussian in the integrand. The contribution of the part $|\underset{\sim}{q}|^2 < \alpha$ is evaluated by distorting the contour to the position L_α. The fall off $e^{-\alpha\gamma\tau}$ is thus obtained.

We will not determine here the maximum value of α for which the properties required on Σ_α are satisfied,* but will prove the existence of a positive α independent of γ such that the surface Σ_α satisfies these properties for all γ smaller than or equal to γ_0.

We denote by $\underset{\sim}{x}$ and $\underset{\sim}{y}$ the real and imaginary parts of $\underset{\sim}{q}$, and choose a coordinate system in which the first axis is along the direction of the vector $\underset{\sim}{d} = \underset{\sim}{u} - \frac{\underset{\sim}{P}}{P_0} u_0$.

Consider the surface L_α obtained by setting $y_2 = y_3 = 0$:

$$L_\alpha \begin{cases} \gamma(\underset{\sim}{x}^2 + y_1^2) + d y_1 - u_0 \,\mathrm{Im}[\varphi(\underset{\sim}{x}+i\underset{\sim}{y})] = \alpha\gamma \\ y_2 = y_3 = 0 \end{cases} \qquad \text{(AIII.3)}$$

where

$$\varphi(\underset{\sim}{q}) = ((\underset{\sim}{P}+\underset{\sim}{q})^2 + m^2)^{\frac{1}{2}} - P_0 - \frac{\underset{\sim}{P}}{P_0}\underset{\sim}{q}$$

If we use the new variables $(\underset{\sim}{x}, z_1 = \frac{y_1}{\gamma})$, Eq. (AIII.3) becomes:

$$\underset{\sim}{x}^2 - \gamma^2 z_1^2 + d z_1 - u_0 z_1 \psi(\underset{\sim}{x}, \gamma z_1) = \alpha \qquad \text{(AIII.4)}$$

where

$$\psi(\underset{\sim}{x}, \gamma z_1) = \frac{P_1 + x_1}{((\underset{\sim}{P}+\underset{\sim}{x})^2 + m^2)^{\frac{1}{2}}} - \frac{P_1}{P_0} + \mathcal{O}(\gamma z_1)\ ,$$

the function $\mathcal{O}(\gamma z_1)$ being of the first order in γz_1.

Choose α sufficiently small so that $|u_0 \psi(\underset{\sim}{x}, 0)| < d$ for $|\underset{\sim}{x}| < \sqrt{\alpha}$. (This is possible since $\psi(0,0) = 0$.)

---

* In non-relativistic quantum mechanics, where the factor $((\underset{\sim}{P}+\underset{\sim}{q})^2+m^2)^{\frac{1}{2}}$ is replaced by $(\underset{\sim}{P}+\underset{\sim}{q})^2/2m$, this maximal value is

$$\alpha = \frac{(\underset{\sim}{u} - (\underset{\sim}{P}/m)u_0)^2}{\gamma^2 + (u_0/2m)^2}$$

(if the domain of analyticity of χ is sufficiently large).

For $\gamma = 0$, Eq. (AIII.4) defines explicitly z_1 as a (positive) function of $\underset{\sim}{x}$ is the region $|\underset{\sim}{x}| \leqq \sqrt{\alpha}$. By virtue of the implicit function theorem, it is easy to show[3] that it defines implicitly z_1 as a function of $\underset{\sim}{x}$ in that region if γ is sufficiently small. We denote by γ_0 any strictly positive number such that $z_1(\underset{\sim}{x})$ is unique and continuous in $|\underset{\sim}{x}| \leqq \sqrt{\alpha}$ for $\gamma \leqq \gamma_0$.

The surface L_α in the region $|\underset{\sim}{x}| \leqq \sqrt{\alpha}$ is thus closed and bounded. If α is chosen sufficiently small, then all the surfaces L_α for $\gamma \leqq \gamma_0$ also lie in the domain of analyticity of χ (and of $((\underset{\sim}{p}+\underset{\sim}{q})^2+m^2)^{\frac{1}{2}}$), and $\hat{f}(u\tau)$ has thus the exponential fall off $e^{-\alpha\gamma\tau}$ for $\gamma \leqq \gamma_0$.

Consider finally a set of χ that have a common support in the region $|\underset{\sim}{q}| \leqq 1$ and a common upper bound K for $\underset{\sim}{q}$ either real or in the complex domain lying between the surfaces $\{\underset{\sim}{q}$ real, $|\underset{\sim}{q}| \leqq \sqrt{\alpha}\}$ and $L_{\alpha\gamma_0}$. We easily obtain the bound:

$$|\hat{f}(u\tau)| < C\, e^{-\alpha\gamma\tau}$$

where $C = K(A_{\alpha\gamma_0} + 4\pi)$, and $A_{\alpha\gamma_0}$ is the area of $L_{\alpha\gamma_0}$ (in the region $|\underset{\sim}{x}| \leqq \sqrt{\alpha}$).

References

1. General references on analytic S-matrix theory:
   a) H. P. Stapp, "Crossing, Hermitian Analyticity and the Connection Between Spin and Statistics," UCRL 16816 April 1966 (to be published in J. Math. Phys.).
   b) G. F. Chew, The Analytic S-matrix (W. A. Benjamin, New York, 1966).
   c) R. J. Eden, P. V. Landshoff, D. I. Olive and J. C. Polkinghorne, The Analytic S-matrix (Cambridge University Press, 1966).

   References concerning more specifically the physical region:
   d) C. Chandler and H. P. Stapp, "Macroscopic Causality Conditions and Properties of Scattering Amplitudes," UCRL 17734 (submitted to J. Math. Phys.).
   e) J. Coester and H. P. Stapp, "Physical Region Discontinuity Equations for Many Particle Scattering Amplitudes," I, II, UCRL 17489 and 17902 (to be published in J. Math. Phys.) and "Physical Region Discontinuity Equations," UCRL 18512 (to be submitted to Phys. Rev.).
   f) M. J. Bloxham, D. I. Olive and J. C. Polkinghorne, "S-Matrix Singularities in the Physical Region," I, II, III Cambridge preprint, June 1967 (to be published in J. Math. Phys.).

g) F. Pahm, "Singularite des Processes de Diffusion Multiple," Ann. Inst. Henri Poincaré, Vol. VI, n. 2, 1967, p. 89.
h) D. Iagolnitzer and H. P. Stapp, "Macroscopic Causality and Physical Region Analyticity," Berkeley preprint.
2. cf. Ref. 1 d).
3. cf. Ref. 1 h).
4. E. H. Wichmann and J. H. Crichton, Phys. Rev. 132, 2788 (1963).
5. J. R. Taylor, Phys. Rev. 142, 1236 (1966).
6. R. Omnes, Phys. Rev. 146, 1123 (1966).
7. M. Kugler and R. Roskies, Phys. Rev. 155, 1685 (1967).
8. J. D. Finley, "Some Implications for Scattering of Short Range Interaction," Ph.D. Thesis, Univ. of Calif., Berkeley.
9. a) F. Pham, cf. Ref. 1 g).
b) H. P. Stapp, Phys. Rev. 139, 257 (1965).
c) G. Wanders, Helv. Phys. Acta 38, 142 (1965).
Refs. a)b)c) contain partial results. For a qualitative discussion see
d) A. Peres, Ann. Phys. 37, 179 (1966).
More complete results are due to H. P. Stapp(private communication).
10. H. P. Stapp, cf. Ref. 9b).
11. F. Pham, cf. Ref. 1 g).
12. G. Wanders, cf. Ref. 9 c).
13. D. Iagolnitzer, "S-matrix Theory and Classical Description of Interactions," SACLAY report 1966 (to be published in J. Math. Phys.).
14. J. Coester and H. P. Stapp, cf. Ref. 1 e).
15. H. P. Stapp in The Trieste Lectures, High Energy Physics and Elementary Particles (IAEA Vienna 1965).
16. H. P. Stapp (private communication).
17. L. Schwartz, Theorie des Distributions (Hermann,Paris,1950).
18. D. Ruelle (private communication).
19. D. N. Williams, J. Math. Phys. 8, 1807 (1967).
20. S. Coleman and R. F. Norton, Nuovo Cimento 38, 438 (1965).
21. H. P. Stapp, J. Math. Phys. 8, 1606 (1967).
22. See for instance "Edge of the Wedge Theorem" in R. F. Streater and A. S. Wightman, PCT, Spin and Statistics and All That (W.A. Benjamin, New York, 1964).
23. D. N. Williams (private communication).
24. F. Pahm, Ref. 1 g).
25. D. Ruelle, Helv. Phys. Acta 35, 147 (1962).

26. See for instance B. A. Fuks, "Introduction to the Theory of Analytic Functions of Several Complex Variables," Vol. 2, American Mathematical Society, Providence, Rhode Island, 1963.
27. M. Froissart, M. L. Goldberger and K. M. Watson, Phys. Rev. 131, 2820 (1963).
28. H. P. Stapp, Phys. Rev. 125, 2134 (1962) and Ref. 1 a).
29. M. J. Bloxham, et al, cf. Ref. 1 f).
30. D. N. Williams (private communication).
31. K. Hepp, in Particle Symmetries and Axiomatic Field Theory, ed. by M. Chretien and S. Deser (Gordon and Breach, N.Y., 1966).
32. J. Bros, Princeton Institute for Advanced Study preprint, 1968.
33. See for instance P. Moussa and R. Stora in the Proceedings of the Hercegnovi International School of Elementary Particles, 1966.
34. J. R. Taylor, Phys. Rev. 140, 187 (1965).
35. D. N. Williams in Lectures in Theoretical Physics, Vol. XB (Gordon and Breach, New York, 1968), ed. Asim O. Barut and Wesley E. Brittin.
36. M. Froissart and J. R. Taylor, Phys. Rev. 153, 1636 (1967).
37. J. R. Taylor, unpublished version of Ref. 5.

NONLINEAR ANALYSIS APPLIED TO S-MATRIX THEORY†‡

Robert L. Warnock
Illinois Institute of Technology
Chicago, Illinois
and
Argonne National Laboratory
Argonne, Illinois

Table of Contents

I. Introduction
II. Fixed Point Theorems; Existence Theorems for Nonlinear Singular Integral Equations
  A. Contraction Mapping Principle
  B. An Application of the Contraction Mapping Principle: Existence Theorem for a Special Nonlinear Singular Integral Equation
  C. Analyticity of Solution with Respect to a Parameter
  D. Brouwer and Schauder Fixed Point Theorems
III. Fixed Point Theorems Applied to the Low Equation
  A. Direct Study of Low Operator as a Singular Integral Operator
  B. Analyticity in Coupling Constant
  C. Study of Low Equation in N/D Formulation
IV. An Approach to the Strong Coupling Regime
  A. Fréchet Derivatives
  B. Implicit Function Theorem
  C. Index of the Fréchet Derivative of the Low Operator; Comments on a Paper of Lovelace
  D. Fréchet Derivative of the N/D Operator; Continuous Dependence of Solutions on Parameters
  E. Newton-Kantorovich Method; Solutions with Successive Increments of the Coupling Constant

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.
‡Work partially supported by the National Science Foundation and performed in part under the auspices of the U. S. Atomic Energy Commission.

F. Newton-Kantorovich Method at a Point where the Fréchet Derivative Lacks an Inverse; Anselone-Moore Change of Parameter
G. Bifurcation Points

I. Introduction

The literature on "dynamical equations" intended to describe the strong interactions of elementary particles is by now very extensive.[1] In reading this work one is struck by the lack of adequate mathematical methods, and the consequent lack of definite conclusions. The main problems concern strongly nonlinear integral equations with singular integrals. The nonlinearity comes about through the combination of crossing symmetry and unitarity. Usually the equations have been solved only after some mutilation of crossing symmetry. Mathematically inclined physicists have not given much attention to this kind of problem, perhaps because there was no reason to believe firmly in the physical validity of the equations. Nevertheless, there are several good reasons to regard dynamical equations as an interesting challenge. (a) The same mathematical problems are likely to occur in more complete dynamical theories, and related problems certainly do occur in other areas of physics. (b) Certain of the equations, particularly those formulated in the S-matrix language, are not really so remote from experiment. It is quite possible that through better experimental data and better mathematics we shall arrive at a believable theory. The theory will probably involve some unknown constants or functions to account for incalculable many-particle states. Even with such empirical elements it is possible for a theoretical scheme to be a considerable advance over mere curve fitting. Many questions may not depend on the small details of multi-particle channels, just as many questions in nuclear physics are answered without referring to the details of the nuclear forces. (c) Some interesting ideas on the origins of symmetry breaking have come from studies of dynamical models.[2] To date these studies are not as penetrating as they could be, partly because the nonlinear aspects of symmetry breaking have not been squarely faced. Certain variational and stability problems are involved, and these have been treated only in a cursory fashion. It seems to me that this field could be especially fertile for the application of suitable nonlinear methods.

During the past two years several papers have appeared in which nonlinear functional analysis is applied to dynamical equations. A nearly complete list of authors is as follows: J. G. Taylor, C. Lovelace, D. Atkinson, A. Ts. Amatuni, H. McDaniel, R. Warnock. Most of this work uses theorems which are commonplace in books on nonlinear methods. In these notes I shall review these standard

theorems, together with some less known but still quite simple facts, and also show how to get existence theorems, various qualitative information, and methods of computation in nonlinear problems of S-matrix theory. The only example of an S-matrix theory that I treat is the one-meson Low equation of the static baryon model. More complicated equations may be approached with essentially the same methods, although the necessary elaborations of technique can be nontrivial.

Some of the main results to date come from fixed point theorems. If φ denotes a set of scattering amplitudes, regarded as elements of an appropriate space, then the S matrix equations have the form

$$\varphi = A\varphi \tag{I.1}$$

where A is a certain nonlinear singular integral operator. Fixed point theorems give conditions for the solubility of Eq. (I.1); i.e., sufficient conditions for the existence of points in the space which are invariant under the operation A.

The first applications of nonlinear analysis to dynamical equations were by J. G. Taylor,[3] who studied off-mass-shell equations of field theory. Taylor reviewed his work at the 1966 meeting of the Boulder Institute.[4] The problems he encountered were more difficult in some respects than those we meet in S-matrix theory; I will not be able to discuss them here.

C. Lovelace[5] investigated on-shell equations of the Chew-Mandelstam type. He used the concept of "index" to determine the dimension of the manifold to which a particular solution belongs. (A solution may belong to an infinite family of solutions, the members of which are specified by n real parameters. n is the dimension of the manifold.) I will give a simplified version of Lovelace's argument in Section IV below.

Lovelace did not look into the question of whether the Chew-Mandelstam equations actually have solutions, but that problem was considered by Atkinson.[6] Atkinson used N/D equations which, because of crossing, form a complicated nonlinear system. By an application of Schauder's fixed point theorem he showed that the N/D equations have a solution when certain subtraction constants (analogous to coupling constants) are sufficiently small. This was not quite enough to show that the Chew-Mandelstam equations have solutions, since the D function could have ghost zeros, but it was an encouraging first attempt at existence problems.[7]

The author[8] investigated Low's one-meson equation for the static baryon model, again by Schauder's theorem, but avoiding the N/D representation and the ghost problem. It was proved that the

Low equation has at least one solution for arbitrary crossing matrix and a wide class of cut-offs, provided the coupling constant is sufficiently small. The solution corresponds to an elementary baryon (i.e., it is not generated dynamically as a bound state). Furthermore, the solution has no resonances.

McDaniel[9] and the author analyzed the Low equation again, but in the N/D formulation. By means of a technique somewhat different from Atkinson's, we were able to rule out ghosts, and to prove that with an arbitrary choice of CDD poles there is a unique solution of the N/D equation within a certain class, provided that the coupling constant and the CDD pole residues are sufficiently small. There is, therefore, an infinite manifold of solutions of the Low equation, corresponding to the various values of the CDD pole positions and residues. The CDD poles are at physical energies in the denominator function and they have small residues, so it follows that they correspond to narrow resonances. Thus there is a solution of the Low equation corresponding to a nearly arbitrary choice of narrow resonances. This analysis is done by both the Schauder fixed point principle and the contraction mapping principle. In the latter case the solutions are actually constructed by a convergent iteration procedure.

Recently Atkinson[10] has carried out a very nice analysis of an "exact" relativistic problem. He proves the existence of an amplitude for scattering of neutral, spin zero mesons which satisfies an unsubtracted Mandelstam representation, crossing symmetry, elastic unitarity in the elastic region, and the unitarity inequalities in the inelastic region. A fixed point theorem is applied to the Mandelstam formula for the elastic part of the double spectral function. Atkinson assumes that a certain inelastic part of the double spectral function is given, so that Mandelstam's formula becomes a nonlinear integral equation for the elastic part. For any inelastic part of a certain class (defined only by continuity, support properties, and bounds) the fixed point theorem shows that there exists an elastic double spectral function satisfying all the necessary conditions. This work shows that the program based on Mandelstam analyticity, unitarity, and crossing is mathematically reasonable. It also illustrates the general viewpoint on dynamical equations that I mentioned above: namely, there are interesting dynamical questions to study even if we do not know all about the inelastic effects.

It should be mentioned that Martin[11] and Newton[12] have applied fixed point theorems to a problem of scattering theory different from those we consider here. They deal with the question of whether the differential cross-section and the unitarity condition together determine the scattering amplitude in the physical region.

Unfortunately all of the work with fixed point theorems only proves existence of solutions which are rather small and uninteresting. For the theorems to be applicable, the coupling constant (or some other measure of the magnitude of amplitudes) must be somewhat smaller than its known experimental value. No "dynamical" resonances are proved to exist. The only resonances obtained so far have narrow widths, and they are put into the equations by hand via CDD poles. Although there is analyticity in angular momentum (in Atkinson's work[10]) there are so far no Regge poles in the right half-plane.[13]

These circumstances seem to indicate a basic limitation of fixed point theorems. The theorems seem not to be effective when the nonlinearities are as severe as they are in strong-interaction physics. They might still play a role as a starting point for a strong-coupling analysis, because if a weak-coupling solution is known it may be possible to follow it into the strong-coupling regime by an iteration procedure involving successive increments of the coupling strength. In the following I will outline a scheme for following a solution from weak to strong coupling. It is based on solution of the equation by successive linearizations. This approach so far yields no existence theorems, but it does give a method for controlled numerical solutions in the strong coupling regime.

Section II introduces the fixed point theorems, and shows how they may be applied to a nonlinear singular integral equation having a structure similar to that of the Low equation.

In Section III our work on the Low equation is outlined. The discussion includes one point of interest that we did not touch in earlier papers; viz., that solutions of the Low equation obtained by iteration are analytic in the coupling constant λ in a neighborhood of $\lambda = 0$.

In Section IV we try to overcome the limitations of the fixed point theorems. Essentially the method is to traverse on foot the more hazardous regions of Banach space. That is to be compared with the armchair mountaineering of the fixed point method. First the Fréchet derivative of a nonlinear operator is defined. In terms of this derivative the implicit function theorem of ordinary analysis is extended to operators in Banach space. It is found that the Fréchet derivative of the nonlinear N/D operator is a compact (i.e., Fredholm) operator, while the derivative of the operator appearing in the Low equation is not compact in general. The index of the latter is computed in a simple way, and a comparison with the work of Lovelace is made. The implicit function theorem is applied to the N/D equations to show that under appropriate conditions their solutions are continuous and differentiable with respect to the coupling constant. This result applies not only to solutions established by the fixed point theorems. Thus when

we have a solution it will normally be possible to change the coupling constant by a small amount and still be assured that a solution exists. The old solution goes continuously into the new solution as the coupling constant is changed continuously. In order to construct the new solution we suggest an application of the Newton-Kantorovich method, which is a generalization to Banach spaces of the familiar Newton method of successive linearizations. This should allow us to construct a sequence of solutions corresponding to larger and larger coupling constants. The sequence will end, however, if we reach a point where the Fréchet derivative does not have an inverse. [Recall that the classical Newton iterates are $x_{n+1} = x_n - f(x_n)/f'(x_n)$, so the procedure fails if $f'(x_n) = 0$]. By means of a technique due to Anselone and Moore[14] it is often possible to go past such points, however. We make an analogy with a problem of elasticity theory considered by these authors. They were able to follow solutions of a nonlinear integral equation to arbitrarily large values of a parameter through an extended Newton method. The same method will probably work for the Low equation in the N/D formulation.

Special interest is attached to a value λ_0 of the coupling constant for which the Fréchet derivative is singular. These points are likely to be "bifurcation points." That means that for any λ close to λ_0 there are two distinct solutions which merge into one as λ tends to λ_0. Physically this would mean two distinct solutions of the Low equation for the same coupling constant, both with the same CDD poles. This would be a previously unsuspected ambiguity, different from the CDD ambiguity. The Fréchet derivative of the N/D operator is a generalization of the Cutkosky-Dashen-Frautschi[15] matrix 1-A. The points where the latter is singular have been emphasized in connection with symmetry breaking. It will be interesting to see what bifurcation points have to do with symmetry breaking.

II. Fixed Point Theorems; Existence Theorems for Nonlinear Singular Integral Equations

A. Contraction Mapping Principle

We are given an equation of the form $\varphi = A\varphi$, where A is an operator (in general nonlinear) on a metric space. The most obvious way to attempt a solution is by iteration. Beginning with an arbitrary element φ_0 of the space one forms the sequence $\varphi_1 = A\varphi_0$, $\varphi_2 = A\varphi_1, \ldots, \varphi_n = A\varphi_{n-1}, \ldots$. If the metric space is complete, then one can give a simple condition on A which will guarantee success of the iterative method. We recall some elementary definitions.

(i) A space K is called a metric space if to every pair φ, ψ of points in K there is assigned a real number $\delta(\varphi, \psi)$ called the distance between φ and ψ such that

(a) $\delta(\varphi,\psi) \geq 0$, and $\delta(\varphi,\psi) = 0$ if and only if $\varphi = \psi$,
(b) $\delta(\varphi,\psi) = \delta(\psi,\varphi)$,
(c) $\delta(\varphi,\chi) + \delta(\chi,\psi) \geq \delta(\varphi,\psi)$.

(ii) A point φ of K will be called a limit of a sequence $\{\varphi_n\}$ in K if $\delta(\varphi_n,\varphi) \to 0$, $n \to \infty$. We write $\varphi_n \to \varphi$.

(iii) A Cauchy sequence is a sequence $\{\varphi_n\}$ such that for any $\epsilon > 0$ there is a positive integer $N(\epsilon)$ such that

$$\delta(\varphi_n,\varphi_m) < \epsilon, \quad n,m > N(\epsilon) .$$

A convergent sequence is necessarily a Cauchy sequence: If $\varphi_n \to \varphi$, then

$$\delta(\varphi_n,\varphi_m) \leq \delta(\varphi_n,\varphi) + \delta(\varphi_m,\varphi) < \epsilon; \quad n,m > N(\epsilon/2).$$

A Cauchy sequence, however, need not have a limit.

(iv) A metric space in which every Cauchy sequence has a limit is called a complete metric space.

The contraction mapping principle (or Banach-Cacciopoli fixed point theorem) is as follows:[37)-49)]

Theorem 1: Suppose that an operator A maps a complete metric space K into itself, and that A satisfies the Lipschitz condition

$$\delta(A\varphi, A\psi) \leq \beta\delta(\varphi,\psi), \quad \beta < 1 , \tag{II.1}$$

for all $\varphi, \psi \in K$, with β independent of φ,ψ. (Such an A is called a contraction mapping on K.) Then the equation $\varphi = A\varphi$ has a unique solution φ in K (i.e., A has a unique fixed point in K). This solution is obtained by iteration $\varphi_n = A\varphi_{n-1}$, beginning with an arbitrary element φ_0 of K. The error at the n-th iteration is bounded in terms of that at the first iteration:

$$\delta(\varphi,\varphi_n) \leq \frac{\beta^n}{1-\beta}\,\delta(\varphi_1,\varphi_0) . \tag{II.2}$$

The proof is quite easy. With any n (II.1) gives

$$\delta(\varphi_{n+1},\varphi_n) \leq \beta\delta(\varphi_n,\varphi_{n-1}) \leq \cdots \leq \beta^n\,\delta(\varphi_1,\varphi_0) . \tag{II.3}$$

To show that we have a Cauchy sequence, consider

$$\begin{aligned}\delta(\varphi_{n+k},\varphi_n) &\leq \delta(\varphi_{n+k},\varphi_{n+k-1}) + \cdots + \delta(\varphi_{n+1},\varphi_n)\\ &\leq (\beta^{n+k-1} + \cdots + \beta^n)\,\delta(\varphi_1,\varphi_0)\\ &< \frac{\beta^n}{1-\beta}\,\delta(\varphi_1,\varphi_0) \quad .\end{aligned} \tag{II.4}$$

Since $\beta^n \to 0$, $\{\varphi_n\}$ is a Cauchy sequence which has a limit φ in our complete space. Moreover, φ is a fixed point of A, as is seen from the inequalities

$$\begin{aligned}\delta(A\varphi,\varphi) &\leq \delta(A\varphi, A\varphi_n) + \delta(A\varphi_n,\varphi) \\ &\leq \beta\,\delta(\varphi,\varphi_n) + \delta(\varphi_{n+1},\varphi) \quad . \end{aligned} \tag{II.5}$$

Since the right-hand side tends to zero as n increases, $\delta(A\varphi,\varphi) = 0$. By (ia) above, $A\varphi = \varphi$. To show that φ is a unique solution, suppose that φ and ψ are both solutions. Then

$$\delta(\varphi, \psi) = \delta(A\varphi, A\psi) \leq \beta\,\delta(\varphi, \psi) \quad . \tag{II.6}$$

Since $\beta < 1$, $\delta(\varphi,\psi) = 0$, or $\varphi = \psi$. The error bound (II.2) is obtained by letting k go to infinity in (II.4).

B. An Application of the Contraction Mapping Principle: Existence Theorem for a Special Nonlinear Singular Integral Equation

As an example we apply the contraction mapping principle to a nonlinear, singular integral equation $\varphi = A\varphi$, where A is defined as follows:

$$\varphi(t) = f(t) + \lambda\, P\int_0^1 \frac{\rho(x)\,\varphi^a(x)\,dx}{x - t}, \quad 0 \leq t \leq 1 \ . \tag{II.7}$$

Here P denotes the Cauchy principal value of the singular integral, λ is a real parameter measuring the strength of the nonlinearity, and $f(x)$ and $\rho(x) \geq 0$ are given real functions which are required to be Hölder-continuous on the interval $[0,1]$. A function $\psi(x)$ is said to be Hölder-continuous or H-continuous on $[0,1]$ if there are positive numbers κ and μ such that

$$|\psi(x) - \psi(t)| \leq \kappa\,|x - t|^{\mu}, \quad x, t \in [0,1] \ . \tag{II.8}$$

The Hölder coefficient is κ and the Hölder exponent is μ. We write $\psi \in H(\mu,\kappa)$ or $\psi \in H(\mu)$. Without losing generality we may require $0 < \mu \leq 1$. Note that an H-continuous function is continuous, but not necessarily continously differentiable. For instance, $x^{\frac{1}{2}}$ is H-continuous on $[0,1]$, but its derivative is not continuous at $x = 0$. To avoid endpoint singularities in (II.7) (which could be allowed with greater effort) we assume that ρ vanishes at the ends: $\rho(0) = \rho(1) = 0$. Then by (II.8) we have

$$\rho(x) = 0(x^{\mu}), \; x \to 0 \; ,$$

$$\rho(x) = 0([1 - x]^{\mu}) \quad , \quad x \to 1 \; . \tag{II.9}$$

if μ is the Hölder exponent of ρ. Equation (II.7) has the essential structure of the equations we encounter in S matrix theory, but it is algebraically simpler. It will be worthwhile to study it in some detail.

In applying the contraction mapping principle to $\varphi = A\varphi$, or in fact for any analysis of such an equation, the first step is to find a space in which the operator A has reasonable properties. That space need not coincide with the whole domain of A.

For our metric space in the contraction mapping we shall take a subset of a certain Banach space. First more definitions.

(v) A normed linear space N is a real or complex linear vector space in which each point φ is assigned a real number $||\varphi||$ called its norm in such a way that

(a) $||\varphi|| \geq 0$,

(b) $||\varphi + \psi|| \leq ||\varphi|| + ||\psi||$,

(c) $||\lambda\varphi|| = |\lambda| \quad ||\varphi||$ for every scalar λ,

(d) $||\varphi|| = 0$ if and only if $\varphi = 0$.

A normed linear space is clearly a metric space with the identification $\delta(\varphi, \psi) = ||\varphi - \psi||$.

(vi) Limits in N will be defined with respect to the norm topology: a sequence $\{\varphi_n\}$ is said to have a limit φ if $||\varphi - \varphi_n||$ tends to zero.

(vii) A Banach space is a normed linear space which is complete in the norm topology.

An example of a Banach space is the space $C[0,1]$ of all real (or all complex) continuous functions on $[0,1]$, with the norm of a function being the maximum of its absolute value:

$$||\varphi|| = \sup_{0 \leq t \leq 1} |\varphi(t)| \quad . \tag{II.10}$$

It is immediate to check that this is a normed linear space, and its completeness is easily proved by means of the Bolzano-Weierstrass theorem. We could try to regard the operator A of (II.7) as an operator on $C[0,1]$, but that would not work since the principal value integral

is not defined on all continuous functions. Next we can try the space $H[0,1]$ of all H-continuous functions on $[0,1]$, keeping the same norm (II.10). This space is also a Banach space; we call it B_1. If $\psi \in H[0,1]$, then the following integral exists for $0 < t < 1$:

$$P\int_0^1 \frac{\psi(x)dx}{x-t} = P\int_0^1 \frac{\psi(x)-\psi(t)}{x-t}\,dx + \psi(t)\,P\int_0^1 \frac{dx}{x-t}\,. \qquad \text{(II.11)}$$

The right side of (II.11) is less than or equal in magnitude to

$$\kappa\int_0^1 |x-t|^{\mu-1}\,dx + |\psi(t)|\left|\log\left(\frac{1-t}{t}\right)\right| < \infty,\quad 0<t<1. \qquad \text{(II.12)}$$

If $\psi(0) = \psi(1) = 0$, then we avoid the logarithmic singularities at $t = 0, 1$ and (II.8) exists on the whole closed interval $0 \leqslant t \leqslant 1$. If ψ has Hölder exponent μ, and $\psi(0) = \psi(1) = 0$, then it is possible to show that the integral (II.8) is itself H-continuous on $[0,1]$. It is H continuous with exponent μ if $\mu < 1$, and with exponent $1 - \epsilon$ if $\mu = 1$, where ϵ is arbitrarily small. For a proof see Muskhelishvili,[16] §20. Now since a product of H-continuous functions is H-continuous, the operator (II.7) maps the space $H[0,1]$ into itself. $H[0,1]$ will in fact be a satisfactory space for our analysis of (II.7), but the norm (II.10) will not be appropriate for an application of the contraction mapping principle. To verify the Lipschitz condition (II.1) we shall have to estimate the integral in (II.7) in terms of the norm. With the norm (II.10) the estimates that we could do would not be good enough. So instead we choose a Banach space B_2 which is $H[0,1]$ with a norm especially suited for estimates of principal value integrals; viz.,

$$\|\varphi\| = \sup_{t\in[0,1]} |\varphi(t)| + H[\varphi],$$

$$H[\varphi] = \sup_{\substack{t,x\in[0,1]\\ t\neq x}} \frac{|\varphi(t)-\varphi(x)|}{|t-x|^{\mu}}\,. \qquad \text{(II.13)}$$

To apply theorem 1 in the case of the operator A of (II.7), we choose a complete metric space K which is a subset of B_2. K consists of all complex functions φ such that

$$|\varphi(t)| \leqslant a,\quad |\varphi(t)-\varphi(x)| \leqslant b|t-x|^{\mu},$$

$$t,x \in [0,1];\quad 0<\mu<1\,. \qquad \text{(II.14)}$$

Here a, b, and μ are fixed numbers, and a and b will eventually be restricted so as to guarantee that A is a contraction mapping on K. The distance in K is given by the norm in B_2: $\delta(\varphi,\psi) = ||\varphi - \psi||$. K is proved to be complete in Ref. 9, Appendix A.

The first step is to show that A maps K into itself, under appropriate restrictions on λ, a and b. Let φ be any element of K, and define the function $\psi = A\varphi$:

$$\psi(t) = f(t) + \lambda P \int_0^1 \frac{\rho(x)\,\varphi^2(x)\,dx}{x - t} \quad . \tag{II.15}$$

To show that $\psi \in K$ we first estimate $|\psi|$ as follows:

$$|\psi(t)| \leq |f(t)| + |\lambda| \left[\int_0^1 \rho(x) \left| \frac{\varphi^2(x) - \varphi^2(t)}{x - t} \right| dx + \varphi^2(t) \left| P \int_0^1 \frac{\rho(x)}{x - t}\, dx \right| \right]$$

$$\leq \sup |f(t)| + |\lambda|\, a^2 \sup \left| P \int_0^1 \frac{\rho(x)}{x - t}\, dx \right|$$

$$+ |\lambda| \int_0^1 \rho(x) \left| \frac{\varphi^2(x) - \varphi^2(t)}{x - t} \right| dx \quad . \tag{II.16}$$

We can now use the H-continuity of φ^2:

$$|\varphi^2(x) - \varphi^2(t)| = |\varphi(x) - \varphi(t)| \cdot |\varphi(x) + \varphi(t)| \leq 2ab\, |x - t|^{\mu} \, . \tag{II.17}$$

Thus from (II.16),

$$|\psi(t)| \leq \sup |f| + |\lambda| \, (a^2 \hat{\rho}_1 + 2ab\hat{\rho}_2) \, , \tag{II.18}$$

$$\hat{\rho}_1 = \sup \left| P \int_0^1 \frac{\rho(x)}{x - t}\, dx \right| ; \; \hat{\rho}_2 = \sup \int_0^1 \rho(x) |x - t|^{\mu - 1}\, dx. \tag{II.19}$$

Next we need an estimate of $\psi(t) - \psi(t')$. We subtract the pole in the integral as in (II.11) and begin as follows:

$$|\psi(t) - \psi(t')| \leq |f(t) - f(t')|$$

$$+ |\lambda| \cdot |\varphi^2(t)\, P \int_0^1 \frac{\rho(x)dx}{x - t} - \varphi^2(t')\, P \int_0^1 \frac{\rho(x)dx}{x - t'}|$$

$$+ |\lambda| \cdot \left| \int_0^1 \frac{\rho(x)[\varphi^2(x) - \varphi^2(t)]dx}{x - t} - \int_0^1 \frac{\rho(x)[\varphi^2(x) - \varphi^2(t')]dx}{x - t'} \right| \, . \tag{II.20}$$

For simplicity in notation we assume that $f(x)$ and $\rho(x)$ both have the same Hölder exponent μ that is involved in (II.14); this is not necessary. From a theorem quoted above, we know that $I(t) = P\int_0^1 \rho(x)dx(x-t)^{-1}$ is H-continuous with exponent μ. Let $r = H[I]$ denote its minimum Hölder coefficient, where H is defined in (II.13). Then the sum of the first two terms of (II.20) is bounded by

$$[|\lambda|\ (2ab\hat{\rho}_1 + a^2 r) + H[f]]\,|t - t'|^{\mu} \quad , \tag{II.21}$$

where

$$r = H[P\int_0^1 dx\ \rho(x)\ (x-t)^{-1}] \quad .$$

This is seen by referring to (II.17). To deal with the last term in (II.20), we first formally extend the region of integration from $[0,1]$ to $\gamma = [-2,3]$. By putting $\rho(x) \equiv 0$ on the added pieces we do not change anything. Suppose $t' \geq t$, and write $t' = t + \theta$. Define a subinterval γ_1 of γ:

$$\gamma_1 = \{x \mid t - 2\theta \leq x \leq t + 2\theta\} \quad . \tag{II.22}$$

If the last term of (II.20) is denoted by $|\lambda|\ |J(t) - J(t')|$ we have $J(t+\theta) - J(t) = \int_{\gamma_1} + \int_{\gamma - \gamma_1}$. The integeral over γ_1 is of order θ^{μ}:

$$\left|\int_{\gamma_1} \rho(x)dx\left[\frac{\varphi^2(x) - \varphi^2(t+\theta)}{x - t - \theta} - \frac{\varphi^2(x) - \varphi^2(t)}{x - t}\right]\right|$$

$$< 2ab\hat{\rho}_3\left[\int_{\gamma_1} |x - t - \theta|^{\mu-1}\,dx + \int_{\gamma_1} |x-t|^{\mu-1}dx\right]$$

$$= \frac{2ab\hat{\rho}_3}{\mu}\,(1 + 3^{\mu} + 2^{1+\mu})\,\theta^{\mu} = 2ab\hat{\rho}_4\theta^{\mu} \quad ,$$

$$\hat{\rho}_3 = \sup\,|\rho(x)|\,,$$

$$\hat{\rho}_4 = \hat{\rho}_3(1 + 3^{\mu} + 2^{1+\mu})/\mu \quad . \tag{II.23}$$

The remaining integral over $\gamma - \gamma_1$ is divided into two parts:

$$I_1 + I_2 = [\varphi^2(t) - \varphi^2(t+\theta)]\int_{\gamma-\gamma_1} \frac{\rho(x)dx}{x - t}$$

$$+ \int_{\gamma-\gamma_1} [\varphi^2(x) - \varphi^2(t+\theta)]\left[\frac{1}{x - t - \theta} - \frac{1}{x - t}\right]\rho(x)dx. \tag{II.24}$$

We have

$$|I_1| \leq 2ab\hat{\rho}_5\,\theta^{\mu}\ ,$$

$$\hat{\rho}_5 = \sup\Big|\int_{\gamma-\gamma_1}\frac{\rho(x)dx}{x-t}\Big|\ . \tag{II.25}$$

For I_2 there is the bound

$$|I_2| \leq 2ab\hat{\rho}_3\,\theta\int_{\gamma-\gamma_1}\frac{dx}{|x-t-\theta|^{1-\mu}|x-t|} =$$

$$2ab\hat{\rho}_3\,\theta\int_{\gamma-\gamma_1}\frac{dx}{|1-\theta/(x-t)|^{1-\mu}|x-t|^{2-\mu}}$$

In this integral $|x-t| \geq 2\theta$, so one has the inequality $|1-\theta/(x-t)|^{-1+\mu} \leq 2^{1-\mu}$. Thus

$$|I_2| \leq 2^{2-\mu}ab\hat{\rho}_3\,\theta\int_{\gamma-\gamma_1}\frac{dx}{|x-t|^{2-\mu}}$$

$$= \frac{2ab\hat{\rho}_3\,\theta^{\mu}}{1-\mu}\{2-(2\theta)^{1-\mu}[(3-t)^{\mu-1}+(2+t)^{\mu-1}]\}$$

$$\leq 2ab\,\theta^{\mu}\hat{\rho}_6\ ,$$

$$\hat{\rho}_6 = \frac{\hat{\rho}_3}{1-\mu}\sup|2-(2\theta)^{1-\mu}[(3-t)^{\mu-1}+(2+t)^{\mu-1}]|\ . \tag{II.26}$$

Thus by assembling (II.20), (II.21), (II.23), (II.25), and (II.26) we have

$$|\psi(t)-\psi(t')| \leq \{H[f]+|\lambda|[a^2r+2ab\hat{\rho}]\}\,|t-t'|^{\mu}\ ,$$

$$\hat{\rho} = \hat{\rho}_1+\hat{\rho}_4+\hat{\rho}_5+\hat{\rho}_6\ . \tag{II.27}$$

Finally, from (II.14), (II.18), and (II.27) we see that sufficient conditions for $A(K) \subset K$ are

$$\sup|f| + |\lambda|(a^2\hat{\rho}_1+2ab\hat{\rho}_2) \leq a\ ,$$

$$H[f]+|\lambda|[a^2r+2ab\hat{\rho}] \leq b\ . \tag{II.28}$$

Certainly there is a choice of a, b, and λ for which conditions (II.28) are satisfied. For instance if we take $a > \sup|f|$ and $b > H[f]$, then (II.28) is satisfied for sufficiently small $|\lambda|$.

Next we turn our attention to verifying the Lipschitz condition (II.1). Let φ_1 and φ_2 be any elements of K, and define $\psi_1 = A\varphi_1$, $\psi_2 = A\varphi_2$. We must find upper bounds for $\sup|\psi_1 - \psi_2|$ and $H[\psi_1 - \psi_2]$. We already know how to bound $\sup|\psi|$ and $H[\psi]$ by making use of the inequalities

$$\varphi^2 \leqslant a^2 ,$$

$$H[\varphi^2] \leqslant 2ab . \tag{II.29}$$

For the new bounds required we can merely take over our previous analysis after noting that the inequalities (II.29) are replaced by

$$|\varphi_1^2 - \varphi_2^2| \leqslant 2a \sup|\varphi_1 - \varphi_2| , \tag{II.30}$$

$$H[\varphi_1^2 - \varphi_2^2] \leqslant 2aH[\varphi_1 - \varphi_2] + 2b \sup|\varphi_1 - \varphi_2| . \tag{II.31}$$

Eq. (II.31) follows from the identity $f(x)g(x) = f(x)(g(x) - g(t)) + g(t)(f(x) - f(t))$, where $f = \varphi_1 + \varphi_2$, $g = \varphi_1 - \varphi_2$. By means of (II.18) and (II.27) we have

$$\sup|\psi_1 - \psi_2| \leqslant 2|\lambda|\{(a\hat{\rho}_1 + b\hat{\rho}_2) \sup|\varphi_1 - \varphi_2| + a\hat{\rho}_2 H[\varphi_1 - \varphi_2]\} , \tag{II.32}$$

$$H[\psi_1 - \psi_2] \leqslant 2|\lambda|\{(ar + b\hat{\rho}) \sup|\varphi_1 - \varphi_2| + a\hat{\rho}H[\varphi_1 - \varphi_2]\} . \tag{II.33}$$

By adding (II.32) and (II.33) we have

$$||\psi_1 - \psi_2|| = ||A\varphi_1 - A\varphi_2|| \leqslant \beta||\varphi_1 - \varphi_2|| ,$$

$$\beta = 2|\lambda| \max[a(\hat{\rho}_1 + r) + b(\hat{\rho}_2 + \hat{\rho}), a(\hat{\rho}_2 + \hat{\rho})] . \tag{II.34}$$

Thus if $|\lambda|$ is sufficiently small, the operator A acting on K obeys the Lipschitz condition with $\beta < 1$.

The result is summarized in Theorem 2: The integral equation

$$\varphi(t) = f(t) + \lambda P \int_0^1 \frac{\rho(x)\varphi^2(x)dx}{x - t} ,$$

$$f, \rho \in H(\mu); \quad \rho(0) = \rho(1) = 0, \quad \rho \geqslant 0 ,$$

has a unique solution φ within the set

$$K = \{\varphi \mid |\varphi| \leq a,\ |\varphi(t) - \varphi(x)| \leq b|t-x|^{\mu}\},\ 0 < \mu < 1\ ,$$

provided a, b, and λ satisfy the inequalities

$$\sup|f| + |\lambda|(a^2\hat{\rho}_1 + 2ab\hat{\rho}_2) \leq a\ , \tag{II.35a}$$

$$H[f] + |\lambda|(a^2 r + 2ab\hat{\rho}) \leq b, \tag{II.35b}$$

$$\beta = 2|\lambda| \max\left[a(\hat{\rho}_1 + r) + b(\hat{\rho}_2 + \hat{\rho}), a(\hat{\rho}_2 + \hat{\rho})\right] < 1\ . \tag{II.35c}$$

Here r, $\hat{\rho}$, $\hat{\rho}_i$ are positive constants depending on $\rho(x)$ only. The solution φ is obtained by iteration $\varphi_n = A\varphi_{n-1}$, beginning with an arbitrary element φ_0 of K; $\varphi_n \to \varphi$ in the sense that $||\varphi_n - \varphi|| \to 0$, where the norm is defined in (II.13). Also $||\varphi_n - \varphi|| \leq \beta^n(1-\beta)^{-1} \cdot ||\varphi_1 - \varphi_0||$, with β defined in (II.35c).

C. Analyticity of Solution with Respect to a Parameter

Since each iterate φ_n is a polynomial in λ, and since the inequalities (II.35) depend on λ only through $|\lambda|$, it seems a good guess that the limit φ would be analytic in λ in some region. Suppose that the inequalities (II.35) are satisfied for some λ such that $|\lambda| = \lambda_0$, and for some $a = a_0$, $b = b_0$. Then they are satisfied for the same a, b at all λ in the circle $|\lambda| \leq \lambda_0$ of the complex plane, and in fact the whole proof of Theorem 2 goes through for such complex λ. We can show by means of Vitali's theorem that the solution $\varphi(x;\lambda)$ referred to in Theorem 2 is actually analytic for $|\lambda| < \lambda_0$, at any fixed x. Vitali's theorem[17] is as follows: If the sequence of functions $\{f_n(z)\}$, analytic in an open connected set D, is uniformly bounded in D (i.e., $\sup_{z \in D}|f_n(z)| \leq M$), and it converges at the points of some set having a limit point in D, then it converges uniformly in every closed, bounded region of D. Consequently, its limit f(z) is analytic in D. Now since $|\varphi_n(x;\lambda)| \leq a_0$ for $|\lambda| \leq \lambda_0$ and all n, and since $\{\varphi_n(x;\lambda)\}$ converges at all λ in the same circle, we conclude immediately that at each x the solution $\varphi(x,\lambda)$ is analytic for $|\lambda| < \lambda_0$.

D. Brouwer and Schauder Fixed Point Theorems

The contraction mapping principle gives existence, uniqueness, and an actual construction of solutions, but it is not effective when the departure from linearity is too large. In the example (II.7) our upper bound for the Lipschitz coefficient β depended linearly on the parameter λ which measures the nonlinearity. It is natural to ask whether we can prove mere existence of solutions under weaker conditions, irrespective of existence questions or a method of constructing solutions. For that purpose we take a topological approach. We shall make use of Schauder's fixed point theorem, which is an infinite-dimensional generalization of the following fixed point theorem due to

Brouwer:[37)-50)]

Theorem 3: Let $f(x)$ be a continuous mapping of the closed unit ball S in R_n into itself. f has at least one fixed point in S.

On the real line R_1 this is intuitively clear: a continuous function $f(x)$ on the interval $[-1,1]$ such that $-1 \leq f(x) \leq 1$ must touch the line $y = x$ at one point x_0 at least, $-1 \leq x_0 \leq 1$. In n-dimensional Euclidean space R_n, the theorem is proved by first noting that if $x \neq f(x)$ for all x in the closed ball, then the distinct points x and $f(x)$ determine a unique straight line which intersects the bounding sphere in a point x^*. It is possible to show that the existence of the continuous mapping $x \to x^*$ for all x leads to a contradiction. The contradiction can be established by a simple application of the divergence theorem, without the topological apparatus that is sometimes used in proofs of Brouwer's theorem; cf. Saaty and Bram,[37)] p. 42. A corollary of Brouwer's theorem is as follows (cf. Saaty and Bram,[37)] p. 43):

Theorem 4: If K is a closed bounded convex subset of R_n, then a continuous mapping of K into itself has at least one fixed point.

(viii) A set K in a linear space is said to be convex if $\lambda x + (1 - \lambda)y \in K$ when $x, y \in K$, $0 \leq \lambda \leq 1$.

The generalization of Brouwer's theorem to Banach spaces must take account of the fact that a bounded set in an infinite dimensional space is not necessarily compact.

(ix) A set K in a metric space N is called compact if every infinite sequence in K has a subsequence which converges to a limit in K.

(x) A set K in a metric space N is called relatively compact if every infinite sequence in K has a subsequence which converges to a limit in N.

(Some authors use the term compact for what is here denoted relatively compact.) Schauder's theorem makes a specific assumption of compactness:[37)-50)]

Theorem 5: If K is a compact convex subset of a Banach space, then a continuous mapping of K into itself has at least one fixed point.

(xi) If $\varphi_n \to \varphi$ implies $A\varphi_n \to A\varphi$, where φ_n and φ belong to the domain of the operator A, then A is said to be continuous. A stronger form of Schauder's theorem is the following:

Theorem 6: Let K be a convex closed subset of a linear normed space. Suppose that A is a continuous mapping of K into itself such that A(K) is compact. Then A has at least one fixed point in K (cf. Cronin[42)]).

A proof of Theorem 5 can be based on Brouwer's theorem by making a sort of uniform approximation of the compact set K by a finite-dimensional subspace. Since the norm metric in a finite-dimensional Banach space is equivalent to a Euclidean metric, Brouwer's theorem can be applied in the subspace. The proof is as follows. First we note that a finite set of points in K can be found such that every point

of K is within a distance ϵ of some point of the set. We denote the points of this set by $\varphi_1, \varphi_2, \ldots \varphi_N$; they form a so-called ϵ-net. For any $\varphi \in K$ we have $||\varphi - \varphi_i|| < \epsilon$ for some i. It is easy to see that an ϵ-net must exist. In the contrary case if $\psi_1, \psi_2, \ldots, \psi_M$ is any finite set there is always some point ψ_{M+1} lying farther than ϵ from any of the ψ_i, $i = 1, \ldots, M$. But similarly there is a ψ_{M+2} lying farther than ϵ from any ψ_i, $i = 1, \ldots, M+1$, and so on. There is an infinite sequence ψ_i, $i = M+1, M+2, \ldots$ such that $||\psi_i - \psi_j|| > \epsilon$, all i,j. This sequence has a subsequence with a limit point ψ, since K is compact. In a ball of radius $\epsilon/3$ about ψ there are infinitely many points of the sequence ψ_i, any pair of them separated by less than $2\epsilon/3$. That contradicts $||\psi_i - \psi_j|| > \epsilon$.

Now by means of the ϵ-net we construct an interesting projection operator P which projects K into a closed convex finite-dimensional subspace of itself in such a way that $||P\varphi - \varphi|| < \epsilon$ for all φ in K. Define the following functions from K into the positive real line.

$$f_i(\varphi) = \begin{cases} \epsilon - ||\varphi - \varphi_i|| & \text{if } ||\varphi - \varphi_i|| < \epsilon, \\ 0 & \text{otherwise .} \end{cases} \tag{II.36}$$

Since $||\varphi - \varphi_i||$ is a continuous function of φ, the f_i are continuous and non-negative. Form also the functions

$$g_i(\varphi) = \frac{f_i(\varphi)}{\sum_{i=1}^{N} f_i(\varphi)} \quad . \tag{II.37}$$

Since at least one of the $f_i(\varphi)$ is positive, $g_i(\varphi)$ is well-defined and continuous. Also,

$$\sum_{i=1}^{N} g_i(\varphi) = 1,$$

$$g_i(\varphi) = 0, \qquad ||\varphi - \varphi_i|| \geq \epsilon \ . \tag{II.38}$$

The desired projection operator is

$$P(\varphi) = \sum_{i=1}^{N} g_i(\varphi)\, \varphi_i \ , \tag{II.39}$$

which takes K into the convex hull K_ϵ of the φ_i (i.e., the set of all linear combinations of the φ_i in which the coefficients add up to 1).

To show that $P(\varphi)$ approximates φ uniformly we note that

$$P(\varphi) - \varphi = \sum g_i(\varphi)(\varphi_i - \varphi) \quad , \tag{II.40}$$

from which follows

$$||P(\varphi) - \varphi|| \leq \sum g_i(\varphi)||\varphi_i - \varphi|| < \varepsilon \sum g_i(\varphi) = \varepsilon . \tag{II.41}$$

With our given operator A we form the continuous mapping

$$\varphi \to P(A(\varphi)), \tag{II.42}$$

which takes the convex hull K_ε of the φ_i into itself. Now K_ε is a bounded convex closed finite-dimensional subset of a Banach space, so by Theorem 4 the mapping $P(A\varphi)$ has a fixed point φ_ε in K_ε.

Finally we consider a sequence of progressively finer ε-nets, with $\varepsilon = 1/n$, n=1, 2, ... The corresponding projection operators and their subspaces will be labeled P_n and K_n respectively, and fixed points in K_n of the maps $\varphi \to P_n(A(\varphi))$ will be denoted $\varphi^{(n)}$. The sequence $\varphi^{(n)}$ has a subsequence with a limit φ in K. Denote this subsequence again by $\varphi^{(n)}$. The limit φ is seen to be the desired fixed point of A, by means of the decomposition

$$A(\varphi) - \varphi = [A(\varphi) - A(\varphi^{(n)})] + [A(\varphi^{(n)}) - P_n(A(\varphi^{(n)}))] + [\varphi^{(n)} - \varphi]. \tag{II.43}$$

The first bracket on the right vanishes in norm with increasing n, because of continuity of A. The norm of the second is less than $\varepsilon = 1/n$, because of (II.41). The last term vanishes in norm by definition of φ. Thus $||A(\varphi) - \varphi|| = 0$, or $A(\varphi) = \varphi$.

To apply Schauder's theorem to concrete integral operators we shall need a criterion for compactness of function spaces. This is provided by Ascoli's theorem, which is proved in almost any text on real analysis.

Theorem 7 (Ascoli): If S is an infinite set of functions on a closed bounded set in R_n, and if S is uniformly bounded and equicontinuous, then S contains a uniformly convergent sequence $\{f_n\}$.

(xii) A set S of functions f(x) defined on a domain D is uniformly bounded if $|f(x)| < M$, $x \in D$, with M independent of f. A set S of functions defined on D is equicontinuous if $|f(x) - f(y)| < \varepsilon$ when $|x - y| < \delta(\varepsilon)$, with δ independent of f.

From Ascoli's theorem it follows that our set K defined in Eq. (II.14) is a compact subset of the Banach space B_1 of Hölder-continuous functions with the norm $||\varphi|| = \sup|\varphi|$. (n.b.: not with the norm (II.13)). Indeed, the two inequalities which define K ensure that its

members are uniformly bounded and equicontinuous. Thus K contains a sequence $\varphi_n(t)$ which converges uniformly to a limit $\varphi(t)$; i.e., $|\varphi(t) - \varphi_n(t)| < \varepsilon$, $n > N(\varepsilon)$, which is the same as saying that $||\varphi - \varphi_n|| \to 0$.

It is now an easy matter to apply Schauder's theorem to our integral equation (II.7). We work in the Banach space B_1 and consider the action of the operator (II.7) on the set K of Eq. (II.14). K is compact and convex, and $A(K) \subset K$ if inequalities (II.35a) and (II.35b) are satisfied. We choose a, b, and λ so that the latter are satisfied. Then Schauder's theorem ensures existence of a fixed point of A, provided A is continuous acting on K. In our earlier work we established continuity with respect to the norm (II.13). With the present norm the integral must be estimated in a slightly different way, as is clear from Eq. (II.32). Let φ and $\bar{\varphi}$ be any two elements of K. We must show that $\sup|\psi - \bar{\psi}| = \sup|A\varphi - A\bar{\varphi}| < \varepsilon$ when $\sup|\varphi - \bar{\varphi}| < \delta$.

We subtract and add a pole term, as usual, to handle the principal value integral:

$$|\psi(t) - \bar{\psi}(t)| < |\lambda| \int_0^1 \rho(x) \left| \frac{\varphi^2(x) - \bar{\varphi}^2(x) - \varphi^2(t) + \bar{\varphi}^2(t)}{x - t} \right|$$

$$+ |\lambda| \; |\varphi^2(t) - \bar{\varphi}^2(t)| \left| \int_0^1 \frac{\rho(x)dx}{x - t} \right| . \qquad \text{(II.44)}$$

The second term in (II.44) is less than a constant times $||\varphi - \bar{\varphi}||$. In the first term we divide the integral into an integral over a small neighborhood Ω of $x = t$, and a remainder. Since φ and $\bar{\varphi}$ are H-continuous, the integral over Ω can be made less than $\varepsilon/2$ by choosing Ω to be sufficiently small. Since in the remaining integral the denominator does not vanish, we can bound the integral by a constant times $||\varphi - \bar{\varphi}||$. This too is less than $\varepsilon/2$ for $||\varphi - \bar{\varphi}||$ sufficiently small, and the continuity of the operator follows.

Without invoking the condition (II.35c) that ensured the contraction property of the operator, we still have an existence theorem:

Theorem 8: Suppose that the conditions of Theorem 2 hold, except for (II.35c). Then the integral equation (II.7) has at least one solution in the set K.

The approach we have taken with Eq. (II.7) is effective with much more general singular, nonlinear integral equations; see, for example, the book of Pogorselski. If end-point singularities are allowed (the singularities that we avoid by taking $\rho(0) = \rho(1) = 1$) then the analysis becomes rather delicate. For that case there are some results in the Russian mathematical literature.[53] The results I have seen do not allow a quadratic nonlinearity, however, so they cannot be applied directly to equations of the Low type. The papers of Ref. 53

nevertheless contain a wealth of technique on singular integrals which might prove to be useful in physical problems.

III. Fixed Point Theorems Applied to the Low Equation

A. Direct Study of Low Operator as a Singular Integral Operator

The one-meson Low equation[18] of the static baryon model is a relatively simple but still non-trivial equation of S matrix theory. It is a good test case for nonlinear methods, because its main features are typical of all S matrix equations. The essential mathematical difficulties are that (i) Cauchy principal value integrals are involved, and (ii) the nonlinearity is very large when coupling constants have their physical values. Especially because of item (ii) the Low equation is not well understood in the very case where it has some relation to physics; for instance, there is up to now neither a thorough mathematical analysis[19] nor a convincing numerical solution[20] when the coupling constant is large enough to produce a dynamical resonance or a bound state. Consequently, we do not know much about the conjectured bootstrap mechanism, even in the case of the Chew-Low theory which was supposed to be the prototype of a bootstrap theory. Thus the Low equation in the strong coupling region remains a mathematical challenge of some physical interest. In this section we apply the fixed point theorems to the Low equation, and find that although they give some interesting information they do not tell us enough about the strong coupling situation. For more detail on the work of this section, see Refs. 8 and 9. In Section IV we try to develop an approach more suitable to strong coupling.

The Low equation is written as follows:

$$f_\alpha(z) = \frac{\lambda_\alpha}{z} + \frac{1}{\pi}\int_1^\infty \frac{d\omega\, \rho(\omega)\left|f_\alpha(\omega + i0)\right|^2}{\omega - z} + \sum_{\beta=1}^{n} c_{\alpha\beta}\,\frac{1}{\pi}\int_1^\infty \frac{d\omega\, \rho(\omega)\left|f_\beta(\omega + i0)\right|^2}{\omega + z} \,. \qquad \text{(III.1)}$$

The meson energy is ω in a system of units where the meson mass is unity. The amplitude $f_\alpha(\omega + i0)$ describes scattering of pseudoscalar mesons in p-waves. In terms of the phase shift $\delta_\alpha(\omega)$ and the cut-off function $v(k)$ the amplitude is defined by

$$f_\alpha(\omega + i0) = \sin\delta_\alpha(\omega)\,\exp i\delta_\alpha(\omega)\,\rho^{-1}(\omega)\,,$$

$$\rho(\omega) = \frac{k^3 v^2(k)}{12\pi}\,,\quad k = (\omega^2 - 1)^{\frac{1}{2}}\,. \qquad \text{(III.2)}$$

The real parameters λ_α are sums of direct and cross channel baryon pole residues:

$$\lambda_\alpha = - g_\alpha^2 + \sum_\beta c_{\alpha\beta}\, g_\beta^2 \quad . \tag{III.3}$$

The crossing matrix $c = [c_{\alpha\beta}] = c^*$ is arbitrary except for the crossing symmetry condition $c^2 = 1$.

In (III.1) we take the limit $z \to \omega + i0$, $\omega > 1$, so that we have a nonlinear, singular integral equation. It is easy to show that an H-continuous solution of the integral equation yields via (III.1) an analytic function $f_\alpha(z)$ with all the desired properties (cf. Ref. 8). After the change of variable $t = 1/\omega$ the integral equation reads

$$\varphi_\alpha(t) = \lambda_\alpha t + i\rho\, \varphi_\alpha \varphi_\alpha^*(t) - \frac{t}{\pi} P \int_0^1 \frac{d\tau}{\tau} \frac{\rho\, \varphi_\alpha \varphi_\alpha^*(\tau)}{\tau - t}$$

$$+ \sum c_{\alpha\beta} \frac{t}{\pi} \int_0^1 \frac{d\tau}{\tau} \frac{\rho\, \varphi_\beta \varphi_\beta^*(\tau)}{\tau + t} \quad , \tag{III.4}$$

where $\varphi_\alpha(t) = f_\alpha(\omega)$, and $\rho(\omega(t))$ is written as $\rho(t)$. Now it is clear that one can apply the same methods that were used in the example (II.15). For instance to apply Schauder's theorem we regard the right hand side of (III.4) as an operator A on the space of complex Hölder-continuous n-component functions $\varphi(t) = [\varphi_1(t), \ldots, \varphi_n(t)]$ with the norm

$$||\varphi|| = \sup_{\substack{0 \le t \le 1 \\ 1 \le \alpha \le n}} |\varphi_\alpha(t)| \quad . \tag{III.5}$$

We again denote this Banach space by B_1. We define a convex, compact subset $K = K(a,b)$ of B_1 by the inequalities

$$|\varphi_\alpha(t)| \le a,$$

$$|\varphi_\alpha(t) - \varphi_\alpha(t')| \le b|t - t'|^\mu, \quad 0 < \mu < 1 .$$

$$t, t' \in [0,1], \quad a = 1, 2, \ldots, n. \tag{III.6}$$

By estimating integrals in much the same way as in Section II, it is possible to show that the operator A maps the set K into itself when the constants $\sup_\alpha |\lambda_\alpha|$, a, and b are sufficiently small. The inequalities

which guarantee $A(K) \subset K$ and which are analogous to Eq. (II.28) have the form

$$\sup|\lambda_\alpha| + M_1 a^2 + M_2 ab \leq a ,$$

$$\sup|\lambda_\alpha| + M_3 a^2 + M_4 ab \leq b . \qquad \text{(III.7)}$$

Here the M_i are positive constants which depend only on the cut-off function $\rho(\omega)$ and the quantity $\sum_\beta |c_{\alpha\beta}|$. The expression for the M_i and and other details of the proof are given in Ref. 8. It was assumed that the cut-off obeys the conditions

$$|\rho(t)/t - \rho(t')/t'| \leq h|t - t'|^\nu, \ \nu \geq \mu, \ t,t' \in [0,1] , \qquad \text{(III.8)}$$

$$\left.\frac{\rho(t)}{t}\right|_{t=0} = \left.\frac{\rho(t)}{t}\right|_{t=1} = 0 . \qquad \text{(III.9)}$$

By combining (III.8) and (III.9) one has $\rho(\omega) = 0(\omega^{-1-\nu})$, $\omega \to \infty$. (These conditions on the cut-off can be weakened, but since we shall get better results by a different method anyhow, we shall not bother to do so.) Because of the quadratic nature in a and b of the coefficients of M_i, it is clear that (III.7) can be satisfied when $\sup|\lambda_\alpha|$, a, and b are small. For instance if $a = b = 2 \sup|\lambda_\alpha|$, then (III.7) takes the form

$$4(M_1 + M_2) \sup|\lambda_\alpha| \leq 1,$$

$$4(M_3 + M_4) \sup|\lambda_\alpha| \leq 1. \qquad \text{(III.10)}$$

Also, the continuity of A acting on K is easily proved as in Section II. By Schauder's theorem it follows that with a cut-off obeying (III.8), (III.9), an arbitrary crossing matrix, and a sufficiently small coupling constant, the Low equation has at least one solution which satisfies inequalities (III.6), (III.7).

With a usual kind of cut-off the allowed range of coupling constants turns out to be disappointingly small. Also, the constant a is so small that resonances cannot occur in the solution under discussion.

The contraction mapping principle may also be applied with little additional effort, although that was not done in Ref. 8. As in Section II we have only to replace the norm (III.5) by

$$||\varphi|| = \sup_{t,\alpha} |\varphi_\alpha(t)| + \sup_{\substack{t,t',\alpha \\ t\neq t'}} \left| \frac{\varphi_\alpha(t) - \varphi_\alpha(t')}{|t - t'|^\mu} \right| . \qquad \text{(III.11)}$$

By making a replacement in the work of Ref. 8 analogous to (II.29) → (II.30), (II.31) we find that for any φ, $\bar{\varphi} \in K$ the Low operator A satisfies the Lipschitz condition

$$||A\varphi - A\bar{\varphi}|| \leq \beta ||\varphi - \bar{\varphi}|| , \qquad \text{(III.12)}$$

$$\beta = \max [aB_1 + bB_2, aB_3] , \qquad \text{(III.13)}$$

where the B_i are positive constants which depend only on the crossing matrix and the cut-off. Thus if a, b, and $\sup|\lambda_\alpha|$ are chosen to satisfy (III.7) and

$$\max[aB_1 + bB_2, aB_3] < 1 , \qquad \text{(III.14)}$$

we have a contraction mapping. In that case the Low equation has a unique solution in the set K of Eq. (III.6), and the solution is obtained by iteration $\varphi_n = A\varphi_{n-1}$ beginning with an arbitrary $\varphi_0 \in K$. Also, $||\varphi - \varphi_n|| \leq \beta^n (1 - \beta)^{-1} ||\varphi_1 - \varphi_0||$.

B. Analyticity in Coupling Constant

This iterative solution of the Low equation is analytic in the coupling constant in a certain circle. We write $\lambda_\alpha = \lambda\mu_\alpha$, μ_α independent of λ, and call λ the coupling constant. In the Chew-Low theory λ is proportional to f^2, while in more general theories it might be a parameter equal to a linear combination of several squared coupling constants [cf. (III.3)]. The argument for analyticity in λ follows Section II. From (III.4) it is clear that every iterate $\varphi^{(n)}$ is a polynomial in λ provided the zeroth iterate is such. The only requirement on $\varphi^{(0)}$ is that it lie in K, so we may choose $\varphi^{(0)}$ to be the Born term: $\varphi^{(0)} = \lambda\mu_\alpha t$. Then $\varphi^{(n)}$ is a polynomial in λ. If λ is such that inequalities (III.7) hold for some $a = a_0$, $b = b_0$, and $|\lambda| = \lambda_0$, we have $|\varphi^{(n)}(t;\lambda)| \leq a_0$ for all n and all λ in the circle $|\lambda| < \lambda_0$. Since $\varphi^{(n)}(t;\lambda)$ converges to a limit $\varphi(t;\lambda)$ for each λ in the circle, Vitali's theorem guarantees that the solution $\varphi(t;\lambda)$ is analytic in λ for each t, $|\lambda| < \lambda_0$.

C. Study of Low Equation in N/D Formulation

In order to extend our work to larger values of λ we can try a different formulation of the problem. Here we may make a vague analogy with potential scattering (or, equivalently, with Fredholm theory). The Born series (Neumann series) of potential scattering converges, in general, only for values of the potential strength $|\lambda|$ smaller than some λ_0^B. The scattering amplitude is analytic for $|\lambda| < \lambda_0^B$, where λ_0^B

is actually the radius of convergence of the power series. Analytic continuation to larger values of $|\lambda|$ is possible by means of Jost's method.[21] One goes over to the Jost functions $f(k;\lambda)$ which are entire functions of λ, and the S matrix is expressed as $S(k;\lambda) = f(k;\lambda)/f(-k;\lambda)$. The poles of S(bound states) which prevent the convergence of the Born series now occur as zeros of $f(-k;\lambda)$. This continuation to the entire λ plane is just the continuation of the Neumann series by means of Fredholm determinants. In the case of the Low equation we also have a power series which converges for $|\lambda| < \lambda_0$, but our λ_0 is by no means the radius of convergence of the series. In fact the contraction mapping theorem alone gives no hint as to how one could find the nearest singularity. Nevertheless, we can still try a generalization of the Jost method, as an attempt to extend the region of λ analyticity or otherwise to improve our results.

The generalized Jost method is based on the N/D representation,[22),23),23)] in which the amplitude $f_\alpha(z)$ of (III.1) is written as

$$f_\alpha(z) = N_\alpha(z)/D_\alpha(z) \ . \qquad \text{(III.15)}$$

$D_\alpha(z)$ is defined (not uniquely) as a function obeying the following conditions: (we drop the subscript α)

(i) $D(z)$ is meromorphic in the z plane with the cut $[1,\infty)$, and is bounded by a power at infinity: $D(z) = 0(|z|^n)$, $|z| \to \infty$,
(ii) $D(z) = D^*(z^*)$,
(iii) D is defined on the cut by continuity, its boundary value $D(\omega + i0)$, $\omega \in [1,\infty)$, is H–continuous in ω on any finite interval, and

$$D_+(\omega) \equiv D(\omega + i0) = |D_+(\omega)|e^{-i\delta(\omega)} \ ; \qquad \text{(III.16)}$$

δ is the phase shift of (III.2).

Since $D_- = e^{2i\delta}D_+$ and $f(z) = f^*(z^*)$,

$$N_+ - N_- = f_+D_+ - f_-D_- = (f_+ - f_- e^{2i\delta})D_+$$
$$= (f_+ - f_+)\, D_+ = 0 \quad , \qquad \text{(III.17)}$$

at any $\omega \geqslant 1$ which is not a pole of D_+. Thus N has no discontinuity over the cut $[1,\infty)$; its only singularities are possible poles, either at infinity or at points where D has a pole not cancelled by a zero of f. If f has a phase shift which is bounded and H–continuous on any finite interval, then the most general D satisfying (i)-(iii) is known to have the following form:[23)]

$$D(z) = R(z)\,\mathfrak{D}(z) = R(z)\exp\left[-\frac{z}{\pi}\int_1^\infty \frac{\delta(\omega)d\omega}{\omega(\omega - z)}\right] \qquad \text{(III.18)}$$

Here $R(z) = R^*(z^*)$ is a rational function. To prove (III.18), one shows that the only singularities of $D/\mathfrak{D}$ are poles, when D is any function satisfying (III.16 i, ii, iii); cf. Ref. 23.

Now suppose we look for solutions of (III.1) which have H-continuous phase shifts (i.e., on any finite interval $\delta(\omega)$ is H-continuous), and which vanish at infinity ($f_{\alpha+}(\omega) \to 0$). Since we shall always assume that $\rho(\omega)$ tends to zero as a power at infinity, the solutions sought have phases which are bounded as well as H-continuous. This follows from (III.2), which shows that the phase shift approaches an integer multiple of π as a power: $\delta(\omega) = n\pi + 0(\omega^{-\alpha})$, $\alpha > 0$. The corresponding $\mathfrak{D}$ has the behavior[23]

$$\mathfrak{D}(\omega) \sim \omega^{\delta(\infty)/\pi} \quad . \qquad \text{(III.19)}$$

There is a canonical class of D functions which is of interest in connection with Levinson's theorem. For a full explanation see Ref. 24. A canonical D here denoted by D_c has a zero for each stable particle pole of the amplitude f, and it has a certain number of simple poles at real points ω_{ci} such that $f(\omega_{ci}) = 0$. One of the ω_{ci} may be at infinity, if the phase tends to its limit from below. $D_c(z)$ has no other zeros or poles, and it behaves either as 1 or as z at infinity, depending on whether one of the ω_{ci} is or is not at infinity. The number of poles of D_c is denoted as n_c, and the number of zeros (the number of stable particles) is denoted by n_b. Under an assumption that 1 is not a characteristic value of the kernel of the N/D integral equation, the following Levinson relation holds:

$$\delta(\infty) - \delta(1) = -\pi(n_b - n_c) \quad . \qquad \text{(III.20)}$$

This relation holds for our solutions of the Low equation without the specific assumption of no unit characteristic value, provided the coupling constant is sufficiently small (cf. Ref. 8).

For the present work we use a different D function which is not canonical in general. We consider a certain class C_m of scattering amplitudes which have an N/D factorization in which D may be represented as

$$D(z) = \prod_{i=1}^{m} (z - \omega_i)^{-1}\,\mathfrak{D}(z) \quad ,$$

$$D(z) \sim 1, \qquad |z| \to \infty \quad , \qquad \text{(III.21)}$$

and in which N is regular at $\omega = \omega_i$, and the phase shift is H-continuous. Since (III.20) has no zeros on the physical sheet, the number m of poles is not necessarily equal to n_c. The class C_m may also be completely characterized by phase shift behavior. It consists of those amplitudes which have H-continuous phase shifts such that $\delta(\infty) - \delta(1) = m\pi \geqslant 0$ (this limit being approached rapidly enough to ensure $\mathcal{S}(z) \sim z^m$), and for which there are at least m finite points $\omega_i > 1$ such that $\sin \delta(\omega_i) = 0$. In the N/D representation based on (III.20), all stable particle poles occur in N. The D function (III.20) has a Cauchy representation as follows:

$$D(z) = 1 - \sum_{i=1}^{m} \frac{c_i}{\omega_i - z} - \frac{1}{\pi} \int_1^\infty \frac{\rho(\omega)\, N(\omega)\, d\omega}{\omega - z} \quad . \tag{III.22}$$

To derive the so-called N/D integral equation for class C_m amplitudes we examine a convenient function $\Lambda(z)$:

$$\Lambda(z) = [f(z) - B(z)]\, D(z) + \frac{1}{\pi} \int_1^\infty \frac{B(\omega)\,\mathrm{Im}D_+(\omega)\,d\omega}{\omega - z}$$

$$= N(z) - B(z)D(z) - \frac{1}{\pi} \int_1^\infty \frac{B(\omega)N(\omega)\,\rho(\omega)\,d\omega}{\omega - z} \quad . \tag{III.23}$$

Here B is the left-cut term of the Low equation, including the direct channel nucleon pole:

$$B_\alpha(z) = \frac{\lambda_\alpha}{z} + \sum_\beta c_{\alpha\beta} \frac{1}{\pi} \int_1^\infty \frac{d\omega\, \rho(\omega)\, |f_\beta(\omega + i0)|^2}{\omega + z} \quad . \tag{III.24}$$

Since D is given by (III.22) it is clear that Λ has no discontinuity over either the right or the left cut, and that Λ vanishes at infinity. Hence it is a rational function having poles only at the poles of -BD; viz.,

$$\Lambda(z) = \sum_{i=1}^{n} \frac{c_i\, B(\omega_i)}{\omega_i - z} \quad . \tag{III.25}$$

Now if we take the limit $z \to \omega + i0$, $\omega \geqslant 1$, and compare the real parts of (III.23) and (III.25) we have

$$N(\omega) = B(\omega) + \sum_{i=1}^{m} c_i \frac{B(\omega_i) - B(\omega)}{\omega_i - \omega} + \frac{1}{\pi}\int_1^\infty \frac{B(\omega) - B(\omega')}{\omega - \omega'} \rho(\omega')N(\omega')d\omega' . \qquad \text{(III.26)}$$

This is the required integral equation. Note carefully its logical status. It is a necessary condition on the N function of any class C_m amplitude, where $-\rho N = \mathrm{Im}D_+$, with D defined in (III.21). If a class C_m amplitude satisfying the Low equation exists, then we know that it has an N function satisfying (III.26) for some appropriate choice of the constants c_i, ω_i. For a given amplitude the N function and the integral equation are not necessarily unique, because there may be more than one choice of the ω_i.

We are mainly concerned with the converse question: after a definite choice of ω_i and c_i, when does a solution of (III.26) yield a class C_m solution of the Low equation? Since B depends on N itself, (III.26) is a nonlinear functional equation for N. Indeed, in (III.24) we must substitute

$$|f_\beta(\omega + i0)|^2 = \frac{[N_\beta(\omega)]^2}{\left[1 - \sum_i \frac{c_{i\beta}}{\omega_{i\beta} - \omega} - \frac{P}{\pi}\int_1^\infty \frac{\rho(\omega')N_\beta(\omega')d\omega'}{\omega' - \omega}\right]^2 + \left[\rho(\omega)N_\beta(\omega)\right]^2} \qquad \text{(III.27)}$$

According to the definition of Λ, the amplitude f is constructed from N by means of the formula

$$f(z) = B(z) + \frac{1}{D(z)}\left[\Lambda(z) + \frac{1}{\pi}\int_1^\infty \frac{\rho(\omega)N(\omega)B(\omega)d\omega}{\omega - z}\right] , \qquad \text{(III.28)}$$

where D and Λ are obtained from (III.22) and (III.25), and B from (III.24) and (III.27). For (III.28) to be a solution of the Low equation [when N solves (III.26), is H-continuous, and vanishes rapidly enough at infinity so that $D(z) \sim 1$], it is sufficient that D(z) have no zero on the physical sheet not shared by $N = fD$; [i.e., not shared by the expression in square brackets in (III.28)]. When that condition is met (III.28) has the proper location of singularities, and it vanishes uniformly as a power at infinity. Since the left cut term is given by (III.24) and (III.27), it follows that f is a solution of the Low equation if the discontinuity over the right cut is consistent with that over the left cut. There is indeed consistency, because when the constraint

of the integral equation (III.26) is combined with (III.28) we see that $f(w + i0) = N(w)/D(w + i0)$.

When D has no zero at all on the physical sheet (not even one shared with N) then (III.28) is a class C_m amplitude. If there is one simultaneous zero of N and D, but no other zero of D, then it is clear from (III.18), (III.19), and (III.22) that $\delta(\infty) - \delta(1) = (m - 1)\pi$, but there are nevertheless m distinct points where $\sin\delta(w_1) = 0$. Thus the amplitude belongs to class C_{m-1} if $m \geqslant 1$. If $m \geqslant 1$ we can then derive an N/D equation for a non-vanishing D function like (III.20) but with only m - 1 poles. On the other hand if m = 0 with one common zero of N and D we have the anomalous situation of an "extinct bound state."[24] By means of fixed point theorems we shall find solutions of (III.26) which have no zeros of D, and hence yield class C_m amplitudes which satisfy the Low equation. Zeros of D not shared by N are called "ghosts."

For simplicity we first study the case m = 0 in which no poles of D occur. The right-hand side of (III.26) is regarded as a nonlinear operator A applied to N; thus (III.26) is written N = AN. As before we work with the variable $t = 1/w$, and consider A first as an operator in the Banach space B_1 of H-continuous functions $\varphi(t) = [\varphi_1(t), \ldots, \varphi_n(t)]$ with norm $||\varphi|| = \sup_\alpha |\varphi_\alpha(t)|$. In this Banach space we apply Schauder's theorem (Theorem 5 of the previous section), choosing for the set K the following compact convex subset of B_1:

$$K = \{\varphi \mid |\varphi_\alpha(t) \leqslant at,\ |\varphi_\alpha(t) - \varphi_\alpha(t')| \leqslant b|t - t'|^\mu,$$
$$0 \leqslant t, t' \leqslant 1, \quad 0 < \mu < 1\}\ . \tag{III.29}$$

Here a, b, and μ are fixed; a and b must eventually be determined so that the conditions of Schauder's theorem can be verified. If $\psi = A\varphi$, where φ is any member of K, then the first step is to show that $\psi \in K$ when a and b are suitably restricted. The key estimate required comes from (III.27):

$$|f|_\varphi^2 = \frac{\varphi^2(t)}{\left[1 + \frac{t}{\pi} P \int_0^1 \frac{\rho(\tau)\varphi(\tau)d\tau}{\tau(\tau - t)}\right]^2 + \rho^2(t)\varphi^2(t)}$$
$$\leqslant \left[\frac{at}{1 - \sup\left|\frac{t}{\pi} P \int_0^1 \frac{\rho(\tau)\varphi(\tau)d\tau}{\tau(\tau - t)}\right|}\right]^2 . \tag{III.30}$$

The integral is easily bounded after subtracting and adding the pole term:

$$\left|\frac{t}{\pi} P \int_0^1 \frac{\rho(\tau)\varphi(\tau)d\tau}{\tau(\tau - t)}\right| \leq b \sup\left|\frac{t}{\pi}\int_0^1 \frac{\rho(\tau)d\tau}{\tau|\tau - t|^{1-\mu}}\right|$$

$$+ a \sup\left|\frac{t^2}{\pi} P \int_0^1 \frac{\rho(\tau)d\tau}{\tau(\tau - t)}\right|$$

$$= bA_1 + aA_2 \ . \qquad \text{(III.31)}$$

The A_i exist if the cut-off satisfies the conditions

$$\text{(i)} \quad |\rho(t) - \rho(\tau)| \leq k|t - \tau|^{\nu} \ ; \quad k, \nu > 0 \ ,$$
$$\text{(ii)} \quad \rho(0) = 0 \ . \qquad \text{(III.32)}$$

It follows from (i) and (ii) together that $\rho(\omega) = 0(\omega^{-\nu})$, $\omega \to \infty$. We assume (III.32) throughout the following. Now suppose that a and b are chosen so that $bA_1 + aA_2 < 1$. Then

$$|f|_\varphi^2 \leq \left[\frac{at}{1 - bA_1 - aA_2}\right]^2 \equiv a^2 \xi(a,b)t^2 \ . \qquad \text{(III.33)}$$

This estimate can be applied in a straightforward way to bound B and the difference quotient of B. It is then an easy exercise[9] to show that A maps K into itself provided the following inequalities hold:

$$[\sup_\alpha|\lambda_\alpha| + \sum_\beta |c_{\alpha\beta}|a^2 \xi A_3][1 + aA_3] \leq a, b \ ,$$

$$A_3 = \frac{1}{\pi}\int_0^1 \rho(t)\, dt \ . \qquad \text{(III.34)}$$

These conditions are simpler, and constitute much less of a restriction on the coupling constant, then the corresponding inequalities (III.7) of our direct attack on the Low eqaution. In the latter the M_i were large numbers containing many terms. The inequalities (III.34) may be satisfied when $\sup|\lambda_\alpha|$ is sufficiently small. If a, b, and the coupling constant are chosen to satisfy (III.34), then Schauder's theorem guarantees that the N/D equation has at least one solution provided A is a continuous operator on K. The continuity of A is easily proved without further assumptions.[9]

To rule out ghosts we use the Phragmén-Lindelöf theorem: Let $f(z)$ be analytic in the half-plane $y > 0$, continuous in the closed half-plane $y \geq 0$, and bounded on the real axis. Suppose that $f(z) = 0(e^{r\beta})$, $\beta < 1$, uniformly in θ for a sequence $r = r_n \to \infty$. Then $|f(z)| \leq M$ for $y \geq 0$. The conditions of this theorem hold for $f(z) = D(z) - 1$, and furthermore

$$|D(\omega + i0) - 1| \leq [(bA_1 + aA_2)^2 + (\sup(\omega^{-1}\rho)a)^2]^{\frac{1}{2}}, \ \omega \geq 1 \ . \quad \text{(III.35)}$$

Hence for a and b sufficiently small, $|D(z) - 1|$ is less than 1 in the entire upper half-plane. Thus there is no ghost in the upper half plane, and because of $D(z^*) = D(z)^*$ also none in the lower half plane, when a and b are chosen appropriately.

When CDD poles are allowed we have to contend with the expression

$$|f|_\varphi^2 = \frac{\varphi^2(t)}{\left[1 + \frac{t\hat{t}c}{\hat{t} - t} + \frac{t}{\pi} P\int_0^1 \frac{\rho(\tau)\varphi(\tau)d\tau}{\tau(\tau - t)}\right]^2 + \rho^2(t)\varphi^2(t)}$$

$$0 < \hat{t} < 1 \ . \quad \text{(III.36)}$$

The case of one CDD pole at $\hat{\omega} = 1/\hat{t}$ is taken as an example. When the residue c is sufficiently small there will always be a zero of the bracketed expression in (III.36) near $t = \hat{t}$. This zero represents a resonance (ReD = 0) if φ is a solution of the N/D equation. At the zero we have only the unitarity bound $|f|_\varphi^2 \leq \rho^{-2}$, instead of the small bound (III.33) proportional to a^2. We define a neighborhood Ω of the pole which is to be treated separately:

$$\Omega = \{t \mid (\hat{\omega} + \epsilon)^{-1} \leq t \leq (\hat{\omega} - \epsilon)^{-1}\},$$

$$\hat{\omega} - \epsilon > 1, \ \epsilon > 0 \ . \quad \text{(III.37)}$$

Then we have the following bounds in Ω and its complement $\bar{\Omega}$:

$$|f(t)|_\varphi^2 \leq \begin{cases} \left[\frac{at}{1 - |c/\epsilon| - aA_2 - bA_1}\right]^2 \equiv a^2\hat{\xi}t^2, \ t \in \bar{\Omega} \ , \\ \rho^{-2}(t) \ , \qquad t \in \Omega \ . \end{cases} \quad \text{(III.38)}$$

We have assumed that

$$|c/\epsilon| + bA_1 + aA_2 < 1 \ . \qquad \text{(III.39)}$$

Now we can proceed as before to get sufficient conditions that A map K into itself. These conditions are

$$[\sup|\lambda_\alpha| + \sum_\beta |c_{\alpha\beta}| (a^2 \hat{\xi}\hat{A}_3 + \frac{2\epsilon}{\pi} \sup_\Omega |\frac{1}{\rho}|)][1 + aA_3 + |c\hat{t}|] \leq a, b,$$

$$\hat{A}_3 = \frac{1}{\pi} \int_{\bar{\Omega}} \rho(\tau)\, d\tau \ . \qquad \text{(III.40)}$$

Inequalities (III.40) can be satisfied when a, b, $\sup|\lambda_\alpha|$, ϵ, and c are sufficiently small. The method has been to take advantage of the smallness of a^2 when $t \in \bar{\Omega}$, and the smallness of Ω when $t \in \Omega$.

To rule out ghosts and to prove the continuity of the operator we must proceed somewhat differently than in the absence of CDD poles. Since we now do not forbid zeros of $[\mathrm{Re}D]_\varphi$ it is possible for the numerator and denominator of (III.36) to vanish simultaneously. This causes difficulty in the proof of continuity (and in the verification of the Lipschitz property of the operator when we apply the contraction mapping theorem). This trouble is bypassed through a further restriction on the set K. In addition to (III.29) we now require

$$|\varphi_\alpha(t) - \lambda_\alpha t| \leq \gamma t, \ \gamma < \inf_\alpha |\lambda_\alpha| \ . \qquad \text{(III.41)}$$

This does not spoil the convexity of K, and it ensures that $\varphi_\alpha(t)$ cannot vanish for $t \neq 0$. At the same time it makes the integral in $[\mathrm{Re}D]_\varphi$ a Herglotz function.[25] Therefore if the CDD residue c_α is given the same sign as φ_α, the D function itself is a Herglotz function. A Herglotz function has no complex zeros, and a possible real zero of D can be ruled out by an inequality similar to (III.35). Thus (III.41) and the proper sign of c_α ensure that there are no ghosts. There is no simultaneous zero of φ and $[\mathrm{Re}D]_\varphi$, since φ has no zero except at $t = 0$, where $[\mathrm{Re}D]_\varphi = 1$.

A careful check shows that the three inequalities which define K can all be satisfied when $\sup|\lambda_\alpha|$, $|c|$, ϵ, a, and b are chosen appropriately. With the same choice of the constants the continuity of A is proved without difficulty. By Schauder's theorem we conclude that the N/D equation has at least one ghost-free solution when $\sup|\lambda_\alpha|$ and the CDD residues c_α are sufficiently small, provided λ_α and c_α have the same sign. We may also apply the contraction mapping principle if the Banach space B_2 is employed. As before B_2 resembles B_1 except that the term $H[\varphi]$ is added to the norm. As is expected the operator A is Lipschitz-continuous with coefficient β less

than one when $\sup|\lambda_\alpha|$ and c_α are sufficiently small. Under those conditions there is a unique solution in K, it is free of ghosts, and it may be obtained by iteration beginning with an arbitrary element of K. By this we mean, of course, that there is a unique solution after the CDD parameters are fixed.

The arguments go through in the same way if there is any finite number of CDD poles. When the residues $c_{\alpha 1}$ of the poles are small, each pole is associated with a narrow resonance. Hence our assertion that there is a solution of the Low equation for a nearly arbitrary choice of narrow resonances.

A proof that the N function obtained by iteration is analytic in λ is not quite so easy as the corresponding proof for the iterative solution of the Low equation. The n-th iterate of the Low equation is a polynomial in λ, while the n-th iterate of the N/D equation is a complicated rational function. The whereabouts of the poles of that function are still unknown.

IV. An Approach to the Strong Coupling Regime

A. Fréchet Derivatives

Our intention in this section is to overcome the limitations of the fixed point theorems. This will require (at least!) some differential calculus in Banach space. First we need an appropriate definition of the derivative of an operator.

Let X and Y be two Banach spaces, real or complex. Let $L(X;Y)$ be the set of all bounded linear operators from X to Y. $L(X,Y)$ is itself a Banach space with the norm

$$||L|| = \sup_{x \in X} \frac{||Lx||}{||x||} \ , \ L \in L(X,Y) \ . \qquad \text{(IV.1)}$$

(i) An operator F from an open set Ω of X into Y is Fréchet-differentiable (or differentiable) at $x \in \Omega$ if there is a linear operator $F'(x) \in L(X,Y)$ such that

$$\lim_{||h|| \to 0} \frac{1}{||h||} ||F(x+h) - F(x) - F'(x)h|| = 0 \ . \qquad \text{(IV.2)}$$

The linear operator $F'(x)$ is called the Fréchet derivative (or derivative) of F at x.

(ii) An operator F from an open set Ω of X into Y is twice differentiable at $x \in \Omega$ if it is differentiable in some ball centered at x and there is a linear operator $F''(x_0) \in L(X, L(X,Y))$ such that

$$\lim_{||h|| \to 0} \frac{1}{||h||} ||F'(x+h) - F'(x) - F''(x)h|| = 0 \ . \qquad \text{(IV.3)}$$

In the numerator of (IV.3) the norm is the norm (IV.1) of $L(X,Y)$. $F''(x)$ is called the second derivative of F at x. Since $F''(x)h \in L(X,Y)$ and $F''(x)hk \in Y$ if $h,k \in X$, it is natural to consider $F''(x)$ as a linear mapping from $X \times X$ into Y. According to (IV.1) the norm of $F''(x)$ is

$$||F''(x)|| = \sup_h \frac{||F''(x)h||}{||h||} = \sup_h \frac{1}{||h||} \sup_k \frac{||F''(x)hk||}{||k||} . \quad (IV.4)$$

Also,

$$||F''(x)hk|| \leq ||F''(x)h|| \cdot ||k|| \leq ||F''(x)|| \cdot ||h|| \cdot ||k|| . \quad (IV.5)$$

If F is a linear operator, $F'(x) = F$ and $F''(x) = 0$ for all $x \in X$. The Fréchet derivatives have the other properties that one would expect; e.g., the chain rule of differentiation holds, as do Taylor formulas, etc.[45),46)]

We obtain the simplest example of a Fréchet derivative by taking $X = R^n$, $Y = R^m$. Then $F:X \to Y$ is given by

$$y^i = f^i(x^1, x^2, \ldots, x^n), \quad i = 1, 2, \ldots, m \quad (IV.6)$$

and the first Fréchet derivative at x_0, if it exists, is the $m \times n$ Jacobian matrix operator

$$F'(x_0) = \left[\frac{\partial f^i}{\partial x^j}\right]_{x=x_0} ;$$

$$(F'(x_0)h)^i = \sum_j \left.\frac{\partial f^i}{\partial x^j}\right|_{x=x_0} h^j . \quad (IV.7)$$

If the second Fréchet derivative at x_0 exists it is the array

$$F''(x_0) = \left[\frac{\partial^2 f^i}{\partial x^j \partial x^k}\right]_{x=x_0} ;$$

$$(F''(x_0)hk)^i = \sum_{\ell,m} \left.\frac{\partial^2 f^i}{\partial x^\ell \partial x^m}\right|_{x=x_0} h^\ell k^m . \quad (IV.8)$$

It should now be clear how to find the derivatives of certain integral operators. For instance let X and Y both be the space of continuous

functions on $[0,1]$ with the norm $||x|| = \sup_{0\leq t\leq 1} |x(t)|$. Let $F:X \to Y$ be the Urysohn operator

$$y(s) = \int_0^1 f(s,t,x(t))dt, \qquad 0 \leq s \leq 1 , \tag{IV.9}$$

where f is continuous in its first two arguments and twice continuously differentiable with respect to the third. Then

$$F'(x)h = \int_0^1 f_x(s,t,x(t))h(t)dt \tag{IV.10}$$

and

$$F''(x)hk = \int_0^1 f_{xx}(s,t,x(t))h(t)k(t)dt \quad . \tag{IV.11}$$

B. Implicit Function Theorem

Let X and Y be Banach spaces, and form the direct product $X \times Y$ consisting of all pairs (x,y), $x \in X$, $y \in Y$. If we linearize $X \times Y$ by defining linear combinations as $\lambda(x,y) + \mu(\overline{x},\overline{y}) = (\lambda x + \mu\overline{x}, \lambda y + \mu\overline{y})$, then $X \times Y$ is a Banach space when suitably normed; e.g., by

$$||(x,y)|| = ||x|| + ||y|| \quad . \tag{IV.12}$$

We may now examine functions $U(x,y)$ of two variables, defined as maps $U:X \times Y \to Z$, where Z is again a Banach space. By fixing y at $y_0 \in Y$, we get a map from X into Z, the Fréchet derivative of which is called the partial-derivative with respect to x. It is a member of $L(X;Z)$ and is denoted by $U_x(x_0;y_0)$. Similarly one has $U_y(x_0,y_0) \in L(Y;Z)$.

We may now inquire about equations of the form

$$U(x,y) = 0 \tag{IV.13}$$

and the question of functions $y = F(x)$ defined by such equations. In S-matrix theory y might be the scattering amplitude, and x might represent input information (coupling constants, masses, elasticity function, etc.) on which y depends. The Hildebrandt-Graves implicit function theorem gives us conditions for existence, uniqueness, and continuity of $F(x)$. It is a precise analog of the ordinary implicit function theorem of advanced calculus.

Theorem 9: (Hildebrandt-Graves)[45),46)] Let $U\Omega \to Z$ be an operator defined on a neighborhood Ω of the point (x_0,y_0) of a product of Banach spaces $X \times Y$, with range in a Banach space Z. Suppose that U meets the following requirements:

(i) U is continuous at (x_0, y_0),
(ii) $U(x_0, y_0) = 0$,
(iii) U_y exists in Ω and is continuous at (x_0, y_0),
(iv) The operator $U_y(x_0, y_0) \in L(\Omega_y, Z)$ has an inverse

$$\Gamma = [U_y(x_0, y_0)]^{-1} \in L(Z, \Omega_y),$$

where

$$\Omega_y = \{y | (x_0, y) \in \Omega\} \quad .$$

Then an operator $F: G \to Y$ exists, where G is some neighborhood of x_0 in X, having the properties

(a) $U(x, F(x)) = 0$, $x \in G$,
(b) $F(x_0,) = y_0$,
(c) F is continuous at x_0.

The operator F is the only one with the properties (a, b, c) in the sense that if there is another operator $\hat{F}$ with the same properties, then there exists an $\epsilon > 0$ such that $F(x) = \hat{F}(x)$ when $||x - x_0|| < \epsilon$. By assuming more in the way of continuity and differentiability of U, we of course get more continuity and differentiability of F. If in Theorem 9 U is assumed to be continuous in all of Ω, then F will be continuous in some neighborhood of x_0. On the other hand if U_x exists in Ω and is continuous at x_0, y_0 then $F'(x_0)$ exists and satisfies

$$U_x(x_0, y_0) + U_y(x_0, y_0) F'(x_0) = 0 \ .$$

C. Index of the Fréchet Derivative of the Low Operator; Comments on a Paper of Lovelace

As an example we identify $U(\lambda, \varphi) = 0$ as the Low equation, where φ is the scattering amplitude and λ the coupling constant as defined in Section III. Thus the space X is one-dimensional. For Y we take the space of Hölder-continuous $2n$-component complex functions $\varphi(t) = [\varphi_1(t), \ldots, \varphi_n(t), \varphi_1(t)^*, \ldots, \varphi_n(t)^*]$ defined on $[0, 1]$. The norm is $||\varphi|| = \sup_{\alpha, t} |\varphi_\alpha(t)|$, and φ_i^* is the complex conjugate of φ_i. The linear operator $U_\varphi(\varphi^0): Y \to Y$ can be read off from the Low equation $\varphi - A(\varphi) = U(\lambda, \varphi) = 0$ as written in (III.4). If $h \in Y$ then the first n components of $U_\varphi(\varphi)h = g$ are as follows:

$$g_\alpha(t) = h_\alpha(t) - i\rho\,\varphi_\alpha h^*_\alpha(t) - i\rho\,\varphi^*_\alpha h_\alpha(t)$$
$$+\frac{t}{\pi} P\int_0^1 \frac{d\tau}{\tau}\,\rho(\tau)\,\frac{[\varphi_\alpha h^*_\alpha(\tau) + \varphi^*_\alpha h_\alpha(\tau)]}{\tau - t}$$
$$-\sum_\beta c_{\alpha\beta}\,\frac{t}{\pi}\int_0^1 \frac{d\tau\,\rho(\tau)}{\tau}\,\frac{[\varphi_\beta h^*_\beta(\tau) + \varphi^*_\beta h_\beta(\tau)]}{\tau + t}\,,$$

$$\alpha = 1,\ldots,n. \qquad \text{(IV.14)}$$

The n equations (IV.14), together with the corresponding complex conjugate equations, define $U_{\varphi\rho}$ in terms of a matrix of linear, singular integral operators. This operator is of a type thoroughly studied by Noether, Vekua, Muskhelishvili, and others; see Muskhelishvili,[16] Chapters 6, 18 (§ 129), and 19. It does not have an inverse in general. In fact we shall find that it fails to have an inverse in some cases of physical interest. The lack of an inverse is related to the existence of CDD manifolds of solutions of the Low equation.

In order that $U_{\varphi\rho}$ have an inverse it is necessary (but not sufficient) that it have index zero. The index κ is defined as $\kappa = k - k'$, where k is the dimension of the null space of $U_{\varphi\rho}$, and k' is the dimension of the null space of its adjoint $U^+_{\varphi\rho}$. Here the adjoint is obtained by transposing (without complex conjugation) the infinite matrix representing $U_{\varphi\rho}$. Since $k = \kappa + k'$, and $k' \geqslant 0$ it is evident that $\kappa = 0$ is necessary for the existence of an inverse. The index can be computed in a simple way, since it is determined by the dominant part $U^0_{\varphi\rho}$ of the operator. Suppose we are given a singular integral operator K, where $K\psi$ has the form

$$K\psi(t) = \sum_{\beta=1}^n A_{\alpha\beta}(t)\psi_\beta(t) + \frac{1}{\pi i}\int_L \sum_{\beta=1}^n \frac{K_{\alpha\beta}(t,\tau)\psi_\beta(\tau)d\tau}{\tau - t}\,, \qquad \text{(IV.15)}$$

$K_{\alpha\beta}$ and $A_{\alpha\beta}$ being given H-continuous functions on L, and L being a simple smooth closed contour in the complex t-plane. Then the dominant part K^0 of K is defined as follows:

$$K^0\psi(t) = \sum_{\beta=1}^n A_{\alpha\beta}(t)\psi_\beta(t) + \sum_{\beta=1}^n \frac{B_{\alpha\beta}(t)}{\pi i}\int_L \frac{\psi_\beta(\tau)d\tau}{\tau - t}\,,$$

$$B_{\alpha\beta}(t) = K_{\alpha\beta}(t,t)\,. \qquad \text{(IV.16)}$$

If the matrices $S = A + B$ and $D = A - B$ are both non-singular on L, then the index of K is

$$\kappa = \frac{1}{2\pi}\left[\arg \frac{\det(A - B)}{\det(A + B)}\right]_L , \qquad \text{(IV.17)}$$

where $[X]_L$ is the change in X in one circuit of L.

Our operator $U_\varphi(\varphi)$ of (IV.14) has the form (IV.15), provided φ is H-continuous, and ρ satisfies conditions (III.8), (III.9). In that case we can let the interval $[0,1]$ be part of a smooth simple closed contour L. On $L - [0,1]$ we set $\rho(t)/t \equiv 0$, so $\int_0^1 = \int_L$. The contribution of the left-cut term (last term in (IV.14)) to $\kappa_{\alpha\beta}$ is proportional to $c_{\alpha\beta} t(\tau - t)/\tau(\tau + t)$, and hence the left-cut term does not affect B. For A and B we have the $2n \times 2n$ matrices

$$A = \left[\begin{array}{c|c} 1 - i\rho\varphi_\alpha^* & -i\rho\varphi_\alpha \\ \hline i\rho\varphi_\alpha^* & 1 + i\rho\varphi_\alpha \end{array}\right] , \quad B = \left[\begin{array}{c|c} i\rho\varphi_\alpha^* & i\rho\varphi_\alpha \\ \hline i\rho\varphi_\alpha^* & i\rho\varphi_\alpha \end{array}\right] , \qquad \text{(IV.18)}$$

where the indicated $n \times n$ blocks are diagonal with diagonal elements corresponding to $\alpha = 1,2,\ldots,n$. The first n rows (columns) correspond to h_α, the remaining n rows (columns) to h_α^*. Recalling the definition (III.2), one sees that

$$\det(A + B) = \exp\left(2i \sum_{\alpha=1}^{n} \delta_\alpha\right),$$

$$\det(A - B) = \exp\left(-2i \sum_{\alpha=1}^{n} \delta_\alpha\right) , \qquad \text{(IV.19)}$$

where δ_α is the phase shift that goes with the solution φ_α. As the theory of the index requires, $A + B$ and $A - B$ are non-singular. By (IV.17) we find that the index is

$$\kappa = \frac{2}{\pi} \sum_\alpha \delta_\alpha(\infty) , \qquad \text{(IV.20)}$$

where $\delta_\alpha(1) = 0$.

If $\varphi_\alpha(t)$ is a solution like that obtained in Section III, and if at least one of the channels has a CDD pole, then $\kappa > 0$ and the Fréchet derivative $U_\varphi(\varphi)$ lacks an inverse. (Recall that the solutions we discussed in Section III all had $\delta_\alpha(\infty) \geq 0$, and $\delta_\alpha(\infty)/\pi = m_\alpha$ was the

number of CDD poles in channel α.) Note also that for the solutions of Section III, $\kappa = 2m$ is also equal to the number of arbitrary parameters that entered the N/D equations; viz., the number m of CDD pole residues plus the number m of CDD pole positions. This count of parameters does not include the stable particle parameters (positions and residues of particle poles) which occur explicitly in the Low equation and in our N/D equation of Section III; they number $2\sum_{\alpha} n_{b\alpha}$.

Lovelace[5] first called attention to the importance of the Fréchet derivative of the integral operator of S matrix equations. He noted, on the basis of a theorem of Vainberg and Trenogin,[26] that the index of the derivative at φ is equal (barring certain degeneracies) to the dimension of the manifold of solutions to which φ belongs. This interpretation of κ emerges rather more explicitly in our N/D formulation. Also, our computation of the index is a good deal simpler than the one of Lovelace. His index differs from ours, by the way, through a term $2\sum_{\alpha} n_{b\alpha}$ since his derivative includes variations of the stable particle pole positions and residues as well as variation of φ.

D. Fréchet Derivative of the N/D Operator; Continuous Dependence of Solutions on Parameters

Since the N/D equations contain explicitly the stable particle parameters and the CDD parameters, it will not be surprising to find that the Fréchet derivative of the nonlinear N/D operator has zero index. In fact it is a compact (i.e., Fredholm) operator: it is continuous and it maps every bounded set of the appropriate Banach space into a relatively compact set. To prove that assertion we regard the N/D operator (defined in Eq. (III.26)) as an operator on the Banach space B_2 [i.e., H-continuous functions with norm given by (III.11)]. Then the Fréchet derivative of $U(N) = N - AN$, a linear operator on the same space, is given by the following expression (in which the derivative is denoted as $U_N(N_0)$, and h is a member of B_2):

$$[U_N(N_0)h]_{\alpha,t} = h_\alpha(t) + \frac{t}{\pi}\int_0^1 \frac{d\tau\,\rho(\tau)}{\tau} G(t,\tau)_{N_0,\alpha} h_\alpha(\tau)$$

$$- 2\sum_\beta c_{\alpha\beta}\frac{t}{\pi}\int_0^1 \frac{d\tau\rho(\tau)}{\tau}[1 - I_\alpha(\tau)]\left\{\frac{1}{\tau+t}\left[\frac{N_0}{|D_0|^2}\left(1 - \frac{\rho^2 N_0^2}{|D_0|^2}\right)\right]_{\beta,\tau}\right.$$

$$\left. + \frac{P}{\pi}\int_0^1 \frac{d\sigma\,\rho(\sigma)}{\tau-\sigma}\,\frac{1}{\sigma+t}\left[\frac{N_0^2 \mathrm{Re} D_0}{|D_0|^4}\right]_{\beta,\sigma}\right\} h_\beta(\tau)\,, \qquad \text{(IV.21)}$$

The definitions of G and I are

$$G(t,\tau)_{N_0,\sigma} = \frac{B_\alpha(t) - B_\alpha(\tau)}{t-\tau}$$

$$= \lambda_\alpha + \sum_\beta \frac{c_{\alpha\beta}}{\pi} \int_0^1 \frac{d\sigma\, \rho(\sigma)}{(\sigma+t)(\sigma+\tau)} \left[\frac{N_0^2}{\rho^2 N_0^2 + (\mathrm{Re}D_0)^2}\right]_\beta ,$$

$$I_\beta(\tau) = \frac{\tau}{\pi}\int_0^1 \frac{dt\rho(t)}{t(t+\tau)}\, N_{0\beta}(t) \ . \qquad \text{(IV.22)}$$

We assume conditions (III.9) on $\rho(t)$, here and in the following. In computing the last term of (IV.21) we have performed a reversal of integration order. In an initial straightforward evaluation of this term (by formal variation of U(N)), h_β occurs inside a principal value integral. In turn the principal value integral is inside a Riemann integral. Since $N_{0\alpha}, h_\alpha \,|D_0|^{-\infty}$ and ρ/t are H-continuous, it is permissible to reverse the order,[27] so that h_α finally occurs under a Riemann integral. The resulting kernel is of Fredholm type.

We rewrite (IV.21) in abbreviated notation:

$$(h + Kh)_{\alpha,t} = h_\alpha(t) + \sum_\beta \int_0^1 K_{\alpha\beta}(t,\tau)\, h_\beta(\tau)d\tau \ . \qquad \text{(IV.23)}$$

Now $U_N(N_0)^{-1}$ exists if $||K|| < 1$,[46] and the latter condition can be met when $||N_0||$, $H[N_0]$, the coupling constant, and the CDD pole residues are sufficiently small. Thus if N_0 is one of the solutions we have discussed in Section III, and the coupling and CDD parameters are sufficiently small, we are assured that the inverse of $U_N(N_0)$ exists. As the coupling constant λ is increased we may reach a value $\lambda = \lambda_0$ where the inverse ceases to exist. Increases in the CDD residues might have the same effect.

We shall be interested in the special case where U_N is singular, but first we shall see what can be learned from the implicit function theorem in the non-singular case. Unlike the Fréchet derivative of the Low operator, the derivative of the N/D operator is expected to have an inverse almost everywhere in the space of the parameters. We write $U(\lambda, N) = N - AN$, and note that the entire λ dependence of U is linear. Separating the term that depends on λ we have

$$U(\lambda,N)_{\alpha,t} = N_\alpha(t) - \lambda\mu_\alpha t \left[1 - \sum_{i=1}^{n_c^\alpha} c_i^\alpha t_i^\alpha - \frac{1}{\pi}\int_0^1 \frac{d\tau\rho(\tau)}{\tau} N_\alpha(\tau)\right]$$

$$+ G_{\alpha,t}(N), \qquad \text{(IV.23)}$$

where $G(N)$ is independent of λ. The coefficient of $-\lambda\mu\, t$ in (IV.23) is just the denominator function $D_\alpha(\omega)$ evaluated at $\omega = 0 (t^{\alpha} \to \infty)$. Consequently, if N is a ghost-free solution of the N/D equations, the coefficient of λ in $U(\lambda, N)$ will not be zero. Now we suppose that

$$U(\lambda_0, N_0) = 0 , \tag{IV.24}$$

where $(\lambda_0, N_0) \in R \times B_2$, and proceed to verify the following properties of $U(\lambda, N)$: Let Ω be some neighborhood of (λ_0, N_0) in $R \times B_2$. Then

(a) $U(\lambda, N)$ is defined and continuous in Ω,

(b) $U_N(\lambda, N)$ exists in Ω and is continuous at (λ_0, N_0),

(c) $U_\lambda(\lambda, N)$ exists in Ω and is continuous at (λ_0, N_0).

From (IV.23) it is clear that proving point (a) comes down to proving continuity of $G(N)$ with respect to N in Ω. For the latter we refer to Section III and, for more detail, Eqs. (IV.19), (IV.20), (IV.21), (IV.22), (IV.30), and (IV.31) of Ref. 9. From the equations quoted, especially (IV.21), it is clear that continuity of $G(N)$ in Ω will follow if we can bound $|D|^{-2} = 1/y$ in Ω. Note that $D_0(t)$ cannot have a zero in the physical region $[0,1]$ unless $N_0(t)$ has a zero at the same point; that is the requirement of unitarity. We shall assume for the following argument that there is no such common zero. Then $|D_0|^{-2}$ is bounded in Ω, and consequently so is $|D|^{-2}$ if Ω is a sufficiently small region. We know this because $||D - \bar{D}|| = 0(||N - \bar{N}||)$, where $N, \bar{N} \in \Omega$ and $D, \bar{D}$ are the corresponding denominator functions. The matter of existence and continuity of $U_N(\lambda, N)$ in Ω is essentially the same story again, in view of the formula (IV.21) of the present work; i.e., we verify continuity by arguments like those of Ref. 9. $U_\lambda(\lambda, N)$, being merely a linear functional of N independent of λ, clearly satisfies condition (c). Now by application of the implicit function theorem we arrive at the following result:

Theorem 10: Let N_0 be an H-continuous solution of the N/D equation $U(\lambda, N) = 0$ corresponding to coupling constant λ_0: $U(\lambda_0, N_0) = 0$. Suppose that $U_N(\lambda_0, N_0)^{-1}$ exists and that N_0 and the corresponding denominator D_0 do not have a common zero on the physical cut $1 \leq \omega < \infty$. Then there is an $\epsilon > 0$ and a function $N(\lambda)$ such that for $\lambda_0 - \epsilon \leq \lambda \leq \lambda_0 + \epsilon$ we have

$$U(\lambda, N(\lambda)) = 0, \tag{IV.25}$$

$N(\lambda)$ being continuous in λ with $N(\lambda_0) = N_0$. Also, $N(\lambda)$ is differentiable at λ_0, and its derivative satisfies the equation

$$U_\lambda(\lambda_0, N_0) + U_N(\lambda_0, N_0) \left.\frac{dN}{d\lambda}\right|_{\lambda_0} = 0 . \tag{IV.26}$$

Furthermore, $N(\lambda)$ is the only function having these properties in some interval $\lambda_0 - \eta \leq \lambda \leq \lambda_0 + \eta$.

In other words, when we sit at a solution where U_N^{-1} exists, we can always increase the coupling constant by a little bit to get a new neighboring solution (which is unique). It is perhaps reasonable to conjecture that the coupling constant may be indefinitely increased until (i) $U_N(\lambda, N)^{-1}$ ceases to exist, or (ii) $||N||$ tends to infinity, or (i) and (ii) occur together.

E. Newton-Kantorovich Method; Solutions with Successive Increments of the Coupling Constant

We now turn to the question of how to compute $N(\lambda)$ when $N(\lambda_0)$ is known, under the assumption that $U_N^{-1}(\lambda_0, N_0)$ exists. The problem of computation at a singularity of U_N will occupy us later. Since $N(\lambda_0)$ will be (in some sense) an approximate solution of $U(\lambda, N) = 0$ when λ is close to λ_0, the question is how to find an exact solution when an approximate solution is known. For this the best known procedure is that of Kantorovich[45] and others: a generalization to arbitrary Banach spaces of the ancient Newton method of successive linearizations. Let $U: X \to Z$ be a mapping of a Banach space X into a Banach space Z. Consider the sequence $x_{n+1} = x_n - [U'(x_n)]^{-1} U(x_n)$, as successive approximations to a solution of $U(x) = 0$.

Theorem 11 (Newton-Kantorovich): Let the following four conditions be satisfied.

(1) $U''(x)$ exists in a ball $||x - x_0|| < \rho$ (ρ defined below), and in this ball

$$||U''(x)|| \leq \kappa ; \tag{IV.27}$$

(2) $U'(x_0)$ has an inverse and

$$||U'(x_0)^{-1}|| \leq \beta ; \tag{IV.28}$$

(3) $$||x_1 - x_0|| \leq \eta ; \tag{IV.29}$$

(4) $$h = \kappa\beta\eta \leq \tfrac{1}{2} \quad . \tag{IV.30}$$

The radius ρ of (1) is $\rho = \eta h^{-1}(1 - \sqrt{1 - 2h})$. Then $U(x) = 0$ has a solution x^* in the closed ball $||x - x_0|| \leq \rho_0$ and the sequence $\{x_n\}$ converges to x^*. At the n-th iteration we have the error estimate

$$||x_n - x^*|| \leq 2^{-n}(2h)^{2n}\frac{\eta}{h} \quad . \tag{IV.31}$$

Instead of (1) let $||U''(x)|| \leq \kappa$ in the ball $||x - x_0|| < \sigma$, where $\sigma = \eta h^{-1}(1 + \sqrt{1 - 2h})$. In that case x^* is a unique solution of $U(x) = 0$ in $||x - x_0|| < \sigma$.

For applications it is important to know that exactly the same statements hold for the sequence $\overline{x}_{n+1} = \overline{x}_n - [U'(x_0)]^{-1}U(\overline{x}_n)$, except that the rate of convergence may be slower than (IV.31) indicates. Instead of (IV.31) one has the bound

$$||x_n - x^*|| \leq [1 - \sqrt{1 - 2h}]^{n+1} \eta/h \ . \qquad \text{(IV.32)}$$

This "modified Newton method" requires only one operator inversion, so it may be more practical in some cases.

Now consider the possibility of generating a "solution curve" $N(\lambda)$ of the N/D equation $U(\lambda, N) \equiv 0$ in the space $R \times B_2$ by successive applications of Newton's method. We begin at a point (λ_0, N_0), where $U(\lambda_0, N_0) = 0$. If we wish to solve $U(\lambda, N) = 0$ for λ close to λ_0, a reasonable candidate for the zeroth approximation is

$$N_0 + \left.\frac{dN}{d\lambda}\right|_{\lambda_0} (\lambda - \lambda_0) \ , \qquad \text{(IV.33)}$$

where $dN/d\lambda$ is obtained from Eq. (IV.26):

$$\left.\frac{dN}{d\lambda}\right|_{\lambda_0} = - U_N(\lambda_0, N_0)^{-1} \ U(\lambda_0, N_0) \ . \qquad \text{(IV.34)}$$

This procedure was suggested by Anselone and Moore,[14] who also considered a formula quadratic in λ. One expects that by taking $|\lambda - \lambda_0|$ sufficiently small, the quantity $||x_1 - x_0||$ will be small enough to guarantee success of Newton's method. A bound on $||U_{NN}(\lambda, N)||$, as in (IV.27), can be established by looking at the explicit formula for $U_{NN}hk$, $(h, k \in B_2)$, provided that $|D|^{-2}$ is bounded in $||N - N_0|| < \rho$. This latter condition was encountered before, in our application of the implicit function theorem. We remark that $U_{NN}hk$ involves at least one of the functions h, k inside a Cauchy principal value integral.

Anselone and Moore[14] have proved a theorem which guarantees that it is possible to take a step in λ of a certain size, if one starts at a point where the Fréchet derivative is non-singular. The theorem also guarantees that the solution curve $N(\lambda)$ generated by successive steps is isolated from other solutions. The theorem is concerned with a class of operators which does not include our N/D operator, however. Although a similar theorem might be established in our case, it would be somewhat academic in the absence of numerical work.

F. Newton-Kantorovich Method at a Point where the Fréchet Derivative Lacks an Inverse; Anselone-Moore Change of Parameter

The paper of Anselone and Moore,[14] which is concerned with Newton's method applied to a problem in elasticity theory, also contains a suggestion for handling the case of a singular first derivative. Both their method and their physical interpretation of the singular derivative seem directly relevant to our problem. If $U_N(\lambda_0, N_0)$ is singular, they study instead of $U(\lambda, N)$ a new operator

$$V(\gamma, N) = U(\lambda(\gamma, N), N) , \qquad \text{(IV.35)}$$

where γ is a new real parameter, and $\lambda(\gamma, N)$ is an appropriately chosen functional of N. Clearly a solution of $V(\gamma, N) = 0$ is a solution of $U(\lambda, N) = 0$ for $\lambda = \lambda(\gamma, N)$. If γ can be varied so as to put $\lambda(\gamma, N)$ in the desired range of λ, we get the required solutions of $U(\lambda, N) = 0$ by solving instead $V(\gamma, N) = 0$. The important point is that $V_N(\gamma_0, N_0)$ may be non-singular when $U_N(\lambda_0, N_0)$ is singular; here $\lambda_0 = \lambda(\gamma_0, N_0)$. Indeed we have

$$V_N(\gamma_0, N_0) = U_N(\lambda_0, N_0) + U_\gamma(\lambda_0, N_0)\lambda_N(\gamma_0, N_0) , \qquad \text{(IV.36)}$$

and the second term on the right may remove the singularity. Presently we shall mention necessary and sufficient conditions for $V_N^{-1}(\gamma_0, N_0)$ to exist, but first we discuss the peculiar behavior of the solution curve $N(\lambda)$ at the singularity of U_N. Suppose that $\gamma = \gamma(\lambda, N)$ is sufficiently differentiable and compute

$$\frac{d\gamma}{d\lambda} = \gamma_\lambda(\lambda, N) + \gamma_N(\lambda, N)\frac{dN}{d\lambda} . \qquad \text{(IV.37)}$$

By (IV.34) it is seen that $dN/d\lambda$ can tend to infinity in norm if λ approaches a value λ_0 where $U_N(\lambda_0, N_0)$ is singular. Hence $d\lambda/d\gamma$ may be zero at the point (λ_0, N_0). Suppose for instance that the curve of λ versus γ [i.e., the inverse of the curve $\gamma(\lambda, N(\lambda))$] is something like the curve shown in Figure 1. The point γ_0 of zero slope corresponds to the singularity of U_N. An interesting observation is that for a value of λ just less than λ_0 there are two points (λ_1, γ_1), (λ_1, γ_2) on the curve, and these will correspond in general to two different solutions $N_1(\lambda_1)$, $N_2(\lambda_1)$ at the same value λ_1 of λ. The situation can be indicated schematically as in Figure 2 where we plot $N(\lambda)$ as a real number rather than as the member of an infinite dimensional space that it really is.

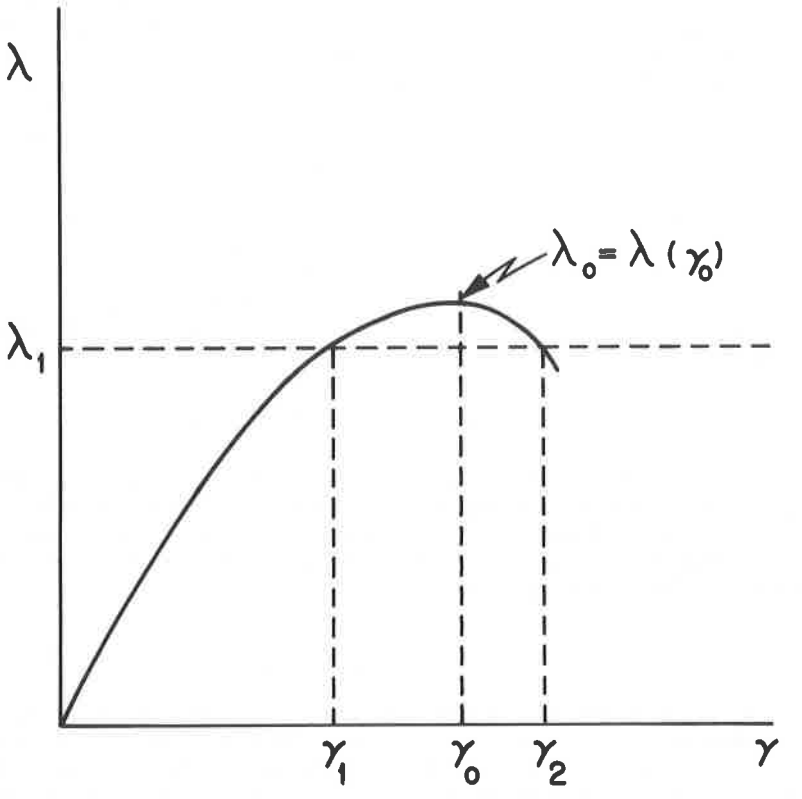


Figure 1

Curves of the type of Figure 1 have been computed numerically by Anselone and Moore in the case of a nonlinear integral equation meant to describe buckling of a thin shallow spherical shell clamped at the edge, under uniform external pressure. The equation reads

$$x_i(s) = g_i(s) + \lambda \int_0^1 L_{i1}(s,t)x_1(t)x_2(t)dt + \lambda \int_0^1 L_{i2}(s,t)\,\tfrac{1}{2}x_1^2(t)\,dt,$$

$$0 \leqslant s \leqslant 1,$$

$$i = 1, 2. \qquad \text{(IV.38)}$$

where x_1 is (slope of displacement function) $\div$ (pressure) and x_2 is (radial stress) $\div$ (pressure). The parameter λ, analogous to our coupling constant, is proportional to the pressure. It lies in the range $0 \leqslant \lambda < \infty$. The functions $g_i(s)$ and $L_{ij}(s)$ are known. For sufficiently small values of λ the fixed point theorems ensure existence of a solution. As in the case of the Low equation the physically interesting situations correspond to larger values of the coupling constant. Beginning with small values of λ, Newton's method is applied with successive increments of λ. That generates a solution curve $x(\lambda) = [x_1(\lambda), x_2(\lambda)]$ which is defined outside the λ interval where the fixed point theorems apply. Eventually a singularity of the Fréchet derivative is encountered, at which point the following change of parameter is invoked:

$$\gamma = \lambda \int_0^1 x_1(t)\, t^2\, dt\ . \qquad \text{(IV.39)}$$

The new parameter γ represents volume displacement of the shell. The Fréchet derivative V_x of $V(\gamma, x) = U(\lambda(\gamma, x), x)$ is found to be non-singular where U_x is singular. By solving $V(\gamma, x) = 0$ while varying γ, the solution curve is continued until a singularity of V_x is reached. That occurs where $d\gamma/d\lambda$ is infinite. At that point one changes back to λ, and so on, until a "global" solution curve is generated for a sufficiently large range of λ. It is not guaranteed in general that this curve will extend to indefinitely large values of λ, since when γ is being used the required changes of γ may be such as to decrease λ. The required changes of γ are those which send us forward rather than backward along the solution curve $x(\lambda)$. That λ may not increase indefinitely is also suggested schematically in Figure 2. In the calculations of Ref. 14, λ did not change monotonically at first, but eventually it did increase monotonically. Several local maxima and minima in the curve of λ versus γ were found. A sample curve from Ref. 14 is shown in Figure 3. Near each local extremum we have pairs of solutions corresponding to the same λ but different γ's, and the two solutions of a pair merge into one at the extremum.

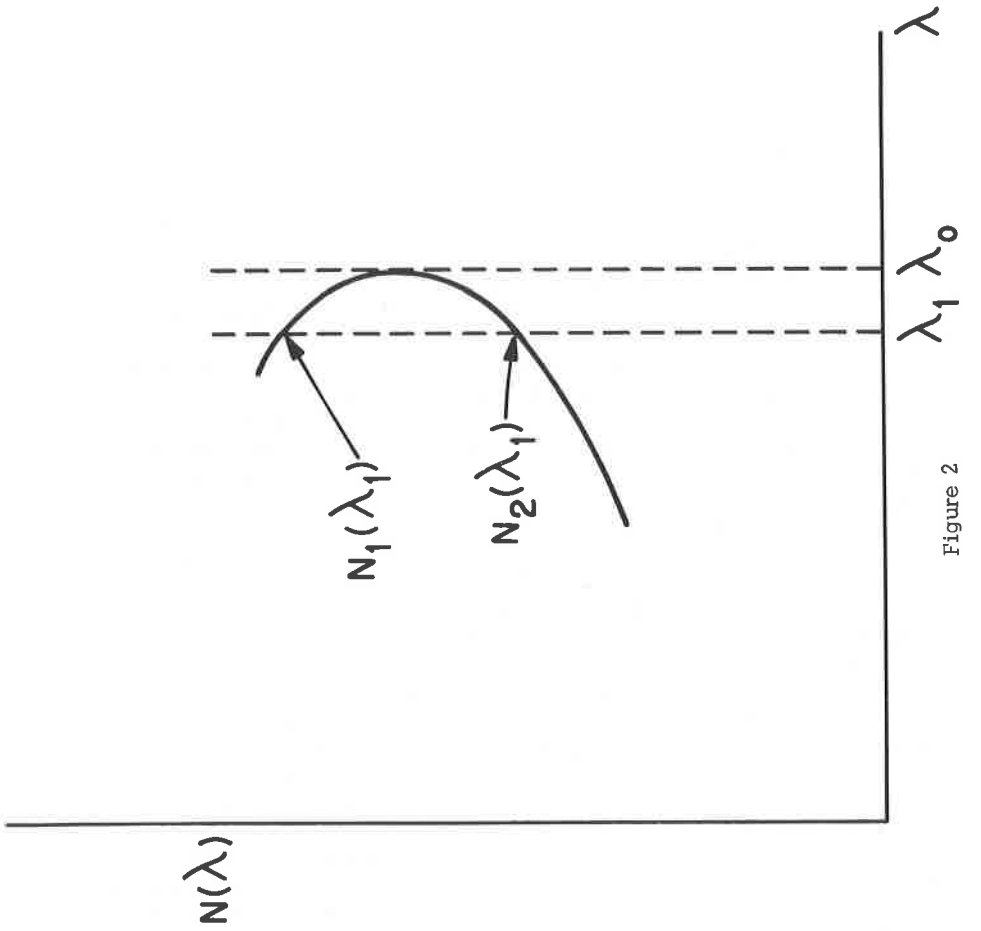


Figure 2

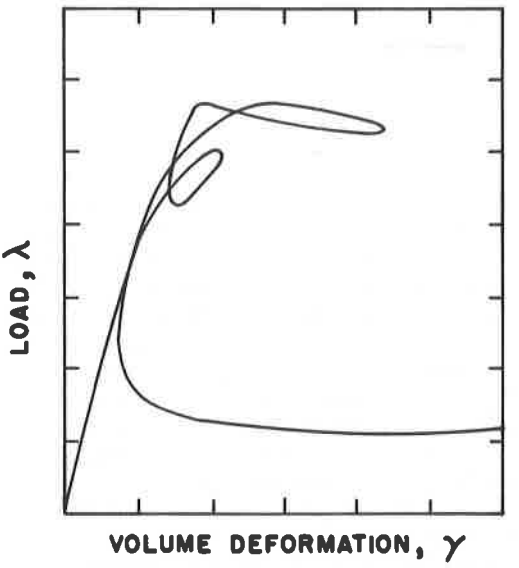


Figure 3

G. Bifurcation Points

Existence of pairs of solutions which merge into single solutions is an example of a general phenomenon which has long been appreciated in the theory of nonlinear equations.

Consider an operator $U(\lambda, x): R \times B \to B$ where $\lambda \in R$ is a real parameter, and B a Banach space. (More generally we could discuss the case where λ belongs to an arbitrary Banach space.) Let x_0 be a solution of $U(\lambda, x) = 0$ for $\lambda = \lambda_0$:

$$U(\lambda_0, x_0) = 0 \ . \tag{IV.41}$$

The point λ_0 in the parameter space is called a bifurcation point if for any $\epsilon, \delta > 0$ there is a λ with

$$|\lambda - \lambda_0| < \epsilon \tag{IV.42}$$

and two distinct vectors x_1, x_2 such that

$$U(\lambda, x_1) = 0 \quad , \quad U(\lambda, x_2) = 0 , \tag{IV.43}$$

$$||x_1 - x_0|| < \delta , \; ||x_2 - x_0|| < \delta \; . \tag{IV.44}$$

In view of (IV.40) it is clear that the λ of each local extremum on the Anselone-Moore curve of λ vs. γ is a bifurcation point with respect to the λ variable. The points of infinite slope are bifurcation points in γ.

Under conditions where the implicit function theorem applies, with $U_x(\lambda_0, x_0)$ non-singular, there cannot be any continuous bifurcation from λ_0. That is to say that in a sufficiently small neighborhood of λ_0 there is only one solution $x(\lambda)$ of $U(\lambda, x) = 0$ which depends continuously on λ in such a way that $x_0 = x(\lambda_0)$. This leaves open the possibility of sequences of solutions $\{x_1(\lambda_i)\}$, $\{x_2(\lambda_i)\}$ with $x_{1,2}(\lambda_i) \to x_0$, $\lambda_i \to \lambda_0$. However, under the rather general circumstances the singularity of U_x is a necessary condition for bifurcation of any sort, continuous or not. The problem of sufficient conditions for bifurcation is more difficult. For more information on bifurcation theory one may refer to a review by Vainberg and Aizengendler,[52] to the books of Krasnosel'skii,[40),41)] or to a recent article by Bancroft et al,[28] which contains further references.

In order to pursue further the analogy between our problem and that of Anselone and Moore, we note that a natural change of variable is suggested when the N/D equation is written in the form (IV.23). In analogy to (IV.40) we can choose

$$\gamma = \lambda D_\beta(0),$$

$$D_\beta(0) = 1 - \left(\sum_{i=1}^{n_c} c_i t_i \right)_\beta - \frac{1}{\pi} \int_0^1 \frac{d\tau \; \rho_\beta(\tau)}{\tau} N_\beta(\tau) \; . \tag{IV.45}$$

At present there seems to be no reason to prefer any particular choice of the channel index β. In terms of γ the N/D equation $V(\gamma, N) = 0$ reads

$$N_\alpha(t) - \gamma \mu_\alpha t \, D_\alpha(0)/D_\beta(0) + G_{\alpha,t}(N) = 0 ,$$

$$\alpha = 1, 2, \ldots n; \; \beta \text{ fixed}. \tag{IV.46}$$

This change of variable has two advantages over others that one might consider. First, $D_\beta(0)$ is non-vanishing as long as we do not have ghosts. (We suppose that channel β does not have a stable particle

corresponding to a zero of D_β.) Thus the change of parameter is always well-defined, since if ghosts appear one throws away the theory or else readjusts the CDD parameters to eliminate them. Second, (IV.45) is a linear relation between γ and λ for fixed N, which means that every solution of $U(\lambda, N) = 0$ is also a solution of $V(\gamma, N) = 0$ for some γ. Indeed if $U(\lambda_0, N_0) = 0$ then $V(\gamma, N_0) = 0$ for $\gamma = \lambda_0 D_\beta(0; N_0)$. This is seen from

$$V(\gamma, N) = U\left(\frac{\gamma}{D_\beta(0;N)}, N\right) . \qquad \text{(IV.47)}$$

The outline of a numerical solution of the Low equation in the N/D formulation might be as follows. Begin with the coupling constant λ and CDD residues $c_{i\alpha}$ so small that the contraction mapping theorem guarantees a unique solution in a certain ball of the Banach space. Generate this solution by iteration, beginning for instance with $N^0_\alpha(t) = \lambda_\alpha t$. Follow this solution as λ (and perhaps the $c_{i\alpha}$) are increased, by successive applications of either the Newton method or ordinary iteration. Presumably the Newton method is eventually necessary as λ increases. The derivative U_N is non-singular for λ and $|c_{i\alpha}|$ less than certain limits, but for large values of the parameters U_N may become singular. In that case one tries to continue with Newton's method after making the change of parameter (IV.45). It would be a good wager that a singularity of U_N does appear, but we would not make a guess as to whether it occurs before ghosts zero of D comes in. (If a ghost tries to enter, by the way, it might possibly be chased away through adjustment of CDD parameters.) If the change of parameter $\lambda \to \gamma$ succeeds, i.e., V_N is non-singular, then one may proceed along the global solution curve $N(\lambda)$. As the curve is followed several singularities of U_N may be encountered, interleaved with singularities of V_N, and at each singularity of U_N one changes from λ to γ, at each singularity of V_N, from γ to λ. Throughout the calculation the phase shifts should be monitored, in order to detect the entry of ghosts. According to the discussion in Section III, appearance of a ghost will be reflected in a change in the asymptotic phase shift behavior. A ghost pole counts as a bound state pole in Levinson's relation.

In connection with the proposed numerical solution there will be two points of particular physical interest. First, is there a ghost-free solution curve $N(\lambda)$ going on to arbitrarily large values of λ? That is, do there exist solutions to the Low equations corresponding to arbitrarily large values of the coupling constant? Second, do there exist bifurcation points and the attendant pairs of solutions corresponding to the same coupling constant? Note that two solutions near a bifurcation point can correspond to the same CDD poles as well as the

same coupling, so in deciding which solution to take as physical, one is dealing with an ambiguity which goes beyond the usual CDD ambiguity.

The question remains of whether V_N will indeed be non-singular when U_N is singular. In a case such as ours where $U_N - 1$ is compact, the following proposition is true.

Theorem 12: Suppose $U_N(\lambda_0, N_0)$ is singular, and $1 - U_N(\lambda_0, N_0)$ is compact. There exist changes of variable such that $V_N(\gamma_0, N_0)$ is non-singular if and only if

(a) the null space of $U_N(\lambda_0, N_0)$ is one-dimensional,

(b) $U_\lambda(\lambda_0, N_0) \notin U_N(\lambda_0, N_0)B$, where B is the Banach space in which N lies.

Under these conditions $V_N(\gamma_0, N_0)$ is non-singular if and only if

(c) $\lambda_N(\gamma_0, N_0)h \neq 0$, where h is a member of the null-space of $U_N(\lambda_0, N_0)$.

This follows rather directly from the definitions; I will not spell out the proof.

In the notation of Eq. (IV.23), we have

$$U_\lambda(\lambda_0, N_0)_{\alpha,t} = -\mu_\alpha t D_\alpha(0)\Big|_{N_0}, \tag{IV.48}$$

whereas with the change of variable (IV.45) we have

$$(\lambda_N(\gamma_0, N_0)h)_\alpha = \gamma D_\beta(0)^{-2}\Big|_{N_0} \frac{1}{\pi}\int_0^1 \frac{d\tau\ \rho(\tau)}{\tau}\ h_\alpha(\tau)\ . \tag{IV.49}$$

In our example where U_N is given by (IV.21) there appears to be no way to verify conditions (a), (b) by analytical arguments. It will be necessary to do the numerical work.

There may be other useful ways to apply the Newton method to the Low equation. In place of the rather painful systematic generation of the solution curve $N(\lambda)$ beginning at small λ, one might hope to guess at the outset an approximate solution close to an actual solution of physical interest. If the guess is close enough, then the Newton method can be used to "trim up" to the actual solution. There are several candidates for the approximate solution to be found in the literature. For instance there are prescriptions for constructing an approximate solution by combining matrix elements of ordinary S matrix perturbation theory. In this category are the "determinantal" expansion,[29),30)] the inverse amplitude expansion,[30)] and the Cini-Fubini or Padé approximant expansion.[31)] For the most part there has been

no investigation of whether these proposals are actually close (in any reasonable sense) to real solutions.

An alternative to these iteration procedures with unknown convergence properties is to take a hint from experiment, and try a saturation of the left cut integral by the observed resonances in zero width approximation. That was done by Chew[32] in the problem of pion-nucleon scattering. He included only the N and N^* poles. His goal was somewhat different from ours, however, in that he doesn't really try to solve the Low equation. He assumes that the D functions in the (11) and (33) states are linear in ω over a sufficiently large region: $D(\omega) = a + b\omega$. The constants a and b are then chosen to be consistent with the bootstrap picture in which both N and N^* are composite states. Chew makes no comment on whether the resulting scheme is consistent (or even whether it should be consistent) with the Low equation. With almost the same viewpoint one can avoid the linear D function assumption , and at the same time take a first step toward consistency with the Low equation. After replacing the left cut by poles one can find a solution of the dispersion relation in closed form by means of the N/D equation; the latter has a degenerate kernel in the pole approximation. Crossing symmetry is then imposed approximately by requiring that the residues and positions of corresponding poles in the direct and crossed channels be in agreement. This can be done with both N and N^* composite, or with one or both elementary (i.e., associated with a CDD pole). In the former case the crossing conditions are very restrictive. They can be satisfied (if at all) only by very special cut-off functions. If such a cut-off can be found, then the corresponding scattering amplitude may be chosen as the starting point of a Newton iteration, with some hope of the iteration converging to an actual bootstrap solution. Huang and Mueller[33] have shown that a bootstrap solution is impossible with a certain class of cut-offs, so the cut-off would have to be different from theirs.

There are strong indications from experiment that the nucleon is not "simply composite";[34] i.e., that it is not generated as a zero of D in a single channel N/D equation without the nucleon (or an equivalent CDD pole) as input. The N/D equation referred to includes inelasticity through an elasticity factor η. Both η and the left-cut term are taken from experiment. The nucleon could still be "multiply composite," which means that it could be generated in corresponding many-channel N/D equations, without the nucleon as input. The Chew-Low theory is a single-channel set-up, however, so the implication of Ref. 34 is that Chew's theory of 1962 is not correct:[35] the nucleon is not a composite state from the one-channel viewpoint. Therefore the solutions of the Low equation of greatest interest may be

those studied in Section III, but continued to the region of bigger coupling constants. Recall that those solutions indeed corresponded to a non-simply-composite nucleon.

In any case the approximation of the left-cut term by n poles can be used as a starting point for a systematic procedure of solution. Then the N/D method and the crossing conditions give n nonlinear algebraic equations for the residues. These algebraic equations may be attacked by the Newton method, perhaps with the Anselone-Moore procedure of advancing a parameter. When n is sufficiently large a solution of the algebraic equations should yield a sufficiently good starting point for a Newton iteration on the N/D equation with the exact left-cut expression. Amatuni[36] suggested this sort of scheme. However, he worked with the inverse amplitude method rather than the N/D equations, and that makes the handling of CDD poles very awkward.

Acknowledgments

Part of this work was done at the Theoretical Study Division of CERN, Geneva, and at the Institute for Theoretical Physics, Boulder. I am indebted to those institutions for hospitality and support.

I enjoyed a collaboration with H. McDaniel, and I have received advice and encouragement from several persons: H. Epstein, D. Atkinson, J. G. Taylor, J. M. Cook, and R. H. Moore. None of these people is responsible for any foolishness that these notes may contain.

References

1. The methods we describe seem to be most effective when applied to on-mass-shell equations of S matrix theory. One of the earliest examples of such an equation is that of Low: F. Low, Phys. Rev. 97, 1392 (1955); G. F. Chew and F. Low, ibid. 101, 1570 (1956); G. C. Wick, Rev. Mod. Phys. 27, 339 (1955); an introduction in textbook style is to be found in Ref. 18. Relativistic cousins of the Low equation are the equations of Chew-Mandelstam and Shirkov: G. F. Chew and S. Mandelstam, Phys. Rev. 119, 467 (1960); G. F. Chew, S. Mandelstam, and H. P. Noyes, ibid., 119, 478 (1960); G. F. Chew and S. Mandelstam, Nuovo Cimento 19, 752 (1961); V. V. Serebryakov and D. V. Shirkov, Forts. d. Phys. 13, 227 (1965) (a review of work by Shirkov et al). Equations incorporating Regge asymptotic behavior have been studied by Chew and collaborators, by Cheng and Sharp, and by Kaus and collaborators; this work appears in the Physical Review, 1963 to the present. For a view of recent work on strong interaction dynamics, including off-shell equations and finite-energy sum rules, see the contributions of Blankenbecler and

Mandelstam in Proceedings of the 1967 International Conference on Particles and Fields, Rochester, N. Y. (Interscience Publishers, New York, 1967).

2. R. E. Cutkosky and P. Tarjanne, Phys. Rev. 132, 1354 (1963); P. Tarjanne, Phys. Letters 18, 183 (1965); K. Maki, Prog. Theoret. Phys. (Kyoto) 31, 698 (1964); K. Nakamura, ibid. 33, 251 (1965); K. Ishida, ibid. 33, 256 (1965); G. Domokos and P. Suranyi, J. Nucl. Phys. (U.S.S.R.) 2, 501 (1965) (translation: Sov. J. Nucl. Phys. 2, 361 (1966)); R. Brout, Nuovo Cimento 67, 932 (1967); P. de Mottoni and E. Fabri, ibid. 54A, 42 (1968); L. Michel and L. A. Radicati, Fifth Coral Gables Conference on Symmetry Principles at High Energy (W. H. Freeman & Co., San Francisco, 1968). L. Michel, Lectures in Theoretical Physics, ed. by K. T. Mahanthappa, Volume XI-A (Gordon and Breach Science Publishers, New York, 1969). A. Pais, Phys. Rev. 173, 1587 (1968). The papers listed above are mostly concerned with group-theoretical and analytical aspects of nonlinear "consistency" equations. Linear questions in symmetry-breaking were studied by Dashen, Frautschi, and others. See Ref. 15. For the field theoretical aspects of symmetry breaking see J. Goldstone, Nuovo Cimento 19, 154 (1961), and many succeeding papers on Goldstone's theorem. A review of the ideas surrounding Goldstone's work has been giving by T. W. B. Kibble, Proceedings of the 1967 International Conference on Particles and Fields, Rochester, N. Y. (Interscience Publishers, New York, 1967).
3. J. G. Taylor, J. Math. Phys. 7, 1720 (1966); 6, 1148 (1965); Proc. Camb. Phil. Soc. 63, 335 (1967).
4. J. G. Taylor, Lectures in Theoretical Physics, edited by W. E. Brittin and A. O. Barut (Gordon and Breach Science Publishers, Inc., New York, 1967), Vol. IX-A, p. 353.
5. C. Lovelace, Commun. Math. Phys. 4, 261 (1967).
6. D. Atkinson, J. Math. Phys. 8, 2281 (1967).
7. It is likely that Atkinson's proof can be rearranged so as to rule out ghosts, provided one is willing to add some extra left-hand poles to satisfy threshold behavior. This matter is being investigated by M. Mashkour and the author.
8. R. L. Warnock, Phys. Rev. 170, 1323 (1968).
9. H. McDaniel and R. L. Warnock, "Resonant Solutions of the Low Equation from Fixed Point Theorems," CERN preprint TH. 905; submitted for publication in the Physical Review.
10. D. Atkinson, Nuc. Phys. 7, B378 (1968); ibid. 8, B377 (1968). See also J. Kupsch, "A Class of Functions That Satisfy Crossing Symmetry and Unitarity," CERN preprint TH. 984, Jan. 15, 1969.

11. A. Martin, Nuovo Cimento 59A, 131 (1969).
12. R. G. Newton, J. Math. Phys. 9, 2050 (1968).
13. Regge poles in the right half-plane require subtractions in the Mandelstam representation. In that case the fixed point method of Ref. 10 is still applicable, but there is a severe difficulty in showing that the inelastic unitarity inequalities can be satisfied. D. Atkinson, private communication, and Lectures in Theoretical Physics, edited by K. T. Mahanthappa et al (Gordon and Breach Science Publishers, Inc., New York, 1969), Volume XI-A.
14. P. M. Anselone and R. H. Moore, J. Math. Anal. and App. 13, 476 (1966).
15. R. F. Dashen and S. C. Frautschi, Phys. Rev. 135, B1190 (1964) 137, B1318, B1331 (1965); 140, B698 (1965); R. F. Sawyer, ibid. 142, 991 (1966).
16. N. I. Muskhelishvili, Singular Integral Equations (P. Noordhoff Ltd., Groningen, Holland, 1953).
17. See almost any text on complex analysis, for instance E. C. Titchmarsh, The Theory of Functions (Oxford University Press, Oxford, 1952).
18. E. M. Henley and W. Thirring, Elementary Quantum Field Theory (McGraw-Hill, New York, 1962), Chap. 18.
19. In the case of a 2 x 2 crossing matrix, however, there are some exact results. See A. W. Martin and W. D. McGlinn, Phys. Rev. 136, B1515 (1964); J. Rothleitner, Z. Physik 177, 287 (1964).
20. Salzman and Salzman solved the Low equation numerically by iterating the inverse amplitude equation. Their method failed beyond a certain strength of the coupling at which zeros of the inverse amplitude (ghosts) developed. G. Salzman and F. Salzman, Phys. Rev. 108, 1619 (1957).
21. R. G. Newton, Scattering Theory of Waves and Particles (McGraw-Hill, New York, 1966).
22. G. F. Chew and S. Mandelstam, Phys. Rev. 119, 467 (1960).
23. G. Frye and R. L. Warnock, Phys. Rev. 130, 478 (1963); R. L. Warnock, ibid. 131, 1320 (1963); Chapter 10 of Lectures in Theoretical High Energy Physics, H. Aly, ed. (Wiley-Interscience, New York, 1968).
24. D. Atkinson and M. B. Halpern, Phys. Rev. 149, 1133 (1966).
25. A Herglotz function is one which maps the upper (or lower) half-plane into itself; cf. Frye and Warnock, Ref. 23.
26. M. M. Vainberg and V. A. Trenogin, Uspekhi Mat. Nauk 17, #2, 13 (1962).
27. See Ref. 16, Chap. 3, Eq. (23.9).
28. S. Bancroft, J. K. Hale, and D. Sweet, J. Differential Equations, 4, 40 (1968).

29. M. Baker, Ann. Phys. (N.Y.) 4, 271 (1958).
30. K. G. Wilson, thesis, California Institute of Technology, 1961, unpublished.
31. M. Cini and S. Fubini, Nuovo Cimento 11, 142 (1954); D. Bessis and M. Pusterla, ibid. 54, 243 (1968).
32. G. F. Chew, Phys. Rev. Letters 9, 233 (1962).
33. K. Huang and A. H. Mueller, Phys. Rev. 140, B365 (1965).
34. G. Bart and R. L. Warnock, to be published.
35. Of course it was never claimed that the simple static model of Ref. 32 would give detailed agreement with experiment. Nevertheless, there seems to be a widespread belief that the single-channel N-N^* reciprocal bootstrap is "essentially correct."
36. A. Ts. Amatuni, Nuovo Cimento 58, 321 (1968).

Monographs and Reviews Concerned Wholly or Partly with Nonlinear Analysis

37. T. L. Saaty and J. Bram, Nonlinear Mathematics (McGraw-Hill, New York, 1964).
38. T. L. Saaty, Modern Nonlinear Equations (McGraw-Hill, New York, 1967).
39. M. A. Krasnosel'skii, "Some Problems of Nonlinear Analysis," American Mathematical Society Translations, Series 2, Volume 10 (American Mathematical Society, Providence, R. I., 1958); translation from Uspehi Mat. Nauk (N.S.) 9, 57 (1954).
40. M. A. Krasnosel'skii, Topological Methods in the Theory of Nonlinear Integral Equations (MacMillan Co., New York, 1964).
41. M. A. Krasnosel'skii, Positive Solutions of Operator Equations (P. Noordhoff Ltd., Groningen, 1964).
42. J. Cronin, "Fixed Points and Topological Degree in Nonlinear Analysis," (American Mathematical Society, Providence, R. I., 1964).
43. P. M. Anselone, editor, Nonlinear Integral Equations (University of Wisconsin Press, Madison, 1964).
44. W. Pogorselski, Integral Equations and Their Applications, Volume I (Pergamon Press, Oxford, 1966).
45. L. V. Kantorovich and G. P. Akilov, Functional Analysis in Normed Spaces (Pergamon Press, Oxford, 1964).
46. L. A. Liusternik and V. J. Sobolev, Elements of Functional Analysis (Hindustan Pub. Corp., Delhi, 1961).
47. N. Dunford and J. T. Schwartz, Linear Operators, Vol. 1 (Interscience Publishers, New York, 1958).
48. L. Collatz, Functional Analysis and Numerical Mathematics (Academic Press, New York, 1966).

49. F. F. Bonsall, "Lectures on Some Fixed Point Theorems of Functional Analysis," notes available from Tata Institute of Fundamental Research, Bombay.
50. A. Granas, "Introduction to Topology of Functional Spaces," (Mathematics Lecture Notes, University of Chicago, 1961).
51. M. M. Vainberg, Variational Methods for the Study of Nonlinear Operators (Holden-Day, San Francisco, 1964).
52. M. M. Vainberg and P. G, Aizengendler, "The Theory and Methods of Investigation of Branch Points of Solutions," in Progress in Mathematics, Volume 2, edited by R. V. Gamkrelidze (Plenum Press, New York, 1968).
53. A. I. Guseinov, Izv. Akad. Nauk (Ser. Mat.) 12,193 (1948); for short reviews of more recent work, mostly in journals that are hard to find in the West, see Math. Reviews 36, #3091, #3092, and 21, #277.
54. Melvyn Berger and Marion Berger, Perspectives in Nonlinearity (W. A. Benjamin, Inc., New York, 1968).
55. J. T. Schwartz, Nonlinear Functional Analysis (Gordon and Breach, Inc., New York, to be published).

TWO-DIMENSIONAL ICE AND FERROELECTRIC MODELS†

Elliott H. Lieb
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts

I. Introduction and Review of the Transfer Matrix Formalism

In these lectures we shall present the solution to the two-dimensional ice and ferroelectric problems. Time does not permit an exposition of all the details, however, nor does it permit us to discuss the relationship between these models and the other soluble two-dimensional models which are the Ising model (Onsager, 1944) and the dimer problem (Kasteleyn and Fisher, 1961). At the same time, we will begin the presentation from a reasonably elementary starting point in order that the reader will not need prior knowledge of the "bag of tricks" which are part of the arsenal of statistical mechanics.

The most important tool, or concept, is the transfer matrix. We shall first expound it in abstract form and then give a simple illustration of it in the one-dimensional Ising model in order that its meaning may thereby be clarified.

Consider a system S (such as a coupled spin system) which may be permitted to assume certain states, or configurations, $C_1, C_2, \ldots$. To each of these total configurations we assign an energy $E(C)$ and we desire to calculate the partition, or generating function

$$Z = \sum_i \exp\left[-\beta\, E(C_i)\right] , \tag{I.1}$$

where $\beta = (kT)^{-1}$ with k = Boltzmann's constant and T = temperature.

Suppose, further, that the system can be decomposed into M disjoint subsystems $S_1, S_2, \ldots, S_M$ so that $S = S_1 \otimes S_2 \otimes \ldots$. The subsystem S_i will have allowed configurations $C_1^i, C_2^i, \ldots$ whence a total configuration C_k may be thought of as specifying <u>any</u> one configuration for each subsystem. This may not always be the case in all problems, but we shall suppose it is true for the system of interest. More formally, the set of all configurations $\{C\}$ is

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

$$\{C\} = \left[C_1^1 \oplus C_2^1 \oplus \ldots\right] \otimes \left[C_1^2 \oplus C_2^2 \oplus \ldots\right] \otimes \ldots \tag{I.2}$$

We shall also make the important assumption that the subsystems S_i are all isomorphic to each other so that we may think of C^i_1 and C^k_1, for example, as being identical. Hence we shall drop the superscript from C. A given total configuration C_α is specified by the integers $\alpha 1, \alpha 2 \ldots$ where

$$C_\alpha = C_{\alpha 1} \otimes C_{\alpha 2} \otimes \ldots \tag{I.3}$$

Now we make the most important assumption of all, namely

$$E(C_\alpha) = \sum_{i=1}^{M} U(C_{\alpha i}) + \sum_{i=1}^{M-1} W(C_{\alpha(i+1)}, C_{\alpha i}) . \tag{I.4}$$

In other words, the total energy is the sum of two parts. The first is a self energy, U, associated with each subsystem, and the second is an energy, W, which depends only on the configurations of neighboring subsystems.

The simplest example is the one-dimensional nearest neighbor Ising model. We have N spins on a line, each of which can take on the value ± 1 and the total energy is

$$E = J \sum_{i=1}^{M-1} s_i s_{i+1} + h \sum_{i=1}^{M} s_i \tag{I.5}$$

where h is the external magnetic field (in appropriate units). In this case the subsystem S_i is identical with spin s_i and there are two allowed configurations for each subsystem (up or down). Clearly (I.5) is the same as (I.4) with

$$U(C_{\alpha i}) = h s_i$$

$$W(C_{\alpha(i+1)}, C_{\alpha i}) = J s_{i+1} s_i \tag{I.6}$$

Returning to (I.4) we shall make one simplification which one can show has no effect in the bulk limit $M \to \infty$. Namely, we suppose that S_M interacts with S_1, so that the last sum in (I.4) goes to M with $C_{\alpha(M+1)} \equiv C_{\alpha 1}$. We can then write (I.1) as

$$Z = \left(\sum_{\alpha M=1}^{A} \cdots \sum_{\alpha 2=1}^{A} \sum_{\alpha 1=1}^{A} \right) T(\alpha 1, \alpha M)\, T(\alpha M, \alpha (M-1)) \cdots T(\alpha 3, \alpha 2) T(\alpha 2, \alpha 1) \tag{I.7}$$

where A = total number of configurations of each subsystem and

$$T(\alpha 2, \alpha 1) = \exp -\beta \left\{ \tfrac{1}{2} U(C_{\alpha 2}) + \tfrac{1}{2} U(C_{\alpha 1}) + W(C_{\alpha 2}, C_{\alpha 1}) \right\} \tag{I.8}$$

It is clear that T is an A-square matrix whose indices are $\alpha 2$ and $\alpha 1$. For the one-dimensional Ising model

$$T(s_2, s_1) = \exp -\beta \left\{ \tfrac{1}{2} h(s_2 + s_1) + J s_2 s_1 \right\} \tag{I.9}$$

If we call C_1 the state in which the spin is up and C_2 the state in which it is down, then $s = 3 - 2\alpha$, and

$$T(\alpha 2, \alpha 1) = \exp -\beta \left\{ \tfrac{1}{2} h(6 - 2\alpha 1 - 2\alpha 2) + J(3 - 2\alpha 1)(3 - 2\alpha 2) \right\} \tag{I.10}$$

or

$$T = \begin{array}{c|cc} \alpha 2 \backslash \alpha 1 & 1 & 2 \\ \hline 1 & e^{-\beta(J+h)} & e^{\beta J} \\ 2 & e^{\beta J} & e^{-\beta(J-h)} \end{array} \tag{I.11}$$

Now (I.7) is immediately recognizable as an iterated matrix product, in fact

$$Z = \text{Trace}\,(A^M) = \sum_{j=1}^{A} (\lambda_j)^M \tag{I.12}$$

where the λ_j are the eigenvalues of T. We will pass to the bulk limit by first letting $M \to \infty$ and keeping the size of each subsystem fixed. Then, (I.12) becomes

$$-\beta F = \lim_{M \to \infty} M^{-1} \ln Z = \ln \lambda_{\max} \tag{I.13}$$

where λ_{max} is the eigenvalue of largest modulus, and F is the free energy per subsystem.

(I.13) requires some comment to make it clear why we have dropped the other eigenvalues. For one thing, (I.8) shows that the matrix elements of T are positive real and the Perron-Fröbenius theorem then tells us that the largest eigenvalue is real and positive (it is also non-degenerate if T does not have too many zero elements). The P-F theorem also tells us that the eigenvector belonging to λ_{max} has all its entries real and non-negative. (This theorem will be important later on because if we can find a non-negative eigenvector we will be certain that it is the maximum one. We also note that the P-F theorem is true even if T is not symmetric.) Armed with this theorem we can state (for M even, at least) that

$$\ell n\, \lambda_{max} \leq M^{-1}\, \ell n\, Z \leq M^{-1}\, (\ell n\, A + M\, \ell n\, \lambda_{max}) \qquad (I.14)$$

and hence (I.13) is true. (Remember that A is fixed and $M \to \infty$.) Strictly speaking, inequality (I.14) is justified if T is symmetric so that all the eigenvalues are real. If T is non-symmetric so that the λ's may be complex, and if T also has too many zeros so that the P-F theorem does not guarantee a unique eigenvalue of largest modulus, we could conceivably encounter difficulty in proving (I.13). In almost all known cases, T is indeed symmetric so that there is no problem. The one known case where T is not symmetric is the KDP ferroelectric model, but there T has sufficiently few zeros that the P-F theorem guarantees a unique eigenvalue of greatest modulus.

To apply (I.13) to the one-dimensional chain we must calculate the eigenvalues of (I.11). These are

$$\lambda_{\pm} = e^{-K} \cosh C \pm \left[e^{-2K} \sinh^2 C + e^{2K} \right]^{\frac{1}{2}} \qquad (I.15)$$

where $K = \beta J$ and $C = \beta h$. The reader can easily check that the largest eigenvalue, λ_+, is accompanied by a positive eigenvector. Thus, using (I.13), the free energy per spin for an infinite system is given by

$$F = - kT\, \ell n(\lambda_+) \qquad (I.16)$$

We note in passing that (I.16) is analytic in the temperature except for $T = 0$ ($\beta = \infty$). For this reason it is sometimes said that one dimensional systems can have a phase transition at zero temperature.

As a further illustration of these ideas we shall set up the transfer matrix for the two-dimensional Ising model but we shall not attempt to diagonalize it. (See Ref. 1 for the additional algebraic tricks

necessary to obtain Onsager's solution.) Here we have a square array of spins with M rows and N columns. The subsystem S_i will be the array of N spins on the ith row. The number of configurations of each row is

$$A = 2^N \tag{I.17}$$

and the index α specifying a configuration is most conveniently thought of as an N-tuple of ± 1's according to whether each spin is up or down. If $\alpha = (s_1, s_2, \ldots, s_n)$ is the configuration of the first row and $\beta = (t_1, t_2, \ldots, t_n)$ is the configuration of the second then

$$W(C_\beta, C_\alpha) = J_1 \sum_{i=1}^{N} s_i t_i \tag{I.18}$$

where J_1 is the coupling constant between rows. The self energy, U, now contains more than merely the interaction with the external magnetic field, for

$$U(C_\alpha) = h \sum_{i=1}^{N} s_i + J_2 \sum_{i=1}^{N-1} s_i s_{i+1} \tag{I.19}$$

We can now define the transfer matrix as in (I.8). An eigenvector, $V = \{v_\alpha\}$, with $\alpha = 1, \ldots, A$, can be thought of as a function of $s_1, \ldots, s_N$ and we will simply denote this by $v_\alpha \to v(s_1, \ldots, s_N)$. To find the free energy we must then calculate the largest eigenvalue of the following linear equation:

$$\lambda v(t_1, \ldots, t_N) = \sum_{s_N = \pm 1} \cdots \sum_{s_1 = \pm 1} v(s_1, \ldots, s_N)$$

$$\times \exp -\beta \left\{ \tfrac{1}{2} h \sum_{i=1}^{N} (s_i + t_i) + J_1 \sum_{i=1}^{N} s_i t_i \right.$$

$$\left. + \tfrac{1}{2} J_2 \sum_{i=1}^{N-1} (s_i s_{i+1} + t_i t_{i+1}) \right\} . \tag{I.20}$$

All the eigenvalues of (I.20) can be found when $h = 0$, but, alas, when $h \neq 0$, none of them can be found because the linear terms in the exponential cause severe complications.

II. Derivation of the Transfer Matrix for the Ferroelectric Problems

We shall begin by discussing the ice problem[2] because that forms the skeleton of the ferroelectric models. In the former we merely count the total number of configurations on a square lattice, subject to a certain constraint while in the latter we must add up those same configurations with a weight (Boltzmann factor).

Each vertex of a square lattice is imagined to be an oxygen atom. On each bond we place one hydrogen atom in either of two positions one third of the way from the terminal oxygens. A disposition of the H atoms is a configuration. If there are N oxygens there will be 2N hydrogens, and the number of configurations (with no further constraint) would be:

$$Z = 2^{2N} = 4^N \tag{II.1}$$

The significance of Z is that

$$S = k \ell n Z \tag{II.2}$$

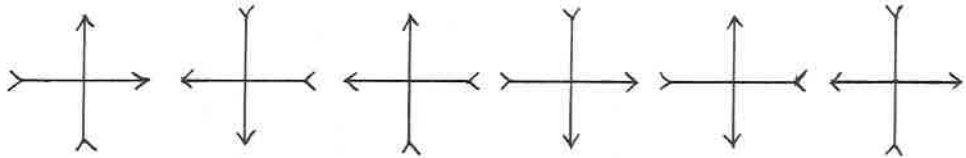
is the entropy of the system. It is called the residual entropy because it exists at T = 0. A residual entropy is actually observed for ice, but it is only N k ℓn (1.5) instead of N k ℓn 4.

To remedy the situation, Pauling proposed the "ice constraint": The only allowed configurations are those in which each oxygen atom is surrounded by two hydrogen atoms close to it and two far from it.

Thus, in the neighborhood of each vertex, only 6 out of a possible 16 configurations are allowed. These are shown in Fig. 1 where an arrow convention has been used to show the position of each hydrogen relative to the bond midpoint. To calculate Z we must calculate the number of ways of placing arrows on a lattice of M rows and N columns so that precisely two arrows point into every vertex.

To set up the transfer matrix we take as subsystem S_i the ith row of vertical arrows and label a particular configuration, α, by an N-tuple of ± 1's according to which arrows are up or down. The rows of horizontal arrows will not be considered explicitly as subsystems for we shall at once sum over all possible arrangements of horizontal arrows in order to provide an effective interaction between adjacent rows of vertical arrows. Thus, let α be the configuration of the first and β be the configuration of the second row of vertical arrows. Define the transfer matrix by

$$T(\beta,\alpha) = \sum \begin{pmatrix} \text{allowed configurations of horizontal} \\ \text{arrows consistent with } \beta \text{ and } \alpha \end{pmatrix} \tag{II.3}$$



	(1)	(2)	(3)	(4)	(5)	(6)
			Vertex Energies			
Ice Model	0	0	0	0	0	0
F Model	ϵ	ϵ	ϵ	ϵ	0	0
KDP Model	0	0	ϵ	ϵ	ϵ	ϵ
External Electric Field	+E	-E	+E	-E	0	0

Fig. 1. The six vertex configurations allowed by the "ice condition" are shown, together with their associated energies appropriate to the three models. Note that $\epsilon > 0$. The last row shows the energy to be added when an external electric field in the vertial direction is present. The arrows indicate the position of a hydrogen atom relative to the bond midpoint.

In (II.3) we mean all configurations such that the ice condition is satisfied at every one of the N intervening vertices, and it is certainly possible that (II.3) is zero for some α and β. In any case, we have

$$Z = \Big(\sum_{\alpha M} \cdots \sum_{\alpha 1}\Big) T(\alpha 1,\alpha M)\, T(\alpha M,\alpha(M-1)) \cdots T(\alpha 2,\alpha 1)$$

$$= \text{Trace}\,(T^M) \to (\lambda_{\max})^M \qquad \text{(II.4)}$$

Next we must evaluate (II.3) explicitly. First observe that

$$T(\alpha,\alpha) = 2, \quad \text{all } \alpha, \qquad \text{(II.5)}$$

because when $\beta = \alpha$ the only choices are either all horizontal arrows to the left or all to the right. If $\beta \neq \alpha$ then at the first position (reading from the left) in which β differs from α we either have $\alpha = +1$, $\beta = -1$ or $\alpha = -1$, $\beta = +1$. In the former case we must have vertex 6 and in the latter vertex 5 of Fig. 1.

In the former case (a +- exchange) the horizontal arrows run outward until they meet another exchange which must necessarily be of the opposite type. In other words, vertices 5 and 6 must alternate on the row. Thus

$$T(\beta,\alpha) = 1 \text{ if the exchanges alternate} \quad (\beta \neq \alpha)$$

$$= 0 \text{ otherwise} \qquad \text{(II.6)}$$

Furthermore, we shall suppose that the last column is connected by a horizontal bond to the first. This means that there must be the same number of both kinds of exchanges if T is to be nonzero. For example, these sequences are allowed: (+-)(-+); (+-)(-+)(+-)(-+); (-+)(+-). These are not allowed: (+-); (+-)(+-); (+-)(-+)(+-).

A trivial, but important corollary of the above is this: Let $n(\alpha)$ be the number of down arrows in α. Then

$$T(\beta,\alpha) = 0 \text{ unless } n(\beta) = n(\alpha)\ . \qquad \text{(II.7)}$$

Consequently, T consists only of diagonal blocks characterized by the index n which can be $0,1,\ldots,N$. It is easy to show that the largest eigenvalue in each n subspace is non-degenerate but we must decide which subspace gives the largest $\lambda_{\max}$. It is possible to prove[2] that as $N \to \infty$ it is the "obvious" subspace $n = \frac{1}{2}N$.

The eigenvalue equation reads:

$$\lambda\, v(\beta) = \sum_{\alpha} T(\beta,\alpha)\; v(\alpha) \quad . \tag{II.8}$$

In any given n-space, a configuration α will have down arrows at certain positions $x_1, x_2, \ldots, x_n$ and we may assume that $1 \le x_1 < x_2 < x_3 < \ldots < x_n \le N$. Thus, $v(\alpha)$ can be thought of as a function of the x's and $v(\alpha) \to f(x_1, x_2, \ldots, x_n)$. We must translate (II.8), by using (II.5) and (II.6), into some convenient equation for f. It turns out to be

$$\lambda f(x_1,\ldots,x_n) = \sum_{y_1=1}^{x_1} \sum_{y_2=x_1}^{x_2} \cdots \sum_{y_n=x_{n-1}}^{x_n} f(y_1,\ldots,y_n)$$

$$+ \sum_{y_1=x_1}^{x_2} \sum_{y_2=x_2}^{x_3} \cdots \sum_{y_n=x_n}^{N} f(y_1,\ldots,y_n)\ , \tag{II.9}$$

with the following addenda:

a. (II.9) is true only for $1 \le x_1 < \ldots < x_n \le N$

b. On the right side of (II.9) there appear terms in which some $y_{i+1} = y_i$. For example $y_2 = y_1 = x_1$ in the first sum. Such terms are to be regarded as non-existent (alternatively, on the right side define $f = 0$ whenever $y_{i+1} = y_i$).

To see that (II.9) is the correct translation of (II.8) observe the following. When each $y_i = x_i$ we have the situation in which $\beta = \alpha$, for the two rows of arrows are identical. (II.9) gives a factor 2 (one from the first sum and one from the second) and this agrees with (II.5). On the other hand, if at least one $y_i \neq x_i$ then the first sum gives all exchanges of the type (-+)(+-) etc., and the second sum gives exchange of the type (+-)(-+) etc. In each case we have a factor of unity in agreement with (II.6). With a little quiet reflection the reader can convince himself of this fact.

In the next section we shall solve (II.9), especially for $n = \frac{1}{2}N$, which is the case of most interest. Here we shall turn to the ferroelectric models[3] and the appropriate analogs of (II.9). To each vertex type (Fig. 1) we associate an energy and thence, by simple addition, we associate an energy to an entire configuration. The allowed configurations are the same as before. The partition function is

$$Z = \sum_{C} \begin{pmatrix} \text{allowed} \\ \text{configurations} \end{pmatrix} \quad \exp\left[-\beta\, E\,(C)\right] \tag{II.10}$$

In the F model (Rys 1963) we assign a positive energy, ϵ, (Fig. 1) to the first 4 vertices, namely to those which have a net electric polarization. This is a model of an antiferroelectric because at low temperatures the most probable configurations will consist predominantly of alternating vertices 5 and 6. The system will therefore resist polarization by an external electric field. Furthermore, just as with the Ising model, we expect to find a critical temperature heralding the onset of long range order.

For the ferroelectric KDP model (Slater 1941) we choose <u>one</u> of the 45° axes as being favored and encourage the vertices which have a polarization in that direction. In Fig. 1 the northeast-southwest axis is shown preferred (vertices 1 and 2). Again, we expect a phase transition to an ordered state.

Finally, for either model, we can add an electric field, E, in the vertical direction. The additional energy is shown (in appropriate units) on the last line of Fig. 1. It is easy to find the effect of E on the eigenvalues. Let $\lambda(n)$ be the largest eigenvalue in each n subspace. In this subspace, the external field energy is the same for all configurations and is given by $E(N - 2n)$ for each row or $E\,M(N - 2n)$ for the whole lattice. Thus $\lambda(E, n)$, the largest eigenvalue with a field, is given by

$$\lambda(E, n) = \lambda(n) \exp\left[-\beta\, E(N - 2n)\right] \quad . \tag{II.11}$$

Although $n = \frac{1}{2}N$ maximizes $\lambda(n)$, we must choose n to maximize the entire right side of (II.11). Thus, as $N, M \to \infty$

$$(N\,M)^{-1} \ell n\, Z = N^{-1} \max_{n} \left\{\ell n\, \lambda(n) - \beta E(N - 2n)\right\} \quad . \tag{II.12}$$

Having disposed of E we can ask for the analogs of (II.3) and (II.9) for the two models. The transfer matrix will be

$$T(\beta, \alpha) = \sum \begin{pmatrix} \text{allowed configurations of horizontal} \\ \text{arrows consistent with } \beta \text{ and } \alpha \end{pmatrix}$$

$$\times \exp\left[-\beta\, E\,(\text{row})\right] \quad . \tag{II.13}$$

Here, E(row) means the total energy (according to the second or third line of Fig. 1) of the row of vertices formed by each assignment of

horizontal arrows on the intervening row. Thus (II.9) will generalize to

$$\lambda f(x_1,\ldots,x_n) = \sum_{y=1}^{x_1} \cdots \sum_{y_n=x_{n-1}}^{x_n} f(y_1,\ldots,y_n) D_1(X,Y)$$

$$+ \sum_{y_1=x_1}^{x_2} \cdots \sum_{y_n=x_n}^{N} f(y_1,\ldots,y_n) D_2(X,Y) \qquad \text{(II.14)}$$

where $X = (x_1, x_2, \ldots, x_n)$ and correspondingly for Y. For the F model we want D_1 and D_2 to have a factor $\exp(-\beta\epsilon)$ for each vertex where a (+-) or a (-+) exchange does not take place. Thus,

$$\text{F model:}\quad D_1 = D_2 = \exp -\beta\epsilon\Big[N-2n+2 \sum_{i=1}^{n} \sum_{j=1}^{n} \delta(x_i - y_j)\Big] \qquad \text{(II.15)}$$

For the KDP model, it is a bit more complicated; with a little thought we derive:

$$\text{KDP model:}\quad D_1 = e^{-\beta\epsilon N}\, U(y_2 - x_1)\, U(y_3 - x_2) \ldots U(y_1 + N - x_n)$$

$$D_2 = e^{-\beta\epsilon N}\, U(y_1 - x_1)\, U(y_2 - x_2) \ldots U(y_n - x_n) \qquad \text{(II.16)}$$

with

$$U(z) = \exp[\beta\epsilon(z-1)], \quad z > 0$$

$$= \exp[\beta\epsilon], \quad z = 0\,. \qquad \text{(II.17)}$$

It is to be remarked that the transfer matrix is no longer symmetric for the KDP model, but no complication is caused thereby for, although T has some zeros in each n subspace, it is easily seen that T^B for some sufficiently large B, has no zeros in every diagonal block. The P-F theorem then guarantees a unique largest eigenvalue.

III. Diagonalization of the Transfer Matrix and Reduction to an Integral Equation

At first let us concentrate on solving the ice problem, Eq. (II.9). The generalization to the more complicated (II.14) is not really hard once the main idea is understood.

There does not seem to be any simple artifice, such as a unitary transformation, to solve (II.9). Instead, we shall make an ansatz for f and then

a. verify that it satisfies (II.9)

b. verify that our f is indeed the maximum eigenvector by showing that it is a positive function and using the P-F theorem.

Our ansatz is:

$$f(x_1,\ldots,x_n) = \sum_P ! \, A(P) \exp\left\{ i \sum_{j=1}^{n} k_{P(j)} x_j \right\} , \qquad \text{(III.1)}$$

where:

1. (III.1) defines f only in the region R: $1 \le x_1 < x_2 < \ldots < x_n \le N$.

2. The set $\{k\} = k_1, k_2, \ldots, k_n$ is some set of distinct wave numbers. They will turn out to be real.

3. The sum in (III.1) is over all n! permutations of $1,\ldots,n$. A typical permutation is denoted by $P = \begin{pmatrix} 1 & 2 & \cdots & n \\ P(1) & P(2) & \cdots & P(n) \end{pmatrix}$.

4. (III.1) thus contains a sum of all the different plane waves we can form by permuting the set $\{k\}$. These are added up with some weight factor called A(P). The problem is to determine the n! coefficients A(P) and the n numbers $\{k\}$ so that (II.9) is satisfied in R.

Consider what happens when n = 1. In this case

$$f(x) = e^{ikx} \qquad \text{(III.2)}$$

Denote by S_1 and S_2 the first and second sums, respectively, in (II.9). Obviously,

$$S_1 = d(k)\left[e^{ik} - e^{ik(x+1)} \right]$$

$$S_2 = d(k)\left[e^{ikx} - e^{ik(N+1)} \right] , \qquad \text{(III.3)}$$

where

$$d(k) = (1 - e^{ik})^{-1} \qquad \text{(III.4)}$$

Now, (III.3) contains one term from each of S_1 and S_2 that is proportional to f(x) and together they are:

$$d(k)(1 - e^{ik}) f(x) = f(x) \qquad \text{(III.5)}$$

It also contains 2 unwanted terms which must be made to cancel. That is, we must choose k so that

$$d(k)\,(e^{ik} - e^{ik(N+1)}) = 0 \qquad \text{(III.6)}$$

Consequently, (II.9) is satisfied with all $\lambda = 1$, and the eigenvectors, (III.2), are determined by

$$k = \frac{2\pi}{N}\,x \text{ (integer)} . \qquad \text{(III.7)}$$

The astute reader will notice that there also is a solution for $k = 0$, whence $S_1 = x$, $S_2 = N - x + 1$ and $\lambda = N + 1$. It turns out that this situation, namely that one member of the set $\{k\}$ is zero, occurs only for n odd. To avoid this complication we shall henceforth only consider the case n = even, but that is really no restriction in the limit $N \to \infty$, $n \to \infty$.

The foregoing example with n = 1 reveals much, but the true nature of the problem is revealed for n = 2. Denote $\{k\}$ simply by k, p and x_1, x_2 simply by x and y. Then,

$$f(x,y) = a\,e^{i(kx+py)} + b\,e^{i(px+ky)} \qquad \text{(III.8)}$$

in the region $1 \le x < y \le N$. To compute S_1 we must be very careful not to count the one point where the two variables are equal. One easily finds:

$$\begin{aligned}
S_1 &= a\Big\{d(k)d(p)\Big[e^{ik} - e^{ik(x+1)}\Big]\Big[e^{ipx} - e^{ip(y+1)}\Big] - e^{i(k+p)x}\Big\} \\
&\quad + b\Big\{d(k)d(p)\Big[e^{ip} - e^{ip(x+1)}\Big]\Big[e^{ikx} - e^{ik(y+1)}\Big] - e^{i(k+p)x}\Big\} \\
S_2 &= a\Big\{d(k)d(p)\Big[e^{ikx} - e^{ik(y+1)}\Big]\Big[e^{ipy} - e^{ip(N+1)}\Big] - e^{i(k+p)y}\Big\} \\
&\quad + b\Big\{d(k)d(p)\Big[e^{ipx} - e^{ip(y+1)}\Big]\Big[e^{iky} - e^{ik(N+1)}\Big] - e^{i(k+p)y}\Big\} .
\end{aligned} \qquad \text{(III.9)}$$

Although (III.9) is quite complicated we first observe that S_1 and S_2, when multiplied out, each contain a pair of terms that, together, are proportional to (III.8), namely $\lambda f(x,y)$ where

$$\lambda = d(k)d(p)\Big[e^{i(k+p)} + 1\Big] . \qquad \text{(III.10)}$$

Thus, λ is our eigenvalue provided we can get the vast number of unwanted terms to disappear.

Next, we observe that S_1 contains a term proportional to exp(ik) (first line) and a term proportional to exp(ip) (second line). Likewise, S_2 contains exp(ip(N+1)) and exp(ik(N+1)). These are factors which depend on where we started and ended the chain. All other factors in

(III.9) would remain the same if the chain ran from 6 to N + 5, for example, instead of from 1 to N. The strategy we shall use is to make the chain independent factors cancel in S_1 and S_2 separately while the chain dependent factors will be made to cancel only when S_1 and S_2 are added together.

The chain independent factors in S_1 (other than those contributing to (III.10)) are all seen to be proportional to $\exp(i(k+p)x)$. To make them vanish we require that

$$a\left\{-d(k)d(p)e^{ik} - 1\right\} + b\left\{-d(k)d(p)e^{ip} - 1\right\} = 0 \ . \quad \text{(III.11)}$$

Likewise, in S_2 the unwanted chain independent factors are proportional to $\exp(i(k+p)y)$ and the condition for their vanishing is the same as (III.11). This is important, otherwise we should have two different equations relating a to b.

Thus, the chain independent factors vanish if, for any choice of k and p, the coefficients a and b are chosen such that

$$a = b\, B(k,p) \quad \text{(III.12)}$$

where

$$B(k,p) = -\frac{1 + e^{i(k+p)} - e^{ik}}{1 + e^{i(k+p)} - e^{ip}} \ . \quad \text{(III.13)}$$

Since the ratio of a to b is now fixed by (III.11) the only variables we can adjust to make the chain dependent terms in $S_1 + S_2$ vanish are k and p. There are two sets of terms: those proportional to $\exp(ik)$ in S_1 or $\exp(ik(N+1))$ in S_2 on the one hand, and those proportional to $\exp(ip)$ or $\exp(ip(N+1))$ on the other hand. To make the first set vanish we require

$$e^{ikN} = \frac{a}{b} = B(k,p) \quad \text{(III.14a)}$$

and to make the second set vanish

$$e^{ipN} = B(k,p)^{-1} = B(p,k) \ . \quad \text{(III.14b)}$$

The conclusion is this: for every pair of numbers k and p satisfying both (III.14a) and (III.14b) we have an eigenvalue given by (III.10). Note that k and p must be distinct, otherwise B = -1 and the vector (III.8) vanishes.

Clearly, if we insert the ansatz (III.1) for arbitrary n into (II.9) a very similar situation will result. We shall merely describe what

happens and assure the reader that a tedious inductive proof[2] verifies the following:

1. S_1 and S_2 each contain one term proportional to (III.1). When added together they give the eigenvalue

$$\lambda = \left\{ \prod_{i=1}^{n} d(k_i) \right\} \left\{ 1 + \exp\left[i \sum_{j=1}^{n} k_j \right] \right\} . \qquad \text{(III.15)}$$

(Cf. (III.10)).

2. The condition for the vanishing of the chain independent, unwanted terms in S_1 is the same as for S_2. It is the following: Let P and Q be any two permutations that are identical except for the ith and (i+1)th positions. Denote $k_{P(i)} = k_{Q(i+1)} = p$ and $k_{P(i+1)} = k_{Q(i)} = q$. Then

$$A(P) = A(Q)\, B(p,q) . \qquad \text{(III.16)}$$

Eq. (III.1) must be true for every pair P and Q in order that such terms vanish and it may be asked whether all these conditions can be simultaneously fulfilled. The answer is yes, for define θ by

$$B(p,q) = - \exp\left[-i\, \theta\, (p,q) \right] \qquad \text{(III.17)}$$

and then define

$$A(P) = (-)^P \exp\left[-\tfrac{1}{2} i \sum_{i<j} \theta\, (k_{P(i)}, k_{P(j)}) \right] \qquad \text{(III.18)}$$

with $(-)^P$ = signature of P.

Since, by definition, $\theta(p,q) = -\theta(q,p)$, it is easily verified that (III.18) satisfies (III.16) for all P and Q.

3. In order to eliminate the chain dependent terms in $S_1 + S_2$ one requires that if

$$P = (P(1), P(2), \ldots, P(n)) \text{ and } Q = (P(2), P(3), \ldots, P(n), P(1))$$

then

$$A(P) = A(Q) \exp(i k_{P(1)} N) . \qquad \text{(III.19)}$$

(III.19) is no surprise for it is an obvious statement of the periodicity of the problem. In other words, (III.19) is easily seen to be equivalent to

$$f(0,x_2,\ldots,x_n) = f(x_2,\ldots,x_n,N) \tag{III.20}$$

with f defined by (III.1). We stress, however, that (III.19) was not guessed at, but came from elimination of unwanted terms. Using (III.16), (III.19) becomes

$$e^{ik_jN} = \prod_{i\neq j} B(k_j,k_i) \tag{III.21}$$

for $j = 1,2,\ldots,n$. These n coupled equations determine the set $\{k\}$ and they in turn determine the eigenvalue.

Returning to the ferroelectric models, one finds[3] that the same ansatz (III.1) still works except for these modifications:

1. (III.16) and (III.21) are still true except that B becomes

$$B(k,p) = -\frac{1 + e^{i(k+p)} - 2\Delta e^{ik}}{1 + e^{i(k+p)} - 2\Delta e^{ip}} \tag{III.22}$$

where

$$\begin{aligned} \Delta &= \tfrac{1}{2} && \text{for ice} \\ &= 1 - \tfrac{1}{2}e^{2K} && \text{for F} \\ &= \tfrac{1}{2}\, e^{K} && \text{for KDP} \end{aligned} \tag{III.23}$$

with $K = \beta\epsilon$.

2. There is also a modification of (III.15):

$$\text{F model: } \lambda = e^{-NK}\prod_{j=1}^{n}\frac{(2\Delta - 1) - e^{ikj}}{1 - e^{ik_j}} + e^{-NK}\prod_{j=1}^{N}\frac{(2\Delta - 1) - e^{-ik_j}}{1 - e^{-ik_j}} \tag{III.24}$$

$$\text{KDP model: } \lambda = \left\{\prod_{j=1}^{n}\left[1 - e^{ik_j+K}\right]^{-1}\right\}\left\{\exp\left(i\sum_{1}^{n}k_j\right) + e^{-KN}\prod_{j=1}^{n}\left[e^{ik_j} - e^{ik_j+2K} + e^{K}\right]\right\} \tag{III.25}$$

Now we are faced with the question of whether (III.21) has a solution and, if so, whether one of them gives the maximum eigenvector. It is very fortunate that one ends up with exactly the same (III.21) in the solution to the anisotropic Heisenberg chain problem. There, one defines an f as in (III.1). The equation for f corresponding to (II.9) is different, and therefore the eigenvalue (III.10), (III.24) or (III.25) is different, but f, and hence (III.21), is the same in the two cases. Also, the ground state f for the H-chain is positive (because of the P-F theorem) and it can be proved[4] that (III.21) has a solution and that one of these gives the ground state of the H-chain. Therefore, that particular solution of (III.21) is the one we are looking for. (Strictly speaking (III.21) has a solution only for $n \leq \frac{1}{2}N$, but since the problem is symmetric under the replacement of n by N - n that is no problem.)

The solution we want is such that no k is zero for n even, and the k's are real for $\Delta < 1$. (The case $\Delta > 1$ arises only for the KDP model at low temperature and we shall consider that case separately later on.) For each k in $\{k\}$ there is also -k and hence $\Sigma_1^n k_j = 0$. Since the variables x_i in (III.1) are integers we may restrict the k's to lie in $(-\pi,\pi)$. As $N \to \infty$ with n/N fixed the n k's will densely fill up some interval (-Q,Q), with $Q \leq \pi$. We will order $\{k\}$ so that $k_1 < k_2 < \ldots < k_n$.

Equating exponents in (III.21) and using (III.17), one has

$$Nk_j = -\tfrac{1}{2}\pi + 2\pi I_j - \sum_{i=1}^{n} \theta(k_j,k_i) \qquad \text{(III.26)}$$

where $\theta(k,k) \equiv 0$ and the I_j are integers. It turns out that (for n even)

$$I_j = -\tfrac{1}{2}n + j \qquad \text{(III.27)}$$

Since $\{k\}$ fills (-Q,Q) densely for large N, define

$$N\,\rho(k)\,dk = \text{number of k's in } (k,k+dk) \qquad \text{(III.28)}$$

whence

$$\int_{-Q}^{Q} \rho(k)dk = \frac{n}{N} = \tfrac{1}{2}(1 - y) \qquad \text{(III.29)}$$

with $y = 1 - 2\,n/N$. Note that $y = 0$ corresponds to the $n = \frac{1}{2}N$ case and that $0 \leq y \leq 1$.

Next let $k_{j+1} - k_j = dk$, take the difference of two successive equations (III.26) with j+1 and j, convert the sum into an integral using (III.28), and obtain

$$1 = 2\pi\, \rho(k) - \int_{-Q}^{Q} \frac{\partial}{\partial k}\, \theta(k,q)\, \rho(q)\, dq \quad . \tag{III.30}$$

(III.30), which holds for k only in (-Q,Q), defines the density function ρ. (III.29) then gives the value of y. The solution to (III.30) depends, of course, on Q, so that if y is regarded as given, we must solve (III.30) for all Q and pick Q so that the integral in (III.29) comes to the desired value. It is also to be noted that ρ depends on the temperature since Δ in (III.23) does and θ is defined in terms of B, (III.17), which in turn depends on Δ as a parameter (III.22). But notice that θ depends only on Δ and not on the particular model under consideration. Furthermore, the previous discussion tells us that ρ must be an even, positive function of k.

In the next section we shall solve (III.30). We conclude here by expressing λ in terms of ρ using (III.15), (III.24) or (III.25). The two terms in (III.15) or (III.24) are equal using the aforementioned properties of $\{k\}$, and a factor of 2 makes no difference in the bulk limit. For the KDP model, the first term in the second parenthesis in (III.25) is unity and the second term goes to zero as $N \to \infty$ when $n < \frac{1}{2}N$. Therefore it may be dropped. Thus,

$$\lim_{N\to\infty} N^{-1} \ell n\, \lambda$$

$$= -\tfrac{1}{2} \int_{-Q}^{Q} \rho(q)\, \ell n(2 - 2\cos q)dq \quad , \qquad \text{for ice} \tag{III.31a}$$

$$= -K + \tfrac{1}{2} \int_{-Q}^{Q} \rho(q) \ell n \left[\frac{(2\Delta-1)^2 - (4\Delta-2)\cos q + 1}{2 - 2\cos q} \right] dq,$$

$$\text{for the F model} \tag{III.31b}$$

$$= -\tfrac{1}{2} \int_{-Q}^{Q} \rho(q)\, \ell n \left[1 + e^{2K} - 2e^{K} \cos q \right] dq \ ,$$

$$\text{for the KDP model.} \tag{III.31c}$$

IV. Solution of the Integral Equation and Physical Results

We are faced with the problem of solving (III.30). Unfortunately, the kernel is not a function of k - q, otherwise we could think in terms of Fourier transforms. In point of fact, from (III.17) and (III.22)

$$\tan \tfrac{1}{2}\theta(p,q) = \frac{\Delta \sin \tfrac{1}{2}(p-q)}{\cos \tfrac{1}{2}(p+q) - \Delta \cos \tfrac{1}{2}(p-q)} \tag{IV.1}$$

Nevertheless, (III.30) can be solved in some cases because over many years much attention has been given to that equation. The most up-to-date summary is given in Ref. 4.

The first trick is to change variables from p to α as follows:

<u>case 1</u>. If $\Delta < -1$, define $\Delta = -\cosh\lambda$ and

$$e^{ip} = \frac{e^{\lambda} - e^{-i\alpha}}{e^{\lambda - i\alpha} - 1} \tag{IV.2}$$

<u>case 2</u>. If $-1 < \Delta < 1$, define $\Delta = -\cos\mu$ and

$$e^{ip} = \frac{e^{i\mu} - e^{\alpha}}{e^{i\mu + \alpha} - 1} \tag{IV.3}$$

In both cases α is real because in the second case one can show that $Q < \pi - \mu$. In the new variables $\rho(p) \to R(\alpha)$ with

$$R(\alpha)\, d\alpha \equiv 2\pi\, \rho(p)\, dp\ . \tag{IV.4}$$

Also, when $p = Q$ then $\alpha = b$.

(III.30) and (III.29), respectively, become

$$R(\alpha) = \xi(\alpha) - \int_{-b}^{b} K(\alpha - \beta)\, R(\beta)\, d\beta \tag{IV.5}$$

$$\pi(1 - y) = \int_{-b}^{b} R(\alpha)\, d\alpha \tag{IV.6}$$

and, almost miraculously, we have achieved the goal of transforming the kernel into a difference kernel. In (IV.5)

$$\begin{aligned} \xi(\alpha) &= \sinh\lambda\, [\cosh\lambda - \cos\alpha]^{-1}\ , \quad \Delta < -1 \\ &= \sin\mu\, [\cosh\alpha - \cos\mu]^{-1}\ , \quad -1 < \Delta < 1 \end{aligned} \tag{IV.7}$$

$$\begin{aligned} K(\alpha) &= \frac{1}{2\pi}\ \frac{\sinh 2\lambda}{\cosh 2\lambda - \cos\alpha}\ , \quad \Delta < -1 \\ &= \frac{1}{2\pi}\ \frac{\sin 2\mu}{\cosh\alpha - \cos 2\mu}\ , \quad -1 < \Delta < 1\ . \end{aligned} \tag{IV.8}$$

In both cases one can show:

a. There is a unique and positive solution to (IV.5).

b. The limit b is a monotonically decreasing function of y with $b = 0$ at $y = 1$. When $y = 0$ then

$$b = \pi \quad , \quad \Delta < -1$$

$$= \infty \quad , \quad -1 < \Delta < 1 \quad . \tag{IV.9}$$

Therefore, for the $y = 0$ case ($n = \frac{1}{2}N$) we can solve (IV.5) by a Fourier series in the first case (because K is periodic) and by a Fourier transform in the second case.

It is easy to find the appropriate Fourier transforms of ξ and K and one finds:

<u>case 1</u>. $\Delta < -1$ and $y = 0$

$$R(\alpha) = \tfrac{1}{2} \sum_{n=-\infty}^{\infty} e^{in\alpha} \operatorname{sech} n\lambda \quad . \tag{IV.10}$$

It is possible to evaluate (IV.10) in terms of elliptic functions but there is no particular point in doing so. In case 2 it is easy to invert the Fourier transform of R to obtain

<u>case 2</u>. $-1 < \Delta < 1$ and $y = 0$

$$R(\alpha) = \frac{\pi}{2\mu} \operatorname{sech} \left(\frac{\pi\alpha}{2\mu}\right) \tag{IV.11}$$

The reader can easily verify (IV.11) as well as the fact that (IV.6) is satisfied with $y = 0$.

Armed with (IV.10) and (IV.11) we can evaluate (III.31) for $y = 0$. Henceforth we shall use the notation

$$\lim_{N \to \infty} N^{-1} \ell n\, \lambda(n) \equiv z(y) \quad . \tag{IV.12}$$

For ice we are interested in $z(0)$ with $\Delta = \frac{1}{2}$, which means case 2 with $\mu = 2\pi/3$. (III.31a) becomes (after changing variables)

$$z(0) = -\frac{3}{16\pi} \int_{-\infty}^{\infty} \frac{d\alpha}{\cosh\,(3\alpha/4)} \ell n \left[1 - \frac{3}{1+2\cosh\alpha}\right]$$

$$= \frac{3}{2} \ell n \frac{4}{3} \qquad \text{for ice} \tag{IV.13}$$

(The details of the evaluation of this integral are given in Ref. 2.)

Returning to Section II we see that the residual entropy of "square ice" is

$$S = N\, k\, \ell n \left(\frac{4}{3}\right)^{\frac{3}{2}} \qquad \text{(IV.14)}$$

where N is the number of ice molecules. Since $\left(\frac{4}{3}\right)^{\frac{3}{2}} = 1.5396\ldots$ we are very close indeed to the experimental value.

We can also insert (IV.10) into (III.31) and perform the integration on α to obtain for $\Delta < -1$ (applicable to the F model only)

$$z(0) = \tfrac{1}{2}\lambda - K + \sum_{n=1}^{\infty} \frac{1}{n}\, e^{-n\lambda} \tanh n\lambda \qquad \text{(IV.15)}$$

Alternatively, we can insert (IV.11) into (III.31) to obtain for $-1 < \Delta < 1$

$$z(0) = -K + \frac{1}{8\mu} \int_{-\infty}^{\infty} \frac{d\alpha}{\cosh(\pi\alpha/2\mu)}\; \ell n \left[\frac{\cosh\alpha - \cos 2\mu}{\cosh\alpha - 1}\right]$$

for the F model (IV.16a)

$$z(0) = \frac{1}{8\mu} \int_{-\infty}^{\infty} \frac{d\alpha}{\cosh(\pi\alpha/2\mu)}\; \ell n \left[\frac{\cosh\alpha - \cos\mu}{\cosh\alpha - \cos 3\mu}\right]$$

for the KDP model. (IV.16b)

Unfortunately neither (IV.15) nor (IV.16) can be evaluated in closed form for arbitrary values of λ and μ (i.e. the temperature). Nevertheless we can discuss the analytic properties of the solution.*

In the general case, with an electric field present (which for convenience we shall take to be negative), the free energy per vertex is given by (I.13) and (II.12) as:

$$-\beta F = \max_{y} \left(z(y) + \beta E y\right) \quad . \qquad \text{(IV.17)}$$

Let us first consider what happens when $E = 0$. One can prove in general that $z(y)$ is a convex, symmetric function of y with a maximum at $y = 0$. This is a very general result that is required for thermodynamic stability. Hence, when $E = 0$, we need to consider $z(0)$ in (IV.17).

F model: When $T = \infty$, $\Delta = \frac{1}{2}$ and, as T decreases, $z(0)$ is given by (IV.17a) until $\Delta = -1$. This is the critical temperature and is given by $e^K = 2$ or

---

* Note added in proof: M. L. Glasser, in a forthcoming paper, has found a closed form evaluation of (IV.16b).

$$\epsilon/\ell n\ 2 = k\ T_c \qquad \text{(IV.18)}$$

Below T_c the free energy is given by (IV.15) which is quite a different expression. At first sight one is inclined to think that some derivative of F with respect to T will be discontinuous, or possibly infinite, at T_c. Surprisingly, such is not the case, for every derivative of F is bounded and continuous at T_c. However, a power series expansion about T_c has a zero radius of convergence. Since $T = T_c$ corresponds to $\mu = 0$ we can expand (IV.16a) in μ and find formally

$$z(0) = -K + 2\ \ell n(\Gamma(\tfrac{1}{4})/\Gamma(\tfrac{3}{4})) + \tfrac{1}{2}\sum_{n=1}^{\infty}[n(2n)!]^{-1}\ B_{2n}\mu^{2n}\left[(-1)^n - |E_{2n}|\right] \qquad \text{(IV.19)}$$

where B and E are Bernoulli and Euler numbers. (IV.19) converges for no value of μ. Thus T_c is a true singularity, but it is an essential singularity and we have here an example of an infinite order phase transition. Worse than that, there is also a cut which is a natural boundary (i.e. no analytic continuation is possible) in the complex T plane.

KDP model: Again, when $T = \infty$, $\Delta = \frac{1}{2}$ and, as T decreases, z(0) is given by (IV.16b) until $\Delta = +1$. The critical temperature at which this happens is the same as for the F model, (IV.18). Why is this the critical temperature? Because (IV.16b) tells us that z(0) is analytic in μ up to $\mu = \pi$, corresponding to $\Delta = 1$. Also, at $\mu = \pi$, $z(0) = 0$. On On the other hand, the partition function is > 2 since there are two ordered states with zero energy and this implies that $z(0) \geqslant 0$. Furthermore, $-\beta F$, and hence z(0), decreases with decreasing temperature since all energies are non-negative. Hence, we conclude

$$z(0) = 0 \quad , \quad \text{for } T < T_c. \qquad \text{(IV.20)}$$

In fact, we can extend (IV.20) by noting that $z(0) \geq z(y)$ for $y \neq 0$ but $z(y) \geq 0$ for the same reason as above. Thus

$$z(y) = 0 \quad , \quad \text{for } T < T_c. \qquad \text{(IV.21)}$$

In this way we have circumvented the problem that (III.21) has no solution with real $\{k\}$ for $\Delta > 1$. In actual fact, what happens is that the ansatz (III.1) and the final equation (III.21) are still correct in this case, but the k's become pure imaginary. It is possible to write an integral equation analogous to (III.30) along the imaginary k axis, but the above argument obviates the need for that.

Since z(0) consists of two separate analytic pieces above and below T_c, there is clearly a phase transition. From (IV.20), the internal energy and the specific heat are identically zero below T_c. In other words, the system is dead. In fact it is frozen into a completely ordered state because (IV.21) and (IV.17) tell us that an infinitesimal electric field will produce perfect polarization, y = 1.

Above T_c, (IV.16b) tells us that the internal energy and the specific heat are not zero. It is easy to differentiate (IV.16b) to find that, just above T_c,

$$U = \tfrac{1}{2}\epsilon \qquad \text{and}$$

$$C \sim \frac{k}{4\pi}\left(\frac{\epsilon}{k\,T_c}\right)^{\frac{3}{2}}\left(1 - \frac{T_c}{T}\right)^{-\frac{1}{2}} \tag{IV.22}$$

where U and C are the internal energy and specific heat, respectively, per vertex. Thus, in zero electric field we have a first order phase transition (latent heat), accompanied by an infinite specific heat on the high temperature side.

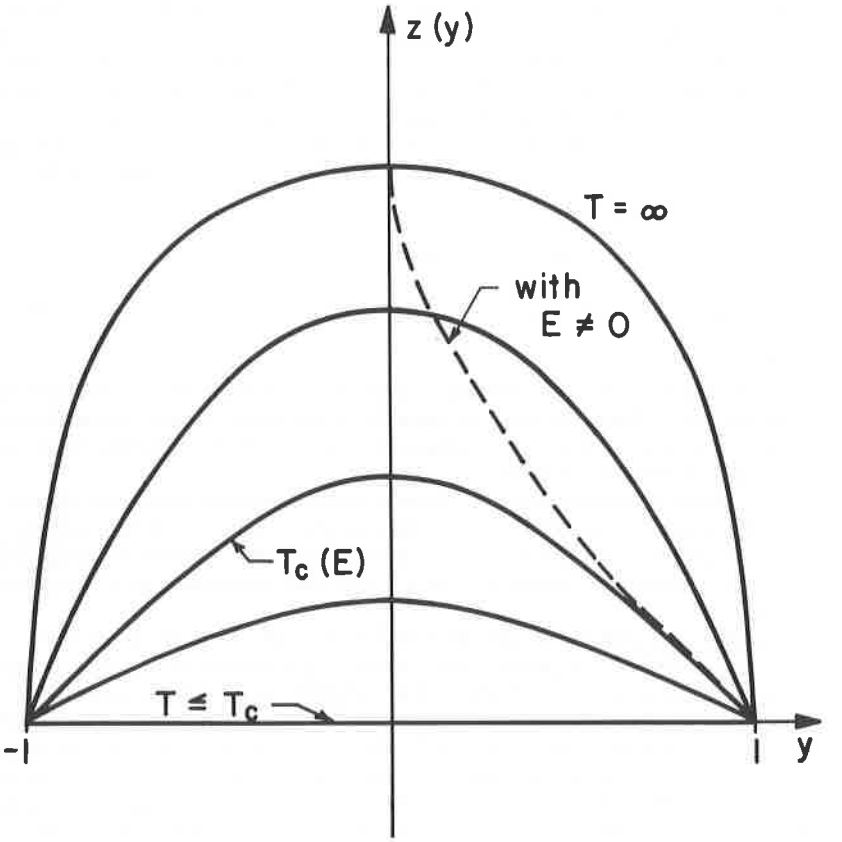
What happens with an electric field is somewhat more complicated, and it is best to refer to Fig. 2 where z(y) is plotted against y for various temperatures. All curves have the correct property that z(y) is convex and symmetric in y, and z(y) increases with temperature.

F model: It can be shown from (IV.5) and (IV.6) that, for $\Delta < -1$ $(T < T_c)$, z(y) has a finite, but non-zero derivative at y = 0. In other words there is a cusp. Furthermore, $\frac{d\,z(y)}{dy}$ at y = 0 becomes more negative as T decreases, and is $-\infty$ at T = 0. With E = 0 we proceed up the y = 0 axis, but with $E \neq 0$, (IV.17) tells us to go along the dotted line in Fig. 2 until T is low enough that the slope at y = 0 is more negative than $-\beta E$. It can also be shown that the slope at y = 1 is infinite so that y < 1 for all E. The value of y for any given T and E is clearly the average polarization per vertex of the crystal and from these considerations we can conclude that:

1. Above T_c the polarization is the usual S-shaped function of E.

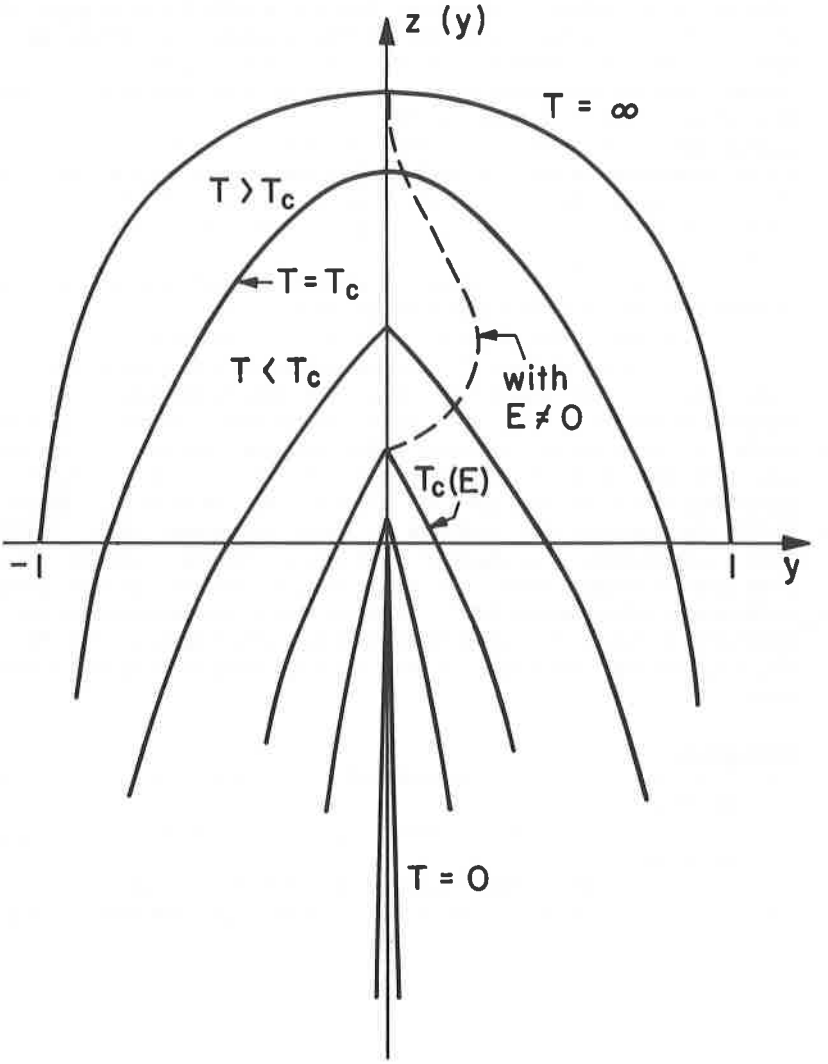
2. Below T_c, the polarization vanishes identically until βE exceeds a critical value given by the slope at y = 0. The polarization then increases with what can be shown to be an infinite slope.

3. The infinite order singularity obtained above occurs only at T_c and only for y = 0. With a fixed electric field, therefore, we cross no singularity until we hit the y = 0 axis at some temperature, $T_c(E)$. This is a critical temperature because the analytic character of F changes abruptly. It can be shown to be second order (no latent heat),



KDP Model

Fig. 2. $z(y) = \lim_{N\to\infty} N^{-1} \ell n\, \lambda(n)_{max}$ is plotted as a function of y = polarization per vertex for the two models. These curves are schematic only. The dashed curve is the trajectory traced out when a fixed electric field is present.



F Model

with C proportional to $[T - T_c(E)]^{-\frac{1}{2}}$ on the high temperature side. Obviously the critical temperature decreases with E and is zero for $E \geq \epsilon$. It is a curious result that all thermodynamic quantities are identical with and without an electric field below $T_c(E)$. It is also curious that we pass from an infinite to a second order phase transition when we apply an electric field.

KDP model. Here there is no cusp, but the novel feature below T_c is, as we explained in (IV.21), that $z(y)$ is identically zero for all y and T. Furthermore, the slope at $y = 1$ is now finite for $T > T_c$ and decreases with T. A typical y trajectory is again shown by a dotted line.

1. Above T_c the polarization is an S-shaped function of E, but it saturates at $y = 1$ for a finite value of E.

2. Below T_c the polarization is perfect for all E.

3. With an electric field we no longer proceed along the $y = 0$ axis, but along the dotted line, until a certain temperature, $T_c(E)$, is reached at which $y = 1$. Then for all lower temperatures, y is always unity. F, once again, consists of two different analytic pieces which join at $T_c(E)$. It can be shown that $C = 0$ and $U = -E$ below $T_c(E)$ since $z(y = 1)$ is clearly zero for all T (cf. (IV.6) et seq.). Above $T_c(E)$, the specific heat still has the same divergence as in (IV.22). Thus, application of an electric field causes the transition to change from first to second order, but it also causes the critical temperature to increase with applied field. Such behaviour is observed experimentally, although for real KDP (potassium dihydrogen phosphate KH_2PO_4) the transition may be only approximately first order in zero field.

References

1. T. D. Schultz, D. C. Mattis and E. H. Lieb, Rev. Mod. Phys. 36, 856 (1964).
2. E. H. Lieb, Phys. Rev. Letters 18, 692 (1967); Phys. Rev. 162, 162 (1967).
3. E. H. Lieb, Phys. Rev. Letters 18, 1046 (1967); 19, 108 (1967).
4. C. N. Yang and C. P. Yang, Phys. Rev. 150, 321 and 327 (1966).

SELF-CONSISTENCY AND STABILITY IN THE VARIATIONAL CALCULATION OF THERMODYNAMIC QUANTITIES†

Albrecht Huber
Department of Mathematical Physics
University of Birmingham
Birmingham, England

Contents

I. Introduction
II. Criteria for Self-Consistency and Stability
III. "Universality" of the Set of Trial Density Matrices
IV. Proof of the Self-Consistency and Stability Theorem
V. The Role of the Duhem-Gibbs Relation
VI. Magnetic Systems
VII. Concluding Remarks

I. Introduction

All thermodynamic equilibrium properties of a macroscopic system can be derived from a single quantity, the so-called thermodynamic potential, which is a function of the independent thermodynamic variables. In the simplest case of a one-component many-particle system, not subjected to external fields and not undergoing bulk motion, the most practical choice for the independent thermodynamic variables consists of the temperature T, the volume V, and the chemical potential μ. The corresponding thermodynamic potential is called the "grand potential" and is, for the case of quantum systems, defined by

$$Y(T,V,\mu) = - k_B T \log \mathrm{Tr}_{\mathcal{F}^\varepsilon} \left\{ \exp\left[-\frac{1}{k_B T} (\mathcal{H} - \mu \mathfrak{n}) \right] \right\} \qquad (I.1)$$

where k_B denotes Boltzmann's constant. The Hamiltonian $\mathcal{H}$ and the total particle number operator $\mathfrak{n}$ in (I.1) are assumed to be written in the formalism of second quantization, and act in the full occupation number Hilbert space, or "Fock space," $\mathcal{F}^\varepsilon$, which contains

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

wavefunctions corresponding to arbitrary total particle number. The superscript ϵ (which may, e.g., be taken to be +1 in the case of Bose-Einstein and -1 in the case of Fermi-Dirac statistics) serves to indicate that only those wavefunctions are admitted in $\mathcal{H}^{\epsilon}$ which have the correct symmetry properties with respect to the interchange of particle coordinates, depending on the nature of the constituent particles of the system.

If the Hamiltonian $\mathcal{K}$ describes a system of interacting particles, it is usually quite impossible to calculate the quantity (I.1) exactly, and one has to take recourse to approximation procedures for its evaluation. Among these, the various forms and modifications of perturbation theory and diagrammatic expansions have been most widely used. There exists, however, an equally appealing and frequently less complicated variational procedure for the approximate calculation of $Y(T,V,\mu)$, which is based on the following fundamental extremum property:

Gibbs-Delbrück-Molière minimum principle for the grand potential. Any self-adjoint linear operator $\tilde{W}$ on $\mathcal{H}^{\epsilon}$ with

$$\tilde{W} \geq 0, \quad \mathrm{Tr}_{\mathcal{H}^{\epsilon}} \tilde{W} = 1 \tag{I.2}$$

satisfies the inequality

$$Y(T,V,\mu) \leq \mathrm{Tr}_{\mathcal{H}^{\epsilon}} \{\mathcal{K}\tilde{W} - \mu\mathcal{N}\tilde{W} + k_B T \tilde{W} \log \tilde{W}\} \; ; \tag{I.3}$$

the equal sign holds if and only if $\tilde{W}$ is equal to the equilibrium density matrix

$$W(T,V,\mu) = \frac{\exp\left[-\frac{1}{k_B T}(\mathcal{K} - \mu\mathcal{N})\right]}{\mathrm{Tr}_{\mathcal{H}^{\epsilon}}\left\{\exp\left[-\frac{1}{k_B T}(\mathcal{K} - \mu\mathcal{N})\right]\right\}} \; . \tag{I.4}$$

Inequality (I.3) is the generalisation of a well-known result of Gibbs[1] to the case of quantum statistical mechanics, and was first established by Delbrück and Molière.[2] It is a special case of a general extremum principle for arbitrary thermodynamic potentials which is of great importance both for the foundations of quantum statistical mechanics[3] and for its application to particular problems.[4]

The minimum property (I.3) suggests a variational procedure for the approximate calculation of $Y(T,V,\mu)$ which bears a close resemblance to the approximate calculation of atomic wavefunctions based on the famous Rayleigh-Ritz variational principle. One chooses an ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$ for $\tilde{W}$ which satisfies the conditions (I.2) and

which contains a set $\{\xi_k\}$ of variational parameters left at one's disposal. The ansatz must be sufficiently simple to permit an exact calculation of the right-hand side of (I.3) while at the same time retaining as much as possible of those correlations which are relevant for a valid description of the system under consideration. One then minimizes the functional on the right-hand side of (I.3) with respect to the ξ_k and takes the minimal value $\hat{Y}(T,V,\mu)$ of the functional as the "best" approximation for $Y(T,V,\mu)$ which can be found in this way.

As always in procedures of this kind, there is no general prescription for how to construct a suitable ansatz for $\tilde{W}$ in the first place. Neither is there an absolute guarantee that the functional which a given ansatz for $\tilde{W}$ generates will possess a minimum, although the fact that the functional is always bounded from below by the true grand potential $Y(T,V,\mu)$ goes a long way to safeguard this point. Moreover, once the minimum of the functional is found it is not at all easy to tell how good an approximation for $Y(T,V,\mu)$ it is, since as long as one does not have, in addition, a lower bound on $Y(T,V,\mu)$ there is no way of estimating the difference $\hat{Y}(T,V,\mu) - Y(T,V,\mu)$. In spite of these uncertainties, the procedure has been applied with good success to a large number of problems in the theory of quantum mechanical many-particle systems.

In this lecture, I shall not be concerned with any of the problems just mentioned, since there is very little that can be said about them in a general way. Instead, I shall discuss a problem which, as we shall see, can be answered in a completely general manner. This problem is: Under what conditions for the ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$ does the minimization procedure outlined above yield an approximation for the grand potential which is self-consistent and thermodynamically stable?

The answer to this question which I shall develop in the course of this lecture can be equally well formulated for completely arbitrary thermodynamic potentials, depending on $r - 1$ intensive and $n - r + 1$ extensive thermodynamic variables ($1 \leq r \leq n$). However, such a formulation is both rather unwieldy and not very appealing to the physicist's imagination. On the other hand, the discussion of the problem for the grand potential permits the general principles behind the argument to become apparent. For these reasons, and because it is by far the most important case, I shall limit most of the discussion to the case of the grand potential.

The next section is devoted to giving a precise meaning to the terms "self-consistency" and "thermodynamic stability."

II. Criteria for Self-Consistency and Stability

In general, the application of the variational principle (I.3) to the approximate calculation of the grand potential results in a final

approximation $\hat{W}(T,V,\mu)$ for the equilibrium density matrix.† From $\hat{W}(T,V,\mu)$, the final approximation $\hat{Y}(T,V,\mu)$ for the grand potential is obtained by

$$\hat{Y}(T,V,\mu) = \mathrm{Tr}_{\mathfrak{f}\,\epsilon} \{\mathcal{H}\hat{W}(T,V,\mu) - \mu\mathcal{N}\hat{W}(T,V,\mu) + k_B T\, \hat{W}(T,V,\mu) \times \log \hat{W}(T,V,\mu)\} . \qquad (II.1)$$

In any situation of this kind, there exists the following possible difficulty which, in the case of the free energy $F(T,V,N)$, was first pointed out and resolved by Wagner and Koppe.[5] If we want to calculate approximations for the entropy S, the average total particle number N, and the internal energy E, we can either form the expressions

$$\hat{S}_1 = -k_B \mathrm{Tr}_{\mathfrak{f}\,\epsilon}\{\hat{W}\log\hat{W}\}\,,\quad \hat{N}_1 = \mathrm{Tr}_{\mathfrak{f}\,\epsilon}\{\mathcal{N}\hat{W}\},\quad \hat{E}_1 = \mathrm{Tr}_{\mathfrak{f}\,\epsilon}\{\mathcal{H}\hat{W}\} \qquad (II.2)$$

which make explicit use of the fact that we have accepted $\hat{W}(T,V,\mu)$ as our approximation for the equilibrium density matrix. It is, however, equally legitimate to form the approximations

$$\hat{S}_2 = -\frac{\partial\hat{Y}}{\partial T}\,,\quad \hat{N}_2 = -\frac{\partial\hat{Y}}{\partial\mu},\quad \hat{E}_2 = \hat{Y} - T\frac{\partial\hat{Y}}{\partial T} - \mu\frac{\partial\hat{Y}}{\partial\mu} \qquad (II.3)$$

in keeping with the usual differential equations of thermodynamics. Neither of the two sets of approximations has an a priori higher claim than the other to represent the quantities in question.

In principle it is also possible, although with less intuitive justification, to adopt hybrid approximations which are formed by some kind of mixture of (II.2) and (II.3). An even greater variety of competing approximations appears if one goes over to other thermodynamic potentials, such as the (Helmholtz) free energy $F(T,V,N)$ or the internal energy $E(S,V,N)$, by means of a Legendre transformation with respect to μ or with respect to μ and T. For these Legendre transformations themselves exist in several versions, and the number

† An exception to this statement occurs if $\hat{Y}(T,V,\mu)$ is not obtained as an actual value of the variational functional for Y, but rather as the limit of a sequence of such values in a limiting process in which no limit for the corresponding sequence of trial density matrices exists. Procedures of this kind will not be considered in the present discussion.

of rivalling expressions that one ends up with for, say, the specific heat at constant volume is quite monstrous.

It follows from these considerations that, unless the two sets of approximations (II.2) and (II.3) are, in fact, equal, the approximation $\hat{W}(T,V,\mu)$ for the equilibrium density matrix has to be rejected as totally inconsistent. We arrive thus at the following self-consistency requirement:

Criterion for self-consistency. An approximation $\hat{W}(T,V,\mu)$ for the equilibrum density matrix $W(T,V,\mu)$ is called self-consistent if and only if it satisfies, for all permitted values of T,V, and μ, the conditions

$$\frac{\partial \hat{Y}}{\partial T} = k_B \operatorname{Tr}_{\mathcal{H}_\varepsilon} \{\hat{W} \log \hat{W}\} \tag{II.4}$$

and

$$\frac{\partial \hat{Y}}{\partial \mu} = - \operatorname{Tr}_{\mathcal{H}_\varepsilon} \{n\hat{W}\} \tag{II.5}$$

where the function $\hat{Y}(T,V,\mu)$ is given by (II.1).

From (II.4), (II.5) and (II.1) it follows immediately that the two definitions for $\hat{E}$ given in (II.2) and (II.3), too, agree, so that this requirement need not be explicitly included in the self-consistency criterion.

The approximations obtained for the various thermodynamic quantities from a $\hat{W}$ which satisfies the self-consistency requirements (II.4) and (II.5) combine the obvious advantages of (II.2) and (II.3). From (II.2) it follows that the approximations for S and N are always non-negative, and that the approximation for $E - \mu N$ is always larger then or equal to the ground state energy of the operator $\mathcal{H} - \mu h$ on $\mathcal{H}_\varepsilon$. From (II.3), on the other hand, it follows that the approximations for S, N and E, together with the approximations for any other thermodynamic quantities derived from them, satisfy all the differential equations (such as the generalised Helmholtz-Gibbs relations, the various Maxwell relations for mixed derivatives, etc.) which, according to classical thermodynamics, are valid for the true grand potential and its derivatives. The only relations of equilibrium thermodynamics the validity of which is not guaranteed by the self-consistency conditions (II.4) and (II.5) are the Duhem-Gibbs relation and the second- and higher-order "fluctuation-response" theorems; to both I shall return later.

There exists, however, a second general property which we would like to demand of the approximative solution $\hat{Y}(T,V,\mu)$. It is well known that the true grand potential obeys the inequality

$$\begin{pmatrix} -\frac{\partial^2 Y}{\partial T^2} & -\frac{\partial^2 Y}{\partial T \partial \mu} \\ -\frac{\partial^2 T}{\partial \mu \partial T} & -\frac{\partial^2 Y}{\partial \mu^2} \end{pmatrix} > 0 \qquad \text{(II.6)}$$

where the > 0 notation means that the matrix $-\mathcal{Y}''$ on the left-hand side of (II.6) is positive definite. Inequality (II.6) is valid for all values of T and μ for which the operator $\mathcal{K} - \mu\mathcal{N}$ is essentially self-adjoint on $\mathcal{F}^{\varepsilon}$ and for which the operator $\exp\left\{\frac{1}{k_B T}(\mathcal{K} - \mu\mathcal{N})\right\}$ has a finite trace, unless there exists a real number λ with $0 \leq \lambda \leq 1$ such that

$$\lambda\mathcal{K} + (1 - \lambda)\mathcal{N} = 0 \; ; \qquad \text{(II.7)}$$

the latter case is of little interest and will be excluded from the discussion.†

To prove (II.6) one notices that $-\mathcal{Y}''$ can be written as

$$-\mathcal{Y}'' = \frac{1}{k_B T^3} \begin{pmatrix} 1 & -\mu \\ 0 & T \end{pmatrix} \mathcal{K}(\mathcal{K},\mathcal{N}) \begin{pmatrix} 1 & 0 \\ -\mu & T \end{pmatrix} \; ; \qquad \text{(II.8)}$$

here, $\mathcal{K}(\mathcal{K},\mathcal{N})$ denotes the covariance matrix

$$\mathcal{K}(\mathcal{K},\mathcal{N}) = \begin{pmatrix} \langle\mathcal{K}^2\rangle - \langle\mathcal{K}\rangle^2 & \langle\mathcal{K}\mathcal{N}\rangle - \langle\mathcal{K}\rangle\langle\mathcal{N}\rangle \\ \langle\mathcal{N}\mathcal{K}\rangle - \langle\mathcal{N}\rangle\langle\mathcal{K}\rangle & \langle\mathcal{N}^2\rangle - \langle\mathcal{N}\rangle^2 \end{pmatrix} \qquad \text{(II.9)}$$

which is by its very nature positive definite (unless (II.7) holds, in which case at least one of the eigenvalues of $\mathcal{K}(\mathcal{K},\mathcal{N})$ is zero). This proof is, incidentally, independent of whether $\mathcal{K}$ and $\mathcal{N}$ commute or not; in the latter case, one makes use of the formula‡

$$\frac{d}{dt} e^{A(t)} = e^{A(t)} \int_0^1 e^{-A(t)s} \frac{dA(t)}{dt} e^{A(t)s} \, ds \qquad \text{(II.10)}$$

for non-commutative one-parameter operator manifolds $A(t)$, and reinterprets the second-order moments in (II.9) as

---

† In the case (II.7), where $\mathcal{K}$ and $\mathcal{N}$ are linearly dependent, the two variables T and μ are no longer independent of each other, and only one intensive variable should have been introduced in the first place.

‡ For derivations and applications of this formula, the reader may consult the survey given by Wilcox.[34]

$$\langle AB \rangle = \mathrm{Tr}_{\mathcal{H}_\varepsilon} \left\{ \int_0^1 A\, W^s\, B\, W^{1-s}\, ds \right\} \qquad \text{(II.11)}$$

where W stands for the equilibrium density matrix defined in (I.4).

Inequality (II.6) states that the true grand potential $Y(T,V,\mu)$ is, for fixed V, a simultaneously strictly concave function of T and μ. This property, which is equivalent to the so-called "thermodynamic stability conditions" for Y in the formulation developed by Münster,[6] has several important consequences. Firstly, it plays a fundamental role for the Legendre transformations of Y with respect to the intensive variables T and μ. These transformations are:

(i) $$K(S,V,\mu) = Y(T_1(S,V,\mu),V,\mu) + T_1(S,V,\mu)S \qquad \text{(II.12)}$$

with

$$-\left[\frac{\partial Y}{\partial T}\right]_{T = T_1(S,V,\mu)} = S\ , \qquad \text{(II.13)}$$

(ii) $$F(T,V,N) = Y(T,V,\mu_1(T,V,N)) + \mu_1(T,V,N)N \qquad \text{(II.14)}$$

with

$$-\left[\frac{\partial Y}{\partial \mu}\right]_{\mu = \mu_1(T,V,N)} = N\ , \qquad \text{(II.15)}$$

and

$$\text{(iii)}\ E(S,V,N) = Y(T_2(S,V,N),V,\mu_2(S,V,N)) + T_2(S,V,N)S + \mu_2(S,V,N)\,N \qquad \text{(II.16)}$$

with

$$-\left[\frac{\partial Y}{\partial T}\right]_{\substack{T = T_2(S,V,N)\\ \mu = \mu_2(S,V,N)}} = S,\quad -\left[\frac{\partial Y}{\partial \mu}\right]_{\substack{T = T_2(S,V,N)\\ \mu = \mu_2(S,V,N)}} = N. \qquad \text{(II.17)}$$

Here, $K(S,V,\mu)$ is the rarely used thermodynamic potential defined by $K = E - \mu N$, while $F(T,V,N)$ denotes the Helmholtz free energy and $E(S,V,T)$ the internal energy. That the equations (II.13) and (II.15) and the two coupled equations (II.17) indeed determine the functions

$T_1(S,V,\mu)$, $\mu_1(T,V,N)$, $T_2(S,V,N)$ and $\mu_2(S,V,N)$ in a unique manner is ensured by the inequalities

$$-\frac{\partial^2 Y}{\partial T^2} > 0 \ , \tag{II.18}$$

$$-\frac{\partial^2 Y}{\partial \mu^2} > 0 \ , \tag{II.19}$$

and

$$\begin{vmatrix} -\frac{\partial^2 Y}{\partial T^2} & -\frac{\partial^2 Y}{\partial T \partial \mu} \\ -\frac{\partial^2 Y}{\partial \mu \partial T} & -\frac{\partial^2 Y}{\partial \mu^2} \end{vmatrix} > 0 \ , \tag{II.20}$$

respectively, the determinant in (II.20) being the Jacobian for the system (II.17). These three inequalities follow immediately from (II.6).

The second, and equally important consequence of (II.6) is that it ensures the positivity of certain second-order static response functions which, on physical grounds, are expected to be positive for any thermodynamically stable system. In the case of the thermodynamic potential (I.1), these quantities are: the specific heat at constant volume, the specific heat at constant pressure, the compressibility at constant temperature, and the compressibility at constant entropy. As we shall see in a later section, these four quantities can all be expressed (the three latter ones with the help of the Duhem-Gibbs relation) in terms of the matrix elements of $-\psi''$, and the expressions found are such that the positivity of $-\psi''$ is carried over to the response functions.

These considerations lead to our second criterion for the approximate thermodynamic potential $\hat{Y}(T,V,\mu)$:

Stability criterion. An approximation $\hat{Y}(T,V,\mu)$ for the grand potential is called thermodynamically stable if and only if

$$\begin{pmatrix} -\frac{\partial^2 \hat{Y}}{\partial T^2} & -\frac{\partial^2 \hat{Y}}{\partial T \partial \mu} \\ -\frac{\partial^2 \hat{Y}}{\partial \mu \partial T} & -\frac{\partial^2 \hat{Y}}{\partial \mu^2} \end{pmatrix} > 0 \tag{II.21}$$

for all permitted values of T, V, and μ.

III. "Universality" of the Set of Trial Density Matrices

Let us imagine that we insert, into the right-hand side of (I.3), an ansatz $W(T,V,\mu;\{\xi_k\})$ for $\tilde{W}$ containing a set of variational parameters $\{\xi_k\}$, where each ξ_k may be restricted to a certain subset R_k of the real axis,† and k takes on all different values in an index set K. If the trial functional $\tilde{Y}(T,V,\mu;\{\xi_k\})$ which this ansatz generates is minimized with respect to the ξ_k, we might expect to be able to prove that the minimum $\hat{Y}$ of $\tilde{Y}$ obeys the self-consistency criterion (II.4), (II.5) and the stability criterion (II.21) by exploiting the fact that, at the minimum, the functional $\tilde{Y}$ is stationary and has a positive definite matrix of second derivatives with respect to the ξ_k. However, it turns out that this is not possible in the general case because of the presence of certain unwanted terms in the derivatives of $\hat{Y}$ with respect to T and μ which stem from the explicit T- and μ-dependence of W. In order to overcome this difficulty we impose a restriction on the ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$.

Let $\mathfrak{w}[\mathcal{H}]$ denote the set of all density matrices on a Hilbert space $\mathcal{H}$, i.e. the set of all self-adjoint linear operators W on $\mathcal{H}$ with

$$W \geq 0 , \qquad \mathrm{Tr}_{\mathcal{H}}\, W = 1 . \tag{III.1}$$

The ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$ then furnishes, for every set of values T, V and μ, a mapping of the set of all permitted parameter sets $\{\xi_k\}$ ($\xi_k \in R_k$, $k \in K$) into $\mathfrak{w}[\mathcal{H}^c]$. Let us denote by $\tilde{\mathfrak{w}}$ the subset of $\mathfrak{w}[\mathcal{H}^c]$ consisting of all images which are generated under this mapping when the ξ_k are allowed to take on all values in their respective ranges R_k. The condition which we want to impose on the ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$ is that the set $\tilde{\mathfrak{w}}$ of trial density matrices is the same for all values of T and μ (although it will necessarily depend on V). We call this condition the Universality condition for the set of trial density matrices. A subset $\tilde{\mathfrak{w}} \subseteq \mathfrak{w}$ is called a universal set of trial density matrices for the variational principle (I.3) if and only if $\tilde{\mathfrak{w}}$ is independent of the intensive thermodynamic variables T and μ.

It should be emphasized that universality of $\tilde{\mathfrak{w}}$ does not mean that the ansatz $\tilde{W}$ must contain no reference to T and μ at all, since it is obviously possible to parametrize even a universal $\tilde{\mathfrak{w}}$ in a T- and μ-dependent fashion. What universality does imply, however, is that any such dependence of the ansatz $\tilde{W}$ on T and μ must be spurious in the following sense. There must exist, for every $k \in K$, a one-to-one mapping $\xi_k \leftrightarrow \xi_k' = f_k(T,V,\mu;\xi_k)$ of R_k onto a subset $R_k' \subseteq R^1$

† Complex-valued variational parameters can, of course, be used, but for the purpose of our discussion we shall assume that they are always written in terms of their real and imaginary parts.

such that the R_k' are independent of T and μ and that, furthermore, a new operator-valued functional $\tilde{W}'(V;\{x_k\})$ $(x_k \in R_k',\ k \in K)$ can be found which satisfies

$$\tilde{W}(T,V,\mu;\{\xi_k\}) = \tilde{W}'(V;\{f_k(T,V,\mu;\xi_k)\}) \qquad \text{(III.2)}$$

for all permitted values of T, V, and μ and all permitted parameter sets $\{\xi_k\}$.

To illustrate this rather abstract remark let us look at a specific example. As was first shown by Mermin,[7] the temperature-dependent Hartree-Fock approximation for a many-particle system interacting via two-body forces can be obtained by minimizing the right-hand side of (I.3) for the ansatz

$$W(T,V,\mu;\{\eta_{\kappa,\kappa'}\}) = \frac{\exp\left[-\frac{1}{k_B T}\left(\sum_{\kappa,\kappa'} \eta_{\kappa,\kappa'} c_\kappa^+ c_{\kappa'} - \mu \sum_\kappa c_\kappa^+ c_\kappa\right)\right]}{\mathrm{Tr}_{\mathfrak{H}}\left\{\exp\left[-\frac{1}{k_B T}\left(\sum_{\kappa,\kappa'} \eta_{\kappa,\kappa'} c_\kappa^+ c_{\kappa'} - \mu \sum_\kappa c_\kappa^+ c_\kappa\right)\right]\right\}} \qquad \text{(III.3)}$$

where the $\eta_{\kappa,\kappa'}$ play the role of variational parameters and are subjected only to the hermiticity condition† $\eta_{\kappa',\kappa} = \eta_{\kappa,\kappa'}^*$. The subscripts κ,κ' stand for sets of quantum numbers characterising single-particle states, and the c_κ^+, c_κ are the creation and annihilation operators for particles in these states. The ansatz (III.3) contains an explicit dependence on T and μ. However, this dependence is spurious, since T and μ occur only in the combination

$$f_{\kappa,\kappa'}(T,V,\mu;\eta_{\kappa,\kappa'}) = \frac{\eta_{\kappa,\kappa'} - \mu\delta_{\kappa,\kappa'}}{k_B T} \equiv \eta'_{\kappa,\kappa'}, \qquad \text{(III.4)}$$

where $\delta_{\kappa,\kappa'}$ stands for the appropriate product of Kronecker symbols. The mapping from $\eta_{\kappa,\kappa'}$ to $\eta'_{\kappa,\kappa'}$, which (III.4) provides is clearly a one-to-one mapping and conserves hermiticity; the new T- and μ-independent parametrization of W is given by

† The fact that the $\eta_{\kappa,\kappa'}$ are not necessarily real contradicts, strictly speaking, our convention of using only real-valued variational parameters. However, this can easily be taken care of by re-writing (III.3) in terms of the real and imaginary parts of the $\eta_{\kappa,\kappa'}$.

$$\tilde{W}'(V;\{\eta'_{\kappa,\kappa'}\}) = \frac{\exp\left(-\sum_{\kappa,\kappa'} \eta'_{\kappa,\kappa'}\, c_\kappa^+ c_{\kappa'}\right)}{\mathrm{Tr}_{\mathcal{H}}\, \varepsilon \left\{\exp\left(-\sum_{\kappa,\kappa'} \eta'_{\kappa,\kappa'}\, c_\kappa^+ c_{\kappa'}\right)\right\}} \quad . \qquad \text{(III.5)}$$

Thus the set of trial density matrices defined by (III.3) is universal in the sense of the definition given above, and the conclusions which we are going to draw by assuming universality of $\tilde{\mathfrak{w}}$ can be applied to the case of the Hartree-Fock approximation.

An entirely analogous situation exists with respect to the Bardeen-Cooper-Schrieffer ("BCS") theory of superconductivity in the formulation given by Valatin[8] and by Blatt, Matsubara and Thompson,[9] and for the so-called "pairing theory" of interacting Bose systems developed by Valatin and Butler,[10] Girardeau and Arnowitt,[11] Tolmachev,[12] Gross,[13] and others. In the case of the BCS-theory, one uses an ansatz for $\tilde{W}$ of the form (III.3) but with the c_κ^+ and c_κ, replaced by the quasiparticle coordinates

$$\tilde{c}_\kappa^+ = c_\kappa^+ \cos s_\kappa - c_{-\kappa} \sin s_\kappa \quad . \qquad \text{(III.6)}$$

In this case, one has $\kappa \equiv (\underset{\sim}{k},\sigma)$, where $\underset{\sim}{k}$ is the linear momentum of the electrons and σ their spin; the s_κ are required to be real and to satisfy the antisymmetry relation $s_\kappa = -s_{-\kappa}$. In the pairing theory for Bose systems one uses again an ansatz for $\tilde{W}$ of the form (III.3), but this time the c_κ^+, c_κ are replaced by

$$\tilde{c}_\kappa^+ = c_\kappa^+ \cosh t_\kappa + c_{-\kappa} \sinh t_\kappa + \lambda \delta_{o,\kappa} \quad . \qquad \text{(III.7)}$$

In this case, κ simply denotes the linear momentum of the particles (provided they have zero spin), and the t_κ are assumed to be real and subjected to the conditions $t_\kappa = t_{-\kappa}$ and $t_o = 0$; the term $\lambda\delta_{o,\kappa}$ serves to take account of the macroscopic occupation of the $\kappa = 0$-state. The new parameters s_κ in (III.6), and t_κ and λ in (III.7) are used as variational parameters in addition to the $\eta_{\kappa,\kappa'}$ already contained in (III.3).

If the original set $\tilde{\mathfrak{w}}(T,V,\mu)$ of trial density matrices is not universal, various methods to construct a universal set from it offer themselves. One idea, which would presumably encounter considerable computational difficulties, would be to choose the set $\bigcap_{T,\mu} \tilde{\mathfrak{w}}(T,V,\mu)$, i.e. the largest common subset of all the $\tilde{\mathfrak{w}}(T,V,\mu)$.

A more practicable alternative would consist in selecting two arbitrary, but fixed values T_o and μ_o and in using the universal set $\tilde{W}(T_o,V,\mu_o)$. The best method, however, seems to be the one proposed by Wagner and Koppe,[5] in which the ansatz $\tilde{W}(T,V,\mu;\{\xi_k\})$ is replaced by $\tilde{W}(T',V,\mu';\{\xi_k\})$, where T' and μ' are two new variational parameters, introduced in addition to the ξ_k and varied simultaneously with and independently of them over a range which corresponds to the physically permitted values of T and μ respectively. This procedure has two great advantages. Firstly, it means that the new set of trial density matrices contains the set originally proposed as a subset, so that the new approximation for the grand potential lies at least as close to the true grand potential as the original one. Secondly, the computation of the trial functional $\tilde{Y}(T,V,\mu;T',\mu',\{\xi_k\})$ is no more difficult than for the original non-universal ansatz, and the procedure of minimization, although certainly more complicated owing to the presence of two additional variational parameters, can generally be expected to be of approximately the same order of computational difficulty as in the original case.

Wagner and Koppe, who considered the case of the free energy $F(T,V,N)$, described this method as "using the temperature as a variational parameter." In the same sense one can, in the case of the grand potential, speak of the variation of temperature and chemical potential, and, in the case of arbitrary thermodynamic potentials, of the variation of all intensive thermodynamic variables. Examples of thermodynamic variables which one may have to treat as variational parameters in this way are the pressure, the magnetic field, the velocity of hydrodynamic flow or the angular velocity of hydrodynamic rotation, to name but a few.

This situation was likened by Wagner and Koppe to the problem of ensuring the validity of the virial theorem in the approximate calculation of atomic or molecular wavefunctions with the help of the Rayleigh-Ritz variational principle. There, the expectation values of the kinetic energy E_{kin} and of the potential energy E_{pot} are known to be related by the virial theorem

$$(\psi, E_{kin}\,\psi) = -\tfrac{1}{2}(\psi, E_{pot}\psi) \qquad \text{(III.8)}$$

for any exact eigenstate ψ of the total Hamiltonian. While (III.8) may be violated for an approximate wavefunction obtained from the variational principle, its validity is restored if the original ansatz $\tilde{\psi}(\underset{\sim}{r})$ is replaced by the ansatz $\eta^{3/2}\,\tilde{\psi}(\eta\,\underset{\sim}{r})$ and the "scaling factor" η is used

as an additional variational parameter in the minimization procedure.† This replacement means that one allows the effective radius of the wavefunction to be determined in a self-consistent manner, rather than to fix it from the outset, even though the radius may be known, e.g., from experiment. Similarly, in replacing T and μ by the variational parameters T′ and μ', one deliberately refrains from using one's knowledge about the actual values of the temperature and the chemical potential in the ansatz for the trial density matrix, and allows these values instead to be determined self-consistently together with the best values for the other variational parameters. As a reward, one obtains an approximation for the grand potential which satisfies the criteria of self-consistency and stability as established in Sec. II above. This will be shown in the next section.

IV. Proof of the Self-Consistency and Stability Theorem

Let us assume that we have constructed, by some method or other, a universal set $\tilde{\mathfrak{W}}(V)$ of trial density matrices which can be parametrized by a set $\{\xi_k\}$ ($\xi_k \in R_k$, $k \in K$) of real-valued parameters. Since $\tilde{\mathfrak{W}}$ is universal, we can choose the parametrization in a T- and μ-independent fashion, which also implies that the ranges R_k of the ξ_k are independent of T and μ. An individual element of $\tilde{\mathfrak{W}}$ is thus written as $\tilde{W}(V;\{\xi_k\})$.

We assume, furthermore, that the expectation values

$$\tilde{E}(V;\{\xi_k\}) = \mathrm{Tr}_{\mathcal{H}\in\mathfrak{E}}\left[\mathcal{K}\tilde{W}(V;\{\xi_k\})\right] , \tag{IV.1}$$

$$\tilde{S}(V;\{\xi_k\}) = -k_B\,\mathrm{Tr}_{\mathcal{H}\in\mathfrak{E}}\left[\tilde{W}(V;\{\xi_k\})\log\tilde{W}(V;\{\xi_k\})\right] , \tag{IV.2}$$

and

† The idea of introducinga scaling factor η as a variational parameter into an atomic trial wavefunction was first proposed by Kellner[14] in his work on the ionisation potential of helium, and was soon applied with great success in calculations by Wang[15] on the hydrogen molecule and by Hylleraas[16] on the ground state of the helium atom. The theorem that minimization with respect to η guarantees the validity of the virial theorem for the approximate wavefunction is due to Fock.[17] The scaling idea is equally successful in the problem of establishing the virial theorem for scattering processes, as was shown by Demkov[18] with the help of variational principles due to Hulthén[19] and to Kohn.[20] For a more detailed discussion of these matters, the reader is referred to an article by Löwdin[21] and to the books by Demkov[22] and by Moiseiwitsch.[23]

$$\tilde{N}(V;\{\xi_k\}) = \mathrm{Tr}_{\not f}\, \epsilon^{[n\tilde{W}(V;\{\xi_k\})]} \qquad \text{(IV.3)}$$

exist and possess continuous first- and second-order derivatives with respect to all the ξ_k. In terms of these expectation values, the trial functional for the grand potential can be written as

$$\tilde{Y}(T,V,\mu;\{\xi_k\}) = \tilde{E}(V;\{\xi_k\}) - T\,\tilde{S}(V;\{\xi_k\}) - \mu\,\tilde{N}(V;\{\xi_k\}) \ . \qquad \text{(IV.4)}$$

Because of (I.3), this functional is bounded from below and thus possesses an infimum. We now make the further assumption that $\tilde{Y}(T,V,\mu;\{\xi_k\})$ assumes its infimum for a set of values $\{\hat{\xi}_k(T,V,\mu)\}$ in $\{\xi_k\}$-space.† Then the $\hat{\xi}_k(T,V,\mu)$ satisfy the conditions

$$\left[\frac{\partial \tilde{Y}}{\partial \xi_i}\right]_{\{\xi_k\} = \{\hat{\xi}_k(T,V,\mu)\}} = 0 \text{ for all } i \in K \ , \qquad \text{(IV.5)}$$

and

$$\left(\left[\frac{\partial^2 \tilde{Y}}{\partial \xi_i \partial \xi_j}\right]_{\{\xi_k\} = \{\hat{\xi}_k(T,V,\mu)\}}\right) \geq 0 \ , \qquad \text{(IV.6)}$$

where the $\geq$-notation means that the left-hand side of (IV.6) is a positive semi-definite matrix, summation kernel, or integral kernel, depending on whether the index set K is finite, or denumerable, or a continuum.

Introducing the abbreviations

$$\hat{E}(T,V,\mu) = \tilde{E}(V;\{\hat{\xi}_k(T,V,\mu)\}) \ , \qquad \text{(IV.7)}$$

$$\hat{S}(T,V,\mu) = \tilde{S}(V;\{\hat{\xi}_k(T,V,\mu)\}) \ , \qquad \text{(IV.8)}$$

$$\hat{N}(T,V,\mu) = \tilde{N}(V;\{\hat{\xi}_k(T,V,\mu)\}) \ , \qquad \text{(IV.9)}$$

and

$$\hat{Y}(T,V,\mu) = \tilde{Y}(T,V,\mu;\{\hat{\xi}_k(T,V,\mu)\}) \ , \qquad \text{(IV.10)}$$

we have obviously

† It is not assumed that the $\hat{\xi}_k(T,V,\mu)$ are unique. If several such sets exist, one of them is selected and denoted by $\hat{\xi}_k(T,V,\mu)$. The results which follow are independent of this selection, provided it is performed in a way in which no artificial discontinuities are introduced.

$$\hat{Y}(T,V,\mu) = \hat{E}(T,V,\mu) - T\,\hat{S}(T,V,\mu) - \mu\,\hat{N}(T,V,\mu) \quad . \tag{IV.11}$$

Furthermore, we obtain

$$\frac{\partial \hat{Y}}{\partial T} = -\hat{S}(T,V,\mu) + \sum_{i \in K} \left[\frac{\partial \tilde{Y}}{\partial \xi_i}\right]_{\{\xi_k\}=\{\hat{\xi}_k(T,V,\mu)\}} \frac{\partial \hat{\xi}_i}{\partial T} \tag{IV.12}$$

and

$$\frac{\partial \hat{Y}}{\partial \mu} = -\hat{N}(T,V,\mu) + \sum_{i \in K} \left[\frac{\partial \tilde{Y}}{\partial \xi_i}\right]_{\{\xi_k\}=\{\hat{\xi}_k(T,V,\mu)\}} \frac{\partial \hat{\xi}_i}{\partial \mu} \quad . \tag{IV.13}$$

Because of (IV.5), the terms in (IV.12) and (IV.13) which stem from the T- and μ-dependence of the $\hat{\xi}_k(T,V,\mu)$ vanish, and we are left with

$$\frac{\partial \hat{Y}}{\partial T} = -\hat{S}(T,V,\mu) \tag{IV.14}$$

and

$$\frac{\partial \hat{Y}}{\partial \mu} = -\hat{N}(T,V,\mu) \quad . \tag{IV.15}$$

Thus our approximation $\hat{Y}(T,V,\mu)$ for the grand potential, together with the approximation $\hat{W}(T,V,\mu) = \tilde{W}(V;\{\hat{\xi}_k(T,V,\mu)\}$ for the equilibrium density matrix, satisfies the self-consistency criterion established in Sec. II.

To tackle the question of stability we notice that, because of (IV.4) and the stationarity condition (IV.5), we have, at the point $\{\xi_k\} = \{\hat{\xi}_k(T,V,\mu)\}$ in $\{\xi_k\}$-space,

$$\frac{\partial \tilde{E}}{\partial \xi_i} - T\frac{\partial \tilde{S}}{\partial \xi_i} - \mu\frac{\partial \tilde{N}}{\partial \xi_i} = 0 \text{ for all } i \in K \quad . \tag{IV.16}$$

This relation is an identity in T and μ, which we are therefore allowed to differentiate with respect to T and μ. Differentiation with respect to T gives

$$\frac{\partial \tilde{S}}{\partial \xi_i} = \sum_{j \in K} \frac{\partial^2 \tilde{Y}}{\partial \xi_i \partial \xi_j} \frac{\partial \hat{\xi}_j}{\partial T} \quad , \tag{IV.17}$$

and differentiation with respect to μ

$$\frac{\partial \tilde{N}}{\partial \xi_i} = \sum_{j \in K} \frac{\partial^2 \tilde{Y}}{\partial \xi_i \partial \xi_j} \frac{\partial \hat{\xi}_j}{\partial \mu} \quad . \tag{IV.18}$$

We now multiply (IV.17) by x and (IV.18) by y--where x and y are arbitrary real variables--and add the results; this yields

$$x\frac{\partial \tilde{S}}{\partial \xi_i} + y\frac{\partial \tilde{N}}{\partial \xi_i} = \sum_{j \in K} \left\{ \frac{\partial^2 \tilde{Y}}{\partial \xi_i \partial \xi_j} \left(x\frac{\partial \hat{\xi}_j}{\partial T} + y\frac{\partial \hat{\xi}_j}{\partial \mu} \right) \right\} \quad . \tag{IV.19}$$

Multiplying (IV.19) by $\left(x\frac{\partial \hat{\xi}_i}{\partial T} + y\frac{\partial \hat{\xi}_i}{\partial \mu} \right)$ and summing over i, we obtain finally

$$\sum_{i \in K} \left\{ \left(x\frac{\partial \hat{\xi}_i}{\partial T} + y\frac{\partial \hat{\xi}_i}{\partial \mu} \right) \left(x\frac{\partial \tilde{S}}{\partial \xi_i} + y\frac{\partial \tilde{N}}{\partial \xi_i} \right) \right\}$$

$$= \sum_{i \in K} \sum_{j \in K} \left\{ \left(x\frac{\partial \hat{\xi}_i}{\partial T} + y\frac{\partial \hat{\xi}_i}{\partial \mu} \right) \frac{\partial^2 \tilde{Y}}{\partial \xi_i \partial \xi_j} \left(x\frac{\partial \hat{\xi}_j}{\partial T} + y\frac{\partial \hat{\xi}_j}{\partial \mu} \right) \right\} \quad . \tag{IV.20}$$

The right-hand side of (IV.20) is a quadratic form in the variables $\left(x\frac{\partial \hat{\xi}_i}{\partial T} + y\frac{\partial \hat{\xi}_i}{\partial \mu} \right)$, the kernel being formed by the second order derivatives of $\tilde{Y}$ with respect to the ξ_i. According to (IV.6), this kernel is positive semi-definite at the point $\{\xi_k\} = \{\hat{\xi}_k(T,V,\mu)\}$ to which the present calculation refers. It follows that

$$\sum_{i \in K} \left\{ \left(x\frac{\partial \hat{\xi}_i}{\partial T} + y\frac{\partial \hat{\xi}_i}{\partial \mu} \right) \left(x\frac{\partial \tilde{S}}{\partial \xi_i} + y\frac{\partial \tilde{N}}{\partial \xi_i} \right) \right\} \geq 0 \quad , \tag{IV.21}$$

which can be re-written as

$$x^2 \frac{\partial \hat{S}}{\partial T} + xy\frac{\partial \hat{N}}{\partial T} + yx\frac{\partial \hat{S}}{\partial \mu} + y^2 \frac{\partial \hat{N}}{\partial \mu} \geq 0 \quad . \tag{IV.22}$$

Because of (IV.14) and (IV.15), this is equivalent to

$$(x,y)\begin{pmatrix} -\frac{\partial^2\hat{Y}}{\partial T^2} & -\frac{\partial^2\hat{Y}}{\partial T\partial\mu} \\ -\frac{\partial^2\hat{Y}}{\partial\mu\partial T} & -\frac{\partial^2\hat{Y}}{\partial\mu^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \geq 0 . \qquad \text{(IV.23)}$$

Since x and y were completely arbitrary real variables, it follows that

$$-\hat{\psi}'' \equiv \begin{pmatrix} -\frac{\partial^2\hat{Y}}{\partial T^2} & -\frac{\partial^2\hat{Y}}{\partial T\partial\mu} \\ -\frac{\partial^2\hat{Y}}{\partial\mu\partial T} & -\frac{\partial^2\hat{Y}}{\partial\mu^2} \end{pmatrix} \geq 0 . \qquad \text{(IV.24)}$$

This result, which I shall refer to as "weak stability," is almost equivalent to our original condition (II.21), but falls short of (II.21) in that it does not exclude the possibility of the lowest eigenvalue of the matrix $-\hat{\psi}''$ being zero. According to (IV.20), the latter is the case if and only if there exist two real numbers x and y with $x^2 + y^2 \neq 0$ such that the vector $\left(x\frac{\partial\hat{\xi}_i}{\partial T} + y\frac{\partial\xi_i}{\partial\mu}\right)$ lies in the null-space of the linear transformation

$$-\tilde{\psi}'' = \left(\left[\frac{\partial^2\tilde{Y}}{\partial\xi_i\partial\xi_j}\right]_{\{\xi_k\}=\{\hat{\xi}_k(T,V,\mu)\}}\right) . \qquad \text{(IV.25)}$$

Because of (IV.19), this condition is equivalent to the condition that the vectors

$$\left(\left[\frac{\partial\tilde{S}}{\partial\xi_i}\right]_{\{\xi_k\}=\{\hat{\xi}_k(T,V,\mu)\}}\right) \qquad \text{(IV.26)}$$

and

$$\left(\left[\frac{\partial\tilde{N}}{\partial\xi_i}\right]_{\{\xi_k\}=\{\hat{\xi}_k(T,V,\mu)\}}\right) \qquad \text{(IV.27)}$$

are co-linear in $\{\xi_k\}$-space.

This condition is certainly fulfilled if $\mathcal{K}$ and $\mathfrak{n}$ are linearly dependent on $\mathfrak{f}, \varepsilon$, as is seen immediately by invoking, once more, the stationarity property (IV.16). This comes as no surprise, since we have already seen that in this case even the true grand potential shows only weak stability. We have excluded this case from the

discussion because a linear dependence between $\mathcal{K}$ and $\mathcal{H}$ is not consistent with the use of T and μ as two separate independent thermodynamic variables and is thus incompatible with the very concept of a grand potential.

The condition derived above is also satisfied whenever the number of independent variational parameters is less than two, since in that case the $\{\xi_k\}$-space is, at most, one-dimensional, and in a one-dimensional space any two vectors are linearly dependent.

However, in serious-minded applications, such as those mentioned at the end of Sec. III, the dimensionality of $\{\xi_k\}$-space is always either very large or denumerably infinite, and it is not anticipated that the two vectors (IV.26) and (IV.27) will ever become colinear in those cases. The most likely situation where one might expect it to occur is at the transition point (or on the transition line) of a phase transition described by a molecular-field-theory type of approximation. There, the lowest eigenvalue of the matrix $-\psi''$ does indeed vanish, and the null-space of the linear transformation (IV.25) is, as a consequence, of higher than zero dimensionality. The relevance of this mathematical feature in phase transition theories of this type seems to me to deserve further elucidation; however, contrary to what one might be tempted to conjecture it does not seem to have the result of making the lowest eigenvalue of $-\hat{\psi}''$ vanish.

We have thus established the following results:

(i) The minimization procedure described above leads to an approximation $\hat{Y}(T,V,\mu)$ for the grand potential and to an approximation $\hat{W}(T,V,\mu)$ for the equilibrium density matrix which satisfy the self-consistency criterion (II.4), (II.5).

(ii) The approximate grand potential $\hat{Y}(T,V,\mu)$ also has the property of weak stability in the sense of (IV.24).

(iii) For $\hat{Y}(T,V,\mu)$ to satisfy the stability criterion (II.21) it is necessary and sufficient that, at the point $\{\xi_k\} = \{\hat{\xi}_k(T,V,\mu)\}$, the vectors $\left(\frac{\partial S}{\partial \xi_i}\right)$ and $\left(\frac{\partial N}{\partial \xi_i}\right)$ are linearly independent.

V. The Role of the Duhem-Gibbs Relation

From the grand potential $Y(T,V,\mu)$, one can calculate the entropy S, the pressure p, and the particle number N by

$$S(T,V,\mu) = -\frac{\partial Y}{\partial T} \quad , \tag{V.1}$$

$$p(T,V,\mu) = -\frac{\partial Y}{\partial V} \quad , \tag{V.2}$$

and

$$N(T,V,\mu) = -\frac{\partial Y}{\partial \mu} \quad . \tag{V.3}$$

With the help of these functions, the following expressions are found for the specific heat at constant volume C_V, the isobaric specific heat C_p, the isothermal compressibility κ_T, and the adiabatic compressibility κ_S:

$$C_V \equiv \left(\frac{\partial E}{\partial T}\right)_{V,N} = T \frac{\frac{\partial(S,N)}{\partial(T,\mu)}}{\left(\frac{\partial N}{\partial \mu}\right)_{T,V}} , \tag{V.4}$$

$$C_p \equiv \left(\frac{\partial H}{\partial T}\right)_{p,N} = T \frac{\frac{\partial(S,p,N)}{\partial(T,V,\mu)}}{\frac{\partial(p,N)}{\partial(V,\mu)}} , \tag{V.5}$$

$$\kappa_T \equiv -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{T,N} = -\frac{1}{V} \frac{\left(\frac{\partial N}{\partial \mu}\right)_{T,V}}{\frac{\partial(p,N)}{\partial(V,\mu)}} , \tag{V.6}$$

and

$$\kappa_S \equiv -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{S,N} = -\frac{1}{V} \frac{\frac{\partial(S,N)}{\partial(T,\mu)}}{\frac{\partial(S,p,N)}{\partial(T,V,\mu)}} . \tag{V.7}$$

For a system to be thermally and mechanically stable, these four quantities have to be non-negative. From (V.4) to (V.7) it is seen that this is indeed the case provided that the functional determinant

$$\frac{\partial(S,-p,N)}{\partial(T,V,\mu)} \tag{V.8}$$

and all its principal minors are non-negative. How is this version of the stability condition related to our original stability condition (II.6)?

The answer to this question lies in the fact that V is the only extensive parameter among the independent thermodynamic variables in Y. In the limit where the system is large and where finite-size effects such as surface contributions to the energy can therefore be expected to become negligible, Y itself should be an extensive quantity and thus a homogeneous function of V of degree 1. This means that in the limit $N \to \infty$, $V \to \infty$, $\frac{N}{V} = \rho = \text{const}$, the true grand potential satisfies the so-called Duhem-Gibbs relation

$$\left(\frac{\partial Y}{\partial V}\right)_{T,\mu} = \frac{Y}{V} . \tag{V.9}$$

The validity of (V.9) can indeed be proved rigorously from the definition of Y, provided that suitable restrictions are imposed on the nature of the forces by which the particles interact and on the shapes of the areas in space which the system is assumed to occupy in the successive stages of the $V \to \infty$ limiting process.[15),16)]

From (V.9) it follows immediately that

$$\left(\frac{\partial p}{\partial V}\right)_{T,\mu} = 0 \quad , \tag{V.10}$$

$$\left(\frac{\partial p}{\partial T}\right)_{V,\mu} = \left(\frac{\partial S}{\partial V}\right)_{T,\mu} = \frac{S}{V} \quad , \tag{V.11}$$

and

$$\left(\frac{\partial p}{\partial \mu}\right)_{T,V} = \left(\frac{\partial N}{\partial V}\right)_{T,\mu} = \frac{N}{V} \quad . \tag{V.12}$$

With the help of these relations, the two stability conditions mentioned above are easily seen to be indeed equivalent, and the expressions (V.4) to (V.7) can be written in terms of the derivatives of Y with respect to T and μ alone. One obtains

$$C_V = T \frac{\det(-\psi'')}{(0,1)(-\psi'')\begin{pmatrix}0\\1\end{pmatrix}} \quad , \tag{V.13}$$

$$C_p = \frac{T}{N^2}(N,-S)(-\psi'')\begin{pmatrix}N\\-S\end{pmatrix} , \tag{V.14}$$

$$\kappa_T = \frac{V}{N^2}(0,1)(-\psi'')\begin{pmatrix}0\\1\end{pmatrix} , \tag{V.15}$$

and

$$\kappa_S = V \frac{\det(-\psi'')}{(N,-S)(-\psi'')\begin{pmatrix}N\\-S\end{pmatrix}} \quad , \tag{V.16}$$

where $-\psi''$ denotes the 2 x 2-matrix on the left-hand side of (II.6). The fact that $-\psi''$ is a positive semi-definite matrix thus ensures that C_V, C_p, κ_T, and κ_S are all non-negative quantities.†

† As long as the system is finite, the matrix $-\psi''$ is in fact, according to (II.6), positive definite and would therefore, if inserted in (V.13) to (V.16), lead to positive (as opposed to merely non-negative) values for C_V, C_p, κ_T, and κ_S. Since, however, in deriving (V.13) to (V.16)

These considerations can at once be carried over to any approximation $\hat{Y}(T,V,\mu)$ for the grand potential which is obtained by the minimization procedure described in the previous sections: If $\hat{Y}(T,V,\mu)$ satisfies the Duhem-Gibbs relation (V.9), then the corresponding approximations $\hat{C}_V$, $\hat{C}_p$, $\hat{\kappa}_T$ and $\hat{\kappa}_S$ can be written in the form (V.13) to (V.16), and the stability result (IV.24) for the matrix $-\psi''$ guarantees that these four quantities are non-negative. We note this as result

(iv) If the approximate grand potential $\hat{Y}(T,V,\mu)$ satisfies the Duhem-Gibbs relation

$$\left(\frac{\partial \hat{Y}}{\partial V}\right)_{T,\mu} = \frac{\hat{Y}}{V} \quad , \tag{V.17}$$

then $\hat{Y}(T,V,\mu)$ describes, on account of its weak stability property (IV.24), a thermally and mechanically stable system in the sense that it satisfies the inequalities

$$\hat{C}_V \geq 0, \quad \hat{C}_p \geq 0, \quad \hat{\kappa}_T \geq 0, \quad \hat{\kappa}_S \geq 0. \tag{V.18}$$

This result shows--to give only one example--that a minimizing approximation $\hat{Y}(T,V,\mu)$ can never lead to the so-called "van der Waals loops" for the isotherms in the pressure vs. volume diagram which are, e.g., displayed by the equations of state of the van der Waals or Dietetrici type. For the existence of such loops implies the occurrence of a region of negative isothermal compressibility, which is excluded by our result (V.18). That this is correct for the true isotherms of any thermodynamic system was first explicitly stated and proved by van Hove.[17] Our theorem (V.18) also generalizes a result first obtained by Mermin[7] according to which the minimizing solution of the Hartree-Fock equations for a many-fermion system interacting via two-body forces always yields a non-negative isothermal compressibility. That a variational procedure of the kind considered here will always yield a non-negative $\hat{C}_V$ was, for the case where the corresponding variational principle for the Helmholtz free energy is used, first shown by Wagner and Koppe.[5]

---

we have made use of the Duhem-Gibbs relation (V.9), which holds rigorously only in the limit $V \to \infty$, we have to interpret (V.13) to (V.16) as formulae which are valid only if, after suitable normalisation, the limit $V \to \infty$ is taken, and as a consequence only the non-negativity of the quantities C_V, C_p, κ_T and κ_S can be inferred.

In concluding this section, it may be remarked that the approximations $\hat{C}_V$, $\hat{C}_p$, $\hat{\kappa}_T$ and $\hat{\kappa}_S$ also satisfy the two well-known thermomechanical relationships

$$\frac{\hat{C}_p}{\hat{C}_V} = \frac{\hat{\kappa}_T}{\hat{\kappa}_S} \tag{V.19}$$

and

$$\hat{C}_p - \hat{C}_V = TV \frac{\hat{\alpha}_p^2}{\hat{\kappa}_T} \quad , \tag{V.20}$$

where $\hat{\alpha}_p$ denotes the corresponding approximation for the isobaric thermal expansion coefficient

$$\alpha_p \equiv \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_{p,N} \quad . \tag{V.21}$$

As a consequence of (V.19) and (V.20), the result (V.18) can be restated as

$$\hat{C}_p \geq \hat{C}_V \geq 0, \qquad \hat{\kappa}_T \geq \hat{\kappa}_S \geq 0 \; . \tag{V.22}$$

VI. Magnetic Systems

While the grand canonical ensemble and the grand potential $Y(T,V,\mu)$ are usually the appropriate concepts for the thermodynamic description of a many-particle system, they are no longer adequate for the problem of a system of interacting spins. In the case of spin systems, one is interested in the influence of an external magnetic field $\underset{\sim}{h}$ on the magnetization $\underset{\sim}{M}$, so that these two quantities appear as an additional pair of conjugate thermodynamic variables. On the other hand, the positions of the spins are frequently assumed to be fixed in space to the sites of a regular lattice, in which case the variables volume and pressure become redundant. The most convenient thermodynamic potential under these circumstances is, as a rule, the so-called Gibbs free energy

$$G(T,\underset{\sim}{h},N) = - k_B T \log \operatorname{Tr}_{\substack{\rho \\ \gamma_N}} \sigma \left\{ \exp \left[-\frac{1}{k_B T} (\mathcal{H} - \underset{\sim}{h} \cdot \underset{\sim}{\mathfrak{m}}) \right] \right\} \quad , \tag{VI.1}$$

where $\mathcal{H}$ denotes the spin-spin interaction Hamiltonian. The quantity $\underset{\sim}{\mathfrak{m}}$ is the total magnetic moment operator and is defined by

$$\underset{\sim}{\mathfrak{m}} = \sum_i m_i \underset{\sim}{\sigma}_i \quad , \tag{VI.2}$$

where m_i denotes the strength of the magnetic moment and $\underset{\sim}{\sigma}_i$ the Pauli spin operator for the spin attached to the i-th lattice site. The summation in (VI.2) ranges over a finite (usually rectangular) one-, two- or three-dimensional array containing N lattice sites; the Hilbert space $\mathcal{H}_N^{\sigma}$ is the $(2\sigma + 1)^N$-dimensional linear vector space consisting of all possible states of the N-spin system in this array.

In this formulation, the particle number rather than a chemical potential is used as the third independent variable. The reason is that, for large systems, at least one of the independent thermodynamic variables must be an extensive quantity, since otherwise the system is not fully specified.[6)]

From $G(T,\underset{\sim}{h},N)$ one obtains the entropy S and the average magnetisation $\underset{\sim}{M}$ as

$$S(T,\underset{\sim}{h},N) = -\frac{\partial G}{\partial T} \tag{VI.3}$$

and

$$\underset{\sim}{M}(T,\underset{\sim}{h},N) = -\frac{\partial G}{\partial \underset{\sim}{h}} \quad . \tag{VI.4}$$

The second-order static response functions which one is mainly interested in are now: the specific heat at constant magnetisation C_M, the specific heat at constant magnetic field C_h, the isothermal magnetic susceptibility χ_T and the adiabatic magnetic susceptibility χ_S. These quantities are given by

$$C_{\underset{\sim}{M}} = \left(\frac{\partial E}{\partial T}\right)_{\underset{\sim}{M},N} = T\,\frac{\dfrac{\partial(S,\underset{\sim}{M})}{\partial(T,\underset{\sim}{h})}}{\left(\dfrac{\partial \underset{\sim}{M}}{\partial \underset{\sim}{h}}\right)} = T\,\frac{\det(-\mathcal{G}'')}{(0,1)(-\mathcal{G}'')\begin{pmatrix}0\\1\end{pmatrix}}, \tag{VI.5}$$

$$C_{\underset{\sim}{h}} = \left(\frac{\partial H}{\partial T}\right)_{\underset{\sim}{h},N} = T\left(\frac{\partial S}{\partial T}\right)_{\underset{\sim}{h},N} = T\,(1,0)(-\mathcal{G}'')\begin{pmatrix}1\\0\end{pmatrix}, \tag{VI.6}$$

$$\chi_T = \left(\frac{\partial \underset{\sim}{M}}{\partial \underset{\sim}{h}}\right)_{T,N} = (0,1)(-\mathcal{G}'')\begin{pmatrix}0\\1\end{pmatrix} \quad , \tag{VI.7}$$

and

$$\chi_S = \left(\frac{\partial \underset{\sim}{M}}{\partial \underset{\sim}{h}}\right)_{S,N} = \frac{\dfrac{\partial(S,\underset{\sim}{M})}{\partial(T,\underset{\sim}{h})}}{\left(\dfrac{\partial S}{\partial T}\right)_{\underset{\sim}{h},N}} = \frac{\det(-\mathcal{G}'')}{(1,0)(-\mathcal{G}'')\begin{pmatrix}1\\0\end{pmatrix}} \quad . \tag{VI.8}$$

Here, $-\mathcal{G}''$ denotes the matrix of second-order derivatives

$$-\mathcal{G}'' \equiv \begin{pmatrix} -\frac{\partial^2 G}{\partial T^2} & -\frac{\partial^2 G}{\partial T \partial \underset{\sim}{h}} \\ -\frac{\partial^2 G}{\partial \underset{\sim}{h} \partial T} & -\frac{\partial^2 G}{\partial \underset{\sim}{h}^2} \end{pmatrix} \tag{VI.9}$$

which is related to the covariance matrix $\mathcal{H}(\mathcal{K},\underset{\sim}{\mathcal{M}})$ in a manner entirely analogous to (II.8), and is thus a positive definite matrix in the case of the true G.

The quantity $G(T,\underset{\sim}{h},N)$ satisfies the inequality

$$G(T,\underset{\sim}{h},N) \leq \mathrm{Tr}_{\mathcal{H}_N^\sigma} \left\{ \mathcal{K}\tilde{W} - \underset{\sim}{h}\cdot\underset{\sim}{\mathcal{M}}\tilde{W} + k_B T \tilde{W} \log \tilde{W} \right\} \tag{VI.10}$$

for all Hermitian matrices $\hat{W}$ acting on $\mathcal{H}_N^\sigma$ which have the property

$$\tilde{W} \geq 0, \quad \mathrm{Tr}_{\mathcal{H}_N^\sigma} \tilde{W} = 1 . \tag{VI.11}$$

Inequality (VI.10) can be used as a variational principle for the approximate calculation of $G(T,\underset{\sim}{h},N)$ in exactly the same way as (I.3) in the case of $Y(T,V,\mu)$, and the results derived in the previous sections about the self-consistency and stability of such an approximation apply equally well to the present case, with all their consequences such as the validity of thermodynamic identities and the uniqueness of the Legendre transformations with respect to the intensive variables T and $\underset{\sim}{h}$. It is to be noticed from (VI.5) to (VI.8) that the second-order response functions are readily expressible in terms of $-\mathcal{G}''$ without the help of a Duhem-Gibbs relation, and can thus be seen to be non-negative for the approximate Gibbs free energy $\hat{G}(T,\underset{\sim}{h},N)$ whether or not the latter is known to satisfy a Duhem-Gibbs relation.

In some respects, the mathematics of the many-spin problem is somewhat simpler than that of the many-particle system, owing to the fact that the Hilbert space in (VI.1) is finite-dimensional. This removes all the existence and convergence difficulties which are present in the earlier case; it also means that the number of real-valued variational parameters in $\tilde{W}$ which can be varied independently is finite, being less than $(2\sigma+1)^{2N}$.

On the other hand, the fact that one has to deal with a discrete lattice structure and, at least by way of estimate, with the boundary conditions for the array of N spins introduces combinatorial problems which are well-known to be formidable even in the case of seemingly simple models.

The variational principle (VI.10) has recently been applied to cluster approximations for the Heisenberg ferromagnet[18] and to the investigation of temperature-dependent crystal field effects in terbium.[19] As an elementary exercise it permits to show that the free energy of the Ising model is higher than that of the Heisenberg model if the exchange coupling constants are the same.[20]

VII. Concluding Remarks

(i) The manner in which the Duhem-Gibbs relation is used in Sec. V needs to be somewhat modified in the case of a Bose system undergoing Bose-Einstein condensation. This is seen most clearly in the case of an ideal Bose gas of particles of mass m, which has the grand potential

$$Y(T,V,\mu) = k_B T \left\{ \log\,(1 - e^{-\gamma}) - \frac{V}{\lambda_T^3}\, g_{5/2}(e^{-\gamma}) \right\} \tag{VII.1}$$

with

$$\gamma = -\frac{\mu}{k_B T} > 0, \qquad \lambda_T = \sqrt{\frac{2\pi\hbar^2}{m\, k_B T}}\ , \tag{VII.2}$$

and

$$g_s(x) = \sum_{\nu=1}^{\infty} \frac{x^\nu}{\nu^s}\ . \tag{VII.3}$$

Although the grand potential (VII.1) satisfies the Duhem-Gibbs relation (V.9) in the limit $V \to \infty$, this relation yields the wrong result for the partial derivatives of Y with respect to T and μ in the region of condensation where $\mu = 0\left(\frac{1}{V}\right)$, because in this region the convergence of $\frac{Y}{V}$ as $V \to \infty$ is so slow that the limit $V \to \infty$ can no longer be interchanged with the differentiation with respect to T or μ. The correct calculation is, of course, elementary and yields the result that, while the formula (V.10) remains unchanged, (V.11) and (V.12) have to be replaced by

$$\left(\frac{\partial p}{\partial T}\right)_{V,\mu} = \left(\frac{\partial N}{\partial V}\right)_{T,\mu} = \frac{S - S_o}{V} \tag{VII.4}$$

and

$$\left(\frac{\partial p}{\partial \mu}\right)_{T,V} = \left(\frac{\partial S}{\partial V}\right)_{T,\mu} = \frac{N - N_o}{V}\ , \tag{VII.5}$$

where S_o and N_o denote the entropy and the number of particles in the condensed phase, respectively, and are given by

$$S_o = k_B \left\{ \frac{\gamma e^{-\gamma}}{1 - e^{-\gamma}} - \log(1 - e^{-\gamma}) \right\} . \qquad \text{(VII.6)}$$

and

$$N_o = \frac{e^{-\gamma}}{1 - e^{-\gamma}} . \qquad \text{(VII.7)}$$

Thus, everywhere in (V.13) to (V.16) S and N have to be replaced by $S - S_o$ and $N - N_o$, respectively. The conclusions which we drew from (V.13) to (V.16) are not affected by this correction. However, in the case of an interacting Bose system, where the calculation can no longer be performed explicitly, this point should be given careful attention.

(ii) If the Duhem-Gibbs relation (V.9) ceases to hold altogether --as is the case, e.g., for the thermodynamics of small systems[30]-- the stability theorem of Sec. IV can no longer be used to establish the non-negativity of the response functions $\hat{C}_V$, $\hat{C}_p$, $\hat{\kappa}_T$ and $\hat{\kappa}_S$. In this case, it becomes necessary to work right from the beginning with a different set of independent thermodynamic variables, viz. T, p and N, and to choose the corresponding statistical equilibrium ensemble ("pressure ensemble"). In this formulation, the appropriate thermodynamic potential is the free enthalpy, or Gibbs free energy, G(T,p,N) defined by

$$G(T,p,N) = - k_B T \log \left\{ \frac{p}{k_B T} \int_o^\infty Z(T,V,N)\, e^{-\frac{pV}{k_B T}} dV \right\} , \qquad \text{(VII.8)}$$

where Z(T,V,N) is the partition function for the canonical ensemble, i.e.

$$Z(T,V,N) = \mathrm{Tr}_{\substack{\mathcal{E} \\ \mathcal{N}}} \left\{ \exp\left(- \frac{1}{k_B T} \mathcal{K} \right) \right\} . \qquad \text{(VII.9)}$$

From G(T,p,N), one obtains the thermodynamic functions

$$S(T,p,N) = - \left(\frac{\partial G}{\partial T} \right)_{p,N} \qquad \text{(VII.10)}$$

and

$$V(T,p,N) = \left(\frac{\partial G}{\partial p}\right)_{T,N} , \qquad \text{(VII.11)}$$

and from these, in turn, the second-order static response functions

$$C_V = T\frac{\frac{\partial(S,V)}{\partial(T,p)}}{\left(\frac{\partial V}{\partial p}\right)_{T,N}} = T\frac{\det(-\mathcal{G}'')}{(0,1)(-\mathcal{G}'')\binom{0}{1}} , \qquad \text{(VII.12)}$$

$$C_p = T\left(\frac{\partial S}{\partial T}\right)_{p,N} = T\,(1,0)(-\mathcal{G}'')\binom{1}{0} , \qquad \text{(VII.13)}$$

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{T,N} = \frac{1}{V}\,(0,1)(-\mathcal{G}'')\binom{0}{1} , \qquad \text{(VII.14)}$$

and

$$\kappa_S = -\frac{1}{V}\,\frac{\frac{\partial(S,V)}{\partial(T,p)}}{\left(\frac{\partial S}{\partial T}\right)_{p,N}} = \frac{1}{V}\,\frac{\det(-\mathcal{G}'')}{(1,0)(-\mathcal{G}'')\binom{1}{0}} . \qquad \text{(VII.15)}$$

Thus the stability condition

$$-\mathcal{G}'' \equiv \begin{pmatrix} -\frac{\partial^2 G}{\partial T^2} & -\frac{\partial^2 G}{\partial T \partial \mu} \\ -\frac{\partial^2 G}{\partial \mu \partial T} & -\frac{\partial^2 G}{\partial \mu^2} \end{pmatrix} > 0 \qquad \text{(VII.16)}$$

alone guarantees that the four quantities (VII.12) to (VII.15) are non-negative, and it is not necessary to invoke a Duhem-Gibbs relation.

However, as far as practical applications are concerned the prospect of having to deal with the (T,p,N)-ensemble is, except in the most elementary cases, rather unattractive for the following reasons. Firstly, the specification of an exact particle number N leads to unwelcome restrictions which result in considerable combinatorial difficulties. Secondly, the variational principle for G which corresponds to the variational principle (I.3) for Y involves the notion of a "volume operator" $\mathcal{V}$, and the introduction and handling of such an operator, although in principle possible, is technically awkward. One of the complications that arise in this context stems from the fact that the canonical partition function depends, as long as the system is

finite, not only on the volume, but on the exact shape of the region in space occupied by the system,† and that, consequently, the shape dependence can be unambiguously expressed in terms of a volume dependence only if the discussion is restricted to regions of some very simple shape such as that of a cube or a sphere.‡

(iii) As was pointed out in the introduction, the derivation of the self-consistency and stability theorem given here can easily be formulated for arbitrary thermodynamic potentials. Such an arbitrary thermodynamic potential Φ is a functionof r - 1 intensive variables α_i (i = 1,...,r-1) and of n-r+1 extensive variables A_j (j=r,...,n), where the α_i and A_j are chosen from a complete set of n pairs of mutually conjugate thermodynamic variables in such a way that each pair is represented exactly once. For the system to be completely specified, at least one of the variables chosen must be an extensive quantity, i.e. one must have $1 \leq r \leq n$. It was shown by Münster[6] that for such a thermodynamic potential there exist two sets of stability conditions, one of which is expressed in terms of the intensive, the other in terms of all but one of the extensive variables. These two conditions can be written as

$$\left(-\frac{\partial^2 \Phi}{\partial \alpha_i \partial \alpha_{i'}}\right) > 0 \ , \ 1 \leq i,i' \leq r-1 \qquad \text{(VII.17)}$$

and

$$\left(\frac{\partial^2 \Phi}{\partial A_j \partial A_{j'}}\right) > 0, \ r \leq j,j' \leq n-1 \ . \qquad \text{(VII.18)}$$

If the methods used in the previous sections for the case of Y are applied to Φ, it is found that only the first of these two stability conditions can be shown to hold for the approximate thermodynamic potential $\hat{\Phi}$, while nothing at all can be said about the second. The reason is that it is not possible to choose the set $\tilde{\mathfrak{W}}$ of trial density matrices independently of the extensive variables A_j, since the necessary conditions

$$\tilde{W} \in \mathfrak{B}[\mathfrak{H}], \ \tilde{W}^+ = \tilde{W}, \ \tilde{W} \geq 0, \ \mathrm{Tr}_{\mathfrak{H}} \ \tilde{W} = 1 \qquad \text{(VII.19)}$$

† For this point, the reader may consult the discussions given by Fisher[25] and by Robinson, Ruelle, and Lanford,[31] where further references are to be found.

‡ Another difficulty in connection with the concept of pressure ensembles has been discussed by Münster[32] and by Sack,[33] where further references are given.

for the elements $\tilde{W}$ of $\tilde{\mathfrak{W}}$ are formulated in terms of a Hilbert space which depends on the values of the extensive variables A_j.

Thus, the interesting question whether $\hat{\Phi}$ also shows stability with respect to its extensive variables in the sense of inequality (VII.18) is left entirely open. The only consolation in this situation comes from the fact that, as a rule, the thermodynamic potentials which contain only one extensive variable, i.e. the ones with $r = n$, are technically the simplest ones to handle and are therefore used with overwhelming preference. For these potentials, the problem of stability with respect to extensive variables does not arise, since for $r = n$ the condition (VII.17), eventually in conjunction with the appropriate Duhem-Gibbs relation, is all that is required for the discussion of questions of stability.

(iv) Another problem that remains open is connected with the "fluctuation-response theorem" (II.8), which relates the matrix $-\Psi''$ to the covariance matrix $\mathfrak{K}(\mathcal{H},\mathfrak{n})$. If the elements of the covariance matrix are calculated with the approximate equilibrium density matrix $\hat{W}$ as statistical operator, but with the true Hamiltonian $\mathcal{H}$ and the true patticle number operator $\mathfrak{n}$ as observables, then the resulting covariance matrix $\hat{\mathfrak{K}}(\mathcal{H},\mathfrak{n})$ is no longer related to $-\hat{\Psi}''$ by formula (II.8). This means that the approximate grand potential $\hat{Y}$ and the approximate equilibrium density matrix $\hat{W}$ are only consistent so far as the first moments of $\hat{W}$ are concerned, but are no longer consistent with regard to second and higher moments. As a consequence, there exists an alternative set of definitions for the approximate quantities $\hat{C}_V$, $\hat{C}_p$, $\hat{\kappa}_T$, and $\hat{\kappa}_S$, viz.

$$\hat{C}_V^{(2)} = \frac{1}{k_B T^2} \frac{\mathrm{var}^2(\mathcal{H})\,\mathrm{var}^2(\mathfrak{n}) - [\mathrm{cov}(\mathcal{H},\mathfrak{n})]^2}{\mathrm{var}^2(\mathfrak{n})}, \qquad \text{(VII.20)}$$

$$\hat{C}_p^{(2)} = \frac{1}{k_B T^2} \mathrm{var}^2\left(\mathcal{H} - \left(\frac{ST}{N} + \mu\right)\mathfrak{n}\right), \qquad \text{(VII.21)}$$

$$\hat{\kappa}_T^{(2)} = \frac{V}{N^2 T} \mathrm{var}^2(\mathfrak{n}) \qquad \text{(VII.22)}$$

and

$$\hat{\kappa}_S^{(2)} = \frac{V}{N^2 T} \frac{\mathrm{var}^2(\mathcal{H})\,\mathrm{var}^2(\mathfrak{n}) - [\mathrm{cov}(\mathcal{H},\mathfrak{n})]^2}{\mathrm{var}^2\left(\mathcal{H} - \left(\frac{ST}{N} + \mu\right)\mathfrak{n}\right)} \qquad \text{(VII.23)}$$

which, as a rule, will not yield the same results as the set (V.13) to (V.16). There remains, thus, a certain amount of ambiguity in the calculation of the approximate thermodynamic functions after all.

It is, of course, to a certain degree a matter of taste as how serious a drawback one considers this. Since neither of the two sets of definitions gives the correct values for the quantities in question, the fact that they do not agree with each other should not be overrated. If one is prepared to accept an approximation at all, one will also have to resign oneself to the fact that, in some higher order, the consistency between the derivatives of $\hat{Y}$ and the moments of $\hat{W}$ will be broken.

Still, there remains the fact that the variational procedure fails to give an unambiguous prescription for the calculation of second-order thermodynamic quantities, and any decision in favour of one set of definitions rather than the other can only be regarded as arbitrary. What is more, the various forms of the fluctuation-response theorem, which link the specific heat at constant volume with the fluctuations of the energy in the case of the canonical ensemble, the isothermal compressibility with the fluctuations of the particle number in the case of the grand canonical ensemble, the isothermal magnetic susceptibility with the fluctuations of the total induced magnetic moment in the magnetic case, etc., are important theorems of great physical significance. They are of the nature of sum rules, and as always in physics one would like to have an approximation which satisfies the sum rules demanded by the exact theory. For these reasons, it would seem worthwhile to study the question whether the variational procedure presented here can be modified in such a way as to satisfy the second-order fluctuation-response theorems.

References

1. J. W. Gibbs, Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics (Charles Scribner's Sons, New York, and Edward Arnold, London, 1902), p. 131-133.
2. M. Delbrück and G. Molière, Abhandl. d. Preuss. Akad. d. Wiss., phys.-math. Kl., Jahrg. 1936, Nr. 1, p. 1-42 (cf.p. 25).
3. A. Huber, in Concepts and Methods in Theoretical Physics--Proceedings of a Conference Held in Honour of Professor Rudolf E. Peierls' 60th Birthday (North-Holland Publishing Company, Amsterdam, 1969).
4. A. Huber, in Mathematical Methods in Solid State and Superfluid Theory--Scottish Universities' Summer School 1967, ed. by R. C. Clark and G. H. Derrick (Oliver & Boyd, Edinburgh, 1969), p. 364-392.
5. F. Wagner and H. Koppe, Z. Naturforsch. 20a, 1553 (1965).
6. A. Münster, Z. Physik 136, 179 (1953).

7. N. D. Mermin, Ann. Phys. (N.Y.) 21, 99 (1963).
8. J. G. Valatin, Nuovo Cimento, ser. X, 7, 843 (1958).
9. J. M. Blatt and T. Matsubara, Progr. Theoret. Phys. (Kyoto) 20, 553 (1958); C. J. Thompson and T. Matsubara, ibid. 29, 494 (1963).
10. J. G. Valatin and D. Butler, Nuovo Cimento, ser. X, 10, 37 (1958).
11. M. Girardeau and R. Arnowitt, Phys. Rev. 113, 755 (1959); M. Girardeau, ibid. 115, 1090 (1959); ibid. 127, 1809 (1962); J. Math. Phys. 3, 131 (1962).
12. V. V. Tolmachev, Dokl. Akad. Nauk SSSR 134, 1324 (1960); 135, 41, 825 (1960) [Engl. transl.: Soviet Physics--Doklady 5, 984, 1190, 1267 (1961)].
13. E. P. Gross, Ann. Phys. (N.Y.) 9, 292 (1960); 19, 234 (1962); 20, 44 (1962); F. M. Eger and E. P. Gross, ibid. 24, 63 (1963); J. Math. Phys. 7, 578 (1966); D. J. Amit and E. P. Gross, Ann. Phys. (N.Y.) 42, 296 (1967).
14. G. W. Kellner, Z. Physik 44, 91 (1927).
15. S. C. Wang, Phys. Rev. 31, 579 (1928).
16. E. A. Hylleraas, Z. Physik 54, 347 (1929).
17. V. Fock, Z. Physik 63, 855 (1930).
18. Yu. N. Demkov, Dokl. Akad. Nauk SSSR 89, 249 (1953).
19. L. Hulthèn, Kungl. Fysiogr. Sällsk. i Lund Förhandl. 14, Nr. 21(1944),p.257-269;Ark. Mat. Astr. Fys. 35A, Nr. 25 (1948) p. 1-14.
20. W. Kohn, Phys. Rev. 74, 1763 (1948).
21. P.-O. Löwdin, J. Mol. Spectroscopy 3, 46 (1959).
22. Yu. N. Demkov, Variational Principles in the Theory of Collisions [Engl. transl.: Pergamon Press, Oxford-London-New York-Paris, 1963].
23. B. L. Moiseiwitsch, Variational Principles (Interscience Publishers, London-New York-Sidney, 1966).
24. D. Ruelle, Helv. Phys. Acta 36, 183, 789 (1963).
25. M. E. Fisher, Arch. Rational Mech. Anal. 17, 377 (1964).
26. L. van Hove, Physica (The Hague) 15, 951 (1949).
27. J. M. Radcliffe, Phys. Rev. 165, 635 (1968).
28. M. S. S. Brooks, D. A. Goodings, and H. I. Ralph, J. Phys. C (Proc. Phys. Soc.), ser. 2, 1, 132, 1596 (1968).
29. K. Huang, Statistical Mechanics (John Wiley & Sons, Inc., New York-London, 1963), p. 223.
30. T. L. Hill, Thermodynamics of Small Systems, Pts. I and II (W. A. Benjamin, Inc., New York-Amsterdam, 1963/64).

31. D.W.Robinson and D. Ruelle, Commun. Math. Phys. 5, 288 (1967); O. E. Lanford III and D. W. Robinson, J. Math. Phys. 9, 1120 (1968); D. W. Robinson, Commun. Math. Phys. 6, 151 (1967).
32. A. Münster, Mol. Phys. 2, 1 (1959).
33. R. A. Sack, Mol. Phys. 2, 8 (1959).
34. R. M. Wilcox, J. Math. Phys. 8, 962 (1967).

COHERENT STATES AND INFRARED DIVERGENCES†

T. W. B. Kibble
Imperial College
London, England

The problem of infrared divergences appears in any quantum field theory involving particles of zero mass and spin one or larger, such as the photon or the graviton. In these lectures we shall consider primarily quantum electrodynamics, but much of our discussion will apply equally well to any other theory involving such particles.

Physically, infrared divergences appear because of the possibility that in a scattering process an infinite number of soft photons may be emitted with finite total energy.

We shall begin by reviewing the conventional formulation of quantum electrodynamics, and the methods which have been developed for handling the infrared divergences. These methods generally involve summing transition probabilities for a scattering process with the emission of all possible numbers of soft photons. They lead to finite scattering probabilities but not to finite scattering amplitudes. Later we shall develop an alternative procedure for handling this problem, in which we use an extended Hilbert space of states which can accommodate states with infinitely many soft photons, as the conventional Fock space cannot. In this space we shall show that it is possible to define a scattering operator all of whose matrix elements are finite.

I. Free Fields and Fields with External Sources

We begin by reviewing briefly the treatment of free field theories and theories involving interaction with an external unquantized source. First we list the notation to be used.

A. Notation

We use units in which Planck's constant divided by 2π and the velocity of light are set equal to unity,

$$c = \hbar = 1 \ .$$

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

Thus all dimensional quantities may be expressed in terms of a single dimensional unit, say that of length. For electromagnetic quantities we use rationalized units, in which no factors of 4π appear in Maxwell's equations. The charge on the electron e is a dimensionless number related to the fine structure constant α by

$$\alpha = \frac{e^2}{4\pi} \approx \frac{1}{137} \quad .$$

Four vectors are denoted by $x = (x^\mu)$ and three vectors by $\underset{\sim}{x} = (x^{\hat{\imath}})$. Greek indices run from 0 to 3 and Latin indices from 1 to 3. We use the summation convention according to which repetition of an index implies summation over all its possible values. Thus four-vector scalar products are written

$$a \cdot b = a_\mu b^\mu$$

where covariant vectors (with lower indices) are related to contravariant vectors (with upper indices) by

$$a_\mu = g_{\mu\nu} a^\nu \quad .$$

The only nonvanishing components of the matrix tensor $g_{\mu\nu}$ are the diagonal elements given by

$$-g_{00} = g_{11} = g_{22} = g_{33} = 1 .$$

Thus

$$a \cdot b = a^0 b_0 + \underset{\sim}{a} \cdot \underset{\sim}{b} = -a^0 b^0 + \underset{\sim}{a} \cdot \underset{\sim}{b}$$

where $\underset{\sim}{a} \cdot \underset{\sim}{b}$ is the ordinary three-vector scalar product

$$\underset{\sim}{a} \cdot \underset{\sim}{b} = a^{\hat{\imath}} b_{\hat{\imath}} = a^1 b^1 + a^2 b^2 + a^3 b^3 \quad .$$

Derivatives are denoted by

$$\partial_\mu = \frac{\partial}{\partial x^\mu} \quad .$$

We write $\partial^2 = \partial_\mu \partial^\mu$ and $\nabla^2 = \partial_{\hat{\imath}} \partial^{\hat{\imath}}$. For integrations we use the notations

$$d^4x = dx^0 \, d^3\underset{\sim}{x}, \quad d^3\underset{\sim}{x} = dx^1 dx^2 dx^3 .$$

The corresponding Dirac delta functions are

$$\delta_4(x) = \delta(x^0)\,\delta_3(\underline{x}) ,$$

$$\delta_3(\underline{x}) = \delta(x^1)\,\delta(x^2)\,\delta(x^3) .$$

If $f(x)$ is any function of the four-vector x, we define its Fourier transform $f(k)$ to be

$$f(k) = \int d^4x\, f(x)\, e^{-ik\cdot x}$$

so that the inverse relation is

$$f(x) = \int \frac{d^4k}{(2\pi)^4}\, f(k)\, e^{ik\cdot x} .$$

We note that if $f(x)$ is real, then $f(k)$ satisfies

$$f(k)^* = f(-k) .$$

We also define the step function

$$\theta(\lambda) = \begin{cases} 1 & \text{if } \lambda > 0, \\ 0 & \text{if } \lambda < 0. \end{cases}$$

B. The Electromagnetic Field

It is convenient to choose a set of independent field variables in terms of which the equations of motion are first-order differential equations. The electromagnetic field is described by a vector $A^\mu(x)$ and an antisymmetric tensor $F^{\mu\nu}(x)$, with the Lagrangian density function

$$\mathcal{L}_{em} = -\tfrac{1}{2}F^{\mu\nu}(\partial_\mu A_\nu - \partial_\nu A_\mu) + \tfrac{1}{4}F^{\mu\nu}F_{\mu\nu} + A_\mu J^\mu \qquad (I.1)$$

where for the moment $J^\mu(x)$ is a classical external current.

The corresponding field equations are

$$\partial_\mu A_\nu - \partial_\nu A_\mu = F_{\mu\nu} . \qquad (I.2)$$

and

$$\partial_\nu F^{\mu\nu} = J^\mu . \qquad (I.3)$$

They are invariant under the gauge transformation

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \lambda(x) \tag{I.4}$$

for any function $\lambda(x)$.

We note that from (I.3) and the antisymmetry of $F^{\mu\nu}$ we may deduce the continuity equation for J,

$$\partial_\mu J^\mu = 0 \ . \tag{I.5}$$

The field equations are mutually consistent only if this condition is satisfied.

The variables whose time derivatives appear in the field equations are A^k and F^{ok}. We may separate the equations into those containing time derivatives,

$$\partial_o A_k = \partial_k A_o + F_{ok},$$

$$\partial_o F^{ko} = -\partial_l F^{kl} + J^k$$

and constraint equations,

$$F_{ij} = \partial_i A_j - \partial_j A_i \tag{I.6}$$

$$\partial_k F^{ok} = J^o \ . \tag{I.7}$$

The first constraint equation determines F_{ij} in terms of the remaining variables. In a theory with mass, the second would do the same for A^o, but with zero mass it becomes a constraint on the variable F^{ok} alone.

Thus not all the variables F^{ok} can be independent dynamical variables. We may separate F^{ok} into transverse and longitudinal parts,

$$F^{ok} = F^{ok}_T - \partial^k \varphi \ ,$$

where

$$\partial_k F^{ok}_T = 0 \ .$$

Then (I.7) yields

$$\nabla^2 \varphi = -J^o,$$

which has the explicit solution

$$\varphi(x) = \int \frac{d^3y}{4\pi|\underset{\sim}{x} - \underset{\sim}{y}|} J^0(y) \Big|_{y^0=x^0} .$$

We may write this more formally as

$$\varphi = -\frac{1}{\nabla^2} J^0 . \tag{I.8}$$

Only the transverse components of $A^{\hbar}$ and $F^{0\hbar}$ are true dynamical variables. The longitudinal component of $F^{0\hbar}$ is an explicit function of J^0, while the longitudinal component of $A^{\hbar}$ is an arbitrary gauge function. The transverse parts satisfy the canonical commutation relations

$$\begin{aligned}[A_T^{\hbar}(x), F_T^{0\ell}(y)]_{x^0=y^0} &= -i[\delta^{\hbar\ell}\delta_3(\underset{\sim}{x} - \underset{\sim}{y})]_T \\ &= -i\left(\delta^{\hbar\ell} - \frac{\partial^{\hbar}\partial^{\ell}}{\nabla^2}\right)\delta_3(\underset{\sim}{x} - \underset{\sim}{y}) \\ &= -i\delta^{\hbar\ell}\delta_3(x - y) - i\partial^{\hbar}\partial^{\ell}\frac{1}{4\pi|\underset{\sim}{x} - \underset{\sim}{y}|} .\end{aligned} \tag{I.9}$$

The longitudinal component of $A^{\hbar}$ may be fixed by imposing any convenient gauge condition. For example the radiation gauge (or Coulomb gauge) is defined by the transversality condition

$$\partial_{\hbar}A^{\hbar} = 0 .$$

In this gauge A^0 becomes an explicit function of the sources

$$A^0 = \varphi = -\frac{1}{\nabla^2} J^0 . \tag{I.10}$$

It represents the classical Coulomb potential.

C. The Free Field

If $J^{\mu} = 0$, then the radiation gauge vector potential satisfies

$$\partial^2 A_{\mu}(x) = 0$$

or in momentum space

$$\hbar^2 A_{\mu}(\hbar) = 0 ,$$

as well as

$$A^0 = 0, \qquad \hbar_j A^j(\hbar) = 0 .$$

Thus $A_\mu(k)$ vanishes except for values of k such that $k^2 = 0$, or in other words $k^0 = \pm |\underset{\sim}{k}|$. To ensure reality of $A_\mu(x)$, we require

$$A_\mu(k)^* = A_\mu(-k) \ .$$

Hence $A_\mu(k)$ must have the form

$$A_\mu(k) = 2\pi\delta(k^2)[\theta(k^0)\, a_\mu(\underset{\sim}{k}) + \theta(-k^0)\, a_\mu(-\underset{\sim}{k})^*]$$

where

$$a^0(\underset{\sim}{k}) = 0, \quad k_j a^j(\underset{\sim}{k}) = 0 \ .$$

Thus

$$A_\mu(x) = A_\mu^{(+)}(x) + A_\mu^{(-)}(x)$$

where

$$A_\mu^{(+)}(x) = A_\mu^{(-)}(x)^* = \int \frac{d^4k}{(2\pi)^4}\, 2\pi\delta(k^2)\,\theta(k^0)\, a_\mu(\underset{\sim}{k})\, e^{ik\cdot x}$$

$$= \int \frac{d^3k}{(2\pi)^3\, 2|\underset{\sim}{k}|}\, a_\mu(\underset{\sim}{k})\, e^{ik\cdot x} \ , \qquad \text{(I.11)}$$

in which $k^0 = |\underset{\sim}{k}|$ in the exponent.

From the canonical commutation relations (I.9) we easily obtain

$$[a_\mu(\underset{\sim}{k}),\, a_\nu(\underset{\sim}{k}')^*] = \gamma_{\mu\nu}(\underset{\sim}{k})(2\pi)^3\, 2|\underset{\sim}{k}|\,\delta_3(\underset{\sim}{k} - \underset{\sim}{k}') \qquad \text{(I.12)}$$

where $\gamma_{\mu\nu}(\underset{\sim}{k})$ is the projector on the two-dimensional space orthogonal to $\underset{\sim}{k}$, whose only nonvanishing components are

$$\gamma_{ij}(\underset{\sim}{k}) = \delta_{ij} - \frac{k_i k_j}{\underset{\sim}{k}^2} \ . \qquad \text{(I.13)}$$

The vacuum state $|0\rangle$ is defined by

$$a_\mu(\underset{\sim}{k})|0\rangle = 0$$

for all $\underset{\sim}{k}$, or equivalently

$$A_\mu^{(+)}(x)|0\rangle = 0 \ .$$

One-photon states are obtained by acting on the vacuum with the creation operators $a_\mu(\underset{\sim}{k})^*$. The most general one-photon state is characterized by a photon wave-function $f_\mu(\underset{\sim}{k})$ satisfying $f^0 = 0$ and $k_j f^j = 0$. It may be written

$$|1,f\rangle = (a^* f)|0\rangle \tag{I.14}$$

where we have introduced the notation

$$a^* f = \int \frac{d^3 k}{(2\pi)^3 \, 2|\underset{\sim}{k}|} a^*_\mu(\underset{\sim}{k}) \, f^\mu(\underset{\sim}{k}) \ . \tag{I.15}$$

The commutation relations (I.12) yield

$$[f^* a, a^* g] = f^* g \tag{I.16}$$

and hence

$$\langle 1,f|1,g\rangle = f^* g. \tag{I.17}$$

The use of the radiation gauge implies a lack of manifest relativistic invariance. However we may overcome this difficulty by considering together a class of <u>physical</u> gauges, each of which is related to the radiation gauge by a gauge function $\lambda(k)$ which is a linear combination of the operators $A_\mu(k)$ for the <u>same</u> value of k. In any such gauge, the operators $a_\mu(\underset{\sim}{k})$ satisfy commutation relations of the form (I.12), where $\gamma_{\mu\nu}(\underset{\sim}{k})$ is some two-dimensional projector orthogonal to k. The one-photon states may be labeled by photon wavefunctions $f^\mu(\underset{\sim}{k})$ chosen to satisfy

$$\gamma_{\mu\nu}(\underset{\sim}{k}) \, f^\nu(\underset{\sim}{k}) = f_\mu(\underset{\sim}{k}) \ .$$

Two functions $f^\mu(\underset{\sim}{k})$ and $g^\mu(\underset{\sim}{k})$ describe the same state in different gauges if their difference is of the form $k^\mu h(\underset{\sim}{k})$.

Since both a_μ and f_μ are orthogonal to k_μ, the addition to either of a term proportional to k_μ does not change $a^* f$, given by (I.15). Thus the states (I.14) and their scalar products (I.17) are gauge-invariant.

Higher states in the photon Fock space are obtained by acting again with the operators $a^* f$. An n-photon state with photon wavefunctions $f_1 \ldots f_n$ is given by

$$|n,f_1 \ldots f_n\rangle = (a^* f_1) \ldots (a^* f_n)|0\rangle \ .$$

It is useful to introduce the unitary operators

$$U(f) = e^{a^* f - f^* a} \tag{I.18}$$

which by using the commutation relations may also be written in the form

$$U(f) = e^{a^* f} e^{-f^* a} e^{-\frac{1}{2} f^* f} .$$

These operators act as translation operators on the $a^\mu(\underset{\sim}{k})$; that is to say they add c-numbers to them:

$$U(f)^{-1}\, a_\mu(\underset{\sim}{k})\, U(f) = a_\mu(\underset{\sim}{k}) + f_\mu(\underset{\sim}{k}) \tag{I.19}$$

or

$$U(f)^{-1}\, (g^* a)\, U(f) = (g^* a) + (g^* f) .$$

It is useful to note that their effect on $A_\mu(x)$ is to add a classical solution of the wave equation,

$$U(f)^{-1}\, A_\mu(x)\, U(f) = A_\mu(x) + F_\mu(x)$$

where $F_\mu = F_\mu^{(+)} + F_\mu^{(-)}$ and

$$F_\mu^{(+)}(x) = F_\mu^{(-)}(x)^* = \int \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|} f_\mu(\underset{\sim}{k})\, e^{i\underset{\sim}{k}\cdot x} . \tag{I.20}$$

Acting on the vacuum state these unitary operators yield the so-called coherent states

$$\begin{aligned} |f\rangle &= U(f)|0\rangle \\ &= \sum_n \frac{1}{n!} |n, f \ldots f\rangle\, e^{-\frac{1}{2} f^* f} . \end{aligned} \tag{I.21}$$

From (I.19) it follows that these are eigenstates of $a_\mu(\underset{\sim}{k})$:

$$a_\mu(\underset{\sim}{k})|f\rangle = |f\rangle f_\mu(\underset{\sim}{k}) .$$

Also the expectation value of $A_\mu(x)$ is the corresponding classical solution,

$$\langle f|A_\mu(x)|f\rangle = F_\mu(x) .$$

It can be shown that the coherent state is the quantum state which most closely approximates the given classical state.

The scalar products of the coherent states are

$$\langle f|g\rangle = e^{f^*g - \frac{1}{2}f^*f - \frac{1}{2}g^*g} \tag{I.22}$$

so that in particular

$$|\langle f|g\rangle|^2 = e^{-(f-g)^*(f-g)} \quad . \tag{I.23}$$

They form a non-orthogonal, overcomplete family of states. That is to say they span the entire Hilbert space but are not linearly independent.

For future reference we note that the expectation value of the total photon number in a coherent state is

$$\langle f|N|f\rangle = f^*f \quad . \tag{I.24}$$

It is easy to see from (I.21) that the photon number obeys a Poisson distribution with this mean.

D. Variation of the External Current

It is often convenient to consider the effect of making small variations in the external current.

Let us recall the results obtained by comparing two quantum systems with the same set of independent dynamical variables $q(t)$ but whose time-development is described by different Hamiltonian functions $H(q)$ and $H(q) + \Delta H(q)$. Let $\alpha(q)$ be some complete set of commuting variables, functions of the variables q at one time, and denote by $|\alpha', t\rangle$ the corresponding eigenstates, defined by

$$\alpha(q(t))|\alpha', t\rangle = |\alpha', t\rangle\alpha' \quad .$$

The time-development of the system is described by the transformation functions $\langle\alpha', t|\alpha'_o, t_o\rangle$. It is easy to prove that for $t > t_o$

$$\langle\alpha', t|\alpha'_o, t_o\rangle_{H+\Delta H} = \langle\alpha', t|\, T \exp\left\{-i\int_{t_o}^{t} dt' \Delta H(q(t'))\right\} |\alpha'_o, t_o\rangle_H \tag{I.25}$$

where the subscripts denote the Hamiltonian which defines the time-development of the system, and T is the time-ordering symbol.

In particular, the change in the transformation function due to an infinitesimal change δH in the Hamiltonian is

$$\delta\langle\alpha', t|\alpha'_o, t_o\rangle = -i\int_{t_o}^{t} dt' \langle\alpha' t|\delta H(q(t'))|\alpha_o t_o\rangle \quad .$$

For a relativistic theory described by a Lagrangian density function $\mathcal{L}(x)$, this becomes

$$\delta\langle\alpha',t|\alpha'_o,t_o\rangle = +i\int_{t>x^o>t_o} d^4x\langle\alpha' t|\delta\mathcal{L}(x)|\alpha_o t_o\rangle . \qquad (I.26)$$

Now let us return to the case of the electromagnetic field. If we restrict the external current J by requiring that it vanish asymptotically for large positive or negative times, then for any state $|\psi\rangle$ of the noninteracting field theory, we can define states $|\psi;\text{in}\rangle$ and $|\psi;\text{out}\rangle$ of the theory with an external source, such that if $\Phi(t)$ is any function of the field variables at time t,

$$\langle\varphi;{}^{\text{in}}_{\text{out}}|\Phi(t)|\psi;{}^{\text{in}}_{\text{out}}\rangle_J \sim \langle\varphi|\Phi(t)|\psi\rangle$$

as $t\to-\infty$ or $t\to+\infty$ respectively. Here the subscript J indicates that the matrix elements on the left are evaluated in the presence of the external current, while that on the right is evaluated with $J=0$.

Let us consider the scattering matrix elements

$$\langle\varphi|S(J)|\psi\rangle = \langle\varphi;\text{out}|\psi;\text{in}\rangle_J . \qquad (I.27)$$

According to (I.26) the change in this matrix element due to a small change in J is given by

$$\delta\langle\varphi;\text{out}|\psi;\text{in}\rangle_J = i\int d^4x\ \delta J^\mu(x)\langle\varphi;\text{out}|A_\mu(x)|\psi;\text{in}\rangle_J ,$$

which we may also write

$$\frac{\delta}{\delta J^\mu(x)}\langle\varphi;\text{out}|\psi;\text{in}\rangle_J = i\langle\varphi;\text{out}|A_\mu(x)|\psi;\text{in}\rangle_J , \qquad (I.28)$$

thus defining the functional derivative $\delta/\delta J^\mu(x)$. It should be noted, however, that since (I.27) is only defined for currents J^μ satisfying $\partial_\mu J^\mu = 0$ (I.28) is only defined up to a gauge transformation. We shall return to this point later.

The matrix element on the right may be obtained by solving the equations of motion with suitable boundary conditions. In the radiation gauge, and in momentum space, A^o is given by (I.10) or

$$A^o(\underset{\sim}{k}) = \frac{1}{\underset{\sim}{k}^2} J^o(\underset{\sim}{k})$$

while A^i satisfies the equation

$$k^2 A^i(k) = J_T^i(k) = \gamma^{ij}(\underset{\sim}{k})\, J_j(k)$$

with γ_{ij} again given by (I.13).

Let us for example consider the vacuum-to-vacuum transition amplitude. Then the boundary conditions to be imposed are that $\langle 0;\text{out}|A_\mu(x)|0;\text{in}\rangle$ have only negative-frequency parts as $x^0 \to \infty$ and only positive-frequency components as $x^0 \to +\infty$. It is therefore given by

$$\langle 0;\text{out}|A_\mu(x)|0;\text{in}\rangle_J = \langle 0;\text{out}|0;\text{in}\rangle_J \int d^4y D_{\mu\nu}(x-y) J^\nu(y) \qquad \text{(I.29)}$$

where

$$D_{\mu\nu}(x-y) = \int \frac{d^4k}{(2\pi)^4} \gamma_{\mu\nu}(k) \frac{e^{ik\cdot(x-y)}}{k^2 - i\varepsilon} \quad ,$$

in which $\gamma_{ij}(k)$ is equal to $\gamma_{ij}(\underset{\sim}{k})$ as given by (I.13) but

$$\gamma_{oo}(k) = -\frac{k^2}{\underset{\sim}{k}^2} \quad .$$

Evidently $\gamma_{\mu\nu}(k)$ reduces to the earlier function $\gamma_{\mu\nu}(\underset{\sim}{k})$ when $k^2 = 0$. It may also be written in the form

$$\gamma_{\mu\nu}(k) = g_{\mu\nu} - k_\mu \ell_\nu(k) - \ell_\mu(k) k_\nu \qquad \text{(I.30)}$$

where

$$\ell_\mu(k) = \frac{(-k_o, \underset{\sim}{k})}{2\underset{\sim}{k}^2} \quad .$$

Using (I.29) we may now integrate the functional differential equation (I.28) and obtain

$$\begin{aligned}\langle 0;\text{out}|0;\text{in}\rangle_J &= \exp \tfrac{1}{2} i \int d^4x\, d^4y\, J^\mu(x)\, D_{\mu\nu}(x-y) J^\nu(y) \\ &= \exp \tfrac{1}{2} i \int \frac{d^4k}{(2\pi)^4} J^\mu(k)^* \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon} J^\nu(k) \quad . \qquad \text{(I.31)}\end{aligned}$$

Now since J^μ satisfies the continuity equation,

$$k_\mu J^\mu(k) = 0 \quad ,$$

it follows that we may drop the terms in $\gamma_{\mu\nu}(k)$ proportional to k_μ or k_ν. In fact the vacuum-to-vacuum transition amplitude, given by (I.31) is

gauge-invariant; its value is unaffected by the choice of the vector $\ell_\mu(k)$ in $\gamma_{\mu\nu}$.

The n-point generalized Green's functions of the theory may be defined by

$$G_{\mu_1\ldots\mu_n}(x_1\ldots x_n) = (-i)^n \frac{\delta}{\delta J_{\mu_1}(x_1)} \cdots \frac{\delta}{\delta J_{\mu_n}(x_n)} \langle 0;\text{out}|\,0;\text{in}\rangle_J \Big|_{J=0} \tag{I.32}$$

However, to define these functional derivatives we must be able to make arbitrary variations of J^μ. Thus we must make a specific choice of the function $\gamma_{\mu\nu}(k)$ in (I.31), and then define by (I.32) the Green's functions in this specific gauge. We may choose any function $\gamma_{\mu\nu}(k)$ of the form (I.30), not merely the restricted class of physical gauge functions discussed earlier. For example, it is often convenient to choose the Lorentz or Feynman gauge defined by the choice $\gamma_{\mu\nu}(k) = g_{\mu\nu}$.

According to (I.24) we may write

$$\langle 0;\text{out}|\,0;\text{in}\rangle_J = \langle 0|\,T \exp i\int d^4x\, \mathfrak{L}_J(x)\,|0\rangle$$

or more generally

$$S(J) = T \exp i\int d^4x\, \mathfrak{L}_J(x)\ , \tag{I.33}$$

where $\mathfrak{L}_J(x)$ is the part of the Lagrangian involving J. If all four components of A_μ were independent dynamical variables, we could write simply

$$\mathfrak{L}_J = A_\mu J^\mu\ .$$

However (I.26) is valid only if ΔH is expressed in terms of independent dynamical variables. To apply it directly, we have to eliminate the constraint variables from the Lagrangian, using (I.6) and (I.8). When this is done we find in the radiation gauge

$$\mathfrak{L}_J = A_k J_T^k + \tfrac{1}{2} J^0 \frac{1}{\nabla^2} J^0\ . \tag{I.34}$$

The quadratic term in J^0 represents the direct Coulomb interaction.

Either by applying Wick's theorem or by solving directly for the scattering matrix elements between coherent states we can remove the time ordering in (I.33) and show that

$$S(J) = \exp i \int d^4x\, J^\mu(x)\, A_\mu(x)$$

$$\exp \tfrac{1}{2} i \int d^4x\, d^4y\, J^\mu(x)\, \bar{D}_{\mu\nu}(x-y)\, J^\nu(y)$$

where $\bar{D}_{\mu\nu}(k) = \frac{\gamma_{\mu\nu}(k)}{k^2}$, the integral being taken in the sense of a principal value. Since A_μ here is a free field, we may recognize the first factor as an operator of the form (I.18). In fact,

$$S(J) = U(j)\, e^{i\sigma} \tag{I.35}$$

where

$$j_\mu(\underset{\sim}{k}) = i\, \gamma_{\mu\nu}(\underset{\sim}{k})\, J^\nu(k)\Big|_{k^0=|\underset{\sim}{k}|}$$

and

$$\sigma = \tfrac{1}{2} \int \frac{d^4k}{(2\pi)^4} \frac{J_\mu(k)^* J^\mu(k)}{k^2} \quad .$$

In this last factor we have dropped the $\gamma_{\mu\nu}$ because of the continuity equation.

A particular consequence of this form is to show that the action of a classical current distribution on a state which is initially the vacuum is to produce a coherent state

$$S(J)|0\rangle = U(j)|0\rangle\, e^{i\sigma} = |j\rangle\, e^{i\sigma} \quad .$$

Because of the extra quadratic term in (I.34), the Green's functions are not precisely equal to the vacuum expectation values of time ordered products. It is sometimes convenient to define a "T^* product" by

$$\langle 0|T^*[A_{\mu_1}(x_1)\ldots A_{\mu_n}(x_n)]|0\rangle = G_{\mu_1\ldots\mu_n}(x_1\ldots x_n) \; .$$

For the independent dynamical variables the T^* product is just equal to the T product. However, for A_o it is not. Indeed for a free field A_o vanishes, but

$$G_{oo}(x,y) = -\,i\, D_{oo}(x-y) = i\, \frac{\delta(x^o - y^o)}{4\pi|\underset{\sim}{x} - \underset{\sim}{y}|} \quad .$$

E. The Dirac Field

The free Dirac field is described by a 4-component spinor $\psi(x)$, with the Lagrangian density function

$$\mathfrak{L} = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi \ .$$

Here the γ^{μ} are 4 x 4 Dirac matrices defined by the anticommutation relations

$$\gamma^{\mu}\gamma^{\nu} + \gamma^{\nu}\gamma^{\mu} = -2g^{\mu\nu} \ . \tag{I.36}$$

The spinor $\overline{\psi}$ is related to ψ by

$$\overline{\psi} = \psi^{\dagger}\beta$$

where β is the hermitizing matrix which satisfies

$$\beta^{\dagger} = \beta \ ,$$

$$\beta\gamma^{\mu}\beta^{-1} = \gamma^{\mu\,\dagger} \ .$$

(In the usual representation, $\beta = \gamma^{0}$.)

The field equation derived from this Lagrangian is

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0 \ .$$

The canonical anticommutation relations are

$$\{\psi(x), \overline{\psi}(y)\}_{x^0=y^0} = \gamma^0\delta_3(\underline{x} - \underline{y}). \tag{I.37}$$

In momentum space the field equation becomes

$$(\not{p} + m)\,\psi(p) = 0 \ .$$

Multiplying by $(-\not{p} + m)$ and using the relation

$$\not{p}^2 = -p^2$$

which follows from (I.32), we find

$$(p^2 + m^2)\,\psi(p) = 0$$

Hence $\psi(p) = 0$ unless $p^0 = \pm\omega_{\underline{p}}$ where $\omega_{\underline{p}} = (\underline{p}^2 + m^2)^{\frac{1}{2}}$. Thus we obtain a decomposition of the field into positive and negative frequency parts

$$\psi(x) = \psi^{(+)}(x) + \psi^{(-)}(x)$$

where

$$\psi^{(\pm)}(x) = \int \frac{d^3 \underset{\sim}{p}}{(2\pi)^3 \, 2\omega_{\underset{\sim}{p}}} \psi^{(\pm)}(\underset{\sim}{p}) \, e^{\mp ip\cdot x}$$

with $p^o = \omega_{\underset{\sim}{p}}$. We may expand $\psi^{(\pm)}(\underset{\sim}{p})$ in the form

$$\psi^{(+)}(\underset{\sim}{p}) = \sum_{r=1}^{2} c_r(\underset{\sim}{p}) \, u_r(\underset{\sim}{p})$$

$$\psi^{(-)}(\underset{\sim}{p}) = \sum_{r=1}^{2} d_r(\underset{\sim}{p})^* v_r(\underset{\sim}{p}) \qquad \text{(I.38)}$$

where $u_r(p)$ and $v_r(p)$ are the independent solutions of the momentum-space Dirac equations

$$(\not{p} + m) \, u_r(\underset{\sim}{p}) = 0 \ ,$$

$$(-\not{p} + m) \, v_r(\underset{\sim}{p}) = 0 \ .$$

These solutions may be normalized so that

$$\bar{u}_r(\underset{\sim}{p}) \, u_s(\underset{\sim}{p}) = \delta_{rs} \, 2m \qquad \text{(I.39)}$$

and

$$\sum_r u_r(\underset{\sim}{p}) \, \bar{u}_r(\underset{\sim}{p}) = \Lambda(p) = m - \not{p} \ . \qquad \text{(I.40)}$$

This matrix $\Lambda(p)$ satisfies

$$\Lambda(p) \, \Lambda(p) = 2m\Lambda(p) \ . \qquad \text{(I.41)}$$

Thus $\Lambda(p)/2m$ is the projector on the 2-dimensional subspace of positive-energy spinors u.

The creation and annihilation operators defined by (I.38) satisfy the anticommutation relations

$$\{c_r(\underset{\sim}{p}), c_s^*(\underset{\sim}{p}')\} = \{d_r(\underset{\sim}{p}), d_s^*(\underset{\sim}{p}')\}$$
$$= (2\pi)^3 \, 2\omega_{\underset{\sim}{p}} \, \delta_3(\underset{\sim}{p} - \underset{\sim}{p}') \, \delta_{rs} \qquad (I.42)$$

with all other anticommutators vanishing.

As in the case of the electromagnetic field we may consider the addition to $\mathfrak{L}$ of a source term

$$+ (\bar{\psi}\eta + \eta\bar{\psi})$$

where η and $\bar{\eta}$ are source functions which anticommute with each other and with ψ and $\bar{\psi}$. The solution of the Dirac equation

$$(-i\gamma^\mu \partial_\mu + m)\psi(x) = \eta(x)$$

with appropriate boundary conditions then yields

$$\langle 0;\text{out} | 0;\text{in}\rangle_{\eta,\bar{\eta}} = \exp i\int d^4x \, d^4y \, \bar{\eta}(x) \, S(x - y) \, \eta(y)$$

with

$$S(p) = \frac{1}{\not{p} + m - i\varepsilon} = \frac{\Lambda(p)}{p^2 + m^2 - i\varepsilon} \quad . \qquad (I.43)$$

The Green's functions of the theory are defined by

$$G(x_1 \ldots x_n; y_1 \ldots y_n) = \langle 0 | T[\psi(x_1) \ldots \psi(x_n)\bar{\psi}(y_n) \ldots \bar{\psi}(y_n)] | 0\rangle$$

$$= (-)^n \frac{\vec{\delta}}{\delta\bar{\eta}(x_1)} \cdots \frac{\vec{\delta}}{\delta\bar{\eta}(x_n)} \langle 0;\text{out} | 0;\text{in}\rangle_{\eta,\bar{\eta}} \frac{\overleftarrow{\delta}}{\delta\eta(y_n)} \cdots \frac{\overleftarrow{\delta}}{\delta\eta(y_1)} \Bigg|_{\eta=\bar{\eta}=0}.$$

There is now no difficulty in identifying the n-point Green's functions with the vacuum expectation value of the corresponding time-ordered product, since all components of ψ and $\bar{\psi}$ are dynamical variables.

II. Quantum Electrodynamics

A. Feynman Rules for Green's Functions

We now turn to the fully interacting theory of quantum electrodynamics described by the Lagrangian density function

$$\mathfrak{L} = \mathfrak{L}_{em} + \bar{\psi}[\gamma^\mu(i\partial_\mu + e_o A_\mu) + m_o]\psi$$

where m_o and e_o are the bare mass and charge of the electron.

The Green's functions of the theory may again be defined as functional derivatives of the vacuum-to-vacuum transition amplitude in the presence of external sources. The equations of motion yield a set of coupled differential equations for these functions which may be converted into coupled integral equations incorporating the causal boundary conditions. These equations may then be iterated to give a solution for the Green's functions in the form of a power series in e_o. The terms in the series may most easily be written down using a graphical representation and a set of rules due to Feynman, which we now quote.

The rules may be given either in coordinate space or in momentum space, but we shall quote the momentum space version, which gives the power series expression for

$$G_{\mu_1 \cdots \mu_m}(p_1 \cdots p_\ell, q_1 \cdots q_\ell, k_1 \cdots k_m) . \tag{II.1}$$

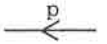
which is the Fourier transform of the general Green's function denoted symbolically by

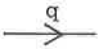
$$\langle 0 | T^*[\psi(x_1) \cdots \psi(x_\ell)\bar{\psi}(y_\ell) \cdots \bar{\psi}(y_1) A_{\mu_1}(z_1) \cdots A_{\mu_m}(z_m)] | 0 \rangle .$$

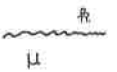
(We define the Fourier transform for the $\bar{\psi}$ variables with the opposite sign in the exponent.) The graphs are constructed from electron lines ($\longleftarrow$) and photon lines ($\sim\sim\sim$). Each n^{th} order graph has n internal vertices, at which three lines meet, as shown:



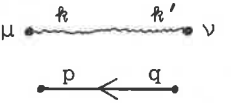
In addition it has $2\ell + m$ external vertices at which a single line terminates. These lines are labeled by the corresponding momenta and the vector indices (and suppressed spinor indices) associated with the external vertices, as shown:

ψ vertex: 

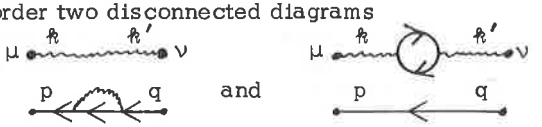
$\bar{\psi}$ vertex: 

A_μ vertex: 

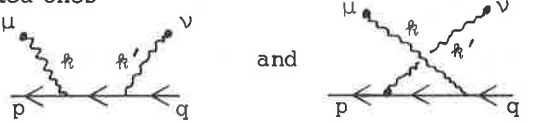
The complete nth order contribution to the Green's function (II.1) is the sum of the contributions of all topologically distinct diagrams of this type. The diagrams need not be completely connected but must have no disconnected pieces without external vertices (the so-called vacuum parts). For example, for the function $G_{\mu\nu}(p,q,k\,k')$ we have in zeroth order one diagram



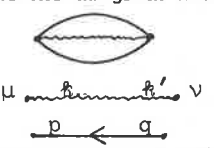
and in second order two disconnected diagrams



and two connected ones

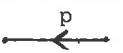


However we do not include the diagram with a completely disconnected piece,



The rules for writing down the contribution of a given diagram are as follows:

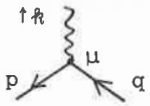
1. Label all lines with a momentum variable, directed along the arrow for electron lines, and with arbitrarily chosen direction for photon lines. By convention, the momentum labels on external photon lines will be directed outwards.
2. Label all internal vertices with a vector index.
3. Associate to each line a factor:



$$-i\,\frac{\Lambda(p)}{p^2 + m^2 - i\varepsilon}$$

$$-i\,\frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon}$$

4. Associate to each internal vertex a factor:

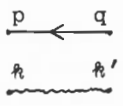


$$i e_0 \gamma^{\mu} (2\pi)^4 \delta_4(p-q+k)$$

5. Order the matrix factors along each electron path (a sequence of electron lines joined at internal vertices) following the arrows from right to left.
6. For any electron path which closes on itself forming a loop, take the trace, and insert a factor of -1.
7. Assign an overall factor of ± 1 according as the permutation required to order the labels $p_1 \ldots p_\ell$ of the ends of the electron paths in the same order as their beginnings $q_1 \ldots q_\ell$ is even or odd.
8. Integrate over all momentum variables of internal lines,

$$\int \frac{d^4 p}{(2\pi)^4} \qquad \text{or} \qquad \int \frac{d^4 k}{(2\pi)^4}$$

9. For any straight-through lines with two external momentum labels insert an extra δ function:



$$(2\pi)^4 \delta_4(p - q)$$

$$(2\pi)^4 \delta_4(k + k')$$

Many of the momentum integrations are eliminated by the momentum-conserving delta functions at the vertices. The number of remaining integrations is the number of independent loops in the diagram, for example 2 in the diagram



B. The Propagator and Mass Renormalization

The complete electron propagator function $S'(p)$ is defined by

$$G(p,q) = -iS'(p)(2\pi)^4 \delta_4(p - q)$$

$$= \int d^4x\, d^4y\, e^{-ip\cdot x + iq\cdot y} \langle 0 | T[\psi(x)\, \bar{\psi}(y)] | 0 \rangle \quad . \qquad \text{(II.2)}$$

Using the translational invariance equation

$$\psi(x) = e^{-iP\cdot y}\, \psi(x - y)\, e^{iP\cdot y} \quad , \qquad \text{(II.3)}$$

where P^μ are the operators representing total energy and momentum, we may perform one of the integrations to obtain a delta function, and thus get

$$-iS'(p) = \int d^4x\, e^{-ip\cdot x} \langle 0|T[\psi(x)\bar{\psi}(0)]|0\rangle \ .$$

Separating the terms in the T-product, inserting a complete set of intermediate states, and using (II.3) again, we find

$$-iS'(p) = \int d^4x\, e^{-ip\cdot x} \{\theta(x^0)\sum_n \langle 0|\psi(0)|n\rangle e^{ip_n\cdot x} \langle n|\bar{\psi}(0)|0\rangle$$

$$-\theta(-x^0)\sum_n \langle 0|\bar{\psi}(0)|n\rangle e^{-ip_n\cdot x} \langle n|\psi(0)|0\rangle\}$$

$$= \sum_n \langle 0|\psi(0)|n\rangle (2\pi)^3 \delta_3(\underline{p} - \underline{p}_n) \frac{i}{p^0 - p_n^0 + i\varepsilon} \langle n|\bar{\psi}(0)|0\rangle$$

$$-\sum_n \langle 0|\bar{\psi}(0)|n\rangle (2\pi)^3 \delta_3(\underline{p} + \underline{p}_n) \frac{-i}{p^0 + p_n^0 - i\varepsilon} \langle n|\psi(0)|0\rangle \ . \qquad \text{(II.4)}$$

Let us consider specifically the contribution from stable single-electron states, $|\underline{p}',\sigma\rangle$ where $\underline{p}'$ denotes the momentum and σ the spin index. From Lorentz covariance one can show that

$$\langle 0|\psi(0)|\underline{p}',\sigma\rangle = Z_2^{\frac{1}{2}}(\underline{p}')\, u_\sigma(\underline{p}')$$

$$\langle \underline{p}',\sigma|\bar{\psi}(0)|0\rangle = Z_2^{\frac{1}{2}}(\underline{p}')\, \bar{u}_\sigma(\underline{p}') \qquad \text{(II.5)}$$

where the coefficient $Z_2^{\frac{1}{2}}$ may depend on $\underline{p}$ if we use a noncovariant gauge in which the Lorentz transformation properties of ψ are complicated. The sum over intermediate states takes the form

$$\int \frac{d^3\underline{p}'}{(2\pi)^3\ 2\omega_{\underline{p}'}} \sum_\sigma$$

and hence the contribution to the first term of (II.4) is

$$Z_2(\underline{p}) \frac{\Lambda(p)}{2\omega_{\underline{p}}} \quad \frac{i}{p^0 - \omega_{\underline{p}} + i\varepsilon} \qquad \text{(II.6)}$$

In the same way the contribution to the second term in (II.4) from single-positron states is

$$Z_2(\underset{\sim}{p})\,\frac{\Lambda(p)}{2\omega_{\underset{\sim}{p}}}\;\frac{-i}{p^\circ+\omega_{\underset{\sim}{p}}-i\epsilon}\quad .$$

Combining the two we find

$$Z_2(\underset{\sim}{p})\,\Lambda(p)\;\frac{-i}{p^2+m^2-i\epsilon}\quad . \tag{II.7}$$

It should be noted that m here refers to the true physical mass of the electron, not to the bare mass m_o which appears in the Lagrangian.

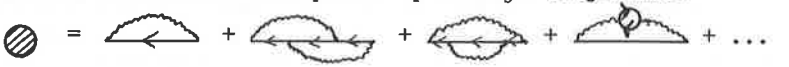
We conclude that the existence of stable single-particle states may be inferred from the appearance of a pole in $S'(p)$. The position of the pole fixes the mass m, and the residue is related to the function $Z_2(\underset{\sim}{p})$.

We may now turn to the diagrams, and examine the structure of the function $G(p,q)$ which we shall denote by a bubble: p ← ○ ← q. It is useful to write it in the form



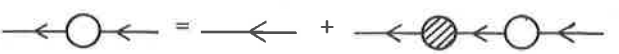
(II.8)

where ◍ denotes the sum of all diagrams which cannot be separated into two disconnected pieces by cutting a single line:



$= -i\Sigma(p)$, say.

The right-hand side of (II.8) is a geometric series which can be summed. It may be written in the form



or, in symbols

$$S'(p) = S(p) - S(p)\Sigma(p)S'(p)\quad .$$

This yields

$$S'(p)^{-1} = S(p)^{-1} + \Sigma(p) = \not{p} + m_o + \Sigma(p).$$

For S' to have a pole this function must vanish. Evidently, its zero need not coincide with $\not{p} + m_o = 0$. Near the position of the zero we must have

$$S'(p)^{-1} \approx Z_2^{-1}(p)(\not{p} + m) \quad .$$

Thus Z_2 is related to the derivative of $\Sigma(p)$ at the position of the zero. We may write

$$S'(p)^{-1} = Z_2^{-1}(p)(\not{p} + m)(1 + \Sigma'(p)) = Z_2^{-1}(p)S^r(p) \quad \text{(II.9)}$$

where $\Sigma'(p)$ must vanish at the pole.

When we actually compute these diagrams, we encounter integrals which diverge for large values of the momenta. However, if we cut off the integrals at some large cut-off mass Λ, and make the separation indicated in (II.9), we find that the only quantities which diverge in the limit $\Lambda \to \infty$ are the so-called renormalization constants

$$\delta m = m - m_o$$

and Z_2^{-1}. The function $\Sigma'(p)$, or $S_r(p)$, is free of such divergences, and this is the function which appears in physical scattering matrix elements.

Unfortunately, however, we encounter divergences of a different kind, the infrared divergences which correspond to integrals that diverge not at large momenta but at $k = 0$. Although $S'(p)$ itself contains no such divergences, they are introduced by the separation in (II.9). Both $Z_2^{-1}(p)$ and $S_r(p)$ are infrared-divergent. These divergences are associated with the vanishing of the photon mass, and may be eliminated by giving the photon a small mass λ. Then in place of divergences we have terms depending on $\ell n\, \lambda$. In the conventional treatment such terms eventually cancel out in expressions for observable scattering probabilities but the fact that they appear only as a result of the way we separate the factors in (II.9) suggests that it should be possible to find an alternative way of proceeding in which they never appear at all.

A similar analysis may be carried through for the photon. There in place of (II.7) we find that the contribution of the single-photon intermediate states is

$$Z_3\, \gamma_{\mu\nu}(k)\, \frac{-i}{k^2 - i\varepsilon} \quad . \qquad \text{(II.10)}$$

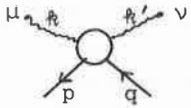
In this case Z_3 is actually a constant, since it can easily be shown to be gauge-invariant. We may factor out this constant, and write (except for irrelevant complications associated with the choice of gauge),

$$D'_{\mu\nu}(k) = Z_3^{-1}\, D^r_{\mu\nu}(k) \quad . \qquad \text{(II.11)}$$

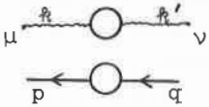
We then find that the only divergences are contained in Z_3^{-1}. In this case, no infrared divergences appear. The physical reason for this is the fact the photon is neutral, as we shall see.

C. Mass-shell Poles in Green's Functions

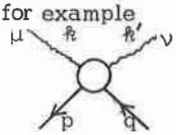
The diagrams contributing to the general Green's function (II.1) may be separated into classes according to the connectivity structure of the diagrams. For example, the diagrams contributing to the Green's function $G_{\mu\nu}(p,q,k\,k')$ are either completely connected, or separate into two disconnected pieces. It is convenient to represent the sum of all connected diagrams by a bubble



The disconnected diagrams are represented by



It is clear from the Feynman rules that the contribution from the diagrams of a given connectivity class contains an overall energy-momentum-conserving delta function factor for each connected piece. Let us define for example



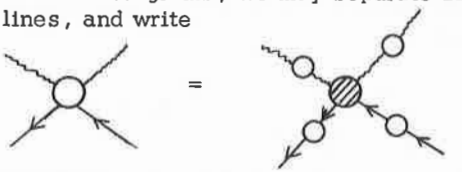
$$= (2\pi)^4\,\delta_4(p+k+k'-q)\,G^c_{\mu\nu}(p,q,k\,k')$$

Then we have the decomposition

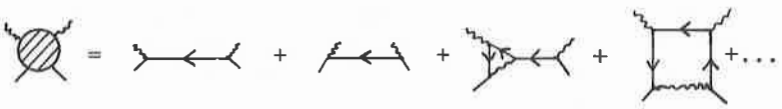
$$G_{\mu\nu}(p,q,k\,k') = (-i)^2\,(2\pi)^4\delta_4(p-q)S'(p)(2\pi)^4\delta_4(k+k')D'_{\mu\nu}(k)$$
$$+ (2\pi)^4\delta_4(p+k+k'-q)\,G^c_{\mu\nu}(p,q,k\,k'), \quad \text{(II.12)}$$

and similarly for other diagrams.

For the connected diagrams, we may separate all corrections to the external lines, and write



where the shaded bubble, like $\Sigma(p)$, has the property that by cutting a single line we cannot separate one of the external lines from the others. Thus



If we define M by



$$= (2\pi)^4 \delta_4(p + k + k' - q) M^{\mu\nu}(p,q,k\,k')$$

then this decomposition corresponds to

$$G^c_{\mu\nu}(p,q,k\,k') = D'_{\mu\rho}(k) D'_{\nu\sigma}(k') S'(p) M^{\rho\sigma}(p,q,k\,k') S'(q) \quad . \quad \text{(II.13)}$$

Clearly this function will have poles on the mass shells, because of the poles in the complete propagator functions. The residues at these poles may be related to matrix elements between certain states.

To see this connection, let us consider the Fourier transform of some time-ordered product,

$$\int d^4x\, e^{-ip\cdot x} \langle 0 | T[\psi(x) \ldots] | 0 \rangle \quad .$$

The singularities of this function in the variable p^0 are related to the asymptotic behaviour of the time-ordered product for large x^0. It is convenient to write it as

$$\int d^4x\, e^{-ip\cdot x} \{ \theta(x^0) \langle 0 | \psi(x) T[\ldots] | 0 \rangle$$

$$- \theta(-x^0) \langle 0 | T[\ldots] \psi(x) | 0 \rangle + \ldots \} \quad . \qquad \text{(II.14)}$$

It is clear that the remaining terms, indicated by dots, are expressible as integrals in which the x^0 integration has a finite range. Such terms are analytic functions of p^0 and do not contribute to the singularities in this variable. So we can forget them.

As in our discussion of the propagator function we may insert a complete set of intermediate states and consider the contribution of the stable single-particle states. For the first term in (II.14) this contribution is found to be

$$\frac{iZ_2^{\frac{1}{2}}(\underline{p})}{2\omega_{\underline{p}}(p^0 - \omega_{\underline{p}} + i\varepsilon)} \sum_\sigma u_\sigma(\underline{p}) \langle \underline{p},\sigma | T[\ldots] | 0 \rangle \quad .$$

Without changing its behaviour near the positive-energy mass shell, we can rewrite this contribution in the form

$$\frac{-iZ_2^{\frac{1}{2}}(\underline{p})}{p^2+m^2-i\varepsilon}\sum_\sigma u_\sigma(\underline{p})\langle \underline{p},\sigma|T[\ldots]|0\rangle\ .$$

Thus to obtain the matrix element

$$\langle \underline{p},\sigma|T[\ldots]|0\rangle$$

we have to isolate the residue of the pole at $p^2+m^2=0$, divide by $-iZ_2^{\frac{1}{2}}$, and then find the coefficient of the appropriate spinor wave function $u_\sigma(\underline{p})$. According to (II.13), the Green's function contains a factor $S'(\underline{p})$, which near the mass shell has the structure

$$-i\frac{Z_2(p)\Lambda(p)}{p^2+m^2-i\varepsilon}=-i\frac{Z_2^{\frac{1}{2}}(p)}{p^2+m^2-i\varepsilon}\sum_\sigma u_\sigma(\underline{p})\,\overline{u}_\sigma(\underline{p})Z_2^{\frac{1}{2}}(p). \quad \text{(II.15)}$$

Hence we have to remove this factor and replace it simply by $Z_2^{\frac{1}{2}}(p)\overline{u}_\sigma(\underline{p})$.

More generally the same rule applies to the reduction of the function

$$\int d^4x\, e^{-ip\cdot x}\langle f;\text{out}|T[\psi(x)\ldots]|i;\text{in}\rangle \quad \text{(II.16)}$$

where $|i;\text{in}\rangle$ and $\langle f;\text{out}|$ are arbitrary initial and final states to a matrix element with one extra particle in the final state,

$$\langle f,\underline{p}\,\sigma;\text{out}|T[\ldots]|i;\text{in}\rangle\ . \quad \text{(II.17)}$$

To see this let us first choose some spinor wave function $\overline{\varphi}(\underline{p})$, and consider the limit for large positive t of the matrix element

$$\lim_{t\to\infty}\frac{1}{2m}\int\frac{d^4p}{(2\pi)^4}\,\overline{\varphi}(\underline{p})\,e^{-i(p^0-\omega_{\underline{p}})t}$$

$$Z(p)^{-\frac{1}{2}}\int d^4x\, e^{-ip\cdot x}\langle f;\text{out}|T[\psi(x)\ldots]|i;\text{in}\rangle\ . \quad \text{(II.18)}$$

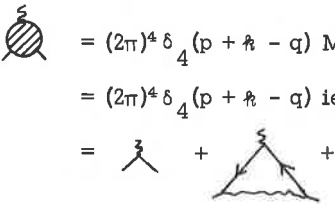
The p^0 integration has the effect of setting $x^0=t$. Thus, as $t\to+\infty$, the values of x^0 which contribute to the integral also tend to $+\infty$, so that $\psi(x)$ may be taken outside the time-ordering symbol. Moreover because of the exponential factor depending on t, the only momentum components which survive in the limit are those with $p^0=\omega_{\underline{p}}$, which came from intermediate states which differ from the final state only in having one extra particle. Thus the limit is simply

$$\langle f,\bar{\varphi};\text{out}|\,T[\ldots]\,|i;\text{in}\rangle$$
$$= \frac{1}{2m}\int \frac{d^3p}{(2\pi)^3\,2w_{\underset{\sim}{p}}}\,\bar{\varphi}(\underset{\sim}{p})\sum_{\sigma} u_{\sigma}(\underset{\sim}{p})\langle f,\underset{\sim}{p}\,\sigma;\text{out}|\,T[\ldots]\,|i;\text{in}\rangle \quad . \qquad \text{(II.19)}$$

In the same way of course to obtain a matrix element with a single additional particle of momentum $\underset{\sim}{q}$ and spin σ in the initial state we have to remove a factor $S'(q)$ and replace it by $Z_2^{\frac{1}{2}}(q)u_{\sigma}(\underset{\sim}{q})$. For a photon of momentum $\hbar$ and polarization vector ϵ, we must remove a factor $D_{\mu\nu}(\hbar)$ and replace it by $Z_3^{\frac{1}{2}}\epsilon_{\nu}(\hbar)$.

D. The Vertex Function and Charge Renormalization

The complete vertex function Γ^{μ} is defined by



$$= (2\pi)^4\delta_4(p+\hbar-q)\,M^{\mu}(p,q,\hbar)$$
$$= (2\pi)^4\delta_4(p+\hbar-q)\,ie_o\Gamma^{\mu}(p,q).$$
$$= \quad + \quad + \ldots$$

It may be regarded as a corrrected form of the bare vertex function γ^{μ}.

The function $Z_1(p)$ is defined by the condition

$$\bar{u}(\underset{\sim}{p})\,\Gamma^{\mu}(p,p)\,u(\underset{\sim}{p}) = Z_1^{-1}(p)\,\bar{u}(\underset{\sim}{p})\,\gamma^{\mu}u(\underset{\sim}{p}) \qquad \text{(II.20)}$$

for $p^2 + m^2 = 0$. This function Z_1 is not, however, a new function as we shall now prove.

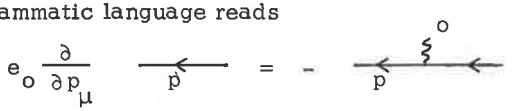
From the identity

$$\frac{\partial}{\partial p_{\mu}}S^{-1}(p) = \frac{\partial}{\partial p_{\mu}}(\not{p} + m_o) = \gamma^{\mu} \qquad \text{(II.21)}$$

we easily derive

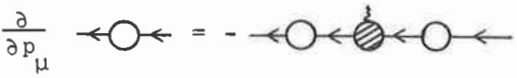
$$\frac{\partial}{\partial p_{\mu}}S(p) = -\,S(p)\,\gamma^{\mu}S(p)$$

which in diagrammatic language reads



$$e_o\frac{\partial}{\partial p_{\mu}} \qquad p \qquad = - \qquad p \qquad .$$

Thus differentiation has the effect of inserting a zero-momentum photon line. Applying this rule to the complete electron propagator we obtain

$e_o \frac{\partial}{\partial p_\mu}$  = -

allowing for the possibility of inserting the line at all possible points. Thus

$$\frac{\partial}{\partial p_\mu} S'(p) = -S'(p)\, \Gamma^\mu(p,p)\, S'(p)$$

or

$$\frac{\partial}{\partial p_\mu} S'^{-1}(p) = \Gamma^\mu(p,p) \ . \tag{II.22}$$

Going to the mass shell we then obtain an identity

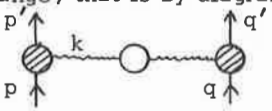
$$Z_2^{-1}(p) = Z_1^{-1}(p) \ . \tag{II.23}$$

We may factor out this function and write

$$\Gamma^\mu(p,q) = Z_2^{-\frac{1}{2}}(p)\, \Gamma_r^\mu(p,q)\, Z_2^{-\frac{1}{2}}(q) \ .$$

As before the diagrams defining Γ^μ are divergent, but all ultraviolet divergences are contained in the factors $Z_2^{-\frac{1}{2}}$. Again, however, this separation introduces infrared divergences which were not present in Γ^μ.

The observed charge of the electron e (like the observed mass m) does not coincide with the parameter e_o appearing in the Lagrangian. To see what the charge really is, we have to recall its definition. The charge is what fixes the strength of the interaction between two particles far apart from one another, or equivalently the interaction for small values of the energy-momentum transfer. Now the scattering of two electrons far apart is dominated by diagrams containing a single photon exchange, that is by diagrams of the form



Using the rules we have obtained this scattering matrix element is, apart from a factor

$$(2\pi)^4 \delta_4(p'+q'-p-q)$$

of the form

$$ie_o^2 Z_2^{\frac{1}{2}}(p)\,\bar{u}(p')\,\Gamma^{\mu}(p',p)\,u(p)\,Z_2^{\frac{1}{2}}(p)\,D'_{\mu\nu}(k)$$

$$Z_2^{\frac{1}{2}}(q')\,\bar{u}(q')\,\Gamma^{\nu}(q',q)\,u(q)\,Z_2^{\frac{1}{2}}(q),$$

where $k = p - p' = q' - q$. Near $k = 0$ it thus has the form

$$ie_o^2\, Z_2(p)\, Z_1^{-1}(p)\,\bar{u}\,\gamma^{\mu}u\,\frac{Z_3\,\gamma_{\mu\nu}(k)}{k^2}\,Z_2(q)\,Z_1^{-1}(q)\,\bar{u}\,\gamma^{\nu}u\,,$$

where we have used (II.20) and the definition of Z_3. Because of (II.23) the factors of Z_2 and Z_1 cancel, and so the matrix element is

$$ie^2\, u\gamma^{\mu}\bar{u}\,\frac{\gamma_{\mu\nu}(k)}{k^2}\,\bar{u}\,\gamma^{\nu}\,u$$

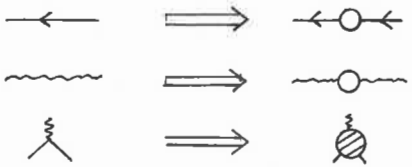
where the physical charge e is given by

$$e^2 = Z_3\, e_o^2 \qquad . \tag{II.24}$$

The only contribution to the charge renormalization comes from the wave function renormalization constant of the photon.

E. Cancellation of Divergences

To any diagram we may associate a skeleton diagram obtained by removing all self-energy parts (connected to the remainder by two lines only) and vertex parts (connected by three lines only). The complete set of diagrams may be obtained by inserting all possible self-energy and vertex parts into these skeleton diagrams. This leads to the replacements



or

$$S \Longrightarrow S' = Z_2\, S_r$$

$$D \Longrightarrow D' = Z_3\, D_r$$

$$e_o\gamma \Longrightarrow e_o\Gamma = Z_3^{-\frac{1}{2}} e\, Z_2^{-\frac{1}{2}}\Gamma_r\, Z_2^{-\frac{1}{2}}\,.$$

It is then easy to see that in any scattering matrix element all the renormalization constants cancel. For each internal line we have a factor of Z which cancels the two factors of $Z^{-\frac{1}{2}}$ associated with the two corresponding vertices. For each external mass-shell line representing a physical particle we have a factor of $Z^{\frac{1}{2}}$ which again cancels the factor of $Z^{-\frac{1}{2}}$ associated with the vertex. Thus we may replace

$$S' \Longrightarrow S_r$$

$$D' \Longrightarrow D_r$$

$$e_o \Gamma \Longrightarrow e\Gamma_r$$

$$Z_2^{\frac{1}{2}} u \Longrightarrow u$$

$$Z_3^{\frac{1}{2}} \epsilon \Longrightarrow \epsilon$$

without changing anything. We then have matrix elements which are free of ultraviolet divergences.

The Green's functions themselves are of course not free of divergences. In that case the external lines are represented not by $Z_2^{\frac{1}{2}} u$ but by $S' = Z_2 S_r$. There is thus an extra factor of $Z_2^{\frac{1}{2}}$ remaining for each such line. To obtain functions free of ultraviolet divergences we have to remove these factors and define for example

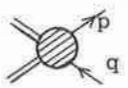
$$G_{\mu\nu}(p,q,\hbar\hbar') = Z_2^{\frac{1}{2}}(p) Z_2^{\frac{1}{2}}(q) Z_3 \, G^r_{\mu\nu}(p,q,\hbar\hbar') \; . \qquad \text{(II.25)}$$

The same factors must be extracted from G^c or M.

However as we remarked earlier this procedure is not altogether satisfactory--quite apart from the unpleasant feature that we are manipulating infinite quantities--because we eliminate divergences of one kind only at the expense of introducing new divergences of a different kind, the infrared divergences.

F. Infrared Divergences

The scattering matrix elements calculated according to the rules we have given and expressed in terms of the physical mass and charge, m and e, of the electron, are now free of ultraviolet divergences, but not of infrared divergences. To see explicitly where these came from let us consider a diagram

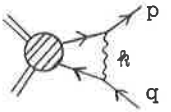


which contributes to some scattering process. Here the double lines indicate other particles which we shall assume to be neutral and with

which we shall not be concerned. We shall denote this contribution by

$$\bar{u}(p)\, M(p,q)\, u(q).$$

Then we can draw another diagram which differs from this one only in having an extra photon line joining the external lines:



According to our rules the contribution of this diagram is

$$-ie^2 \int \frac{d^4k}{(2\pi)^4} \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon}\, \bar{u}(p)\gamma^\mu \frac{1}{\not p - \not k + m - i\varepsilon}\, M(p-k, q-k) \frac{1}{\not q - \not k + m - i\varepsilon}\, \gamma^\nu u(q) \tag{II.26}$$

Now let us consider the behaviour of the integrand near $k = 0$. Since M is presumably a slowly varying function we may replace $p - k$ by p and $q - k$ by q and take it outside the integral. The $q - k$ propagator is

$$\frac{m - \not q + \not k}{(q-k)^2 + m^2} \quad .$$

The denominator is

$$q^2 + m^2 - 2q\cdot k + k^2 = -2q\cdot k + k^2 \quad .$$

Thus keeping only the leading terms in k it becomes

$$\frac{m - \not q}{-2q\cdot k - i\varepsilon}$$

But using the commutation relations of the Dirac matrices

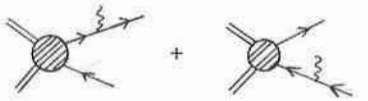
$$(m - \not q)\gamma^\nu u(q) = [\gamma^\nu(m + \not q) - 2q^\nu] u(q) = -2q^\nu u(q) \quad .$$

Thus we obtain for the contribution to (II.26) from a small region Ω of the k space around $k = 0$ precisely the original matrix element multiplied by an additional factor

$$B(p,q) = -ie^2 \int_\Omega \frac{d^4k}{(2\pi)^4} \frac{p^\mu}{p\cdot k + i\varepsilon} \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon} \frac{q^\nu}{q\cdot k + i\varepsilon}$$

It is easily seen that this integral diverges at $k = 0$ because it behaves like $\int d^4k/k^4$. This divergence is associated with the vanishing of the photon mass. If we were to replace k^2 in the denominator by $k^2 + \lambda^2$, we would obtain a convergent integral which behaves for small λ like $\ell n(\lambda/K)$ where K is the cutoff defined by the size of Ω.

A similar divergence appears in a slightly different way if we compute the probability of emitting an additional soft photon. This process is described in lowest order by the diagrams



By the same reasoning this matrix element is given by

$$\bar{u}(\underset{\sim}{p})\, M(p,q)\, u(\underset{\sim}{q})\, e\left(\frac{q^\mu}{q\cdot k} - \frac{p^\mu}{p\cdot k}\right) \quad . \tag{II.27}$$

To find the probability P_1 of emission of a photon with momentum in the range Ω we have to square this matrix element and integrate over k:

$$|\bar{u}(\underset{\sim}{p})M(p,q)u(\underset{\sim}{q})|^2\; e^2 \int_\Omega \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|}\left(\frac{q^\mu}{q\cdot k} - \frac{p^\mu}{p\cdot k}\right)\gamma_{\mu\nu}(k)\left(\frac{q^\nu}{q\cdot k} - \frac{p^\mu}{p\cdot k}\right) \tag{II.28}$$

The factor $\gamma_{\mu\nu}(k)$ arises from the sum over polarizations of the emitted photon. On the other hand, to order e^2 the probability P_0 for the process without photon emission is given by adding the two diagrams and squaring:

$$|\bar{u}(\underset{\sim}{p})M(p,q)u(\underset{\sim}{q})|^2\, |1 + B(p,q)|^2$$

$$= |\bar{u}(\underset{\sim}{p})M(p,q)u(\underset{\sim}{q})|^2\, (1 + 2\mathrm{Re}\, B(p,q))$$

But it is easy to see that the real part of B comes from the δ-function term in the denominator $1/(k^2 - i\varepsilon)$, namely

$$\mathrm{Re}\, B(p,q) = e^2 \int_\Omega \frac{d^4k}{(2\pi)^4}\, \frac{p^\mu}{p\cdot k}\, \gamma_{\mu\nu}(k)\, \pi\delta(k^2)\, \frac{q^\nu}{q\cdot k}$$

$$= e^2 \int_\Omega \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|}\, \frac{p^\mu}{p\cdot k}\, \gamma_{\mu\nu}(k)\, \frac{q^\nu}{q\cdot k} \quad .$$

This term therefore cancels the cross term between q and p in (II.28). The quadratic terms can be shown to be cancelled by the contributions

to the renormalization functions $Z_2^{\frac{1}{2}}(p)$ and $Z_2^{\frac{1}{2}}(q)$. Thus $P_0 + P_1$ is actually finite to order e^2.

G. Removal of Infrared Divergences

This cancellation occurs to all orders. The conventional prescription for handling infrared divergences, which was developed particularly by Yennie, Frautschi and Suura and by Eriksson, may be described as follows.

First we give the photon a small fictitious mass λ so as to remove the infrared divergences in individual diagrams. Then we compute the scattering probability $P_n(\lambda, \Delta E)$ for a given process with emission of n soft photons of total energy less than ΔE. The total scattering probability independent of the number of emitted soft photons is

$$P(\lambda, \Delta E) = \sum_{n=0}^{\infty} P_n(\lambda, \Delta E) \quad . \tag{II.29}$$

In this sum all terms proportional to $\ell n\, \lambda$ cancel, and we obtain a finite limit

$$P(\Delta E) = \lim_{\lambda \to 0} P(\lambda, \Delta E) \ .$$

Since in any given experimental situation there is also some small energy loss to soft photons which would be undetectable, this quantity is the observable scattering probability.

Let us now examine this process in more detail. It is easy to show that in the limit $\lambda \to 0$ the individual terms in the sum (II.29) all vanish:

$$\lim_{\lambda \to 0} P_n(\lambda, \Delta E) = 0 \ . \tag{II.30}$$

In each order of perturbation theory, the contribution is infrared divergent, but it turns out that the series is of the form $e^{\alpha \ell n \lambda}$ which tends to zero as $\lambda \to 0$.

It is also possible to show that the expectation value of the number of emitted photons is actually infinite, in the sense that

$$\lim_{\lambda \to 0} \sum_{n=0}^{\infty} n P_n(\lambda, \Delta E) = \infty \ . \tag{II.31}$$

The physical origin of the phenomenon of infrared divergences may be traced to an invalid assumption which is implicit in the conventional perturbation calculation, namely that the asymptotic states

of the theory are states in the Fock space containing a finite total number of particles. In fact, if we start with such a state, then in the subsequent scattering process an infinite number of soft photons will in general be emitted, so that the final state does not really belong to the Fock space. Divergences appear because we insist on trying to characterize the final state by its components in a Fock space basis. As is shown by (II.31) these components are in fact all zero, so that the state does not belong to this space at all.

Mathematically, the infrared divergences appear because we have insisted on trying to evaluate the residues of mass-shell poles of functions which actually do not have poles. In a theory with massless particles, there is no gap in the mass spectrum above the electron mass m. The continuum, corresponding to states containing an electron and soft photons, comes right down to m. Thus instead of an isolated pole we actually get a branch cut starting at the mass shell, as we shall see later.

To obtain a more satisfactory treatment of the infrared divergences we need to extend our Hilbert space so that it can accommodate states of infinitely many photons. The way to do this will be our next topic.

As in a conventional field theory, the Green's functions should contain complete information about the theory. However we must drop the invalid assumption that the asymptotic states necessarily belong to the Fock space, and instead seek to determine the structure of the space of asymptotic states from the nature of the singularities in the Green's functions. We should be able in this way to determine what the asymptotic states really are and to evaluate the scattering matrix elements between them.

III. Generalized Coherent States

We now digress from our main discussion to describe the construction of a Hilbert space of states of the free electromagnetic field which can contain infinitely many soft photons. We shall discuss the interaction with an external current $J^{\mu}(x)$ which does not vanish asymptotically. In such a case the scattering matrix elements as usually defined are infrared divergent. We shall show that on the larger Hilbert space a scattering operator may be defined all of whose matrix elements are finite. Later we shall extend the discussion to quantum electrodynamics

A. States of Infinite Photon Number

It is clearly difficult to extend the conventional Fock space basis so as to include states with infinitely many photons. However a more convenient choice is available, namely the coherent states $|f\rangle$ which form a basis, albeit not an orthonormal one in the Fock space.

From (I.24) we see that in order to accommodate states in which the expectation value of the photon number is infinite we have to define coherent states corresponding to wave functions f such that

$$f^* f = \infty \ .$$

It is clear from (I.23) that if

$$(f - g)^* (f - g) = \infty$$

then the coherent states $|f\rangle$ and $|g\rangle$ should be orthogonal. Thus we may expect the space spanned by such states to separate into orthogonal subspaces such that in some sense states in the same subspace differ only by finite numbers of photons.

There is, however, a difficulty which arises from the fact that in the expression for the scalar product

$$\langle f|g\rangle = e^{f^* g - \frac{1}{2} f^* f - \frac{1}{2} g^* g} \qquad \text{(III.1)}$$

the imaginary part of the exponent may diverge even though the real part converges, so that the scalar product has formally the value $e^{i\infty}$.

To indicate more clearly the nature of this difficulty, it is convenient to go over to a discrete representation. Let us suppose that we have a set of modes labeled by a parameter k, with creation and annihilation operators satisfying

$$[a_k, a_l^*] = \delta_{kl}$$

Then the coherent states are labeled by sequences

$$f = \{f_k\}$$

of complex numbers. They are defined by

$$|f\rangle = U(f)|0\rangle \ ,$$

with

$$U(f) = \exp \sum_k (a_k^* f_k - f_k^* a_k) \ .$$

Because of this exponential structure, the coherent states have a product form

$$|f\rangle = |f_1\rangle\ |f_2\rangle \ldots = \prod_{k} |f_k\rangle \tag{III.2}$$

Their scalar products are

$$\langle f|g\rangle = \prod_{k} \langle f_k|g_k\rangle = \exp \sum_{k} (f_k^* g_k - \tfrac{1}{2}|f_k|^2 - \tfrac{1}{2}|g_k|^2) \ . \tag{III.3}$$

So long as we stick to states such that

$$\sum_{k} |f_k|^2 < \infty \tag{III.4}$$

these scalar products are always finite.

However when we relax the condition (III.4) we run into trouble. For let us choose a sequence $\lambda = \{\lambda_k\}$ of real numbers and consider the state

$$|f,\lambda\rangle = \prod_{k} |f_k\rangle\ e^{i\lambda_k} \ .$$

Since this state differs from $|f\rangle$ only by a phase factor for each mode, we shall expect it to be physically indistinguishable. However, the scalar product of the two states is

$$\langle f|f,\lambda\rangle = \prod_{k} \langle f_k|f_k\rangle\ e^{i\lambda_k}$$

$$= \exp i \sum_{k} \lambda_k \ .$$

If the sum is finite,

$$\sum_{k} \lambda_k = \lambda < \infty \ ,$$

then the scalar product is just a phase factor, $e^{i\lambda}$. However, if the sum diverges, then formally

$$\langle f|f,\lambda\rangle = e^{i\infty} \ .$$

B. Infinite Tensor Products

This is a mathematical problem, which was in fact solved by von Neumann in a classic paper on infinite tensor products. Consider a sequence of Hilbert spaces $\mathcal{K}_k$. (In our case, $\mathcal{K}_k$ is the space of states of the kth mode.) Then von Neumann's definition of the infinite tensor product

$$\mathcal{K} = \prod_k \otimes \mathcal{K}_k$$

is equivalent to the following.

Every sequence of normalized vectors $|\psi_k\rangle \in \mathcal{K}_k$ defines a product vector

$$|\psi\rangle = \prod_k |\psi_k\rangle \quad .$$

We then define equivalence classes of these product vectors. Two product vectors $|\varphi\rangle$ and $|\psi\rangle$ are equivalent if and only if

$$\sum_k |1 - \langle\varphi_k|\psi_k\rangle| < \infty \ .$$

Their scalar product is then defined to be the convergent product

$$\langle\varphi|\psi\rangle = \prod_k \langle\varphi_k|\psi_k\rangle \quad . \tag{III.5}$$

By convention the scalar product of two inequivalent product vectors, represented formally by a divergent product, is zero. Finally the space $\mathcal{K}$ is defined as the space of all finite linear combinations of such products, together with the limits of convergent sequences, convergence being defined of course in terms of the scalar product (III.5). It was proved by von Neumann that the space so constructed is a Hilbert space. It is nonseparable, which means that it does not possess a countable basis. The subspace $\mathcal{K}(\psi)$ spanned by all product vectors equivalent to a given product vector $|\psi\rangle$ is, however, a separable space. The nonseparable space $\mathcal{K}$ is an uncountably infinite direct sum of these subspaces.

If for each k, U_k is a unitary operator on $\mathcal{K}_k$, then we can define a tensor product operator

$$U = \prod_k U_k$$

by

$$U|\psi\rangle = U\prod_{k} |\psi_k\rangle = \prod_{k} U_k|\psi_k\rangle \quad .$$

It is easy to prove that U is a unitary operator on $\mathcal{H}$.

In particular, if $\lambda = \{\lambda_k\}$ is a sequence of real numbers, we may define

$$V_k(\lambda_k) = e^{i\lambda_k}$$

and

$$V(\lambda) = \prod_{k} V_k(\lambda_k) \quad . \tag{III.6}$$

If the series $\sum\lambda_k$ is absolutely convergent, with

$$\sum_{k} \lambda_k = \lambda \quad ,$$

then

$$V(\lambda) = e^{i\lambda} \quad .$$

However, if the series is not absolutely convergent, then $V(\lambda)$ is not a multiple of the identity.

We may now apply these ideas to coherent states. We set

$$U(f) = \prod_{k} U_k(f_k) \tag{III.7}$$

with

$$U_k(f_k) = \exp(a_k^* f_k - f_k^* a_k) \quad .$$

Then we define generalized coherent states by

$$|f,\lambda\rangle = U(f)\, V(\lambda)\, |0\rangle \, .$$

The unitary operators $U(f)\, V(\lambda)$ satisfy the multiplication laws

$$V(\lambda)\,V(\mu) = V(\lambda+\mu)$$

$$U(f)\,V(\lambda) = V(\lambda)\,U(f)$$

$$U(f)\,U(g) = U(f+g)\,V\left(\frac{g^{*}f - f^{*}g}{2i}\right) \quad . \qquad \text{(III.8)}$$

The scalar products of the generalized coherent states are given by

$$\langle f,\lambda \,|\, g\ \mu\rangle = \exp\sum_{\underline{k}} (f^{*}_{\underline{k}} g_{\underline{k}} - \tfrac{1}{2} f^{*}_{\underline{k}} f_{\underline{k}} - \tfrac{1}{2} g^{*}_{\underline{k}} g_{\underline{k}} - i\lambda_{\underline{k}} + i\mu_{\underline{k}}) \qquad \text{(III.9)}$$

if the series in the exponent is absolutely convergent, and is zero otherwise. Note that all terms are combined before the summation over $\underline{k}$ is performed.

The extra label λ in these states is physically irrelevant. The vectors $|f,\lambda\rangle$ and $|f,0\rangle$ describe the same physical state, just as in the familiar case the vectors $|f\rangle e^{i\lambda}$ and $|f\rangle$ do. It would be reasonable to define a generalized ray defined by a vector $|\psi\rangle$ as the space spanned by all vectors of the form $V(\lambda)|\psi\rangle$.

C. Generalized Coherent States of the Electromagnetic Field

So far we have used a discrete mode representation, but it is not difficult to rewrite the results in a form appropriate to an integral representation.

We consider a class of function $f_{\mu}(\underline{k})$ which are no longer required to satisfy

$$f^{*}f = \int \frac{d^3\underline{k}}{(2\pi)^3\, 2|\underline{k}|}\ f^{*}_{\mu}(\underline{k})\, f^{\mu}(\underline{k}) < \infty \ .$$

However, to ensure that the total energy of the soft photons is finite we impose the condition

$$\int \frac{d^3\underline{k}}{(2\pi)^3\, 2|\underline{k}|} \left(\frac{|\underline{k}|}{|\underline{k}|+1}\right) f^{*}_{r}(\underline{k})\, f^{r}(\underline{k}) < \infty \ . \qquad \text{(III.10)}$$

We have inserted the denominator so as to avoid imposing extra conditions on f at large values of $|\underline{k}|$. Even in the usual Fock space, there are some coherent states for which the total energy of the hard photons is infinite, even though the expectation value of the total number is finite. Instead of sequences of real numbers, we now consider real measurable functions $\lambda(\underline{k})$ satisfying the condition

$$\int \frac{d^3\underline{k}}{(2\pi)^3\, 2|\underline{k}|} \left(\frac{|\underline{k}|}{|\underline{k}|+1}\right) |\lambda(\underline{k})| < \infty \quad . \qquad \text{(III.11)}$$

For each such f and λ we introduce formally a generalized coherent state $|f,\lambda\rangle$. Two such states are equivalent

$$|f,\lambda\rangle \sim |g,\mu\rangle$$

if and only if the integral

$$f^*g - \tfrac{1}{2}f^*f - \tfrac{1}{2}g^*g - i\lambda + i\mu$$
$$= \int \frac{d^3\underset{\sim}{k}}{(2\pi)^3 2|\underset{\sim}{k}|} [f^*_\mu(\underset{\sim}{k})g^\mu(\underset{\sim}{k}) - \tfrac{1}{2}f^*_\mu(\underset{\sim}{k})f^\mu(\underset{\sim}{k}) - \tfrac{1}{2}g_\mu(\underset{\sim}{k})g^\mu(\underset{\sim}{k}) - i\lambda(\underset{\sim}{k}) + i\mu(\underset{\sim}{k})] \tag{III.12}$$

is (absolutely) convergent. If they are equivalent their scalar product is given by

$$\langle f,\lambda|g,\mu\rangle = \exp(f^*g - \tfrac{1}{2}f^*f - \tfrac{1}{2}g^*g - i\lambda + i\mu) \ . \tag{III.13}$$

If they are inequivalent, their scalar product is zero by definition.

We may now construct as before a nonseparable Hilbert space $\mathcal{H}_{IR}$ which is the completion of the set of all finite linear combinations of generalized coherent states. It is not hard to prove that it is indeed a Hilbert space. It is the direct sum of uncountably many separable Hilbert spaces, each spanned by generalized coherent states of one specific equivalence class.

We shall frequently write down exponential expressions of the form (III.13). In every case, we apply the same convention. All terms in the exponent are to be combined before doing the integration over $\underset{\sim}{k}$. Then if the integral converges, it yields the required result (whether or not the integrals of individual terms would converge). If it diverges, the exponential has the value zero.

D. Scattering by an External Current

We now return to the problem discussed earlier of the scattering by an external unquantized current $J_\mu(x)$. Previously we restricted this current by the condition that it vanish for large positive and negative times. However we shall now drop this condition and suppose instead that there are classical charged particles persisting at infinite times. So long as these particles are asymptotically unaccelerated we should still expect to be able to define asymptotic states $|\psi;\text{in}\rangle$ and $|\psi;\text{out}\rangle$ of the radiation field corresponding to any state $|\psi\rangle$ of the noninteracting fields. However it is no longer legitimate to assume that the total number of photons radiated is finite, and so we cannot restrict the states to belong to the Fock space.

Instead we shall work in the space $\mathcal{H}_{IR}$ spanned by the generalized coherent states, and consider the scattering matrix elements

$$\langle f,\lambda | S(J) | g,\mu \rangle = \langle f,\lambda;\text{out} | g,\mu;\text{in} \rangle_J . \qquad \text{(III.14)}$$

Formally the calculation proceeds exactly as before, and we find that a solution of the functional differential equation satisfied by this matrix element is given by

$$S(J) = U(j)\, V(\sigma) \qquad \text{(III.15)}$$

where, as before,

$$j_\mu(\underset{\sim}{k}) = i\gamma_{\mu\nu}(k)\, J^\nu(k) \Big|_{k^\circ = |\underset{\sim}{k}|} \qquad \text{(III.16)}$$

and formally

$$V(\sigma) = \exp \tfrac{1}{2} i \int \frac{d^4k}{(2\pi)^4} \frac{J_\mu(k)^* J^\mu(k)}{k^2} \qquad \text{(III.17)}$$

Of course $V(\sigma)$ is not in general of the form $e^{i\sigma}$, because this integral may diverge, but it nevertheless has a well-defined meaning as an operator on $\mathcal{H}_{IR}$.

This solution makes sense only if the functions j and σ belong to the classes we have defined. However, whereas before we required $J_\mu(x)$ to vanish for large times, we now need only a much less stringent condition. Actually it is still not possible to choose a current corresponding asymptotically to a classical point particle of definite momentum, but it is sufficient to integrate over a small range of values of the momentum.

Since the operators $U(j)$ and $V(\sigma)$ are unitary operators on $\mathcal{H}_{IR}$, the operator $S(J)$ is clearly unitary. We can prove from the multiplication laws (III.8) that

$$S(J) | g,\mu \rangle = \Big| g + j,\ \mu + \sigma + \frac{g^* j - j^* g}{2i} \Big\rangle \quad . \qquad \text{(III.18)}$$

It is not difficult to show that the unitary operators U and V map equivalent vectors onto equivalent vectors, that is to say if

$$| f,\lambda \rangle \sim | g,\mu \rangle$$

then also

$$S(J) | f,\lambda \rangle \sim S(J) | g,\mu \rangle \quad .$$

Thus the operator $S(J)$ maps each separable subspace of $\mathcal{H}_{IR}$ spanned by an equivalence class of generalized coherent states onto another

such subspace, though not in general the same one. In particular, if the initial state belongs to the equivalence class of the vacuum, that is to say if it is in the Fock space, then the final state belongs to the same equivalence class as the state $|j,\sigma\rangle$.

We conclude therefore that by choosing asymptotic states from the space $\mathcal{K}_{IR}$ rather than from the Fock space we are enabled to define a unitary scattering operator even in the presence of currents which do not vanish for large times. When interpreted according to the rules we have given, all the matrix elements of this operator are finite. For a given initial state, they all vanish except for those to final states in some definite equivalence class, although not the same class as that of the initial state. Finally we note that any of these generalized coherent states may be produced from the vacuum by some appropriately chosen current.

E. Momentum Spectrum of Soft-Photon Coherent States

It will be convenient to introduce a purely conventional separation between hard photons and soft photons. We divide the four-dimensional k space into complementary regions Ω^h and Ω^s. The soft-photon region Ω^s is defined by the conditions

$$|\underset{\sim}{k}| < K, \quad |k^0| < K^0 \tag{III.19}$$

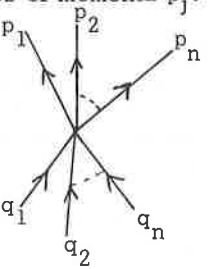
where K and K^0 are constants chosen to satisfy

$$K << K^0 << m$$

but otherwise arbitrary. The purpose of the first inequality is purely technical. It will allow us to extend certain integrals over k^0 to the entire real axis and hence evaluate them by contour integrations.

Let us now examine a specific class of generalize
states which will play an important role in our later work.

To introduce these states let us consider the classical current corresponding to n incoming particles of momenta q_j which scatter at $x = 0$ into n outgoing particles of momenta p_j. The corresponding



current density is

$$j^{\mu}(x) = \sum_j \int_0^{\infty} d\tau\, e \frac{p_j^{\mu}}{m} \delta_4 \left(x - \frac{p_j}{m}\tau\right) + \sum_j \int_{-\infty}^{0} d\tau\, e \frac{q_j^{\mu}}{m} \delta_4 \left(x - \frac{q_j}{m}\tau\right)$$

or, in momentum space

$$J^{\mu}(k) = -ie \sum_j \frac{p_j^{\mu}}{p_j \cdot k - i\epsilon} + ie \sum_j \frac{q_j^{\mu}}{q_j \cdot k + i\epsilon} \quad . \tag{III.20}$$

We note that this current satisfies the continuity equation

$$k_{\mu} J^{\mu}(k) = 0 \ .$$

The hard photon emission in a scattering process depends on the details of the scattering but the soft-photon emission does not. Thus for the soft photons it should be sufficient to consider this specific case. Since we are interested only in the soft photons we shall set

$$J^{\mu}(k) = 0 \qquad \text{for } k \in \Omega^h \ .$$

As we have already remarked the current $J^{\mu}(k)$ actually does not belong to the class of currents we have considered. The reason for this is closely connected with the infinite self-energy of a classical point particle. If we try to compute the phase σ, we obtain an integral involving double-pole terms of the form

$$e^2 \int_{\Omega^s} \frac{d^4k}{(2\pi)^4} \frac{p^2}{(p \cdot k + i\epsilon)(p \cdot k - i\epsilon)} \frac{1}{k^2} \ .$$

This is unacceptable because it is not singular only at $k = 0$ but for other values of k also. The k^0 contour is pinched between coincident poles, and gives an infinite result even for $\underset{\sim}{k} \neq \underset{\sim}{0}$. This difficulty can be avoided by spreading out the point particles, integrating over a range of values of each p_j and q_j. However, it will turn out that in the full quantum theory, these infinities are removed in any case (essentially by mass renormalization), so we shall ignore the undefined phase σ and concentrate on the photon wave function

$$j_\mu(\underset{\sim}{k}) = e\,\gamma_{\mu\nu}(\underset{\sim}{k})\left(\sum_{j=1}^{n}\frac{p_j^{\nu}}{p_j\cdot k} - \sum_{j}\frac{q_j^{\nu}}{q_j\cdot k}\right) \qquad \text{(III.21)}$$

In particular let us examine the momentum spectrum of the state $|j,\sigma\rangle$, which is actually independent of σ.

In general, the behaviour of a coherent state under translations is given by

$$e^{-iP\cdot x}\,|j,\sigma\rangle = |(x)\,j,\sigma\rangle \qquad \text{(III.22)}$$

where $(x)j$ is the translated wave function defined by

$$(x)j^{\mu}(\underset{\sim}{k}) = j^{\mu}(\underset{\sim}{k})\,e^{-ik\cdot x} \quad . \qquad \text{(III.23)}$$

The momentum spectrum of the state is characterized by the distribution function

$$\rho(k) = \int d^4x\, e^{-ik\cdot x}\,\langle j,\sigma|\,e^{iP\cdot x}\,|j,\sigma\rangle$$

$$= \int d^4x\, e^{-ik\cdot x}\,\langle (x)j,\sigma\,|\,j,\sigma\rangle \quad .$$

Thus using the scalar product formula,

$$\rho(k) = \int d^4x\, e^{-ik\cdot x}\, e^{L(x)} \qquad \text{(III.24)}$$

where

$$L(x) = (x)j^{*}j - \tfrac{1}{2}(x)j^{*}(x)j - \tfrac{1}{2}j^{*}j$$

$$= (x)j^{*}j - j^{*}j \quad . \qquad \text{(III.25)}$$

As we noted above, this distribution is independent of the phase σ. We also may easily verify that $\rho(k)$ is positive definite, vanishes outside the forward light cone, and satisfies the normalization condition

$$\int \frac{d^4k}{(2\pi)^4}\,\rho(k) = 1 \quad .$$

We now return to the specific state described the wave function j of (III.21), and examine the behaviour of $\rho(k)$ for small values of k.

It is convenient to change the notation slightly. We may treat incoming and outgoing particles together if we define $p_{n+1}\ldots p_{2n} = -q_{n+1}\ldots -q_{2n}$ and assign charges $e_j = e\,(j=1\ldots n)$ and $e_j = -e$ $(j=n+1\ldots 2n)$. Then

$$L(x) = \sum_{i,j=1}^{2n} e_i e_j \int_{\Omega_s} \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|} \frac{p_i^{\mu}\gamma_{\mu\nu}(k)p_j^{\nu}}{(p_i\cdot k)(p_j\cdot k)} \left(e^{ik\cdot x} - 1\right)$$

$$= \sum_{i,j=1}^{2n} L_{ij}(x) \quad , \tag{III. 6}$$

say.

It is easy to see that L(x) itself is gauge-invariant. However the separation into terms L_{ij} depends on the choice of gauge.

The integral is convergent at $k = \underset{\sim}{0}$ because of the factor $e^{ik\cdot x}-1$ which vanishes there. It behaves like

$$\int_0^K \frac{dk}{k} \left(e^{ikr} - 1\right) \sim -\ell n(Kr) + \text{constant}.$$

In fact by explicit calculation we find that

$$L(x) \approx \tfrac{1}{2}\xi\, \ell n(K^2x^2) + f(\underset{\sim}{x}/x^0) \tag{III.27}$$

where the function f depends only on the direction of the four-vector x, not on its magnitude. The parameter ξ may be written

$$\xi = \sum_{i,j=1}^{2n} \xi_{ij} \quad . \tag{III.28}$$

Although ξ is gauge invariant, the individual parameters ξ_{ij} are not. We find in the Lorentz gauge

$$\xi_{ij}^{L} = \frac{e_i e_j}{4\pi^2}\, \Phi(u_{ij}) \tag{III.29}$$

where u_{ij} is the relative velocity of particles i and j,

$$u_{ij} = \left(1 - \frac{m^4}{(p_i\cdot p_j)^2}\right)^{\frac{1}{2}}$$

and

$$\Phi(u) = \frac{1}{2u}\, \ell n\, \frac{1+u}{1-u} \quad .$$

In the radiation gauge, on the other hand

$$\xi_{ij}^{R} = \frac{e_i e_j}{4\pi^2}\left[\Phi(u_{ij}) - \Phi(u_i) - \Phi(u_j) + 1\right] \qquad \text{(III.30)}$$

where $u_i = |\underline{p}_i/p_i^\circ|$. It is easy to check the gauge invariance directly, using the fact that

$$\sum_j \frac{e_i e_j}{4\pi^2}\,\Phi(u_i) = \frac{e_i}{4\pi^2}\,\Phi(u_i)\sum e_j = 0 \quad .$$

We see that in the radiation gauge $\xi_{ij}^{R} = 0$ when either $u_i = 0$ or $u_j = 0$. A particle at rest does not contribute to the radiation in this gauge.

The asymptotic behaviour of $L(x)$ determines the nature of the singularity of $\rho(k)$ at $k = 0$. We find that as $k \to 0$ from any direction within the forward light cone,

$$\rho(k) = \frac{1}{(-k^2)^2}\left(\frac{-k^2}{K^2}\right)^{\frac{1}{2}\xi} g(\underline{k}/k^\circ) \qquad \text{(III.31)}$$

where g is again a function only of the direction of k. It may be shown that ξ is always negative, so that ρ has an integrable singularity. It is possible to obtain an explicit formula for the energy spectrum. The probability that the total energy is less than $E(\leq K)$ is

$$P(E) = \int \frac{dk}{(2\pi)^4}\,\rho(k)\,\theta(E - k^\circ) = \frac{e^{\gamma\xi}}{(-\xi)!}\left(\frac{E}{K}\right)^{-\xi}, \qquad \text{(III.32)}$$

where γ is Euler's constant and $(-\xi)! = \Gamma(1 - \xi)$.

IV. Mass-shell Singularities of Green's Functions

We now return to the fully quantized theory of interacting electromagnetic and Dirac fields. In order to be able to evaluate scattering matrix elements, we first have to determine the appropriate space of asymptotic states, without making the a priori assumption that these states must contain finitely many particles. To do this we begin by studying the singularity structure of the Green's functions of the theory.

A. Separation of Soft-Photon Contributions

The propagator function corresponding to each internal photon line may be written as a sum of two parts

$$D_{\mu\nu}^{h,s}(x) = \int_{\Omega^{h,s}} \frac{d^4k}{(2\pi)^4}\,\frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon}\,e^{ik\cdot x} \quad .$$

It may be noted that $D^s_{\mu\nu}(x)$, being the Fourier transform of a distribution which vanishes outside a finite region of k space, is an entire analytic function of x, and thus in particular has no singularity on the light cone $x^2 = 0$. This may be confirmed by explicit evaluation of the integral.

We shall separate the contributions to each internal photon line from $D^h_{\mu\nu}$ and $D^s_{\mu\nu}$, and regard them as contributions to distinct Feynman diagrams, in which the internal photon lines have an extra label, "hard" or "soft."

We know that when k^2 is close to zero, the only effect of adding all radiative corrections to an internal photon line is to multiply the propagator functions by the overall factor Z_3, which is absorbed by charge renormalization at the vertices to which the line is attached. Within Ω^s, k^2 is always small. Thus if we work in terms of the renormalized charge then we need not include any radiative corrections to the internal soft-photon lines.

The general Green's function of the theory has the form

$$G_{\mu_1\ldots\mu_m}(p_1\ldots p_\ell, q_1\ldots q_\ell, k_1\ldots k_m)$$

$$= \int dz_1\ldots dz_m \exp(-i\sum_j k_j\cdot z_j)\ (-i)^m \frac{\delta}{\delta J^{\mu_1}(z_1)}\ldots\frac{\delta}{\delta J^{\mu_m}(z_m)}$$

$$G(p_1\ldots p_\ell, q_1\ldots q_\ell | J)\Big|_{J=0} \qquad \text{(IV.1)}$$

where $G(p_1\ldots p_\ell, q_1\ldots q_\ell | J)$ is the Fourier transform of

$$\langle 0;\text{out}|\, T[\psi(x_1)\ldots\psi(x_\ell)\bar{\psi}(y_\ell)\ldots\bar{\psi}(y_1)]\,|0;\text{in}\rangle_J\ . \qquad \text{(IV.2)}$$

So long as we consider only currents $j^\mu(x)$ for which these vacuum-to-vacuum matrix elements can be expanded in a functional Taylor series in powers of J, then it is entirely a matter of convenience whether we choose to work with the functionals of J or with the Green's functions which are the coefficients in this expansion. However we shall wish to be able to consider external currents of the class discussed earlier, for which this series expansion is invalid. Hence for the soft-photon part of the electromagnetic field we retain the functional form. We shall consider the functions

$$G_{\mu_1\ldots\mu_m}(p_1\ldots p_\ell, q_1\ldots q_\ell, k_1\ldots k_m | J) \qquad \text{(IV.3)}$$

where $k_1 \ldots k_m$ are momenta in Ω^h, and J is a soft external current, that is to say one whose Fourier transform vanishes outside Ω^s.

We shall be interested in the behaviour of these functions for values of the momenta close to their mass shells,

$$p_j^2 \approx -m^2 , \quad q_j^2 \approx -m^2 , \quad k_j^2 \approx 0 .$$

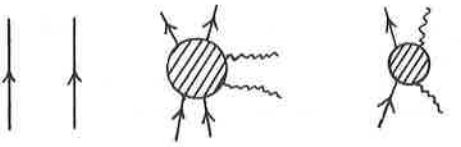
In any Feynman diagram which contributes to (IV.3) we may identify a core diagram obtained by discarding all internal soft-photon lines (and setting $J = 0$). We shall denote by G^h the sum of all contributions of core diagrams. Then the complete set of diagrams may be recovered by inserting soft-photon internal lines into the core diagrams in all possible ways. When we do this, however, we need not actually include all possible insertions because many of them yield contributions which are negligible when integrated over the small region Ω^s. If both ends of an internal soft-photon line are attached to internal lines of the core diagrams, we obtain an integral which behaves like $\int d^4k/k^2$ and is therefore proportional to K^2 and thus negligible. If one end is attached to an external line, then we introduce an extra propagator with momentum close to its mass shell. As $p^2 \to -m^2$, the propagator develops a $1/k$ singularity. Such contributions are still proportional to K in the limit and may therefore be neglected. Only if both ends are attached to external lines do we obtain a contribution which diverges in the limit as the momenta approach their mass shells and which cannot be neglected. In the case of interactions with the external current J, the role of one of the propagator functions is played by the Fourier transform of the current, $J^\mu(k)$. Thus if this function is no more singular at $k = 0$ than $1/k$ (as we shall assume) it again follows that we need only consider insertions in the external lines of the core diagrams.

B. Structure of the Core Diagrams

We begin by examining the core diagrams. It is convenient to separate them into classes according to their connectivity structure. Symbolically we write

$$G^h = \sum \prod_\alpha (2\pi)^4 \, \delta_4(P_\alpha) \, G_\alpha^{h,c} , \qquad \text{(IV.4)}$$

where each term represents the contributions of diagrams with one particular connectivity structure, and α labels the connected pieces. The momentum P_α is the sum of the external momenta attached to the αth connected piece, and $G_\alpha^{h,c}$ is the sum of the contribution of all connected core diagrams for this piece. A typical term in the sum is illustrated in the figure:



Next we isolate the propagator factors associated with the external lines. Since we are interested in values of the momenta near their mass shells we may write

$$S'^{h}(p) = \frac{Z_2^h(p)\,\Lambda(p)}{p^2 + m^2 - i\epsilon} \quad , \tag{IV.5}$$

where the superscript h indicates that the function includes the contribution of all hard-photon radiative corrections, but not of soft-photon radiative corrections. The function $Z_2^h(p)$ the wave function renormalization constant with soft-photon contributions removed contains the entire ultraviolet-divergent part of Z_2, but is free of infrared divergences. It depends on p even in a covariant gauge, because of the noncovariance associated with the separation of hard and soft photons.

For photon propagators, we may write

$$D'^{h}_{\mu\nu}(k) = D'_{\mu\nu}(k) = \frac{Z_3}{k^2 - i\epsilon} \quad , \tag{IV.6}$$

because Z_3 is not infrared divergent.

For straight-through lines we have a single factor $-iS'^{h}(p)$ or $-iD'^{h}_{\mu\nu}(k)$. For other connected pieces, we have one such factor for each external line, and a function M_α^h which represents the contributions of the connected core diagrams with external lines removed. For such pieces, one half of the wave function renormalization factor may be absorbed in the corresponding function M_α^h, to cancel the vertex renormalization function, or to contribute to the charge renormalization. The renormalized functions M_α^h including these factors should be finite, when expressed in terms of the physical charge e. For each external line we are left with a factor

$$\frac{[Z_2^h(p)]^{\frac{1}{2}}\Lambda(p)}{p^2 + m^2 - i\epsilon} \qquad \text{or} \qquad \frac{Z_3^{\frac{1}{2}}\gamma_{\mu\nu}(k)}{k^2 - i\epsilon} \quad .$$

It may be noted that in every case we have precisely one factor $Z^{\frac{1}{2}}$ for each external momentum, including those attached to straight-through lines. It will be convenient to write G^h in a form in which

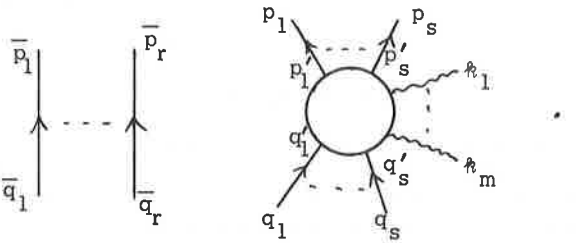
the mass-shell poles are explicitly separated from the remainder. Let us define the function

$$\Delta^\circ(p_1 \ldots p_n, q_1 \ldots q_n) = \prod_{j=1}^{n} \Delta^\circ(p_j, q_j) \tag{IV.7}$$

with

$$\Delta^\circ(p,q) = -i(2\pi)^4 \, \delta_4(p-q) \, \frac{1}{p^2 + m^2 - i\epsilon} \ .$$

To simplify the notation let us assume that in the region of interest the only core diagrams which contribute significantly are those of one particular structure, with r straight-through lines connecting $\bar{p}_1 \ldots \bar{p}_r$ to $\bar{q}_1 \ldots \bar{q}_r$, and otherwise connected, as indicated in the diagram



Here the primed variables are dummy variables introduced to allow us to write

$$G^h(\bar{p}\,p, \bar{q}\,q, k) = \prod \{Z_2^h(\bar{p}_j)\,\Lambda(\bar{p}_j)\}$$

$$\prod \{[Z_2^h(p_j)]^{\frac{1}{2}}\,\Lambda(p_j)\} \prod \left\{\frac{-iZ_3^{\frac{1}{2}}\,\gamma(k_j)}{k_j^2 - i\epsilon}\right\}$$

$$\int \prod \frac{dp_j'}{(2\pi)^4} \, \frac{dq_j'}{(2\pi)^4} \, \Delta^\circ(\bar{p}\,p\,q', \bar{q}\,p'q)$$

$$(2\pi)^4 \, \delta_4(\Sigma p' - \Sigma q' + \Sigma k) M^h(p'\,q'\,k)$$

$$\prod \{[Z_2^h(q_j)]^{\frac{1}{2}}\,\Lambda(q_j)\} \tag{IV.8}$$

where p stands for $p_1 \ldots p_s$, and so on, and we have suppressed the vector indices. As we noted before, the function M^h is free of divergences and finite on the mass shells. The purpose of writing G^h in

this form is that we shall be able to show that the only effect of including soft-photon corrections is to modify the function Δ°.

C. Insertion of Soft-Photon Contributions

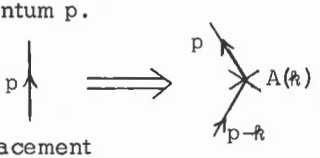
We now have to insert all possible soft-photon internal lines into the external lines of the core diagrams, and include also the effect of the external current J.

To do this it is convenient to use a functional technique. We introduce a classical external potential $A_{\mu}(x)$, with Fourier transform vanishing outside Ω^{s}, and first evaluate the Green's function $G^{h}(\ldots|A)$ in the presence of this external potential. To do this we have to insert vertices at which A appears in all possible ways in the external lines of the diagrams. Then the required Green's function with soft-photon contributions is obtained by acting with a functional differential operator,

$$G(\ldots|J) = \exp\left\{-\tfrac{1}{2}i\int d^4x d^4y \frac{\delta}{\delta A_{\mu}(x)} D^{s}_{\mu\nu}(x-y) \frac{\delta}{\delta A_{\nu}(y)}\right\}$$
$$\exp\{i\int d^4x\, J^{\mu}(x) A_{\mu}(x)\}\, G^{h}(\ldots|A)\Big|_{A=0} . \qquad \text{(IV.9)}$$

In diagrammatic language, the factor $\exp(i\int JA)$ introduces vertices at which the external current J interacts with the external potential A. The functional differential operator represented by the first factor then connects these, and the vertices inserted in the external lines, by soft-photon internal lines in all possible ways.

To evaluate $G^{h}(\ldots|A)$ we first consider the insertion of a vertex at which $A(k)$ appears, with small momentum transfer k, into an external line of momentum p.



The effect is the replacement

$$\frac{\Lambda(p)}{m^2+p^2-i\varepsilon} \Longrightarrow \frac{\Lambda(p)}{m^2+p^2-i\varepsilon}\, e\gamma_{\mu} A^{\mu}(k) \frac{\Lambda(p-k)}{m^2+(p-k)^2-i\varepsilon} .$$

For small k, we may neglect the quadratic term in k in the denominator and the linear term in the numerator, so that we have to consider

$$\frac{m-\not{p}}{m^2+p^2-i\varepsilon}\, e\gamma_{\mu} A^{\mu}(k) \frac{m-\not{p}}{m^2+p^2-2p\cdot k-i\varepsilon} .$$

Now using the anticommutation rules of Dirac matrices,

$$(m - \not{p})\gamma_\mu (m - \not{p}) = (m - \not{p})[(m + \not{p})\gamma_\mu + 2p_\mu]$$
$$= (m^2 + p^2)\gamma_\mu + 2p_\mu (m - \not{p}) \quad .$$

The first term gives a contribution which is less singular in the limit as the momentum approaches its mass shell, and thus may be neglected. Thus we obtain

$$\frac{\Lambda(p)}{m^2 + p^2 - i\varepsilon} \quad \frac{2ep\cdot A(k)}{m^2 + p^2 - 2p\cdot k - i\varepsilon} \quad . \qquad \text{(IV.10)}$$

In fact, the insertion of a soft-photon vertex in a charged-particle line of any spin has the effect of multiplying the propagator function by this same factor independent of the spin of the particle involved.

Now let us consider the sum of all such insertions in an external line:



The effect is to replace

$$-i(2\pi)^4 \delta_4(p - q) \frac{\Lambda(p)}{m^2 + p^2 - i\varepsilon} = \Lambda(p)\, \Delta^\circ(p,q)$$

by $\Lambda(p)\Delta^\circ(p,q|A)$ where

$$\Delta^\circ(p,q|A) = -i\sum_{n=0}^{\infty} \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_n}{(2\pi)^4} (2\pi)^4 \delta\Big(p - q - \sum_{j=1}^{n} k_j\Big)$$

$$\frac{1}{m^2 + p^2 - i\varepsilon} \prod_{r=1}^{n} \frac{2ep\cdot A(k_r)}{m^2 + p^2 - 2p\cdot k_1 - \ldots - 2p\cdot k_r - i\varepsilon} \quad .$$

We now insert a Fourier integral representation for the delta function,

$$\int d^4y \exp\Big[-i(p - q)\cdot y + i\sum_j k_j \cdot y\Big]$$

and an integral representation for each denominator,

$$\frac{1}{m^2 + p^2 - i\varepsilon} = i \int_0^\infty d\tau \; e^{-i\tau (m^2 + p^2)} \; .$$

Thus we obtain in the integrand an exponential factor

$$\exp \{ -i \sum_{r=0}^{n} \tau_r (m^2 + p^2 - 2p \cdot k_1 - \ldots - 2p \cdot k_r) \} \; .$$

Next we make the change of variables to

$$\sigma_r = \sum_{s=r}^{n} \tau_s$$

so that this part of the integrand becomes

$$\int_0^\infty d\sigma_0 \int_0^{\sigma_0} d\sigma_1 \cdots \int_0^{\sigma_{n-1}} d\sigma_n \exp \{ -i\sigma_0 (m^2 + p^2) + 2i \sum_{r=1}^{n} \sigma_r p \cdot k_r \}$$

or, symmetrizing in these n variables,

$$\frac{1}{n!} \int_0^\infty d\sigma_0 \int_0^{\sigma_0} d\sigma_1 \cdots \int_0^{\sigma_0} d\sigma_n \exp \{ - i \sigma_0 (m^2 + p^2) + 2i \sum_{r=1}^{n} \sigma_r p \cdot k_r \} \; .$$

Then we can sum the series to an exponential form and obtain

$$\Delta^\circ (p, q | A) = \int d^4 y \; e^{-i(p-q) \cdot y} \int_0^\infty d\sigma_0 \; e^{-i\sigma_0 (m^2 + p^2)}$$

$$\exp \{ i \int \frac{d^4 k}{(2\pi)^4} \; 2ep \cdot A(k) \; e^{ik \cdot y} \int_0^{\sigma_0} d\sigma' \; e^{2i\sigma' p \cdot k} \}$$

$$= \int d^4 y \; e^{-i(p-q) \cdot y} \int_0^\infty d\sigma \; e^{-i\sigma (m^2 + p^2)}$$

$$\exp \{ \int \frac{d^4 k}{(2\pi)^4} \; \frac{ep \cdot A(k)}{p \cdot k} \; e^{ik \cdot y} \; (e^{2i\sigma p \cdot k} - 1) \} \; .$$

It should be noted that despite the denominator $p \cdot k$ the integral in the exponent is convergent because of the factor $(e^{2i\sigma p \cdot k} - 1)$.

We could also have expressed each Feynman denominator in terms of q rather than p. Then we would have obtained essentially

the same expression, but with p replaced by -q in the k integral, and e replaced by -e.

This same modification is required for each external or straight-through line. Thus we conclude that to obtain $G^h(\ldots|A)$ from $G^h(\ldots)$ we have to replace Δ° by

$$\Delta^\circ(p_1\ldots p_n, q_1\ldots q_n|A) = \prod_{j=1}^{n} \Delta^\circ(p_j, q_j|A) \quad .$$

Because the dependence on A is purely in the exponent, we can easily perform the functional differentiations indicated in (IV.9), using the identity

$$\exp\{-\tfrac{1}{2}i\int \frac{\delta}{\delta A} D \frac{\delta}{\delta A}\} \exp\{i\int JA\}\Big|_{A=0}$$

$$= \exp\{\tfrac{1}{2}i\int J D J\} \quad .$$

Thus we find that $G(\ldots|J)$ has the same structure as $G^h(\ldots)$ but with the function $\Delta^\circ(\ldots)$ replaced by

$$\Delta^s(p_1\ldots p_n, q_1\ldots q_n|J) = \int d^4y_1\ldots d^4y_n \int_0^\infty d\sigma_1\ldots d\sigma_n$$

$$\exp\{-i\sum_j (p_j - q_j)\cdot y_j - i\sum_j \sigma_j(m^2 + p_j^2)\} \exp X \qquad \text{(IV.11)}$$

where

$$X = \tfrac{1}{2} i \int_\Omega^s \frac{d^4k}{(2\pi)^4} I^\mu(k)^* \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon} I^\nu(k) \qquad \text{(IV.12)}$$

and

$$I^\mu(k) = J^\mu(k) + i\sum_{j=1} \frac{e p_j^\mu}{p_j\cdot k} e^{-ik\cdot y_j} (e^{-2i\sigma_j p_j\cdot k} - 1) \; . \qquad \text{(IV.13)}$$

This function Δ^s has a singularity in each momentum variable p_j at the mass shell $p_j^2 = -m^2$, which is a modified form of the mass-shell pole of Δ°. This singularity is governed by the asymptotic behaviour of the integrand for large σ_j. It also has a singularity near $p_j = q_j$, which is a modified form of the delta-function singularity of Δ°, and which is governed by the asymptotic behaviour for large y_j.

In each term of the expansion, the integration over y_j sets $p_j - q_j$ equal to a sum of soft-photon momenta. Thus Δ^s falls off rapidly with increasing values of $p_j - q_j$.

D. Mass-shell Singularities

Let us now consider specifically the mass-shell singularities of Δ^s. For simplicity, we shall set $J = 0$. To investigate the structure of these singularities, we have to examine the asymptotic behaviour of the exponent X for large σ_j. Let us concentrate on the term which arises from the contribution of a specific part of particles, namely

$$X_{ij} = \tfrac{1}{2} i e^2 \int_{\Omega^s} \frac{d^4k}{(2\pi)^4} \frac{p_i^{\mu} \gamma_{\mu\nu}(k) p_j^{\nu}}{(k^2 - i\epsilon)}$$

$$e^{ik\cdot(y_i - y_j)} \left(\frac{e^{2i\sigma_i p_i \cdot k} - 1}{p_i \cdot k} \right) \left(\frac{e^{2i\sigma_j p_j \cdot k} - 1}{p_j \cdot k} \right) \quad \text{(IV.14)}$$

The case $i = j$ will be considered separately later.

It is easy to see that the behaviour of this function depends on the order in which we allow σ_i and σ_j to approach infinity. Qualitatively, its behaviour is similar to that of the function

$$\ln \frac{Km(\sigma_i + \sigma_j)}{\sigma_i \sigma_j} \quad .$$

Suppose we first let $\sigma_j \to \infty$. Then the function remains finite. If we later let $\sigma_i \to \infty$ it diverges logarithmically. The effect of allowing σ_j to approach infinity is described by the replacement

$$\frac{e^{-2i\sigma_j p_j \cdot k} - 1}{p_j \cdot k} \longrightarrow - \frac{1}{p_j \cdot k - i\epsilon} \quad .$$

It is clear that we may set $y_i - y_j = 0$. For if we write

$$e^{ik\cdot y} = (e^{ik\cdot y} - 1) + 1 \quad ,$$

then the first term vanishes at $k = 0$ and thus yields a contribution which remains finite in the limit. Thus we have to consider

$$X_{ij}(\sigma_i,\infty) = -\tfrac{1}{2}ie^2 \int_\Omega s \frac{d^4k}{(2\pi)^4} \frac{p_i^\mu \gamma_{\mu\nu}(k) p_j^\nu}{(p_i\cdot k)(k^2-i\epsilon)(p_j\cdot k - i\epsilon)} (e^{2i\sigma_i p_i\cdot k} - 1) \quad \text{(IV.15)}$$

The behaviour of this function depends on the relative signs of p_i° and p_j°, that is to say on whether we are interested in the singularity at the positive-energy or negative-energy mass shell. Because of symmetry there is no essential loss of generality in assuming that $p_i^\circ > 0$. Then we may complete the k° contour in the lower half k° plane, and use contour integration.

In general we obtain two terms. The first arises from the pole at $k^\circ = |\underset{\sim}{k}|$, and is

$$X_{ij}^{(1)}(\sigma_i,\infty) = +\tfrac{1}{2}e^2 \int_\Omega s \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|} \frac{p_i^\mu \gamma_{\mu\nu} p_j^\nu}{(p_i\cdot k)(p_j\cdot k)} (e^{2i\sigma_i p_i\cdot k} - 1) \quad . \text{(IV.16)}$$

This is precisely the function $\frac{1}{2}L_{ij}(x)$ which we considered earlier, evaluated for $x = 2\sigma_i p_i$. Thus its asymptotic behaviour is given by

$$X_{ij}^{(1)}(\sigma_i,\infty) \approx \tfrac{1}{2}\xi_{ij} \ln(\mathrm{Km}\ \sigma_i) + \text{constant},$$

as $\sigma_i \to \infty$ where ξ_{ij} is the parameter introduced earlier.

The second term comes from the pole at

$$p_j^\circ k^\circ = \underset{\sim}{p_j}\cdot\underset{\sim}{k} - i\epsilon \quad .$$

It lies below the real axis, and therefore inside the contour if $p_j^\circ > 0$, but is excluded if $p_j^\circ < 0$. Hence we obtain the contribution

$$X_{ij}^{(2)}(\sigma_i,\infty) = \theta(p_i^\circ p_j^\circ)\tfrac{1}{2}e^2 \int_\Omega s \frac{d^4k}{(2\pi)^4} \frac{p_i^\mu \gamma_{\mu\nu}(k) p_j^\nu}{(p_i\cdot k)(k^2)} 2\pi\delta(p_j\cdot k)(e^{2i\sigma_i p_i\cdot k} - 1) \quad \text{(IV.17)}$$

By making the transformation $k \to -k$, it is easy to see that this term is actually purely imaginary. It also behaves logarithmically as $\sigma_i \to \infty$, and by explicit calculation we find

$$X_{ij}^{(2)}(\sigma_i,\infty) \approx -\tfrac{1}{2}i\theta(p_i^\circ p_j^\circ)\eta_{ij} \ln(\mathrm{Km}_i\sigma_i) + \text{constant}$$

The parameter η_{ij}, unlike ξ_{ij}, is gauge-invariant, because gauge terms in $\gamma_{\mu\nu}$ proportional to k_ν do not contribute while those proportional to k_μ contribute only finite terms. We find

$$\eta_{ij} = \frac{e^2}{4\pi u_{ij}} \quad .$$

We note that, also unlike ξ_{ij}, η_{ij} diverges as the relative velocity approaches zero. This divergence is closely related to the divergent phase of the nonrelativistic Coulomb scattering amplitude. It also indicates the need for special treatment of the case $i = j$.

We may write in general

$$X_{ij}(\sigma_i, \infty) \approx \tfrac{1}{2} \zeta_{ij} \ell n (Km\sigma_i) + \text{constant} \qquad \text{(IV.18)}$$

where the complete parameter ζ_{ij} is given by

$$\zeta_{ij} = \xi_{ij} - i\theta(p_i^\circ p_j^\circ)\, \eta_{ij} \, . \qquad \text{(IV.19)}$$

We have still to consider the special case $i = j$. Here of course we cannot let σ_i and σ_j tend to infinity successively. The function we have to consider is

$$X_{ii}(\sigma_i) = \tfrac{1}{2} ie^2 \int_{\Omega_S} \frac{d^4 k}{(2\pi)^4} \frac{p_i^\mu \gamma_{\mu\nu}(k) p_i^\nu}{(p_i \cdot k)^2 (k^2 - i\varepsilon)} (e^{2i\sigma_i p_i \cdot k} - 1)(e^{-2i\sigma_i p_i \cdot k} - 1) \, . \qquad \text{(IV.20)}$$

If we distort the k° contour to pass around the point where $p_i \cdot k = 0$, we may separate the various terms. Then making the transformation $k \to -k$ in two of the four terms we obtain

$$X_{ij}(\sigma_i) = \tfrac{1}{2} ie^2 \int_{\Omega_S} \frac{d^4 k}{(2\pi)^4} \frac{p_i^\mu \gamma_{\mu\nu}(k) p_i^\nu}{k^2 - i\varepsilon} (e^{2i\sigma_i p_i \cdot k} - 1) \left\{ \frac{1}{(p_i \cdot k + i\varepsilon)^2} + \frac{1}{(p_i \cdot k - i\varepsilon)^2} \right\} \, .$$

As before we may now complete the k° contour in the lower half plane, and obtain two contributions. The first, from the pole at $k^\circ = |\underset{\sim}{k}|$ has exactly the same form as before but with an extra factor of 2 whose origin may be traced to the presence of two nonoscillatory terms in the integrand rather than merely one.

The second term arises from the double pole at $p_i \cdot k = 0$. It can be obtained correctly from $X_{ij}^{(2)}$ by taking the limit $p_j \to p_i$. Then we have in the integrand the factor

$$\lim_{p_i\cdot k \to 0} \frac{e^{2i\sigma_i p_i\cdot k} - 1}{p_i\cdot k} = 2i\sigma_i \quad .$$

Hence the dependence of this term on σ_i is now linear. It is in fact easy to evaluate it explicitly and we obtain

$$X_{ii}^{(2)}(\sigma_i) = -i\frac{e^2 m}{2\pi^2 p_i^0}\,\Phi(u_i)(Km\sigma_i) \quad .$$

Formally this term represents the soft-photon contribution to the mass renormalization, as may be seen from the fact that its effect would be to shift the position of the mass-shell singularity away from $p_i^2 = -m^2$. However, it is in any case proportional to K, and therefore negligible, as it must be since the mass renormalization constant is free of infrared divergences.

Dropping this term, we are left with

$$X_{ii}(\sigma_i) \approx \xi_{ii}\,\ell n(Km\sigma_i) + \text{constant} \quad . \tag{IV.21}$$

It is now straightforward to substitute this asymptotic behaviour into our expression for Δ^S, and obtain the mass-shell singularity structure. Let us consider first the singularity as $p_1^2 \to -m_1^2$. This is governed by the asymptotic behaviour of the integrand as $\sigma_1 \to \infty$ with all other σ_j held fixed. The only term in the exponent X which does not remain finite is X_{11}. Thus the relevant part of the integral is

$$\int_0^\infty d\sigma_1\, e^{-i\sigma_1(m^2+p^2)}\, e^{\xi_{11}\ell n(Km\sigma_1)}$$
$$= \int_0^\infty d\sigma_1\, e^{-i\sigma_1(m^2+p^2)}\,(Km\,\sigma_1)^{\xi_{11}} \quad .$$

This is a standard integral, namely

$$\frac{\xi_{11}!}{mK}\left(\frac{-imK}{m^2+p_1^2-i\varepsilon}\right)^{1+\xi_{11}} \quad . \tag{IV.22}$$

Thus Δ^S has a branch point on the mass shell. In the radiation gauge ξ_{11} is negative, and the singularity is less severe than a pole, which as we shall see is necessary on physical grounds in any physical gauge. (In the Lorentz gauge on the other hand, the singularity is more severe than a simple pole.)

Now let us isolate the coefficient of this branch-point singularity p_1, and then consider the limit as $p_2^2 \to -m_2^2$. This time we have two terms to consider, X_{22} and X_{12}, and the structure we obtain is

$$\frac{(\xi_{22} + \zeta_{12})!}{mK} \left(\frac{-imK}{m^2 + p_2^2 - i\varepsilon} \right)^{1+\xi_{22}+\zeta_{12}} . \qquad \text{(IV.23)}$$

Thus we see that the singularity in p_2 is changed by the fact that we have already gone to the mass-shell limit in p_1. This behaviour is quite different from that in a theory without massless particles, where the singularity is always a simple pole regardless of the order in which the limits are taken.

As we noted earlier, in the radiation gauge there is one special value of the momentum, namely $\underset{\sim}{p}_1 = \underset{\sim}{0}$ at which $\xi_{11} = 0$ and so Δ^s has a simple pole. On the other hand, if $\underset{\sim}{p}_2 = 0$ then although ξ_{22} and ξ_{12} both vanish the imaginary part of the exponent η_{12} need not, so that we get a form like

$$\left(\frac{-imK}{m^2 + p_2^2 - i\varepsilon} \right)^{1-i\eta_{12}} .$$

V. Asymptotic States

Now that we know the structure of the Green's functions, we may go on to investigate the nature of the asymptotic states of the theory. The major problem is that of defining states containing electrons and positrons, since the Green's functions for the corresponding fields do not have mass-shell poles in general. However, we shall find it convenient to begin by discussing the soft-photon asymptotic states, that is to say the states which can be produced from the vacuum by the action of operators constructed from the soft-photon part of the electromagnetic field, or equivalently by interaction with some soft-photon external current.

A. Soft-photon States

Let us consider first the special case of the Green's function $G(J)$ with no external lines,

$$G(J) = \langle 0;\text{out} \,|\, 0;\text{in} \rangle_J .$$

Since the only soft-photon insertions according to our rules are in external lines of core diagrams, there are no such contributions in the present case. The only contribution comes from soft-photon lines with both ends terminating at vertices where the current J appears, so that we obtain

$$\langle 0;\text{out}|\,0;\text{in}\rangle_J = \exp \tfrac{1}{2}i \int_{\Omega^S} \frac{d^4k}{(2\pi)^4}\, J^\mu(k)^*\, \frac{\gamma_{\mu\nu}(k)}{k^2 - i\epsilon}\, J^\nu(k) \ . \qquad (V.1)$$

This is of course precisely the expression we obtained earlier for the case of a free field in interaction with an external current. For this special case the interaction with the quantized electron field has no effect.

This result suggests that the structure of the states containing only soft photons should be identical with that in free-field case. In particular we should be able to define asymptotic soft-photon generalized coherent states $|f,\lambda;\text{in}\rangle$ and $|f,\lambda;\text{out}\rangle$.

Let us choose a photon wave function $f^\mu(\underset{\sim}{k})$ and consider a current

$$J_f^\mu(t,\underset{\sim}{k}) = -\chi(k)\, f^\mu(\underset{\sim}{k})\, e^{i(k^\circ - |\underset{\sim}{k}|)t} + i\chi(-k)^*\, f^\mu(-\underset{\sim}{k})^*\, e^{i(k^\circ + |\underset{\sim}{k}|)t} \ , \qquad (V.2)$$

where $\chi(k)$ is a bounded function which vanishes outside Ω^S and within it satisfies the conditions

$$\chi(k) = \begin{cases} 1 \text{ for } k^\circ = |\underset{\sim}{k}| \ , \\ 0 \text{ for } k^\circ = -|\underset{\sim}{k}| \ . \end{cases} \qquad (V.3)$$

In the free-field case this current acting on the vacuum produces a generalized coherent state $|f,\lambda_t\rangle$ for every value of t. The phase label λ_t depends on t, but tends to a definite limit as $t \to \pm\infty$, namely

$$\lambda = \lim_{t\to\pm\infty} \lambda_t = \int \frac{d^4k}{(2\pi)^4}\, \frac{|\chi(k)|^2}{k^2}\, f_\mu(\underset{\sim}{k})^*\, f^\mu(\underset{\sim}{k}) \ .$$

(As usual this integral may diverge at $k = 0$ and is then to be interpreted according to the rules introduced earlier.)

Now let us turn to the quantized theory and consider a Green's function of the form

$$\langle 0;\text{out}|\,T[\ldots]\,|0;\text{in}\rangle_{J+J_f(t)}$$

$$= \langle 0;\text{out}|\,T[\ldots \exp\{i\textstyle\int d^4x\, J_f^\mu(t,x)\, A_\mu(x)\}]\,|\,0;\text{in}\rangle_J \ .$$

As $t \to -\infty$, it follows from the structure of the current $J^\mu(t,x)$ that the values of x° which contribute here also tend to $-\infty$. Hence the

time-ordered exponential factor involving the current $J_f(t)$ may be taken outside the time-ordering symbol to the right and allowed to act directly on the vacuum. It should then produce a generalized coherent state. Hence it is natural to define the asymptotic generalized coherent state $|f,\lambda;in\rangle$ by the limiting formula

$$\langle 0;out|\,T[\ldots]\,|f,\lambda;in\rangle_J = \lim_{t\to-\infty} \langle 0;out|\,T[\ldots]\,|0;in\rangle_{J+J_f(t)} \qquad (V.4)$$

It is of course necessary to show that this limit exists, and depends on J_f only through the functions f and λ. But this is easy to do. The matrix element on the right is given by the general formula obtained earlier, but with the exponent X in the integrand replaced by

$$X(t) = \tfrac{1}{2}i \int \frac{d^4k}{(2\pi)^4} [I^\mu(k) + J_f^\mu(t,k)]^* \frac{\gamma_{\mu\nu}(k)}{k^2 - i\epsilon} [I^\nu(k) + J_f^\nu(t,k)] \qquad (V.5)$$

Now in the limit $t \to -\infty$, the exponential factors involving the time t ensure that in the cross terms between I and $J_f(t)$ the only momentum components which survive are those with $k^0 = \pm|\underset{\sim}{k}|$. The quadratic term involving J yields a term f^*f and also the imaginary part λ, so that we have

$$\lim_{t\to-\infty} X(t) = X + I^*f - \tfrac{1}{2}f^*f + i\lambda \ , \qquad (V.6)$$

where we have used again the formal notation for three-dimensional k integrals.

Thus our limiting formula (V.4) does indeed characterize a well-defined state $|f,\lambda;in\rangle$.

We may now repeat the argument to define asymptotic out states by limits as $t \to +\infty$. Hence we obtain an expression for the general matrix element

$$G_{f\lambda,g\mu}(\ldots|J) = \langle f,\lambda;out|\,T[\ldots]\,|g,\mu;in\rangle_J \ . \qquad (V.7)$$

This function has again the same structure as before but with an exponent $X_{f\lambda,g\mu}$ obtained by a double limiting procedure, which turns out to be

$$X_{f\lambda,g\mu} = X + (f^*g - \tfrac{1}{2}f^*f - \tfrac{1}{2}g^*g - i\lambda + i\mu) + i(f^*I + I^*g) \quad . \tag{V.8}$$

We note that formally

$$\exp X_{f\lambda,g\mu} = \langle f,\lambda \,|\, g,\mu\rangle \exp\left[X + i(f^*I + I^*g)\right] , \tag{V.9}$$

although as usual the factors here may separately diverge and must be combined before the k integration is performed.

It is easy to check that this structure is just what we would have obtained by making a formal expansion of the generalized coherent states in terms of n-photon states, and applying the usual rules of perturbation theory. The factor $\langle f,\lambda \,|\, g,\mu\rangle$ contains the straight-through soft-photon lines, while $\exp(if^*I)$ and $\exp(iI^*g)$ represent respectively the soft-photon emission and absorption vertices. However our formula has a significance even when the expansion in terms of n-photon states is invalid.

As a special case of our formula we have an expression for the transition amplitude between arbitrary generalized coherent states, namely

$$\langle f,\lambda;\mathrm{out} \,|\, g,\mu;\mathrm{in}\rangle_J = \langle f,\lambda \,|\, g,\mu\rangle \exp\left\{ i(f^*J + J^*g) + \tfrac{1}{2}i \int_{\Omega_S} \frac{d^4k}{(2\pi)^4} J^\mu(k)^* \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon} J^\nu(k) \right\} \tag{V.10}$$

This agrees precisely with the formula obtained earlier for a free field interacting only with an external source J, thus confirming that the interaction with the quantized electron field is ineffective in pure soft-photon states.

In particular, setting $J = 0$ we obtain

$$\langle f,\lambda;\mathrm{out} \,|\, g,\mu;\mathrm{in}\rangle_{J=0} = \langle f,\lambda \,|\, g,\mu\rangle \quad . \tag{V.11}$$

This shows that in the absence of an external current the soft-photon states are stable and do not scatter. It means that in the case $J = 0$ we need not distinguish the states $|f,\lambda;\mathrm{out}\rangle$ and $|f,\lambda;\mathrm{in}\rangle$. We may write simply

$$G_{f\lambda,g\mu}(\ldots) = \langle f,\lambda \,|\, T[\ldots] \,|\, g,\mu\rangle \quad .$$

The soft-photon external current J has now served its main purpose of allowing us to create soft-photon states, and for most purposes we may from now on set it equal to zero.

B. The Two-Point Function

We now turn to the crucial problem of defining states containing a single electron and, possibly, some soft photons.

To gain some insight into the structure to be expected, let us begin by examining the two-point function

$$G_{f\lambda,g\mu}(p,q) = \int d^4x d^4y\, e^{-ip\cdot x+iq\cdot y} \langle f,\lambda \,|\, T[\psi(x)\,\bar{\psi}(y)] \,|\, g,\mu\rangle \ .$$

In this case the only core diagrams which can contribute are those with a single straight-through line:



Hence our general formula reduces to

$$G_{f\lambda,g\mu}(p,q) = Z_2^h(p)\,\Lambda(p)\,\Delta^s_{f\lambda,g\mu}(p,q) \qquad (V.12)$$

where

$$\Delta^s_{f\lambda,g\mu}(p,q) = \int d^4y\, e^{-i(p-q)\cdot y} \int_0^\infty d\sigma\, e^{-i\sigma(m^2+p^2)} \exp X_{f\lambda,g\mu} \ .$$

In this case the exponent X contains only a single term X_{11}. As we saw in discussing the mass-shell singularity of the Green's function, the contribution to this term from the double pole at $p\cdot k = 0$ is negligible, and we may keep only the contribution from the pole at $k^\circ = |\underset{\sim}{k}|$, namely

$$X = e^2 \int \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|} \ \frac{p^\mu \gamma_{\mu\nu}(\underset{\sim}{k}) p^\nu}{(p\cdot k)^2} \left(e^{2i\sigma p\cdot k} - 1\right) \ .$$

It will be convenient to define the special photon wave functions associated with an electron of momentum p, namely

$$s_{p\mu}(\underset{\sim}{k}) = e\gamma_{\mu\nu}(\underset{\sim}{k}) \frac{p^\nu}{p\cdot k} \ . \qquad (V.13)$$

We recall also the notation for a translated wave function

$$(x)s_{p\mu}(\underset{\sim}{k}) = s_{p\mu}(\underset{\sim}{k})\, e^{-ik\cdot x} .$$

Then it is clear that

$$X = (2\sigma p)\, s_p^{*} s_p - s_p^{*} s_p .$$

There is also now only a single term in the current $I^{\mu}(k)$, which reduces for $k^2 = 0$ to

$$\gamma_{\mu\nu}(k) I^{\nu}(k) = i(\gamma + 2\sigma p) s_{p\mu} - i(\gamma) s_{p\mu} .$$

Thus the remaining terms in $X_{f\lambda, g\mu}$ may be written as

$$i(f^{*} I + I^{*} g) = - f^{*}(\gamma + 2\sigma p) s_p + f^{*}(\gamma) s_p + (\gamma + 2\sigma p) s_p^{*} g - (\gamma) s_p^{*} g .$$

We may expect to be able to define states $|\underset{\sim}{p};f,\lambda\rangle$ containing a single electron of momentum $\underset{\sim}{p}$ (and suppressed spin index) and soft photons in a generalized coherent state $|f,\lambda\rangle$. The effect on a soft-photon state of the part of $\psi(x)$ with momentum close to the mass shell must be expected to be to create a single electron but also to modify the soft-photon state in some way. This effect may be characterized by giving the matrix elements of $\psi(0)$ between all such states: $\langle\alpha|\psi(0)|\underset{\sim}{p};\beta\rangle$ where $|\alpha\rangle$ and $|\beta\rangle$ are arbitrary linear combinations of coherent states.

Before discussing the general case, let us examine the special case $\underset{\sim}{p} = 0$. In the radiation gauge we then have $s_p = 0$, so that the function Δ reduces to

$$\Delta^{s}_{f\lambda, g\mu}(p,q)\Big|_{\underset{\sim}{p}=\underset{\sim}{0}} = \langle f,\lambda | g,\mu\rangle (2\pi)^4 \delta_4(p-q) \frac{-i}{m^2 + p^2 - i\epsilon}\Bigg|_{\underset{\sim}{p}=\underset{\sim}{0}} \qquad (V.14)$$

In this case the Green's function does have a pole, and its discontinuity function is proportional to $\delta(m^2 + p^2)$. This must mean that the component of $\psi(x)$ with zero spatial momentum can only create a single electron without changing the soft-photon state. It is easy to verify that the discontinuity function of G is correctly reproduced if we assume that near the mass shell the only intermediate states which can contribute are of the form $|\underset{\sim}{p};\alpha\rangle$ and that

$$\langle \alpha | \psi(0) | \underset{\sim}{0};\beta \rangle = [Z_2^h(\underset{\sim}{0})]^{\frac{1}{2}} u(\underset{\sim}{0}) \langle \alpha | \beta \rangle \quad ,$$

$$\langle \underset{\sim}{0};\alpha | \bar{\psi}(0) | \beta \rangle = [Z_2^h(\underset{\sim}{0})]^{\frac{1}{2}} \bar{u}(\underset{\sim}{0}) \langle \alpha | \beta \rangle \quad . \tag{V.15}$$

These formulas effectively define the states $|\underset{\sim}{0};\alpha\rangle$. It should be noted that the definition involves an arbitrary phase convention. We could modify the states by adding any arbitrarily chosen generalized phase λ to the phase label of the coherent states without affecting anything. However this choice is clearly the natural one.

Having defined the states with zero three-momentum, we could in principle define all other states by applying Lorentz transformations. However this is hard to do in practice because of the complicated Lorentz-transformation properties of the fields in the radiation gauge. It is easier to proceed directly to obtain a corresponding formula for arbitrary momentum, and then verify its consistency with Lorentz invariance.

To obtain such a formula it is convenient to apply certain transformations to the function Δ^s, to obtain a form which is more symmetrical as between p and q. We first introduce a new variable of integration x together with a delta function $\delta_4(x-y-2\sigma p)$, which we shall represent by a Fourier integral:

$$\int d^4x \int \frac{d^4k}{(2\pi)^4} e^{-ik\cdot(x-y-2\sigma p)} \quad .$$

Then we may replace $y+2\sigma p$ by x and write

$$i(f^*I+I^*g) = -f^*(x)s_p + f^*(y)s_p + (x)s_p{}^*g - (y)s_p{}^*g \quad .$$

Moreover using the identity

$$(x)s_p{}^*(y)s_p = (x-y)s_p{}^*s_p \quad ,$$

we may also write

$$X = (x-y)\,s_p{}^*s_p - s_p{}^*s_p = (x)s_p{}^*(y)s_p - \tfrac{1}{2}(x)s_p{}^*(x)s_p - \tfrac{1}{2}(y)s_p{}^*(y)s_p \quad .$$

The variable σ now no longer appears in the exponent $X_{f\lambda,g\mu}$, so that we may perform the σ integration, obtaining

$$\int_0^\infty d\sigma\, e^{-i\sigma(m^2+p^2-2p\cdot k)} = \frac{-i}{m^2+p^2-2p\cdot k-i\varepsilon} .$$

We obtain

$$\Delta^s_{f\lambda,g\mu}(p,q) = -i\int d^4x d^4y \int \frac{d^4k}{(2\pi)^4} \frac{e^{-ik\cdot x - i(p-q-k)\cdot y}}{m^2+p^2-2p\cdot k-i\varepsilon}$$

$$\langle f,\lambda \,|\, g,\mu\rangle \exp\{ (x)s_p^*(y)s_p - \tfrac{1}{2}(x)s_p^*(x)s_p - \tfrac{1}{2}(y)s_p^*(y)s_p$$

$$- f^*(x)s_p + f^*(y)s_p + (x)s_p^* g - (y)s_p^* g \} . \qquad (V.16)$$

Now using the multiplication law of the unitary operators $U(f)$ repeatedly it is not hard to prove the identity

$$\langle f,\lambda \,|\, U(h_1)U(h_2) \,|\, g,\mu\rangle$$

$$= \langle f,\lambda \,|\, g,\mu\rangle \exp\{ -h_1^* h_2 - \tfrac{1}{2}h_1^* h_1 - \tfrac{1}{2}h_2^* h_2$$

$$+ f^* h_1 + f^* h_2 - h_1^* g - h_2^* g \} .$$

Hence we may apply this identity to write the exponential factor in the integrand as

$$\langle f,\lambda \,|\, U(-(x)s_p)\, U((y)s_p) \,|\, g,\mu\rangle .$$

Next, we observe that k may be treated as small. For in every term of the expansion of the exponential, the x integration would set k equal to a sum of soft-photon momenta. Hence it is consistent with our approximation scheme to add a term k^2 to the denominator, so that it becomes $m^2+(p-k)^2-i\varepsilon$. We then achieve a more symmetrical form by introducing a new variable $\ell = p - k$ in place of k. Finally we note that in the slowly-varying functions $Z_2^h(p)$, $\Lambda(p)$, and s_p it is permissible to replace p by ℓ. Thus we achieve the form

$$G_{f\lambda,g\mu}(p,q) = -i\int d^4x d^4y \int \frac{d^4\ell}{(2\pi)^4} \frac{e^{-i(p-\ell)\cdot x - i(\ell-q)\cdot y}}{m^2+\ell^2-i\varepsilon}$$

$$Z_2^h(\ell)\,\Lambda(\ell)\langle f,\lambda \,|\, U(-(x)s_\ell)U((y)s_\ell) \,|\, g,\mu\rangle . \qquad (V.17)$$

The corresponding discontuity function is obtained by replacing the pole by its delta-function part, and is

$$\text{disc } G_{f\lambda,g\mu}(p,q) = \int d^4x d^4y\, e^{-ip\cdot x+iq\cdot y} \int \frac{d^3\underset{\sim}{\ell}}{(2\pi)^3\, 2\omega_{\underset{\sim}{\ell}}} e^{i\underset{\sim}{\ell}\cdot(x-y)}$$

$$Z_2^h(\ell)\,\Lambda(\ell)\langle f,\lambda\,|\,U(-(x)s_{\underset{\sim}{\ell}})U((y)s_{\underset{\sim}{\ell}})\,|\,g,\mu\rangle \ .$$

On the other hand, directly from the definition of G we have

$$\text{disc } G_{f\lambda,g\mu}(p,q) = \int d^4x d^4y\, e^{-ip\cdot x+iq\cdot y}$$

$$\sum_n \langle f,\lambda\,|\,\psi(x)\,|\,n\rangle\langle n\,|\,\overline{\psi}(y)\,|\,g,\mu\rangle \ . \qquad \text{(V.18)}$$

Comparison of these two formulas suggests that the intermediate states which can contribute for momenta in the neighbourhood of the mass shell are states of the form $|\underset{\sim}{\ell},\alpha\rangle$ where $\underset{\sim}{\ell}$ is an electron momentum and α a soft-photon state. Choosing some complete set of soft-photon states $|\alpha\rangle$, we may write

$$\text{disc } G_{f\lambda,g\mu}(p,q) = \int d^4x d^4y\, e^{-ip\cdot x+iq\cdot y}$$

$$\int \frac{d^3\underset{\sim}{\ell}}{(2\pi)^3\, 2\omega_{\underset{\sim}{\ell}}} \sum_{\text{spins}} \sum_{\alpha} \langle f,\lambda\,|\,\psi(x)\,|\,\underset{\sim}{\ell};\alpha\rangle\langle\underset{\sim}{\ell};\alpha\,|\,\overline{\psi}(y)\,|\,g,\mu\rangle \ . \qquad \text{(V.19)}$$

It is clear that our two formulas for the discontinuity function agree if we set

$$\langle f,\lambda\,|\,\psi(0)\,|\,\underset{\sim}{\ell};\alpha\rangle = [Z_2^h(\ell)]^{\frac{1}{2}} u(\underset{\sim}{\ell})\,\langle f,\lambda\,|\,U(-s_{\underset{\sim}{\ell}})\,|\,\alpha\rangle \ . \qquad \text{(V.20)}$$

For then by translational invariance we have more generally

$$\langle f,\lambda\,|\,\psi(x)\,|\,\underset{\sim}{\ell};\alpha\rangle = [Z_2^h(\ell)]^{\frac{1}{2}}\, u(\underset{\sim}{\ell})e^{i\underset{\sim}{\ell}\cdot x}\langle f,\lambda\,|\,U(-(x)s_{\underset{\sim}{\ell}})\,|\,\alpha\rangle \ . \qquad \text{(V.21)}$$

and correspondingly

$$\langle\underset{\sim}{\ell};\alpha\,|\,\overline{\psi}(y)\,|\,g,\mu\rangle = [Z_2^h(\underset{\sim}{\ell})]^{\frac{1}{2}}\,\overline{u}(\ell)e^{-i\underset{\sim}{\ell}\cdot y}\langle\alpha\,|\,U((y)s_{\underset{\sim}{\ell}})\,|\,g,\mu\rangle \ . \qquad \text{(V.22)}$$

We see then that when the operator $\psi(0)$ creates a single electron it also modifies the soft photon state by adding in a soft-photon coherent state described by the wave function $-s_{\underset{\sim}{\ell}}$. It should be noted that this formula for the matrix element of ψ between states with no electrons and one electron reduces correctly to the formula we obtained earlier in the case $\underset{\sim}{\ell} = \underset{\sim}{0}$.

There is of course an undetermined phase in the state $|\underset{\sim}{\ell};\alpha\rangle$. It might happen that we should include not only an operator U but also an operator V which changes the phase label of the states. However, we are not at liberty to choose this phase arbitrarily, if we want our states to transform correctly under Lorentz transformations. In fact, it is possible to show that the choice we have made is the correct one by examining the behaviour under infinitesimal Lorentz transformations. However, the proof will be omitted here. It will be possible to check later that the scattering matrix elements we obtain are correctly Lorentz invariant.

Another point which should be noted is that we have not inserted any in or out labels on the one-electron states. We have implicitly assumed that the states containing one electron and soft photons are stable. This is physically plausible. For the soft photons can only transfer small amounts of energy and momentum to the electron which is therefore effectively unaccelerated and does not radiate. It may also be verified later when we evaluate scattering matrix elements between these states.

C. One-Particle Reduction Formulas

The discussion so far has been rather heuristic in nature, but we can now give a precise definition of the states containing one electron and soft photons, using a limiting formula of the kind used to define soft photon states.

The general state we seek to define is a linear combination of the states $\langle \underset{\sim}{p};\alpha;\text{out}|$, say

$$\langle \bar{\varphi};\text{out}| = \frac{1}{2m}\int \frac{d^3\underset{\sim}{p}}{(2\pi)^3\, 2\omega_{\underset{\sim}{p}}}\, \bar{\varphi}(\underset{\sim}{p}) \sum_{\text{spins}} u(\underset{\sim}{p})\, \langle \underset{\sim}{p};\alpha_{\underset{\sim}{p}};\text{out}| \quad .$$

Here $\bar{\varphi}$ is some spinor wave function, and for each $\underset{\sim}{p}$, $\alpha_{\underset{\sim}{p}}$ is some soft-photon state, in general a linear combination of generalized coherent states.

To define this state we have to give a prescription for calculating the matrix elements

$$\langle \bar{\varphi};\text{out}|\, T[\ldots]\, |g,\mu\rangle \quad .$$

We can in fact easily write down such a formula. Only a slight modification is needed of our earlier formula applicable to the case of theories without massless particles. We write

$$\langle\bar{\varphi};\text{out}|\,T[\ldots]\,|g,\mu\rangle$$

$$= \lim_{t\to\infty} \frac{1}{2m}\int \frac{d^4p}{(2\pi)^4}\,\bar{\varphi}(p)\,e^{-i(p^\circ-\omega_p)t}$$

$$[Z_2^h(p)]^{-\frac{1}{2}}\int d^4x_1 e^{-ip\cdot x_1}\,\langle\alpha_{p,x_1}|\,T[\psi(x_1)\ldots]\,|g,\mu\rangle\ . \tag{V.23}$$

There are only two differences between this formula and the earlier one, (II.18). Firstly, we have replaced the renormalization constant Z_2 by Z_2^h. Secondly, there is an additional dependence on x in the state $\langle\alpha_{p,x}|$ in the matrix element of the integrand. This state is defined by

$$\langle\alpha_{p,x_1}| = \langle\alpha_p|\,U((x)s_p) \qquad . \tag{V.24}$$

It is now straightforward to substitute our formula for the Green's formula into the right-hand side of this equation, and to show that the limit does indeed exist. We shall not go through the algebra in detail but merely indicate the important steps.

It is convenient to apply to the variables with label 1 the same transformations that we used in discussing the two-point function, replacing the integration over σ_1 by integrations over x_1 and ℓ_1. Taking the Fourier transform on p_1, we obtain an expression for

$$\langle f,\lambda|\,T[\psi(x_1)\ldots]\,|g,\mu\rangle\ .$$

From the structure of our limiting formula, it is clear that we are interested in the asymptotic behaviour of this function for large values of x_1°. In this limit we find that the dependence on the state $\langle f,\lambda|$ and the variable x_1 is in the combination

$$\langle f,\lambda|\,U(-(x_1)s_{\ell_1}) \qquad .$$

From the form of the x_1 integration it follows that $p-\ell_1$ may be treated as small. Thus in the slowly-varying function s_{ℓ_1} we may replace ℓ_1 by p. Then the factor $U(-(x_1)s_p)$ cancels the factor $U((x_1)s_p)$ in $\langle\alpha_{p,x_1}|$, and the x_1 integration has the effect of setting $\ell_1 = p$.

Thus although the matrix element between fixed soft-photon states does not have a pole, we find that the matrix element in which we modify the state in a specific way depending on the coordinate x_1 does have a pole. It follows that the limit does exist.

The limiting formula yields of course an expression for the matrix element

$$\langle \underline{p}_1;f,\lambda;\text{out}|\,T[\ldots]\,|g,\mu\rangle \quad . \tag{V.25}$$

It is not hard to guess how this function is related to the Green's function

$$G_{f\lambda\,,g\mu}(p_1\ldots) \quad . \tag{V.26}$$

In the familar case of a theory without massless particles, we have to remove the factor of $[Z_2(p_1)]^{\frac{1}{2}}$, replace $\Lambda(p_1)$ by $\bar{u}(p_1)$, delete the factor $(m^2 + p_1^2 - i\varepsilon)^{-1}$, and take the limit $p_1^2 \to -m^2$. In the present case, we have again to remove the factor $[Z_2^h(p_1)]^{\frac{1}{2}}$ and replace $\Lambda(p_1)$ by $\bar{u}(p_1)$. The elimination of the mass-shell singularity is achieved by replacing $\Delta^s(p_1\ldots)$ by a new function $\Delta^s(\hat{p}_1\ldots)$, where the hat denotes a variable on its mass shell. To get $\Delta^s(\hat{p}_1\ldots)$ from $\Delta^s(p_1\ldots)$ we have to delete the integration over σ_1 and the exponential factor $e^{i\sigma_1(m^2+p_1^2)}$ [or equivalently, to delete the integrations over x_1 and ℓ_1, and the denominator $-i(m^2+\ell_1^2 - i\varepsilon)^{-1}$, and set $\ell_1 = p_1$]. In the current $I^\mu(k)$ the term depending on p_1 must be replaced by its limiting form for $\sigma_1 \to \infty$ (or $x_1^\circ \to \infty$), namely

$$\frac{iep_1^{\mu}}{p_1\cdot k}\, e^{-ik\cdot y_1}\,(e^{-2i\sigma_1 p_1\cdot k} - 1) \Rightarrow -i\,\frac{ep_1^{\mu}}{p_1\cdot k - i\varepsilon}\, e^{-ik\cdot y_1} \quad . \tag{V.27}$$

There is one further modification which must be made, which concerns the term in X which is quadratic in this term in the current, namely X_{11}. If we were to substitute this limiting form directly into X_{11} we should obtain an undefined integral because of the factors $(p_1\cdot k - i\varepsilon)(p_1\cdot k + i\varepsilon)$ in the denominator. However as we saw earlier in discussing the mass-shell singularities of the Green's functions, the contribution of the double pole is in fact proportional to K and therefore negligible. To obtain a well-defined result, we must remove this term before going to the limit, and retain only the contribution from the pole at $k^2 = 0$, which may be written as

$$-\tfrac{1}{2}s_{p_1}^{*}\, s_{p_1} \quad .$$

D. Straight-Through Lines

We may now proceed in exactly the same way to define in states $|\underline{q}_1;g,\mu;\text{in}\rangle$, and thus obtain an expression for the matrix elements between such states,

$$\langle \underset{\sim}{p}_1;f,\lambda;\text{out}|\, T[\ldots]\,|\underset{\sim}{q}_1;g,\mu;\text{in}\rangle \quad . \tag{V.28}$$

There is one special case which must be considered separately, namely the contributions to such a matrix element which came from core diagrams with a straight-through line joining p_1 to q_1.

We may recall that in the familar case of a theory without soft photons, such a line in the Green's function involves a factor

$$\frac{-i}{m^2 + p^2 - i\varepsilon}\ (2\pi)^4\ \delta_4(p - q) \quad ,$$

so that when we have isolated the contribution of the pole in p, we have simply

$$(2\pi)^4\ \delta_4(p - q) \ .$$

Since $p^2 = -m^2$, we may write this delta function as the product of a mass-shell delta function for q and a three-dimensional delta function,

$$\theta(q^\circ)\ 2\pi\delta(q^2 + m^2)(2\pi)^3\, 2\omega_{\underset{\sim}{p}}\, \delta_3(\underset{\sim}{p} - \underset{\sim}{q}) \quad .$$

Then to go to the matrix element involving a physical particle of momentum $\underset{\sim}{q}$ in the initial state, we do not have to remove a factor $-i(q^2 + m^2)^{-1}$ but rather a delta function factor $\theta(q^\circ)2\pi\delta(q^2 + m^2)$, leaving the three-dimensional delta function.

In the present theory we must again expect to have to remove a mass-shell delta function, although the remaining matrix element will no longer involve a three-dimensional delta function, since some momentum may be transferred to other particles by soft-photon exchange.

To isolate the matrix element in question we again have to study the asymptotic behaviour of the integrand of the function

$$\Delta^s(\hat{p}_1\ldots, q_1\ldots) \ ,$$

this time as a function of y_1 as $y_1^\circ \to -\infty$. Now however the dependence on y_1 in the current $\hat{I}^\mu(k)$ is in the term

$$-i\,\frac{ep_1^{\ \mu}}{p_1\cdot k - i\varepsilon}\ e^{-ik\cdot y_1} \quad .$$

In a sense the imaginary part of the denominator here has the "wrong" sign. That is to say, if it were of the form $p_1\cdot k + i\varepsilon$, the pole would lie outside the k° contour and would not contribute. Thus we may replace this term by

$$ep_1^{\mu}\, 2\pi\, \delta(p_1\cdot k)\, e^{-ik\cdot y_1} \quad .$$

Because of this delta function, the integrand is now independent of the component of y_1 in the direction of p_1. Hence in the integral

$$\int d^4 y_1\, e^{-i(p_1-q_1)\cdot y_1} \ldots$$

the integration over this component of y_1 yields a delta function

$$\delta(p_1\cdot(p_1-q_1)) \quad .$$

Since $p_1 - q_1$ is small, this delta function is effectively equivalent to a mass-shell delta function for q_1,

$$\theta(q_1^{\circ})\, \delta(q_1^2 + m^2) \quad .$$

It may also be noted that since p_1 is a timelike vector the argument $p_1\cdot k$ of the delta function cannot vanish when $k^2 = 0$. Thus this term in $\Gamma^{\mu}(k)$ does not contribute to the mass-shell parts of the exponent X, and in particular to the photon emission and absorption parts. It can contribute only to cross terms with other electron lines. In particular, we find that when no other operators appear between the states, we have simply

$$\langle \underline{p}_1;f,\lambda;\text{out}|\underline{q}_1;g,\mu;\text{in}\rangle$$
$$= \langle \underline{p}_1|\underline{q}_1\rangle\langle f,\lambda|g,\mu\rangle \quad . \qquad \text{(V.28)}$$

This result shows that these states are actually stable, as we assumed earlier. For states involving only one electron and soft photons we need not distinguish between in and out states.

E. Multiparticle States

We can also of course go on to define states containing more than one electron, such as $\langle \underline{p}_1, \underline{p}_2;f,\lambda;\text{out}|$.

There is one new feature which appears here. If we start from the function

$$\Delta^s(\hat{p}_1\, p_2 \ldots)$$

and examine the asymptotic behaviour of the integrand as $x_2^{\circ} \to \infty$, we find in the exponent cross terms between the particles 1 and 2 of the form

$$-i\int_{\Omega_S}\frac{d^4k}{(2\pi)^4}\,\frac{ep_1^{\mu}}{p_1\cdot k+i\varepsilon}\,e^{ik\cdot y_1}\,\frac{\gamma_{\mu\nu}(k)}{k^2-i\varepsilon}\,\frac{ep_2^{\nu}}{p_2\cdot k-i\varepsilon}\,e^{-ik\cdot x_2}\quad .$$

Now, however, closing the contour in the lower half plane as usual we obtain a contribution from the pole at $p_1\cdot k = 0$.

When we go through the analysis we find that this term represents an effect on particle 1 of the current of particle 2, in much the same way that the unitary operator $U(-(x_2)s_{p_2})$ represents its effect on the soft-photon states. We find that the matrix element between one-electron and two-electron states

$$\langle \underline{p}_1;\alpha\,|\,\psi(x_2)\,|\,\underline{p}_1',\underline{p}_2;\beta;\text{out}\rangle$$

is not proportional to a delta function $\delta_3(\underline{p}_1-\underline{p}_1')$. Rather we have an extra factor which may be written

$$\Delta^S(\hat{\underline{p}}_1;\hat{p}_1'\,|\,I_{(2)})^*$$

where $I_{(2)}$ is the "current" of particle 2,

$$I_{(2)}{}^{\mu}(k) = -i\frac{ep_2^{\mu}}{p_2\cdot k-i\varepsilon}\,e^{-ik\cdot x_2}\quad .$$

The net effect, however, is to give essentially the same formula for the matrix elements between such states. Instead of going through this development in detail, we shall simply quote the resulting formula for the scattering matrix elements.

VI. Scattering Matrix Elements

The application of the reduction formulas we have developed allows us to write down an expression for an arbitrary scattering matrix element. To be specific let us consider the scattering matrix element between two states, each containing n electrons and some soft photons,

$$\langle \underline{p}_1\ldots\underline{p}_n;f,\lambda\,|\,S\,|\,\underline{q}_1\ldots\underline{q}_n;g,\mu\rangle$$
$$=\langle \underline{p}_1\ldots\underline{p}_n;f,\lambda;\text{out}\,|\,\underline{q}_1\ldots\underline{q}_n;g,\mu;\text{in}\rangle\quad . \qquad \text{(VI.1)}$$

We could of course also deal with matrix elements between states containing positrons or hard photons with obvious modifications.

We first write down the expression obtained by direct application of the reduction formulas. In order to compare this expression

with the corresponding one in a theory without massless particles, it will be useful to transform it to a more transparent form in which in particular its gauge invariance will be clear.

A. Rules for Obtaining Scattering Matrix Elements

To obtain an expression for the matrix element in question we have to start with the corresponding generalized Green's function

$$G_{f\lambda,g\mu}(p_1\cdots p_n,\ q_1\cdots q_n) \quad ,$$

and decompose it according to the connectivity structure of the core diagrams which contribute.

Let us consider the contribution of diagrams with one particular connectivity structure, in which there are r straight-through lines, joining $p_1\cdots p_r$ to $q_1\cdots q_r$. Then the soft-photon contribution to the Green's function is contained in the function

$$\Delta^s_{f\lambda,g\mu}(p_1\cdots p_r p_{r+1}\cdots p_n q'_{r+1}\cdots q'_n;$$
$$q_1\cdots q_r p'_{r+1}\cdots p'_n q_{r+1}\cdots q_n) \quad . \qquad \text{(VI.2)}$$

The corresponding contribution to the scattering matrix element is obtained by making the following changes:

(i) Delete the factor $[Z_2^h(p)]^{\frac{1}{2}}$ or $[Z_2^h(q)]^{\frac{1}{2}}$ for each external line.

(ii) Replace the factor $\Lambda(p_j)$ or $\Lambda(q_j)$ by $\bar{u}(p_j)$ or $u(q_j)$.

(iii) Replace the function Δ^s by a new function in which the first r lines have both ends on the mass shell and the remainder have one end on the mass shell, namely

$$\Delta^s_{f\lambda,g\mu}(\hat{p}_1\cdots\hat{p}_r\hat{p}_{r+1}\cdots\hat{p}_n q'_{r+1}\cdots q'_n;$$
$$\hat{q}_1\cdots\hat{q}_r p'_{r+1}\cdots p'_n\hat{q}_{r+1}\cdots q_n)\prod_{j=1}^{r} 2\pi\,\delta(q_j^2+m^2)$$
$$=\int\prod_{j=1}^{r} d^4y_j\, e^{-i(p_j-q_j)\cdot y_j}\prod_{j=r+1} d^4y_j d^4y'_j e^{-i(p_j-p'_j)\cdot y_j - i(q'_j-q_j)\cdot y'_j}$$
$$\langle f,\lambda\,|\,g,\mu\rangle\,\exp\{i(f^*I+I^*g)+\tfrac{1}{2}i\int_{\Omega^s}{}' \frac{d^4k}{(2\pi)^4} I^\mu(k)^*\frac{\gamma_{\mu\nu}(k)}{k^2-i\varepsilon}I^\nu(k)\} \quad , \qquad \text{(VI.3)}$$

where

$$I^{\mu}(k) = \sum_{j=1}^{r} ep_j^{\mu}\, 2\pi\delta(p_j\cdot k)\, e^{-ik\cdot y_j} - i\sum_{j=r+1}^{n} \frac{ep_j^{\mu}}{p_j\cdot k - i\varepsilon}\, e^{-ik\cdot y_j} + i\sum_{j=r+1}^{n} \frac{eq_j^{\mu}}{q_j\cdot k + i\varepsilon}\, e^{-ik\cdot y_j} \qquad \text{(VI.6)}$$

The prime on the integral signifies the special treatment which is required for the "diagonal" terms involving double poles.

If we label the connected pieces of the core diagrams by a label α, then we have for each a contribution of the form

$$(2\pi)^4\delta\Big(\sum_{(\alpha)} p' - \sum_{(\alpha)} q'\Big)\, M_{\alpha}^{h}(p'\ q')$$

where $\sum_{(\alpha)}$ denotes the sum of all momenta attached to this connected piece. Since the momenta $p_j - p_j'$ and $q_j' - q_j$ are small, we may make the replacement

$$M_{\alpha}^{h}(p'\ q') \Longrightarrow M_{\alpha}^{h}(p\ q)$$

so that the matrix element contains an overall factor

$$\prod_i \bar{u}(p_j) \prod_{\alpha} M_{\alpha}^{h}(p\ q) \prod_i u(q_j)$$

This may be recognized as being equal to the corresponding contribution to the scattering matrix element

$$\langle p_1 \cdots p_n | S^h | q_1 \cdots q_n \rangle$$

without soft-photon contributions, with the delta function factors associated with each connected piece removed. In fact, this non-soft-photon scattering matrix element has a contribution from diagrams of this connectivity structure of the form

$$\langle p_1 \cdots p_n | S^h | q_1 \cdots q_n \rangle = \prod_j \bar{u}(p_j) \prod_{\alpha} M_{\alpha}^{h}(p\ q)\, 2\pi\delta_4\Big(\sum_{(\alpha)} p - \sum_{(\alpha)} q\Big) \prod_j u(q_j). \qquad \text{(VI.5)}$$

Of course for a straight-through line, we should have simply

$$\bar{u}(\underline{p})\, u(\underline{q})\, (2\pi)^3\, 2\omega_{\underline{p}}\, \delta_3(\underline{p} - \underline{q}) \quad .$$

To accommodate straight-through lines within the same general formula we must make the formal identification

$$2m\, M_\alpha^h(p\ q)(2\pi)^4 \delta_4(p - q) = (2\pi)^3 2\omega_{\underline{p}}\, \delta_3(\underline{p} - \underline{q})$$

or

$$2m\, M_\alpha^h(p\ q)(2\pi)\, \delta(q^2 + m^2) = 1 \ . \tag{VI.6}$$

The factor of 2m arises from the fact that there is only one factor $\Lambda(p)$ for each straight-through line, which must be written in the form

$$\Lambda(p) = \Lambda(p)\, \frac{1}{2m}\, \Lambda(q) \quad ,$$

before applying our rules.

Now let us return to our formula including soft-photon contributions, and insert a Fourier integral representation for each delta function,

$$\int d^4x_\alpha \exp\Big[-i\Big(\sum_{(\alpha)} p' - \sum_{(\alpha)} q'\Big)\cdot x_\alpha\Big] \ .$$

Then we may perform the integrations over each p_j' and q_j'. The effect is simply to set each y_j or y_j' equal to the corresponding x_α. For straight-through lines we already have only a single integral over y_j to perform, and so writing x_α in place of this y_j we arrive at the formula

$$\langle \underline{p}_1 \cdots \underline{p}_n; f, \lambda; \text{out} | \underline{q}_1 \cdots \underline{q}_n; g, \mu; \text{in} \rangle$$

$$= \prod_j \bar{u}(\underline{p}_j) \prod_\alpha M_\alpha^h(p\ q) \prod_j u(\underline{q}_j) \int \prod_\alpha d^4x_\alpha \exp\Big[-i\Big(\sum_{(\alpha)} p - \sum_{(\alpha)} q\Big)\cdot x_\alpha\Big]$$

$$\langle f, \lambda | g, \mu \rangle \exp\{i(f^* I + I^* g) + \tfrac{1}{2} i \int_{\Omega^S}' \frac{d^4k}{(2\pi)^4} I^\mu(k)^* \frac{\gamma_{\mu\nu}(k)}{k^2 - i\varepsilon} I^\nu(k)\} \ , \tag{VI.7}$$

where now

$$I^\mu(k) = \sum_\alpha I_\alpha^\mu(k)\, e^{-ik\cdot x_\alpha} \tag{VI.8}$$

with

$$I_\alpha^\mu(k) = -i\sum_{(\alpha)} \frac{ep_j^\mu}{p_j\cdot k - i\varepsilon} + i\sum_{(\alpha)} \frac{eq_j^\mu}{q_j\cdot k + i\varepsilon} \quad . \tag{VI.9}$$

We note that this is correct even for a straight-through line. For in that case we have just one p_j and one q_j, with $q_j - p_j$ small. Hence replacing q_j by p_j we get

$$I_\alpha^\mu(k) = -i\frac{ep^\mu}{p\cdot k - i\varepsilon} + i\frac{ep^\mu}{p\cdot k + i\varepsilon} = ep^\mu\, 2\pi\,\delta(p\cdot k) \quad , \tag{VI.10}$$

which is correct.

We also note that each current I_α^μ is conserved,

$$k_\mu I_\alpha^\mu(k) = 0 \ .$$

Hence we may drop the factor $\gamma_{\mu\nu}(k)$ in the last term. Since $\gamma_{\mu\nu}(k)$ then no longer appears anywhere, our formula is manifestly gauge-invariant.

Except for the prime on the integral in (VI.7), the matrix element $\langle f,\lambda\,|\,g,\mu\rangle$ and the following exponential factor are just equal to the scattering matrix element in the presence of an external current I,

$$\langle f,\lambda;\mathrm{out}|g,\mu;\mathrm{in}\rangle_I = \langle f,\lambda\,|\,S(I)\,|\,g,\mu\rangle \ .$$

As we remarked earlier, the current I does not in fact belong to the class of currents studied earlier, and the prime on the integral is needed to give it a meaning. However we can define a new unitary operator

$$S'(I) = U(j)\,V(\sigma') \tag{VI.11}$$

where

$$j_\mu(\underset{\sim}{k}) = i\,\gamma_{\mu\nu}(\underset{\sim}{k})\,I^\nu(k)\Big|_{k^0=|\underset{\sim}{k}|} \tag{VI.12}$$

and

$$\sigma' = \tfrac{1}{2}\int_\Omega{}^{\!\prime}{}_S \frac{d^4k}{(2\pi)^4}\,\frac{I^\mu(k)^*\,I_\mu(k)}{k^2} \quad . \tag{VI.13}$$

Here the prime has the simple significance that we are required to omit the contributions of "diagonal" terms.

Then we obtain simply

$$\langle \underline{p}_1 \cdots \underline{p}_n; f,\lambda; \text{out} | \underline{q}_1 \cdots \underline{q}_n; g,\mu; \text{in} \rangle$$

$$= \sum \prod_j \bar{u}(\underline{p}_j) \prod_\alpha M_\alpha^h(p,q) \prod u(\underline{q}_j)$$

$$\int \prod_\alpha d^4x_\alpha \exp\left[-i\left(\sum_{(\alpha)} p - \sum_{(\alpha)} q\right)\cdot x_\alpha\right]$$

$$\langle f,\lambda | S'(I) | g,\mu \rangle \ , \qquad \text{(VI.14)}$$

where the first sum is over all connectivity structures.

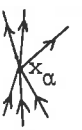
It should be noted that the rule for obtaining this matrix element from the corresponding one without soft photons is to replace the product of momentum-conserving delta functions by this soft-photon factor:

$$\prod_\alpha (2\pi)^4 \, \delta_4\left(\sum_{(\alpha)} p - \sum_{(\alpha)} q\right)$$

$$\Rightarrow \int \prod_\alpha d^4x_\alpha \exp\left[-i\left(\sum_{(\alpha)} p - \sum_{(\alpha)} q\right)\cdot x_\alpha\right] \langle f,\lambda | S'(I) | g,\mu \rangle \ . \qquad \text{(VI.15)}$$

This factor takes account of the momentum transferred to and from soft photons, and also the momentum transferred between electrons by soft-photon exchange.

Physically we may interpret I_α^μ as the classical current corresponding to the scattering processs represented by the αth connected piece of the diagram, occurring at the point x_α. We have to integrate over all possible positions x_α at which the scattering can occur.



B. Properties of the Scattering Operator

The formulas we have obtained define a scattering operator on the space spanned by the asymptotic states

$$|\underline{p}_1 \cdots \underline{p}_n; \underline{q}_1 \cdots \underline{q}_m; \underline{k}_1 \cdots \underline{k}_\ell; f, \lambda\rangle \quad .$$

This space is the direct product $\mathcal{H}_o^h \otimes \mathcal{H}_{IR}^s$ where $\mathcal{H}_o^h$ is the Fock space of all particles other than soft photons, and $\mathcal{H}_{IR}^s$ is the nonseparable Hilbert space spanned by the soft-photon coherent states.

Let us now list the properties we should expect this operator to possess, and discuss them in turn.

i) Gauge Invariance

We have already seen that this property follows from the fact that the current I is conserved.

ii) Crossing Symmetry

According to this principle the amplitude for a process involving emission of a positron of momentum q is obtained from the corresponding amplitude for absorption of an electron of momentum q by making the replacements $q \to -q$ and $u(\underline{q}) \to v(\underline{q})$. We have not explictly constructed the positron scattering amplitudes, but it is easy to do so by following the same line of argument, and to verify that this property is indeed satisfied. Indeed, the core diagrams clearly have the required property, and the only other place in which the momentum q appears is in the current I. For an incoming electron of momentum q the contribution is

$$i\frac{eq^\mu}{q\cdot k + i\epsilon} \quad .$$

Making the transformation $q \to -q$, we obtain

$$i\frac{eq^\mu}{q\cdot k - i\epsilon} \quad ,$$

which differs from the contribution of an outgoing electron only by an overall minus sign. This is correct because of the opposite sign of the charge for a positron.

iii) Translational Invariance

Our chosen states are not eigenstates of energy and momentum, and so we do not expect an overall energy-momentum conserving delta function. The expression of translational invariance is

$$\langle \underline{p}_1 \cdots \underline{p}_n; (x) f, \lambda; \text{out} | \underline{q}_1 \cdots \underline{q}_n; (x) g, \mu; \text{in}\rangle \, e^{i(\Sigma p - \Sigma q)\cdot x}$$
$$= \langle \underline{p}_1 \cdots \underline{p}_n; f, \lambda; \text{out} | \underline{q}_1 \cdots \underline{q}_n; g, \mu; \text{in}\rangle \quad . \qquad \text{(VI.16)}$$

It is easy to verify that this relation is satisfied, using the identity

$$\langle (x)f,\lambda | S'((x)I) | (x)g,\mu \rangle = \langle f,\lambda | S'(I) | g,\mu \rangle$$

which is readily established.

If we were to project out from the coherent states states of definite energy and momentum, we should of course obtain an overall energy-momentum-conserving delta function.

iv) Lorentz Invariance

The invariance of the S operator under infinitesimal Lorentz transformations may be established in a similar way, using the covariance of $S'(I)$ under such transformations. We shall not go through the algebra in detail. The invariance under finite Lorentz transformations is somewhat complicated by the fact that our separation between hard and soft photons is non-covariant. In particular this means that a soft-photon coherent state goes over under a finite Lorentz transformation into a state which contains some hard photons. What we have to prove in fact is that the expression we have obtained is unaffected by the choice of the soft-photon region Ω^S (so long as it is small, of course). This is related to the next property we require:

v) Independence of the Soft-Photon Cutoff K

Clearly since K is a purely conventional parameter, the operator S cannot depend on it. It is necessary for consistency to establish that this is indeed the case.

Before discussing this point in detail let us first complete our list of expected properties.

vi) Unitarity

Although we have defined a set of asymptotic states we have yet to show that they constitute a complete set. To do this we have to show that the operator S is unitary on the space $\mathcal{H}_o^n \otimes \mathcal{W}_{IR}^S$. We shall return to this problem later.

vii) Relationship to Physical Measurements

We have not yet shown how to compute the scattering probabilities of given physical processes. It is necessary for consistency that we should be able to reproduce the results which have been obtained by other methods and which are known to be in substantial agreement with experiment.

There is one other important property which is a feature of this method but not of others and which we therefore include in our list, namely

viii) Finiteness of Scattering Amplitudes

We developed earlier rules for interpreting such expressions as $\langle f,\lambda | S'(I) | g,\mu \rangle$. We may recall that if the exponent is represented by a divergent integral, then the matrix element is assigned the value zero. With this convention, all the scattering matrix elements we

have obtained are finite. Indeed for a given initial state, they are all zero except for final states in a very restricted class. Clearly we can define equivalence classes of the asymptotic states, such that all final states for which the matrix element is nonvanishing belong to one equivalence class. We shall return to this classification later.

C. Dependence on the Soft-Photon Cutoff

Now let us return to point number (v).

Let us consider a small change $\delta\Omega^S$ in the soft-photon region Ω^S. What we have to show is that the scattering operator S is the same whether we choose to regard as soft the photons in Ω^S or in $\Omega^{S\prime} = \Omega^S + \delta\Omega^S$.

The proof of this may be separated into two distinct parts, concerned with the external soft photons in the coherent states and with the internal soft-photon lines, respectively.

Let us examine first the external soft photons. To do this we first choose a state $|f,\lambda\rangle'$ which relative to $\Omega^{S\prime}$ is a pure soft-photon generalized coherent state. Without essential loss of generality we may suppose that the function $\lambda(\underset{\sim}{k})$ which defines the generalized phase of the state vanishes in $\delta\Omega^S$, since the part of $\lambda(\underset{\sim}{k})$ in this region contributes only a finite phase factor. Then relative to Ω^S, the state $|f,\lambda\rangle$ is not a pure soft-photon state but contains a small component containing one hard photon. There are of course also small components containing two or more hard photons, but these are of order $(\delta\Omega^S)^2$ or higher and therefore negligible. To first order in $\delta\Omega^S$, however, we also have to take account of the changed normalization of the state, and thus obtain

$$|f,\lambda\rangle' = |f,\lambda\rangle \left\{ 1 - \tfrac{1}{2} \int_{\delta\Omega} \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|}\, f_\mu(\underset{\sim}{k})^* f^\mu(\underset{\sim}{k}) \right\}$$

$$+ \int_{\delta\Omega} \frac{d^3\underset{\sim}{k}}{(2\pi)^3\, 2|\underset{\sim}{k}|}\, |\underset{\sim}{k}_\mu; f,\lambda\rangle\, f^\mu(\underset{\sim}{k}) \qquad \text{(VI.17)}$$

in which the subscript μ on $\underset{\sim}{k}$ indicates the polarization index of the single hard photon.

Now let us consider some scattering matrix element in which the asymptotic soft-photon states are

$$\text{`}\langle f,\lambda| \qquad \text{and} \qquad |g,\mu\rangle' \quad .$$

In terms of $\Omega^{S\prime}$, the contribution to the integrand of our expression for the scattering matrix element from the parts of f and g within the small region of $\delta\Omega$ is

$$\int_{\delta\Omega} \frac{d^3k}{(2\pi)^3 2|\underline{k}|} \{ f^* g - \tfrac{1}{2} f^* f - \tfrac{1}{2} g^* g + i f^* I + i I^* g \} . \qquad \text{(VI.18)}$$

In terms of Ω^S, the difference between this matrix element and the matrix element between

$$\langle f,\lambda | \quad \text{and} \quad | g,\mu \rangle$$

comes from the extra terms in (VI.17). It is easy to see that the changes in the normalization factors reproduce the terms $-\frac{1}{2} f^* f$ and $-\frac{1}{2} g^* g$ in (VI.18). Next we consider the contribution which comes from inserting the states containing a single hard photon on both sides. The only part of this contribution which is of order $\delta\Omega$ is that which arises from diagrams with a straight-through line joining the two hard photons. This correctly reproduces the term $f^* g$. Finally we consider the cross terms between the states containing no hard photons and one, representing single-photon emission or absorption. Even though we choose to regard the photon as hard, its momentum is still small compared to the electron mass, and we may therefore make the same approximations as before. We need only insert the hard-photon vertex in the external lines of the core diagrams, and it is easy to see that in this way we recover the contribution of the terms $i f^* I + i I^* g$.

A very similar argument may be used to deal with the internal soft-photon lines. In terms of $\Omega^{S'}$ the contribution of internal lines with momenta in the region $\delta\Omega$ is

$$\tfrac{1}{2} i \int'_{\delta\Omega} \frac{d^4k}{(2\pi)^4} \frac{I^\mu(k) I_\mu(k)}{k^2 - i\varepsilon} . \qquad \text{(VI.19)}$$

In terms of Ω^S, on the other hand, these photons are hard, and this contribution must be counted as a part of the core diagrams. To find the contribution from hard-photon internal lines with momentum in the region $\delta\Omega$, we would have to insert such internal lines in all possible ways into the core diagrams. We may again restrict attention to insertions with both ends on external lines of the core diagrams. In this way it is not hard to show that we recover this same expression. The only point of difficulty is associated with the prime on the integral which in terms of Ω^S comes from the fact that for an insertion with both ends in the same external line we are required to subtract out the contribution to the mass renormalization constant. The contribution to the

wave-function renormalization constant $[Z_2^h(p)]^{\frac{1}{2}}$ is removed automatically by the factor of $\frac{1}{2}$.

D. Equivalence Classes

Next let us examine the conditions under which a scattering matrix element can be nonzero.

Since any soft-photon integral involving a factor of the form $(e^{-ik\cdot x}-1)$ which vanishes at $k = 0$ is automatically convergent, it follows that to find the divergent part of the integral we may always replace a factor $e^{-ik\cdot x}$ by 1. Hence the matrix element

$$\langle \underline{p}_1 \ldots \underline{p}_n; f, \lambda; \text{out} | \underline{q}_1 \ldots \underline{q}_n; g, \mu; \text{in} \rangle \tag{VI.20}$$

can be nonzero if and only if the same is true of

$$\langle f, \lambda | S'(I) | g, \mu \rangle$$

with I evaluated with all x_α set equal to zero. We may write

$$I = I_{\text{out}} + I_{\text{in}}$$

where

$$I_{\text{out}\,\mu}(k) = -ie\gamma_{\mu\nu}(k) \sum \frac{p_j^{\mu}}{p_j \cdot k - i\epsilon}$$

$$I_{\text{in}\,\mu}(k) = ie\gamma_{\mu\nu}(k) \sum \frac{q_j^{\mu}}{q_j \cdot k + i\epsilon} \quad .$$

We note that although I is conserved, I_{out} and I_{in} are not, so that the separation between the two is not gauge-invariant.

Introducing as before the notation

$$s_{p\mu}(k) = \gamma_{\mu\nu}(k) \frac{ep^{\nu}}{p \cdot k}$$

we may write

$$S'(I) = U\Big(\sum_j s_{p_j} - \sum_j s_{q_j}\Big) V(\sigma') \quad . \tag{VI.21}$$

The contribution to σ' from a cross term between p_i and q_j is

$$\sigma_{p_i q_j} = -e^2 \,\mathrm{Re} \int_{\Omega_s} \frac{d^4 k}{(2\pi)^4} \frac{p_i^{\mu} \gamma_{\mu\nu}(k) q_j^{\nu}}{(p_i \cdot k + i\varepsilon)(k^2 - i\varepsilon)(q_j \cdot k + i\varepsilon)}$$

But because both denominators $(p_i \cdot k + i\varepsilon)$ and $(q_j \cdot k + i\varepsilon)$ have the same sign of the imaginary part, we may complete the k^0 contour in the upper half plane, and exclude both poles. Thus we get only a contribution from the pole at $k^0 = -|\underset{\sim}{k}|$, which is pure imaginary, so that this contribution vanishes.

On the other hand in the term involving p_i and p_j we may make the replacement

$$\frac{1}{p_j \cdot k - i\varepsilon} \to \frac{1}{p_j \cdot k - i\varepsilon} - \frac{1}{p_j \cdot k + i\varepsilon} = 2\pi i\, \delta(p_j \cdot k) \quad .$$

Then since k^2 cannot vanish when $p_j \cdot k = 0$, we may discard the $i\varepsilon$ in that denominator. Therefore the real part of the integral comes only from the delta-function part of the denominator $(p_i \cdot k + i\varepsilon)$, so that we obtain

$$\sigma_{p_i p_j} = \tfrac{1}{2} e^2 p_i \cdot p_j \int_{\Omega_s} \frac{d^4 k}{(2\pi)^4}\, 2\pi\delta(p_i \cdot k) \frac{1}{k^2}\, 2\pi\delta(p_j \cdot k) \quad .$$

This term represents the effect of interchange of soft photons between the electrons i and j, with momenta orthogonal to both p_i and p_j, which thus leave both particles on their mass shells. It corresponds to the usual Coulomb interaction, which is well known to give rise to a formally divergent phase factor.

Because all the wave functions s_{p_i} are real, we may write

$$U\Big(\sum_j s_{p_j} - \sum_j s_{q_j}\Big) = U\Big(\sum s_{p_j}\Big)\, U\Big(-\sum s_{q_j}\Big) \quad .$$

Hence we obtain

$$S'(I) = S'(I_{out})\, S'(I_{in}) \tag{VI.22}$$

with

$$S'(I_{out}) = U\Big(\sum s_{p_j}\Big)\, V\Big(\sum_{i<j} \sigma_{p_i p_j}\Big)$$

$$S'(I_{in}) = U\Big(-\sum s_{q_j}\Big)\, V\Big(\sum_{i<j} \sigma_{q_i q_j}\Big)$$

It follows that the matrix element

$$\langle \underset{\sim}{p}_1 \cdots \underset{\sim}{p}_n; f,\lambda; \text{out} | \underset{\sim}{q}_1 \cdots \underset{\sim}{q}_n; g,\mu; \text{in} \rangle$$

can be nonzero if and only if the same is true of the matrix element

$$\langle f,\lambda | S'(I_{\text{out}})\, S'(I_{\text{in}}) | g,\mu \rangle$$

or, in terms of the definition of equivalence introduced earlier for soft-photon states if and only if

$$S'(I_{\text{out}})^{-1} |f,\lambda\rangle \sim S'(I_{\text{in}}) |g,\mu\rangle$$

or, more explicitly, if

$$U\Big(-\sum s_{p_j}\Big)\, V\Big(-\sum_{i<j} \sigma_{p_i p_j}\Big) |f,\lambda\rangle$$

$$\sim U\Big(-\sum s_{q_j}\Big)\, V\Big(\sum_{i<j} \sigma_{q_i q_j}\Big) |g,\mu\rangle \ . \qquad \text{(VI.23)}$$

Hence we may define the equivalence classes of the states containing n electrons to be those of the corresponding soft-photon states, defined by the correspondences

$$|\underset{\sim}{p}_1 \cdots \underset{\sim}{p}_n; f,\lambda; \text{out}\rangle \Rightarrow U\Big(-\sum s_{q_j}\Big)\, V\Big(-\sum_{i<j} \sigma_{p_i p_j}\Big) |f,\lambda\rangle$$

$$|\underset{\sim}{q}_1 \cdots \underset{\sim}{q}_n; g,\mu; \text{in}\rangle \Rightarrow U\Big(-\sum s_{q_j}\Big)\, V\Big(+\sum_{i<j} \sigma_{q_i q_j}\Big) |g,\mu\rangle \ . \qquad \text{(VI.24)}$$

If we choose an initial state in one particular equivalence class of in states, then the only nonvanishing scattering matrix elements are those to final states in the corresponding equivalence class of out states. It should be noted, however, that in contrast to the case of scattering by an external field discussed earlier, the decomposition of $\mathcal{H}_0^h \otimes \mathcal{H}_{IR}^s$ into equivalence classes is different for the in and out states, because the phase factors appear with opposite signs. If two in states are equivalent it does not follow that the out states with the same labels are equivalent in general.

Although as we have seen the separation of $S'(I)$ into the product $S'(I_{\text{out}})S'(I_{\text{in}})$ is not gauge-invariant, it is nevertheless true that

the definition of equivalence for states with the same total charge is gauge-invariant. These are of course the only states between which the operator S can have nonvanishing matrix elements. However, we can also define equivalence for states with different total charges by the correspondences we have introduced. For such states the concept of equivalence is gauge-dependent. It is not hard to give a physical interpretation to this extended concept of equivalence. It is easy to check by going back to our earlier formulas that a matrix element of the form

$$\langle \underset{\sim}{p}_1 \ldots \underset{\sim}{p}_m ; f, \lambda ; \text{out} | \, T[\ldots] \, | \underset{\sim}{q}_1 \ldots \underset{\sim}{q}_n ; g, \mu ; \text{in} \rangle$$

where the dots denote any finite product of field operators can be nonzero only if the states are equivalent. Since the field operators are gauge-dependent, it is natural that this definition of equivalence should also be gauge-dependent.

E. Relationship to Physical Measurements

We now examine point number (vii).

For simplicity, let us suppose that we are interested in a region of momenta in which only completely connected core diagrams can contribute.

Let us introduce the total momenta

$$p = \sum_j p_j \qquad\qquad q = \sum_j q_j$$

and write the integration over the final-state phase space of the electrons in the form

$$\prod_j \frac{d^3 \underset{\sim}{p}_j}{(2\pi)^3 \, 2\omega_{\underset{\sim}{p}_j}} = \frac{d^4 p}{(2\pi)^4} \, d\Omega'$$

where $d\Omega'$ represents the integration over the remaining $3n - 4$ variables. Since the total momentum transfer to soft photons $q - p$ is necessarily small, we may write the function M^h as a function of q, Ω' and Ω, independent of p.

If we had no soft-photon contributions we should obtain a scattering matrix element of the form

$$\langle \underset{\sim}{p}_1 \ldots \underset{\sim}{p}_n ; \text{out} | \underset{\sim}{q}_1 \ldots \underset{\sim}{q}_n ; \text{in} \rangle^h = (2\pi)^4 \delta_4(p - q) \langle \Omega' | S^h(q) | \Omega \rangle \quad .$$

The transition probability per unit space-time volume into a set of final states defined by a region Ω' would then be

$$w^h(\Omega') = \int_{\Omega'} d\Omega' \, |\langle \Omega' | S^h(q) | \Omega \rangle|^2 \qquad \text{(VI.25)}$$

What we want to do is to compare the true transition probability with w^h.

In the true theory the scattering matrix element may be written in the form

$$\langle \underline{p}_1 \ldots \underline{p}_n; f, \lambda; \text{out} | \underline{q}_1 \ldots \underline{q}_n; g, \mu; \text{in} \rangle$$
$$= \langle \Omega' | S^h(q) | \Omega \rangle \int d^4x e^{-i(p-q)\cdot x} \langle f, \lambda | S'((x)I) | g, \mu \rangle \quad .$$

Now let us suppose that the range of final state is defined by specifying some range Δ of the total momentum transfer $\hbar = q - p$ to soft photons, and some range Ω' of the other final state variables. We are assuming here that no direct measurement is made of the final soft-photon state. To compute this transition probability we have to take the squared modulus of the matrix element and sum over a complete set of soft-photon states $|\alpha\rangle$. This yields

$$\int_{\Omega'} d\Omega' \, |\langle \Omega' | S^h(q) | \Omega \rangle|^2$$
$$\int_{\Delta} \frac{d^4\hbar}{(2\pi)^4} \sum_{\alpha} \left| \int d^4x \, e^{i\hbar\cdot x} \langle \alpha | S'((x)I) | g, \mu \rangle \right|^2 \quad .$$

Now the sum over α may be written

$$\sum_{\alpha} \left| \int d^4x e^{i\hbar\cdot x} \langle \alpha | S'((x)I) | g, \mu \rangle \right|^2$$
$$= \int d^4x d^4x' e^{i\hbar\cdot(x-x')} \langle g, \mu | S'^{-1}((x')I) \, S'((x)I) | g, \mu \rangle$$

First let us suppose that no measurement of the total momentum transfer to soft photons is made, so that the integral over Δ is over all $\hbar$ space. Then the $\hbar$ integration has the effect of setting $x = x'$. But then the two factors S'^{-1} and S' cancel, and we are left with $\langle g, \mu | g, \mu \rangle = 1$. In this case we have simply

$$w(\Omega') = w^h(\Omega') \qquad \text{(VI.26)}$$

This is a very interesting result. The probability of any scattering process in which we make no observation of the soft photons, or of the total momentum transfer to soft photons, is precisely equal to the corresponding probability calculated by discarding all soft-photon contributions.

Next, let us take another special case, and suppose that the initial soft-photon state is the vacuum, so that $g = 0$. Then from translational invariance it follows that

$$\langle 0 | S'^{-1}((x')I)\, S'((x)I) | 0 \rangle$$

$$= \langle 0 | S'^{-1}((x'-x)I) S'(I) | 0 \rangle$$

We again obtain a result proportional to the total space-time volume, which may be written

$$w(\Omega', \Delta) = w^h(\Omega') \int_\Delta \frac{d^4k}{(2\pi)^4}\, \rho(k) \qquad \text{(VI.27)}$$

where

$$\rho(k) = \int d^4x e^{-ik\cdot x} \langle 0 | S'^{-1}((x)I)\, S'(I) | 0 \rangle \quad . \qquad \text{(VI.28)}$$

This function $\rho(k)$ may be recognized as being the momentum distribution function associated with the coherent state $S'(I)|0\rangle$, which is characterized by the photon wave function

$$\sum s_{p_j} - \sum s_{q_j} \quad .$$

In particular, if Δ is defined by a limitation on the total energy transfer to soft photons, $k^\circ < E$, we find

$$w(\Omega', \Delta) = w^h(\Omega) \int \frac{d^4k}{(2\pi)^4}\, \rho(k)\, \theta(E - k^\circ)$$

$$= w^h(\Omega) \frac{e^{\gamma\xi}}{(-\xi)!} \left(\frac{E}{K}\right)^{-\xi} \qquad \text{(VI.29)}$$

where ξ is the parameter introduced earlier. This form is in agreement with the one which has been obtained by other methods.

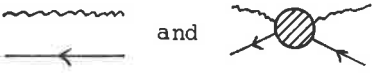
In the general case, in which the initial state is <u>not</u> the vacuum, the transition probability is not simply proportional to the space-time volume. This is as one should expect since clearly soft-photon absorption can only occur if the interaction occurs in a volume where

the soft-photon field is appreciable. The general formula can be used to find the effect of an initial soft-photon state different from the vacuum, even one in which the expectation value of the total photon number is infinite, whereas conventional methods cannot easily handle such problems.

F. Unitarity

We have still to prove that the scattering operator we have defined is unitary. We shall not give a general proof, but rather examine a special case which will be sufficient to illustrate the general method.

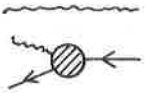
Let us consider an initial state containing one electron, one hard photon and some soft photons. If the final state is of the same type, then we have two connectivity classes to consider:



For other final states, there is in general only one class, for example



(although if one of the outgoing photons is nearly soft we may have a small contribution from diagrams with the structure



which we shall ignore).

For final states containing, say one electron and two hard photons, we have to integrate over the nine-dimensional phase space of these particles, and sum over all final soft-photon states. It is possible to choose as before four of the integration variables to be the components of the total energy-momentum vector p. Thus we have to consider the integral

$$\int \frac{d^4p}{(2\pi)^4} \int d\Omega' \; |\bar{u}\, M^h\, u|^2$$

$$\sum_\alpha \left| \int d^4x e^{-i(p-q)\cdot x} \langle \alpha | S'((x)I) | f,\lambda \rangle \right|^2 ,$$

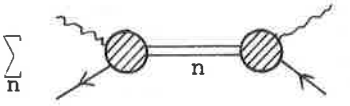
where we have chosen to indicate the dependence of I on the variable x explicitly.

Now since the total momentum transfer to soft photons is necessarily small, it is possible to write the non-soft-photon part of the matrix element, $\bar{u} M^h u$ in a form which is independent of p, but depends only on the initial momenta and the other five final-state variables, by replacing p wherever it appears by q. The same may be done for the dependence of I on p. Then p appears only in the exponential factor $e^{-ip\cdot x}$. Hence the effect of the p integration is to set the two variables x, which appear in the matrix element and its complex conjugate, equal. Then we can perform the sum over α to obtain

$$\langle f,\lambda | S'((x)I)^{-1} S'((x)I) | f,\lambda \rangle = 1 .$$

Thus there is no remaining dependence on the initial soft-photon state. What is left is simply a contribution to the unitarity integral for a theory without soft photons, that is to say a contribution to a matrix element of $S^{h*}S^{h*}$.

We now sum over all possible final states, a process we may represent diagrammatically by



omitting only the contribution to the two-particle term arising from the case where both the contributions to S and S^* correspond to disconnected diagrams:



When we sum over the soft-photon part of the final state, we obtain in every case the corresponding contribution to the matrix element of $S^{h*}S^h$. The omission of the totally disconnected diagrams means that we obtain the matrix element of

$$S^{h*}S^h - 1 ,$$

which of course vanishes by virtue of the fact that the scattering operator S^h for a theory without soft photons is unitary.

It remains only to show that the contribution of the disconnected diagrams reproduces correctly the matrix element of the right-hand side of the unitarity equation

$$S^*S = 1 .$$

This is especially easy to prove directly for the case we have considered, because of the fact that there is only one charged-particle line. The integration over the momentum of the final electron may be written

$$\int \frac{d^4p}{(2\pi)^4}\,\theta(p^0)\,\delta(p^2 + m^2) \quad .$$

If we did not have the mass-shell delta-function the integration over p would again have the effect of setting the values of x equal in the two factors. However this delta function is equivalent to $(1/2m)\delta(p\cdot q)$. The remaining three-dimensional integration has the effect of setting equal the components of x orthogonal to q. But as we have seen, for a straight-through line, the matrix element is in fact independent of the component of x in the direction of q, so that this is actually sufficient to ensure the cancellation of the soft-photon factors.

For more complicated processes, for which various connectivity structures can arise, the argument is somewhat more complicated, but it can be carried through by similar methods. The result is to prove the unitarity of the scattering operator S.

G. Conclusions

We have thus shown that quantum electrodynamics may be treated according to the same general principles that are applied to other field theories. If we avoid making the a priori assumption that the asymptotic states must belong to the Fock space but instead allow the theory itself to determine the appropriate space of asymptotic states, then we can obtain scattering matrix elements from the Green's functions of the theory. The scattering operator so defined has all the properties to be expected of it. Its matrix elements yield expressions for physical scattering probabilities which are in agreement with those obtained by other methods, but which are in fact more general inasmuch as we can investigate such questions as the dependence of scattering probabilities on the initial soft-photon state, even when this state contains infinitely many photons.

References

Quantum Electrodynamics and Field Theory

1. J. M. Jauch and F. Rohrlich, The Theory of Photons and Electrons (Addison-Wesley Pub. Co., Reading, Mass., 1955).
2. K. Johnson, Brandeis Summer Institute in Theoretical Physics (1964) Vol. 2: Lectures on Particles and Field Theory (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1965).
3. H. Lehmann, K. Symanzik and W. Zimmermann, Nuovo Cimento 1, 205 (1955).

Infrared Divergences

1. F. Bloch and A. Nordsieck, Phys. Rev. 52, 54 (1937).
2. D. Yennie, S. Frautschi and H. Suura, Ann. Phys. (N.Y.) 13, 379 (1961).
3. K. E. Eriksson, Nuovo Cimento 19, 1010 (1961).
4. J. M. Jauch and F. Rohrlich, Helv. Phys. Acta 27, 613 (1954).
5. E. L. Lomon, Nuc. Phys. 1, 101 (1956); Phys. Rev. 113, 726 (1959).
6. E.L. Lomon and R. Perrin, Ann. Phys. (N.Y.) 33, 328 (1965).
7. K. T. Mahanthappa, Phys. Rev. 126, 329 (1961).
8. S. Weinberg, Phys. Rev. 140, B516 (1965).
9. J. Tarski, J. Math. Phys. 7, 560 (1966).
10. R. J. Glauber, Phys. Rev. 84, 395 (1951).

Coherent States

1. R. J. Glauber, Phys. Rev. 130, 2529 (1963); 131, 2766B (1963).
2. S. S. Schweber, J. Math. Phys. 3, 831 (1962).
3. V. Bargmann, Proc. Natl. Acad. Sci.U.S. 48, 199 (1962).
4. L. I. Schiff, Quantum Mechanics , 2nd ed. (McGraw-Hill, New York, 1955) p. 67-69.
5. T. W. B. Kibble, Cargèse Lectures in Physics, Vol. 2 (Gordon and Breach, New York, 1968).

Infinite Tensor Products

1. J. von Neumann, Composito Math. 6, 1 (1938).
2. J. R. Klauder and J. McKenna, J. Math. Phys. 6, 68 (1965).
3. J. R. Klauder, J. McKenna, and E. J. Woods, J. Math. Phys. 7, 822 (1966).

Coherent States and Infrared Divergences

1. K. O. Friedrichs, Mathematical Aspects of the Quantum Theory of Fields (Interscience Publishers, New York, 1953).
2. V. Chung, Phys. Rev. 140, B1110 (1965).
3. T. W. B. Kibble, J. Math. Phys. 9, 315 (1968); Phys. Rev. 173, 1527; 174, 1882; and 175, 1624 (1968).
4. J. K. Storrow, Nuovo Cimento 54A, 15 (1968).

SOLUTIONS OF BETHE-SALPETER EQUATIONS†

Antonio Pagnamenta
Department of Physics
Rutgers University
New Brunswick, New Jersey

Contents

I. Introduction
II. General Discussion of the BS Equation
III. Relationship to the Schrödinger Equation
IV. The Structure of the BS Kernel
V. Two and Three Particle Unitarity
VI. The Alternative
VII. An Analytic Solution
VIII. An Expansion Around the Analytic Solution
IX. The Bound State Region
X. Regge and Toller Trajectories
XI. The Elastic Region
XII. Solutions in X-Space
XIII. Various Approximations
XIV. Relativistic Three Particle Equations
XV. The Relativistic Quark Model
XVI. Discussion

I. Introduction

Some people despise the Bethe-Salpeter (BS) equation. They say it cannot be derived in a rigorous way from field theory or from a Hamiltonian without making generous approximations. This can be done for the Schrödinger equation, for instance. Others realize that the Hamiltonian from which one obtains a non-relativistic equation is an even less satisfying approximation to begin with. Over the last few years the second group has gotten a big boost from various researchers who have found in the BS equation an interesting relativistic and covariant model which, in many ways, is quite realistic. It has been used to study new ideas and has brought many surprises.

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

The BS equation provides a relativistic description of the dynamics of a democratic two-body system. Early interest in this equation came from the necessity to study the positronium[1),2),3)] and the deuteron.[4)] The early period was also very exciting because Wick[5)] and Cutkosky[6)] presented analytic solutions for special cases. In both systems the two particles are of comparable mass. But in the positronium the interaction is electromagnetic, that is so weak that one or two terms are sufficient. The binding in the deuteron is so weak that somehow the Schrödinger equation seems to give satisfactory results in spite of great efforts.[4)] Today we are more interested in tightly bound systems and are wondering what the BS equation might tell us about the quark model.[7)] A more abstract line of investigation has been started by Gell-Mann and Low[8)] and Mandelstam.[9)] This is continued in recent studies of the behaviour and presence of Regge trajectories and daughter trajectories in the BS equation. The entire problem would lose a lot of its interest if we had not available numerical methods and were restricted to the few analytic solutions available. In saying this I am well aware of the preferability of an analytic solution and sometimes I will prefer a quickly available approximate but analytic solution to a difficult to obtain accurate one. But restricting oneself to the analytic solutions alone and dropping the computer work could be compared to restricting algebra to the ring of integers where one could make use of the entire field of reals. Clearly solving the equation can only be part of the work. It is just as necessary to study the origin and the consequences of the equation as well as uniqueness and analytic properties of the solutions.

These lectures will be concerned mainly with numerical solutions but I will not bother you with dozens of tables. I will, however, pay some attention to the special problems which you encounter when you attempt a numerical solution as distinguished from an analytical one. I will therefore deal mostly with that part of the calculation which has been termed trivial by some authors. It is interesting to remark that most of these authors have never successfully obtained a numerical result. The difficulties to overcome in putting an equation onto a computer are of entirely a different nature than the ones you face in an analytic discussion. A good example is given by the numerical integration of a logarithmic infinity. Such a singularity is square integrable and therefore an integral equation with a kernel containing such an infinity can well be Fredholm. You would be disappointed, however, about the result if you were to put this equation straight onto the computer. The results would be oscillate in a wide range, its value would depend crucially on how close to the singularity you happen to place a meshpoint. Take for example the integral

$$\int_0^1 dx \log x = -1 \tag{I.1}$$

and integrate it by the 100 point trapezoidal rule. The error in the result can easily be as much as 5%. Make the simple substitution

$$x = y^2 , \tag{I.2}$$

and you find

$$4 \int_0^1 dy\, y \log y \tag{I.3}$$

an integral which you can do by 3 point (!) Gaussian integration. You obtain accurately 4 significant figures. We will see that it is even more important but also more difficult to remove such singularities from the kernels of integral equations.

II. General Discussion of the BS Equation

In this section I will try to give some justification for studying the BS equation. I will put it into the context of more familar and easier soluble equations of motion. By giving several, all of them heuristic, derivations I will show as accurately as possible what is being kept and what has been neglected in writing down the BS equation.

In non-relativistic quantum mechanics we write down the Schrödinger equation

$$\left[\frac{p^2}{2m} + V(\underset{\sim}{x})\right] \psi_o(\underset{\sim}{x}) = E\, \psi_o(\underset{\sim}{x}) \tag{II.1}$$

for the stationary wave function of a particle with mass m. This equation remains valid for a non-relativistic two particle system if we interpret m as the reduced mass m_r

$$m_r = \frac{m_1\, m_2}{m_1 + m_2} \quad . \tag{II.2}$$

We can take the Fourier transform of Eq. (II.1) and write it in momentum space. For a bound system of two particles with masses m_1 and m_2 and with binding energy

$$B = 2m - E = m_1 + m_2 - E \tag{II.3}$$

it becomes

$$\left[B + \frac{p^2}{2m_r}\right]\tilde{\psi}_o(\underset{\sim}{p}) = \int d^3k\, V(\underset{\sim}{p}-\underset{\sim}{k})\,\tilde{\psi}_o(\underset{\sim}{k}) \qquad (II.4)$$

$\tilde{\psi}(\underset{\sim}{p})$ is the Fourier transform but I will drop the ~ from now on.

These equations are still non-relativistic. We can help here. For a spin $\frac{1}{2}$ particle we are used to seeing the Dirac equation

$$[\underset{\sim}{\alpha}\cdot\underset{\sim}{p} + \beta m]\psi_o(x) = V(x)\,\psi_o(x) \qquad (II.5)$$

For a spin 0 particle we have the Klein-Gordon equation

$$[\Box_x + m^2]\psi_o(x) = V(x)\psi_o(x) \qquad (II.6)$$

In (II.5) $\underset{\sim}{\alpha}, \beta$ are the four-dimensional Pauli matrices. x denotes the four vector $(x_o, \underset{\sim}{x})$. We use the time favoring metric (+---) and let

$$\Box_x = \partial_\mu \partial^\mu = \frac{\partial^2}{\partial x_o^2} - \underset{\sim}{\nabla}^2 \quad . \qquad (II.7)$$

We can take the four-dimensional Fourier transform of (II.6) to get

$$[p^2 - m^2]\psi_o(p) = \int d^4k\, V(q-k)\psi_o(k) \qquad (II.8)$$

where $p = (p_o, \underset{\sim}{p})$ and $p^2 = p_o{}^2 - \underset{\sim}{p}^2$. But now we realize that we do not know what is V(q), the interaction, as a function of the four vector q. In addition all the above equations are one-particle equations only. This is fine as long as you describe a hydrogen atom with its monarchistic center. One commits no great error by keeping the center static. None of these equations are any good for the positronium, the deuteron or to describe relativistic proton-proton collisions.

We may try to write a wave function for both particles and introduce a Klein-Gordon operator for each one. This leads to

$$[\Box_1 + m_1^2][\Box_2 + m_2{}^2]\psi_o(x_1, x_2) = I(x_1, x_2)\psi_o(x_1, x_2) \quad , \qquad (II.9)$$

which is the Bethe-Salpeter[2] equation. This includes the fact that we do not know $I(x_1, x_2)$. At least for $I \equiv 0$ this equation describes the motion of two free particles.

If you wish an additive set of equations you may carry out the following consideration. Assume you have a spinless particle "one" in the potential of particle "two":

$$[\Box_1 + m_1^2]\psi_1(x_1) = V(x_1, x_2)\psi_1(x_1).$$

The same holds for particle "two" in the field of "one":

$$[\Box_2 - m_2^2]\psi_2(x_2) = V(x_2, x_1)\psi_2(x_2) .$$

Multiplying the two equations together you find essentially (II.9).

What is the interaction? In non-relativistic nuclear physics the Yukawa potential derived from the exchange of a single particle has been quite successful.

$$V(|\underline{r}|) = \pm g^2 \frac{e^{-\mu|\underline{r}|}}{|r|} \quad . \tag{II.10}$$

Here g^2, the coupling constant, gives the strength of the interaction and μ^{-1} its range. The + sign is for repulsion, the (-) for attraction. The 3-dimensional Fourier transform is

$$V(\underline{q}) = \frac{\pm g^2}{\underline{q}^2 + \mu^2} \quad . \tag{II.11}$$

This can be considered as the $q_0 = 0$ limit of the covariant expression

$$V(q) = \frac{\mp g^2}{q^2 - \mu^2} \tag{II.12}$$

which is the interaction used in the so-called ladder approximation to the BS equation.

We can also bring (II.9) into p-space. Translation invariance requires both $I(x_1, x_2)$ and $\psi(x_1, x_2)$ to depend on x_1 and x_2 through their difference $x_1 \cdot x_2$ only. This leads to conservation of momentum which we take care of by setting

$$p_1 + p_2 = 2p$$

$$p_1 - p_2 = 2q. \tag{II.13}$$

We obtain with the attractive (+) sign of the potential (II.12)

$$[(p+q)^2 - m_1^2][(p-q)^2 - m_2^2]\,\psi_0(p,q) = \frac{ig^2}{(2\pi)^4}\int \frac{d^4k\,\psi_0(p,k)}{(q-k)^2 - \mu^2} . \tag{II.14}$$

This is the form of the equation as given by Nambu.[1] Let us introduce the substitution

$$[(p+q)^2 - m_1^2][p-q)^2 - m_2^2]\psi_0(p,q) = M_0(p,q) , \tag{II.15}$$

then the Klein-Gordon factors become propagators under the integral and we find

$$M_o(p,q) = \frac{ig^2}{(2\pi)^4} \int \frac{d^4k\, M_o(p,k)}{[(q-k)^2 - \mu^2][(p+k)^2 - m^2][(p-k)^2 - m^2]} \,. \tag{II.16}$$

Here all the denominators have to be interpreted as causal propagators which is achieved by giving each mass an infinitesimal negative imaginary part. This equation is shown graphically in Figure 1.

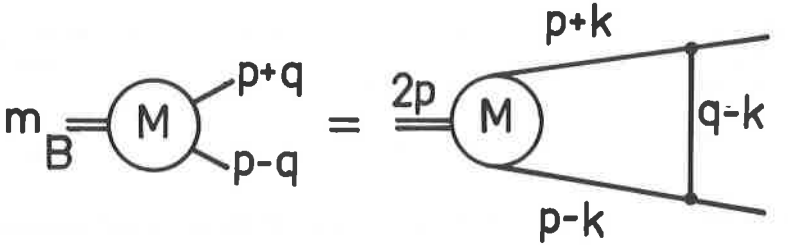


Fig. 1. The homogeneous BS equation.

Unless I am interested in the unequal mass case I will take $m_1 = m_2 = m$. Eq. (II.14) gives a relativistic generalization $\psi_o(p,q)$ of the bound state wave function

$$\psi_E(\underset{\sim}{q}) = \psi_o(p, q_o = 0, \underset{\sim}{q}) \,. \tag{II.17}$$

The factor i in (II.14) and (II.16) is a consequence of the Minkowsky metric; it disappears when you use an Euclidean metric.

The equation is covariant, therefore we can study it in any frame. In the center of mass, the rest frame of the bound state, we have

$$p = (p_o, \underset{\sim}{0}) = \left(\frac{m_B}{2}, \underset{\sim}{0}\right) \tag{II.18}$$

where m_B is the mass of the bound state.

A similar equation holds for the scattering of two particles. If we consider only the iterated presence of a one particle exchange potential (II.12) we get the ladder series (Fig. 2)

$$M = V + VG_oV + VG_oVG_oV + \ldots \,. \tag{II.19}$$

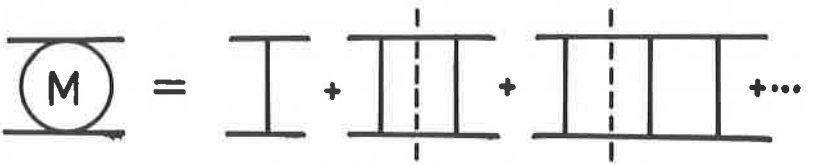


Fig. 2. The ladder series for the off shell scattering amplitude.

G_o describes the free propagation of the two particles between successive actions of the potential. This series may converge for weak enough coupling but we can sum it exactly and then consider the resulting integral equation as defining M for any coupling. This gives (Fig. 3)

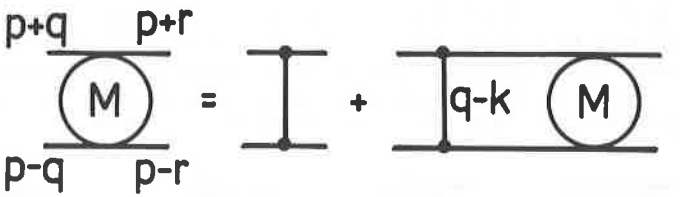


Fig. 3. The ladder approximation to the BS equation as obtained by summing all the graphs in Fig. 2.

$$M(p,q,r) = \frac{g^2}{(q-r)^2-\mu^2} + \frac{ig^2}{(2\pi)^4}\int \frac{d^4k\, M(p,k,r)}{[(q-k)^2-\mu^2][(p+k)^2-m^2][(p-k)^2-m^2]} \qquad \text{(II.20)}$$

This is the inhomogeneous or scattering counterpart of Eq. (II.16).

Equation (II.20) also follows from the $g\varphi^3$: theory with the Feynman rules:

$$\text{internal propagator} = \frac{-i(2\pi)^4}{p^2-m^2}$$

$$\text{vertex} = g\,\delta(p+q+k)/(2\pi)^4$$

$$\text{circle} = -(2\pi)^8\, M(p\pm q, p\pm r) \quad , \qquad \text{(II.21)}$$

and if one drops all nonladder graphs. This actually means the neglect of many more graphs than have been kept. The only justification here is that otherwise we would not obtain such a nice equation.

A partial justification for this approximation is given by Taylor's[10] non-perturbative derivation of the BS equation. The LSZ[11] reduction formula gives the T-matrix as the mass shell limit of the time ordered products or their vacuum expectation values. For the scattering matrix we go out from the four point function

$$M = \langle 0 | T(\varphi(x_1)\,\varphi(x_2)\,\varphi(x_3)\,\varphi(x_4)) | 0 \rangle \quad , \qquad \text{(II.22)}$$

while for the bound state wave function we go out from

$$\psi = \langle 0 | T(\varphi(x_1)\,\varphi(x_2)) | B \rangle \quad .$$

Unitarity of the S-operator leads to the τ-function equations, a non-linear and infinite coupled set. Taylor and Symanzik[12] have given a systematic linearization of these equations. I will begin with the M function from which I assume all self energy insertions in the external legs have been removed by premultiplication with appropriate factors. Then we insert, into the first possible position from the right, a one particle intermediate state. The element then decomposes into two parts (Fig. 4). The part on the right in Fig. 4 contains no

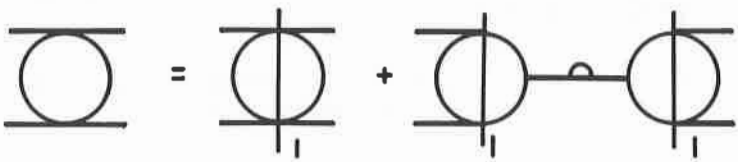


Fig. 4. Exposition of the single particle intermediate state and the one particle irreducible amplitudes.

more stable one particle intermediate state. In the first term we have exposed this state explicitly. The little bubble in the propagator has been inserted to indicate that this is a full propagator including all self energy insertions. Note that both vertices in this term are one particle irreducible. This follows simply because we could have started exposing the one particle state from the left. Clearly sometimes conservation laws may prevent us from having such a state at all. The recognition and exposition of this one particle state is necessary in principle. It distinguishes particle physics from potential scattering where the number of particles remains constant. That nothing happens if we do not expose this one particle state before going to the BS equation is a consequence of the Fredholm alternative which I will discuss below.

To get the BS equation for the one particle irreducible part we expose in it the first two body intermediate state from the right. (Fig. 5.) Note that the matrix elements denoted by bar-2 in Fig. 5 do not

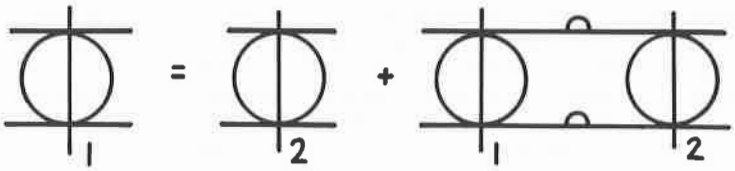


Fig. 5. The BS equation for the one particle irreducible amplitude as obtained by exposing the first two particle intermediate state from the right. The two parallel propagators are full propagators and the two particle irreducible functions are taken as the potentials.

contain any one or two particle intermediate states. We could now go on and expose the 3-particle state and so on. We arrive at the BS approximation if we interpret the 2-irreducible expression as input or potential V_2. We get the BS equation in the bare ladder approximation if we replace V_2 by (II.12) and drop the self energy corrections in the two propagators. This equation then has the Lippman-Schwinger form

$$M = V_2 + iV_2 G_2 M \ . \qquad \text{(II.23)}$$

Equation (II.20) is of interest in the scattering region and for not too high energies because we have neglected multiparticle intermediate states. As pointed out by Gell-Mann and Low[8] $M(p;q,r)$ also has all the bound state poles. If (II.16) has an eigenvalue corresponding to a bound state of mass m_B we can factor out its propagator which is a function of the total energy s only. In the neighbourhood of a bound state the off shell amplitude factorizes and for an s-wave bound state can be written

$$M(p,q,r) = \frac{\psi(q^2)\,\psi(r^2)}{s - m_B^2} + \text{terms regular at } s = m_B^2 \ . \qquad \text{(II.24)}$$

which is recognized as being identical to Taylor's decomposition Fig. 4. Therefore if we go with the total energy $s = 4p^2$ near such a pole only the two terms in (II.17) which contain M become large while the Born term remains finite and can be neglected. More rigorously: If we multiply (II.17) by $s - m_B^2$ and take the limit $s \to m_B^2$, the homogeneous eigenvalue equation which has the significance of a bound state equation results.

III. Relationship to the Schrödinger Equation

Because the BS equation (II.17) is quite complicated to solve it is most useful to see exactly what we would neglect in order to get a simpler equation and eventually back to the Schrödinger equation. We will give such a derivation and make as few approximations as possible. We do not use the Wick rotation. In the course of this derivation we will arrive at an intermediate equation[7] which, after partial wave projection, is a one dimensional integral equation and has still relativistic kinematics. This equation is considered to give a better approximation to the BS equation because no weak binding condition has been imposed. The exchange propagator in Eq. (II.16) has the explicit form

$$(q_o - k_o)^2 - (\underset{\sim}{q} - \underset{\sim}{k})^2 - \mu^2 \qquad \text{(III.1)}$$

and it is here that the right hand side of (III.1) has its only q_o-dependence. The first approximation which we introduce is to neglect retardation by putting

$$q_o = k_o \quad . \qquad \text{(III.2)}$$

We get from (II.17)

$$\varphi(p,\underset{\sim}{q}) = \frac{-ig^2}{(2\pi)^4} \int \frac{d^3k\, \varphi(p,\underset{\sim}{k})}{(\underset{\sim}{q}-\underset{\sim}{k})^2+\mu^2} \int_{-\infty}^{\infty} \frac{dk_o}{[(p+k)^2-m^2][(p-k)^2-m^2]} \qquad \text{(III.3)}$$

Here we have made use of the fact that after the approximation (III.2) has been made $\psi_o(p,q) = \varphi(p,\underset{\sim}{q})$ no longer depends on q_o, therefore $\psi(p,k) = \varphi(p,\underset{\sim}{k})$ no longer depends on k_o. It can therefore be taken out of the k_o integral. The last integral can be done by contour integration. Setting $\omega^2 = \underset{\sim}{k}^2 + m^2$ we find, suppressing the p-dependence,

$$\varphi(\underset{\sim}{q}) = \frac{-g^2}{4(2\pi)^3} \int \frac{d^3k\, \varphi(\underset{\sim}{k})}{[(q-k)^2+\mu^2]\, \omega\, (p^2-\omega^2)} \quad . \qquad \text{(III.4)}$$

This equation no longer contains the relative $q_o - k_o$ coordinate and after partial wave projection it is a one dimensional integral equation only. We will have to make several additional approximations to get to the Schrödinger equation. (III.4) is therefore of great interest for the discussion of relativistic problems. Our derivation[7] has been carried out in the bound state region, $s < 4m^2$. Blankenbecler and Sugar[13] found the inhomogeneous counterpart in the scattering region by constructing a unitary approximation to the BS equation as follows: The only energy (s) dependence of (II.18) is in the two parallel propagators. Cutkosky's rules tell us that we get the discontinuity of M

across its elastic cut by replacing these propagators by $(2\pi i)^2$ times the corresponding δ functions. Blankenbecler and Sugar observe that we get a function with the same discontinuity if we write for the two particle Green's function

$$G_2 = 2\pi i \int_{4m^2}^{\infty} \frac{ds'}{s'-s}\, \delta[(p+q)^2 - m^2]\,\delta[(p-q)^2 - m^2] + R_2 . \quad \text{(III.5)}$$

Here $s = 4p^2$ is the total center of mass energy. The unknown remainder R_2 does not have the two particle cut and is dropped. They get their approximate equation by substituting this in place of the free Green's function in the BS equation. The product of the two δ-functions can be simplified to read

$$\delta[(p+k)^2 - m^2]\,\delta[(p-k)^2 - m^2] = \frac{1}{8p_0 w}\, \delta(k_0)\, \delta(w - p_0) . \quad \text{(III.6)}$$

With this we can do the integral in (III.6) and find the approximate Green's function

$$G_2' = \frac{2\pi i\, \delta(k_0)}{4w(w^2 - p^2)} . \quad \text{(III.7)}$$

Substitution into (II.17) leads to (III.4) with an inhomogeneous term.

To proceed with our derivation of the Schrödinger equation we now make the weak binding assumption

$$2p_0 \sim 2m$$

or

$$p_0 = m - \frac{B}{2}, \qquad B \ll m , \quad \text{(III.8)}$$

where the binding energy B is given by

$$-B = \sqrt{s} - (m_1 + m_2)^2 , \quad B \geqslant 0 . \quad \text{(III.9)}$$

The last denominator in (III.4)

$$p^2 - w^2 = p^2 - m^2 - \underset{\sim}{k}^2 = (p-m)(p+m) - \underset{\sim}{k}^2$$

can now be approximated by

$$-\frac{B}{2}\, 2m - \underset{\sim}{k}^2$$

$$= -(mB + \underset{\sim}{k}^2) = -m\left(B + \frac{\underset{\sim}{k}^2}{2m_r}\right) , \quad \text{(III.10)}$$

where m_r is the reduced mass (II.2). To get finally the Schrödinger equation we remove the last trace of relativistic kinematics by using the static (!) limit to write

$$\omega^{-1} \sim m^{-1} \tag{III.11}$$

and find

$$\varphi(\underset{\sim}{q}) = \lambda' \int \frac{d^3k}{(\underset{\sim}{q}-\underset{\sim}{k})^3 + \mu^3} \cdot \frac{\varphi(\underset{\sim}{k})}{B + \frac{\underset{\sim}{k}^2}{2m_r}} \tag{III.12}$$

with $\lambda' = g^3 [\, 4m^3 (2\pi)^3 \,]^{-1}$. The wavefunction

$$\psi(\underset{\sim}{q}) = \varphi(\underset{\sim}{q}) \left[B + \frac{\underset{\sim}{q}^2}{2m_r} \right]$$

is now seen to satisfy (II.4). Aside from dropping the relativistic kinematics it was the weak binding assumption which has led from (III.4) to (II.4).

IV. The Structure of the BS Kernel

Before we can give meaningful solutions of the BS equation we have to study the analytic properties of the kernel as a function of the variables of integration, especially k_0, and of the parametric energy p_0. Also we will see that even after partial wave analysis we are still left with a two dimensional integral equation. This could be solved by Gaussian integration with reasonable accuracy if the kernel function were reasonably smooth. But the kernel has actually an infinite number of singularities either in the k-k_0 plane of integration or nearby. Fortunately only a few of them bother us at low energies. Therefore we will study first the location of these most relevant singularities.

Each of the propagators leads to two poles in k_0. The parallel ones give the four poles at

$$k_0 = -p - \omega + i\varepsilon,\ p - \omega + i\varepsilon \tag{IV.1}$$

and

$$k_0 = -p + \omega - i\varepsilon,\ p + \omega - i\varepsilon\ . \tag{IV.2}$$

Here we have given explicitly the $\pm i\varepsilon$ of the causal propagators. The poles (IV.1) are in the upper, the ones in (IV.2) in the lower k_0-plane. Their position is independent of the external variable q_0 but they move as we change the energy p_0. Figure 6 shows the position of these four

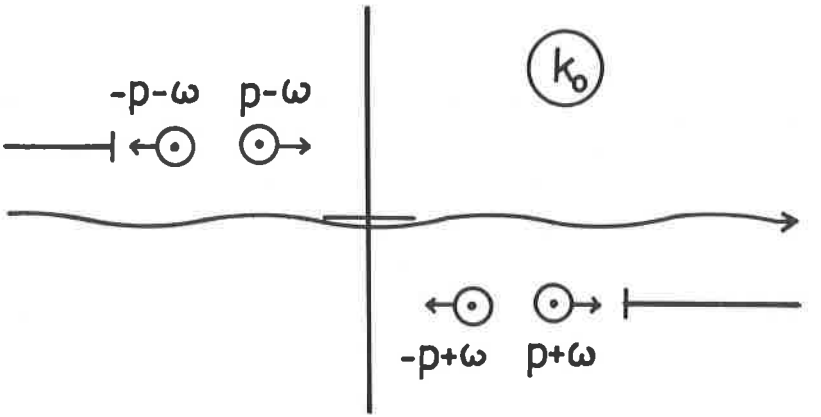


Fig. 6. The position of the contour of integration and of the singularities of the kernel of the BS equation in the k_o-plane. The little arrows show how the four poles move as p_o is increased.

poles in the bound state region $2p_o < 2m$ or $s < 4m^2$. The arrows indicate their motion as we increase s (or p_o). The two poles which are closest to the imaginary axis pinch the contour of integration when

$$p - \omega = - p + \omega \quad . \tag{IV.3}$$

This can happen first at

$$2p_o = 2m \tag{IV.4}$$

and produces the elastic threshold at $s = 4m^2$ in $\psi(q,s)$. There is a similar pinch of the other two poles for $p_o = -m$. For $s < 4m^2$ we have no singularities in quadrants I and III (Fig. 6). For $4m^2 < s < (2m+\mu)^2$ we find a pole each in quadrant I and III.

There are also two poles from the exchange propagator at

$$k_o = q_o \pm \sqrt{(\underset{\sim}{q} - \underset{\sim}{k})^2 + \mu^2} \quad . \tag{IV.5}$$

Their position is independent of p_o but depends on the variables q_o, $\underset{\sim}{q}$. Again the μ^2 has a $-i\varepsilon$. The inelastic threshold in p_o is generated if the pole in the first quadrant moves on, as we increase p_o, to pinch

the contour of integration against one of the branch points created by (IV.5). This happens when

$$q_0 + \sqrt{(\underset{\sim}{q}-\underset{\sim}{k})^2 + \mu^2} = p_0 - \sqrt{\underset{\sim}{k}^2 + m^2} \quad , \tag{IV.6}$$

an expression which has to be minimized with respect to $\underset{\sim}{q}\cdot\underset{\sim}{k}$. After Wick rotation $q_0 \to iq_4$ and we see that the singularities of the exchange propagator can be on the real axis only for $q_0 = 0$. This creates the two cuts shown in Fig. 6. As the pole moves on it creates the higher thresholds in p_0 and q_0 corresponding to more exchanged particles being able to come onto the mass shell.

The easiest method of removing the singularities from the path of integration is the so-called Wick rotation.[5] One rotates the contour of integration in a clockwise to the imaginary axis. This can be done without difficulties only if two conditions are fulfilled:

1. The infinite part of the contour does not contribute.
2. No singularities in quadrants I and III.

In the present equation power counting insures us that the integrand falls off fast enough so that the infinite circle does not contribute. This is not known,[14] for the case of two spin $\frac{1}{2}$ particles. As long as we are in the bound state region we can rotate the contour without difficulty because we have shown that there are no singularities in quadrants I and III. In the elastic region the rotation can still be carried out but one has to take care of the residues, containing an unknown function, which are picked up. After the rotation one returns to real variables of integration by the substitutions

$$\begin{aligned} q_0 &= iq_4 \\ k_0 &= ik_4 \end{aligned} \tag{IV.7}$$

which rotate the entire k_0-plane. Note that the substitution $p_0 \to ip_4$ can be carried out but is not needed.

For a more detailed discussion we refer to the literature.[15]

V. Two and Three Particle Unitarity

It is interesting to see that the BS equation satisfies elastic unitarity and if the self energy insertions in the parallel propagators are kept even three particle unitarity.[15),16] This is quite surprising since the BS equation is not an exact consequence of a hermitian Hamiltonian as is the case for the Schrödinger equation or the Vθ-equation in the Lee model.

To prove elastic unitarity we rewrite (II.18) in operator notation

$$M = I + iIGM \ . \tag{V.1}$$

This has the formal solution

$$M = (1-iIG)^{-1} I. \tag{V.2}$$

Here the s dependence of M is entirely in the G-function of the resolvent. We also have

$$M^*(s) = (1 + iIG^*)^{-1} I, \tag{V.3}$$

for the function M(s) evaluated at a point across the elastic cut for $4m^2 < s < (2m+\mu)^2$. We compute the discontinuity across this cut as

$$M - M^* = \frac{1}{1-iGI} I - \frac{1}{1+iIG^*} I \ . \tag{V.4}$$

Inserting unity in the proper places this can be rewritten as

$$\frac{1}{1-iIG}(1+iG^*I)\frac{1}{1+iG^*I} I - \frac{1}{1-iIG}(1-iIG)\frac{1}{1+iIG^*} I \ .$$

Using (IV.3) for the last terms we find

$$M - M^* = \frac{1}{1-iIG}[1+iIG^* - 1+iIG] M^* \tag{V.5}$$

The real part of G is

$$2ReG = G + G^* = (2\pi i)^2 \delta(p_1^2 - m^2)\, \delta(p_2^2 - m^2) \tag{V.6}$$

where one has to take proper care of the coinciding singularities in the principal values of G. After cancellation we find from (IV.5) and (IV.6)

$$M(s) - M^*(s) = - 4\pi^2 \int d^4k \, M(p,k,r)\, \delta\delta M^*(p,k,r) \ , \tag{V.7}$$

which is the expression of elastic unitarity; top line of Fig. 7. Note that for the partial wave amplitude the integrals in (IV.7) can be carried out.

For $s > (2m+\mu)^2$ unitarity requires that the discontinuity across the inelastic cut be given by the entire expression in Fig. 7. We see that this would lead out of the framework of the BS equation in that it involves a production amplitude. We cannot neglect this, however, because we certainly do have three particle intermediate states. Now

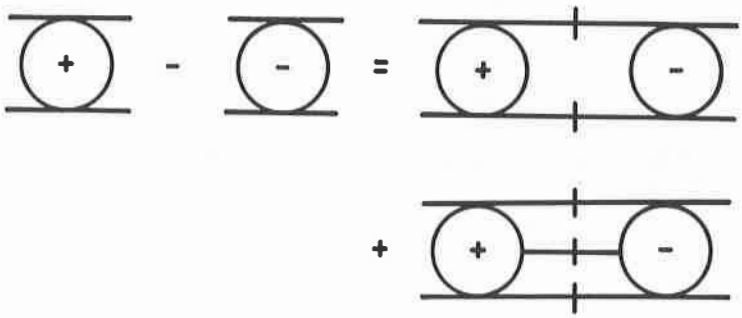


Fig. 7. The discontinuity equation. The top line gives the discontinuity in the elastic region. The second line has to be added in the singly inelastic region.

we can write the most general production amplitude as the sum of two terms. Fig. 8. The first one contains any contribution which can be factorized into an elastic amplitude times a vertex. The second term is called irreducible because it contains all the terms which cannot be factorized. For the BS equation the irreducible part is neglected, hence has been put equal to zero. Therefore we get to the unitarity relation all the possible contributions from the first diagram in Fig. 8. These are:

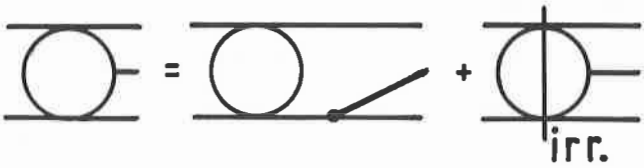


Fig. 8. Separation of the production amplitude into a part which factorizes and a part which cannot be written in this form.

a. the self energy insertions (which we should have kept in the first place)

b. the exchange terms

as shown in Fig. 9. Therefore if we retain the renormalized self energy insertions in the parallel propagators our equation will satisfy inelastic unitarity. This guarantees that just the proper amount of particles get lost into the inelastic channel and we have

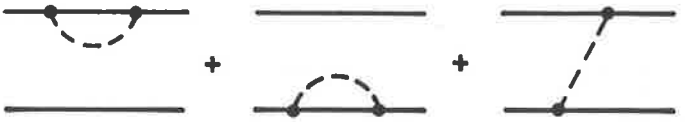


Fig. 9. All intermediate states contributing to three particle unitarity in the BS equation.

$$\sigma_{el} \leq \sigma_{tot} \quad . \tag{V.8}$$

For the φ^3 theory the propagator obtained by summing all the self energy bubble graphs can be renormalized and written as

$$\Delta_c'(q^2) = [\, q^2 - m^2 + i\epsilon \,]^{-1} [\, 1 + (q^2 - m^2) A(q^2) \,]^{-1}$$

with

$$A(q^2) = \lambda \int_{(m+\mu)^2}^{\infty} \frac{ds}{s - q^2 - i\epsilon} \cdot \frac{\sqrt{\frac{s-(m+\mu)^2}{s}}}{(s-m^2)^2} \quad . \tag{V.9}$$

VI. The Alternative

The BS equation is a linear integral equation. Therefore if we find a special solution of the inhomogeneous equation we could always add any solution of the homogeneous equation multiplied with an arbitrary coefficient. This makes the solution non-unique. If the equation has a compact kernel one has the Fredholm alternative which says that either the inhomogeneous equation has a unique solution or else the homogeneous equation has a solution.

The kernel of the BS equation as it stands is not square integrable even in the bound state region because the contour of integration encounters singularities. Therefore Fredholm theory has nothing to say. It can also be shown easily by examples that singular integral equations need not have a discrete spectrum in λ. It is therefore of utmost importance that one can arrive at the same alternative by alternate means and it is interesting to see that is is unitarity which requires the BS equation to have a reasonable spectrum.

a) In the bound state region for $s < 4m^2$ we can carry out a rotation of the k_o-contour of integration. This leaves p_o real. Then none of the propagators becomes singular in (II.17) on the path of integration. The equation is now Fredholm and it has a real kernel at least in the equal mass case. The kernel can be shown to be self adjoint

(it is real and symmetric) therefore we have a real and discrete eigenvalue spectrum; both in λ and p_o^2. Note that also negative values of p_o^2 are permitted.

b) In the elastic region, for $4m^2 < s < (2m+\mu)^2$ the kernel becomes complex but here it cannot have any eigenvalues. If we had such a solution of the homogeneous equation we could use it to construct a bound state propagator (II.24). Suppose we had an s-wave pole (the others work similarly). We have then[17]

$$M_o = \frac{\Gamma^2}{s - m_B^2 + i\epsilon} \tag{VI.1}$$

which is not of the N/D form required by unitarity because it gives

$$|M|^2 = \frac{\Gamma^4}{|s - m_B^2|^2} \tag{VI.2}$$

and

$$\text{Im } M_o = \Gamma^2 \, \delta(s - m_B^2) \quad . \tag{VI.3}$$

Collecting we find that the solution of the BS equation is unique, a statement which is proven up to the singly inelastic region and thus it makes sense to try to apply numerical methods to its solution.

VII. An Analytic Solution

Wick[5] gave exact analytic solutions for the case of zero mass exchange and zero total energy:

$$\mu = 0 \tag{VII.1}$$

$$p = (0, \underset{\sim}{0}) \quad . \tag{VII.2}$$

In this case the BS equation becomes

$$(q^2 - m^2)^2 \varphi_o(q) = \frac{ig^2}{(2\pi)^4} \int \frac{d^4k}{(q-k)^2} \, \varphi_o(k) \quad . \tag{VII.3}$$

This equation has full $O(3,1)$ and, after Wick rotation, $O(4)$ symmetry and its solutions can be classified accordingly. Wick[5] gave the full set of solutions. We need here the ground state only which we assume to be spherically symmetric.

Eq. (VII.3) is solved by acting on it with the 4-dimensional Laplacian and observing the identity

$$\Box_{q^2}(q^2 - k^2)^2 = -4\pi^2 i\,\delta^{(4)}(q-k) \qquad \text{(VII.4)}$$

and, with $s = q^2$,

$$\Box_{q^2} = 4s\frac{\partial^2}{\partial s^2} - 8\frac{\partial}{\partial s} + \text{angular terms} . \qquad \text{(VII.5)}$$

The angular terms are of interest only if one studies excited states.

We introduce the dimensionless coupling λ by

$$m\lambda = g^2(4\pi)^{-2} \qquad \text{(VII.6)}$$

and find the partial differential equation

$$\Box_{q^2}\varphi_o(q) = 4\lambda\,\varphi_o(q) . \qquad \text{(VII.7)}$$

For the ground state this is solved by the ansatz

$$\varphi_o(q) = (q^2 - m^2)^{-3} \qquad \text{(VII.8)}$$

if the characteristic equation

$$\lambda = 2 \qquad \text{(VII.9)}$$

is fulfilled.

In the next section we will use this analytic solution to produce a much needed and very accurate approximate solution for nonzero but very small exchange mass as well as small but nonzero bound state mass.

VIII. An Expansion Around the Analytic Solution

For the situation where the exchange mass μ is nonzero but much smaller than the propagator mass m ($m/\mu > 20$) the potential in the BS equation becomes near singular. One has then two characteristic dimensions in the problem; the effective range of integration which is a few m^2 and the region over which the integrand is large which is a few μ^2. This sets severe limitations on the direct numerical methods we discuss below. It is therefore important that we can give[14] a simple expansion around Wick's solution which gives quite accurate results.

We write the BS equation in the general form

$$[(p+q)^2 - m_1^2][(p-q)^2 - m_2^2]\,\varphi(q) = \frac{i\lambda m^2}{\pi^2}\int\frac{d^4k\,\varphi(k)}{(q-k)^2 - \mu^2} , \qquad \text{(VIII.1)}$$

with the slightly different masses

$$m_1 = m + \Delta_1$$

$$m_2 = m + \Delta_2 \ . \qquad \text{(VIII.2)}$$

We assume that p_0, μ, Δ_1 and Δ_2 are much smaller than m. For $p_0 = \mu = \Delta_1 = \Delta_2 = 0$ Wick's solution (VII.8) satisfies Eq. (VIII.1) and if these parameters are nonzero but small we can expect to get a good approximation for λ if we replace φ by φ_0 in Eq. (VIII.1). In order to reduce Eq. (VIII.1) to an algebraic equation, we multiply both sides by a function of q^2 and integrate over d^4q. This function has to ensure convergence and for convenience we have chosen φ_0 (VII.8), but the final result depends little on the specific form of this function. We observe that in Eq. (VIII.1), μ and $m_B = 2p_0$ occur quadratically, while Δ_1 and Δ_2 enter linearly (and, for symmetry reasons, only as their sum). We therefore seek a perturbative formula of the form

$$\lambda(\mu^2, m_B^2, \Delta_1+\Delta_2) = \lambda(0,0,0) + \left(\frac{m_B}{m}\right)\lambda_1 + \left(\frac{\Delta_1+\Delta_2}{m}\right)\lambda_2 + \left(\frac{\mu}{m}\right)^2\lambda_3$$

$$+ \mathcal{O}\left(m_B^4, \mu^4, m_1^2, m_2^2\right) . \qquad \text{(VIII.3)}$$

$\lambda(0,0,0)$ is of course 2, which is Wick's value. By carrying out the integrals arising from the left-hand side of Eq. (VIII.1) as Feynman integrals, we obtain for the second and third terms

$$\lambda_1 = -\tfrac{2}{5} \ , \quad \lambda_2 = 2. \qquad \text{(VIII.4)}$$

The fourth term is somewhat tricky to evaluate since the inner integrand in

$$\lambda(\mu) = \frac{\pi^4}{6m^4}\left[\int \frac{d^4q}{(q^2-m^2)^3}\int \frac{d^4k}{(k^2-m^2)^3}\cdot\frac{1}{(q-k)^2-\mu^2}\right]^{-1} \qquad \text{(VIII.5)}$$

has no analytic expansion around $\mu = 0$. This is understandable since for small μ Eq. (VIII.5) is expected to behave like $\mu^2 \log\mu^2$ due to the singularity in the exchange propagator. We shall therefore calculate $\lambda(\mu) - \lambda(0)$, to be identified with the fourth term of Eq. (VIII.3). By going through the standard techniques, we obtain from (VIII.5)

$$\lambda(\mu) = \frac{1}{3m^4}\left[\int_0^1\int_0^1 \frac{dx\,dy\,x^3y^3(1-x)(1-y)}{[m^2x(1-xy)+\mu^2(1-x)(1-y)]^3}\right]^{-1} , \qquad \text{(VIII.6)}$$

which is evaluated numerically and fitted to a linear dependence on $\mu^2/m^2 \log\mu^2/m^2$ which is valid over a wide range of values of μ/m.

We arrive at the relationship

$$\lambda = 2.0 - \frac{2}{5}\left(\frac{m_B}{m}\right)^2 + 2\,\frac{(\Delta_1+\Delta_2)}{m} + \frac{\mu^2}{m^2}\left[a \log \frac{\mu^2}{m^2} + b\right], \qquad \text{(VIII.7)}$$

with the values

$$a = -0.85, \qquad b = 1.69 . \qquad \text{(VIII.8)}$$

Relation (VIII.8) is algebraic and describes very accurately the connection between the coupling λ and the mass of the bound state. Moreover, since we did not use the Wick rotation we could apply it to the situation where we have different masses in the parallel propagators. We see that Eq. (VIII.7) predicts a larger coupling for a more tightly bound system and also a larger coupling for a more massive exchange particle. This is in accordance with what we expect since, for a massive exchange particle, the potential becomes less singular. We may comment that for large values of m/μ the perturbative formula is more reliable and more accurate than a numerical solution of Eq. (VIII.1), while for $m \sim \mu$ it should not be used. In the special case of $\mu = 0$ a similar formula was apparently used by Gürsey, Lee and Nauenberg[18] in deriving a quadratic mass formula for mesons.

IX. The Bound State Region

In this section I will discuss some concrete numerical techniques which one can apply to solve the BS equation in the bound state region.[7] We go out from the homogeneous equation

$$[(p+q)^2 - m^2][(p-q)^2 - m^2]\,\psi(p,q) = \frac{ig^2}{(2\pi)^4}\int \frac{d^4k\,\psi(p,k)}{(q-k)^2-\mu^2} , \qquad \text{(IX.1)}$$

which can be considered as a relativistic bound state equation. It is an eigenvalue problem in p_o if g is given. We introduce again the dimensionless coupling constant λ by (VII.6) and try to find $\lambda(p_o)$. To solve the eigenvalue problem one does not have to solve the entire equation (IX.1) but has only to study the Fredholm denominator. The solution of (IX.1) or finding ψ is then another step.

First we observe that Eq. (IX.1) is unchanged under the substitution $q_o \to -q_o$. Therefore if $\psi(\underline{q}, q_o)$ solves it so does $\psi(\underline{q}, -q_o)$. We can take sum and difference and have both an even and odd (in q_o) solution.

$$\psi(\underline{q}, q_o) = \pm\psi(\underline{q}, -q_o) \qquad . \qquad \text{(IX.2)}$$

The incoming particles are on the mass shell if

$$q_o = 0$$

$$\underline{q}^2 = (s - 4m^2)/4 \ . \qquad \text{(IX.3)}$$

Note that on the mass shell where $q_o = 0$ the odd solution vanishes. Nevertheless it may find some physical interest but we shall restrict ourselves here to the even solution.

Since angular momentum is conserved in Eq. (IX.1) we can reduce the number of variables of integration by carrying out a partial wave analysis. Here we have two possibilities. Either we expand in 3-dimensional harmonics or, after Wick rotation, in 4-dimensional spherical harmonics. The first one leads to a single but two dimensional integral equation; the second one to an infinitely coupled set of one dimensional equations. We study the second expansion later and proceed with an ordinary partial wave analysis. Let

$$\psi_\ell(q_o\ , |\underline{q}|; s) = \tfrac{1}{2} \int_{-1}^{1} dz\ \psi(q,p)\ P_\ell(z) \qquad \text{(IX.4)}$$

and

$$\psi(q,p) = \sum_{\ell=0}^{\infty} (2\ell+1)\ \psi_\ell(q_o\ , |\underline{q}|; s)\ P_\ell(z) \qquad \text{(IX.5)}$$

where the $P_\ell(z)$, $z = \cos\theta$ are the Legendre polynomials. Using their orthonormality Eq. (IX.1) becomes

$$[\,(p+q)^2 - m^2\,][\,(p-q)^2 - m^2\,]\ \psi_\ell(q_o\ , |\underline{q}|; p_o)$$

$$= \frac{i\lambda m^2}{\pi^2} \int \frac{d^4k\ P_\ell(z)}{(q-k)^2 - \mu^2}\ \psi_\ell(k_o\ , |\underline{k}|; p_o)\ . \qquad \text{(IX.6)}$$

The angular integrations can be done leading to the two variable integral equation

$$[\,(p+q)^2 - m^2\,][\,(p-q)^2 - m^2\,]\,\psi_\ell(q_o, |\underline{q}|; p_o)$$

$$= \int_{-\infty}^{\infty} dk_o \int_0^{\infty} dk\, \frac{|\underline{k}|}{|\underline{q}|}\, I_\ell(q_o\ , |\underline{q}|; k_o\ , |\underline{k}|; p_o)\, \psi_\ell(k_o\ , |\underline{k}|; p_o)\ , \qquad \text{(IX.7)}$$

where we have set

$$I_\ell(q_o\ , |\underline{q}|; k_o\ , |\underline{k}|; p_o) = \frac{2\lambda m^2}{\pi}\, Q_\ell\left(\frac{\underline{q}^2 + \underline{k}^2 - (q_o - k_o)^2 + \mu^2 - i\varepsilon}{2|\underline{q}|\ |\underline{k}|}\right). \qquad \text{(IX.8)}$$

Now we carry out the Wick rotation whence the product on the left becomes

$$[(p_o+q_o)^2-\omega^2][(p_o-q_o)^2-\omega^2] \to (q_4^2+\omega^2-p_o^2)^2+4p_o^2q_4^2 \geq 0 \ . \qquad (IX.9)$$

This is a real form only in the equal mass case. For $s < 4m^2$ it is positive definite. The BS equation becomes

$$[(-p_o^2+q_4^2+q^2+m^2)^2-4p_o^2(q^2+m^2)]\,\psi_\ell(iq_4,|\underset{\sim}{q}|;p_o)$$
$$=\int_o^\infty dk_4 \int_o^\infty dk \frac{|\underset{\sim}{k}|}{|\underset{\sim}{q}|}[I_\ell(iq_4,|\underset{\sim}{q}|;ik_4,|\underset{\sim}{k}|)-I_\ell(-iq_4,|\underset{\sim}{q}|;-ik_4,|\underset{\sim}{k}|)]$$
$$\times \psi_\ell(ik_4,|\underset{\sim}{k}|;p_o) \ . \qquad (IX.10)$$

Here we have used the evenness of $\psi(q_o \to -q_o)$ (IX.2) to restrict the integration in k_o from $0 \to \infty$. The potential, which is now Euclidean, and therefore the kernel in (IX.10) is real.

The Wick rotation and the manipulations described above have left us with a smooth kernel. Relation (IX.2) has greatly reduced the number of meshpoints needed. For use on the computer the equation is approximated by a finite set of coupled algebraic equations ($q_{oi} \to i$, $q_j \to j$)

$$\psi_{ij} = \lambda \sum_{\ell,m} K_{ij,\ell m}\, w_\ell w_m\, \psi_{\ell m} \ . \qquad (IX.11)$$

The w_i are the Gaussian weights. Matrix inversion and iteration work best if the kernel matrix is symmetric. There are even eigenvalue routines available for symmetric matrices. Therefore we define φ by

$$\varphi(q_4,q) = q[(m^2+p^2+q_4^2+\underset{\sim}{q}^2)^2-4p^2(q^2+m^2)]^{\frac{1}{2}}(w_i w_j)^{\frac{1}{2}}\psi(iq_4,q) \quad (IX.12)$$

where the variables are understood to be discrete. The resulting kernel matrix is symmetric and has finite elements. The substitution (IX.12) also removes the factor k/q in (IX.10).

In order to solve this equation on a computer we first map the infinite quadrant of integration into a finite one. From the resulting square ($0 < x < 1$, $0 < y < 1$) we select a finite number of meshpoints. Their number is severely limited by computer memory (storage of the large kernel matrix) and computer time (inversion of a matrix or evaluation of a determinant). A typical matrix which can still be inverted conveniently on present day computers is of the order of 100 x 100. This leaves us 10 points only per variable to describe the unknown

function. The maximal kernel we have used in this calculation was represented by a 12 x 12 x 12 x 12 array which was mapped into a 144 x 144 matrix. If the function $\psi(q_0, q; p_0)$ is not varying too wildly which is the case for not too much unequal masses m, μ and a finite distance below the elastic threshold we can expect quite accurate results by the use of Gaussian integration. This achieves that the integral and therefore the external function are represented by n points with the accuracy of a polynomial fit of order $2n - 1$. In view of the few meshpoints available Gaussian integration seems indispensable.

The smallest coupling λ (largest eigenvalue λ^{-1}) to get a bound state at a given energy squared can be found by iteration. The other eigenvalues can be found by plotting the determinant of the kernel matrix and locating its zeros,

$$\text{Det}\ (1 - \lambda K)_{\ell}(s) = 0 \ . \qquad \text{(IX.13)}$$

The results are presented in Fig. 10. Note that the curves continue smoothly through $s = 0$. Therefore there exist ghost states, with imaginary mass, for strong enough couplings.

X. Regge and Toller Trajectories

Toller[19] and Freedman and Wang[20] have shown that each parent Regge trajectory in an invariant scattering amplitude implies an infinite set of so-called daughter trajectories. Toller derived this result on the basis of $O(3,1)$ symmetry at (his) $t = 0$ while Freedman and Wang used $O(4)$ symmetry at $t = 0$.

It is well known that $O(3,1)$ and $O(4)$ are continuously connected by a transformation in the complex Lorentz group CL. On the other hand, the two groups are not trivially equivalent if applied to scattering functions. Among other things this is realized by observing that we can go from the original BS equation in Minkowsky space to the one in Euclidean space by carrying out a Wick rotation, even for the spinless case, without complications only if we are below the elastic threshold. In the elastic region, for the scattering amplitude and $\sqrt{s} > (m_1 + m_2)$, this is no longer possible without changing the equation. This is no contradiction to the above statement if we realize that it is only the Wightman function which is invariant under CL but not the time ordered products from which we get the scattering amplitudes.

I will present a short study of the trajectories as derived numerically from the BS equation. I find it useful, following to some extent Chung and Snider[21] to study both the 3-dimensional and the 4-dimensional partial wave analyses.

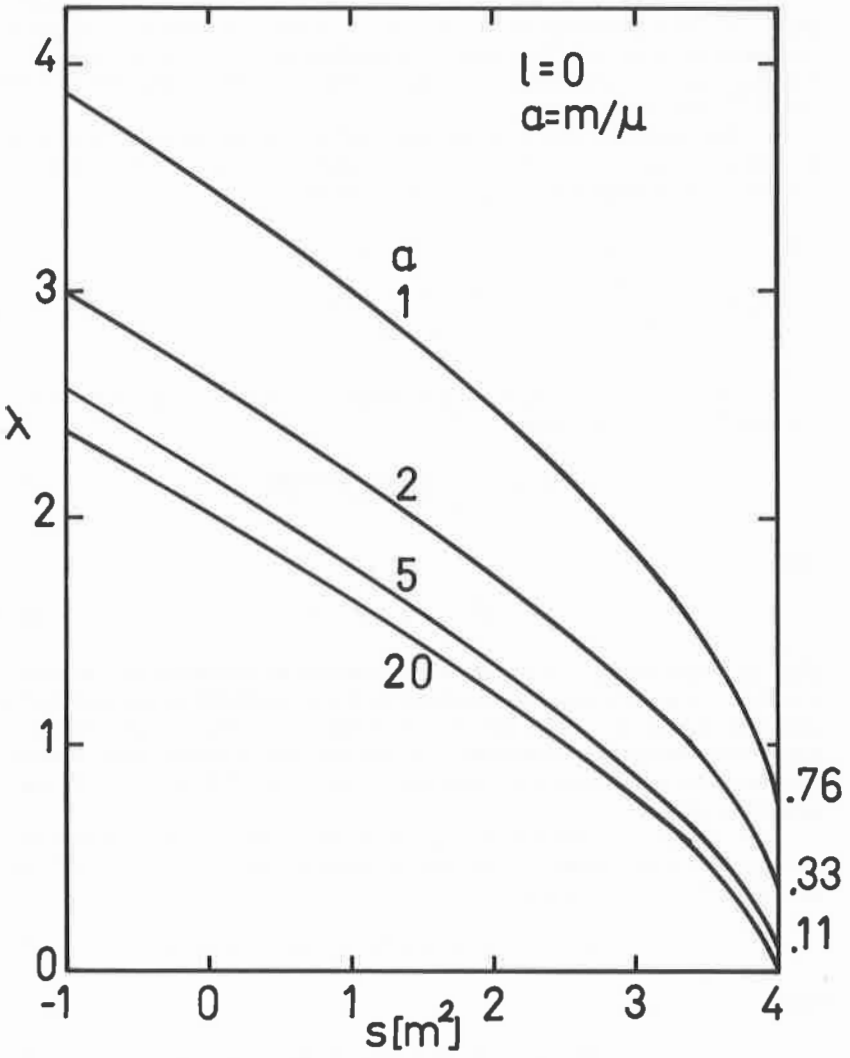


Fig. 10. The coupling $\lambda(s)$ for an s-wave bound state as a function of $s = m_0^2$. The curves have a smooth continuation to negative s.

Numerical studies for the ladder BS equation have been reported[21),22)] and perturbation studies of the first few daughter trajectories have been given[23)] by Fredholm method in the complex ℓ-plane. For the case of zero mass exchange a power series expansion is available[24)] near zero energy.

The Regge trajectories can be considered as the locations of the bound state poles for complex ℓ. Therefore I can go out from the ordinary partial wave amplitude derived earlier.

$$[(p+q)^2 - m_1^2][(p-q)^2 - m_2^2]\,\psi_\ell(q_0, |\underline{q}|; p_0)$$

$$= \frac{i\lambda m^2}{\pi^3} \int_{-\infty}^{\infty} dk_0 \int_0^{\infty} dk\, Q_\ell\left(\frac{\underline{q}^2 + \underline{k}^2 + \mu^2 - (q_0 - k_0)^2 - i\varepsilon}{2|\underline{q}|\,|\underline{k}|}\right) \psi_\ell(k_0, |\underline{k}|, p_0). \quad (X.1)$$

All we need is a numerically tractable integral representation for the Q_ℓ. This is given by[25)]

$$Q_\nu(z) = \int_0^{z^-} \frac{h^\nu\, dh}{\sqrt{1 - 2zh + h^2}} \quad (X.2)$$

with

$$z^\pm = z \pm \sqrt{z^2 - 1} \quad . \quad (X.3)$$

This is valid for Re $\nu > 0$ and can be integrated numerically for real $\ell > 0$. The equation is then solved with the methods presented earlier and reproduces the upper half $\ell > 0$ of Fig. 11. The integral values agree with earlier calculations. While this calculation does contain the daughter trajectories it does not show their classification in the analytic form.

In order to show the existence of daughter trajectories and to get small s and negative ℓ we use 4-dimensional harmonics. We go out from the Wick rotated equation

$$L(q,s)\, M(q,q';s) = I(q,q') + \int d^4 k_E\, I(q,k)\, M(k,q';s) \quad (X.4)$$

with

$$L(q,s) = [(p_0 + iq_4)^2 - \omega_1^2][(p_0 - iq_4)^2 - \omega_2^2] \quad (X.5)$$

$$\omega_1^2 = m_1^2 + \underline{q}^2$$

$$\omega_2^2 = m_2^2 + \underline{q}^2 \quad (X.6)$$

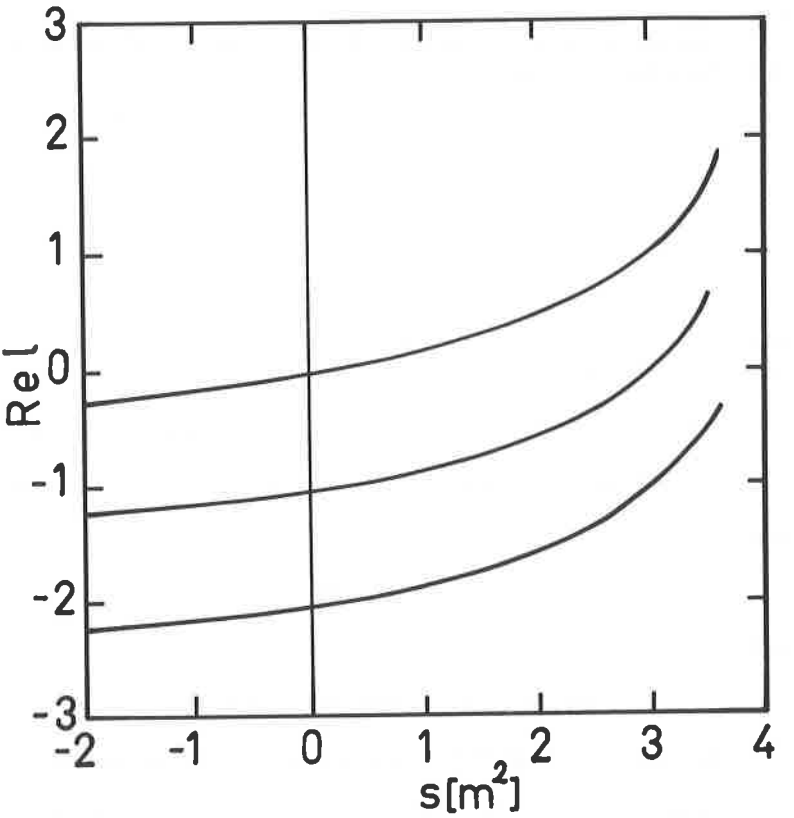


Fig. 11. One Regge and some daughter trajectories.

and

$$I(q,k) = \frac{\lambda m^2}{\pi^2} \frac{1}{(q-k)^2 + \mu^2} \ . \qquad (X.7)$$

Note that L is no longer real. We multiply it out to find

$$L(q,s) = (p_0^2 - q_4^2 - w_1^2)(p_0^2 - q_4^2 - w_2^2) + 4p_0^2 q_4^2 + 2ip_0 q_4 (m_1^2 - m_2^2) \ . \qquad (X.8)$$

In the unequal mass case this is complex for $s > 0$ and real for $s < 0$ ($\sqrt{s} = 2p_0$).

Let the length of a vector in E(4) be $q = (q_4^2 + \underline{q}^2)^{\frac{1}{2}}$ and introduce the polar coordinates

$$q_4 = q \cos\psi$$

$$q_3 = q \sin\psi \cos\theta$$

$$q_2 = q \sin\psi \sin\theta \cos\varphi$$

$$q_1 = q \sin\psi \sin\theta \sin\varphi \ . \qquad \text{(X.9)}$$

Here θ and φ are the old three dimensional angles while ψ denotes the new degree of freedom. We find

$$L(q,s) = (q^2 - s/4 + m_1^2)(q^2 - s/4 + m_2^2) + q^2 s \cos^2\psi + iq\sqrt{s}(m_1^2 - m_2^2)\cos\psi$$

$$= F_0(q_1^2 s) + q^2 s \cos^2\psi + iq\sqrt{s}\,\Delta\cos\psi \ , \qquad \text{(X.10)}$$

where

$$\Delta = m_1^2 - m_2^2$$

and

$$F(q^2, s) = (q^2 - s/4 + m_1^2)(q^2 - s/4 + m_2^2) \ .$$

We expand $M(q,q's)$ in terms of four dimensional harmonics

$$M(q,q';s) = \sum_{nn'\ell m} Y_{n\ell m}(\Omega_q) M_{nn'\ell m}(q,q';s)\, Y_{nn'\ell m}(\Omega_{q'}) . \qquad \text{(X.11)}$$

The Y are orthogonal over a four dimensional unit sphere and obey a simple addition theorem. They are given in terms of the three dimensional harmonics $y_{\ell m}$ by

$$Y_{n\ell m}(\psi,\theta,\varphi) = \zeta_n^{\ell}(\psi)\, y_{\ell m}(\theta,\varphi) \qquad \text{(X.12)}$$

and

$$\zeta_n^{\ell}(\psi) = \left[\frac{2^{2\ell+1}}{\pi}\,\frac{(n+1)(n-\ell)!}{(n+\ell+1)!}\right]^{\frac{1}{2}} \ell!\, \sin^{\ell}\psi\ C_{n-\ell}^{\ell+1}(\cos\psi) \ . \qquad \text{(X.13)}$$

The functions $C_n^m(x)$, $x = \cos\psi$, are the Gegenbauer functions. They are polynomials in x only if

$$n - m = K = \text{integer} \quad . \tag{X.14}$$

For m = 1 they can be obtained from the generating formula

$$(1 - 2xt + t^2)^{-1} = \sum_{n=0}^{\infty} C_n(x)\, t^n \quad . \tag{X.15}$$

Using the addition theorem we find that the exchange propagator has the expansion

$$\frac{1}{(q-k)^2+\mu^2} = \sum_{n\ell m} Y_{n\ell m}(\Omega_q) I_n(q,k) Y_{n\ell m}(\Omega_k) \tag{X.16}$$

with

$$I_n(q,k) = \frac{8\pi^2}{n+1} \left(\frac{r-s}{r+s}\right)^n \frac{1}{(rs)^2} \tag{X.17}$$

and

$$r = [\,(q+k)^2+\mu^2\,]^{\frac{1}{2}}$$
$$s = [\,(q-k)^2+\mu^2\,]^{\frac{1}{2}} \quad . \tag{X.18}$$

This is verified easiest by writing the exchange propagator as

$$\begin{aligned} \frac{1}{(q-k)^2+\mu^2} &= \frac{1}{q^2+k^2+\mu^2-2qk\cos\psi} \\ &= \frac{1}{2qk} \cdot \frac{1}{\frac{1+t^2}{2t} - x} \\ &= \frac{1}{2qk} \cdot \frac{2t}{1-2xt+t^2} \quad . \end{aligned} \tag{X.19}$$

This leaves us with the identification of

$$\frac{1+t^2}{2t} = \frac{q^2+k^2+\mu^2}{2qk} \tag{X.20}$$

which leads to (X.17,18).

We substitute our expansions into the BS equation which becomes the infinite set

$$F_0(q_1^2 s)M^{\ell}_{nn'} + q^2 s\left(f^{\ell}_{nn}M^{\ell}_{nn'} + f^{\ell}_{n,n+2}M^{\ell}_{n+2,n'} + f^{\ell}_{n,n-2}M^{\ell}_{n-2,n'}\right)$$
$$+ iq\sqrt{s}\,\Delta\left(g^{\ell}_{n,n-1}M^{\ell}_{n-1,n'} + g^{\ell}_{n,n+1}M^{\ell}_{n+1,n'}\right)$$
$$= -\delta_{nn'}I_n + \int_0^{\infty} k^3 dk\, I_n(q,k)M_{nn'}(k,q') \quad . \qquad (X.21)$$

The coefficients can be evaluated from the trigonometric integrals

$$f^{\ell}_{nn'} \equiv \int d\Omega\, Y_{n\ell m}(\Omega)\cos^2\psi\, Y^{*}_{n\ell m}(\Omega)$$
$$= A^{\ell}_n A^{\ell}_{n+1}\delta_{n',n+2} + \left[(A^{\ell}_n)^2 + (A^{\ell}_{n-1})^2\right]\delta_{n',n} + A^{\ell}_{n-1}A^{\ell}_{n-2}\delta_{n',n-2} \quad . \qquad (X.22)$$

$$g^{\ell}_{nn'} \equiv \int d\Omega\, Y_{n\ell m}(\Omega)\cos\psi\, Y^{*}_{n\ell m}(\Omega)$$
$$= A^{\ell}_n\delta_{n',n+1} + A^{\ell}_{n-1}\delta_{n',n-1} \qquad (X.23)$$

where

$$A^{\ell}_n = \left[\frac{(n-\ell+1)(n+\ell+2)}{4(n+1)(n+2)}\right]^{\frac{1}{2}} \quad . \qquad (X.24)$$

The coefficients A^{ℓ}_n are analytic functions in n and ℓ. We can therefore continue Eq. (X.21) to some complex values of n,ℓ; keeping the difference $K = n - \ell$ a fixed integer.

We recover the 3-dimensional angular momentum expansion by summing over all n,n' which are allowed in (X.21)

$$M_{\ell}(q_0,q;q_0',q';s) = \sum_{n=\ell}^{\infty}\sum_{n'=\ell}^{\infty}\zeta^{\ell}_n(\psi)M^{\ell}_{nn'}\,\zeta^{\ell}_{n'}(\psi) \quad . \qquad (X.25)$$

For $s = 0$ the set (X.21) reduces to the diagonal term

$$(q+m_1^2)(q^2+m_2^2)M_{nn}(q,q') = -I_n + \int_0^{\infty} k^3 dk\, I_n(q,k)M_{nn}(k,q') \qquad (X.26)$$

which is independent of ℓ. We assume now that the solution of (X.26) has a Regge pole at $s = 0$ in M_{nn} for $n = n_0$. Eq. (X.25) shows that

this M_{nn} is present in any M_ℓ for $\ell < n_0$. This implies that there is also a pole in each amplitude M_ℓ for

$$\ell = n_0 , n_0 - 1 , n_0 - 2 , \ldots \quad . \tag{X.27}$$

This is the daughter series. For $s = 0$ therefore the separation of the daughter trajectories is $\Delta\ell = 1$.

Note that the unequal mass has made the left-hand side of (X.4) complex for $s > 0$. To find the Regge trajectories we need not solve the equation but merely evaluate the Fredholm determinant. This is, in contradistinction to the numerator, a function of s only and is independent of q. It can be found from the real equation obtained by letting $q \to iq$. Since the integral equations are one dimensional we can select 24 Gaussian meshpoints and have still sufficient computer memory left to take 3 - 5 n-values. For small s-values, $|s| < m^2$ this gives about 4 accurate figures.

The result of such a calculation is presented in Fig. 11 where we have taken $m_1 = m_2 = 10\mu$ and fixed the coupling λ such that the parent trajectory gives an $s = 0$, $\ell = 0$ bound state. We attribute the fact that our trajectories are nearly parallel to the near singular potential we have used. The daughters do not have exactly the same slope at $s = 0$ so that the unit spacing rule only holds at $s = 0$. For $s \to -\infty$ the daughters seem to aim at the negative integer ℓ-values. The fact that the slopes at $s = 0$ are not all the same has led to the speculation[22] that some trajectories might intersect.

XI. The Elastic Region

In the region of elastic scattering we are mainly interested in finding the phase shifts δ_ℓ. Kemmer and Salam[26] have first shown that the Wick rotation can still be carried out in the elastic region in p-space at the price of picking up residual terms. I will show below the problems one encounters in treating the equation numerically in this region. These phase shifts, of course, have first been calculated by Schwartz and Zemach[27] in x-space by the use of variational methods. (See next section.) By matrix inversion in p-space we cannot match their accuracy but the mere fact that these phase shifts have been obtained by numerical methods only makes it desirable to have an alternate and completely independent derivation. We fully agree with their results. Levine, Tjon and Wright[16] have extended the p-space method into the lowest inelastic region. Unfortunately these authors do not reveal how they have taken care of the logarithmic singularity coming from the inelastic threshold nor do they tell

whether and how they have included the additional discontinuity across the inelastic cut.

The BS equation for the ℓth partial wave projection of the fully off shell amplitude M_ℓ can be written

$$M_\ell(q,q_o;q',q_o';p_o) = I_\ell(q,q_o;q',q_o') + i\int_{-\infty}^{\infty} dk_o \int_o^{\infty} dk\, I_\ell(q,q_o;k,k_o)G_2(k,k_o;p_o)M_\ell(k,k_o;q',q_o';p_o) \tag{XI.1}$$

with I_ℓ given by (IX.8). The Green's function G_2, which we take here as the product of two free propagators is

$$G_2(q,q_o;p_o) = [(q_o+p_o)^2 - \underline{q}^2 - m^2 + i\epsilon]^{-1}[(q_o-p_o)^2 - \underline{q}^2 - m^2 + i\epsilon]^{-1} . \tag{XI.2}$$

In the center of mass the total energy is

$$s^{\frac{1}{2}} = 2p_o \tag{XI.3}$$

since

$$p = (p_o, \underline{0}) . \tag{XI.4}$$

We find the on shell t-matrix or scattering amplitude by putting both the initial and final legs on the mass shell which, in the equal mass case, is defined by

$$(p \pm q)^2 = m^2 , \tag{XI.5}$$

or

$$q_o = 0 \tag{XI.6}$$

$$|\underline{q}| = [s/4 - m^2]^{\frac{1}{2}} \equiv \hat{q} . \tag{XI.7}$$

There is an identical set of conditions (XI.5-7) for q_1', q'. The exact relationship between our off shell M_ℓ and $t_\ell(s)$ and the phase shifts can be found by looking at the unitarity equation (V.7). It is

$$t_\ell(s) = \frac{4}{\pi^3}\sqrt{\frac{s-4m^2}{s}}\, e^{i\delta_\ell} \sin\delta_\ell = M_\ell(\hat{q},o;\hat{q},o;\frac{\sqrt{s}}{2}) . \tag{XI.8}$$

Numerical solution of Eq. (XI.1) is complicated by the fact that two out of the four poles in G_2 can pinch the contour of integration

(Fig. 6) and then are in quadrants I and III. This pinch introduces a double pole (!) just at the edge of the region of integration. It can be reduced to a single pole by using the factorization of Noyes[28] and Kowalsky.[29] This has also the advantage of bringing the entire equation into an N/D form.

Eq. (XI.1) can also be written for the half on shell amplitude

$$M_\ell(q,q_0;p_0) \equiv M_\ell(q,q_0;\hat{q},o;p_0) \quad . \tag{XI.9}$$

Let us put

$$M_\ell(q_0,q;p_0) \doteq t_\ell(s)\, f_\ell(q,q_0) \tag{XI.10}$$

and observe that (XI.8) implies

$$f(\hat{q},o) \equiv 1 \quad . \tag{XI.11}$$

Substitute (XI.10) into (XI.1) and observe that $t_\ell(s)$ can be taken out of the integral.

$$t_\ell(s) f_\ell(q,q_0) = I_\ell(q,q_0;\hat{q},o) + i t_\ell(s) \int_{-\infty}^{\infty} dk_o \int_0^{\infty} dk\, I_\ell(q,q_0;k,k_o)$$

$$\times\, G_2(k,k_o;p_o)\, f_\ell(k,k_o) \quad . \tag{XI.12}$$

First we evaluate this equation for $q_0 = 0$, $|\underline{q}| = \hat{q}$.

$$t_\ell(s) = \frac{I_\ell(\hat{q},o;\hat{q},o)}{D_\ell(s)} \equiv \frac{N_\ell(s)}{D_\ell(s)} \quad , \tag{XI.13}$$

where $N_\ell(s) = I_\ell(\hat{q},o,\hat{q},o)$ and

$$D_\ell(s) = 1 - i \int_{-\infty}^{\infty} dk_o \int_0^{\infty} dk\, I_\ell(\hat{q},o;k,k_o) G_2(k,k_o;p_o) f_\ell(k,k_o) \tag{XI.14}$$

which contains the unknown function f_ℓ and is shown in Fig. 12.



Fig. 12. Graphical representation of the integral for the denominator function.

We obtain an integral equation for f_ℓ if we substitute t_ℓ from (XI.13) and (XI.14) into (XI.12). Dropping all ℓ- and some p_0-dependence we get

$$f_\ell(q_0,q) = \frac{I_\ell(q,q_0;\hat{q},o)}{N(s)} + i\int_{-\infty}^{\infty} dk_0 \int_0^{\infty} dk\, \bar{V}(q,q_0;k,k_0)G_2(k,k_0,p)f_\ell(k,k_0). \tag{XI.15}$$

with

$$\bar{V}(q,q_0;k,k_0) = I(q,q_0;k,k_0) - \frac{I(o,\hat{q},k,k_0)I(q,q_0,o,\hat{q})}{N(s)} \tag{XI.16}$$

The potential $\bar{V}$ has a simple zero just at the position where G_2 develops its double pole. The additional pole of the kernel in (XI.15) is not removed by the Wick rotation. It can be removed by contour deformation or by an additional subtraction. The expressions get so lengthy, however, that I will not reproduce them here. Also it is safest to let the computer carry out these substitutions. To remove the contour away from the standard singularities (Fig. 6) we carry out the Wick rotation in (XI.15) which gives

$$f(q,iq_4) = \frac{I(q,iq_4;\hat{q},o)}{N(s)} - \int_{-\infty}^{\infty} dk_4 \int_0^{\infty} dk\, \bar{V}(q,iq_4,k,ik_4)G_2 f(k,ik_4)$$

$$+ \frac{\pi}{4p_0}\int_0^{\hat{q}} dk\, [\bar{V}(q,iq_4,k,\bar{k_0}) + \bar{V}(q,iq_4,k,-\bar{k_0})]\frac{g(k)\ \theta(\hat{q})}{k_0\,\omega}\ . \tag{XI.17}$$

The second integral comes from circling the Green's function poles at

$$k_0 = \pm\,\bar{k_0}\,, \bar{k_0} = p - \omega\,,\ \omega^2 = \underline{k}^2 + m^2 \tag{XI.18}$$

in quadrants I and III. The Heaviside function $\theta(\hat{q})$ is to indicate that this term is present only in the elastic region. We have also introduced the new unknown function

$$g(k) \equiv f(k,\bar{k_0}(k)) \tag{XI.19}$$

and we can get an integral equation for it by evaluating (XI.17) at $iq_4 = q_0$. This leads to two coupled sets for f and g. The presence of the additional function g(k) can be absorbed by memory space on the computer because it is a function of one variable only. One can use the symmetry of f in $q_0 \to -q_0$ to reduce the k_0-integration in (XI.17) from $0 \to \infty$.

We can also carry out the Wick rotation in (XI.14) but there is still the double pole from the elastic pinch. It can be removed by two successive subtractions. The solution of (XI.15) and its pendant give the functions f and g just at the points needed to do the integrals for D. The kernels are nonsingular as long as we stay a finite distance below the inelastic threshold. The result of such a calculation for s-waves is presented by the solid lines in Fig. 13. Note that none of the s-wave phase shifts increases through $\pi/2$ corresponding to the absence of a centrifugal barrier which could trap an s-wave resonance.

XII. Solution of the BS Equation in x-Space

Solving the BS equation in x-space has several advantages. The main thing is that in the elastic region in x-space the Wick rotation can be carried out and the elastic phase shifts are real. One finds a fourth order differential equation. After imposing proper boundary conditions to eliminate the unwanted solutions one has therefore only one real equation to solve instead of the complex valued set we saw in p-space. Most of the solutions in x-space have been obtained by the use of a variational principle either in integral or in differential form. The application of variational methods should not be restricted to x-space, however. These methods have been explored first by Vosko,[30] Schwartz[31] and by Schwartz and Zemach.[27] Their technique generally requires quite a bit more analytic work on the matrix elements than matrix inversion but it saves a considerable amount of computer time. The results are of unique accuracy.

Since the variational principle demands, at present at least, a positive (negative) definite kernel the method goes out from the Wick rotated equation. It assumes therefore its validity which has been proven only for the spinless case. Only in the equal mass situation the differential operator remains real after Wick rotation. On the other hand the accuracy of the method allows to solve the inhomogeneous integral equation below threshold where the kernel is regular and then make a numerical extrapolation[32] into the elastic region. It is hoped that this technique can eventually be extended into the lower inelastic region.

It will be difficult to improve on the accuracy obtained by the above mentioned authors. Therefore I will present their method although with some minor modifications concerning the variation of the nonlinear parameter added by Ladányi.[33] I will restrict myself to the very simplest case in the bound state region.

In x-space we had the BS equation written as

$$(\Box_1 + m_1^2)(\Box_2 + m_2^2)\,\psi(x_1, x_2) = V(x_1 - x_2)\,\psi(x_1, x_2) \quad . \qquad \text{(XII.1)}$$

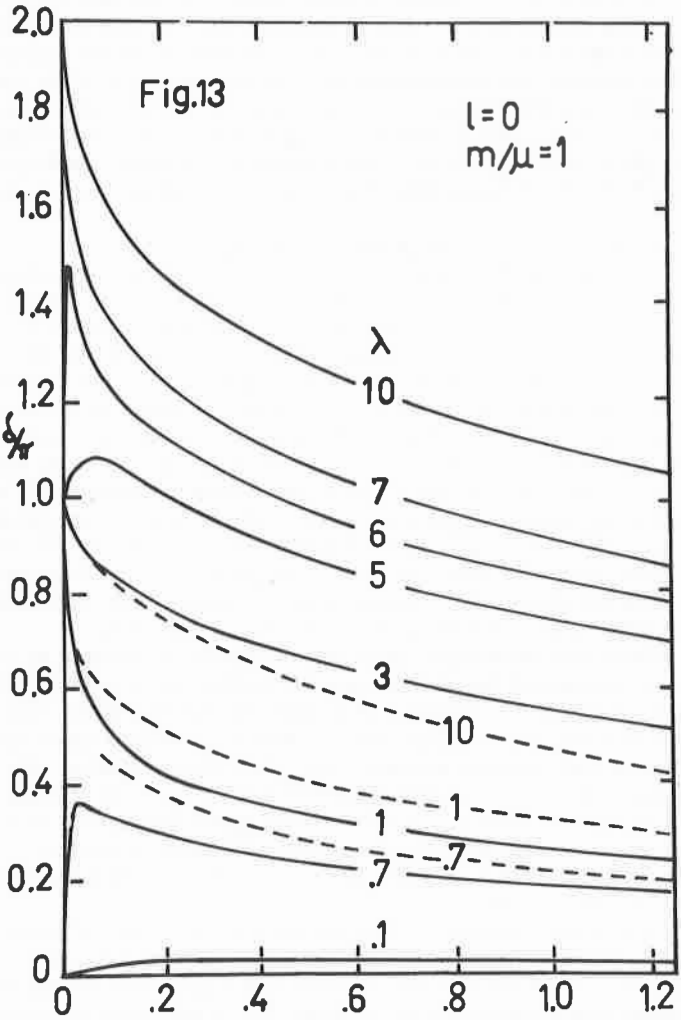


Fig. 13. The s-wave phase shifts in the elastic region (solid lines). The dotted lines have been obtained from the CPT approximation (Fig. 14).

One eliminates the center of mass coordinate by putting

$$x_1 = X + \frac{x}{2}$$

$$x_2 = X - \frac{x}{2}$$

which leads to

$$\frac{\partial}{\partial x^{\mu}_{1,2}} = \frac{\partial}{\partial X^{\mu}} \pm \tfrac{1}{2}\frac{\partial}{\partial x^{\mu}} \quad .$$

in the equal mass case only. Eq. (XII.1) can then be brought into the form

$$L\psi = \lambda V \psi$$

where

$$L = (-\Box + m^2 - p_o^2)^2 - 4p_o^2 \frac{\partial^2}{\partial x_4^2} \quad .$$

Our p_o is related to Wick's η by $p_o = \eta m$. The expressions

$$\Box = \sum_{i=1}^{4} \frac{\partial^2}{\partial x_i^2}$$

and

$$r = \left(\sum_{i=1}^{4} x_i^2 \right)^{\frac{1}{2}}$$

show that we use Euclidean metric with a positive volume element d^4x. The potential $V(r)$ is the four dimensional Fourier transform of the exchange interaction.

$$V_{\mu}(r) = \frac{1}{\pi^2} \int d^4q \, \frac{e^{-iqx}}{q^2 + \mu^2} = \frac{4\mu}{r} K_1(\mu r) \quad .$$

K_1 is the modified Bessel function with the properties

$$K_1(x) \xrightarrow[x \to 0]{} \frac{1}{x} + \tfrac{1}{2} x \ell n \, x + \ldots$$

$$\xrightarrow[x \to 0]{} \left(\frac{\pi}{2x}\right)^{\frac{1}{2}} e^{-x} + \ldots \quad .$$

For intermediate values one can use more terms in the asymptotic series.25) We use an interesting approximation given by Vosko[30]

$$xK_1(x) \cong \frac{1}{1+a}\left[(1+x)e^{-x} + a(1+2x)e^{-2x}\right]$$

with a = .66746. This parametrization which may be considered as an effective potential is extremely useful because it allows to evaluate all the integrals needed explicitly. The error is less than 5%.

In the bound state region $p_o < m$, Eq. (XII.1) is an eigenvalue problem for λ. Since we have used the Euclidean metric both L and V are positive definite and, in the equal mass case, selfadjoint operators. It follows that the Rayleigh quotient

$$[\lambda] \equiv \frac{\int d^4x\, \psi^*(x)\, L\, \psi(x)}{\int d^4x\, \psi^*(x)\, V\, \psi(x)}$$

leads to stationary approximations to the true value of λ. The error committed is of second order in the error in the trial functions to be substituted for ψ.

The invariance group of Eq. (XII.1) is O(3) only plus a special reflection group in $(x_4 \to -x_4)$. That is Eq. (XII.1) is invariant under rotations in 3-dimensional space only. In the strong binding limit $p_o = 0$ it is invariant under O(4). Nevertheless it is useful to expand ψ in terms of 4-dimensional spherical harmonics. We select the polar coordinates

$$x_4 = r\cos\psi$$

$$x_3 = r\sin\psi\cos\theta$$

$$x_2 = r\sin\psi\sin\theta\cos\varphi$$

$$x_1 = r\sin\psi\sin\theta\sin\varphi$$

and represent ψ as an infinite sum over terms of the form

$$f(r)\, Y_{n\ell m}(\psi,\theta,\varphi) \quad .$$

The $Y_{\ell mn}$ are the four dimensional spherical harmonics and are orthonormal as

$$\int_0^\infty d\psi \sin^2\psi \int_0^\pi d\theta \sin\theta \int_0^{2\pi} d\varphi\, Y^*_{n\ell m}(\psi,\theta,\varphi) Y_{n',\ell',m'}(\psi,\theta,\varphi) = \delta_{nn'}\delta_{\ell\ell'}\delta_{mm'} \, .$$

The matrix elements of L can be calculated by using the relations

$$\Box f(r)\, Y_{n\ell m} = Y_{n\ell m}\left(\frac{\partial^2}{\partial r^2} + \frac{3}{r}\frac{\partial}{\partial r} - \frac{n(n+2)}{r^2}\right) f(r)$$

$$\frac{\partial}{\partial x_4} = \cos\psi \frac{\partial}{\partial r} - \frac{1}{r}\sin\psi \frac{\partial}{\partial \psi}$$

$$\cos\psi\, Y_{n\ell m} = A^{\ell}_{n}\, Y_{n-1,\ell,m} + A^{\ell}_{n+1}\, Y_{n+1,\ell,m}$$

$$\sin\theta \frac{\partial}{\partial\theta}\, Y_{n\ell m} = -(n+2)\, A^{\ell}_{n}\, Y_{n-1,\ell,m} + nA^{\ell}_{n+1}\, Y_{n+1,\ell,m}$$

where

$$A^{\ell}_{n} = \left(\frac{(n-\ell)(n+\ell+1)}{4n(n+1)}\right)^{\frac{1}{2}} \qquad \text{for } n > \ell$$

$$A^{\ell}_{n} = 0 \qquad \text{for } n \le \ell \ .$$

The eigen-solutions are characterized by their reflection properties under $x_4 \to -x_4$ (even $\sigma = +1$, odd $\sigma = -1$) as well as ℓ, m.

The explicit form of the trial functions used is

$$\psi'_{\ell m\sigma} = \sum_{n\ge\ell}^{N(\sigma)} a_{\ell\sigma}(n)\, R^{n}\, e^{-\alpha n}\, Y_{n\ell m}$$

where the summation is over even n ($\sigma = +1$) or odd n ($\sigma = -1$) only. We have introduced two variational parameters in each term of the above series. The matrix elements of the potential can be calculated by use of the Vosko approximation or by numerical integration. This evaluation of the matrix elements is probably the most tedious part of the variational calculation and would take too long to be presented here. Some details of it are given by Vosko, Schwartz and Ladányi.

We may conclude by mentioning that the simple trial function

$$\psi'_{oo+} = a\, e^{-\alpha r}\, Y_{ooo}$$

can be evaluated by hand and leads to the surprisingly reasonable value $\lambda = 3.5$ for a $p_0 = 0$, $m = \mu$ bound state while the exact value is $\lambda = 3.419$.

XIII. Various Approximations to the Bethe-Salpeter Equation

The need for approximate solutions to the BS equation is still urgent in the wide areas where neither the variational technique nor the method of matrix inversion can be applied easily. Therefore a number of authors[34),35),36),13)] have proposed recently a variety of such approximations and tested their results in the case of the scattering of spinless isoscalar particles interacting through the exchange of a similar particle for which exact numerical results are available.[31)] The Padé approximant[37)] was proposed by Nuttall[35)] as an alternate to the CPT approximation.[34)] Vasavada[36)] proposed a method known from the one dimensional N/D equations and pointed out the need for correct threshold behavior of the approximate solutions. We show by the use of the Kowalsky-Noyes[28),29)] technique how to relate these different approximations and present a generator for the Padé series solution to the BS equation. All these approximate solutions satisfy elastic unitarity.

This time we write the inhomogeneous BS equation for the partial wave amplitude as

$$M_\ell(\vec{q},\vec{q}') = B_\ell(\vec{q},\vec{q}') + \frac{i}{4\pi^3}\,\hat{q}^{2\ell}\int_{-\infty}^{\infty} dk_o \int_0^{\infty} k^2 dk\, B_\ell(\vec{q},\vec{k})\, G(k) M_\ell(\vec{k},\vec{q}) \tag{XIII.1}$$

where the single particle exchange term is

$$B_\ell(\vec{q},\vec{k}) = -\frac{g^2}{2q^{2\ell}}\int_{-1}^{1} \frac{dz\, P_\ell(z)}{(q-k)^2 - m^2} \tag{XIII.2}$$

with the two particle Green's function of (XI.2). Here $\vec{q}$ denotes the pair $(q_o, |\underline{q}|)$ and q,q' are the initial and final four momenta. The hat denotes the mass shell values

$$\hat{q}_o = 0,\ \hat{q} \equiv |\hat{q}| = \tfrac{1}{2}[s-4m^2]^{\frac{1}{2}} \quad . \tag{XIII.3}$$

Following Vasavada[36)] we have introduced the factor $\hat{q}^{2\ell}$ which is a function of s only. This leaves M_ℓ and B_ℓ finite at threshold and implies that $M_\ell(\vec{q},\vec{q}')$ with all four legs on shell will have the correct threshold behavior in all the approximations we are considering. With this we mean that all the approximate solutions lead for $\hat{q} \to 0$ to $\delta_\ell \sim \hat{q}^{2\ell+1}$, the behavior found for the solutions of the Schrödinger equation with a finite range potential. $M_\ell(\hat{q},\hat{q}) \equiv M_\ell(s)$ is now related to the phase shifts by

$$M_\ell(s) = -\,4\pi\,\frac{\sqrt{s}}{\hat{q}^{2\ell+1}}\; e^{i\delta_\ell}\; \sin\delta_\ell \quad . \tag{XIII.4}$$

The mass shell value of $B_\ell(\vec{q},\vec{q}')$ is given by

$$B_\ell(s) = -\frac{q^2}{2\hat{q}^{2\ell+2}} Q_\ell\left(1+\frac{m^2}{2\hat{q}^2}\right) \qquad \text{(XIII.5)}$$

and the fact that for large z, the Legendre polynomials of the second kind, $Q_\ell(z) \sim z^{-\ell-1}$ insures $B_\ell(s \to 4m^2)$ to be finite.

The CPT approximation is obtained by removing $M_\ell(k,\hat{q})$ from the integral in (XIII.1) at the mass shell which is the point of maximum singularity of G(k) in the $(k_o, |\underline{k}|)$ plane. This falls into the region of integration only when $s > 4m^2$. We obtain the convenient ND^{-1} form

$$M_\ell(s) = B_\ell(s)\left\{1 - \frac{i}{4\pi^3}\int dk_o\, \underline{k}^2\, dk\hat{q}^{2\ell} B_\ell(\hat{q},\vec{k})\, G(k)\right\}^{-1}. \qquad \text{(XIII.6)}$$

In this approximation which is equivalent to the first order Fredholm denominator explicit exposition of the factor $\hat{q}^{2\ell}$ does not affect the calculation of the phase shifts.[34] This form shows (Fig. 14) how the triangle graph can appear in the denominator of the solution.



Fig. 14. An approximate D-function obtained by setting f = 1 in (XI.16) or Fig. 12. Note the presence of the triangle graph in the denominator.

The [1,1] Padé approximant[37] is found to be

$$M_\ell(s) = B_\ell(s)\left\{1 - \hat{q}^{2\ell}\frac{i}{4\pi^3}\int dk_o\, \underline{k}^2\, dk\, B_\ell^{\,2}(\hat{q},\vec{k})\, G(k)\, B_\ell^{-1}(s)\right\}^{-1}, \qquad \text{(XIII.7)}$$

again the factor $\hat{q}^{2\ell+2}$ provides the correct threshold behavior and the integral in (XIII.7) exists for all ℓ.

To demonstrate the relation between (XIII.6) and (XIII.7) we factor the on shell amplitude from the half off shell amplitude as proposed by Kowalsky and Noyes[28),29)]

$$M_\ell(\vec{q},\hat{q}) = M_\ell(s)\, f(\vec{q}) \qquad \text{(XIII.8)}$$

with $f(\hat{q}) = 1$. Substitution into (XIII.1) leads to

$$M_\ell(s) = B_\ell(s)\left\{1 - \hat{q}^{2\ell}\,\frac{i}{4\pi^3}\int dk_o\,\underline{k}^2\,dk\,B_\ell(\hat{q},\vec{k})G(k)f(\vec{k})\right\}^{-1} \qquad \text{(XIII.9)}$$

where $f(\vec{q})$ is given as the solution of

$$f(q) = \frac{B_\ell(\vec{q},\hat{q})}{B_\ell(s)} + \hat{q}^{2\ell}\,\frac{i}{4\pi^3}\int dk_o\,\underline{k}^2\,dk\,f(\vec{k})G(k)$$
$$\left\{B_\ell(\vec{q},\vec{k}) - B_\ell(\hat{q},\vec{k})\,\frac{B_\ell(\vec{q},\hat{q})}{B_\ell(s)}\right\} \quad . \qquad \text{(XIII.10)}$$

The CPT approximation is obtained from (XIII.9) by setting $f(\vec{k}) = 1$, its mass shell value, in (XIII.10). The [1,1] Padé approximant is obtained by retaining for $f(\vec{k})$ the entire inhomogeneous term in (XIII.10) by substituting in (XIII.9) $f(\vec{k}) = B_\ell(\vec{k},\hat{q})/B_\ell(s)$. We note that the [1,1] Padé approximant retains more of the off shell behavior than does the CPT approximation; this is the reason for the slightly more accurate numerical results obtained from (XIII.7).

The [1,N] Padé series for $M_\ell(s)$ is defined as

$$M_\ell(s) = B_\ell(s)\left[\sum_{n=0}^{N}\lambda^n b_n\right]^{-1} \qquad \text{(XIII.11)}$$

where $\lambda = g^2(4\pi)^{-2}$ is the coupling contained in B_ℓ. The b_n are to be determined so as to reproduce an iteration series of $M_\ell(s)$ up to the order N in λ. We see that the (N-1)th iterate of (XIII.10) inserted into (XIII.9) yields the [1,N] Padé series and that

$$M_\ell(s) = \lim_{N\to\infty}[1,N] \quad . \qquad \text{(XIII.12)}$$

This latter series, which represents an iteration in the denominator of M_ℓ only, maintains unitarity and is expected to be better convergent than an iteration of (XIII.1) because the norm of the kernel in (XIII.10) is generally smaller than the one in (XIII.1). At each finite stage it is different from the Fredholm series. The elastic threshold is contained in (XIII.9) and $f(\vec{k})$ does not have the elastic cut. It is therefore easy to see at each iteration of f in (XIII.10), which corresponds to a particular Padé approximant, what if any additional thresholds have been retained. The CPT approximation contains the triangle graph in the denominator (Fig. 14) while the [1,1] Padé approximant leads to a box diagram. The expressions in (XIII.6) and (XIII.7) differ from the triangle and box only in that the exchange propagators are

replaced by their partial wave projections. We note that Vasavada's approximation[36] is exactly the [1,1] Padé approximant.

While the above approximations lead to algebraic solutions containing some trace of the relative time, Blankenbecler and Sugar[13] (BBS) proposed a different method. They approximate the kernel of (XIII.1) by removing the relative time in such a way as to maintain elastic unitarity and then solve the resulting one dimensional integral equation numerically exact. We remark that the CPT, the [1,1] Padé and the BBS methods all satisfy elastic unitarity for all $s > 4m$. That the BBS equation does not contain any inelastic thresholds is realized directly by observing that the neglect of the relative time has made the exchange potential nonsingular ($m_{ex} \neq 0$. The dotted lines in Fig. 13 show the results obtained from the CPT approximation. For couplings $\lambda < .7$ they are reasonably accurate. For larger couplings one feels the influence of the fact that this approximation leads to a monotonic D-function and hence can give at most one bound state. The BBS equation leads to a spectrum which is quite close to the one of the BS equation[7] provided the effective coupling constant is renormalized by roughly a factor of $\frac{1}{2}$. The much smaller coupling constant is a consequence of the fact that the BBS equation develops singularities at $s = 0$ and on the left hand cut which are not there in the BS equation.[38]

XIV. Relativistic Three Particle Equations

J. G. Taylor[39),40)] has given a systematic discussion on how to obtain the relativistic three particle analog to the BS equation. The Taylor equations are linear off shell equations. They are derived for the three body irreducible part of the off shell amplitude. For a detailed discussion I recommend his papers. In a heuristic way they can be obtained, under the restriction to two body interactions, by writing down all possible linear diagrams.

This reads

$$T = V + V\, G_3\, T \qquad \text{(XIV.1)}$$

where T is the three body T-matrix $\langle q_1' q_2' q_3' | T | q_1 q_2 q_3 \rangle$. V is the potential which lets interact two out of the three particles at a time. One particle runs through unaffected but may be involved in a successive collision. G_0 is the three particle propagator and should be the product of three full propagators including all possible self energy insertions. Here we take it simply as the product of three free particle propagators

$$G_3 = G_0 = \prod_{i=1}^{3} [k_i^2 - m_i^2]^{-1} \quad . \qquad \text{(XIV.2)}$$

From now on I will restrict myself to the three pi-meson problem.

The above equations have at least the same drawbacks as their nonrelativistic analogues. They contain δ-functions which run through and prevent any iteration from converging. They contain the potentials about which we know very little. To the general solution we can superimpose any homogeneous solution of a two body subproblem. Following Faddeev, who first carried out a systematic analysis for the nonrelativisitic equations we eliminate the potentials. V has the form

$$V = V_{12} + V_{23} + V_{31} \tag{XIV.3}$$

since we have restricted ourselves to two body interactions. With this Eq. (XIV.1) implies that T itself is the sum of three terms

$$T = T_1 + T_2 + T_3 \tag{XIV.4}$$

where T_1 satisfies the equation

$$T_1 = V_{23} + V_{23} G_0 T \tag{XIV.5}$$

and a similar set for T_2 and T_3. We assume that the two body off shell scattering amplitude satisfies a BS equation as discussed earlier. In the three particle space then this entire equation is simply multiplied by a propagator for the third particle (Fig. 3). $t_{23} \sim M(q_1', q_2', q_1, q_2)[q_3^2 - m^2]^{-1}$.

$$t_{23} = V_{23} + t_{23} G_0 V_{23} \quad . \tag{XIV.6}$$

Let us rearrange this equation as

$$V_{23} = t_{23} = t_{23} G_0 V_{23} \quad . \tag{XIV.7}$$

This we substitute into Taylor's equations, for T, say, and rearrange

$$T_1 = t_{23} - t_{23} G_0 V_{23} + (t_{23} - t_{23} G_0 V_{23}) G_0 T \tag{XIV.8}$$

$$T_1 = t_{23} + t_{23} G_0 T - t_{23} G_0 (V_{23} + V_{23} G_0 T) \quad . \tag{XIV.9}$$

The last bracket is T_1 as seen from Eq. (XIV.5) and we get

$$T_1 = t_{23} + t_{23} G_0 (t - T_1) \quad . \tag{XIV.10}$$

But $T - T_1 = T_2 + T_3$ and we find

$$T_1 = t_{23} + t_{23} G_0 (T_2 + T_3) \ . \qquad \text{(XIV.11)}$$

This is the first one (Fig. 15) of a set of relativistic Faddeev type equations. These equations give the three body T-matrix as a function

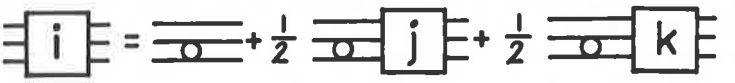


Fig. 15. The relativistic Faddeev equations. i,j,k = cyclic.

of the off shell two body amplitudes. The complete set reads

$$\begin{aligned} T_1 &= t_{23} + t_{23} G_0 (T_2 + T_3) \\ T_2 &= t_{31} + t_{31} G_0 (T_3 + T_1) \\ T_3 &= t_{12} + t_{12} G_0 (T_1 + T_2) \ . \end{aligned} \qquad \text{(XIV.12)}$$

It does not contain any diagonal δ-functions which could run through on iteration. Taylor has given a more general discussion which includes the presence of an irreducible 3-body potential V_3.

These equations are fully relativistic and manifestly covariant. As they stand they are much too complicated to be solved even on a computer and even after partial wave analysis. The usual approach is to do first a 3-body angular momentum analysis. This destroys covariance because it has to be done in the center of mass of all three particles. Then one has to do a two body partial wave analysis for each two body amplitude. This is done in the center of mass frame for each two particle subsystem and since the equations are no longer covariant the necessary transformations lead to complications. At the end one still has too many variables of integration and has to select a single partial wave, into which a separable Breit-Wigner form is substituted. This destroys much of the value of the 3-body angular momentum analysis.

Alessandrini and Omnes[41] and Freedman, Lovelace and Namyslovsky[42] have therefore proposed to substitute the BS kernel by the Blankenbecler-Sugar kernel in these equations. This removes two relative times and reduces therefore the number of independent variables of integration. Still the separable t-matrix has to be used. Basdevant and Kreps[43] have carried out some interesting calculations with these equations. But the mass shell restrictions imposed are too severe[38] so that these equations fail precisely if one wants to look for

the ω as a ρ-π bound state or resonance and in similar interesting problems.

We would like to outline here an alternative approach which, if carried out, would maintain more of the off shell properties of the relativistic set. It maintains covariance as long as possible and shows the close relationship of the relativistic three body problem with the BS equation in the inelastic region. We have not yet carried through these calculations to the very end and I will use drastic approximations in this first go around. This seems worthwhile, however, because the 3-body amplitude has a quite complicated analytic structure even in the low energy domain we are interested in.

We replace the t-matrix in Eq. (XIV.12) by a covariant separable expression (Fig. 16).



Fig. 16. The covariant separable approximation (ρ-pole) to the ππ-amplitude.

$$t(p_1, p_2, q_1, q_2) = \Gamma(p_1, p_2) D^{-1}((p_1+p_2)^2) \Gamma(q_1, q_2). \quad \text{(XIV.13)}$$

This approximation is a good one if there is a dominant resonance in one partial wave as for instance the ρ in the ππ-system. While this expression is an approximation it can be given quite a realistic analytic structure. We require for the t-matrix

1. Elastic right hand cut.
2. The ρ-pole is on the second sheet.
3. Elastic unitarity.

This is shown in Fig. 17a and b. For a realistic calculation one will have to include some remnance of the left hand cut or the crossed channel ρ-pole.

We use the Källén-Lehmann representation to write

$$t_{\pi\pi}(s) = \frac{-g^2}{D(p^2)} = \frac{-g^2}{\pi} \int_{4\mu^2}^{\infty} \frac{\sigma(k^2)dk^2}{k^2 - p^2} \quad . \quad \text{(XIV.14)}$$

Here we have restricted ourselves to an s-wave resonance while the ρ is actually p-wave. Since the ρ is unstable and there is no stable ππ-bound state we have written an unsubtracted dispersion relation with no pole terms. Comparing the discontinuities we find

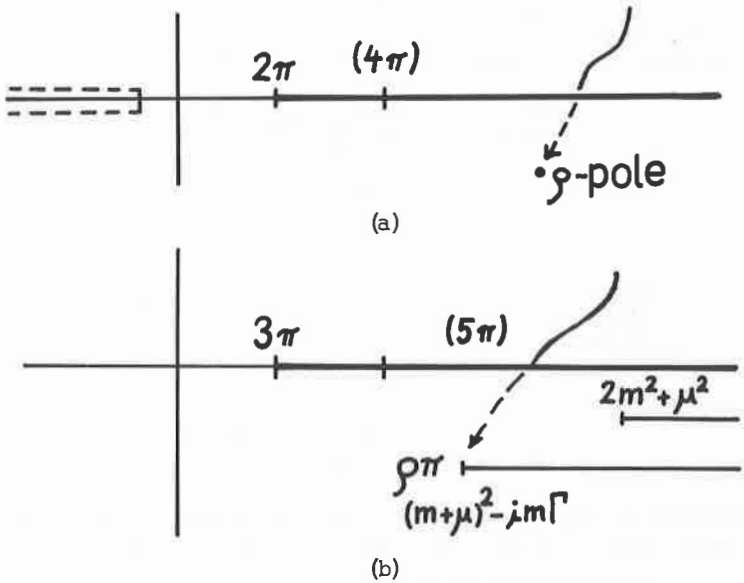


Fig. 17. a) Analytic structure of the approximate ππ-amplitude in the complex s_{12} plane.
b) Analytic structure of our three body D-function in the complex s plane.

$$t_{\pi\pi}(p^2) = \frac{1}{\pi} \int_{4\mu^2}^{\infty} \frac{\text{Im } H(k^2)}{|H(k^2)|^2} \frac{dk^2}{k^2 - p^2} \quad . \tag{XIV.15}$$

One can find an expression for $D(p^2)$ by summing and renormalizing all the self energy bubbles for 2 pi-mesons (Eq. (V.9)). It is interesting to see that $D(p^2)$ can be approximated quite well by

$$\text{Re } H(p^2) = p^2 - m_\zeta^2$$

$$\text{Im } H(p^2) = g^2 \sqrt{p^2 - 4\mu^2} \quad . \tag{XIV.16}$$

This has the correct analytic structure, satisfies elastic unitarity and has only two free parameters g and m_ρ which are fixed by fitting position and width of the ρ-meson.

Substitution of our separable t-matrix into the 3-body equations shows that each incoming ππ-pair has to go through a ρ. Therefore

one can factor off a vertex and the relevant part of the 3π amplitude is given by a $\rho\pi$-amplitude (Fig. 18a).

We see that the $\rho\pi$-amplitude satisfies a BS type equation although for stable unstable particle scattering (Fig. 18b). We have

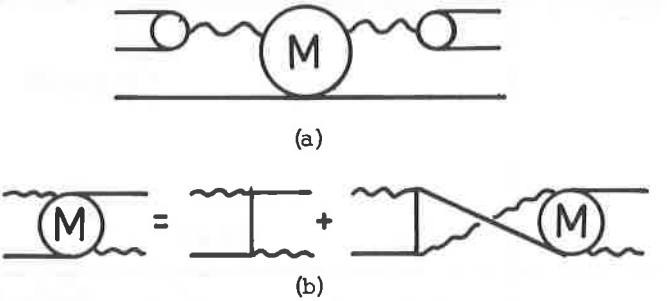


Fig. 18. a) The three π amplitude in the $\rho\pi$ approximation. Each pair of π has to go through a ρ.
b) The $\rho\pi$ BS equation.

crossed the propagators to indicate that the exchanged pi-meson is a u-channel pole. The equation reads, again for an s-wave ρ,

$$M(p,q,r) = I(q,r) + \frac{ig^2}{(2\pi)^4}\int \frac{d^4k\, D_\rho((p+k)^2)M(p,k,r)}{[(q+k)^2-\mu^2][(p-k)^2-\mu^2]} \qquad \text{(XIV.17)}$$

Analogous to our earlier discussion one can show that this equation satisfies 3-particle unitarity since we have kept the self energy terms in the ρ-propagator. Partial wave projection and substitution of the Lehmann representation gives

$$M_\ell(p_o,q_o,q) = I_\ell(q_o,q;s) + \frac{ig^2}{(2\pi)^4}\int_{4\mu^2}^{\infty} dm^2\,\sigma(m^2)\int \frac{d^4k\, M_\ell(p_o,k_o,k)P_\ell(z)}{[(q+k)^2-\mu^2][(p-k)^2-\mu^2][(p+k)^2-m^2]}\,. \qquad \text{(XIV.18)}$$

This equation is at present not tractable numerically. One can show that the ρ-pole is still in the domain of integration for the ω-energy. One can therefore get a meaningful one dimensional integral equation by evaluating the M under the integral at the ρ-pole.

For the present discussion I will make a more drastic approximation. I will evaluate M_ℓ at the point at which the singularities of the parallel pi- and ρ-propagators cross. This is only in the region

of integration for energies higher than $\sqrt{s} > m_\rho + \mu$, and I cannot expect to get the ω in this way. Now I can take the M_ℓ out of the integral and get an algebraic equation which leads to the N/D form

$$M_\ell = \frac{N_\ell}{D_\ell(s)} \quad . \tag{XIV.19}$$

It is interesting that even this approximate equation satisfies 3-particle unitarity. To find the resonance position we study $D_\ell(s)$ only. It contains the triangle graph. We find

$$D_o(s) = 1 - \lambda \int_{4\mu^2}^{\infty} \sigma(m^2)\, f(m^2, s)\, dm^2 \quad , \tag{XIV.20}$$

where f is the standard triangle function. Evaluation of this integral is simplified considerably by doing a partial integration:

$$\int_{4\mu^2}^{\infty} \sum (m^2) \frac{\partial f}{\partial m^2} (m^2, s)\, dm^2 \tag{XIV.21}$$

with

$$\sum (m^2) = \int_{4\mu^2}^{m^2} \sigma(p^2)\, dp^2 \tag{XIV.22}$$

and we have made use of the fact that the limits do not contribute. We recall that the triangle graph cannot be evaluated explictly, that is one will have one integration left, while its derivative has an explicit analytic form. It is straightforward to show that this D-function has the expected analytic properties (Fig. 17b), i.e. a $\rho\pi$-cut on the second sheet and a Landau (Peierls) singularity near the physical region. Note, however, that both these singularities are in the denominator function and hence will not directly cause a resonance. Whether and for what values of the parameters they are able to generate a resonance indirectly is an open question and will have to wait for a numerical study which we hope to carry out in the near future.

XV. The Relativistic Quark Model

As an application of these numerical results, which is quite hypothetical however, we might try to see what the BS equation has to say about a relativistic quark model.[7] We restrict ourselves here to the model which considers the mesons as bound states of a quark antiquark system.

At first we neglect the spin of the quarks. Then we have just our bound state Eq. (II.14) with the coupling λ and the quark mass m as free parameters. Since we want the known mesons to be stable against decay into quarks we take for a start m to be two nucleon masses. We want to get the π or ρ as a low lying s-wave state (the spin of the ρ is presumably made up by addition of the quark spins); we fix λ. With this we can solve for the position of the $\ell = 1,2,3$ bound states. We find (Fig. 19) that for this single particle exchange potential the masses lie much too high. In other words, we find a Regge trajectory which is much too flat instead of the infinitely rising straight line on which one finds the resonances.[44] We have presented the most favorable case in Fig. 19; all the variations of the different parameters make things worse.[14] What would be needed is a potential that is less attractive in the s-wave only and so would get the higher partial waves lower. The BS equation does sum up an infinite set of particle exchange diagrams but we have neglected the sum total of all sorts of crossed irreducible terms. Nobody knows the effect of these. One might try to replace them by a short range repulsive potential. This would give an additional free parameter and would allow to get a steeper trajectory. Such a repulsive hard core acts as a cut-off and we found that for the spin $\frac{1}{2}$ problem such a cut-off is probably required to make the equation well defined.

We have also looked at the situation[14] where both quarks have spin $\frac{1}{2}$. The question whether or not the homogeneous spin $\frac{1}{2}$ equation needs a cut-off or not has not been settled and is closely connected with the possibility of carrying out the Wick rotation. We found numerically that we could get a well defined solution only after introducing a high energy cut-off. The solutions depend strongly on this cut-off value. However, after fixing this cut-off at a reasonable value one can solve the set of coupled equations and one finds that both for pseudo scalar and for vector exchange the 0^- state is more strongly bound than the 1^- and one could easily adapt the parameters such as to fit the ρ–π mass difference. For the exchange of a pseudo scalar one finds a crossover in that only for tightly bound systems the 0^- mass is lower than the 1^- while for weakly bound systems the situation is reversed.

More detailed numerical calculations are being carried out. We might point out that the Blankenbecler-Sugar Eq. (III.4) is much simpler to solve and might give a quite accurate qualitative picture of the situation here.

XVI. Discussion

In these lectures I have not presented much new material but have concentrated on putting together a wide collection of statements

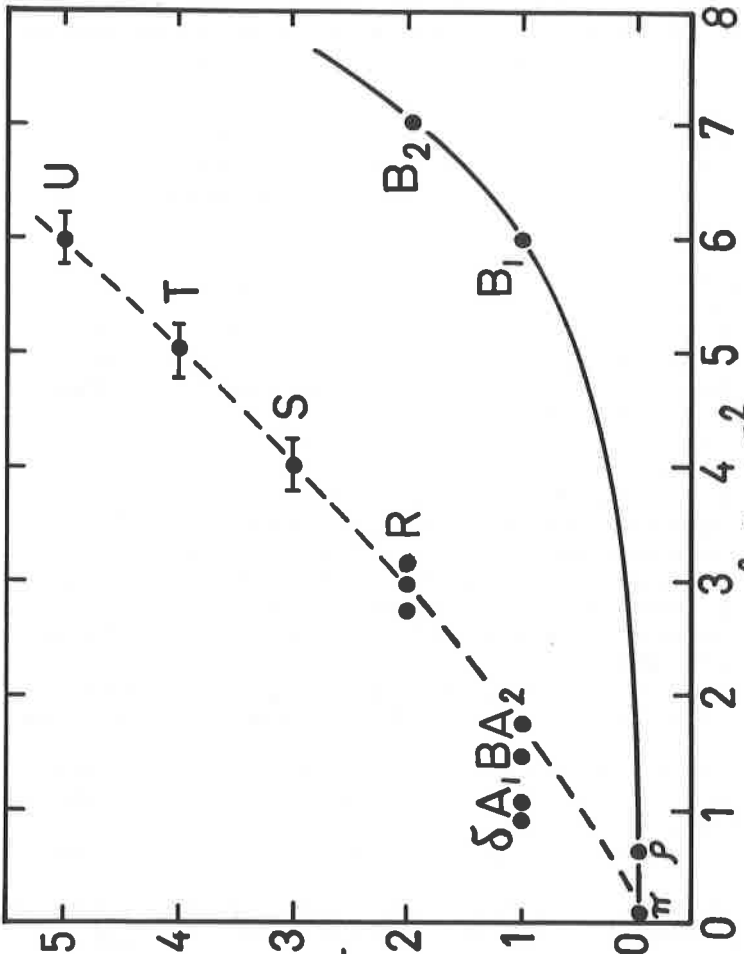


Fig. 19. Comparison of the realistic meson spectrum (Ref. 44) with the principal Regge trajectory in the BS equation.

on the BS equations which are well known to the experts. I have necessarily been short and I have omitted wide and important areas. To find the fine points in each case I refer to the list of original references. I have tried to point out open questions here and there and if my lectures have helped to get one of you interested in the problems around the BS type equations or have made it easier to penetrate the highly specialized literature, my goal has been achieved.

Acknowledgment

I want to take this opportunity to thank my collaborators who have taught me many of the things I have presented here; especially Drs. Harold Cohen, David R. Harrington, P. Narayanaswamy and John G. Taylor.

References

1. Y. Nambu, Prog. Theor. Phys. 5, 614 (1950).
2. E. E. Salpeter and H. A. Bethe, Phys. Rev. 84, 1232 (1951).
3. E. E. Salpeter, Phys. Rev. 87, 328 (1952); R. Karplus and A. Klein, Phys. Rev. 87, 848 (1952); G. Tiktopulos, J. Math. Phys. 6, 573 (1965).
4. M. Gourdin, Nuovo Cimento 7, 338 (1958).
5. G. C. Wick, Phys. Rev. 96, 1124 (1954).
6. R. E. Cutkosky, Phys. Rev. 96, 1135 (1954).
7. A. Pagnamenta, Nuovo Cimento 53, 30 (1968).
8. M. Gell-Mann and F. Low, Phys. Rev. 84, 350 (1957).
9. S. Mandelstam, Proc. Roy. Soc. (London) A233, 248 (1955).
10. J. G. Taylor, Nuovo Cimento Suppl. 1, 857 (1964).
11. H. Lehmann, K. Symanzik and W. Zimmermann, Nuovo Cimento 1, 209 (1955).
12. K. Symanzik, in Lectures on High Energy Physics, Herzegnovi. Ed. by B. Taksic (Gordon and Breach, Science Publishers, New York, 1966).
13. R. Blankenbecler and R. Sugar, Phys. Rev. 142, 1951 (1966).
14. P. Narayanaswamy and A. Pagnamenta, Nuovo Cimento 53A, 635 (1968), and Phys. Rev. 172, 1750 (1968).
15. A. Pagnamenta and J. G. Taylor, Phys. Rev. Letters 17, 258 (1966).
16. M. T. Levine, T. A. Tjon and T. Wright, Phys. Rev. Letters 16, 462 (1966); Phys. Rev. 154, 1433 (1967); Phys. Rev. 157, 1416 (1967).
17. M. M. Broido and J. G. Taylor, J. Math. Phys. 10, 184 (1969).
18. F. Gürsey, T. D. Lee and M. Nauenberg, Phys. Rev. 135, B647 (1964).
19. M. Toller, Nuovo Cimento 53, 671 (1968) and earlier reports.

20. D. Z. Freedman and T. M. Wang, Phys. Rev. Letters 17, 569 (1966) and Phys. Rev. 153, 1596 (1967).
21. V. Chung and D. R. Snider, Phys. Rev. 162, 1639 (1967).
22. R. E. Cutkosky and B. B. Deo, Phys. Rev. Letters 19, 1256 (1967).
23. R. Sawyer, Phys. Rev. 167, 1372 (1968).
24. H. T. W. Müller, Nucl. Phys. B5, 277 (1968).
25. M. Abramovitz and I. A. Stegun, eds., Handbook of Mathematical Functions, National Bureau of Standards, Wash., D.C. (1964).
26. W. Kemmer and Abdus Salam, Proc. Roy. Soc. (London) A230, 266 (1955).
27. C. Schwartz and C. Zemach, Phys. Rev. 141, 1454 (1966).
28. H. P. Noyes, Phys. Rev. Letters 15, 538 (1965).
29. K. L. Kowalsky, Phys. Rev. Letters 15, 789 (1965).
30. S. H. Vosko, J. Math. Phys. 1, 505 (1960).
31. C. Schwartz, Phys. Rev. 137, B717 (1965).
32. L. Schlessinger, University of California, Berkeley, preprint.
33. K. Ladányi, Nuovo Cimento 56A, 173 (1968).
34. H. Cohen, A. Pagnamenta and J. G. Taylor, Nuovo Cimento 50, 796 (1967). This work is referred to as CPT in the text.
35. J. Nuttall, Phys. Letters 23, 494 (1966).
36. K. Vasavada, Phys. Rev. 165, 1830 (1968).
37. G. Baker, Jr., Advances in Theoret. Phys. 1, 1 (1965).
38. T. L. Basdevant and R. L. Omnès, Phys. Rev. Letters 17, 775 (1966).
39. J. G. Taylor, Ref. 10, paper III pp. 945 ff.
40. J. G. Taylor, Phys. Rev. 150, 1321 (1966).
41. V. A. Alessandrini and R. L. Omnès, Phys. Rev. 139, B167 (1965).
42. D. Z. Freedman, C. Lovelace and T. M. Namyslovsky, Nuovo Cimento 43, 258 (1966).
43. T. L. Basdevant and R. E. Kreps, Phys. Rev. 141, 1398 (1966).
44. M. Focacci, W. Kienzle, B. Levrat, B. Maglic and M. Martin, Phys. Rev. Letters 17, 890 (1966).

ON WIGHTMAN'S THEORY OF QUANTIZED FIELDS†

Walter Wyss‡
Princeton University
Princeton, New Jersey

Contents

I. The Neutral Scalar Field
  A. Wightman Axioms
  B. A *-algebra
II. Borchers Algebra $\mathfrak{A}$
  A. Topological Properties of $\mathfrak{A}$ and $\mathfrak{A}'$
  B. Algebraic Properties of $\mathfrak{A}$ and its Idealstructure
  C. Positive Functionals and States on $\mathfrak{A}$
  D. A Family of Wightman States
  E. Some Theorems
III. Scattering Theory
  A. Haag-Ruelle Asymptotic Condition
  B. LSZ Asymptotic Condition
A. Appendix: Topological Vectorspaces

Introduction

Different mathematical concepts will give us different insights into a physical problem. These lectures intend to illustrate this fact in the case of Wightman's quantum field theory. It is well known that a set of Wightman functions gives us a Hilbertspace and an algebra of field operators.[1),2)] On the other hand a set of Wightman functions can be looked at as a state on a certain *-algebra.[3)]

Besides the discussion of algebraic aspects of the neutral scalar field we will present some of its Hilbertspace results, in particular the scattering theory.[4),5)]

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

‡Present address: Department of Physics and Astrophysics and Department of Mathematics, University of Colorado, Boulder, Colorado.

I. The Neutral Scalar Field

A. Wightman Axioms

We first discuss the free neutral scalar field because its properties will give us an idea of what the axioms for a realistic field should be. The Fock formalism for the free neutral scalar field consists in constructing a complex Hilbertspace and operators on it. Doing so a unitary continuous representation of the inhomogeneous Lorentz group will appear.

Denote by

$$d\mu(p) = \theta(p)\delta(p^2 - m^2)dp, \quad \theta(p) = \begin{cases} 1, & p^0 > 0 \\ 0, & p^0 < 0 \end{cases} \tag{I.1}$$

a Lorentz invariant measure on the forward mass hyperboloid

$$H_+(m) = \{p \in \mathbb{R}^4;\ p^2 = m^2,\ p^0 > 0\} \tag{I.2}$$

in momentum space. For simplicity we only treat the case of $m > 0$.

To construct the above mentioned Hilbertspace $\mathcal{H}$ we need the Hilbertspace $\mathcal{H}_1$, consisting of all functions ψ_1 on $\mathbb{R}^4$ satisfying

$$||\psi_1||^2 = \int |\psi_1(p)|^2 \, d\mu(p) < \infty \tag{I.3}$$

i.e. $\mathcal{H}_1 = \mathcal{L}_2(\mu)$, and $\mathcal{H}_n$, the Hilbertspace of all symmetric functions ψ_n on $\mathbb{R}^{4n}$ with

$$||\psi_n||^2 = \int |\psi_n(p_1, \ldots, p_n)|^2 \, d\mu(p_1) \ldots d\mu(p_n) < \infty \ . \tag{I.4}$$

Furthermore let $\mathcal{H}_0 = \mathbb{C}$, the one dimensional Hilbertspace of complex numbers.

$\mathcal{H}$ consists now of all sequences

$$\psi = \{\psi_0, \psi_1, \ldots\}, \qquad \psi_n \in \mathcal{H}_n \tag{I.5}$$

with the property that

$$||\psi||^2 = \sum_{n=0}^{\infty} ||\psi_n||^2 < \infty \ . \tag{I.6}$$

This is a norm and can be derived from the scalar product

$$(\phi, \psi) = \sum_{n=0}^{\infty} (\phi_n, \psi_n) \ , \tag{I.7}$$

where

$$(\phi_n, \psi_n) = \int \overline{\phi_n(p_1 \ldots p_n)}\ \psi_n(p_1 \ldots p_n) d\mu(p_1) \ldots d\mu(p_n) \qquad (I.8)$$

is the scalar product in $\mathcal{K}_n$.

$\mathcal{K}$ carries a continuous unitary representation of the restricted inhomogeneous Lorentz group $P_+^\uparrow$, namely

$$(U(a,\Lambda)\psi)_o = \psi_o \qquad (I.9)$$

$$(U(a,\Lambda)\psi)_n(p_1 \ldots p_n) = e^{ia \sum_{k=1}^{n} p_k}\ \psi_n(\Lambda^{-1} p_1, \ldots, \Lambda^{-1} p_n) \qquad (I.10)$$

Due to the Lorentz invariance of $d\mu(p)$ we have

$$||U(a,\Lambda)\psi|| = ||\psi|| \qquad (I.11)$$

The representation of the translation group

$$T(a) = U(a,1) \qquad (I.12)$$

has a spectral decomposition

$$T(a) = \int e^{ipa}\, dE(p) \quad , \qquad (I.13)$$

with

$$\left(\int \chi(p) dE(p) \psi\right)_n (p_1, \ldots p_n) = \chi(p_1 + p_2 + \ldots + p_n)\, \psi_n(p_1, \ldots, p_n) \qquad (I.14)$$

for any bounded measurable function χ.

The spectrum of the energy-momentum operator (infinitesimal generators of the translations)

$$P = \int p\, dE(p) \qquad (I.15)$$

consists of an isolated point $p = 0$ and of all the points

$$p = p_1 + p_2 + \ldots + p_n \quad , \ p_n \in H_+(m), \ n = 1, 2, \ldots \qquad (I.16)$$

Hence the spectrum is contained in $\bar{V}_+$, the closure of the forward lightcone.

For every function f on $\mathbb{R}^4$, satisfying

$$||f||^2 = \int \{ |f(p)|^2 + |f(-p)|^2 \}\, d\mu(p) < \infty \quad , \qquad (I.17)$$

we define a linear operator $A(f)$, the field operator, acting on $\mathcal{K}$ by

$$(A(f)\psi)_n(p_1,\ldots,p_n) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} f(p_k)\psi_{n-1}(p_1,\ldots\hat{p}_k,\ldots p_n)$$

$$+\sqrt{n+1}\int f(-p)\psi_{n+1}(p,p_1,\ldots,p_n)d\mu(p), \qquad (I.18)$$

where $\hat{p}_k$ means deletion of the variable p_k.

A(f) is an unbounded operator on $\mathcal{H}$ but it is clearly defined on the dense domain D given by

$$D = \bigoplus_{n=0}^{\infty} D_n \qquad (I.19)$$

where D_n consists of all symmetric Schwartz testfunctions[6] on $H_+(m)^{\otimes n}$. The direct sum means that each element of D is a terminating sequence.

We now like to list some of the properties of D. For $f \in \mathcal{S}(\mathbb{R}^4)$ we find that

$$A(f)D \subset D \qquad (I.20)$$

Furthermore D contains the vacuumstate Ω

$$\Omega = \{1,0,0,\ldots\} \qquad (I.21)$$

and is Lorentz invariant

$$U(a,\Lambda)D \subset D \ . \qquad (I.22)$$

Take $\phi,\psi \in D$ and $f \in \mathcal{S}(\mathbb{R}^4)$. Then

$$(\phi,A(f)\psi) = (A(f^*)\phi,\psi) \qquad (I.23)$$

with

$$(f^*)(p) = \overline{f(-p)} \qquad (I.24)$$

A straightforward calculation shows that on D with $f,g \in \mathcal{S}(\mathbb{R}^4)$

$$[A(f),A(g)] = \int \{f(-p)g(p) - f(p)g(-p)\}\, d\mu(p) \qquad (I.25)$$

$$= \int \tilde{f}(x)\,\Delta(x-y)\,\tilde{g}(y)\,dxdy \qquad (I.26)$$

where

$$\tilde{f}(x) = (2\pi)^{-2} \int e^{-ipx} f(p)\, dp \tag{I.27}$$

and

$$\Delta(z) = (2\pi)^{-4} \int e^{ipz} \epsilon(p)\delta(p^2 - m^2)dp \tag{I.28}$$

$$\epsilon(p) = \theta(p) - \theta(-p) \tag{I.29}$$

It is well known that

$$\Delta(z) = 0 \qquad \text{for } z^2 < 0 \tag{I.30}$$

and thus

$$[A(f), A(g)] = 0 \tag{I.31}$$

if $\tilde{f}$ and $\tilde{g}$ have spacelike separated support.

The representation of the Lorentz group gives us the following transformation property for the field operator

$$U(a,\Lambda)\, A(f)\, U(a,\Lambda)^{-1} = A((a,\Lambda)f) \tag{I.32}$$

with

$$((a,\Lambda)f)(p) = e^{ipa} f(\Lambda^{-1}p) \tag{I.33}$$

Finally, by applying the algebra of the operators $A(f)$, $f \in \mathcal{J}(\mathbb{R}^4)$, on Ω, we get a dense domain in $\mathcal{H}$. We say that Ω is a cyclic vector for the algebra of field operators.

We are now prepared to state the Wightman axioms for the neutral scalar field.

Axiom 1: The physical states are vectors in a complex Hilbertspace.

Axiom 2: To each $f \in \mathcal{J}(\mathbb{R}^4)$ there is an operator $A(f)$ on $\mathcal{H}$ with a dense domain D of definition which is independent of f.

For $\phi, \psi \in D$

$$(\phi, A(f)\psi) \tag{I.34}$$

is a continuous linear functional on $\mathcal{J}(\mathbb{R}^4)$ and

$$(\phi, A(f)\psi) = (A(f^*)\phi, \psi) \quad . \tag{I.35}$$

Finally we require that

$$A(f)D \subset D \tag{I.36}$$

Axiom 3: There exists a strongly continuous unitary representation $U(a,\Lambda)$ of the restricted inhomogeneous Lorentz group satisfying

$$U(a,\Lambda)\, A(f)\, U(a,\Lambda)^{-1} = A((a,\Lambda)f) \tag{I.37}$$

Furthermore

$$U(a,\Lambda)D \subset D \quad . \tag{I.38}$$

Axiom 4: $p = 0$ is an isolated eigenvalue of the energy-momentum operator P; the corresponding eigenspace is one dimensional and is spanned by the vacuum Ω, $\Omega \in D$. The rest of the spectrum of P is contained in the forward light cone.

Axiom 5: If f and g have spacelike separated support then

$$[A(f), A(g)]D = 0 \tag{I.39}$$

Axiom 6: The vacuum is cyclic with respect to the algebra of field operators.

Remarks:

a) We have to keep in mind that physical states actually correspond to rays in a Hilberspace.

b) We choose $\mathcal{J}$ as testfunctionspace because the Fourier transformation is a topological isomorphism of $\mathcal{J}$. For the use of other testfunction spaces see A. M. Jaffe.[7]

c) The content of Axiom 4 is called the spectrum condition and Axiom 5 is the locality assumption.

d) The axioms do not contain the notion of a particle. However a particle interpretation can be associated with our field theory; it will be discussed at the end of these lectures.

e) The axioms may be translated into properties of the Wightman functions

$$W_n(f_1 \otimes f_2 \otimes \ldots \otimes f_n) = (\Omega, A(f_1)\ldots A(f_n)\Omega) \quad . \tag{I.40}$$

The nuclear theorem tells us that W_n is uniquely extendable to a continuous linear functional on $\mathcal{J}(\mathbb{R}^{4n})$. The knowledge of these Wightman functions in return allows us uniquely to reconstruct the Hilbertspace $\mathcal{H}$ and the field operators $A(f)$. Thus we have two ways to analyse our axioms.

B. A *-algebra

The fact that the field operators form an algebra on D and that the Wightman functions

$$W_n(f_1 \otimes \ldots \otimes f_n) = (\Omega, A(f_1)\ldots A(f_n)\Omega) \tag{I.41}$$

can uniquely be extended to a continuous linear functional on $\mathcal{S}(\mathbb{R}^{4n})$ enables us to describe a Wightman theory by a functional on a topological *-algebra $\mathfrak{A}$.

Let $\mathcal{S}_0 = \mathbb{C}$ and $\mathcal{S}_n = \mathcal{S}(\mathbb{R}^{4n})$. As a vectorspace $\mathfrak{A}$ is the topological direct sum

$$\mathfrak{A} = \bigoplus_{n=0}^{\infty} \mathcal{S}_n \tag{I.42}$$

An element $f \in \mathfrak{A}$ thus is a terminating sequence

$$f = \{f_0, f_1, \ldots\} \tag{I.43}$$

with $f_n \in \mathcal{S}_n$.

We now equip $\mathfrak{A}$ with a noncommutative multiplication, defined by

$$(fg)_n(p_1, \ldots, p_n) = \sum_{k=0}^{n} f_k(p_1, \ldots, p_k) g_{n-k}(p_{k+1}, \ldots, p_n) \tag{I.44}$$

and an involution

$$(f^*)_n(p_1, \ldots, p_n) = \overline{f_n(-p_n, \ldots, -p_1)} \tag{I.45}$$

$$(f^*)_0 = \overline{f_0} \tag{I.46}$$

There is a unity for the multiplication, namely

$$1 = \{1, 0, \ldots\} \tag{I.47}$$

and hence $\mathfrak{A}$ is a *-algebra with unit element; we call it the Borchers algebra.

The Lorentz group acts on $\mathfrak{A}$ as automorphisms

$$((a, \Lambda)f)_n(p_1, \ldots, p_n) = e^{ia \sum_{k=1}^{n} p_k} f_n(\Lambda^{-1} p_1, \ldots, \Lambda^{-1} p_n) \tag{I.48}$$

$$((a, \Lambda)f)_0 = f_0 \tag{I.49}$$

Another automorphism is provided by the Fourier transform

$$(Ff)_n(x_1,\ldots,x_n) = (2\pi)^{-2n}\int e^{-i\sum_{k=1}^{n} p_k x_k} f_n(p_1,\ldots,p_n)dp_1\ldots dp_n \tag{I.48}$$

$$(Ff)_o = f_o \tag{I.49}$$

A Wightman theory now can be described by a continuous linear functional on $\mathfrak{A}$ satisfying certain restrictions. Because of the continuity we have to examine the topology of $\mathfrak{A}$.

Remark:

a) The construction of Borchers algebra could be based on any other testfunction space E. The Fourier transform then would map $\mathfrak{A}$ onto the Borchers algebra that is based on the Fourier image of E.

II. Borchers Algebra $\mathfrak{A}$

In this section we will study the topological and algebraic properties of Borchers algebra and show how to describe a Wightman theory by means of states on $\mathfrak{A}$.

A. Topological Properties of $\mathfrak{A}$ and $\mathfrak{A}'$

We now are in a position to determine the topological structure of

$$\mathfrak{A} = \bigoplus_{n=0}^{\infty} \mathcal{S}_n \quad , \tag{II.1}$$

where

$$\mathcal{S}_o = \mathbb{C} \quad , \quad \mathcal{S}_n = \mathcal{S}(\mathbb{R}^{4n}) \quad , \tag{II.2}$$

and of

$$\mathfrak{A}' = \prod_{n=0}^{\infty} \mathcal{S}_n' \quad . \tag{II.3}$$

First let us discuss $\mathcal{S}(\mathbb{R}^N)$ and $\mathcal{S}'(\mathbb{R}^N)$.

Definition II.1: In $\mathcal{S}(\mathbb{R}^N)$ the topology τ is given by the following countable set of seminorms

$$p_k(\varphi) = \sup_{\substack{\ell+|m|\le k \\ x\in\mathbb{R}^N}} |(1+||x||^2)^{\ell/2} D^m\varphi| \quad , \quad k = 0,1,\ldots \tag{II.4}$$

where $||x||$ is the Euclidean norm in $\mathbb{R}^N$ and D^m a differential monomial of order $|m|$.

Lemma II.1: $\mathcal{S}(\mathbb{R}^N)$ is a F-space.

Proof: See Ref. 6. ■

Lemma II.2: $\mathcal{S}(\mathbb{R}^N)$ is a nuclear space.

Proof: See Ref. 10. ■

Theorem II.1:

a) $\mathcal{S}(\mathbb{R}^N)$ is reflexive and separable.

b) $\mathcal{S}(\mathbb{R}^N)$ is barrelled.

c) On $\mathcal{S}(\mathbb{R}^N)$ the topology τ coincides with the Mackey and the strong topology.

d) $\mathcal{S}(\mathbb{R}^N)$ is bornological.

e) In $\mathcal{S}(\mathbb{R}^N)$ the closed bounded sets are compact.

f) $\mathcal{S}(\mathbb{R}^N)$ has no bounded neighborhoods.

g) The Fourier transform is a topological isomorphism on $\mathcal{S}(\mathbb{R}^N)$.

Proof:

a) See theorem A.15a).

b) See theorem A. 12c).

c) See theorem A. 12a).

d) See theorem A. 13d).

e) $\mathcal{S}(\mathbb{R}^N)$ is complete and its bounded sets are precompact, see theorem A. 15b).

f) Combine theorems A. 15c) and A. 11a).

g) See Ref. 6. ■

Theorem II.2: $\mathcal{S}'(\mathbb{R}^N)$ has the following properties:

a) In the strong topology it is nuclear and complete.

b) It is weak sequentially complete.

Proof:

a) See theorem A. 15d) and Ref. 6.

b) See theorem A. 12b). ■

Theorem II.3: $\mathfrak{A}$ now has the following properties:

a) It is a LF-space and thus complete.

b) It is nuclear, barrelled and bornological.

c) It is reflexive.

d) It is separable.

Proof:

a) See lemma A.8.

b) See theorems II. 15g), A. 12d) and A. 13c).

c) See theorem A. 10a).

d) $\mathfrak{A}$ is a countable topological sum of separable spaces. ■

Theorem II.4: $\mathfrak{A}'$ is

a) nuclear and complete.

b) weak sequentially complete.

Proof:

a) See theorem A. 15g).

b) See theorem A. 10b). ■

Remark:

a) All these topological properties are essential for an integral representation of certain functionals in $\mathfrak{A}'$.[11]

B. Algebraic Properties of $\mathfrak{A}$ and Its Idealstructure

Since $\mathfrak{A}$ has the structure of a topological sum and a polynomial like multiplication, we can expect some special algebraic properties.

Definition II.2: An element $f \in \mathfrak{A}$ satisfying $f^* = f$ is called Hermitian. We denote the real vectorspace of Hermitian elements by ϑ.

Lemma II.3:

$$\mathfrak{A} = \vartheta + i\vartheta \tag{II.5}$$

Proof:

$$f = g + ih \tag{II.6}$$

where

$$g = \tfrac{1}{2}(f + f^*) \tag{II.7}$$

$$h = \frac{1}{2i}(f - f^*) \tag{II.8}$$

Definition II.3: We call

$$\pi = \left\{ n \in \mathfrak{A};\ n = \sum \lambda_i f_i^* f_i,\ \lambda_i \geq 0,\ f_i \in \mathfrak{A} \right\} \quad . \tag{II.9}$$

where the sum only extends over finitely many terms, the positive cone and its element the positive elements of $\mathfrak{A}$.

Lemma II.4: $\pi \subset \vartheta$ and ϑ can be written as

$$\vartheta = \pi - \pi \tag{II.10}$$

Proof: Take $f = f^*$, $f \in \vartheta$. Then

$$f = \tfrac{1}{4}[(1+f^*)(1+f) - (1-f)^*(1-f)] \quad ■ \tag{II.11}$$

Lemma II.5: π is a proper cone with vertex at 0 and introduces a partial ordering on $\mathfrak{A}$, restricted to the field $\mathbb{R}$.

Proof: Take $f \in \pi$ and $-f \in \pi$. This is only possible if $f = 0$. The partial ordering is given by

$$f \leq g \Leftrightarrow g - f \in \pi \quad ■ \tag{II.12}$$

Lemma II.6: π has no interior points.
Proof: Every element of π has an even highest component, and then every neighborhood U of $f \in \mathfrak{A}$ contains elements with odd highest component, so $U \not\subseteq \pi$. Hence no point of π is interior. ∎
Definition II.4: The center $\mathfrak{Z}$ of an algebra consists of those elements of the algebra that commute with the whole algebra.
Definition II.5: An element $f \in \mathfrak{A}$ is called an idempotent if $ff = f$.
Theorem II.5: $\mathfrak{A}$ has the following algebraic properties:

a) $\mathfrak{A}$ has no divisors of zero.
b) The center of $\mathfrak{A}$ is $\mathbb{C}$.
c) No element outside the center has an inverse.
d) 0 and 1 are the only idempotents and are obviously in the center.

Proof: a) The condition $fg = 0$, $f \neq 0$ reads in components

$$(fg)_n(p_1, \ldots, p_n) = \sum_{k=0}^{n} f_k(p_1, \ldots, p_k) g_{n-k}(p_{k+1}, \ldots, p_k) \qquad (II.13)$$
$$n = 0, 1, \ldots$$

From this we easily deduce by induction that $g = 0$.

b) The center of $\mathfrak{A}$ consists of all elements f such that

$$fg = gf \quad , \quad \forall g \in \mathfrak{A} \qquad (II.14)$$

Writing down the components of this equation shows that $f = \lambda \cdot 1$, $\lambda \in C$.

c) Take $f \notin \mathbb{C}$, which means that at least for one $k \neq 0$, . $f_k(p_1, \ldots, p_k) \neq 0$. From

$$fg = 1 \qquad (II.15)$$

we see that g is not a terminating sequence and thus $g \notin \mathfrak{A}$.

d) The equation $ff = f$ has only two solutions, namely $f = 0$ and $f = 1$. ∎

These properties are reflected in the idealstructure of $\mathfrak{A}$.[12]
Definition II. 6: The Jacobson radical $\mathfrak{R}$ of $\mathfrak{A}$ is the intersection of all maximal two-sided ideals.

If $\mathfrak{R} = 0$, we call $\mathfrak{A}$ semisimple.

Theorem II.6: $\mathfrak{A}$ is semisimple.
Proof: It can be shown[13] that

$$\mathfrak{R} = \{ g \in \mathfrak{A}; \quad \exists (1 + fg)^{-1} \quad \forall f \in \mathfrak{A} \} . \qquad (II.16)$$

Due to theorem II.5c) $\mathfrak{R} = 0$. □

Definition II.7: A two-sided ideal P is called primitive if it is the biggest two-sided ideal contained in a maximal leftideal I.

$$P = \{f \in \mathfrak{U};\ f\,\mathfrak{U} \subset I\} \tag{II.17}$$

Denote the set of primitive ideals of $\mathfrak{U}$ by $P(\mathfrak{U})$.

Lemma II.7: Let A and B be any two-sided ideals and $P \in P(\mathfrak{U})$. Then

$$AB \subseteq P \tag{II.18}$$

implies

$$A \subseteq P \quad \text{or} \quad B \subseteq P$$

Proof: Let P be the biggest two-sided ideal contained in the maximal leftideal I. Take

$$B \not\subseteq P \quad \text{and} \quad A \not\subseteq P$$

and let

$$H = \mathfrak{U}/I \tag{II.19}$$

Denote the elements of H by $\epsilon(f)$. Since I is maximal, the left module H has no invariant subspaces and only

$$P\,\epsilon\,(f) = 0 \tag{II.20}$$

Thus $\exists\ \epsilon(f) \in H$ with $B\,\epsilon(f) \neq 0$ and since this is an invariant subspace $B\,\epsilon(f) = H$. There also is an $\epsilon(g) \in H$ with $A\,\epsilon(g) \neq 0$ and hence

$$AB\,\epsilon(f) \neq 0$$

or

$$AB \not\subseteq P.\ \square$$

Definition II.8: Let A be a subset of $P(\mathfrak{U})$. A closure operation has to satisfy:

a) $\bar{\phi} = \phi$ (II.21)

b) $\bar{A} \supseteq A$ (II.22)

c) $\bar{\bar{A}} = A$ (II.23)

d) $\overline{A \cup B} = \bar{A} \cup \bar{B}$ (II.24)

It is well known that a closure operation defines a topology.

Lemma II.8: Take $A \subset P(\mathfrak{A})$ and let

$$D_A = \cap \{P;\ P \in A\} \tag{II.25}$$

Then

$$\bar{A} = \{P \in P(\mathfrak{A});\quad P \supseteq D_A\} \tag{II.26}$$

is a closure operation.

Proof: a),b),c) are trivial. For d) we have to use that $C \supseteq D$ implies $\bar{C} \supseteq \bar{D}$. Thus

$$\overline{A \cup B} \supseteq \bar{A} \cup \bar{B} \tag{II.27}$$

Now let $P \in \overline{A \cup B}$. Then $P \supseteq D_{A \cup B} = D_A \cap D_B$. Since D_A and D_B are ideals we have

$$D_A\, D_B \subseteq D_A \cap D_B \subseteq P\ ,$$

leading to

$$D_A \subseteq P \quad \text{or} \quad D_B \subseteq P$$

Hence $P \in \bar{A}$ or $P \in \bar{B}$, and this means

$$P \in \bar{A} \cup \bar{B}$$

or

$$\overline{A \cup B} \subseteq \bar{A} \cup \bar{B} \quad \blacksquare \tag{II.28}$$

$P(\mathfrak{A})$ equipped with the above hull-kernel topology is called the structurespace of $\mathfrak{A}$.

Theorem II.7: The structurespace of $\mathfrak{A}$ is compact and connected.

Proof: The compactness follows from the fact that $\mathfrak{A}$ has an identity and the connectedness because $\mathfrak{A}$ has only two central idempotents.[12] ■

Theorem II.8: $\mathfrak{A}$ has no minimal ideals.

Proof: Suppose I is a minimal leftideal in $\mathfrak{A}$. For $f \in I$, $f \neq 0$, we have $If \subset I$; but minimality requires $If = I$. Hence $\exists\ g \in I$ such that $gf = f$. This only holds if $g = 1$. Since I is proper, $1 \notin I$. ■

Definition II.9: If S is a subset of $\mathfrak{A}$, then

$$R(S) = \{f \in \mathfrak{A};\ Sf = 0\} \tag{II.29}$$

is called the right-annihilator and

$$L(S) = \{f \in \mathfrak{A};\ fS = 0\} \tag{II.30}$$

the left-annihilator of S. An algebra with identity, where the left-annihilator of any subset is generated by an idempotent, is called a Baer algebra.

Theorem II.9: $\mathfrak{A}$ is a Baer algebra.

Proof: This follows from Theorem II.5a) and d). ■

Remark:

a) We do not yet know the physical significance of the special properties of the structurespace $P(\mathfrak{A})$.

C. Positive Functionals and States on $\mathfrak{A}$

A continuous linear functional T on $\mathfrak{A}$ is called positive if

$$T(f) \geq 0 \quad \forall\, f \in \pi \tag{II.31}$$

The continuous linear positive functionals on $\mathfrak{A}$ form a proper cone π' in $\mathfrak{A}'$.

Lemma II.9: Each $T \in \pi'$ satisfies

$$\text{a)}\quad T(f^*) = \overline{T(f)} \qquad \text{(reality)} \tag{II.32}$$

$$\text{b)}\quad |T(f^*g)|^2 \leq T(f^*f)\, T(g^*g) \qquad \text{(Schwarz inequality)} \tag{II.33}$$

Proof: See Neumark.[13] ■

Definition II.10: A continuous linear positive functional $T \in \pi'$ with $T(1) = 1$ is called a state on $\mathfrak{A}$.

Definition II.11: A nonempty convex subset β of π' is called a base for π' if each nonzero element $T \in \pi'$ has a unique representation of the form $T = \lambda W$ with $\lambda > 0$, $W \in \beta$.

Lemma II.10: The set of states on $\mathfrak{A}$

$$\sigma = \{T \in \pi';\ T(1) = 1\} \tag{II.34}$$

is a base for π'.

Proof: Take $T \in \pi'$, $T \neq 0$. Then we always can normalize T because $T(1) \neq 0$. This follows from the Schwarz inequality that says

$$|T(f)|^2 \leq T(1)\, T(f^*f) \qquad \forall\, f \in \mathfrak{A} \tag{II.35}$$

The convexity of σ is obvious. ■

To every $T \in \pi'$ we associated a leftideal I(T) in $\mathfrak{A}$ by

$$I(T) = \{f \in \mathfrak{A};\ T(f^* f) = 0\} \tag{II.36}$$

and a two-sided ideal E(T), namely the biggest two-sided ideal contained in I(T);

$$E(T) = \{f \in I(T);\ f\,\mathfrak{A} \subset I(T)\} \tag{II.37}$$

in short

$$E(T) = I(T) : \mathfrak{A} \tag{II.38}$$

We call I(T) the leftkernel of T and E(T) the dynamical ideal of T. Observe that I(T) generally has not a direct sum structure.

Lemma II.11: The map $T \to I(T)$ has the following properties:

a) $I(\lambda T) = I(T), \quad \lambda > 0$ (II.39)

b) $I(T_1 + T_2) = I(T_1) \cap I(T_2)$ (II.40)

c) I(T) is an order ideal. (II.41)

d) I(T) is closed. (II.42)

Proof:

a) and b) are trivial.

c) We have to show that if $f,g \in I(T)$ and $f \leq h \leq g$, then $h \in I(T)$ too. Thus

$$h = f + \alpha \quad , \quad \alpha \in \pi \tag{II.43}$$

and

$$h = g - \beta \quad , \quad \beta \in \pi \tag{II.44}$$

Hence

$$T(h) = T(\alpha) \geq 0 \tag{II.45}$$

and

$$T(h) = -\,T(\beta) \leq 0 \tag{II.46}$$

or

$$T(h) = 0 \tag{II.47}$$

In return this means

$$T(\alpha) = T(\beta) = 0 \tag{II.48}$$

and since $\alpha, \beta \in \pi$ we get

$$T(h^* h) = 0 \tag{II.49}$$

d) To prove that $I(T)$ is closed, we use the representation

$$I(T) = \{f;\ T(gf) = 0 \quad \forall\, g \in \mathfrak{A}\} \tag{II.50}$$

$$= \bigcap_{g \in \mathfrak{A}} \{f;\ T(gf) = 0\} \tag{II.51}$$

But for a fixed $g \in \mathfrak{A}$

$$\{f;\ T(gf) = 0\} \tag{II.52}$$

is closed because T is continuous. ∎

Every element $T \in \pi'$ gives rise to the exact sequence

$$0 \to I(T) \to \mathfrak{A} \xrightarrow{\varepsilon_T} H(T) \equiv \mathfrak{A}/I(T) \to 0 \tag{II.53}$$

$H(T)$ is a topological left module.

In addition, $T \in \pi'$ gives us a topological algebra

$$R(T) = \mathfrak{A}/E(T)\ , \tag{II.54}$$

which we shall call the dynamical algebra of T.

Definition II.12: We call $T \in \pi'$ irreducible if $T(T)$ is a maximal left-ideal. $E(T)$ then is primitive and thus an element of the structure space $P(\mathfrak{A})$.

Lemma II.12: If $T \in \pi'$ is irreducible, then

$$\mathfrak{A} = \mathfrak{A}f + I(T) \quad , f \notin I(T) \tag{II.55}$$

i.e. for $\varepsilon_T(f) \neq 0$, we have

$$H(T) = \mathfrak{A}\, \varepsilon_T(f) \quad . \tag{II.56}$$

Proof: For $f \notin I(T)$, $\mathfrak{A}f + I(T)$ is a leftideal that properly contains the maximal $I(T)$ and thus has to be equal to $\mathfrak{A}$. ∎

Remarks:

a) The construction of $H(T)$ is the famous Gelfand-Segal construction. $H(T)$ becomes an inner product space by

$$(\varepsilon_T(f),\ \varepsilon_T(g)) = T(f^* g) \tag{II.57}$$

and its completion relative to the norm

$$||\epsilon_T(f)|| = \sqrt{T(f^*f)} \tag{II.58}$$

becomes a Hilbertspace.

b) We say that T satisfies ideal-conditions, labeled by the index α, if there are certain leftideals I_α with

$$I_\alpha \subset I(T), \quad \forall\, \alpha\ . \tag{II.59}$$

c) Even though π has no interior point, it is still possible to give an integral representation for elements in π' by means of conical measures. This follows from the fact that π' is complete in the weak topology $\sigma(\mathfrak{A}', \mathfrak{A})$ of $\mathfrak{A}'$. We thus have a similar situation as for positive functionals on a B^*-algebra.[14]

D. A Family of Wightman States

A Wightman theory for a neutral scalar field can be described by a continuous linear positive functional W on $\mathfrak{A}$ that satisfies certain restrictions. First of all we have the Lorentz invariance

$$W((a,\Lambda)f) = W(f) \quad , \quad W \in \pi' \tag{II.60}$$

The spectrum condition and locality can be formulated as ideal-conditions by

$$I_1 = \{f \in \mathfrak{A};\ f_0 = 0,\ f_n(p_1, \ldots, p_n) = 0 \quad \text{if } q_k = \sum_{\ell=k}^{n} p_\ell \in \bar{V}_+ \quad \forall\, k,\ 1 \le k \le n, \quad \forall\, n\} \tag{II.61}$$

and I_2, being the two-sided ideal generated by elements of the form

$$(Ff)_n(x_1, \ldots, x_n) = (F\varphi)_n(x_1, \ldots \underline{x}, \ldots x_n) - (F\varphi)_n(x_1, \ldots \underline{x_\pi}, \ldots x_n), \tag{II.62}$$

where $\underline{x}$ is the subset $\{x_\ell, x_{\ell+1}, \ldots, x_{\ell+k}\}$ of $\{x_1, \ldots, x_n\}$ for any ℓ and k, and π is a permutation of the elements in $\underline{x}$; furthermore

$$(F\varphi)_n(x_1, \ldots \underline{x}, \ldots x_n) = 0 \text{ if } (x_i - x_j)^2 > 0 \quad \forall\, x_i, x_j \in \underline{x},\ i \neq j \tag{II.63}$$

Definition II.13: We denote by

$$\omega = \{W \in \pi';\ W((a,\Lambda)f) = W(f),\ I_1, I_2 \subset I(\omega)\}\ , \tag{II.64}$$

the set of all Wightman functionals, and call it the Wightman cone.

$$\omega_1 = \{W \in \omega;\ W(1) = 1\} \tag{II.65}$$

thus is the set of all Wightman states; clearly $\omega_1 \subset \sigma$.

To get used to the notion of leftkernels and dynamical ideals, we now look at a family of Wightman states. We first map $\mathfrak{A}$ into another graded vectorspace $\mathfrak{C}$ by

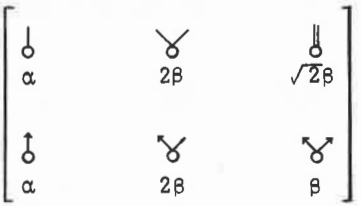
$$C_{\alpha\beta}: \mathfrak{A} \to \mathfrak{C} \tag{II.66}$$

where

$$f_n(p_1, \ldots, p_n) \to F_m(p_1, \ldots, p_m) \tag{II.67}$$

in a way we are going to specify graphically.

$f_n(p_1, \ldots, p_n)$ will be represented by n vertices. We get its image by inserting the elements of the following box



between these n vertices in a way that the solid lines end at vertices. The number of arrows corresponds to the number of variables in $F_m(p_1, \ldots p_n)$. The analytic form of the graphs is determined if we know the meaning of the solid lines and arrows in the simplest cases.



$$: f_2(p_1, p_2) \to \alpha^2 f_2(p_1, p_2) \tag{II.6 }$$



$$: f_2(p_1, p_2) \to \alpha^2 \int d\mu(p)\, f_2(-p, p) \tag{II.70}$$



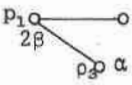
$$: f_2(p_1, p_2) \to 2\alpha\beta \int d\mu(p)\, f_2(p_1 - p, p) \tag{II.71}$$



$$: f_2(p_1, p_2) \to 2\beta^2 \int d\mu(p) d\mu(q) f_2(-p-q, p+q) \tag{II.72}$$



$$: f_1(p_1) \to \beta\, f_1(p_1 + p_2) \tag{II.73}$$



$$: f_3(p_1, p_2, p_3) \to 2\alpha^2\beta \int d\mu(p)d\mu(q)f_3(-p-q, p, q) \quad \text{(II.74)}$$

All these integrals make sense because we only integrate over the forward masshyperboloid $H_+(m)$

Denote by

$$\begin{bmatrix} k & \ell & m \\ r & s & t \end{bmatrix} \quad \text{(II.75)}$$

the graph, that shows the elements of box (II.68) k, ℓ, m, r, s, t times, respectively.

Definition II.14: The map $C_{\alpha\beta}$ (II.66) is defined by

$$(C_{\alpha\beta}f)_n(p_1, \ldots, p_n) = \sum_{\substack{r+s+2t=n \\ k,\ell,m,r,s,t}} \left(\begin{bmatrix} k & \ell & m \\ r & s & t \end{bmatrix} f_{k+\ell+m+r+s+t} \right)(p_1, \ldots, p_n) \quad \text{(II.76)}$$

Definition II.15: Let

$$\text{sym}: \quad \mathfrak{A} \to \mathfrak{A} \quad \text{(II.77)}$$

be the map

$$(\text{sym } f)_n(p_1, \ldots, p_n) = \frac{1}{n!} \sum_{\pi \in \gamma_n} f_n(p_{\pi(1)}, \ldots, p_{\pi(n)}) \quad \text{(II.78)}$$

Theorem II.10: The map $C_{\alpha\beta}$ is linear and has the property

$$(\text{sym } C_{\alpha\beta}fg)_n(p_1, \ldots, p_n) = \sum_{k=0}^{\infty} \int d\mu(q_1) \ldots d\mu(q_k) \sum_{\ell=0}^{n} \frac{(\ell+k)!(n-\ell+k)!}{n!k!}$$

$$\sum_{\{r_1, \ldots, r_n\}} (\text{sym } C_{\alpha\beta}f)_{k+\ell}(p_{r_1}, \ldots, p_{r_\ell}, -q_1, \ldots, -q_k) \cdot$$

$$(\text{sym } C_{\alpha\beta}g)_{k+n-\ell}(p_{r_{\ell+1}}, \ldots, p_{r_n}, -q_1, \ldots, -q_k) \quad \text{(II.79)}$$

where the sum $\sum_{\{r_1,\ldots,r_n\}}$ extends over all partitions $\{r_1,\ldots,r_\ell\}\{r_{\ell+1},\ldots,r_n\}$ of $\{r_1,\ldots,r_n\}$.

Proof: For simplicity reasons we give the proof only for the case $\beta = 0$.

$$
\begin{aligned}
&(\mathrm{sym}\, C_{\alpha o} fg)(p_1,\ldots,p_n) \\
&\quad = \mathrm{sym} \sum_{k=0}^{\infty} \begin{bmatrix} k & 0 & 0 \\ n & 0 & 0 \end{bmatrix} (fg)_{k+n} \quad (p_1,\ldots,p_n) \\
&\quad = \mathrm{sym} \sum_{\ell=0}^{\infty} \begin{bmatrix} 2\ell & 0 & 0 \\ n & 0 & 0 \end{bmatrix} \sum_{\substack{s+r+k=\ell \\ n_1+n_2=n}} f_{2s+n_1+k}\, g_{2r+n_2+k} \quad (p_1,\ldots,p_n) \\
&\quad = \sum_{k=0}^{\infty} \int d\mu(q_1)\ldots d\mu(q_k) \sum_{\ell=0}^{n} \sum_{\{r_1,\ldots r_n\}} \frac{1}{k!\,n!}\,(\ell+k)!\,(n-\ell+k)! \\
&\qquad (\mathrm{sym}\, C_{\alpha o} f)_{k+\ell}\,(p_{r_1},\ldots,p_{r_\ell},-q_1,\ldots,-q_k) \\
&\qquad (\mathrm{sym}\, C_{\alpha o} g)_{n+k-\ell}(p_{r_{\ell+1}},\ldots,p_{r_n},q_1,\ldots,q_k) \;\blacksquare
\end{aligned}
\tag{II.80}
$$

Lemma II.13: The functional

$$W_{\alpha\beta}(f) = (C_{\alpha\beta} f)_o \tag{II.81}$$

is a state on $\mathfrak{A}$.

Proof: The continuity and normalization $W_{\alpha\beta}(1) = 1$ are obvious. For the positivity we get from theorem II.10

$$
\begin{aligned}
W_{\alpha\beta}(f^* f) &= (C_{\alpha\beta} f^* f)_o \\
&= \sum_{k=0}^{\infty} \int d\mu(q_1)\ldots d\mu(q_k)\, k!\, (\mathrm{sym}\, C_{\alpha\beta} f^*)_k(-q_1,\ldots,-q_k) \\
&\qquad (\mathrm{sym}\, C_{\alpha\beta} f)_k(q_1,\ldots,q_k) \\
&= \sum_{k=0}^{\infty} \int d\mu(q_1)\ldots d\mu(q_k)\, k!\, |(\mathrm{sym}\, C_{\alpha\beta} f)_k(q_1,\ldots,q_k)|^2 \qquad (II.82)
\end{aligned}
$$

which is clearly positive definite. ∎

It is clear from where we got these states. They correspond to the Wightman fields $\alpha A(x) + \beta{:}A^2{:}(x)$, where $A(x)$ is a free field with mass m. Thus the states of Lemma II.13 are Wightman states.

We now are interested in their leftkernels and dynamical ideals.

Lemma II.14:

$$\text{a)}\quad f \in I(W_{\alpha\beta}) \Leftrightarrow (\text{sym } C_{\alpha\beta} f)_n(p_1, \ldots, p_n) = 0$$

$$\text{for } p_1, \ldots, p_n \in H_+(m),\ \forall\, n \qquad \text{(II.83)}$$

$$\text{b)}\quad f \in E(W_{\alpha\beta}) \Leftrightarrow (\text{sym } C_{\alpha\beta} f)_n(p_1, \ldots, p_n) = 0$$

$$\text{for } p_1, \ldots, p_n \in H_+(m) \cup (-H_+(m)),\ \forall n \qquad \text{(II.84)}$$

Proof: This is a direct consequence of (II.36), (II.37) and (II.82). ∎

Since $I(W_{\alpha\beta})$ is a Garding domain for the inhomogeneous Lorentz group, we have with $f \in I(W_{\alpha\beta})$

$$(a,1)f - f \in I(W_{\alpha\beta}) \qquad \text{(II.85)}$$

or

$$Pf \in I(W_{\alpha\beta}) \qquad \text{(II.86)}$$

where

$$(Pf)_n(p_1, \ldots, p_n) = (p_1 + \ldots + p_n)\, f_n(p_1, \ldots, p_n) \qquad \text{(II.87)}$$

We can define an operator P on the graded vectorspace $\mathfrak{G}$ by

$$(PF)_m(p_1, \ldots, p_m) = (p_1 + \ldots + p_m)\, F_m(p_1, \ldots, p_m) \quad , \qquad \text{(II.88)}$$

and easily observe that

$$C_{\alpha\beta} Pf = P\, C_{\alpha\beta} f \quad \blacksquare \qquad \text{(II.89)}$$

Definition II.16: For $W \in \pi'$, a vector $\epsilon_W(f) \in H(W)$ is called translation invariant provided

$$\epsilon_W((a,1)f) = \epsilon_W(f) \qquad \text{(II.90)}$$

Lemma II.15: The Wightman states $W_{\alpha\beta}$ have only one translation invariant vector.

Proof: A translation invariant vector $\epsilon_{W_{\alpha\beta}}(f)$ is equivalent to the equation

$$Pf \in I(W_{\alpha\beta}) \ . \tag{II.91}$$

According to lemma II.14 and (II.89) this means

$$(\text{sym } P\, C_{\alpha\beta}\, f)_n(p_1, \ldots, p_n) = 0 \quad \forall\, p_k \in H_+(m) \quad k = 1, \ldots, n \tag{II.92}$$

and f_0 arbitrary.

But $\sum_{l=1}^{n} P_l \neq 0$ if $p_l \in H_+(m)$ and thus $f \in I(W_{\alpha\beta})$ and f_0 arbitrary.

So $\epsilon_{W_{\alpha\beta}}(1)$ is the only normalized translation invariant vector. ∎

Remark:

For two different Wightman states $W_{\alpha\beta}$ and $W_{\alpha'\beta'}$ there is no direct way to compare their corresponding leftkernels or dynamical ideals. The elements of the Borchers class of the free field have a complicated relation amongst their leftideals and dynamical ideals.

E. Some Theorems

We are going to discuss some results of Borchers' paper[3] and a uniqueness theorem for Wightman states.

Lemma II.16: Let

$$M = \{f \in \mathfrak{A};\ D^m f_n(p_1, \ldots p_n) = 0 \text{ if } p_n,\ p_n+p_{n-1}, \ldots, p_n+\ldots+p_2 \in \overline{V_+}\ p_n+p_{n-1}+\ldots+p_1 = 0\} \tag{II.93}$$

This is a subspace of $\mathfrak{A}$.

Every translation invariant functional on $\mathfrak{A}$ that satisfies the spectrum condition, vanishes on M; i.e. $T \in \mathfrak{A}'$ with

$$\text{i)}\quad T((a,1)f) = T(f) \tag{II.94}$$

$$\text{ii)}\quad T(I_1) = 0 \tag{II.95}$$

satisfies

$$T(M) = 0 \tag{II.96}$$

Proof: From $T(I_1) = 0$ it follows that supp $T_n(p_1, \ldots, p_n)$ is contained in the region

$$p_n,\ p_n+p_{n-1}, \ldots, p_n+\ldots+p_1 \in \overline{V_+} \ .$$

Translation invariance on the other hand requires that the support of T_n is contained in the submanifold $p_n+\ldots+p_1 = 0$. ∎

We know already that a Wightman functional W gives us an exact sequence

$$0 \longrightarrow I(W) \longrightarrow \mathfrak{A} \xrightarrow{\epsilon_W} H(W) = \mathfrak{A}/I(\omega) \longrightarrow 0, \qquad \text{(II.97)}$$

where H(W) is an $\mathfrak{A}$-leftmodule and can be equipped with the inner product

$$(\epsilon_W(f), \epsilon_W(g)) = W(f^* g) \qquad \text{(II.98)}$$

and isometric maps

$$U(a,\Lambda) : H(W) \longrightarrow H(W) , \qquad \text{(II.99)}$$

defined by

$$U(a,\Lambda)\, \epsilon_W(f) = \epsilon_W((a,\Lambda)f) . \qquad \text{(II.100)}$$

Definition II.17: We call a linear map

$$B : H(W) \longrightarrow H(W) \qquad \text{(II.101)}$$

satisfying

$$Bf\, \epsilon_W(g) = f\, B\, \epsilon_W(g), \ \forall\, f,g \in \mathfrak{A} \qquad \text{(II.102)}$$

an $\mathfrak{A}$-endomorphism of the leftmodule H(W).

Definition II.18: Any vector $\epsilon_W(f) \in H(W)$, satisfying

$$\epsilon_W((a,\Lambda)f) = \epsilon_W(f), \quad \forall\, (a,\Lambda) \in P_+^{\uparrow} \qquad \text{(II.103)}$$

is called a vacuum.

Theorem II.11: Any translation invariant vector $\epsilon_W(f) \in H(W)$, i.e.

$$\epsilon_W((a,1)f) = \epsilon_W(f) , \quad \forall\, (a,1) , \qquad \text{(II.104)}$$

is already a vacuum.

Proof: See Ref. 3. The proof uses all the conditions that the Wightman functional W has to satisfy. ∎

Theorem II.12: Every $\mathfrak{A}$-endomorphism of H(W) commutes with the representation $U(a,\Lambda)$ of the Lorentz group.

Proof: Let B be an $\mathfrak{A}$-endomorphism and define

$$B(x) = U(x,1)\ B\ U(x,1)^{-1} \tag{II.105}$$

Then $B(x)$ is again an $\mathfrak{A}$-endomorphism. The functional

$$T(gf) = (\epsilon_W(1),\ \int B(x)\ \tilde{g}(x)\ dx\ \epsilon_W(f)) \tag{II.106}$$

can be extended to a continuous linear functional on $\mathfrak{A}$ that is translation invariant and satisfies the spectrum condition. Thus the support of $T_n(p_1,\ldots,p_n)$ is contained in the region

$$p_n, p_n+p_{n-1},\ldots,p_n+\ldots+p_2 \in \bar{V}_+,\ p_1+\ldots+p_n = 0 \tag{II.107}$$

Due to the fact that $B(x)$ is an $\mathfrak{A}$-endomorphism, the support is also contained in the region

$$p_1, p_1+p_n,\ldots,\ p_1+p_n+\ldots+p_3\ \in\ \bar{V}_+,\ p_1+\ldots+p_n = 0 \tag{II.108}$$

Hence $p_1 = 0$ and this means that $B(x) = B\ \forall x$, or B commutes with the representation of the translations. Now let $\epsilon_W(f)$ be a cyclic vacuum. Then $B\ \epsilon_W(f)$ is translation invariant and hence invariant under the representation of the Lorentz group, i.e. $B\ \epsilon_W(f)$ is also a vacuum.

It is now easy to derive that

$$U(0,\Lambda)\ B\ U(0,\Lambda)^{-1}\ g\ \epsilon_W(f) = B\ g\ \epsilon_W(f) \tag{II.109}$$

Since $\epsilon_W(f)$ is cyclic, we find that B commutes with the representation of the Lorentz group. ∎

Theorem II.13: The $\mathfrak{A}$-endomorphism ring of H(W) is Abelian.

Proof: The proof uses the PCT-theorem and the transitivity of local commutativity.[3] ∎

Theorem II.14: Let W_1, W_2 be two Wightman states, satisfying

$$I(W_1) = I(W_2)\quad . \tag{II.110}$$

If the leftmodule

$$H(W_1) = H(W_2) \equiv H \tag{II.111}$$

has only one vacuum, then

$$W_1 = W_2 \tag{II.112}$$

Proof: Let

$$\varepsilon : \mathfrak{A} \to H = \mathfrak{A}/I(W_1) \tag{II.113}$$

be the canonical quotient map.

On H we have now two inner products, namely

$$(\varepsilon(f), \varepsilon(g))_1 = W_1(f^* g) \tag{II.114}$$

and

$$(\varepsilon(f), \varepsilon(g))_2 = W_2(f^* g) \tag{II.115}$$

We define now a map

$$V: \quad H \longrightarrow H \tag{II.116}$$

by

$$(\varepsilon(f), \varepsilon(g))_1 = (\varepsilon(f), V\,\varepsilon(g))_2 \ . \tag{II.117}$$

It is easy to verify that V is an $\mathfrak{A}$-endomorphism and thus maps a vacuum into another vacuum. Since we have only one vacuum, it must be $\varepsilon(1)$ and we have

$$V\,\varepsilon(1) = \varepsilon(1) \ . \tag{II.118}$$

Thus

$$\begin{aligned} W_1(f) &= (\varepsilon(f^*), \varepsilon(1))_1 \\ &= (\varepsilon(f^*), V\,\varepsilon(1))_2 = (\varepsilon(f^*), \varepsilon(1))_2 \\ &= W_2(f) \qquad \blacksquare \end{aligned} \tag{II.119}$$

III. Scattering Theory

The Wightman axioms, as we mentioned earlier, do not deal with the notion of a particle. Nevertheless Haag[15] and Ruelle[16] discovered that under detailed assumptions on the spectrum of the energy-momentum operator, the Hilbertspace formalism of a Wightman theory gives rise to an asymptotic particle interpretation. This is the content of the Haag-Ruelle asymptotic condition. With its help

Hepp[17] could prove the LSZ asymptotic condition[18] on nonoverlapping states, i.e. a convergence theorem for field operators to asymptotically free fields. Consequently he gets reduction formulae for S-matrix elements between nonoverlapping states. On the other hand, by assuming the existence of retarded products of field operators, Steinmann[19] got reduction formulae also on overlapping states; he does not use the LSZ asymptotic condition.

From now on we assume that the support of the spectral measure for the translation operator $T(a) = \int e^{ipa}\, dE(p)$ is

$$\operatorname{supp} dE(p) = \{0\} \cup \{p; p^2 = m^2, p_0 > 0\} \cup \{p; p^2 \geq 4m^2, p^0 > 0\}.$$

The eigenspaces of the mass operator $M^2 = P^2$, belonging to the eigenvalues 0, m^2, will be denoted by $\mathcal{H}_0$, $\mathcal{H}_1$. $\mathcal{H}_0$ is one-dimensional and spanned by the vacuum Ω and $\mathcal{H}_1$ is the one-particle space, carrying the irreducible representation $[m,0]$ of the Lorentz group. Furthermore we assume that the field operator $A(p)$ applied on Ω creates a one-particle state.

A. Haag-Ruelle Asymptotic Condition

Without proofs, we first state two lemmas on which the convergence theorems are based.

Lemma III.1:[20]

Let $f(t,\vec{x})$ be a solution of the Klein-Gordon equation

$$(\Box + m^2)f = 0\,, \quad \Box = \left(\frac{\partial}{\partial t}\right)^2 - \sum_{k=1}^{3} \left(\frac{\partial}{\partial x^k}\right)^2 \tag{III.1}$$

satisfying

$$f(0,\vec{x}) \in \mathcal{S}(\mathbb{R}^3)\,, \quad \frac{\partial f}{\partial t}(0,\vec{x}) \in \mathcal{S}(\mathbb{R}^3)\,, \tag{III.2}$$

then

$$f(t,\vec{x}) \in \mathcal{S}(\mathbb{R}^3) \quad \forall\, t \tag{III.3}$$

and

$$|f(t,\vec{x})| < A\ (1 + |t|)^{-\frac{3}{2}} \tag{III.4}$$

$$\int |f(t,\vec{x})| < B\ (1 + |t|)^{\frac{3}{2}} \tag{III.5}$$

Definition III.1:

Let $W = \{W_0, W_1, \ldots\} \in \mathfrak{U}'$, i.e. that is a sequence of distributions with $W_0 \in \mathbb{C}$ and $W_n \in \mathcal{S}'(\mathbb{R}^{4n})$. For two such sequences T and W define a new sequence $T \otimes W$ by

$$(T \otimes W)_0 = 0 \tag{III.6}$$

$$(T \otimes W)_1(x_1) = 0 \tag{III.7}$$

$$(T \otimes W)_n(x_1, \ldots, x_n) = \sum_{k=1}^{n-1} T_k(\underline{x_k})\, W_{n-k}(\underline{x_{n-k}}), \tag{III.8}$$

where

$$\underline{x_n} = \{x_{i_1}, \ldots, x_{i_k}\} \subset \{x_1, \ldots, x_n\} \tag{III.9}$$

is an ordered subset of $\{x_1, \ldots, x_n\}$ and the sum extends over all these subsets.

For a sequence of distributions $W \in \mathfrak{U}'$ we define a truncated sequence W^T by

$$W^T = \ell n(1 + W) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} W^{\otimes n} \tag{III.10}$$

Lemma III.2: Let W be a Wightman state with a mass gap spectrum condition. With the variables

$$\xi_k = x_k - x_{k-1}, \quad q_k = \sum_{\ell=k}^{n} p_\ell , \tag{III.11}$$

the translation invariance of W gives the representation

$$(\tilde{W}^T)_n(x_1, \ldots, x_n) = \tilde{R}_n(\xi_2, \ldots, \xi_n) \tag{III.12}$$

$$(W^T)_n(p_1, \ldots, p_n) = (2\pi)^4 \delta(q_1) R_n(q_2, \ldots, q_n) \tag{III.13}$$

For $\varphi \in \mathcal{S}(\mathbb{R}^{n-1})$ we then have

$$\int dq_2^{\circ} \ldots dq_n^{\circ}\ \varphi(q_2^{\circ}, \ldots, q_n^{\circ})\, R_n(q_2, \ldots, q_n) \in \mathcal{S}(\mathbb{R}^{3(n-1)}) \tag{III.14}$$

We are now prepared to prove the strong Haag-Ruelle asymptotic condition.

Theorem III.1: Let $\mathcal{S}(G)$ be the space of testfunctions $f \in \mathcal{S}$ with support in

$$G = \{p;\ 0 \leq p^2 \leq 4m^2\ ,\ p^0 > 0\} \quad . \tag{III.15}$$

For $f_i \in \mathcal{S}(G)$ we define the one parameter family of operators

$$A_t(f_i) = \int dp\, A(p)\, f_i^*(p)\, e^{-i(p^0 - \omega(\vec{p}))t} \tag{III.16}$$

and the net

$$\phi_t = A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n)\, \Omega \quad . \tag{III.17}$$

$A_t^{(*)}(f_i)$ means either $A_t(f_i)$ or $A_t^*(f_i)$. Furthermore let

$$\hat{f}_i(\vec{p}) = f_i(\omega(\vec{p}), \vec{p}) \tag{III.18}$$

With these notations, there exist two free neutral scalar fields $A_+(\hat{f})$ and $A_-(\hat{f})$ such that in the strong Hilbertspace topology

$$\lim_{t \to -\infty} \phi_t = A_-^{(*)}(\hat{f}_1) \ldots A_-^{(*)}(\hat{f}_n)\, \Omega \tag{III.19}$$

$$\lim_{t \to \infty} \phi_t = A_+^{(*)}(\hat{f}_1) \ldots A_+^{(*)}(\hat{f}_n)\, \Omega \tag{III.20}$$

Proof: We first prove that ϕ_t is a Cauchy net and thus converges in $\mathcal{H}$. This will be achieved by developing $\left|\left|\frac{d}{dt}\phi_t\right|\right|^2$ into truncated expectation values. A typical term in this expansion is

$$T_n = \int dp_1 \ldots dp_n\, g_1(p_1) \ldots g_n(p_n) \prod_{k=1}^{n} e^{\mp i[p_k^0 - \omega(\vec{p}_k)]t} \times (\Omega, A^{(*)}(p_1) \ldots A^{(*)}(p_n)\Omega)^T \tag{III.21}$$

where $g_k(p_n)$ stands for either $f_k^{(*)}(p_k)$ or $\mp i(p_k^0 - \omega(\vec{p}_k))f_k^{(*)}(p_k)$. Because

$$A_t(f)\, \Omega = \frac{d}{dt}\, A_t^*(f)\, \Omega = 0 \tag{III.22}$$

each term in the above expansion will have a T_n with $n \geq 3$. Put now

$$g(p_1^0, \ldots, p_n^0; \vec{k}_1, \ldots, \vec{k}_n) = g_1(p_1^0, \vec{k}_1) \ldots g_n(p_n^0, \vec{k}_n) \tag{III.23}$$

and observe that $g \in \mathcal{S}(\mathbb{R}^{4n})$. Then, due to translation invariance

$$T_n = \int dp_1 \ldots dp_n \, g(p_1^\circ, \ldots, p_n^\circ; \vec{k}_1, \ldots, \vec{k}_n) e^{i[\pm\omega(\vec{k}_1)\pm\ldots\pm\omega(\vec{k}_n)]t}$$

$$\delta(\vec{p}_1 - \vec{k}_1)\ldots\delta(\vec{p}_n - \vec{k}_n) d\vec{k}_1 \ldots d\vec{k}_n (\Omega, A^{(*)}(p_1)\ldots A^{(*)}(p_n)\Omega)^T \quad \text{(III.24)}$$

$$= \int d\vec{p}_1 \ldots d\vec{p}_n \, dp_1^\circ \ldots dp_n^\circ \; g(p_1^\circ, \ldots, p_n^\circ; \vec{k}_1, \ldots, \vec{k}_n)$$

$$e^{i[\pm\omega(\vec{k}_1)\pm\ldots\pm\omega(\vec{k}_n)]t} \; (2\pi)^{-3n} e^{i(\vec{p}_1 - \vec{k}_1)\vec{x}_1 + \ldots + i(\vec{p}_n - \vec{k}_n)\vec{x}_n}$$

$$d\vec{k}_1 \ldots d\vec{k}_n \, d\vec{x}_1 \ldots d\vec{x}_n \, (\Omega, A^{(*)}(p_1^\circ, \vec{p}_1)\ldots A^{(*)}(p_n^\circ, \vec{p}_n)\Omega)^T \quad \text{(III.25)}$$

Changing to the new variables

$$q_k = p_k + p_{k+1} + \ldots + p_n \,, \quad p_k = q_k - q_{k+1} \,, \quad \text{(III.26)}$$

we get according to lemma III.2:

$$T_n = (2\pi)^{-3n} \int dq_1^\circ \ldots dq_n^\circ \, d\vec{q}_1 \ldots d\vec{q}_n \, \hat{g}(q_1^\circ, \ldots q_n^\circ; \vec{k}_1, \ldots, \vec{k}_n)$$

$$e^{i[\pm\omega(\vec{k}_1)\pm\ldots\pm\omega(\vec{k}_n)]t} \; e^{-i[\vec{x}_1\vec{k}_1 + \ldots + \vec{x}_n\vec{k}_n]}$$

$$e^{i\vec{x}_1(\vec{q}_1 - \vec{q}_2) + \ldots + i\vec{x}_n\vec{q}_n} \; (2\pi)^4 \delta(q_1) R_n(q_2, \ldots, q_n)$$

$$d\vec{k}_1 \ldots d\vec{k}_n \, d\vec{x}_1 \ldots d\vec{x}_n \quad \text{(III.27)}$$

We still have $\hat{g} \in \mathcal{S}(\mathbb{R}^{4n})$ and by defining

$$\chi(\vec{q}_2, \ldots, \vec{q}_n; \vec{k}_1, \ldots, \vec{k}_n) = \int dq_2^\circ \ldots dq_n^\circ \; \hat{g}(0, q_2^\circ, \ldots, q_n^\circ; \vec{k}_1, \ldots, \vec{k}_n)$$

$$R_n(q_2, \ldots, q_n) \quad \text{(III.28)}$$

we have $\chi \in \mathcal{S}(\mathbb{R}^{3(2n-1)})$ and

$$T_n = (2\pi)^{4-3n} \int d\vec{q}_1 \ldots d\vec{q}_n \, \delta(\vec{q}_1) \, e^{i[\pm\omega(\vec{k}_1)\pm\ldots\pm\omega(\vec{k}_n)]t}$$

$$e^{-i[\vec{x}_1\vec{k}_1 + \ldots + \vec{x}_n\vec{k}_n]} \; e^{i\vec{x}_1(\vec{q}_1 - \vec{q}_2) + \ldots + i\vec{x}_n\vec{q}_n}$$

$$\chi(\vec{q}_2, \ldots, \vec{q}_n; \vec{k}_1, \ldots, \vec{k}_n) d\vec{k}_1 \ldots d\vec{k}_n \, d\vec{x}_1 \ldots d\vec{x}_n \quad \text{(III.29)}$$

$$= (2\pi)^{4-3n} \int d\vec{q}_1 \ldots d\vec{q}_n \, \delta(\vec{q}_1) \, e^{i\vec{x}_1(\vec{q}_1 - \vec{q}_2) + \ldots + i\vec{x}_n\vec{q}_n}$$

$$\chi'(\vec{q}_2, \ldots, \vec{q}_n; t, \vec{x}_1; \ldots; t, \vec{x}_n) \, d\vec{x}_1 \ldots d\vec{x}_n \,, \qquad \text{(III.30)}$$

where $\chi'(\vec{q}_2, \ldots, \vec{q}_n; t, \vec{x}_1; \ldots; t, \vec{x}_n)$ is in every variable $(t, \vec{x}_n)$ a smooth solution of the Klein-Gordon equation. We again change variables to

$$\vec{p}_k = \vec{q}_k - \vec{q}_{k+1} \; ; \; \vec{q}_k = \vec{p}_k + \ldots + \vec{p}_n \qquad \text{(III.31)}$$

and obtain

$$T_n = (2\pi)^{4-3n} \int d\vec{p}_1 \ldots d\vec{p}_n \, \delta(\vec{p}_1 + \ldots + \vec{p}_n) \, e^{i[\vec{p}_1\vec{x}_1 + \ldots + \vec{p}_n\vec{x}_n]}$$

$$\hat{\chi}'(\vec{p}_2, \ldots, \vec{p}_n; t, \vec{x}_1; \ldots; t, \vec{x}_n) \, d\vec{x}_1 \ldots d\vec{x}_n \qquad \text{(III.32)}$$

$$= (2\pi)^{4-3n} \int d\vec{p}_2 \ldots d\vec{p}_n \, e^{i[\vec{x}_2\vec{p}_2 + \ldots + \vec{x}_n\vec{p}_n]} \, \hat{\chi}'(\vec{p}_2, \ldots, \vec{p}_n; t, \vec{x}_1; \ldots; t, \vec{x}_n)$$

$$e^{-i\vec{x}_1(\vec{p}_2 + \ldots + \vec{p}_n)} \, d\vec{x}_1 \ldots d\vec{x}_n \qquad \text{(III.33)}$$

$$|T_n| \leq A \int d\vec{p}_2 \ldots d\vec{p}_n \, e^{i[\vec{x}_2\vec{p}_2 + \ldots + \vec{x}_n\vec{p}_n]} \hat{\chi}'(\vec{p}_2, \ldots, \vec{p}_n; t, \vec{x}_1, \ldots, t, \vec{x}_n)$$

$$d\vec{x}_1 \ldots d\vec{x}_n \qquad \text{(III.34)}$$

Integration over $d\vec{x}_1$ and applying lemma III.1 gives

$$|T_n| \leq B(1+|t|)^{\frac{3}{2}} \int d\vec{p}_2 \ldots d\vec{p}_n \, e^{i[\vec{x}_2\vec{p}_2 + \ldots + \vec{x}_n\vec{p}_n]}$$

$$\varphi(\vec{p}_2, \ldots \vec{p}_n; t, \vec{x}_2; \ldots t, \vec{x}_n) \, d\vec{x}_2 \ldots d\vec{x}_n \qquad \text{(III.35)}$$

$$\leq C(1+|t|)^{\frac{3}{2}} (1+|t|)^{-\frac{3}{2}(n-1)} \int d\vec{p}_2 \ldots d\vec{p}_n \, e^{i[\vec{x}_2\vec{p}_2 + \ldots + \vec{x}_n\vec{p}_n]}$$

$$\psi(\vec{p}_2, \ldots, \vec{p}_n) \, d\vec{x}_2 \ldots d\vec{x}_n \qquad \text{(III.36)}$$

$$|T_n| \leq D(1+|t|)^{-\frac{3n}{2}+3} \qquad \text{(III.37)}$$

We know that in the above mentioned expansion the nonvanishing terms will consist of at least two 3-point functions of a n-point function with $n \geq 4$. Thus

$$\left\| \frac{d}{dt} \, T \right\| \leq E(1 + |t|)^{-\frac{3}{2}} \qquad \text{(III.38)}$$

and

$$\|\phi_{t_1} - \phi_{t_2}\| = \|\int_{t_2}^{t_1} \frac{d}{dt}\phi_t \cdot dt\| \leq \int_{t_2}^{t_1} \|\frac{d}{dt}\phi_t\| \, dt$$

$$\leq E\int_{t_2}^{t_1} (1 + |t|)^{-\frac{3}{2}} \, dt$$

$$\leq F\{(1 + |t_1|)^{-\frac{1}{2}} - (1 + |t_2|)^{-\frac{1}{2}}\} \, , \qquad \text{(III.39)}$$

which proves that ϕ_t is a Cauchy net.

Now we have to show that ϕ_t converges to a free field state. Let

$$\lim_{t \to \pm\infty} A_t^*(f_1)\ldots A_t^*(f_n)\Omega = \phi_\pm(\hat{f}_1, \ldots, \hat{f}_n) \qquad \text{(III.40)}$$

and

$$D_\pm = \langle \Omega, \phi_\pm(\hat{f}_1, \ldots, \hat{f}_n); \, n = 1,2,\ldots \rangle \qquad \text{(III.41)}$$

their corresponding spans. On $D_\pm$ we define operators by

$$A_\pm^{(*)}(\hat{f})\,\phi_\pm(\hat{f}_1, \ldots, \hat{f}_n) = \lim_{t \to \pm\infty} A_t^{(*)}(f)\, A_t^*(f_1)\ldots A_t^*(f_n)\Omega \qquad \text{(III.42)}$$

Computing

$$\lim_{t \to \pm\infty} \|A_t^*(f)\, A_t^*(f_1)\ldots A_t^*(f_n)\Omega\|^2 \qquad \text{(III.43)}$$

by an expansion into truncated expectation values, one finds that only terms of the form

$$(\Omega, A_t(f) A_t^*(g)\Omega) = (f,g) \, , \qquad \text{(III.44)}$$

with

$$(f,g) = \int d\mu(p) f^*(p) g(p) \qquad \text{(III.45)}$$

survive and that (III.43) has the same form as the corresponding free field expectation values. Thus (III.42) consistently defines free field operators. ∎

Definition III.2: Denote the closure of $D_\pm$ by $\mathcal{H}_\pm$. We say that a Wightman theory is asymptotically complete, if the corresponding Hilbertspace $\mathcal{H}$ satisfies

$$\mathcal{H} = \mathcal{H}_+ = \mathcal{H}_- \tag{III.46}$$

Definition III.3: The S-matrix is given by

$$S\,\phi_t(\hat{f}_1, \ldots, \hat{f}_n) = \phi_-(\hat{f}_1 \ldots, \hat{f}_n) \quad . \tag{III.47}$$

It is an isometric operator from $\mathcal{H}_+$ onto $\mathcal{H}_-$. If the theory is asymptotically complete, then S is unitary.

If we do not restrict the support of f in $A_t(f)$ then the Haag-Ruelle asymptotic condition only holds in the weak Hilbertspace topology.

Theorem III.2: Let $A_t(f)$ be as in theorem III.1, but $f \in \mathcal{J}(\mathbb{R}^4)$. Then

$$\lim_{t \to \pm\infty} A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n)\Omega = A_\pm^{(*)}(\hat{f}_1) \ldots A_\pm^{(*)}(\hat{f}_n)\,\Omega \tag{III.48}$$

holds in the weak topology of $\mathcal{H}_\pm$, i.e.

$$\begin{aligned} &\lim_{t \to \pm\infty} (\phi_\pm, A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n)\Omega) \\ &= (\phi_\pm, A_\pm^{(*)}(\hat{f}_1) \ldots A_\pm^{(*)}(\hat{f}_n)\,\Omega), \quad \phi_\pm \in \mathcal{H}_\pm \end{aligned} \tag{III.49}$$

Proof: Define

$$\phi_\pm(\hat{f}_1, \ldots, \hat{f}_n) = \lim_{t \to \pm\infty} A_t^*(\alpha f_1) \ldots A_t^*(\alpha f_n)\,\Omega \quad , \tag{III.50}$$

where

$$\alpha \in O_M(G), \ \alpha(p) = 1 \ \ \forall\, p \in H_+ \tag{III.51}$$

Now choose sequences $\{\phi_{\pm\nu}\}$ in $D_\pm$, which converge to $\phi_\pm$, and a Haag-Ruelle net $\{\phi_{t,\nu}\}$ that converges to $\phi_{\pm\nu}$. Then

$$\begin{aligned} &|(\phi_\pm, A_t^*(f_1) \ldots A_t^*(f_n)\Omega - \phi_\pm(\hat{f}_1, \ldots, \hat{f}_n))| \\ &\quad \le |(\phi_\pm - \phi_{\pm\nu}, A_t^*(f_1) \ldots A_t^*(f_n)\Omega)| \\ &\quad + |(\phi_{\pm\nu} - \phi_{t,\nu}, A_t^*(f_1) \ldots A_t^*(f_n)\Omega)| \\ &\quad + |(\phi_{t,\nu}, A_t^*(f_1) \ldots A_t^*(f_n)\Omega) - (\phi_{\pm\nu}, \phi_\pm(\hat{f}_1, \ldots, \hat{f}_n))| \\ &\quad + |(\phi_{\pm\nu} - \phi_\pm, \phi_\pm(\hat{f}_1, \ldots, \hat{f}_n))| \end{aligned} \tag{III.52}$$

By developing into truncated expectation values it can be shown[5] that $||A_t^*(f_1)\ldots A_t^*(f_n)\,\Omega||$ is bounded in t, and that

$$\lim_{t\to\pm\infty}(A_t^*(g_1)\ldots A_t^*(g_m)\,\Omega,\ A_t^*(f_1)\ldots A_t^*(f_n)\,\Omega)$$
$$=(\phi_\pm(\hat{g}_1,\ldots,\hat{g}_m),\phi_\pm(\hat{f}_1,\ldots,\hat{f}_n))\ ,\qquad g_k\in\mathcal{S}(G)\ . \quad \text{(III.53)}$$

But this implies

$$|(\phi_\pm,\ A_t^*(f_1)\ldots A_t^*(f_n)\,\Omega-\phi_\pm(\hat{f}_1,\ldots,\hat{f}_n))|<\epsilon\ . \quad \text{(III.54)}$$

(III.49) is now a straightforward consequence. ∎

B. LSZ Asymptotic Condition

Definition III.4: For $f\in\mathcal{S}(\mathbb{R}^4)$ let

$$\hat{f}(\vec{p})=f(\omega(\vec{p}),\ \vec{p}),\quad \omega(\vec{p})=\sqrt{\vec{p}^2+m^2} \quad \text{(III.55)}$$

$f,g\in\mathcal{S}(\mathbb{R}^4)$ are called nonoverlapping if

$$\operatorname{supp}\hat{f}\cap\operatorname{supp}\hat{g}=\phi \quad \text{(III.56)}$$

Theorem III.3:

Take $\{f_i\}\subset\mathcal{S}(G)$ of pairwise nonoverlapping functions and let

$$\phi_t=A_t^*(f_1)\ldots A_t^{(*)}(f_n)\,\Omega \quad \text{(III.57)}$$

as in (III.17). Then

$$||\frac{d}{dt}\phi_t||^2\in\mathcal{S}(\mathbb{R}^1)\ . \quad \text{(III.58)}$$

Proof:

We develop $||\frac{d}{dt}\phi_t||^2$ into truncated expectation values and observe that the only contributing terms are

$$(\Omega,A_t^{(\cdot)(*)}(f_{i_1})\ldots A_t^{(\cdot)(*)}(f_{i_k})\,\Omega)^T \quad \text{(III.59)}$$

with $k\geq 3$ and at most two timederivatives$^{(\cdot)}$. Due to translation invariance (III.59) becomes as in (III.21)

$$\int dp_1\ldots dp_k\ g_{i_1}(p_1)\ldots g_{i_k}(p_k)\ e^{i[\pm\omega(\vec{p}_1)\pm\ldots\pm\omega(\vec{p}_k)]t}$$
$$W_k^T(p_1,\ldots,p_k) \quad \text{(III.60)}$$

Look now at two variables p_ℓ, p_m with the property that their corresponding $\omega(\vec{p}_\ell)$, $\omega(\vec{p}_m)$ have the same signature in the exponent of (III.60). Using translation invariance again we can write

$$p_\ell = p_\ell(p_1, \ldots, \hat{p}_\ell, \ldots, p_k) = -\sum_{\substack{i=1 \\ i \neq \ell}}^{k} p_i \ , \tag{III.61}$$

where $\hat{p}_\ell$ means deletion of the variable p_ℓ.

Let

$$\Omega(\vec{p}_1, \ldots, \hat{\vec{p}}_\ell, \ldots, \vec{p}_k) = \pm\omega(\vec{p}_1) \pm \ldots \pm \omega\Big(\sum_{\substack{i=1 \\ i \neq \ell}}^{k} \vec{p}_i\Big) \pm \ldots \pm \omega(\vec{p}_k) \tag{III.62}$$

and

$$\chi(\vec{p}_1, \ldots, \hat{\vec{p}}_\ell, \ldots, \vec{p}_k) = \int dp_\ell \, g_{i_\ell}(p_\ell) \prod_{\substack{n=1 \\ n \neq \ell}}^{k} dp_n \, f_{i_n}^{(*)}(p_n)$$

$$W_k^{\,T}(p_1, \ldots, p_k) \tag{III.63}$$

(III.60) then reads

$$\int d\vec{p}_1 \ldots d\hat{\vec{p}}_\ell \ldots d\vec{p}_k \, \chi(\vec{p}_1, \ldots, \hat{\vec{p}}_\ell, \ldots, \vec{p}_k) \, e^{i\Omega(\vec{p}_1, \ldots, \hat{\vec{p}}_\ell, \ldots, \vec{p}_k)t} \tag{III.64}$$

According to lemma III.2

$$\chi \in \mathcal{J}(\mathbb{R}^{3(k-1)}) \tag{III.65}$$

Due to the nonoverlapping of $f_{i_\ell}^{(*)}$, $f_{i_m}^{(*)}$ we find that

$$\frac{\partial \Omega}{\partial \vec{p}_m} = \pm\{\omega^{-1}(\vec{p}_m)\vec{p}_m + \omega^{-1}(\vec{p}_1 + \ldots + \hat{\vec{p}}_\ell + \ldots \vec{p}_k) \cdot (\vec{p}_1 + \ldots + \hat{\vec{p}}_\ell + \ldots + \vec{p}_k) \neq 0 \tag{III.66}$$

on supp χ. Hence we can replace one of the variables $p_m^{\,i}$, $1 \leq i \leq 3$, by Ω and (III.64) then becomes a Fourier transform of a function in $\mathcal{J}$ with respect to Ω. This proves the above theorem. The proof can be extended to nonoverlapping $f_n(p_1, \ldots, p_n)$ that are not necessarily of the form $f_1 \otimes \ldots \otimes f_n$. ■

Definition III.5: Let $f_n \in \mathcal{J}(G^{\otimes n})$ be nonoverlapping and

$$\phi_{\pm}(\hat{f}_n) = \lim_{t \to \pm\infty} \int dp_1 \ldots dp_n \, f_n(p_1, \ldots, p_n) \, e^{i \sum_{k=1}^{n} [p_k^0 - \omega(p_k)]t} A^*(p_1) \ldots A^*(p_n) \Omega \,. \quad \text{(III.67)}$$

$$D_{\pm 0} = \langle \Omega, \, \phi_{\pm}(\hat{f}_n), \, n = 1,2,\ldots \rangle \quad \text{(III.68)}$$

then stands for the corresponding spans.

We are now in a position to prove the LSZ asymptotic condition on $D_{\pm 0}$.

Theorem III.4: Let $\phi_{\pm} \in D_{\pm 0}$ and $\{f_1\} \subset \mathcal{J}(G)$. Then

$$\lim_{t \to \pm\infty} A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n) \phi_{\pm} = A_{\pm}^{(*)}(\hat{f}_1) \ldots A_{\pm}^{(*)}(\hat{f}_n) \, \phi_{\pm} \quad \text{(III.69)}$$

holds in the strong topology.

If $\{f_1\} \subset \mathcal{J}(\mathbb{R}^4)$, then the above limit exists in the weak topology of $\mathcal{H}_{\pm}$, e.g.

$$\lim_{t \to \pm\infty} (\psi_{\pm}, A_t^*(f_1) \ldots A_t^*(f_n) \phi_{\pm}) = (\psi_{\pm}, A_{\pm}^*(\hat{f}_1) \ldots A_{\pm}^*(\hat{f}_n) \phi_{\pm}) \,, \quad \forall \, \psi_{\pm} \in \mathcal{H}_{\pm} \quad \text{(III.70)}$$

Proof:[17] Let $\phi_{\pm}(\hat{g}_m) \in D_{\pm 0}$. One first shows that the field operator is defined on $D_{\pm 0}$ and then approximates $\phi_{\pm}(\hat{g}_m)$ by

$$\phi_+(g_m) = \int dp_1 \ldots dp_m \, f_m(p_1, \ldots, p_m) \, e^{i \sum_{k=1}^{m} [p_k^0 - \omega(\vec{p}_k)]t} A^*(p_1) \ldots A^*(p_m) \Omega \quad \text{(III.71)}$$

using

$$\begin{aligned} &\| A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n) \phi_+(\hat{g}_m) - A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n) \phi_t(g_m) \| \\ &\leq \int_t^{\infty} ds \, \| A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n) \frac{d}{ds} \phi_s(g_m) \| \\ &\leq \int_t^{\infty} ds \, \| \frac{d}{ds} \phi_s(g_m) \|^{\frac{1}{2}} \| \{A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n)\}^* A_t^{(*)}(f_1) \ldots A_t^{(*)}(f_n) \\ &\qquad \frac{d}{ds} \phi_s(g_m) \|^{\frac{1}{2}} \end{aligned} \quad \text{(III.72)}$$

With the help of theorem III.3 and the strong and weak Haag-Ruelle asymptotic condition, theorem III.4 becomes obvious. ∎

Acknowledgments

I would like to thank Profs. V. Bargmann, E. Nelson and A. S. Wightman for many helpful discussions during my stay at Princeton University, and Profs. A. Ramsay and A. S. Wightman for critical comments on the composition of these lecture notes. Furthermore I am grateful for financial support from the Schweizerische Nationalfonds.

A. Appendix: Topological Vectorspaces

We collect some facts about Hausdorff locally convex topological vectorspaces, abbreviated by LCTVS. For the proofs see Refs. 8, 9, and 10.

Definition A.1: A subset A of a vectorspace E is called absolutely convex if with $x, y \in A$ and $\lambda, \mu \in \mathbb{C}$, satisfying $|\lambda| + |\mu| \leq 1$, we have $\lambda x + \mu y \in A$.

Definition A.2: A subset A of a vectorspace E is called absorbing if for every $x \in E$ $\exists \lambda > 0$ such that $x \in \mu A$ $\forall \mu$ with $|\mu| \geq \lambda$.

Definition A.3: A topological vectorspace $E(\tau)$ is a vectorspace in which the algebraic operations are continuous. In a LCTVS a neighborhood base of zero consists of absolutely convex sets.

Theorem A.1: A LCTVS has a neighborhood base $\mathfrak{U}(0)$ of zero with the following properties:

a) For $U, V \in \mathfrak{U}$ $\exists W \in \mathfrak{U}$ such that $W \subseteq U \cap V$.
b) With $U \in \mathfrak{U}$ and $\lambda \neq 0$ we have $\lambda U \in \mathfrak{U}$.
c) Every $U \in \mathfrak{U}$ is absolutely convex and absorbing.

Theorem A.2: A LCTVS has a neighborhood base of zero consisting of closed sets and satisfying the properties of theorem A.1.

Definition A.4: A seminorm p on a vectorspace E is a function

$$p: E \to \mathbb{R}$$

satisfying

a) $p(x) \geq 0$
b) $p(\lambda x) = |\lambda| \, p(x)$
c) $p(x+y) \leq p(x) + p(y)$

Theorem A.3: Let p be a seminorm on a vectorspace E and $\alpha > 0$. Then

$$\{x;\ p(x) < \alpha\} \quad \text{and} \quad \{x;\ p(x) \leq \alpha\}$$

are absolutely convex and absorbing. If A is an absolutely convex set and absorbing then

$$p_A(x) = \inf \{\lambda;\ \lambda > 0,\ x \in \lambda A\}$$

is a seminorm, called the support function of A. It has the property

$$\{x;\ p(x) < 1\} \subseteq A \subseteq \{x;\ p(x) \leq 1\} \ .$$

Theorem A.4: Let A be an absolutely convex and absorbing subset in a LCTVS and p_A its support function. p_A is continuous if A is a neighborhood. In this case the interior of A is $\{x;\ p(x) < 1\}$ and the closure of A is $\{x;\ p(x) \leq 1\}$.

Definition A.5: A barrel is a closed, absolutely convex and absorbing set.

Corollary A.1: According to theorem A.2, a LCTVS has a neighborhood base of zero consisting of barrels.

Theorem A.5: If $\mathcal{P}$ is a set of seminorms on a vectorspace E then there is a weakest topology on E in which all these seminorms are continuous. In this topology E is a LCTVS and a base of closed neighborhoods of zero is given by

$$\{x;\ \sup_{1 \leq i \leq n} p_i(x) \leq \varepsilon,\ \varepsilon > 0,\ p_i \in \mathcal{P}\} \ .$$

Definition A.6: Bounded sets:

i) Let $A, B \subset E(\tau)$. We say that A absorbs B if $\exists\ \alpha > 0$ such that $B \subset \lambda A$, $\forall \lambda$ with $|\lambda| \geq \alpha$.

ii) $A \subset E$ is called bounded if every neighborhood of zero absorbs A.

Definition A.7: Precompact sets:

i) Let U be a neighborhood of zero in a LCTVS. $A \subset E$ is called small of order U if $x - y \in U$, $\forall x, y \in A$.

ii) $A \subset E$ is called precompact if for every neighborhood U, A can be covered by finitely many A_k that are all small of order U.

Lemma A.1: A precompact set is bounded.

Definition A.8: Compact sets:

$A \subset E$ is compact if every open covering of A contains a finite subcovering.

Definition A.9: A pair of vectorspaces (E,F) is called a dual pair if there is a bilinear form $\langle x, y \rangle$ on $E \times F$ satisfying

i) $\langle x_0, y \rangle = 0 \quad \forall y \in F \Rightarrow x_0 = 0$.

ii) $\langle x, y_0 \rangle = 0 \quad \forall x \in E \Rightarrow y_0 = 0$.

Lemma A.2: Let $E(\tau)$ be a LCTVS. Denote by E^* the algebraic dual of E and by E' the topological dual of E. Then (E, E^*) and (E, E') are dual pairs.

Definition A.10: Let (E,F) be a dual pair. Every $y \in F$ gives rise to a seminorm

$$p_y(x) = |\langle x, y \rangle| \ .$$

The weakest topology in which all these seminorms are continuous is called the weak topology on E and denoted by $\sigma(E,F)$.

Lemma A.3: Let (E,F) be a dual pair. Equip E with the weak topology $\sigma(E,F)$. Then

$$E'(\sigma(E,F)) = F .$$

Definition A.11: Let (E,F) be a dual pair. Every locally convex topology on E under which the dual of E is F, is called a topology of the dual pair.

Definition A.12: Let (E,F) be a dual pair. For a subset $A \subset E$

$$A^\circ = \{y \in F;\ |\langle x,y\rangle| \leq 1,\ \forall x \in A\}$$

is called the polar of A.

Lemma A.4: A polar A° has the following properties:

a) A° is absolutely convex and $\sigma(F,E)$-closed

b) $A \subseteq B \Rightarrow B^\circ \subseteq A^\circ$

c) $(\lambda A)^\circ = \frac{1}{|\lambda|} A^\circ$

d) $(\cup A_\alpha)^\circ = \cap A_\alpha^\circ$

Definition A.13: Let $\mathfrak{G}$ be a family of weakly bounded sets in E with

i) $A,B \in \mathfrak{G} \Rightarrow \exists\ C \in \mathfrak{G}$ with $A \cup B \subseteq C$

ii) $A \in \mathfrak{G},\ \lambda \in \mathbb{C} \Rightarrow \lambda A \in \mathfrak{G}$

iii) $\bigcup_{A \in \mathfrak{G}} A$ spans E.

then we call $\mathfrak{G}$ saturated.

Theorem A.6: Let (E,F) be a dual pair and $\mathfrak{G}$ a saturated family on E. Then for $A \in \mathfrak{G}$, A° is absolutely convex and absorbing. Thus there exists a weakest topology in which all the A°, $A \in \mathfrak{G}$, are neighborhoods. We call this topology the topology of $\mathfrak{G}$–convergence or $\mathfrak{G}$-topology. Every topology of the dual pair is a $\mathfrak{G}$-topology.

Definition A.14: The strongest $\mathfrak{G}$-topology, where $\mathfrak{G}$ consists of all weakly bounded sets is called the strong topology on F and denoted by $\beta(F,E)$. If $\mathfrak{G}$ consists of all finite subsets of E, then we get the weak topology $\sigma(F,E)$. In general the strong topology is not a topology of the dual pair.

Theorem A.7: (Mackey-Arens). Let (E,F) be a dual pair. The strongest topology of the dual pair is given by a $\mathfrak{G}$-topology, where $\mathfrak{G}$ consists of all $\sigma(F,E)$-compact sets of F. It is called the Mackey topology and denoted by $\tau(E,F)$.

Lemma A.5: Let $E(\tau)$ be a LCTVS and E' its topological dual. Take U, a neighborhood of zero. Then U° is $\sigma(E',E)$-compact.

Lemma A.6: A lot of topological properties only depend on the dual pair. In every topology of a dual pair we have

a) The closed convex sets are the same.

b) The bounded sets are the same.

Definition A.15: Completeness:

a) An index set I is directed if for certain pairs of elements $\alpha,\beta \in I$ we have a relation $\alpha \leq \beta$ satisfying

ii) $\alpha \leq \beta,\ \beta \leq \alpha \Rightarrow \alpha = \beta$

i) $\alpha \leq \alpha$

vi) $\alpha \leq \beta,\ \beta \leq \gamma \Rightarrow \alpha \leq \gamma$

iv) $\alpha,\beta \in I \Rightarrow \exists \gamma \in I$ with $\alpha \leq \gamma,\ \beta \leq \gamma$

b) Let I be a directed set. $\{x_\alpha\} \in E,\ \alpha \in I$ is a directed system of vectors in E if the correspondence $\alpha \longleftrightarrow x_\alpha$ is unique.

c) A directed system converges to $x \in E$ if for every neighborhood U of the neighborhood base of x, $\exists\, \alpha_o \in I$ with $x_\alpha \in U\ \ \forall\, \alpha \geq \alpha_o$. We write $\lim x_\alpha = x$.

d) $\{x_\alpha\}$ is called a Cauchy net in a LCTVS E if for every neighborhood U of zero, $\exists\ \alpha_o \in I$ with $x_\alpha - x_\beta \in U,\ \forall\, \alpha,\beta \geq \alpha_o$.

Every convergent directed system is a Cauchy net. If the converse is true then E is called complete.

e) Sometimes we have to deal with LCTVS that are not complete but only sequentially complete, i.e. only the Cauchy sequences converge but not necessarily the Cauchy nets.

Lemma A.7: A metric space is complete if it is sequentially complete.

Definition A.16: A complete metric LCTVS is called a Fréchet space (F-space) and a complete normed space is a Banach space (B-space).

Theorem A.8: Take the dual pair (E,E^*). Then E^* is $\sigma(E^*,E)$-complete.

Definition A.17: Topological product and topological sums. For a sequence of LCTVS $E_i(\tau_i)$ with neighborhood base of zero $\mathfrak{U}_i$, we define

a) the topological product

$$E(\tau) = \prod E_i(\tau_i)$$

with elements $x = \{x_i\}$ and neighborhoods of zero

$$U = \{U_i\}$$

where only a finite number of U_i are not identical E_i.

The corresponding seminorms are

$$p_U(x) = \sup_i \ p_{U_i}(x_i)\ .$$

b) the topological sum

$$E(\tau) = \oplus\ E_i(\tau_i)$$

with elements

$$x = \{x_i\}\ ,$$

where only a finite number of x_i are not zero. The neighborhoods of zero are given by

$$U = \{U_i\}$$

and the corresponding seminorms

$$p_U(x) = \sup_i\ p_{U_i}(x_i)\ .$$

c) If $E_n \subset E_{n+1}$ and τ_{n+1} induces τ_n, then

$$E = \oplus\ E_n$$

is called a strict inductive limit.

Theorem A.9: The topological sum and the topological product are in duality.

a) Let $E(\tau) = \Pi\, E_\alpha(\tau_\alpha)$, then

$$E' = \oplus\ E_\alpha'$$

and

$$E(\xi(E,E')) = \Pi\, E_\alpha(\xi(E_\alpha, E_\alpha'))$$

for ξ being the weak, Mackey or strong topology.

b) Let $E(\tau) = \oplus E_\alpha(\tau_\alpha)$, then

$$E' = \Pi\, E_\alpha'$$

and

$$E(\xi(E,E')) = \oplus\ E_\alpha(\xi(E_\alpha', E_\alpha'))$$

for ξ being the Mackey or strong topology. For the weak topology, the above relation is only true if there are finitely many E_α.

Lemma A.8: The strict inductive limit of F-spaces is called a LF-space; they are complete but not metrizable.

Definition A.18: $E'' = (E'(\beta(E',E)))'$ is called the bidual of E.

$E(\tau)$ is called reflexive if $E'' = E$ and $\beta(E'',E') = \tau$.

Theorem A.10:

a) The topological sum and topological product of reflexive spaces are reflexive.

b) The topological sum and topological product of sequentially complete spaces are sequentially complete.

Theorem A.11:

a) If a LCTVS has a bounded neighborhood, then it is normable.

b) If a LCTVS has a precompact neighborhood, then it is finite dimensional.

With the above concepts we now will discuss certain special spaces.

Definition A.19: Barrelled spaces.

A LCTVS $E(\tau)$ is called barrelled if its barrels form a base of neighborhoods.

Theorem A.12:

a) If $E(\tau)$ is barrelled then

$$\tau = \tau(E,E') = \beta(E,E')$$

b) Let $E(\tau)$ be barrelled. Then E' is weakly sequentially complete.

c) All F-spaces are barrelled.

d) The topological sum and topological product of barrelled spaces are barrelled.

e) Every quotient space of a barrelled space is barrelled.

Definition A.20: Bornological spaces.

A LCTVS $E(\tau)$ is called bornological if every absolutely convex set that absorbs all bounded sets of E, is a neighborhood of zero.

Theorem A.13:

a) A LCTVS $E(\tau)$ is bornological if every bounded linear map into any LCTVS is continuous.

b) Every sequentially complete bornological space is barrelled.

c) The topological sum and topological product of countably many bornological spaces are bornological.

d) Every metrizable LCTVS is bornological.

e) Every quotient space of a bornological space is bornological.

Definition A.21: Nuclear spaces.

a) Let $E(\tau)$ be a LCTVS. A subset A in E' is called equicontinuous if for each neighborhood W in $\mathbb{C}$ there is a neighborhood U in $E(\tau)$ such that

$$\bigcup_{T \in A} T(U) \subset W .$$

b) Let E and F be two LCTVS. A continuous linear map

$$T: E \to F$$

is nuclear provided it can be written in the form

$$Tx = \sum_{n=1}^{\infty} \lambda_n \langle a_n, x \rangle \, y_n \qquad \forall x \in E$$

with $\{a_n\}$ being an equicontinuous set in E', $\{y_n\}$ a bounded sequence in F, and $\sum_{n=1}^{\infty} |\lambda_n| < \infty$.

c) A LCTVS $E(\tau)$ is nuclear if every continuous linear map of $E(\tau)$ into a B-space is nuclear.

Theorem A.14: A LCTVS E is nuclear if a neighborhood base of zero $\mathfrak{U}(0)$ has the following property: for every $U \in \mathfrak{U}(0)$ $\exists\, V \in \mathfrak{U}(0)$ and a positive Radon measure μ on the weakly compact polar V°, such that

$$p_U(x) \leq \int_{V^\circ} |\langle a, x \rangle| \, d\mu(a) \quad \forall x \in E$$

Theorem A.15:

a) All nuclear F-spaces are reflexive and separable.

b) In every nuclear LCTVS, all bounded sets are precompact.

c) A nuclear normed space is finite dimensional.

d) A F-space is nuclear if its strong dual is nuclear.

e) Each linear subspace of a nuclear LCTVS is nuclear.

f) The quotient of a nuclear LCTVS relative a closed subspace is nuclear.

g) The topological sum and topological product of nuclear spaces are again nuclear.

Remarks:

a) All our LCTVS are assumed to be Hausdorff. It then turns out that the topological sum and topological product are Hausdorff too.

b) We only dealt with countable topological sums and products. Some of the above results however still hold in the uncountable case.

References

1. A. S. Wightman, Phys. Rev. 101, 860 (1956).
2. R. F. Streater and A. S. Wightman, PCT, Spin,Statistics, and All That (Benjamin, New York, 1964).

3. H. J. Borchers, Nuovo Cimento 24, 214 (1962).
4. R. Jost, The General Theory of Quantized Fields (Am. Math. Soc., Providence, 1965).
5. K. Hepp, Brandeis University Summer Institute 1965 (Gordon and Breach, New York, 1966).
6. L. Schwartz, Théorie des Distributions I,II (Hermann, Paris, 1957/59).
7. A. M. Jaffe, Phys. Rev. 158, 1454 (1967).
8. G. Köthe, Topologische Lineare Räumme I (Springer-Verlag, 1960).
9. A. P. Robertson and W. J. Robertson, Topological Vector Spaces (Cambridge University Press, Cambridge, 1964).
10. A. Pietsch, Nukleare Lokalkouvexe Räume, (Akademie-Verlag, Berlin, 1965).
11. G. Choquet, Princeton Lecture Notes on Analysis, 1967.
12. N. Jacobson, Structure of Rings (American Mathematical Society Colloquium Publication, 1964).
13. M. A. Neumark, Normierte Algebreu (Deutsches Verlag der Wissenschaften, 1959).
14. O. Lanford and D. Ruelle, J. Math. Phys. 8, 1460 (1967).
15. R. Haag, Phys. Rev. 112, 669 (1958).
16. D. Ruelle, Helv. Phys. Acta 35, 147 (1962).
17. K. Hepp, Commun. Math. Phys. 1, 95 (1965).
18. H. Lehmann, K. Symanzik and W. Zimmermann, Nuovo Cimento 1, 205 (1955) and Nuovo Cimento 6, 319 (1957).
19. O. Steinmann, Commun. Math. Phys. 10, 245 (1968).
20. R. Jost, Helv. Phys. Acta 39, 21 (1966).

SOME CONSIDERATIONS ON CONFIGURATION SPACE ANALYTICITY IN GENERAL QUANTUM FIELD THEORY†

A. C. Manoharan
Department of Physics
University of Wisconsin
Milwaukee, Wisconsin

Contents

Introduction
I. Bounds on Holomorphy Envelopes
II. A Simplified 4-point Function Domain
III. Bounds Continued
IV. CP^1T in Odd-dimensional Spacetime
V. Remarks on Primitive Domain Boundaries
Conclusion
References

Introduction

A. We would like to discuss in these lectures some topics concerning configuration space (x-space) analyticity within the framework of the general theory of quantum fields. This will be with the use of the Wightman formulation[1),2)] which Walter Wyss has introduced to you already. If he has shown you how "elaborate" the theory turns out to be, let me also show you how "instructive" it can be.

The formulation is in terms of analytic Wightman (W) functions which have as boundary values vacuum expectation values of products of field operators. This enables one to mathematically analyse physically reasonable features of successful field theories.

There arise problems on the nature of the domains of analyticity and construction of envelopes of holomorphy for these functions. These questions interest the physicist when he wants to look at dispersion relations or even the CPT theorem. They also interest the mathematician because they furnish interesting examples to study in the theory of several complex variables. This is a rather new subject,

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

and it is only fairly recently that people have begun to think of several complex variables as forming part of the mathematical methods of theoretical physics. The topics dealt with here are selected to convey briefly an optimum amount of information in a manner that the directors of this Institute might like.

Basic information on notation, definitions and important known results which we need to use are contained in the Introduction.

In Section I we shall see, in terms of upper bounds, how holomorphy envelopes for W functions of different orders are related. This can be done in a manner which does not involve explicit constructions of envelopes of holomorphy. The results are valid for arbitrary spinor fields and apply also to arbitrary spacetime dimension.

The simplifications due to 2-dimensional spacetime will be considered in Section II, together with some typical partial analytic extensions of the 4-point function domain. Knowledge of the above material will enable some remarks to be made in Section III concerning some possible conjectures for partial analytic extensions and also Wu's conjecture (derived from perturbation theory) for the 4-point function. In Section IV we shall look at some physical consequences of analyticity, namely, CPT, in a novel way, that is with regard to the evenness or oddness of the dimension of spacetime. When the dimension of spacetime exceeds 2, extra types of boundaries appear for extended tubes beyond those dealt with in Section II. Remarks on this matter are made in Section V, where we consider 3-dimensional spacetime for simplicity and show how not only extra boundaries but even more boundaries show up in a manner which is also true for 4-dimensional spacetime.

B. Notation

1. Dimension of spacetime: S i.e. 1 time dimension and S - 1 space dimensions.
2. Scalar product: $(\xi,\xi') = \xi^{\mu}\xi'_{\mu} = \xi^{o}\xi'^{o} - \sum_{k=1}^{S-1} \xi^{k}\xi',^{k}$

$$= g_{\mu\nu}\,\xi^{\mu}\xi',^{\nu} \ .$$

where $g^{\mu\nu} = \mathrm{diag}\,(1,-1,\ldots-1)$.

3. Forward / Backward } Cone: $V_{\pm} = \{\xi;(\xi,\xi) > 0,\ \xi^{o} \gtrless 0\}$.
4. Forward tube domain: $\mathcal{T}_n = \mathcal{T}_n(\zeta_1,\ldots\zeta_n)$

$$=\{(\zeta_1,\ldots\zeta_n)\in \mathbb{C}^{Sn};\ \zeta_i=\xi_i-i\eta_i,$$

$$\eta_i \in V_+,\ i = 1,\ldots n\}.$$

5. Proper orthochronous Lorentz group: $L_+^\uparrow$.
Covering group of $L_+^\uparrow$: $\begin{cases} \tilde{L}_+^\uparrow = \text{2 sheeted (spinor) covering group for } S > 3 \\ L_+^\uparrow \text{ for } S = 2. \end{cases}$
(so $\tilde{L}_+^\uparrow$ is the universal covering group for $S \geq 4$).
Complex Lorentz group (identity component): $L_+(C)$.
Inhomogeneous (spinor) group: The semi-direct product $T \wedge \tilde{L}_+^\uparrow$ where T is the translation group.
6. Extended Tube (not necessarily a tube): $\mathfrak{T}_n' = \bigcup_{\Lambda \in L_+(C)} \Lambda \mathfrak{T}_n$.

C. Preliminaries

This is all material you will find in the books.[1),2)] We shall sketch it here so as to use it to discuss what is not in the books.

The n point Wightman (W) distribution is the vacuum expectation value

$$W_n(x_1, x_2, \ldots x_n) = \langle 0 | \varphi_1(x_1) \varphi_2(x_2) \ldots \varphi_n(x_n) | 0 \rangle$$

where $x_1, x_2, \ldots x_n$ are n spacetime points and $\varphi_1, \varphi_2, \ldots \varphi_n$ are field operators (operator valued continuous linear functionals)[3)] transforming according to representations of the inhomogeneous spinor group $T \wedge \tilde{L}_+^\uparrow$. Spinor indices have been suppressed. The notion of "field" together with field transformation properties we shall refer to as the physical assumption of Lorentz invariance.

We shall have the usual physical assumptions that go with the fact that we are doing quantum mechanics, i.e. the underlying Hilbert Space, except for the positive definite scalar product in the space. Consequences of the latter will not be discussed in Section I, II and III but indirectly in IV. Except in Section IV, the discussion is within the so-called "linear program" since nonlinear relations between different order Wightman distributions will not be exploited.

Another physical assumption is that the support of the spectrum of the momentum operator is within the closed forward light cone $\bar{V}_+$; this can be referred to as "reasonable mass spectrum." We would like to point out that further spectral restrictions will not change, within the linear program, the x-space analyticity domains.

The final assumption we incorporate is microscopic causality, that field operators commute (anti-commute) at spacelike separations --referred to as local commutativity. This can be compactly written $[\varphi(x), \varphi(y)]_\pm = 0$, for $(x - y)^2 < 0$.

On account of translation invariance and the support properties of the spectrum one has the well known result that:

The W distribution $\mathbb{W}_n(x_1, \ldots x_n)$ is the boundary value of an analytic function (called Wightman (W) function) $W_n(\zeta_1, \ldots \zeta_{n-1})$ holomorphic in the tube domain $\mathcal{T}_{n-1}(\zeta_1, \ldots \zeta_{n-1})$ where $\zeta_i = \xi_i - i\eta_i$, $\xi_i = x_i - x_{i+1}$, $i = 1, \ldots n-1$.

Here one has "proceeded to the quotient" in the translation invariance since the difference variables $\xi_i = x_i - x_{i+1}$ are used. It helps sometimes to preserve the symmetry in the x_i by taking $\mathbb{W}_n(x_1, \ldots x_n)$ as the boundary value of an analytic function $\mathbb{W}_n^n(z_1, \ldots z_n)$ holomorphic in the domain

$$\mathcal{O}_n(z_1, \ldots z_n) \equiv \mathcal{O}_n(z_1 - z_2, \ldots z_{n-1} - z_n)$$

$$= \{(z_1, \ldots z_n);\ (z_1 - z_2, \ldots z_{n-1} - z_n) \in \mathcal{T}_{n-1}\}\ .$$

It is a deep theorem due to Bargmann, Hall and Wightman (and Jost generalized it for $S \neq 4$) that W_n can be analytically extended into the extended tube domain $\mathcal{T}'_{n-1}(\zeta_1, \ldots \zeta_{n-1})$ or equivalently $\mathbb{W}_n$ into $\mathcal{O}'_n(z_1, \ldots z_n) \equiv \mathcal{O}'_n(z_1 - z_2, \ldots z_{n-1} - z_n) = \{(z_1, \ldots z_n);\ (z_1 - z_2, \ldots z_{n-1} - z_n) \in \mathcal{T}'_{n-1}\}$. This arises from the complex extension of the group $L_+^\uparrow$ to $L_+(C)$.

The use of the locality assumption produces a further analytic extension of the domain $\mathcal{O}'_n$ into the totally symmetric domain, the union of permuted extended tubes $\mathcal{O}_n^P(z_1, \ldots z_n) = \bigcup_{\sigma \in \mathcal{S}_n} \mathcal{O}'_n(z_{\sigma(1)}, \ldots z_{\sigma(n)})$ where σ denotes permutations of the symmetric group $\mathcal{S}_n$ on n symbols.

Although in many simpler cases $\mathcal{O}'_n$ is a natural domain of holomorphy, (this is known for all n when $S = 2$, and for $n \leq S+1$ when $S \geq 3$) the domain $\mathcal{O}_n^P$ for $n > 2$ is not a natural domain of holomorphy.

Thus an analytic extension problem arises for $\mathcal{O}_n^P$, that is, one is interested in its holomorphy envelope $H(\mathcal{O}_n^P)$ i.e. the largest domain into which all functions holomorphic in $\mathcal{O}_n^P$ can be analytically continued in a single valued manner. From past experience this has turned out to be an extremely difficult problem which is one reason why it is worth discussing. $H(\mathcal{O}_2^P) = \mathcal{O}_2^P$ trivially, $H(\mathcal{O}_3^P) = KW$ (say) was computed (a highly nontrivial problem) by Källén and Wightman.[5]

I. Bounds on Holomorphy Envelopes

A. Now we shall discuss (within the linear program) how without explicitly constructing envelopes of holomorphy one can make statements about what analyticity will not come out.

Let $\mathbb{C}^{Sr}(z_1, \ldots z_r)$ be the space of Sr complex variables, and x denote the topological product. Then we have

Theorem 1 $$H[\mathfrak{S}_n^P(z_1, \ldots z_n)] \subset \bigcap_{\sigma \in \mathfrak{S}_n} \{ \mathbb{C}^{Sr}(z_{\sigma(1)}, \ldots z_{\sigma(r)}) \times H[\mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \ldots z_{\sigma(n)})] \} \quad \forall\ 0 < r < n,$$

relative to $H(\mathfrak{S}_n^P)$.

The proof is available elsewhere[4] and we shall just describe the results here.

B. One of the complications that arise in the construction of holomorphy envelopes is that H(D) of a schlicht (single sheeted) domain D may not be schlicht. It is not known at present whether $H(\mathfrak{S}_n^P), n \geq 4$ is schlicht but we know that $H(\mathfrak{S}_3^P) = KW$ is schlicht. For schlicht domains the symbol $\subset$ denotes inclusion. For non-schlicht domains we have to state first what kind of domain we are talking about. It is enough to work within the class of locally schlicht Riemann domains because that is the nature of the domain $\mathfrak{S}_n^P$ which we start from, and we are assured that there is a unique holomorphy envelope within this class. Then $\subset$ denotes that the projection into $\mathbb{C}^{Sn}$ of the Riemann domain or covering space on the left hand side is contained within that of the right.

The relative intersection means that only the component of the intersection on the right hand side connected to $\mathfrak{S}_n^P$ applies. The theorem thus states that $H(\mathfrak{S}_n^P)$ is bounded (with respect to the projection) by the relative intersection of all permuted cylinders having lower order function domains as base.

This means in particular that for $n > 3$ no analytic extension is possible (within the linear program) through the cuts (the 2-point function singularities) nor beyond the intersection of permuted Källén–Wightman (KW) domains for the 3-point function.

It might be remarked that it has helped to preserve the symmetry by "leaving in" the translational invariance as well as the homogeneous Lorentz invariance. If one works with the Lorentz invariants it is necessary to consider a more complicated type of analytic space.

C. Proof of Theorem (sketch)

Defn Holomorphy envelope of D: $H(D) = \bigcap_{f \in F} H_f$, relative to D, where H_f is the holomorphy domain for the function f and F is the set of all functions holomorphic in D.

We use 3 lemmas:

Lemma 2: If $D_1 \subset D_2$ then $H(D_1) \subset H(D_2)$.

Proof: Functions holomorphic in D_2 are also holomorphic in D_1 and this follows directly from the definition. ∎

Lemma 3: Consider 2 domains $D = D(z_1, \ldots z_\ell)$ and $E = E(w_1, \ldots w_m)$. Then $H(D \times E) = H(D) \times H(E)$.

Lemma 1: $\mathfrak{S}_n^P(z_1, \ldots z_n) \subset \mathbb{C}^{Sr}(z_{\sigma(1)}, \ldots z_{\sigma(r)}) \times \mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \ldots z_{\sigma(n)})$, $\forall\, \sigma \in \mathfrak{S}_n$, $\forall\, 0 < r < n$.

The proof involves the structure of $\mathfrak{S}_n$ and $\mathfrak{S}_n'$.

Proof of theorem:

Using lemmas 1 and 2,

$$H[\mathfrak{S}_n^P(z_1, \ldots z_n)] \subset H[\mathbb{C}^{Sr}(z_{\sigma(1)}, \ldots z_{\sigma(r)}) \times \mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \ldots z_{\sigma(n)})]$$

$$\forall\, \sigma \in \mathfrak{S}_n, \qquad \forall\, 0 < r < n .$$

Since $H(\mathbb{C}^{Sr}) = \mathbb{C}^{Sr}$, we have from lemma 3

$$H[\mathfrak{S}_n^P(z_1, \ldots z_n)] \subset \mathbb{C}^{Sr}(z_{\sigma(1)}, \ldots z_{\sigma(r)}) \times H[\mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \ldots z_{\sigma(n)})]$$

$$\forall\, \sigma \in \mathfrak{S}_n .$$

Thus

$$H[\mathfrak{S}_n^P(z_1, \ldots z_n)] \subset \bigcap_{\sigma \in \mathfrak{S}_n} \{\mathbb{C}^{Sr}(z_{\sigma(1)}, \ldots z_{\sigma(r)}) \times H[\mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \ldots z_{\sigma(n)})]\}$$

$\forall\, 0 < r < n$, relative to $H[\mathfrak{S}_n^P(z_1, \ldots z_n)]$. ∎

II. A Simplified 4-point Function Domain

A. Now we shall look at some examples of analytic extensions of a domain. Here we encounter apparent singularities which turn out to be partly removable. A good example is the 4-point function. This can be generally considered for $S \geqslant 3$ with equivalent results on account of the fact that $\mathfrak{S}_4'$ has the same types of boundary hypersurfaces for all $S \geqslant 3$. Taking $S = 2$ leaves out the DANAD boundary and makes the problem much simpler.

We mention in this connection that people have also found it instructive to study quantum field theory models for $S = 2$ because they can go further in solving them.

B. We take characteristic coordinates[6] $r_i = z_i^0 + z_i^1$, $s_i = z_i^0 - z_i^1$, $i = 1 \ldots, 4$. In these coordinates $\Lambda \in L_+(C)$ is diagonal with $\Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$, $\lambda \in \mathbb{C}$, $\lambda \neq 0$. Also if λ is real and > 0 then $\Lambda \in L_+^{\uparrow}$.

Let us define $t_i = -1/s_i$, $u_i = -\bar{s}_i$, $r_{ij} = r_i - r_j$, $s_{ij} = s_i - s_j$, $u_{ij} = u_i - u_j$, $t_{ij} = -1/s_{ij}$. Then $\mathcal{O}_4(z_1, z_2, z_3, z_4)$ is represented by Fig. 1 which is an Argand diagram where all the "vectors" $r_{12}, r_{23}, r_{34}, u_{12}, u_{23}, u_{34}$ lie in the open lower half plane.

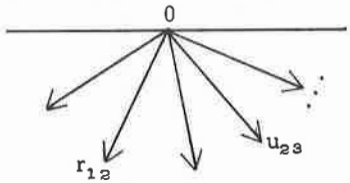


Fig. 1. A point of $\mathcal{O}_4$.

Applying $\Lambda \in L_+(C)$ to such a point rotates and lengthens or shortens these vectors so that we get a point of $\mathcal{O}_4'(z_1, z_2, z_3, z_4)$ which can be represented by stating that all the "vectors" lie on one side of some open half plane bounded by a straight line through the origin. Hence the boundary $\partial\mathcal{O}_4'$ of $\mathcal{O}_4'$ consists of the following analytic hypersurfaces.

$$6 \text{ "cuts," } C_{ij} \equiv C_{ij,ij}: r_{ij} = -\rho u_{ij} \qquad i \neq j,$$

$$\left.\begin{aligned} S_{ij,k\ell}: r_{ij} &= -\rho r_{k\ell} \\ S_{ij',k\ell'}: u_{ij} &= -\rho u_{k\ell} \\ F_{ij,k\ell'}: r_{ij} &= -\rho u_{k\ell} \end{aligned}\right\} \quad i \neq j,\ k \neq \ell,\ ij \neq k\ell \neq ji$$

where ρ is a positive parameter (separate for each surface).

$\mathcal{O}_4'$ is a natural domain of holomorphy, but $\mathcal{O}_4^P$ which consists of a union of 24 permuted domains (which coincide in pairs due to CPT) is not so. It does not help very much to make a head-on attack by combining all 12 domains and trying to make analytic extensions. But fortunately Streater found a way of finding holomorphy envelopes of the union of a few domains at a time as well as conjectures for other sets of domains. So we begin with the union of all Streater domains into which analytic extensions have already been made.[7]

C. The problem is still very complicated, and we shall only show here how analytic extensions[8] can be made in some typical simple cases, i.e. in certain directions, up to the boundary of the upper bound given by theorem I.

Table I defines 9 Streater domains α_i, β_i, ν_i, $i = 1,2,3$. Each is in the form of an intersection of extended tube domains. For example

i	α_i	β_i	γ_i
1	(13,23,14,24) ∩ (12) ∩ (34)	(24,34,21,31) ∩ (23) ∩ (14)	(12,32,14,34) ∩ (13) ∩ (24)
2	(13,23,41,42) ∩ (12)	(24,34,12,13) ∩ (23)	(12,32,41,43) ∩ (13)
3	(13,32,14,42) ∩ (34)	(24,43,21,13) ∩ (14)	(12,23,14,43) ∩ (24)

Table I. Streater Domains

(13,23,14,24) denotes the extended tube domain $\mathcal{T}_4'(z_1-z_3, z_2-z_3, z_1-z_4, z_2-z_4)$ which is characterized by stating that $r_{13}, r_{23}, r_{14}, r_{24}, u_{13}, u_{23}, u_{14}, u_{24}$ all lie in an open half plane with 0 on the boundary. Let us also abbreviate r_i, u_i, r_{ij}, u_{ij} by i, i', ij, ij' respectively.

In Fig. 2 all the complex variables but u_4 are fixed in a certain manner shown (generally off boundary hypersurfaces in themselves) and the contributing domains are plotted in the u_4 plane. The curved arrows indicate singularity regions (other than cuts) and the boundary of the union of domains is shaded on the (apparent) singularity side. We note that 3 out of the 6 cuts appear in such a diagram.

The boundary contains indented corners $F_{34',14} \cap F_{24',43}$, $F_{14',24} \cap F_{24',21}$ and we can apply the Kantensatz. Each corner is not an analytic manifold and therefore is not a possible boundary point for a natural domain of holomorphy.

D. In checking for analytic manifolds it helps to read this off from our notation by the

Rule[19] : Consider a (non-trivial) corner of the form $A_{a,b} \cap B_{c,d}$ where A,B denote any C, S or F analytic hypersurfaces, and a,b,c,d denote primed or unprimed double indices in our notation for these surfaces. Then the corner is an analytic manifold if one of the subscripts of A is equivalent to $\pm$ one of B (e.g. if $a \equiv -c$, $b \not\equiv \pm d$) and not an analytic manifold if otherwise (e.g. $12 \equiv -21$ but $23 \not\equiv \pm 23'$).

Proof: Apply the general treatment of the Kantensatz in Wightman's Les Houches Lectures.[9] ∎

It was noticed by Källén and Wightman for the 3-point function in a situation where a corner shows up in a figure of this sort that the infinitesimal analytic extension into the corner provided by the Kantensatz can be generalized (using more profound techniques) to an analytic extension up to the convex closure. In fact this works here too, as we shall see below, and we can extend through the two corners up to the cuts as shown in the lower left of Fig. 2. In fact there are still corners, which do not appear on the figure, which permit a further analytic extension into the shaded domain between the cuts, entering, of course, through the gap between the cuts. The final domain in the restriction of Fig. 2 has only the cuts as singularities.

If we recall[6] that the 3-point function has no singularities, except cuts, in this kind of diagram, when the cuts do not intersect, then the analytic extension has been made here (in this restriction) up to the intersection of 3-point function domains, $\bigcap_{\sigma \in \mathfrak{S}_4} KW(z_{\sigma(1)}, z_{\sigma(2)}, z_{\sigma(3)})$ which is as far as we can go by Theorem I.

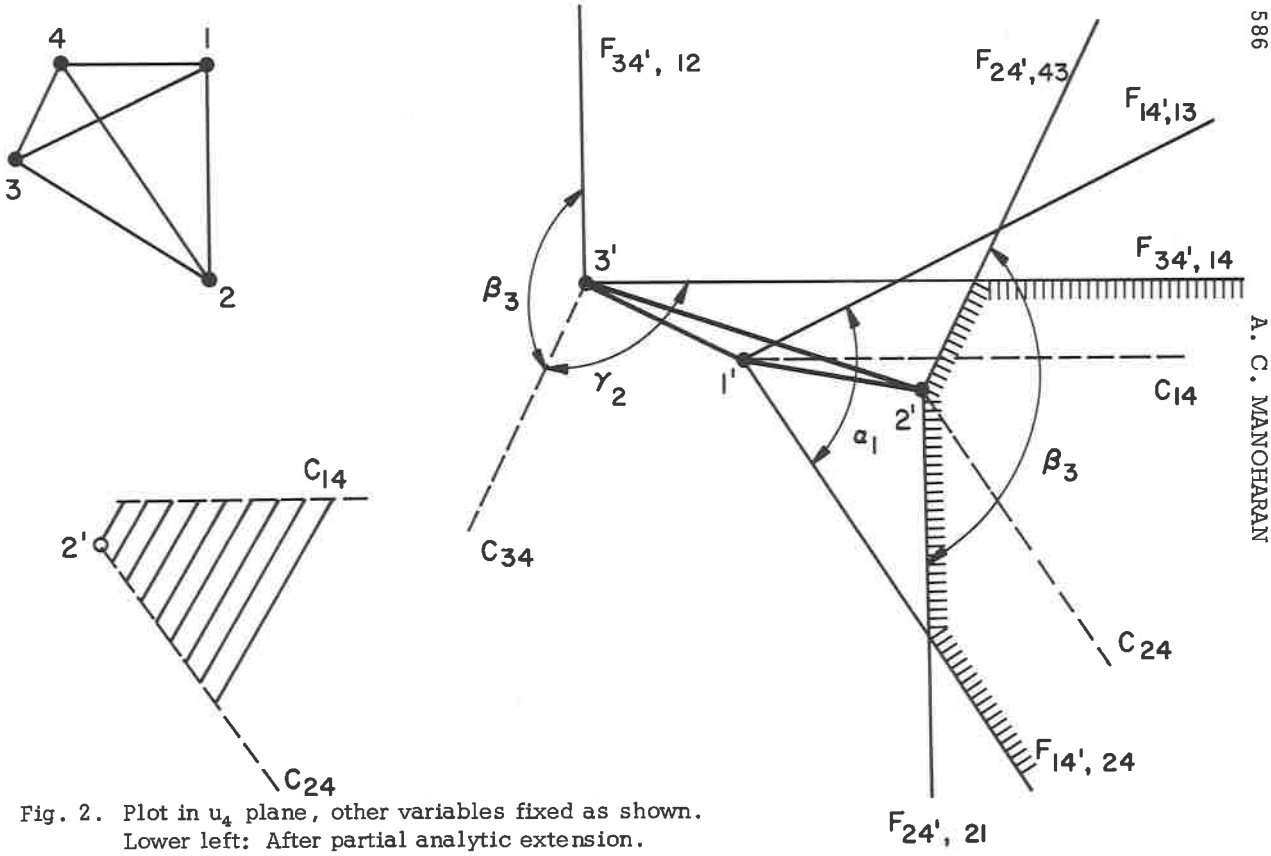


Fig. 2. Plot in u_4 plane, other variables fixed as shown. Lower left: After partial analytic extension.

E. Proof of Analytic Extension

We sketch, as a typical case, the extension through the corner $F_{34',14} \cap F_{24',43}$ up to the convex closure in the figure. Consider the domain D_1 =

$$\begin{array}{lllllll} & 31' < 32' < 12' & < 41 & < 31 < 34 < 21 < 24 < 23 \\ \cap & & 41 & < 24' < 34 \\ \cap & 12' < 34' < 41 & & \\ \cap & & 41 < 14' & < 34 \end{array}$$

where $41 < 24' < 34$ means $\exists\, \theta$ such that

$$\theta < \arg 41 < \arg 24' < \arg 34 < \theta + \pi .$$

So D_1 is an intersection of 4 subdomains of extended tubes. Also D_1 contains the convex closure we are interested in. Since $12' + 24' = 14'$, $14'$ can be regarded as a dependent variable, which also satisfies the convexity condition $12' < 14' < 24'$. (We have left in all the the variables for symmetry purposes.)

We are actually working[8] in a complex projective space $\mathbb{P}^7 = \mathbb{C}^{8*}/\mathbb{C}^*$ where $\mathbb{C}^{n*} = \mathbb{C}^n - \{(0, \ldots 0)\}$ and we define Lorentz invariants or ratios of homogeneous coordinates by

$$d = t_{24}/r_{41}\,, \qquad e = r_{34}/r_{41}\,, \qquad f = r_{34}/t_{34}$$

with other independent variables being fixed in D_1.

Ruelle's lemma[6]: If the function $\varphi(d,e,f)$ of three complex variables is holomorphic in the product P of the 3 upper half planes, except perhaps for $\arg d \le \arg e \le \arg f$, then φ is holomorphic in P.

Proof of lemma: Take new variables $\log d$, $\log e$, $\log f$. Then φ in the new variables is holomorphic in a tube domain having as base the cube in Fig. 3 less the dotted prism. These singularities are removable by taking the convex closure using the tube theorem. ∎

The lemma is not immediately applicable even for the 3-point function. In that case, there are cuts present as additional singularities, but they do not intersect the singularities that are to be eliminated. The tube theorem is inapplicable but by use[10] of the strong continuity theorem of Bremermann the singularities (except the cuts) can be eliminated. Thus one has a "strong Ruelle Lemma."

For the situation here we need apply a much stronger Ruelle lemma (see Fig. 4) where there are other singularities present together with cuts, but which leave enough room to complete into the triangle shaped region by means of analytic discs of the form indicated by the dotted line. This is possible for the domain D_1 and hence the singularities in D_1 are eliminated.

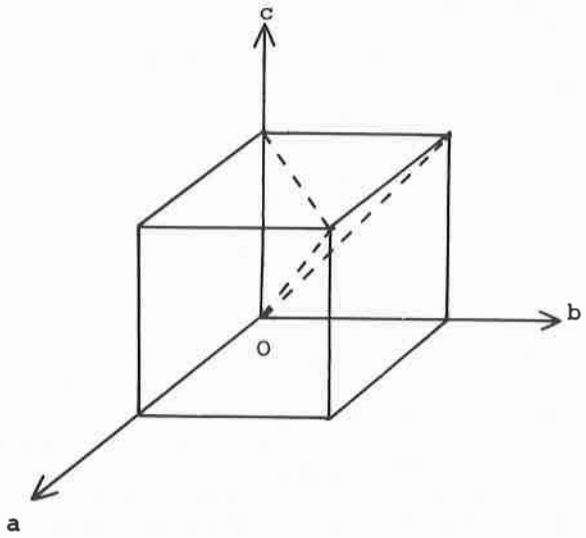


Fig. 3. Ruelle's cube with a = arg d, b = arg e, c = arg f.

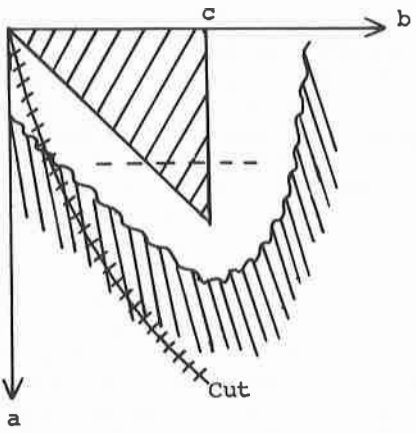


Fig. 4. The last figure with certain additional singularities, for fixed c.

Fig. 5 depicts another case. Here the shaded singularities above the cut C_{14} are removable. Those below, in D, are not immediately removable. However it has been possible[8] to extend the domain in some directions up to the $\mathfrak{F}$ boundaries of the 3-point function. These boundaries are known to be very troublesome ones to deal with.

III. Bounds Continued

A. Using theorem I of Sec. I and our knowledge about diagrams of domains in Sec. II for S = 2 we have

Theorem II: It is impossible to extend the 4 point function domain $\mathfrak{S}_4{}^P$ for $S \geqslant 2$ (in particular for S = 4) into the entire domain $D^1_{R\cdot A} = D_{R\cdot A}(12,13,14) = \{(z_1,z_2,z_3,z_4);(z_1-z_2,z_1-z_3,z_1-z_4) \in \mathfrak{F}_3{}'(z_1-z_2,z_1-z_3,z_1-z_4)\}$. (It is understood that all the 4 point function cuts do not belong to $D^1_{R\cdot A}$.)

Proof: By permutation symmetry of $\mathfrak{S}_4{}^P$ it is enough to consider $D^2_{R\cdot A} = D_{R\cdot A}(23,24,21)$ in place of $D^1_{R\cdot A}$. First let S = 2. In Fig. 6 are shown the singularities of $D^2_{R\cdot A}$ and the 3 point domain KW(134) in the variables z_1, z_3, z_4. Since $D^2_{R\cdot A}$ does not have any singularities, except for the cuts, within the singularities of KW(134) we have points of holomorphy inside $D^2_{R\cdot A}$ which are outside the upper bound of theorem I. Thus the theorem is proved for S = 2.

For $S \geqslant 3$, in addition to the hypersurface boundaries C, S, F for $D^2_{R\cdot A}$ there is an extra one of the type DANAD. Other boundary points are limit points of these hypersurfaces since they are of lower dimension. We simply consider the restriction of $D^2_{R\cdot A}$ to the analytic manifold $z_i{}^2 = z_i{}^3 = \ldots = z_i{}^{s-1} = 0$, $i = 1, \ldots 4$. The DANAD vanishes without trace except for S, F surfaces and we have Fig. 6 again.

B. Discussion

Firstly we note that if we replace $\mathrm{Re}(z_1-z_i)$ by p_i in $D^1_{R\cdot A}$ we have a primitive extended tube domain for retarded and advanced functions r_1, a_1 of LSZ in p space.[11] The apparent similarity prompts the notation. (If one begins with the support properties in x-space of $\tilde{r}_1, \tilde{a}_1$ one has analyticity in the appropriate tubes for r_1, a_1 in p space.) The similarity holds also for the 3-point function where we have (in x-space) $D_{R\cdot A}(12,13)$. But this is actually a Streater domain (together with cuts) obtained in the form

$$H[\mathfrak{S}_3{}'(12,23) \cup \mathfrak{S}_3{}'(13,32)] = \mathfrak{S}_3'(12,13) \cap \mathfrak{S}_2'(23)$$

and we can and do have holomorphy in the entire domain, whereas this does not happen for the 4-point function.

Secondly we recall that some time ago, by looking at possible Landau singularities of the crossed box diagram in p-space it was conjectured by Wu[12] that the DANAD boundary for $D^1_{R\cdot A}$ (or

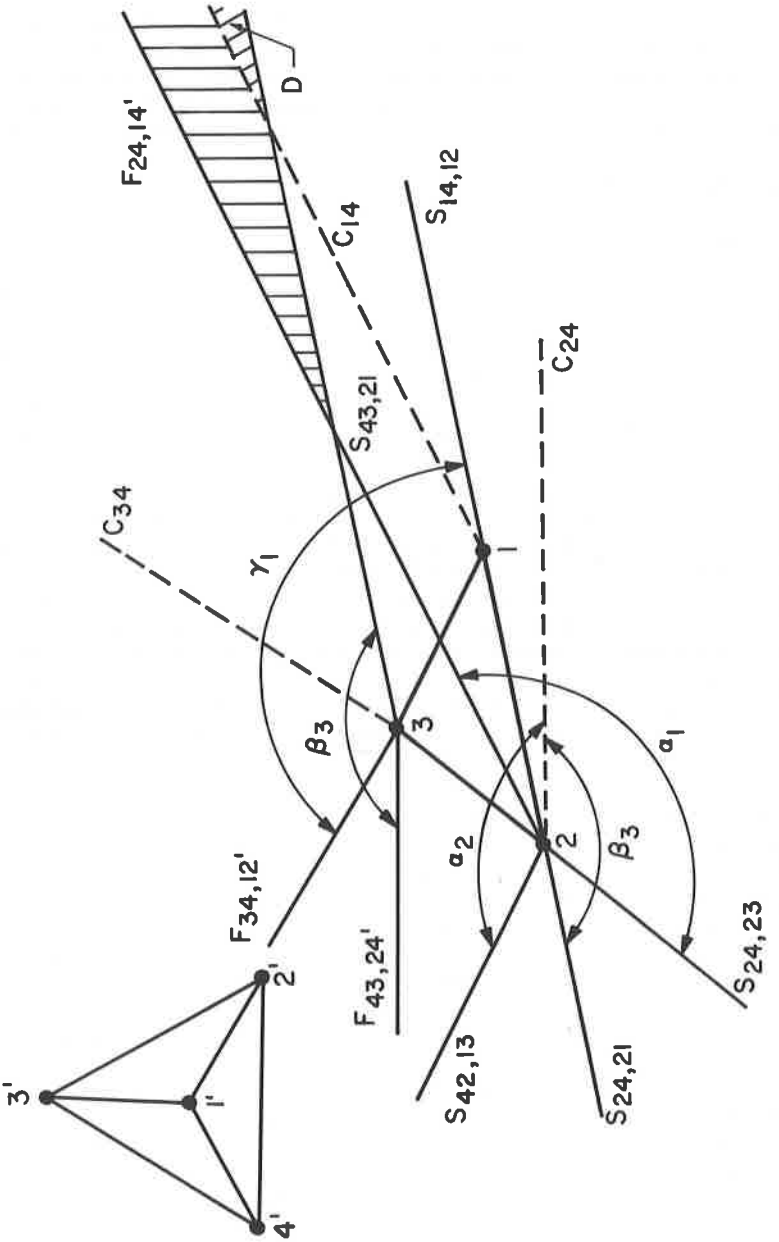


Fig. 5. Plot in r_4 plane, other variables fixed as shown.

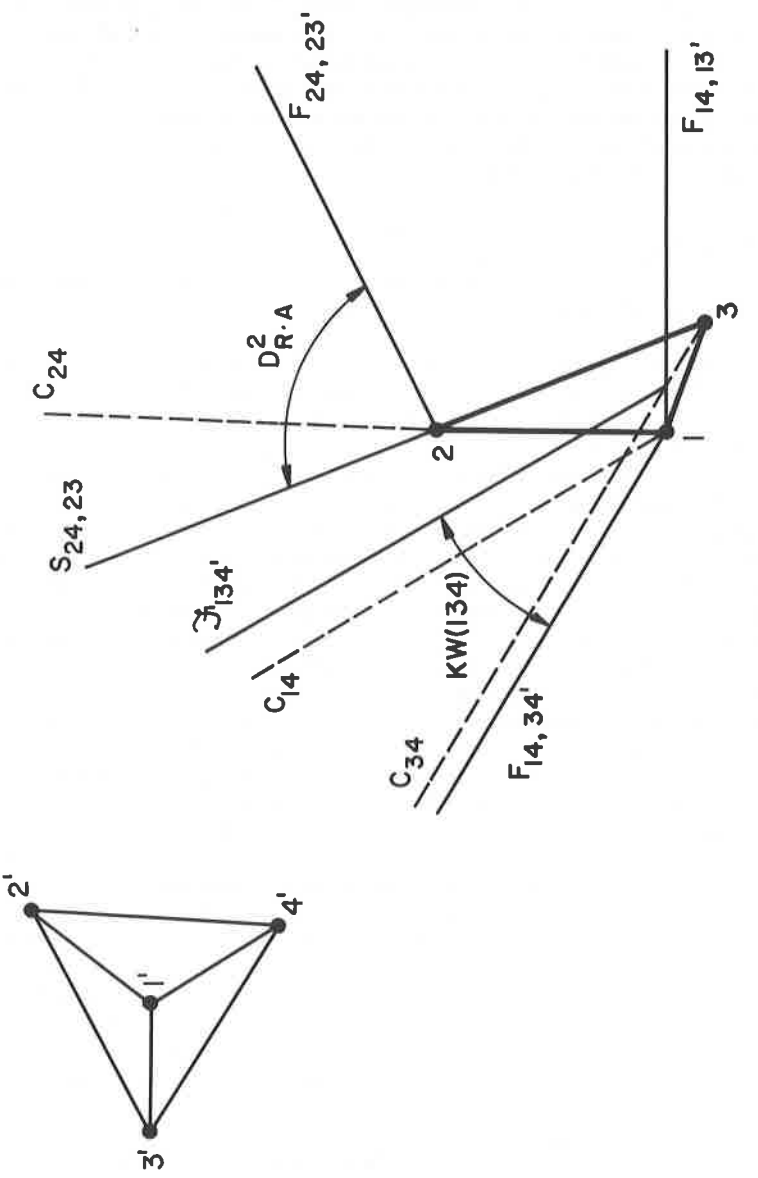


Fig. 6. Plot in r_4 plane.

permutations) is the main boundary of $H(\mathfrak{S}_4{}^P)$ (which is a statement for x-space). This conjecture gives an upper bound, which if valid, is further narrowed down by our proved bound of Theorem I in accordance with Theorem II. However Wu's work has been criticized[13] on the grounds that he carried over the strong symmetry between x- and p-space analyticity which applies to the 3-point function unjustifiably to the case of the 4-point function.

IV. CP^1T in Odd Dimensional Spacetime

A. Under the permutation TCP (tri-cresyl-phosphate), this additive was first introduced by the Shell Oil Company as a cleaning agent for the lead which goes into the antiknock additives put in to increase the octane rating of gasoline.

We are concerned here with the CPT theorem in the form of Lüders,[14] Pauli[15] and Jost.[16] This was formulated for S = 4. We wish to see whether or not the theorem holds with modifications for S = an odd number.

As is well known the CPT theorem for S = 4 is closely connected with the fact that $PT \in L_+(C)$, for S = even, the connected complex extension of $L_+^\uparrow$, although neither P nor $T \in L_+^\uparrow$. We can begin with a relativistic theory invariant (or covariant) under $L_+^\uparrow$ and not necessarily possessing the reflection invariances C, P, T separately, but find that the theory is indeed invariant under the product reflection CPT. As pointed out by Jost[16] CPT is a direct consequence of the fundamental physical assumptions of Wightman's quantum field theory with the local commutativity assumption replaced by one of only weak local commutativity in neighbourhoods of Jost points. The results of Bargmann-Hall-Wightman on the extension of Wightman functions from the domain $\mathfrak{S}_n$ which is invariant under $L_+^\uparrow$ into $\mathfrak{S}_n'$ which is invariant under $L_+(C)$, also plays a crucial role in the proof of the CPT theorem.

Now $\Lambda = PT$ has $\det \Lambda = +1$ for S = even but $\det \Lambda = -1$ for S = odd. Thus $\Lambda \notin L_+(C)$ for S = odd and one will not expect[19] CPT invariance for S = odd for the same reasons as one expects it for S = even.

Furthermore, $P \in L_+^\uparrow$ for S = odd (it is simply a space rotation) and we automatically have P invariance following from $L_+^\uparrow$ invariance. So it is natural to take, instead of P which reflects all space coordinates, the reflection P^1 of 1 space coordinate only (which can be taken as the first one). $P^1T \in L_+(C)$ for all S and P^1 is obtained from P by a space rotation for S = even; thus the CPT theorem for S = even is also a CP^1T theorem. Hence we shall look for CP^1T invariance for S = odd (also).

B. Scalar Fields

Let us first consider scalar fields and let us look at Lüders form of the theorem. We note, in passing, that this proof is accomplished

in the interaction picture, by showing invariance of the free field equation and commutation relations and the Hamiltonian density.

C. Lüders CP^1T for Scalar Fields, All S

With the space vector $\underset{\sim}{r} = (r_1, r_2, \ldots r_{s-1})$ we associate $P^1\underset{\sim}{r}=-\underset{\sim}{r}'$, $\underset{\sim}{r}' = (+r_1, -r_2, \ldots -r_{s-1})$. HC denotes Hermitian conjugation, and SR "strong" reflection i.e. P^1T so that $(SR)(\underset{\sim}{r},t) = (-\underset{\sim}{r}', -t)$. Then we have the definitions

$$\mathcal{P}^1\varphi(\underset{\sim}{r},t)(\mathcal{P}^1)^{-1} = \eta_{P^1}\varphi(-\underset{\sim}{r}',t)$$

$$C\,\varphi(\underset{\sim}{r},t)\,C^{-1} = \eta_C\varphi^*(\underset{\sim}{r},t)$$

$$\mathcal{T}\,\varphi^*(\underset{\sim}{r},t)\mathcal{T}^{-1} = \eta_T^*\varphi^*(\underset{\sim}{r},-t) \text{ and complex conjugate c-numbers since T is antiunitary.}$$

Then

$$\varphi(\underset{\sim}{r},t) \xrightarrow{P^1} \eta_{P^1}(-\underset{\sim}{r}',t) \xrightarrow{C} \eta_C\eta_{P^1}\varphi^*(-\underset{\sim}{r}',t) \xrightarrow{T} \eta_T^*\eta_C^*\eta_{P^1}^*\varphi^*(-\underset{\sim}{r}',-t) \quad .$$

But

$$\varphi(\underset{\sim}{r},t) \xrightarrow{SR} \varphi(-\underset{\sim}{r}',-t)$$

and

$$\varphi(-\underset{\sim}{r}',-t) \xrightarrow{HC} \varphi^*(-\underset{\sim}{r}',-t) \quad .$$

Thus with a choice of phase factors such that $\eta_T^*\eta_C^*\eta_{P^1}^* = 1$ we have $TCP^1 = (HC)(SR)$ so that the theorem holds for scalar fields.

D. Jost CP^1T for Scalar Fields, All S

We recall his arguments[16] with slight modifications. Weak local commutativity (WLC), in a real neighbourhood of a Jost point expressed in terms of W distributions is

$$\langle 0|\varphi(x_1)\varphi(x_2)\ldots\varphi(x_n)|0\rangle = \langle 0|\varphi(x_n)\ldots\varphi(x_2)\varphi(x_1)|0\rangle \qquad (1)$$

or

$$W(\xi_1,\xi_2,\ldots\xi_{n-1}) = W(-\xi_{n-1},\ldots -\xi_2,-\xi_1). \qquad (2)$$

(The different scalars φ can represent different fields). This relation can be analytically continued from the real neighbourhood to a complex neighbourhood and hence to the extended tube $\mathcal{T}'_{n-1}$ in the form

$$W(\zeta_1,\zeta_2,\ldots\zeta_{n-1}) = W(-\zeta_{n-1},\ldots-\zeta_2,-\zeta_1) \quad . \qquad (3)$$

Using $L_+(C)$ invariance and the fact that $P^1 T \in L_+(C)$ we get

$$W(\zeta_1, \zeta_2, \ldots \zeta_{n-1}) = W(\zeta'_{n-1}, \ldots \zeta'_2, \zeta'_1) \tag{4}$$

where

$$\zeta_k^{,\mu} = \tau^{\mu} \zeta_k^{\mu}, \quad \tau^{\mu} = \begin{cases} +1 & \text{for } \mu = 0,1 \\ -1 & \text{otherwise.} \end{cases}$$

Now if $(\zeta_1, \zeta_2, \ldots \zeta_{n-1}) \in \mathfrak{T}_{n-1}$, then $(\zeta_{n-1}, \ldots \zeta_2, \zeta_1)$ also $\in \mathfrak{T}_{n-1}$ and $(\zeta'_{n-1}, \ldots \zeta'_2, \zeta'_1)$ also $\in \mathfrak{T}_{n-1}$, where, in the latter statement we have used the fact that $\mathfrak{T}_{n-1}$ is invariant under reflections of any number of space coordinates. Taking real boundary values $\zeta_k \to \xi_k$, $\zeta'_k \to \xi'_k$ in (4), we get

$$W(\xi_1, \xi_2, \ldots \xi_{n-1}) = W(\xi'_{n-1}, \xi'_{n-2}, \ldots \xi'_2, \xi'_1). \tag{5}$$

Equivalently,

$$\langle 0 | \varphi(x_1) \ldots \varphi(x_n) | 0 \rangle = \langle 0 | \varphi(-x'_n) \ldots \varphi(-x'_2)\, \varphi(-x'_1) | 0 \rangle \tag{6}$$

$$= \langle 0 | \varphi^*(-x'_1) \varphi^*(-x'_2) \ldots \varphi^*(-x'_n) | 0 \rangle^* \tag{7}$$

and Eq. (7) is recognized as the symmetry property that corresponds to an antiunitary involution $\theta = CP^1 T$. The result we have for W distributions is also a result for the field theory since by the reconstruction theorem the W distributions together with the physical postulates completely determine the theory.

E. The Case of Arbitrary Spinor Fields

Let us look at the Dirac equation for S = 3. According to Lüders' definition of C,T the free field Dirac equation is C,T invariant if there exists unitary matrices C,T such that

$$C \gamma_\mu C^{-1} = -\gamma_\mu^{\,T} \tag{8}$$

$$T \gamma_\mu T^{-1} = \gamma_\mu^{\,T}. \tag{9}$$

For S = 3, in a representation where γ_0 is diagonal we can take $(\gamma_0, \gamma_1, \gamma_2) = (\sigma_3, i\sigma_2, -i\sigma_1)$ where the σ's are in the standard Pauli representation. The only other linearly independent matrix is σ_0 and we see that Eq. (8) is satisfied by taking $C = \sigma_2$, but there is no solution of Eq. (9). So the Dirac equation is C invariant but not T invariant.[17] The P transformation is

$$P\,\psi(\underset{\sim}{r},t)\,P^{-1} = \eta_P''\,\gamma_0\,\psi(-\underset{\sim}{r},t)$$

so that we should analogously define P^1 by

$$P^1\,\psi(\underset{\sim}{r},t)(P^1)^{-1} = \eta_{P^1}''P^1\psi(-\underset{\sim}{r}',t)$$

where we would like to have a matrix P^1 satisfying

$$(P^1)^{-1}\,\gamma_\mu P^1 = \begin{cases}\gamma_\mu & \mu \neq 1\\ -\gamma_\mu & \mu = 1\end{cases} \quad . \tag{10}$$

We can take $P^1 = \gamma_0\gamma_5$ for S = even and then we would have P^1 invariance. But for S = 3 there is no matrix with the property $\{\gamma_\mu,\gamma_5\} = 0$ of γ_5 and (10) cannot be satisfied. Thus there is no P^1 invariance in the above sense.

This does not however exclude the possibility of the existence of an antiunitary inversion θ of the nature of CP^1T. The moral is that one must look for CP^1T rather than CPT with suitable definitions of these transformations.

V. Remarks on Primitive Domain Boundaries

In this section we consider the boundaries of the extended tube $\mathfrak{T}_n'$ in terms of the image $\mathfrak{m}_n = Z\,\mathfrak{T}_n'$ of $\mathfrak{T}_n'$ under the mapping

$$Z:\{\zeta_1,\ldots\zeta_n\}\to\{Z_{ij} = (\zeta_i,\zeta_j),\ i,j = 1,\ldots n\}$$

of vectors into L_+ invariants. $\mathfrak{m}_n$ is called a primitive domain because local commutativity has not yet been used, and is not used in this section.

It is often stated, though not in print, that in 4-spacetime all n-point function primitive domain boundaries, for $n > 5$, are of 5-point type. We note that the extended tube $\mathfrak{T}_4$ for the 5-point function has 4 difference vectors and in 4 dimensions any larger number of vectors is linearly dependent. Hence this depends on what one means by 5-point type. We shall work in terms of a reasonable definition[18] in terms of natural stability properties under complex Lorentz transformations and show that the belief is not true with such a definition.

To simplify the discussion we take an example here of the 5-point function (n = 5) in 3-dimensional spacetime (S = 3) and show that in addition to the C,S,F and DANAD hypersurface boundaries for $\mathfrak{m}_5$, there is one more.[19] The case S = 4 can also be treated somewhat similarly.[20] We shall also assume invariance under L_+ rather only $L_+^\uparrow$ for simplicity.

One begins with the non-trivial result[18] that $\partial \mathfrak{M}_n = Z(\partial \mathfrak{T}_n') = Z\, B_n$ where $B_n = (\partial \mathfrak{T}_n') \cap (\partial\, \mathfrak{T}_n)$ and classifies points of $\partial \mathfrak{M}_n$ by splitting B_n into disjoint sets: $B_n = \bigcup_{m=1}^{n} B_n^{\,m}$ where $B_n^{\,m} \cap B_n^{\,m'} = \phi$ for $m \neq m'$.

Definition: A point $P \equiv (\zeta_1, \ldots \zeta_n) \in B_n^{\,m}$ if and only if

a) $P \in \partial \mathfrak{T}_n$

b) For some m element subset $\zeta_{\sigma(1)}, \ldots \zeta_{\sigma(m)}$ of $\zeta_1, \ldots \zeta_n$ there exists no $\Lambda \in L_+(C)$ such that $(\Lambda\zeta_{\sigma(1)}, \ldots \Lambda\zeta_{\sigma(m)}) \in \mathfrak{T}_m$ whereas

c) For every $(m-1)$ element subset $\zeta_{\rho(1)}, \ldots \zeta_{\rho(m-1)}$ of $\zeta_1, \ldots \zeta_n$ there exists a $\Lambda \in L_+(C)$ such that $(\Lambda\zeta_{\rho(1)}, \ldots \Lambda\zeta_{\rho(m-1)}) \in \mathfrak{T}_{m-1}$.

Theorem III: The set $B_4^{\,4}$, for $S = 3$, is non empty.

Proof: Table II gives a point in it (P say). It is necessary to check that every 3 element subset of the $\xi_i \wedge \eta_i$ $i = 1, \ldots 4$ is linearly independent (see column 4) and that $\exists\ \lambda_i > 0$ such that $\sum_{i=1}^{4} \lambda_i \xi_i \wedge \eta_i = 0$. ∎

i	ξ_i	η_i	$\xi_i \wedge \eta_i$	λ_i
1	(0,1,-1)	(t,1,1)	(-t,t,2)	1
2	(0,1,1)	(t,1,-1)	(-t,-t,-2)	1
3	(0,-1,1)	(t,-1,-1)	(t,t,-2)	1
4	(0,-1,-1)	(t,-1,1)	(t,-t,2)	1

Table II. A point of $B_4^{\,4}$ $(t = \sqrt{2})$

One has to show that a neighbourhood restricted to $B_4^{\,4}$, of P maps into a hypersurface piece on $\partial \mathfrak{M}_4$. This can be done.[19),20)] We only sketch the arguments here. $\mathfrak{M}_4$ in a neighbourhood of $ZP = Q$ (say) is a space of 9 complex variables and thus of (real) dimension 18.

The question as to whether the piece has the (real) dimension 17 to be a hypersurface or not can be reduced to whether the rank of a (17 x 18) Jacobian matrix J at P is maximal or not. Deleting the last column one gets a determinant $|\det J_{17}| = 8589934592$. [This can be evaluated using a standard Gauss elimination or pivotal condensation process combined with the fact that P and J have been chosen to give an integer as the result.] This is recognized to be not such an

unfamiliar number after all; in fact it is 2^{33}. Since this is a non-zero number one concludes that we have indeed an extra hypersurface.

Conclusion

Thus we have made an excursion through different dimensions of spacetime studying what analyticity isn't there in some cases, what analyticity there appears not to be but is actually there in other cases, a possible way out if it turns out that CPT is not an exact conservation law of Nature, and some folklore to think about as you roam around the ghost towns of the Colorado mountains.

References

1. R. F. Streater and A. S. Wightman, PCT Spin and Statistics and All That (W. A. Benjamin, Inc., New York, 1964).
2. R. Jost, The General Theory of Quantized Fields (American Mathematical Society, 1965).
3. A. S. Wightman and L. Gårding, Arkiv för Fys. 28, 129 (1964).
4. A. C. Manoharan, J. Math. Phys. 8, 2026 (196).
5. G. Källén and A. S. Wightman, Mat. Fys. Skr. Dan. Vid. Selsk. 1, No. 6 (1958).
6. D. Ruelle, Helv. Phys. Acta 34, 587 (1961).
7. R. F. Streater, Proc. Roy. Soc. A256, 39 (1960) and Nuovo Cimento 15, 937 (1960); R. B. Lavine, Senior Thesis, Princeton University (1961); S. M. Gersten, Senior Thesis, Princeton University (1961).
8. A. C. Manoharan, in Proceedings of the Indiana Symposium on Complex Analytic Methods in Mathematical Physics (to appear); Bull. Am. Phys. Soc. 13, No. 4, 686 (1968).
9. A. S. Wightman, in Dispersion Relations and Elementary Particles (Les Houches lectures) (John Wiley and Sons, Inc., New York, 1960), p. 229.
10. W. S. Brown, Ph.D. Thesis, Princeton University (1961).
11. J. Bros, H. Epstein and V. Glaser, Nuovo Cimento 31, 1265 (1964).
12. A. C. Wu, Phys. Rev. 135 B, 222, 230 (1964).
13. M. Minami and N. Nakanishi, Kyoto University preprint (1968).
14. G. Lüders, Ann. Phys. 2, 1 (1957).
15. W. Pauli, in Niels Bohr and the Development of Physics, W. Pauli, ed. (McGraw-Hill, New York, 1955).
16. R. Jost, Helv. Phys. Acta 30, 409 (1957).
17. S. P. Rosen, Purdue University preprint (1968).
18. A. S. Wightman, J. Ind. Math. Soc., 24, 625 (1960).
19. A. C. Manoharan, Ph.D. Thesis, Brandeis University (1963).
20. A. C. Manoharan, Nuovo Cimento 36, 1283 (1965).

EXACT SOLUTIONS FOR SPIN AMPLITUDES IN EXTERNAL OSCILLATING FIELDS†‡

Neil Ashby
Department of Physics and Astrophysics
University of Colorado
Boulder, Colorado

I. Introduction

In these lectures we shall give a method for exact solution of the equations of motion for the spin amplitudes of a particle interacting with an external magnetic field having an arbitrarily large oscillating component. The development of this method occurred as an outgrowth of the study of propagation of electromagnetic waves through a gas of spins whose energy levels are split by a weak static field. The method is quite general and should have many interesting applications to spin resonance problems, as for example one can study shifts in resonance frequencies as a function of angle of orientation, polarization or amplitude of the applied field. A method for exact solution of a time-dependent Schrödinger equation is also of considerable interest for its own sake.

For the sake of simplicity we shall restrict the present discussion to the cases of spin $\frac{1}{2}$ and spin 1 particles interacting with the applied field $\vec{B}(t)$ by means of their magnetic moments

$$\vec{\mu}_{op} = \gamma \vec{s} \tag{1}$$

where $\vec{s}$ is the spin angular momentum operator and γ is the gyromagnetic ratio. The Hamiltonian of the problem is then

$$H = H(t) = -\vec{\mu}_{op} \cdot \vec{B}(t) \quad . \tag{2}$$

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.

‡This work supported in part by Office of Aerospace Research, under contract AF33(615)1652, Applied Mathematics Division, Wright-Patterson Air Force Base, Ohio, through PEC Research Associates, Boulder, Colorado 80302.

We shall--again for simplicity--assume $\vec{B}$ contains oscillating components of a single fundamental frequency ω,

$$B_z = B + b(e^{-i\omega t} + e^{i\omega t})$$
$$B_x + iB_y = b^+ e^{i\omega t} + b^- e^{-i\omega t}$$
$$B_x - iBy = b^{+*} e^{-i\omega t} + b^{-*} e^{i\omega t} \qquad (3)$$

where B, b are real and $B > 0$. The static part of $\vec{B}$ thus determines our choice of z-direction. A fundamental property of the magnetic field is its periodicity in time: $\vec{B}(t + 2\pi/\omega) = \vec{B}(t)$; it is upon this property that the present method of solution is based. Higher harmonics of the frequency ω may also be included in $\vec{B}$, but the analysis is more complicated.

Exact solutions for the spin amplitudes in certain special cases of the above problem are well known and can be found in texts. For example if $b = b^+ = 0$, we have the standard problem of a circularly polarized oscillating field at right angles to the static field.

We shall state the precise mathematical problem to be solved for the spin $\frac{1}{2}$ case: let $a^+(t)$, $a^-(t)$ be the spin-up, spin-down amplitudes of the particle in the Schrödinger wavefunction. Since any solution may be written as a linear combination of independent solutions, we wish to find two linearly independent solutions of the time-dependent equations,

$$i\hbar \frac{d}{dt} \begin{bmatrix} a^+ \\ a^- \end{bmatrix} = H(t) \begin{bmatrix} a^+ \\ a^- \end{bmatrix} \qquad (4)$$

or writing out the equations in component form,

$$i\dot{a}^+ = -\tfrac{1}{2}\gamma[B_z a^+ + (B_x - iB_y)\, a^-]$$
$$i\dot{a}^- = -\tfrac{1}{2}\gamma[-B_z a^- + (B_x + iB_y)\, a^+] \; . \qquad (5)$$

This is a system of linear differential equations with periodic coefficients; such systems have received considerable attention from mathematicians, and some properties of the solutions are well known. Consequently the first part of this presentation will consist of collecting in convenient form, for use in this specific problem, some of these things which are already known.[1] The principal theorem we need is the generalization of Floquet's theorem to the case of systems of linear differential equations with periodic coefficients. The method of

solution of Eqs. (5) which will be discussed here is a generalization of Hill's method of solution of a single second-order differential equation with periodic coefficients; the analysis is considerably more complicated than that required for Hill's equation.[2)]

II. Formal Properties of Exact Solutions--Floquet's Theorem

It will save writing to introduce a matrix notation. Let $a^{\pm}_{(1)}$ and $a^{\pm}_{(2)}$ be the amplitudes for the two linearly independent solutions, and introduce the "integral matrix"

$$X(t) = \begin{bmatrix} a^{+}_{(1)} & a^{+}_{(2)} \\ a^{-}_{(1)} & a^{-}_{(2)} \end{bmatrix} \tag{6}$$

in which each column comprises a single one of the independent solutions. (The results in this section may then be extended to the case of arbitrary spin by a trivial generalization.) An integral matrix which reduces to the unit matrix E at t = 0 is called a "normalized integral matrix" and is denoted by $X_E(t)$;

$$X_E(t = 0) = E. \tag{7}$$

The equations satisfied by both independent solutions may then be written compactly in matrix notation as

$$i\hbar \frac{dX}{dt} = HX \quad . \tag{8}$$

Our first step is to show that detX(t) is constant in time. Differentiating the determinant of X with respect to time, using Eq. (8) and the properties of determinants, we find

$$i\hbar \frac{d\ \det X}{dt} = (H_{11} + H_{22})\ \det X = (\text{Trace } H)\ \det X = 0,$$

since the spin matrices are traceless. Thus

$$\det X(t) = \text{const.} = \det X(t = 0) \ . \tag{9}$$

The determinant cannot vanish as the solutions would then be linearly dependent. Hence X(t) is at all times nonsingular. For the normalized integral matrix we have for all times

$$\det X_E(t) = 1 \quad . \tag{10}$$

We may next state that if $\tilde{X}(t)$ is any nonsingular matrix solution of Eq. (8), then the most general solution is of the form $X = \tilde{X}C$ where C is an arbitrary constant matrix. To show this, multiply the equation $i\hbar \frac{d\tilde{X}}{dt} = \tilde{X}$ on the right by C; then since C is constant it follows that $\tilde{X}C$ is a solution of Eq. (8). On the other hand, if X is an arbitrary solution of Eq. (8), then

$$i\hbar \frac{dX}{dt} = i\hbar \frac{d}{dt}(\tilde{X}\tilde{X}^{-1}X) = i\hbar \frac{d\tilde{X}}{dt}\tilde{X}^{-1}X + \tilde{X}\, i\hbar \frac{d}{dt}(\tilde{X}^{-1}X)$$

$$= HX + \tilde{X}\, i\hbar \frac{d}{dt}(\tilde{X}^{-1}X) \quad .$$

Hence since $\tilde{X}$ is nonsingular

$$\frac{d}{dt}(\tilde{X}^{-1}X) = 0$$

$$\tilde{X}^{-1}X = \text{const.} = C$$

and so

$$X = \tilde{X}C \quad . \tag{11}$$

From this result we can obtain the periodicity property of the solution X, for since $H(t + 2\pi/\omega) = H(t)$, if we replace t by $t + 2\pi/\omega$ in Eq. (8) we obtain

$$i\hbar \frac{dX(t + 2\pi/\omega)}{dt} = H(t)\, X(t + 2\pi/\omega)$$

and so $X(t + 2\pi/\omega)$ is an integral matrix if $X(t)$ is. But then by the result in Eq. (11),

$$X(t + 2\pi/\omega) = X(t)\, C$$

where C is a constant matrix. In particular, for the normalized integral matrix

$$X_E(t + 2\pi/\omega) = X_E(t)\, D \tag{13}$$

where the constant matrix D has the properties

$$\det D = 1 \tag{14}$$

$$D^{\dagger} = D^{-1} \quad . \tag{15}$$

To prove (14), take determinants of both sides of Eq. (13) and use Eq. (10). To prove (15), we calculate

$$i\hbar \frac{d}{dt}(X_E^\dagger X_E) = X_E^\dagger\, i\hbar \frac{dX_E}{dt} + i\hbar \frac{dX_E^\dagger}{dt} X_E$$

$$= X_E^\dagger H X_E - X_E^\dagger H^\dagger X_E = 0$$

since H is Hermitian. Thus using the initial condition, Eq. (7), we obtain

$$X_E^\dagger(t)\, X_E(t) = E \tag{16}$$

which expresses orthonormality of the two independent solutions in X_E. From Eqs. (13) and (16) it then follows that

$$E = X_E^\dagger(t + 2\pi/\omega)\, X_E(t + 2\pi/\omega) = D^\dagger X_E^\dagger(t)\, X_E(t) D = D^\dagger D$$

and thus D is unitary, establishing Eq. (15).

It is a well-known result of matrix theory that any unitary matrix may be diagonalized by a unitary transformation, and the diagonal elements so obtained have absolute magnitude unity. We may use this to construct matrix solutions satisfying Floquet's theorem: there exists a unitary matrix P such that

$$P^{-1} D P = \mathrm{diag}\left[e^{-2\pi i \mu_1/\omega},\ e^{-2\pi i \mu_2/\omega}\right] \equiv \Lambda \tag{17}$$

where Λ is the diagonalized matrix, and μ_1, μ_2 are real. P has at least the indeterminacy of an arbitrary phase factor.

We choose an integral matrix to be

$$X(t) = X_E(t) P \quad . \tag{18}$$

Then we have

$$X(t + 2\pi/\omega) = X_E(t) D P = X_E(t)\, P\, P^{-1} D P$$

$$= X(t)\, \Lambda \quad . \tag{19}$$

Thus the chosen matrix X has the property that on increasing t by $2\pi/\omega$, the effect is to multiply X(t) on the right by a diagonal matrix. The effect on a particular one of the independent solutions in X(t) is to multiply all components by the same phase factor.

The matrix Λ has unit determinant, for taking determinants of both sides of Eq. (19), and using Eq. (9), we get $\det X = \det X \det \Lambda$,

so det $\Lambda = 1$. Further, from Eq. (18) it follows that $X^\dagger X = E$, hence $|\det X|^2 = 1$ or $\det X = e^{i\varphi}$. By proper choice of phase factor in P, the integral matrix X can then be chosen such that $\det X = 1$.

Summarizing these results, we may say that there exists an integral matrix X(t) such that

$$X^\dagger X = E \tag{20a}$$

$$\det X = 1, \tag{20b}$$

$$X(t + 2\pi/\omega) = X(t)\Lambda \tag{20c}$$

where Λ is a diagonal matrix having the properties

$$\det \Lambda = 1, \tag{20d}$$

$$\Lambda_{ij} = \delta_{ij} e^{-2\pi i \mu_j/\omega} \quad . \tag{20e}$$

We next define a matrix L(t) by the equation

$$L(t) \equiv X(t) \operatorname{diag}\left[e^{i\mu_1 t}, e^{i\mu_2 t}\right] \quad . \tag{21}$$

Then making the replacement $t \to t + 2\pi/\omega$, we find

$$L(t + 2\pi/\omega) = X(t)\Lambda \operatorname{diag}\left[e^{i\mu_1 t} e^{2\pi i \mu_1/\omega}, e^{i\mu_2 t} e^{2\pi i \mu_2/\omega}\right]$$

$$= L(t) \quad . \tag{22}$$

So L(t) is periodic with period $2\pi/\omega$. Thus

$$X(t) = L(t) \operatorname{diag}\left[e^{-i\mu_1 t}, e^{-i\mu_2 t}\right] \tag{23}$$

This is the required generalization of Floquet's theorem, namely that a linearly independent solution can be chosen equal to a periodic function, multiplied by an exponential factor of the form $e^{-i\mu_i t}$. The key to actually finding the solutions lies in knowing the numerical values of the roots μ_i. Once these are known, the problem is essentially solved.

III. Solutions for Roots--Spin $\frac{1}{2}$ Case

Our task is now to derive a usable expression for the determination of the roots μ_i. We may note in passing that Eq. (20d) requires $\mu_2 = -\mu_1$, so only one such root need be found.

Eq. (20a) implies $L^{\dagger}L = E$, or extracting from this matrix equation the ith diagonal element,

$$\sum_j |L_{ij}|^2 = 1 \ .$$

Hence $|L_{ij}|^2$ is bounded from above for every i and j, so a convergent Fourier series for the periodic function $L_{ij}(t)$ exists:

$$L_{ij}(t) = \sum_{n=-\infty}^{\infty} a_{ij(n)}\, e^{-in\omega t} \ . \tag{24}$$

The resulting form for the jth column of X is therefore

$$X_{ij}(t) = \sum_{n=-\infty}^{\infty} a_{ij(n)}\, e^{-i(n\omega+\mu_j)t} \ . \tag{25}$$

In Eq. (25) as i varies we obtain expressions for the various spin amplitudes in the jth independent solution. Dropping the index j and specializing to spin $\frac{1}{2}$, we obtain

$$a^{+}(t) = \sum_{n=-\infty}^{\infty} a_n^{+}\, e^{-i(n\omega+\mu)t}$$

$$a^{-}(t) = \sum_{n=-\infty}^{\infty} a_n^{-}\, e^{-i(n\omega+\mu)t} \ . \tag{26}$$

Writing out the equations of motion,

$$\begin{aligned} i\dot{a}^{+} &= -\tfrac{1}{2}\gamma B a^{+} - \tfrac{1}{2}\gamma b(e^{i\omega t} + e^{-i\omega t})\, a^{+} - \tfrac{1}{2}\gamma(b^{+*} e^{-i\omega t} + b^{-*} e^{i\omega t})\, a^{-} \\ i\dot{a}^{-} &= +\tfrac{1}{2}\gamma B a^{-} + \tfrac{1}{2}\gamma b(e^{i\omega t} + e^{-i\omega t})\, a^{-} - \tfrac{1}{2}\gamma(b^{+}\, e^{i\omega t} + b^{-} e^{-i\omega t})\, a^{+} . \end{aligned} \tag{27}$$

Upon substituting the expressions (26) into (27) and equating coefficients of $e^{-in\omega t}$, we obtain the following two recurrence relations for the coefficients $a_n^{\pm}$:

$$\begin{aligned} (n\omega + \mu)\, a_n^{+} &= -\tfrac{1}{2}\gamma B a_n^{+} - \tfrac{1}{2}\gamma b(a_{n+1}^{+} + a_{n-1}^{+}) - \tfrac{1}{2}\gamma b^{+*} a_{n-1}^{-} - \tfrac{1}{2}\gamma b^{-*} a_{n-1}^{-} \\ (n\omega + \mu)\, a_n^{-} &= +\tfrac{1}{2}\gamma B a_n^{-} + \tfrac{1}{2}\gamma b(a_{n+1}^{-} + a_{n-1}^{-}) - \tfrac{1}{2}\gamma b^{+} a_{n+1}^{+} - \tfrac{1}{2}\gamma b^{-} a_{n-1}^{+} \ . \end{aligned} \tag{28}$$

These form an infinite set of coupled linear homogeneous equations in the unknowns $a_n^{\pm}$, and for a nontrivial solution to exist the determinant formed from the coefficients of these unknowns must vanish. It is this condition which determines the roots μ. We construct the determinant by enumerating the equations and coefficients in some arbitrarily chosen order. We write the equations in the order $n = -m$, for a_{-m}^{+}; $n = -m$, for a_{-m}^{-}; $n = -m + 1$, for a_{-m+1}^{+}; ...; $n = 0$, for a_0^{+}; $n = 0$, for a_0^{-}; ...; $n = m$ for a_m^{+}; $n = m$ for a_m^{-}. The coefficients are written in the order corresponding to $\ldots a_{-1}^{+}, a_{-1}^{-}, a_0^{+}, a_0^{-}, a_{+1}^{+}, a_{+1}^{-}, \ldots$, and we adopt the convention that when the set of equations is truncated we always do it by setting $a_n^{\pm} = 0$ for $|n| > M$ for some positive integer M. There are then $2 + 4M$ equations for $2 + 4M$ unknowns.

The matrix of coefficients may then be most compactly written in block form:

$$\mathcal{G} = \begin{bmatrix} \ddots & & & & & \\ & G_{-2} & G^{\dagger} & 0 & 0 & 0 \\ & G & G_{-1} & G^{\dagger} & 0 & 0 \\ & 0 & G & G_0 & G^{\dagger} & 0 \\ & 0 & 0 & G & G_1 & G^{\dagger} \\ & 0 & 0 & 0 & G & G_2 \\ & & & & & \ddots \end{bmatrix} \tag{29}$$

where G_n is a two-by-two matrix of the diagonal form

$$G_n = \begin{bmatrix} \mu + n\omega + \gamma B/2 & 0 \\ 0 & \mu + n\omega - \gamma B/2 \end{bmatrix} \tag{30}$$

and $G^{\dagger}$ is the Hermitian conjugate of the two-by-two matrix G:

$$G = \tfrac{1}{2}\gamma \begin{bmatrix} b & b^{+*} \\ b^{-} & -b \end{bmatrix} \tag{31}$$

In this form the determinant of the coefficients diverges, so we study instead the convergent determinant obtained by dividing through the a_n^{+} row by $n\omega - \gamma B/2$ and the a_n^{-} row by $n\omega + \gamma B/2$. This corresponds to dividing each equation in the set (29) by a constant. The resulting

determinant is denoted by $\Delta(\mu)$. We write out explicitly below the form of $\Delta(\mu)$ for the first few centrally located elements:

$$\Delta(\mu)=\det\begin{bmatrix}
\ddots & & & & & & \\
\frac{\mu-\omega+\frac{1}{2}\gamma B}{-\omega+\frac{1}{2}\gamma B} & 0 & \frac{\frac{1}{2}\gamma b}{-\omega+\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{-*}}{-\omega+\frac{\gamma B}{2}} & 0 & 0 \\
0 & \frac{\mu-\omega-\frac{1}{2}\gamma B}{-\omega-\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{+}}{-\omega-\frac{1}{2}\gamma B} & \frac{-\frac{1}{2}\gamma b}{-\omega-\frac{1}{2}\gamma B} & 0 & 0 \\
\frac{\frac{1}{2}\gamma b}{\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{+*}}{\frac{1}{2}\gamma B} & \frac{\mu+\frac{1}{2}\gamma B}{\frac{1}{2}\gamma B} & 0 & \frac{\frac{1}{2}\gamma b}{\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{-*}}{\frac{1}{2}\gamma B} \\
\frac{\frac{1}{2}\gamma b^{-}}{-\frac{1}{2}\gamma B} & \frac{-\frac{1}{2}\gamma b}{-\frac{1}{2}\gamma B} & & \frac{\mu-\frac{1}{2}\gamma B}{-\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{+}}{-\frac{1}{2}\gamma B} & \frac{-\frac{1}{2}\gamma b}{-\frac{1}{2}\gamma B} \\
0 & 0 & \frac{\frac{1}{2}\gamma b}{\omega+\frac{1}{2}\gamma B} & \frac{\frac{1}{2}\gamma b^{+*}}{\omega+\frac{1}{2}\gamma B} & \frac{\mu+\omega+\frac{1}{2}\gamma B}{\omega+\frac{1}{2}\gamma B} & 0 \\
0 & 0 & \frac{\frac{1}{2}\gamma b^{-}}{\omega-\frac{1}{2}\gamma B} & \frac{-\frac{1}{2}\gamma b}{\omega-\frac{1}{2}\gamma B} & 0 & \frac{\mu+\omega-\frac{1}{2}\gamma B}{\omega-\frac{1}{2}\gamma B} & \ddots
\end{bmatrix} \tag{32}$$

We also define a determinant denoted by $\Delta_1(\mu)$, obtained from the matrix in $\Delta(\mu)$ by dividing every row by its diagonal element so that the diagonal elements of $\Delta_1(\mu)$ are all unity. Then

$$\Delta(\mu) = \Delta_1(\mu) \lim_{M\to\infty} \prod_{n=-M}^{+M} \left[\frac{(\mu+n\omega+\frac{1}{2}\gamma B)(\mu+n\omega-\frac{1}{2}\gamma B)}{(n\omega+\frac{1}{2}\gamma B)(n\omega-\frac{1}{2}\gamma B)}\right] \tag{33}$$

The product in Eq. (33) may be evaluated using the identity

$$\sin z = z \prod_{n\neq 0} (1 + z/n\pi) \quad , \tag{34}$$

so that Eq. (33) becomes

$$\Delta(\mu) = -\Delta_1(\mu)\frac{\sin\frac{\pi}{\omega}\left(\mu+\frac{1}{2}\gamma B\right)\sin\frac{\pi}{\omega}\left(\mu-\frac{1}{2}\gamma B\right)}{\sin^2\frac{\pi\gamma B}{2\omega}} \quad . \tag{35}$$

The problem is thus reduced to the study of $\Delta_1(\mu)$. Although we know that the solutions of $\Delta(\mu) = 0$ must give real μ, we shall study the behavior of the determinant $\Delta_1(\mu)$ as a function of the complex

variable μ. We shall prove a number of important properties of which the first is

1. The value of $\Delta_1(\mu)$ is not changed if the elements are changed by making either or both of the replacements:

$$\text{a)}\quad b \to -b$$

$$\text{b)}\quad b^{\pm} \to -b^{\pm},\ b^{\pm *} \to -b^{\pm *} \quad . \tag{36}$$

This is an important lemma which expresses a symmetry property of $\Delta_1(\mu)$ occurring in spin problems. With this lemma we will be able to construct a proof that $\Delta_1(\mu)$ is absolutely convergent. The proof of (36) is by mathematical induction. We imagine the infinite determinant is truncated by setting $a_n^{\pm} = 0$, $|n| > M$, and we examine the upper left corner of the determinant so obtained. The elements are a very special case of the general form

$$\Delta_1(\mu) = \det \begin{bmatrix} 1 & f_2^- b b^{-*} + f_2^+ b b^{+*} & f_3 b & f_4 b^{-*} & 0 & 0 \\ g_1^- b b^- + g_1^+ b b^+ & 1 & g_3 b^+ & g_4 b & 0 & 0 \\ h_1 b & h_2 b^{+*} & 1 & 0 & h_5 b & h_6 b^{-*} \\ m_1 b^- & m_2 b & 0 & 1 & & \\ 0 & 0 & & & & \\ \vdots & & & & & \end{bmatrix}$$

where the functions f_i, $f_i^{\pm}$, g_i, $g_i^{\pm}$, h_i, m_i, etc., for $i = 1, \ldots \infty$ are assumed to satisfy the replacement property (36). Of course, at the very first stage many of these functions are identically zero, or do not depend on b, $b^{\pm}$ at all. (For example $f_2^{\pm} = g_1^{\pm} = 0$.) We now proceed to reduce the order of the determinant by one by subtracting multiples of the first row from rows 2, 3, 4, and then factoring out of each row the value of that row's diagonal element, which will satisfy (36). The result is

$$F' \begin{bmatrix} 1 & f_2'^{-}b^{-}+f_2'^{+}b^{+} & f_3'b & 0 & 0 & \cdots \\ g_1'^{-}b^{-*}+g_1'^{+}b^{+*} & 1 & g_3'bb^{-*} & g_4'b & g_5'b^{-*} & \cdots \\ h_1'b & h_2'bb^{-} & 1 & h_4'b^{+} & h_5'b & \cdots \\ 0 & m_2'b & m_3'b^{+*} & 1 & 0 & \\ \vdots & & & & & \end{bmatrix}$$

where it is easily seen that all the functions F', f_i', g_i', $f_i'^{\pm}$, $g_i'^{\pm}$, h_i', etc., satisfy (36). A similar reduction on rows two and three of this determinant yields a determinant of order two less than that from which we started, of the form

$$F'' \begin{bmatrix} 1 & f_2''^{-}bb^{-*}+f_2''^{+}bb^{+*} & f_3''b & f_4''b^{-*} & 0 & \cdots \\ g_1''^{-}bb^{-}+g''^{+}bb^{+} & 1 & g_3''b^{+} & g_4''b & 0 & \cdots \\ h_1''b & h_2''b^{+*} & 1 & 0 & h_5''b & \\ m_1''b^{-} & m_2''b & 0 & 1 & m_5''b^{+} & \cdots \\ 0 & 0 & & & & \\ \vdots & & & & & \end{bmatrix}$$

where all of the doubly primed functions satisfy (36). But this is of the same form as the determinant we started with. Repetition of the above sequence of steps results ultimately in reduction to a two-by-two matrix of the form

$$F''' \begin{bmatrix} 1 & f_2'''^{-}bb^{-*}+f_2'''^{+}bb^{+*} \\ g_1'''^{-}bb^{-}+g_1'''^{+}bb^{+} & 1 \end{bmatrix}$$

where the triply-primed functions satisfy (36). But this determinant obviously satisfies (36). Thus, since the property is satisfied for a determinant of order 2+4(M+1) if it is satisfied for a determinant of order 2+4M, the lemma is proved.

2. $\Delta_1(\mu)$ converges absolutely.

We begin the proof by observing that the series $\sum_{-\infty}^{\infty}[(n+z_1)(n+z_2)]^{-1}$ converges absolutely. We can then add several such series so that the sum

$$\sum_{\epsilon_1,\epsilon_2} \sum_{n=-\infty}^{\infty} [(\mu + n\omega + \epsilon_1\gamma B)(\mu + (n+1)\omega + \epsilon_2\gamma B)]^{-1} \quad ,$$

(where $\epsilon_1 = +\frac{1}{2}, -\frac{1}{2}$, $\epsilon_2 = +\frac{1}{2}, -\frac{1}{2}$), is absolutely convergent. The sum of terms

$$S_b \equiv 2b^2 + 3|b^+|^2 + 3|b^-|^2 + 2|b^+b^-|$$

certainly exists, so the sum of terms

$$\left(\frac{\gamma}{2}\right)^2 \sum_{\epsilon_1,\epsilon_2} \sum_{n=-\infty}^{\infty} S_b[(\mu + n\omega + \epsilon_1\gamma B)(\mu + (n+1)\omega + \epsilon_2\gamma B)]^{-1}$$

is absolutely convergent. So we may multiply out all these terms and rearrange them at will. A basic theorem on infinite products then allows us to state that the product

$$\overline{P_M} = \prod_{\epsilon_1,\epsilon_2} \prod_{n=-M}^{+M-1} \left\{1 + \left(\frac{\gamma}{2}\right)^2 S_b |[\mu+n\omega+\epsilon_1\gamma B)(\mu+n\omega+\omega+\epsilon_2\gamma B)]^{-1}|\right\}$$

tends to a limit as $M \to \infty$.

The lemma (36) means that the various terms in the expansion occur in such a way that <u>pairs</u> of off-diagonal elements always occur multiplied together in the expansion of the determinant. For example if the off-diagonal element $\frac{-\frac{1}{2}\gamma b}{\mu + n\omega - \frac{\gamma B}{2}}$ occurs in a term in the determinant, then the diagonal element unity two rows further down, which is in the same column, cannot appear as a multiplicative factor in the same term. Hence the denominator $\mu+n\omega+\omega-\frac{\gamma B}{2}$ <u>must</u> appear, as it occurs in all other nonvanishing terms in that row. Otherwise that term in the determinant would vanish anyway. The sum S_b constructed above is just the sum of all conceivable products of numerators satisfying Eq. (36). Thus if we examine the terms in the expansion of the product

$$P_M = \prod_{\epsilon_1,\epsilon_2} \prod_{n=-M}^{M-1} \left\{1+(b^2+b^{2}+|b^+|^2+|b^-|^2+b^{+*}b^-+b^{-*}b^++b^{+2}+b^{+*2}+b^{-2}+b^{-*2}) \right.$$

$$\left. \times \frac{(\frac{1}{2}\gamma)^2}{[(\mu+n\omega+\epsilon_1\ \gamma B)(\mu+n\omega+\omega+\epsilon_2\ \gamma B)]} \right\}$$

we see that if some of these terms are replaced by zero, and others have their signs changed, we can obtain all the terms in the expansion of $\Delta_1(\mu)^{(M)}$, which is the truncated determinant $\Delta_1(\mu)$. (Actually we have many more terms in P_M than appear in the determinant.) Hence to each term in the expansion of $\Delta_1(\mu)^{(M)}$ there corresponds, in the expansion of $\overline{P}_M$, a term of equal or greater modulus. $\Delta_1(\mu)^{(M+L)} - \Delta_1(\mu)^{(M)}$, with L an integer, represents the sum of those terms in the determinant $\Delta_1(\mu)^{(M+L)}$ which vanish when the off-diagonal elements in the rows corresponding to $a_n^{\pm}$, $|n| = M+1, \ldots M+L$ are replaced by zero, and to each of these terms there corresponds a term of equal or greater modulus in $\overline{P}_{M+L} - \overline{P}_M$. Hence

$$|\Delta_1(\mu)^{(M+L)} - \Delta_1(\mu)^{(M)}| \leq \overline{P}_{M+L} - \overline{P}_M .$$

Therefore, since $\overline{P}_M$ tends to a limit as $M \to \infty$, so also does $\Delta_1(\mu)^{(M)}$.

3. $\lim_{\mathrm{Im}\,\mu \to \pm\infty} \Delta_1(\mu) = 1$. (37)

This is fairly obvious as $\Delta_1(\mu)$ converges absolutely, in this limit

$$\Delta_1(\mu) = 1 + 0\left(\frac{1}{\mu^2}\right) \to 1.$$

4. $\Delta_1(\mu+\omega) = \Delta_1(\mu)$. (38)

This follows by shifting the infinite determinant $\Delta_1(\mu+\omega)$ down two rows and to the right two columns. Hence both $\Delta_1(\mu)$ and, from Eq. (35), $\Delta(\mu)$ are periodic functions of μ with period ω.

5. $\Delta_1(-\mu) = \Delta_1(\mu)$, for μ real. (39)

We recall that when forming $\Delta_1(\mu)^{(M)}$ from the coefficients of Eqs. (28), we divided by the product

$$R_M = \prod_{n=-M}^{+M} (\mu + n\omega + \tfrac{1}{2}\gamma B)(\mu + n\omega - \tfrac{1}{2}\gamma B)$$

$$= \left(\mu^2 - \left(\frac{\gamma B}{2}\right)^2\right) \prod_{n=1}^{M} [\mu^2 - (n\omega + \tfrac{1}{2}\gamma B)^2][\mu^2 - (n\omega - \tfrac{1}{2}\gamma B)^2]$$

which is even in μ. So for convenience we may examine the parity of the determinant $R_M \Delta_1(\mu)^{(M)}$, which is identical to $\mathfrak{C}_\downarrow^{(M)}$ of Eq. (29). We perform the following sequence of operations on this determinant:

a). Change $\mu \to -\mu$. This gives $R_M \Delta_1(-\mu)^{(M)}$.

b). Interchange rows and columns in each two-by-two block. This involves an even number of interchanges so does not change the determinant's value.

c). Interchange rows of blocks and columns of blocks (again, this is equivalent to an even number of row and column interchanges), so that the coefficients corresponding to -n stand in place of those for +n, for $|n| = 1,2,\ldots M$.

d). Replace $b^{\pm} \to -b^{\pm}$, and take the complex conjugate of the result. By (36), and since $\mathcal{G}^{(M)}$ is Hermitian, this does not change the value of the determinant.

Under these operations, every diagonal element changes sign:

$$\mu + n\omega \pm \tfrac{1}{2}\gamma B \to -\mu + n\omega \mp \tfrac{1}{2}\gamma B \to -\mu - n\omega \mp \tfrac{1}{2}\gamma B = -(\mu + n\omega \pm \tfrac{1}{2}\gamma B) \ .$$

Secondly, every element involving b is changed in sign

$$\pm\tfrac{1}{2}\gamma b \to \mp\tfrac{1}{2}\gamma b = -(\pm\tfrac{1}{2}\gamma b) \quad .$$

Thirdly, every element involving b^+ or b^- is changed in sign:

$$\frac{\gamma}{2} b^- \to \frac{\gamma}{2} b^{+*} \to \frac{\gamma}{2} b^{-*} \to -\frac{\gamma}{2} b^{-*} \to -\frac{\gamma}{2} b^-$$

and similarly for the other terms b^+, b^{-*}, b^{+*} .

Thus by all these operations every row in the determinant is changed in sign. But there are an even number of rows, hence

$$R_M \Delta_1(-\mu)^{(M)} = R_M \Delta_1(\mu)^{(M)}$$

and as $M \to \infty$, the theorem is established.

6. $\Delta_1(\mu)$ has simple poles at $\mu + n\omega \pm \frac{1}{2}\gamma B = 0$ for all n.

We may now use all of these properties to extract the μ-dependence of $\Delta_1(\mu)$. Let us define

$$f(\mu) = \Delta_1(\mu) - \sum_{n=-\infty}^{\infty} \frac{A}{\mu + n\omega - \frac{1}{2}\gamma B} - \sum_{n=-\infty}^{\infty} \frac{B}{\mu + n\omega + \frac{1}{2}\gamma B} \qquad (40)$$

where A and B are constants (i.e., not functions of μ) chosen such that $f(\mu)$ has no poles at $\mu = \pm\frac{1}{2}\gamma B$. Then since the subtracted terms are periodic in μ with period ω, $f(\mu)$ has no poles anywhere in the complex μ-plane. Since $\Delta_1(\mu)$ is an even function of μ, the subtracted terms must also be even in μ, and it is easily seen that this requires

$$A + B = 0 \quad . \qquad (41)$$

Then we may perform the summations indicated on the subtracted terms in Eq. (40) using the identity

$$\sum_{n=-\infty}^{\infty} \frac{1}{n+z} = \pi \frac{\cos \pi z}{\sin \pi z} \quad , \tag{42}$$

the result is

$$f(\mu) = \Delta_1(\mu) - \frac{A\pi}{\omega}\left[\frac{\cos\frac{\pi}{\omega}\left(\mu-\frac{\gamma B}{2}\right)}{\sin\frac{\pi}{\omega}\left(\mu-\frac{\gamma B}{2}\right)} - \frac{\cos\frac{\pi}{\omega}\left(\mu+\frac{\gamma B}{2}\right)}{\sin\frac{\pi}{\omega}\left(\mu+\frac{\gamma B}{2}\right)}\right] . \tag{43}$$

From this it follows that, upon taking the limit as Im $\mu \to \pm\infty$, from Eq. (37)

$$\lim_{\mathrm{Im}\mu\to\pm\infty} f(\mu) = 1 \tag{44}$$

According to Liouville's theorem, since $f(\mu)$ is analytic for all μ, and is obviously bounded as Im $\mu \to \pm\infty$, $f(\mu)$ is a constant. From Eq. (44), the constant is unity, so

$$\Delta_1(\mu) = 1 + \frac{\pi A}{\omega}\left[\frac{\cos\frac{\pi}{\omega}\left(\mu-\frac{1}{2}\gamma B\right)}{\sin\frac{\pi}{\omega}\left(\mu-\frac{1}{2}\gamma B\right)} - \frac{\cos\frac{\pi}{\omega}\left(\mu+\frac{\gamma B}{2}\right)}{\sin\frac{\pi}{\omega}\left(\mu+\frac{\gamma B}{2}\right)}\right]$$

$$= 1 + \frac{2\pi A}{\omega}\left[\frac{\sin \pi\gamma B/2\omega \cos \pi\gamma B/2\omega}{\sin^2\frac{\pi\mu}{\omega} - \sin^2 \pi\gamma B/2\omega}\right] . \tag{45}$$

This is an exact result which displays the complete dependence of $\Delta_1(\mu)$ on μ. In Eq. (45) we have yet to determine the constant A. If we can evaluate the determinant $\Delta_1(\mu)$ at any value of μ, then A can be found. For example setting $\mu = 0$ in Eq. (45),

$$\Delta_1(0) = 1 - \frac{2\pi A}{\omega}\frac{\cos \pi\gamma B/2\omega}{\sin \pi\gamma B/2\omega}$$

and thus for A,

$$\frac{\pi A}{\omega} = \tfrac{1}{2}[1 - \Delta_1(0)]\frac{\sin \pi\gamma B/2\omega}{\cos \pi\gamma B/2\omega} \tag{46}$$

Combining Eqs. (35), (45), and (46), we obtain after some algebra the following expression for $\Delta(\mu)$:

$$\Delta(\mu) = \Delta_1(0) - \sin^2(\mu\pi/\omega)/\sin^2\pi\gamma B/2\omega \qquad (47)$$

The roots are obtained from the simple formula $\Delta(\mu) = 0$, or

$$\sin^2\mu\pi/\omega = \Delta_1(0)\sin^2(\pi\gamma B/2\omega) \quad . \qquad (48)$$

This is the basic exact result for spin $\frac{1}{2}$. To find μ it is necessary to evaluate for a given frequency ω, field B, and coefficients b, $b^{\pm}$, one infinite determinant. Then solution of Eq. (48) gives the roots. Since the only nonvanishing elements of the determinant occur in a few rows parallel to the main diagonal, the computation of $\Delta_1(0)^{(M)}$ is not difficult; computer programs for this may be written which occupy only a few lines, even though the order of $\Delta_1(0)^{(M)}$ is chosen to be several thousand.

To take some simple examples, suppose $b = b^{\pm} = 0$. There is then no oscillating field, and $\Delta_1(0) = 1$. Then $\mu = \pm\gamma B/2$.

In the case $b = b^{+} = 0$ (circularly polarized oscillating field at right angles to static field), it is easily seen that $\Delta_1(0)$ reduces immediately to an infinite product of the following form

$$\Delta_1(0) = \prod_{n=-\infty}^{\infty}\left(1 - \frac{|\gamma b^{-}/2|^2}{(n\omega+\frac{1}{2}\gamma B)(n\omega+\omega-\frac{1}{2}\gamma B)}\right)$$

Each factor in the above product can be further factorized:

$$= \prod_{n} \frac{\left(n+\frac{1}{2}+\frac{1}{2}\sqrt{(1-\frac{\gamma B}{\omega})^2+|\frac{\gamma b^{-}}{\omega}|^2}\right)\left(n+\frac{1}{2}-\frac{1}{2}\sqrt{(1-\frac{\gamma B}{\omega})^2+|\frac{\gamma b^{-}}{\omega}|^2}\right)}{\left(n+\frac{\gamma B}{2\omega}\right)\left(n+1-\frac{\gamma B}{2\omega}\right)}$$

and using the product formula (34), this becomes

$$\Delta_1(0) = \sin^2\left(\frac{\pi}{2}+\frac{\pi}{2}\sqrt{(1-\frac{\gamma B}{\omega})^2+|\frac{\gamma b^{-}}{\omega}|^2}\right) / \sin^2\pi\gamma B/2\omega \quad .$$

Thus we find

$$\mu = \pm\tfrac{1}{2}\left(\omega + \sqrt{(\omega-\gamma B)^2+\gamma^2|b^{-}|^2}\right) \quad , \qquad (46)$$

a well-known result. No other inequivalent cases in which the exact explicit form of μ can be obtained analytically are known. If $b \neq 0$, or if b^{-} and b^{+} are both non-zero the entire infinite determinant must be evaluated numerically; if $b^{+} = b^{-}$, and $b = 0$, the determinant can

be reduced to something resembling a continued fraction. It is worth noting that the form of $\Delta_1(0)$ lends itself to scaling in the sense that the only independent parameters are $\lambda = \omega_0/\omega$, and $\beta^{\pm} = \gamma b^{\pm}/\omega$, $\beta = \gamma b/\omega$, where $\omega_0 = \gamma B$ is the resonance frequency for vanishingly small applied field.

If the applied oscillating field is small compared to the static field, only a few rows and columns in $\Delta_1(0)$ need be retained. The results will be equivalent to those obtained from perturbation theory. In such cases it is probably easier to use perturbation theory. However if the oscillating field is not small compared to the static field, it is much more efficient to calculate $\Delta_1(0)$ and use Eq. (48).

Having found the roots μ, it is a straightforward but tedious calculation to obtain all the coefficients $a_n^{\pm}$. For the spin $\frac{1}{2}$ case, it is only necessary to find one solution $[a^+(t), a^-(t)]$, as a second solution independent of the first is $[-a^{-*}(t), a^{+*}(t)]$. To prove this, take complex conjugates of the equations of motion, Eq. (4), and use the facts that the Hamiltonian is Hermitian and traceless to show that $[-a^{-*}, a^{+*}]$ is indeed a solution. Suppose these two solutions to be linearly dependent. Then there exist constants α and β such that

$$\alpha a^+ + \beta(-a^{-*}) = 0$$

$$\alpha a^- + \beta a^{+*} = 0 \ .$$

But then for nontrivial values of α and β to exist, the determinant of the coefficients of α and β has to vanish:

$$\det \begin{bmatrix} a^+ & -a^{-*} \\ a^- & a^{+*} \end{bmatrix} = 0$$

or

$$|a^+|^2 + |a^-|^2 = 0$$

an impossibility. Thus the two solutions are linearly independent.

Having found an integral matrix, the normalized integral matrix may be obtained from Eq. (18), for setting $t = 0$ we have

$$X(0) = P$$

and so

$$X_E(t) = X(t)P^{-1} = X(t)\,X^{\dagger}(0) \ . \qquad (47)$$

The solutions obtained from the normalized integral matrix are the most convenient for the expression of particular solutions with given initial conditions, e.g., as when we know a particle is in the + state at t = 0, the transition probability will be proportional to the resulting $|a^-(t)|^2$.

IV. Solutions for Roots--Spin 1 Case

The case of integral spin is sufficiently different in some of its details to require separate discussion. For the case of spin 1, we take the Hamiltonian matrix to be

$$H = -\gamma\hbar \begin{bmatrix} B_z & \frac{1}{\sqrt{2}}(B_x - iB_y) & 0 \\ \frac{1}{\sqrt{2}}(B_x + iB_y) & 0 & \frac{1}{\sqrt{2}}(B_x - iB_y) \\ 0 & \frac{1}{\sqrt{2}}(B_x + iB_y) & -B_z \end{bmatrix} \tag{48}$$

The implication of Floquet's theorem for the spin amplitudes a^+, a^o, a^- is that these amplitudes can be chosen to have the form

$$\begin{bmatrix} a^+(t) \\ a^o(t) \\ a^-(t) \end{bmatrix} = e^{-i\mu t} \begin{bmatrix} \sum_n a_n^+ e^{-in\omega t} \\ \sum_n a_n^o e^{-in\omega t} \\ \sum_n a_n^- e^{-in\omega t} \end{bmatrix} . \tag{49}$$

and we expect to find three roots for μ, corresponding to three independent solutions. Upon substituting the above assumed form into the Schrödinger equation, the following recurrence relations are obtained:

$$\begin{aligned} (n\omega+\mu)a_n^+ &= -\gamma B a_n^+ - \gamma b(a_{n+1}^+ + a_{n-1}^+) - \frac{1}{\sqrt{2}}\gamma(b^{+*}a_{n-1}^o + b^{-*}a_{n+1}^o) \\ (n\omega+\mu)a_n^o &= -\frac{1}{\sqrt{2}}\gamma(b^+a_{n+1}^+ + b^-a_{n-1}^+) - \frac{1}{\sqrt{2}}\gamma(b^{+*}a_{n-1}^- + b^{-*}a_{n+1}^-) \\ (n\omega+\mu)a_n^- &= \gamma B a_n^- + \gamma b(a_{n+1}^- + a_{n-1}^-) - \frac{1}{\sqrt{2}}\gamma(b^+a_{n+1}^o + b^-a_{n-1}^o) . \end{aligned} \tag{50}$$

We divide each of the above equations by $n\omega+\gamma B$, $n\omega$ (if $n \neq 0$) and $n\omega-\gamma B$, respectively and arrange the matrix of coefficients in order corresponding to n increasing from top to bottom and from left to right, and for a given n, +0- in order from top to bottom and from left to right. The determinant of the resulting matrix we denote by $\Delta(\mu)$. We seek solutions of $\Delta(\mu) = 0$. We define a determinant $\Delta_1(\mu)$ obtained by dividing each row of $\Delta(\mu)$ by the diagonal element in that row, so that $\Delta_1(\mu)$ has unit diagonal elements. Then the relation between $\Delta(\mu)$ and $\Delta_1(\mu)$ is

$$\Delta(\mu) = \Delta_1(\mu)\cdot\mu\cdot \prod_n \frac{\mu+n\omega+\gamma B}{n\omega+\gamma B} \prod_n \frac{\mu+n\omega-\gamma B}{n\omega-\gamma B} \cdot \prod_{n\neq 0} \frac{\mu+n\omega}{n\omega}$$

$$= -\frac{\omega}{\pi} \sin\frac{\mu\pi}{\omega} \frac{\sin\frac{\pi}{\omega}(\mu-\gamma B)\,\sin\frac{\pi}{\omega}(\mu+\gamma B)}{\sin^2 \pi\gamma B/\omega} \Delta_1(\mu) \qquad (51)$$

with the aid of Eq. (34).

We may now proceed with a methodical analysis of $\Delta_1(\mu)$ for this case just as for the spin $\frac{1}{2}$ case. Most of the details of the proofs are quite similar and will not be given here. The first nine rows and columns of $\Delta_1(\mu)$ appear as follows:

1	0	0	$\frac{\gamma b}{\mu-\omega+\gamma B}$	$\frac{\gamma b^{-*}/\sqrt{2}}{\mu-\omega+\gamma B}$	0	0	0	0
0	1	0	$\frac{\gamma b^{+}/\sqrt{2}}{\mu-\omega}$	0	$\frac{\gamma b^{-*}/\sqrt{2}}{\mu-\omega}$	0	0	0
0	0	1	0	$\frac{\gamma b^{+}/\sqrt{2}}{\mu-\omega-\gamma B}$	$\frac{-\gamma b}{\mu-\omega-\gamma B}$	0	0	0
$\frac{\gamma b}{\mu+\gamma B}$	$\frac{\gamma b^{**}/\sqrt{2}}{\mu+\gamma B}$	0	1	0	0	$\frac{\gamma b}{\mu+\gamma B}$	$\frac{\gamma b^{-*}/\sqrt{2}}{\mu+\gamma B}$	0
$\frac{\gamma b^{-}/\sqrt{2}}{\mu}$	0	$\frac{\gamma b^{+*}/\sqrt{2}}{\mu}$	0	1	0	$\frac{\gamma b^{+}/\sqrt{2}}{\mu}$	0	$\frac{\gamma b^{-*}/\sqrt{2}}{\mu}$
0	$\frac{\gamma b^{-}/\sqrt{2}}{\mu-\gamma B}$	$\frac{-\gamma b}{\mu-\gamma B}$	0	0	1	0	$\frac{\gamma b^{+}/\sqrt{2}}{\mu-\gamma B}$	$\frac{-\gamma b}{\mu-\gamma B}$
0	0	0	$\frac{\gamma b}{\mu+\omega+\gamma B}$	$\frac{\gamma b^{+*}/\sqrt{2}}{\mu+\omega+\gamma B}$	0	1	0	0
0	0	0	$\frac{\gamma b^{-}/\sqrt{2}}{\mu+\omega}$	0	$\frac{\gamma b^{+*}/\sqrt{2}}{\mu+\omega}$	0	1	0
0	0	0	0	$\frac{\gamma b^{-}/\sqrt{2}}{\mu+\omega-\gamma B}$	$\frac{-\gamma b}{\mu+\omega-\gamma B}$	0	0	1

The following essential properties of $\Delta_1(\mu)$ may then be established:

1. The value of $\Delta_1(\mu)$ is unchanged under either or both of the replacements:

$$\text{a) } b \to -b \qquad \text{b) } b^{\pm} \to -b^{\pm},\ b^{\pm *} \to -b^{\pm *} \tag{52}$$

2. $\Delta_1(\mu)$ converges absolutely.

3. $$\lim_{\mathrm{Im}\mu \to \pm\infty} \Delta_1(\mu) = 1 \tag{53}$$

4. $$\Delta_1(\mu+w) = \Delta_1(\mu) \tag{54}$$

5. $$\Delta_1(-\mu) = \Delta_1(\mu), \text{ for } \mu \text{ real}. \tag{55}$$

The proof of the invariance property (52) is again by induction, starting from an assumed form of the determinant corresponding to a truncated system of equations such that $a_n^{\pm} = a_n^{o} = 0$ for $|n| > M$. The corresponding truncated determinant is denoted by $\Delta_1(\mu)^{(M)}$. The form is (in the upper left corner)

$$\begin{vmatrix}
1 & f_2 bb^{-*} & f_3 & f_4 b & f_5 b^{-*} & 0 & \cdots \\
g_1 bb^{-} & 1 & g_3 bb^{+*} & g_4 b^{+} & 0 & g_6 b^{-*} & \\
h_1 & h_2 bb^{+} & 1 & 0 & h_5 b^{+} & h_6 b & \\
r_1 b & r_2 b^{+*} & 0 & 1 & 0 & 0 & \\
s_1 b^{-} & 0 & s_3 b^{+*} & 0 & 1 & 0 & \\
0 & t_2 b^{-} & t_3 b & 0 & 0 & 1 & \\
\vdots & & & & & &
\end{vmatrix}$$

where the functions $f_i, g_i, h_i, r_i, s_i, \ldots, i = 1,2,\ldots$ satisfy Eq. (52). Reduction of the order of this determinant by three results in a determinant with an upper left corner of the same form, multiplied by a function satisfying Eq. (52). Having proved (52), a proof of absolute convergence may then be constructed.

In the proof of evenness, Eq. (55), corresponding to step b) in the proof of Eq. (39) we interchange only the outermost two rows and columns in each three-by-three block.

6. $\Delta_1(\mu)$ has single poles at $\mu+n\omega = 0$, $\mu+n\omega \pm \gamma B = 0$, for all n. (Although at this point we must allow that some elements of $\Delta_1(\mu)$ have simple poles at $\mu+n\omega = 0$, we shall show in a moment that the structure of $\Delta_1(\mu)$ is such that there are actually no poles at these points.) Let us define the function

$$f(\mu) = \Delta_1(\mu) - \sum_n \frac{A}{\mu+n\omega} - \sum_n \frac{B}{\mu+n\omega+\gamma B} - \sum_n \frac{C}{\mu+n\omega-\gamma B} \tag{56}$$

where A, B, and C are chosen so that $f(\mu)$ has no poles at $\mu = 0$, or $\mu = \pm\gamma B$. Then since $f(\mu)$ and $\Delta_1(\mu)$ are periodic in μ with period ω, $f(\mu)$ will have no poles anywhere in the complex μ-plane. Since $f(\mu)$ is also bounded as $\mathrm{Im}\,\mu \to \pm\infty$, $f(\mu)$ must be a constant. Performing the sums in Eq. (56), we find

$$f = \Delta_1(\mu) - \frac{A\pi}{\omega}\frac{\cos \mu\pi/\omega}{\sin \mu\pi/\omega} - \frac{B\pi}{\omega}\frac{\cos(\mu+\gamma B)\pi/\omega}{\sin(\mu+\gamma B)\pi/\omega} - \frac{C\pi}{\omega}\frac{\cos(\mu-\gamma B)\pi/\omega}{\sin(\mu-\gamma B)\pi/\omega} \tag{57}$$

Since both f and $\Delta_1(\mu)$ are then even functions of μ, the remaining terms in Eq. (57) must constitute an even function of μ. After some algebra this condition reduces to the two equations

$$\begin{aligned} A\cos^2\frac{\gamma B\pi}{\omega} + B + C &= 0 \\ -A - B - C &= 0 . \end{aligned} \tag{58}$$

Hence if $\omega \neq \gamma B$, we get $A = 0$, $B + C = 0$. (If $\omega = \gamma B$, then examination of $\Delta_1(\mu)$ indicates the presence of double poles and a different analysis is necessary.)

Assuming henceforth that $\omega \neq \gamma B$, Eq. (57) becomes upon using Eq. (53):

$$\Delta_1(\mu) = \frac{\pi B}{\omega}\left[\frac{\cos(\mu+\gamma B)\pi/\omega}{\sin(\mu+\gamma B)\pi/\omega} - \frac{\cos(\mu-\gamma B)\pi/\omega}{\sin(\mu-\gamma B)\pi/\omega}\right] + 1 . \tag{59}$$

The fact $A = 0$ indicates that, although some of the elements of $\Delta_1(\mu)$ blow up at $\mu = -n\omega$, $\Delta_1(\mu)$ itself is well-behaved at such points. We may illustrate this by taking the determinant $\Delta_1(\mu)^{(1)}$ truncated by keeping only the first nine rows and columns, for $b = b^- = 0$. Evaluation yields

$$\begin{aligned} \Delta_1(\mu)^{(1)} = &\left[1 - \frac{\gamma|b^+|^2/2}{(\mu-\omega)(\mu-\gamma B)}\right]\left[1 - \frac{\gamma|b^+|^2/2}{(\mu+\omega)(\mu+\gamma B)}\right] \\ &\times\left[1 - \frac{\gamma|b^+|^2/2}{\mu}\left\{\frac{1}{\mu+\omega-\gamma B} + \frac{1}{\mu-\omega+\gamma B}\right\}\right] \end{aligned} \tag{60}$$

Thus although there are elements proportional to μ^{-1} in $\Delta_1(\mu)$, the coefficient multiplying such terms in the expansion of the determinant is proportional to μ:

$$\frac{1}{\mu+\omega-\gamma B} + \frac{1}{\mu-\omega+\gamma B} = \frac{2\mu}{\mu^2-(\omega-\gamma B)^2}$$

Hence $\Delta_1(\mu)^{(1)}$ exists as $\mu \to 0$. This happens for every n, $\mu \to -n\omega$, in $\Delta_1(\mu)$.

This means that it is not convenient, for the evaluation of the constant B in Eq. (59), to choose the particular value $\mu = 0$. Instead a better choice appears to be $\mu = \omega/2$. This yields

$$\frac{\pi B}{\omega} = \tfrac{1}{2}\left(1 - \Delta_1(\omega/2)\right) \frac{\cos\pi\gamma B/\omega}{\sin\pi\gamma B/\omega} \quad . \tag{61}$$

Combining Eqs. (61), (59), and (51), we find

$$\Delta(\mu) = -\frac{\omega}{\pi} \sin\mu\pi/\omega \left[\Delta_1\left(\frac{\omega}{2}\right) \cos^2\frac{\pi\gamma B}{\omega} - \cos^2\frac{\pi\mu}{\omega}\right] / \sin^2\frac{\pi\gamma B}{\omega} \tag{62}$$

and $\Delta(\mu) = 0$ yields the roots

$$\mu = 0,$$

$$\cos^2\pi\mu/\omega = \cos^2\frac{\pi\gamma B}{\omega} \; \Delta_1\left(\frac{\omega}{2}\right) . \tag{63}$$

$\mu = 0$ is always an exact root. The other two roots are opposite in sign.

The above arguments can be extended to higher spins. For spin 3/2 and spin 1, two undetermined constants are involved and two determinants must be evaluated.

References

1. N. P. Erugin, Linear Systems of Ordinary Differential Equations (Academic Press, New York, 1966).
2. E. T. Whittaker and G. N. Watson, Modern Analysis, 4th ed. (Cambridge, 1950), p. 413-417.

HIGH ENERGY SCATTERING IN THE QUARK MODEL†

David R. Harrington
and
Antonio Pagnamenta‡
Department of Physics
Rutgers University
New Brunswick, New Jersey

I. Introduction

The existence or nonexistence of the quarks as particles is one of the most exciting questions of present day high energy physics. There have been other great ideas in physics which had to be abandoned eventually.

Phlogiston--Ether--Quarks (?)

Do the quarks belong to this fatal series? We recall that the ether, before it was abolished,had to have the most surprising properties.

There are really two ways in which we can attack this question seriously. Firstly we can wait until larger accelerators become available; this will take at least a few years. In the meantime we may pursue the second route, namely pursue further the kind of analysis which has led to the question of quarks in the first place. Let us assume their existence, in some form, and ask what observable consequences they would have observable at the energies available today. There seems to be one fundamental difference between the first and second method which we should emphasize at the outset. The accelerator will presumably look for free quarks while at present we have to restrict ourselves to bound quarks. The two kinds may have quite different properties due to the large binding energy involved. One might even go to the extreme situation and consider the possibility that only bound quarks exist but not free quarks.

Present evidence for the quark model is based on quite weak statements like mass formulae, ratios of magnetic moments and sum rules for total cross sections. While the total cross sections seem

---

†Presented at the THEORETICAL PHYSICS INSTITUTE, University of Colorado, Summer 1968.
‡Reporting.

to flatten out at high energies one has found recently that the differential cross sections show considerable structure. Here we show that if appears possible to explain this structure also on the basis of the quark model.

It seems to be generally true both in hadron-nucleus scattering and (as we show below) in hadron-hadron scattering in the quark model, that the (N+1)th order scattering amplitude is much smaller than the Nth order scattering amplitude near the forward direction, but decreases more slowly with increasing momentum transfer, eventually becoming the larger of the two. One therefore expects, and in most cases finds, the differential cross section to be divided into regions of increasingly gradual decrease, with perhaps some additional structure in the region where the slope changes due to interference between two amplitudes.

Here we present a more quantitative discussion of multiple-scattering effects in the quark model using the high energy multiple scattering expansion developed by Glauber. We are of course hampered by the absence of any direct information on the quark-quark scattering amplitude and are forced to guess its properties from the shape of the near-forward differential cross sections, the total cross sections and the ratio of real to imaginary parts of the forward amplitudes for hadron-hadron scattering. We attempt to minimize this difficulty by fitting NN, $\bar{N}N$ and πN scattering simultaneously, reducing the ratio of the number of free parameters to amount of experimental information. We find that quite satisfactory fits are possible, and in the process obtain some information about the interaction between quarks at high energy.

After presenting the basic idea on proton deuteron scattering we begin with a brief derivation of the multiple scattering expansion for the scattering of two composite particles, obtaining the scattering amplitudes in terms of quark-quark scattering amplitudes and the generalized form factors of the composite hadrons. This derivation is a straightforward application of the Glauber method. We assume simple forms for the quark-quark amplitudes and the pion and nucleon form factors and obtain fits to the experimental cross sections. The parameters found are used to estimate the corrections to total cross sections. We conclude with a discussion of our results and suggest additional experimental and theoretical tests for the quark model.

More details and a more complete set of references can be found in a published article.[1)]

II. Pd Scattering

In order to introduce in a simple way the mechanism we use to study the scattering of two hadronic quark clusters and to show that

the secondary slopes in the differential cross sections are indicative of a substructure we will expose first the main ideas on the well-accepted Glauber model[2] for high energy proton deuteron or pion deuteron scattering.

At high energies where the wavelength of the incoming particle is much shorter than the separation of the P and the N in the deuteron one might take the Pd amplitude as the sum of the PN and the PP amplitudes which leads to the linear sum rule

$$\sigma(\mathrm{Pd}) = \sigma(\mathrm{PP}) + \sigma(\mathrm{PN}) \quad . \tag{II.1}$$

It was realized quickly that this leads to too large a Pd cross section. Experiment: $\sigma(\mathrm{PN}) = 42$ mb, $\sigma(\mathrm{PP}) = 48$ mb, while $\sigma(\mathrm{Pd}) = 82$ mb.

Glauber[2] showed that an eclipse or double scattering term had to be included and that this term which can be obtained from the assumption of additivity of the phase shifts would give considerable improvement. The exact expression for the Pd scattering amplitude which can be found in the papers of Glauber[2] and Franco[3] and which we will derive later in this lecture is

$$F(\mathrm{Pd},\vec{k},\vec{\Delta}) = f(\mathrm{PP},\vec{\Delta})S_d(\vec{\Delta}) + f(\mathrm{PN},\vec{\Delta})S_d(\vec{\Delta}) + \frac{2i}{4\pi}\int d^2\delta\ f(\mathrm{PP},\tfrac{\Delta}{2}+\delta)\ S_d(\tfrac{\Delta}{2}+\delta,\tfrac{\Delta}{2}-\delta)\ f(\mathrm{PN},\tfrac{\Delta}{2}-\delta). \tag{II.2}$$

Here the f are the individual nucleon-nucleon scattering amplitudes and S_d is the deuteron strong interaction form factor. Eq. (II.2) is represented in Fig. 1.



Fig. 1. Single and double scattering terms in the Pd amplitude.

Eq. (II.2) has important consequences both for the total and the differential cross section. The linear sum rule (II.1) is considerably improved by adding the double scattering correction which follows from (II.2). Using the optical theorem

$$\sigma = 4\pi\ \mathrm{Im}\ F(0)$$

and $S(0) = 1$, as well as $S(\delta, -\delta) = S(2\delta)$ we find

$$F(Pd,0) = f(PP,0) + f(PN,0) + \frac{2i}{4\pi}\int d^2\delta\ f(PP,\delta)\ f(PN,\delta)\ S_d(2\delta) \qquad (II.3)$$

or

$$\sigma(Pd) = \sigma(PP) + \sigma(PN) + 2Re\int d^2\delta\ f(PP,\delta)\ f(PN,\delta)\ S_d(2\delta)\ . \qquad (II.4)$$

Realizing that high energy amplitudes are (nearly) purely imaginary we see that the double scattering correction to the total cross section is actually negative.

In order to understand the differential cross section we reason as follows. It is only for small impact parameters, or nearly head-on collisions that we expect double scattering to be important. Therefore for large impact parameters single scattering dominates; this also corresponds, classically, to small deflections or small $|t|$ (t = momentum transfer squared). This situation has been plotted in Fig. 2a which has been adapted schematically from Franco.[3]

It is the secondary slope for large $|t|$ which is indicative of the composite structure of the deuteron. Whether there is a dip or a break in the transition region is a question of the phase difference between the single and double scattering terms and is of lesser importance here.

In Fig. 2a we have plotted a typical πp differential cross section. The similarity of the structure to the Pd cross section is striking. Whenever we see a structure in a scattering cross section we cannot avoid the impression of seeing the effect of an underlying substructure. In nuclear physics this substructure is well known; in high energy hadron physics we have to guess. We took the similarity of the two Figs. 2 as an incentive to investigate whether the quark model might provide this substructure and whether it can explain the observed structure of the differential cross sections.

III. The Multiple Scattering Expansion

In the quark model the scattering of two hadrons becomes the scattering of two composite systems, each described by an internal wave function. We assume this can be accurately described by the eikonal method, as developed by Glauber, which has been applied to elementary particle-nucleus scattering with considerable success. In this section we develop the extension of this method to the scattering of two composite systems. The analysis is the same for nucleus-nucleus scattering but, for definiteness, we shall use the language of the quark model.

We consider the process

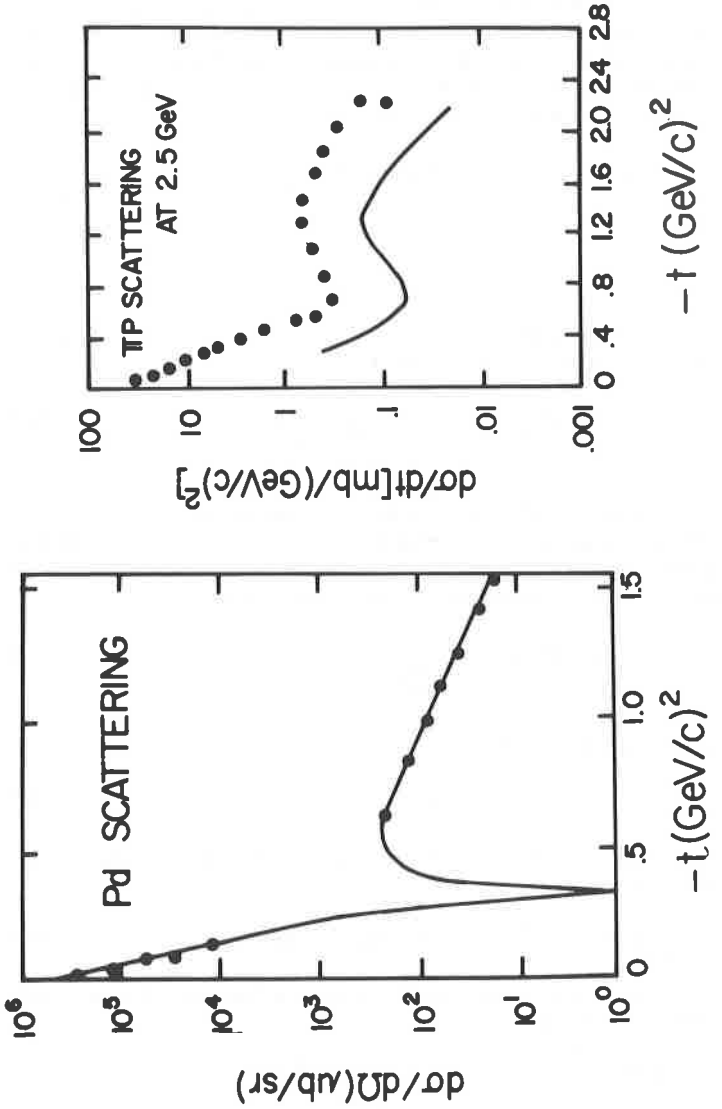


Fig. 2. Comparison of Pd scattering (2a) and πP scattering (2b).

$$A(\underline{p}) + B(-\underline{p}) \to A(\underline{p}+\underline{\Delta}) + B(\underline{p}-\underline{\Delta}) \tag{III.1}$$

where the hadrons A and B are assumed to be composed of n_A and n_B quarks, respectively, with internal wavefunctions $\psi_A(\{\underline{r}\})$ and $\psi_B(\{\underline{r}'\})$. The amplitude for this process will depend upon $f_{ii'}(\underline{\Delta})$, the amplitude for the scattering of the two free quarks i and i', for which we chose the impact parameter representation

$$f_{ii'}(\underline{\Delta}) = \pi^{-1} \int d^2 b \exp(-i\underline{\Delta}\cdot\underline{b})\, \gamma_{ii'}(\underline{b}) \ . \tag{III.2}$$

The function $\gamma_{ii'}(\underline{b})$ is given by the inverse transform

$$\gamma_{ii'}(\underline{b}) = (4\pi)^{-1} \int d^2\Delta \ \exp(i\underline{\Delta}\cdot\underline{b})\, f_{ii'}(\underline{\Delta}) \tag{III.3}$$

and is equivalent to $(2i)^{-1}[\exp(2i\delta_\ell)-1]$, where δ_ℓ is the phase shift for angular momentum $\ell \simeq pb$, so that

$$\exp[2i\delta_{ii'}] = 1 + 2i\gamma_{ii'}(\underline{b}_{ii'}) \ . \tag{III.4}$$

We assume each hadron can be reasonably well described as a composite of quarks with non-relativistic internal motions. The eikonal method in potential theory then suggest that for high energies and small angles the complete phase shift for AB scattering is simply the sum of the corresponding quark-quark phase shifts:

$$\delta_{AB}(\underline{b};\{\underline{r}\},\{\underline{r}'\}) = \sum_{ii'} \delta_{ii'}(\underline{b}_{ii'}) \ , \tag{III.5}$$

where, as shown in Fig. 3,

$$\underline{b}_{ii'}\cdot\underline{\Delta} = (\underline{b} - \underline{r}_i - \underline{r}_{i'})\cdot\underline{\Delta} \tag{III.6}$$

relates $\underline{b}_{ii'}$ to $\underline{b}$, $\underline{r}_i$, and $\underline{r}_{i'}$. If, in analogy with (III.3), we write for the full amplitude

$$F_{AB}(\underline{\Delta}) = \pi^{-1} \int d^2 b \exp(-i\underline{\Delta}\cdot\underline{b})\, \Gamma_{AB}(\underline{b}) \ , \tag{III.7}$$

then, using the internal wavefunctions to average over the positions of the quarks, we have

$$\Gamma_{AB}(\underline{b}) = \int d\tau_A \, d\tau_B \, \psi_A^{+}(\{\underline{r}_i\})\, \psi_B^{+}(\{\underline{r}_{i'}\}) \times \Gamma_{AB}(\underline{b};\{\underline{r}_i\},\{\underline{r}_{i'}\})\, \psi_A(\{\underline{r}_i\})\, \psi_B(\{\underline{r}_{i'}\}) \ , \tag{III.8}$$

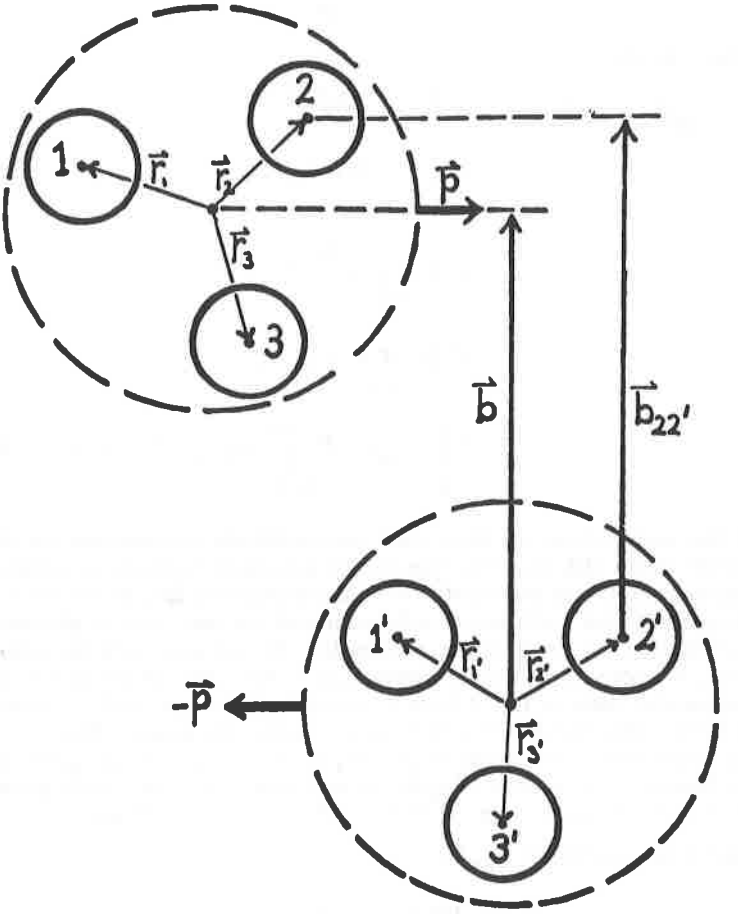


Fig. 3. The collision of two composite systems, showing the internal coordinates $\underset{\sim}{r}_i$ and $\underset{\sim}{r}'_i$ and the relation between the impact parameters $\underset{\sim}{b}$ and $\underset{\sim}{b}_{ij'}$.

and, using (III.5) and (III.4),

$$\Gamma_{AB}(\underline{b}_j,\{\underline{r}_i\},\{\underline{r}_{i'}\}) = (2i)^{-1}\left\{\prod_{ii'}[1 + 2i\gamma_{ii'}(\underline{b}_{ii'})] - 1\right\}, \qquad \text{(III.9)}$$

where in turn

$$\begin{aligned}\Gamma_{AB}(b_{ii'},\{\underline{r}_i\},\{\underline{r}_{i'}\}) &= \frac{1}{2i}\left(e^{2i\delta_{AB}} - 1\right)\\ &= \frac{1}{2i}\left(e^{2i\sum\delta_{ii'}} - 1\right)\\ &= \frac{1}{2i}\left(\prod_{ii'} e^{2i\delta_{ii'}} - 1\right)\\ &= \frac{1}{2i}\left[\prod_{ii'}(1 + 2i\gamma_{ii'}) - 1\right]\\ &= \sum_{ii'}\gamma_{ii'} + 2i\sum_{ii'jj'}{}'\gamma_{ii'jj'} + \cdots . \qquad \text{(III.10)}\end{aligned}$$

In the second line we have used the additivity assumption for the phase shifts (III.5). The sum in the exponent becomes a product of exponentials. In the fourth line we substituted (III.4) for the s-matrix element for the individual collisions and the last line is obtained by multiplying out the product, canceling the one and ordering into single-, double-, and higher-scattering terms. The prime on the sum in the second term in (III.10) is to indicate that each product occurs only once, and that the two factors are never the same. The terms in the expansion correspond to single, double, ..., and n_An_B-th order scattering, the number of n-th order scattering terms being given by the binomial coefficient $\binom{n_An_B}{n}$. With our normalization for the scattering amplitude $F_{AB}(\Delta)$,

$$\frac{d\sigma}{d\Delta^2}(AB) = \pi|F_{AB}(\Delta)|^2 \qquad \text{(III.11)}$$

and

$$\sigma(AB) = 4\pi\ \mathrm{Im}\ F_{AB}(0) . \qquad \text{(III.12)}$$

Combining (III.7), (III.8), (III.10), and (III.2) we find the following expansion for F_{AB} in powers of the quark-quark scattering amplitudes:

$$F_{AB}(\underline{\Delta}) = \sum_{ii'} \int d^2\Delta_1\ \delta(\underline{\Delta} - \underline{\Delta}_1)\, S_{A,i}(\underline{\Delta}_1)\, f_{ii'}(\underline{\Delta}_1)\, S_{B,i'}(\underline{\Delta}_1)$$

$$+ (21/4\pi) \sum_{ii',jj'}{}' \int d^2\Delta_1\, d^2\Delta_2\ \delta^2(\underline{\Delta} - \underline{\Delta}_1 - \underline{\Delta}_2)$$

$$\times S_{A,ij}(\underline{\Delta}_1, \underline{\Delta}_2)\, f_{ii'}(\underline{\Delta}_1)\, f_{jj'}(\underline{\Delta}_2)\, S_{B,i'j'}(\underline{\Delta}_1, \underline{\Delta}_2)$$

$$+ (21/4\pi)^2 \sum_{ii',jj',kk'}{}' \int d^2\Delta_1\, d^2\Delta_2\, d^2\Delta_3\ \delta^2(\underline{\Delta} - \underline{\Delta}_1 - \underline{\Delta}_2 - \underline{\Delta}_3)$$

$$\times S_{A,ijk}(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)\, f_{ii'}(\underline{\Delta}_1)\, f_{jj'}(\underline{\Delta}_2)\, f_{kk'}(\underline{\Delta}_3)\, S_{B,i'j'k'}(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)$$

$$+ \ldots \qquad \text{(III.13)}$$

Here the functions S are generalized form factors for the hadrons, determined by their internal wavefunctions. For example

$$S_{A,ijk}(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) = \int d\tau_A\, \psi_A^{+}(\{\underline{r}_i\}) \times \exp\left[-i(\underline{\Delta}_1 \cdot \underline{r}_i + \underline{\Delta}_2 \cdot \underline{r}_j + \underline{\Delta}_3 \cdot \underline{r}_k)\right] \psi_A(\{\underline{r}_i\}). \qquad \text{(III.14)}$$

It should be noted, however, that i, j, and k need not be all different, and since ψ_A is an internal wavefunction,

$$d\tau_A = \prod_{i=1}^{n_A} d^3 r_i\, \delta^3\left[n_A^{-1}\left(\sum_i \underline{r}_i\right)\right] .$$

A typical term in (III.13) is represented graphically in Fig. 4.

The application of (III.13) to hadron-hadron scattering in the quark model will differ from its application to nucleus-nucleus scattering mainly in that we have no direct information on quark-quark scattering, whereas nucleon-nucleon scattering has been studied extensively. We shall have to attempt to guess the properties of $f_{ii'}(\underline{\Delta})$ from the observed properties of hadron-hadron scattering.

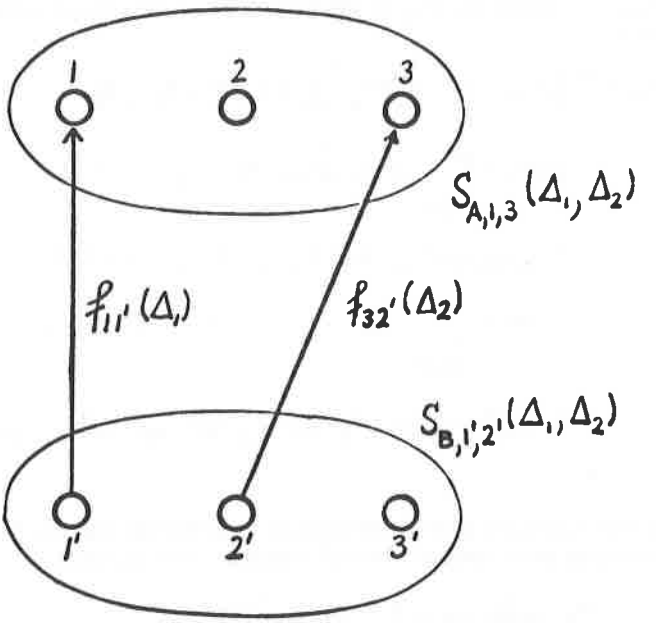


Fig. 4. Graphical representation of a double scattering term in the multiple scattering expansion (III.11), with $i = 1$, $j = 3$, $i' = 1'$ and $j' = 2'$.

We should also note here that the above formulas are true only if there is no spin or isospin flip in quark-quark scattering. They can be modified to include these effects, at least approximately, and this would certainly be necessary in order to treat such inelastic processes as $\pi + N \to \rho + \Delta$. It seems likely, however, that the probabilities for spin and isospin flip are small at high energies, and furthermore we want to avoid introducing a large number of parameters at this early stage. In the calculations below we shall therefore proceed as if the spin and isospin of the individual quarks were strictly conserved.

IV. Differential Cross Sections

In this section we want to see whether the detailed structure of the differential cross sections for hadron collisions can be explained consistently by taking into account multiple scattering effects in the

quark model if, in the absence of any direct information, we make the simplest assumptions for the forms of the quark-quark scattering amplitudes and the form factors. To reduce the ratio of the number of parameters to amount of experimental data we consider nucleon-nucleon (NN), nucleon-anti-nucleon ($\bar{N}N$), and pion-nucleon scattering simultaneously.

For NN and $\bar{N}N$ scattering we have $n_A = n_B = 3$, and therefore 9 first order, 36 second order, and 84 third order terms in the expansion (III.12). For πN scattering $n_A = 2$ and $n_B = 3$, giving 6, 15, and 20 first, second, and third order terms respectively. The nature of the various terms contributing to a given order of scattering can be quite different. To clarify this point, and as an aid to keeping the variables straight, we have introduced the graphs in Fig. 5 to represent the NN, $\bar{N}N$, and πN amplitudes. Each circle in the upper line represents a quark in hadron A, while the circles in the lower line correspond to quarks in B. The quarks and anti-quarks are represented by open and solid circles respectively. The integer in front of each graph indicates the number of equivalent terms of that type contributing to the total amplitude. A line between two quarks indicates a collision between them, transferring momentum $\underline{\Delta}_i$, and is to be associated with a scatter-amplitude $f(\Delta_i)$ (for quark-quark) or $\bar{f}(\Delta_i)$ (for quark-anti-quark). The particular version of the form factors S_A and S_B to be used can be determined from the number of lines attached to each quark and the corresponding momentum transferred to it. Note that because backward scattering is ignored no pair of quarks is connected more than once.

The explicit expressions for the hadron scattering amplitudes are found as

$$\begin{aligned} F_{NN}(\underline{\Delta}) &= 9\, f(\Delta)\, S_N^{\,2}(\underline{\Delta},0,0) \\ &+ (2i/4\pi)\int d^2\Delta_1 d^2\Delta_2\, \delta^2(\underline{\Delta}-\underline{\Delta}_1-\underline{\Delta}_2)\, f(\underline{\Delta}_1)\, f(\underline{\Delta}_2) \\ &\times [18\, S_N^{\,2}(\underline{\Delta}_1,\underline{\Delta}_2,0) + 18\, S_N(\underline{\Delta},0,0)\, S_N(\underline{\Delta}_1,\underline{\Delta}_2,0)] \\ &+ (2i/4\pi)^2 \int d^2\Delta_1\, d^2\Delta_2\, d^2\Delta_3\, \delta^2(\underline{\Delta}-\underline{\Delta}_1-\underline{\Delta}_2-\underline{\Delta}_3) f(\underline{\Delta}_1) f(\underline{\Delta}_2) f(\underline{\Delta}_3) \\ &\times [6\, S_N^{\,2}(\underline{\Delta}_1,\underline{\Delta}_2,\underline{\Delta}_3) + 36\, S_N(\underline{\Delta}_1,\underline{\Delta}_2,\underline{\Delta}_3)\, S_N(\underline{\Delta}_1+\underline{\Delta}_2,\underline{\Delta}_3,0) \\ &+ 36\, S_N(\underline{\Delta}_1+\underline{\Delta}_2,\underline{\Delta}_3,0)\, S_N(\underline{\Delta}_1,\underline{\Delta}_2+\underline{\Delta}_3,0) \\ &+ 6\, S_N(\underline{\Delta},0,0)\, S_N(\underline{\Delta}_1,\underline{\Delta}_2,\underline{\Delta}_3)] \end{aligned} \qquad \text{(IV.1)}$$

$$F_{\bar{N}N}(\underline{\Delta}) = F_{NN}(\Delta) \qquad \text{with} \qquad f(\Delta_i) \to \bar{f}(\Delta_i) \qquad \text{(IV.2)}$$

and

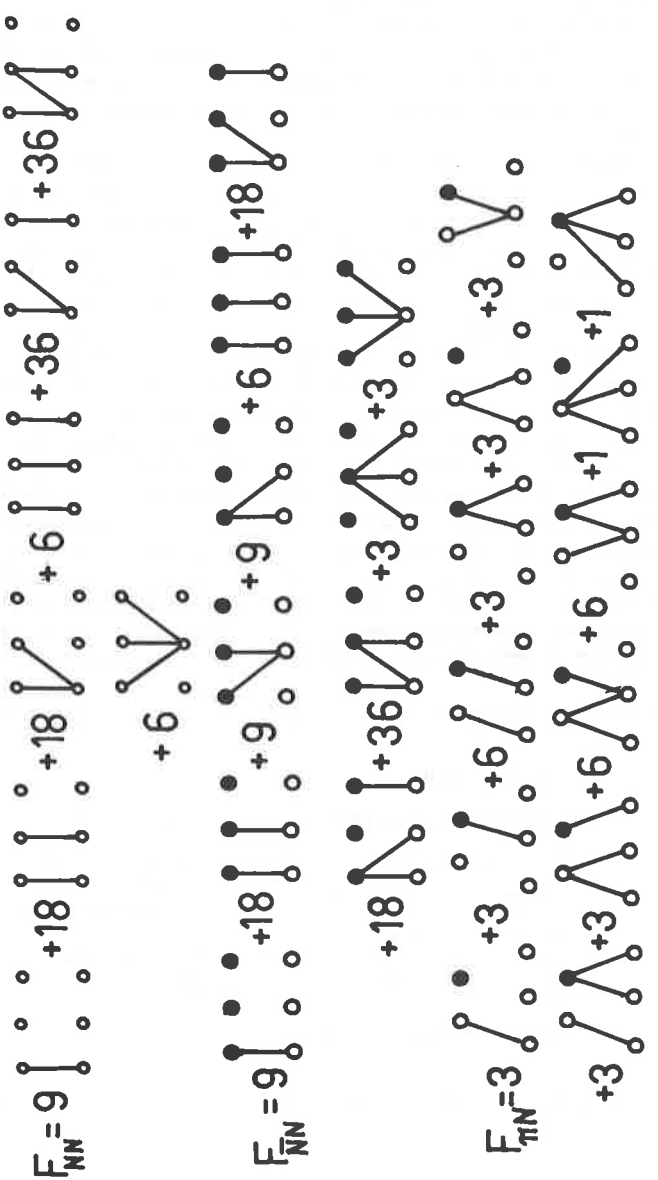


Fig. 5. Diagrams representing the multiple scattering expansions for the NN, $\bar{N}N$, and πN scattering amplitudes. The integers indicate the number of equivalent terms of a given type.

$$
\begin{aligned}
F_{\pi N}(\underline{\Delta}) &= 3[f(\underline{\Delta}) + \bar{f}(\underline{\Delta})]\, S_N(\underline{\Delta},0,0)\, S_\pi(\underline{\Delta},0) \\
&+ (2i/4\pi) \int d^2\Delta_1\, d^2\Delta_2\, \delta^2(\underline{\Delta} - \underline{\Delta}_1 - \underline{\Delta}_2) \\
&\times \{3\, f(\underline{\Delta}_1)\, \bar{f}(\underline{\Delta}_2)\, [2\, S_N(\underline{\Delta}_1, \underline{\Delta}_2, 0) + S_N(\underline{\Delta},0,0)]\, S_\pi(\underline{\Delta}_1, \underline{\Delta}_2) \\
&+ 3[f(\underline{\Delta}_1)\, f(\underline{\Delta}_2) + \bar{f}(\underline{\Delta}_1)\, \bar{f}(\underline{\Delta}_2)]\, S_N(\underline{\Delta}_1, \underline{\Delta}_2, 0)\, S_\pi(\underline{\Delta},0)\} \\
&+ (2i/4\pi)^2 \int d^2\Delta_1\, d^2\Delta_2\, d^2\Delta_3\, \delta^2(\underline{\Delta} - \underline{\Delta}_1 - \underline{\Delta}_2 - \underline{\Delta}_3) \\
&\times \{3[\bar{f}(\underline{\Delta}_1)\, f(\underline{\Delta}_2)\, f(\underline{\Delta}_3) + f(\underline{\Delta}_1)\, \bar{f}(\underline{\Delta}_2)\, \bar{f}(\underline{\Delta}_3)] \\
&\times [S_N(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) + 2S_N(\underline{\Delta}_1 + \underline{\Delta}_2, \underline{\Delta}_3, 0)] S_\pi(\underline{\Delta}_1, \underline{\Delta}_2 + \underline{\Delta}_3) \\
&+ [f(\underline{\Delta}_1)\, f(\underline{\Delta}_2)\, f(\underline{\Delta}_3) + \bar{f}(\underline{\Delta}_1)\, \bar{f}(\underline{\Delta}_2)\, \bar{f}(\underline{\Delta}_3)] \\
&\times S_N(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)\, S_\pi(\underline{\Delta},0)\} \quad . \qquad \text{(IV.3)}
\end{aligned}
$$

For convenience we use the notation

$$S_N(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) \equiv S_{N,123}(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) \qquad \text{(IV.4)}$$

and

$$S_\pi(\underline{\Delta}_1, \underline{\Delta}_2) \equiv S_{\pi,12}(\underline{\Delta}_1, \underline{\Delta}_2) \qquad \text{(IV.5)}$$

where $S_{N,123}$ and $S_{\pi,12}$ are defined by (III.13). These form factors must be completely symmetric under interchange of their arguments, and translational invariance requires that

$$S_N(\underline{\Delta}_1+\underline{q}, \underline{\Delta}_2+\underline{q}, \underline{\Delta}_3+\underline{q}) = S_N(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) \qquad \text{(IV.6)}$$

and

$$S_\pi(\underline{\Delta}_1+\underline{q}, \underline{\Delta}_2+\underline{q}) = S_\pi(\underline{\Delta}_1, \underline{\Delta}_2) \quad . \qquad \text{(IV.7)}$$

In order to test the expressions (IV.1), (IV.2), (IV.3) against the available experimental data we must choose parametrizations for the scattering amplitudes and form factors. For the former we assume

$$
\begin{aligned}
f(\Delta) &= f(0) \exp[-\tfrac{1}{8} b^2 \Delta^2] \\
\bar{f}(\Delta) &= \bar{f}(0) \exp[-\tfrac{1}{8} \bar{b}^2 \Delta^2] \quad , \qquad \text{(IV.8)}
\end{aligned}
$$

and take b^2 and $\bar{b}^2$ as well as $f(0)$ and $\bar{f}(0)$ to be complex numbers. We found that it was impossible to avoid an interference minimum in the differential cross section and at the same time match the phase for the forward proton-proton amplitude if the phase of $f(\Delta)$ were constant. Giving b^2 and $\bar{b}^2$ imaginary parts is perhaps the simplest way to allow the phases of $f(\Delta)$ and $\bar{f}(\Delta)$ to change with Δ.

For the pion and nucleon form factors we take the expressions

$$S_\pi(\underline{\Delta}_1, \underline{\Delta}_2) = \exp[-\tfrac{1}{4} a_\pi^2 (\underline{\Delta}_1 - \underline{\Delta}_2)^2] \tag{IV.9}$$

and

$$S_N(\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3) = \exp\{-\tfrac{1}{4} a_N^2 [(\underline{\Delta}_1 - \underline{\Delta}_2)^2 + (\underline{\Delta}_2 - \underline{\Delta}_3)^2 + (\underline{\Delta}_3 - \underline{\Delta}_1)^2]\} \tag{IV.10}$$

corresponding to Gaussian wavefunctions. We should point out that (IV.10) arises from a completely symmetric spatial wavefunction, while the standard version of the quark model requires complete antisymmetry and therefore a form factor which takes negative values. This point has also come up with regard to the electromagnetic form factor of the nucleon $F_N(\Delta)$. In the quark model (leaving out inessential complications)

$$F_N(\Delta) = F_q(\Delta)\, S_N(\Delta, 0, 0) \tag{IV.11}$$

where $F_q(\Delta)$ is the electromagnetic form factor of the quarks. Extensive experimental studies have revealed no zeros in $F_N(\Delta)$ up to $\Delta^2 \approx 25$ $(\mathrm{GeV}/c)^2$, implying that if $S_N(\Delta, 0, 0)$ has any zeros they must be at quite large values of Δ. Although the extrapolation to three variables is a bit risky we feel that (IV.10) is a reasonable way of parametrizing S_N when the arguments are not too large. Eqs. (IV.1), (IV.2), and (IV.3) show that for small Δ the differential cross sections go as S_N^4 which in the point quark limit is equal to F_N^4.

With the above expressions for the amplitudes and form factors all the integrals involved in (IV.1), (IV.2), and (IV.3) can be done analytically (I do not reproduce the lengthy expressions here). The differential cross sections for NN, $\bar{N}N$ and πN scattering are then determined by 10 real parameters:

$$f(0) \equiv f_R + i f_I \tag{IV.12}$$

$$\bar{f}(0) \equiv \bar{f}_R + i \bar{f}_I \tag{IV.13}$$

$$b^2 \equiv (b^2)_R + i(b^2)_I \tag{IV.14}$$

$$\bar{b}^2 \equiv (\bar{b}^2)_R + i(\bar{b}^2)_I \quad , \tag{IV.15}$$

$$a_N^2 \quad ,$$

and

$$a_\pi^2 \quad .$$

We wish to emphasize that the parameters $f(0)$, $\bar{f}(0)$, b^2 and $\bar{b}^2$ in general will be energy dependent. Judging from the behavior of hadron-hadron scattering this dependence becomes weaker as the energy increases, but at presently accessible energies cannot be ignored. This in principle limits us to treating NN and $\bar{N}N$ scattering at the same energy, and πN scattering at some definite related energy. There is unfortunately no convincing prescription relating the πN energy to the NN, $\bar{N}N$ energy, although the former should presumably be taken somewhat smaller than the latter.

Our fitting procedure goes roughly as follows: First we choose trial values for

$$R \equiv (b^2)_R / a_N^2 \tag{IV.16}$$

$(b^2)_I$ and $(\bar{b}^2)_I$. Once these are fixed we determine the remaining parameters from the experimental values for $\sigma(NN)$, $\sigma(\bar{N}N)$, $\alpha(NN)$, $\alpha(\bar{N}N)$, $B^2(NN)$, $B^2(\bar{N}N)$ and $B^2(\pi N)$, where

$$\alpha(AB) \equiv \text{Re } F_{AB}(0) / \text{Im } F_{AB}(0) \tag{IV.17}$$

and $B^2(AB)$ is defined by

$$\frac{d\sigma}{d\Delta^2}(AB) \propto \exp\left[-B^2(AB)\,\Delta^2\right] \tag{IV.18}$$

for small Δ^2. (The experimental values for $\alpha(\pi N)$ and $\sigma(\pi N)$ are of essentially no use in fixing R, $(b^2)_I$, and $(\bar{b}^2)_I$, but are reasonably well fit by the calculated values.) If multiple scattering effects were ignored the relations used would be simply

$$f(0) = (36\pi)^{-1}\,\sigma(NN)\,[i + \alpha(NN)] \tag{IV.19}$$

$$\bar{f}(0) = (36\pi)^{-1}\,\sigma(\bar{N}N)\,[i + \alpha(\bar{N}N)] \tag{IV.20}$$

$$2a_N^2 + (b^2)_R = B^2(NN) \quad , \tag{IV.21}$$

$$2a_N^2 + (\bar{b}^2)_R = B^2(\bar{N}N) \quad , \tag{IV.22}$$

and

$$a_N^2 + \tfrac{1}{2}[a_\pi^2 + (b^2)_R + (\bar{b}^2)_R] = B^2(\pi N) \ . \qquad \text{(IV.23)}$$

We begin with values determined by these relations and correct iteratively for multiple scattering effects, typical corrections being of the order of 20%. The differential cross sections are then calculated on an electronic computer using complex arithmetic and compared with experiment. By repeating this procedure for many sets of values for R, $(b^2)_I$ and $(\bar{b}^2)_I$ we determine which set gives the best overall fit to the experimental data. Below we discuss the problems associated with fitting the individual processes and present the results.

A. Proton-Proton Scattering

Since there seems to be no convenient data spanning a wide enough range of momentum transfer at a single energy, we have used an analysis given by Krisch,[4] who has shown that most of the presently available experimental points fall on a single curve when plotted against $\beta^2 p^2 \sin^2\theta$ ($\beta = v/c$), and when a reasonable adjustment for the effects of the identity of the particles is made. We have used this result to construct an "experimental" plot of $\frac{d\sigma}{d\Delta^2}$ versus Δ^2 (which we interpret as $p_\perp^2 = p^2 \sin^2\theta$) for $p_{Lab} = 15$ GeV/c. This is the highest momentum for which experimentally determined values for $\sigma(\bar{p}p)$, $\alpha(\bar{p}p)$ and $B^2(\bar{p}p)$ are available. For a given R and $(b^2)_I$ we choose the remaining parameters to reproduce $\sigma(pp) = 39$ mb, $B^2(pp) = 10(\text{GeV/c})^{-2}$ and $\alpha(pp) = -0.2$. We had to take $(b^2)_I$ in the neighborhood of -6.0 $(\text{GeV/c})^{-2}$ or $+12.0$ $(\text{GeV/c})^{-2}$ to eliminate the dip due to interference between single and double scattering, and R had to be taken in the neighborhood of 1.2 to give the correct magnitude in the triple scattering region. The weak binding or point quark limit, $R = 0$, gives a differential cross section which is very flat and too large in the triple scattering region, while the strong binding limit, $R = \infty$, gives differential cross sections fit which are too small at large Δ^2. Our most satisfactory is shown in Fig. 6. The parameters used were

$$f_I = 1.1 \ (\text{GeV/c})^{-2}$$

$$f_R = -.18 \ (\text{GeV/c})^{-2}$$

$$(b^2)_R = 3.25 \ (\text{GeV/c})^{-2}$$

$$(b^2)_I = 6.0 \ (\text{GeV/c})^{-2}$$

and

$$a_N^2 = 2.7 \ (\text{GeV/c})^{-2} \ . \qquad \text{(IV.24)}$$

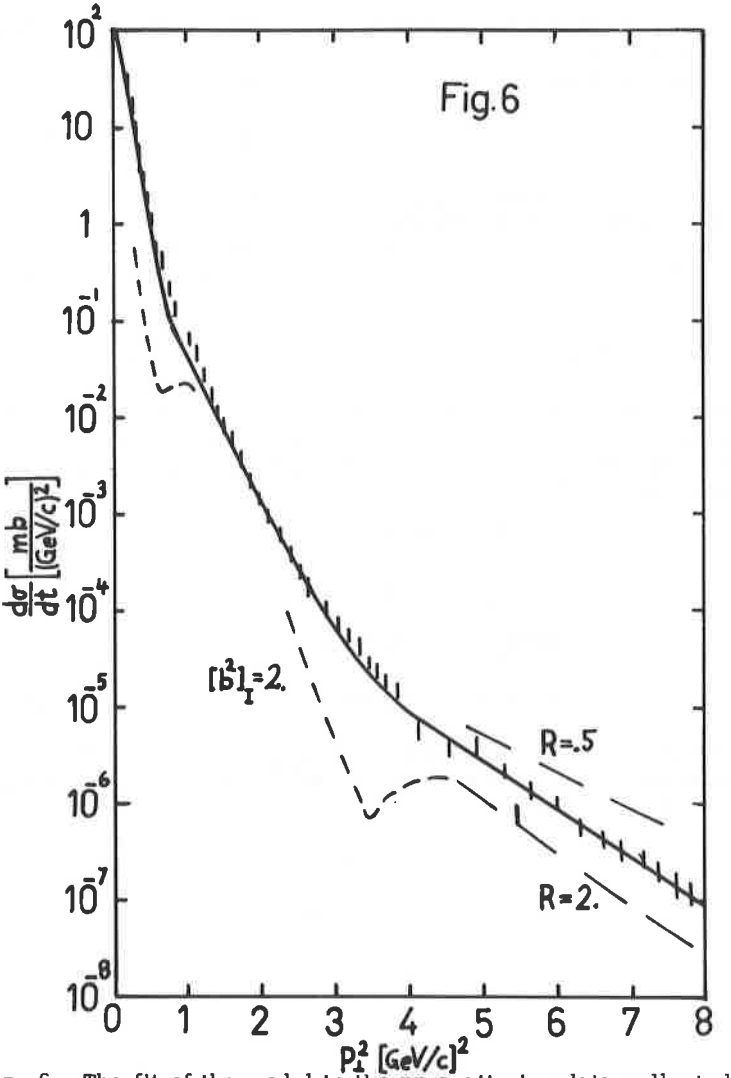


Fig. 6. The fit of the model to the pp scattering data collected and analyzed in Ref. 4. The experimental points reduced to p_{Lab}=15 GeV/c are represented by the vertical error bars. The short-dashed curve shows the influence of a variation in $(b^2)_I$, the long-dashed curves the influence of variation in $R=(b^2)_R/a_N^2$. The solid curve was drawn using the parameters given in the text.

In Fig. 7 we show the dependence of $\alpha(PP)$ on Δ^2. It is interesting to note that taking merely the Pomeranchuk trajectory in the Regge pole model would lead to a behaviour of the form

$$\alpha(PP) \sim - \cot \frac{\pi}{2} \alpha_{Pom.}(\Delta^2) \qquad \text{(IV.25)}$$

which for a slowly varying $\alpha_{Pom.}$ would reproduce our fit.

B. Proton-Antiproton Scattering

Unfortunately no large-angle data are available at 15 GeV/c, so we begin by testing our double and triple scattering contributions at lower energies where such data is available. We keep a_N^2 fixed at 2.7 $(GeV/c)^{-2}$, the value which we found to give a good fit to the proton-proton differential cross section. The parameters $\bar{f}_R$, $\bar{f}_I$, and $(\bar{b}^2)_R$ are chosen to fit $\sigma(\bar{p}p)$, $B^2(\bar{p}p)$, and $(\bar{p}p)$ respectively at the appropriate energy. This leaves only $(\bar{b}^2)_I$ with which to fit the shape of the differential cross section away from the forward direction. The most satisfactory fit at $p_{Lab} = 3.99$ GeV/c[28] is shown in Fig. 7. It is obtained with

$$\begin{aligned} f_I &= 2.40 \ (GeV/c)^{-2} \\ f_R &= -.53 \ (GeV/c)^{-2} \\ (\bar{b}^2)_R &= 5.1 \ (GeV/c)^{-2} \\ (\bar{b}^2)_I &= -1.0 \ (GeV/c)^{-2} \quad . \end{aligned} \qquad \text{(IV.26)}$$

We also present a calculation for $p_{Lab} = 6.0$ GeV/c where we use the experimental values $\sigma(\bar{p}p) = 50$ mb, $\alpha(\bar{p}p) = -0.05$ and $B^2(pp) = 9.0 \ (GeV/c)^{-2}$. We find

$$\begin{aligned} \bar{f}_I &= 1.59 \ (GeV/c)^{-2} \\ \bar{f}_R &= -.21 \ (GeV/c)^{-2} \\ (\bar{b}^2)_R &= 4.9 \ (GeV/c)^{-2} \end{aligned} \qquad \text{(IV.27)}$$

and use

$$(\bar{b}^2)_I = 0.0 \ .$$

We have not quite enough data to fix $(\bar{b}^2)_I$ at 15 GeV/c in this way. If we use a value which allows a satisfactory fit to the πp-data we predict the wide angle differential cross section shown as the dashed curve in Fig. 8, which is obtained with

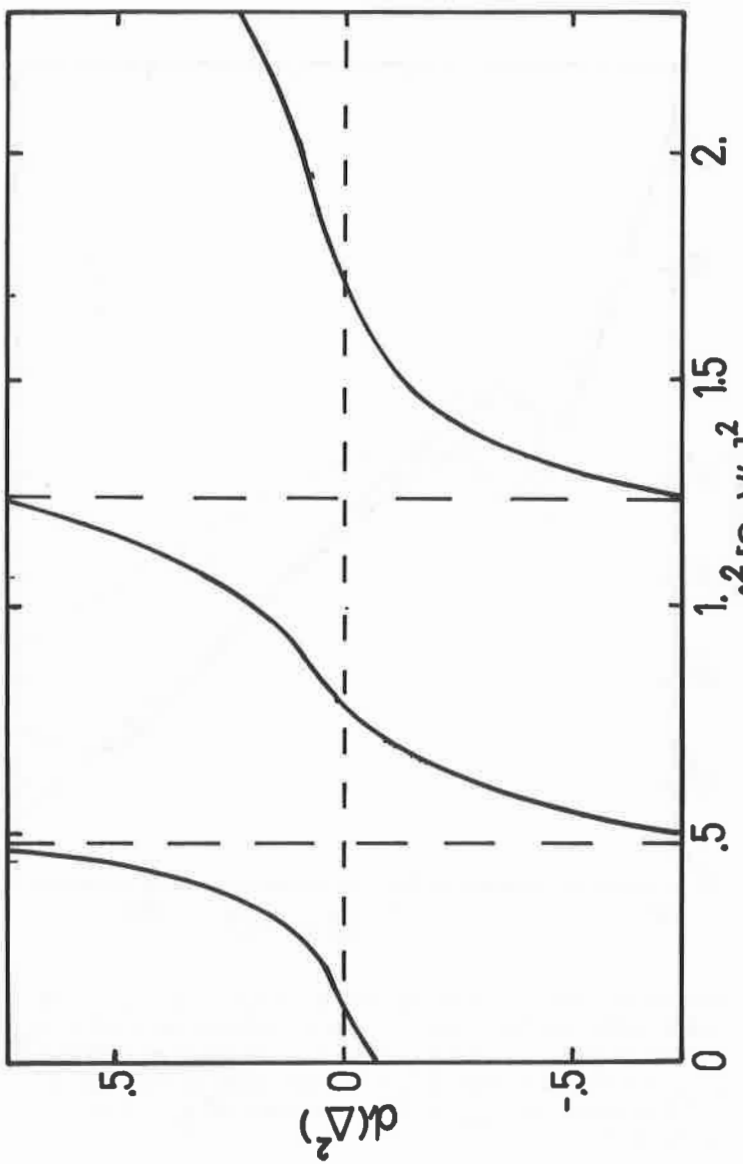


Fig. 7. The ratio α for the NN scattering amplitude as a function of $p_\perp^2$, calculated using the parameters given in the text.

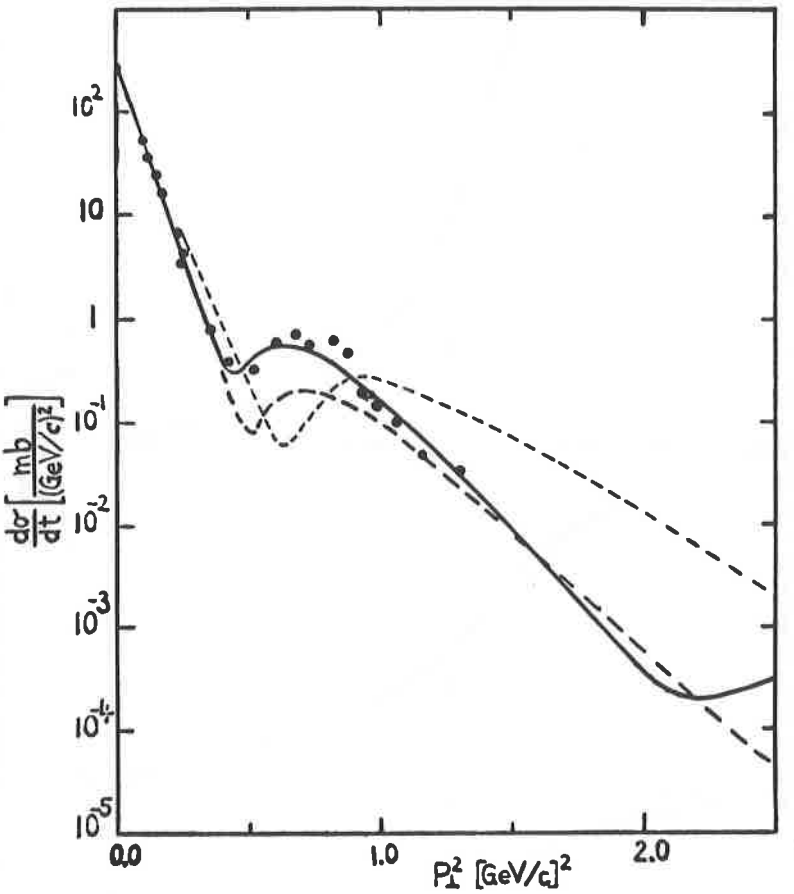


Fig. 8. The $\bar{p}p$ differential cross section as a function of $p_\perp^2$. The solid curve has been fitted to the data (solid circles) at p_{Lab} = 3.66 GeV/c. The havy dashed line shows a calculation for p_{Lab} = 6.0 GeV/c with $(b^2)_I = 0$. The light dashed curve is the prediction for p_{Lab} = 15 GeV/c using $(\bar{b}^2)_I = -3.0$ $(GeV/c)^{-2}$.

$$\bar{f}_I = 1.4\ (\mathrm{GeV/c})^{-2}$$

$$\bar{f}_R = -.47\ (\mathrm{GeV/c})^{-2}$$

$$(\bar{b}^2)_R = 1.59\ (\mathrm{GeV/c})^{-2}$$

$$(\bar{b}^2)_I = -3.0\ (\mathrm{GeV/c})^{-2} \quad . \qquad \text{(IV.28)}$$

Note that the antishrinkage which is barely noticeable in the forward peak becomes very large in the triple scattering region.

C. Pion-Proton Scattering

Having found these sets of parameters fitting proton-proton and proton-antiproton scattering at 15 GeV/c we can use them to calculate the πp differential cross section. The only remaining free parameters are $a_\pi{}^2$ (which is chosen to fit the forward slope $B^2(\pi p)$ and $(\bar{b}^2)_I$; the latter is free only to the extent that we lack wide-angle $\bar{p}p$ data. The question immediately arises of what πp energy we should try to fit, and as noted above, there seems to be no unambiguous answer (although it presumably must be less than 15 GeV/c). Fortunately $B^2(\pi N)$ is almost energy independent above 4.0 GeV/c, and gives $a_\pi{}^2 \simeq 4.5\ (\mathrm{GeV/c})^{-2}$ with only a very weak dependence on $(\bar{b}^2)_I$. Taking $(\bar{b}^2)_I = -3.0\ (\mathrm{GeV/c})^{-2}$ gives the differential cross section shown in Fig. 9, along with some experimental data at 2.5, 2.0, 6.0 GeV/c and 8 GeV/c. The plot indicates a secondary dip above $p_\perp{}^2 = 2.0\ (\mathrm{GeV/c})^2$. These values also predict a total cross section $\sigma(pp) = 31.0$ mb which is slightly too large.

Our calculations also seem to fit reasonably well some preliminary data for high $|t|$-values at 6.0 and 8.0 GeV/c.

V. Total Cross Sections

If multiple scattering effects are ignored the quark model predicts a number of "additive" or linear sum rules for total cross sections. Multiple scattering produces corrections to each total cross section and therefore to the sum rules. These corrections appear as a byproduct of our differential cross section fits; we have only to use the optical theorem III.12 and calculate $F_{AB}(0)$ using the parameters found. Since our parameters are not necessarily unique we record here the formulas for the total cross sections with our parametrization including all double scattering corrections (the triple scattering corrections are generally no more than 1%):

$$\sigma(NN) = 36\pi\, f_I + 4\pi\ \mathrm{Re}\ f^2(0)\left[18(2b^2+6a_N{}^2)^{-1} + 18(2b^2+3a_N{}^2)^{-1}\right] \qquad \text{(V.1)}$$

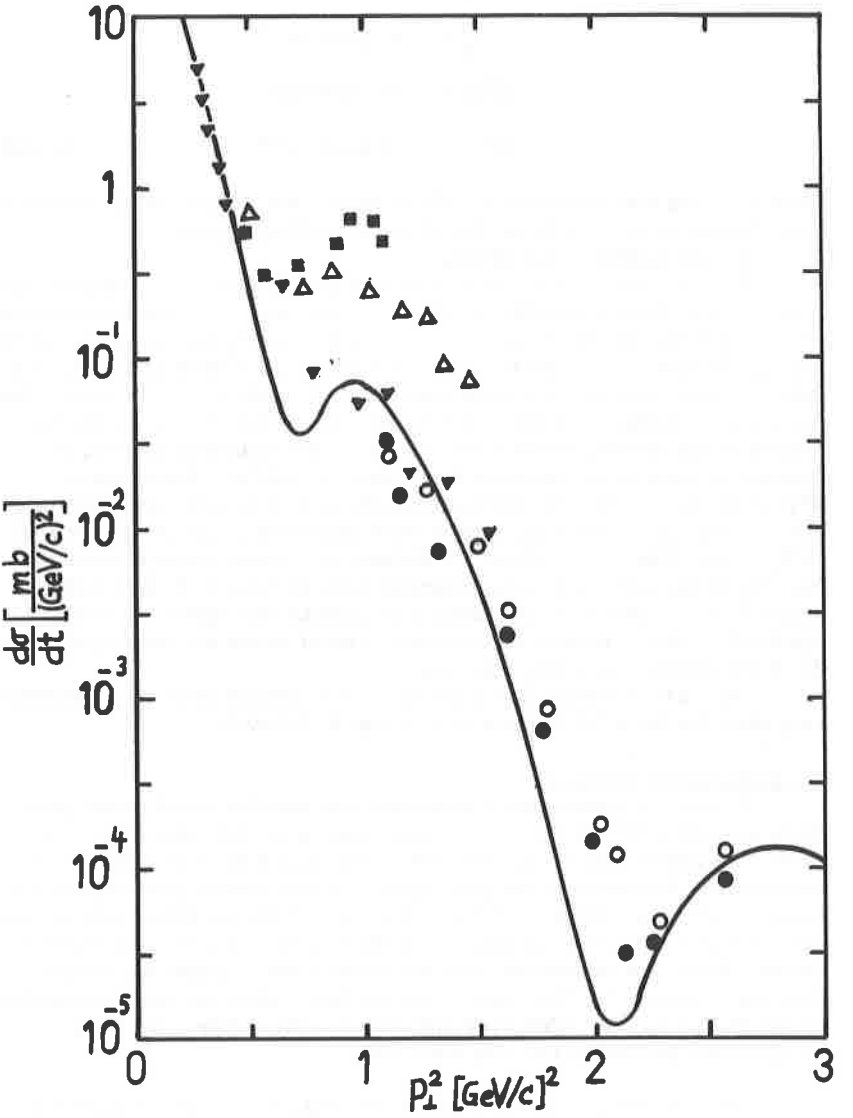


Fig. 9. (See next page for caption)

Fig. 9. The differential cross section for πp scattering as a function of $p_\perp^2$. The data are for $\pi^- p$ at 2.5 GeV/c (open triangles), at 3.0 GeV/c (squares), at 6.0 GeV/c (solid triangles) and at 8.0 GeV/c (solid circles); for $\pi^+ p$ at 8.0 GeV/c (open circles). The curve is the prediction of the model with a_π^2 taken to fit the initial slope, $(b^2)_I = 3.0$ (GeV/c)$^{-2}$ and the remaining parameters taken from pp and $p\bar{p}$ data at 15 GeV/c. It predicts a second dip above $p^2 = 2.0$ (GeV/c)^2.

$$\sigma(\bar{N}N) = 36\pi\,\bar{f}_I + 4\pi \mathrm{Re}\,\overline{f(0)}^2\left[18(2\bar{b}^2+6a_N^2)^{-1}+18(2\bar{b}^2+3a_N^2)^{-1}\right] \qquad (V.2)$$

$$\begin{aligned}\sigma(\pi N) = {} & 12\pi(f_I+\bar{f}_I) \\ & + 4\pi\,\mathrm{Re}\left[6\,f(0)\,\bar{f}(0)\,(b^2+\bar{b}^2+2a_\pi^2+3a_N^2)^{-1}\right. \\ & + 3\,f(0)^2\,(2b^2+3a_N^2)^{-1} \\ & + 3\,\overline{f(0)}^2\,(2\bar{b}^2+3a_N^2)^{-1} \\ & \left. + 3\,f(0)\overline{f(0)}(b^2+\bar{b}^2+2a_\pi^2)^{-1}\right] \qquad (V.3)\end{aligned}$$

It should be remembered that in the above formulas $f(0)^2$, $f(0)\bar{f}(0)$ and $\bar{f}(0)^2$ are close to being purely negative real numbers and b^2 and $\bar{b}^2$ have nonzero imaginary parts. Such imaginary parts generally decrease the magnitude of the double scattering corrections by increasing the magnitude of the denominators and by changing the phase of the double scattering amplitudes.

For our 15 GeV/c parameters, writing the results as the sum of single and double scattering contributions, we have

$$\sigma(NN) = 48\text{ mb} - 9\text{mb} = 39\text{ mb} \qquad (V.4)$$

and

$$\sigma(\bar{N}N) = 63\text{ mb} - 13\text{ mb} = 50\text{ mb} \qquad (V.5)$$

while for the corresponding πN energy

$$\sigma(\pi N) = 37.7\text{ mb} - 6.0\text{ mb} = 31.7\text{ mb}\,. \qquad (V.6)$$

The double scattering corrections reduce the first order total cross sections by about 20%. Our values for $\sigma(NN)$ and $\sigma(\bar{N}N)$ are taken

from experiment at about 15 GeV/c, while $\sigma(\pi)$ is calculated using the parameters which fit these 15 GeV/c NN and $\bar{N}N$ data. The value of 31.7 mb indicates that either the πN energy corresponding to 15 GeV/c laboratory momentum is very low, in the neighborhood of 3 or 4 GeV/c, or that our total cross section results are in error by a few millibarns, perhaps because we have ignored corrections such as spin and isospin dependence.

Finally, we record the quark-quark and quark-anti-quark total cross sections which we found from the 15 GeV/c parameters:

$$\sigma(qq) = 4\pi f_I = 5.4 \text{ mb} \tag{V.7}$$

and

$$\sigma(\bar{q}q) = 4\pi \bar{f}_I = 7.0 \text{ mb} \tag{V.8}$$

As we would expect from comparing the NN and $\bar{N}N$ total cross sections, $\sigma(\bar{q}q)$ is somewhat larger than $\sigma(qq)$, presumably because of the large number of annihilation channels.

VI. Processes Requiring Double Charge Exchange

Of special interest to our model are processes which require double charge or double strangeness exchange or double spin flip. There is an entire series of such reactions but we will study here only the process

$$K^- P \to \pi^+ \Sigma^- \tag{VI.1}$$

for which some experimental data are available. All these processes namely have very small cross sections and are therefore not studied frequently. They also have a large backward and a minor forward peak (for the π^+ here). The backward peak can be explained qualitatively at least as due to the exchange of a doubly charged baryon (Fig. 10). Since there exists no doubly charged meson the forward peak is not explained.

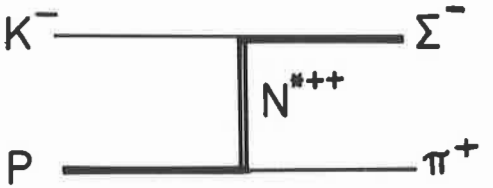


Fig. 10. The double charge exchange process $K^- P \to \pi^+ \Sigma^-$ can be mediated by an N^{*++}. This is dominant at backward angles only.

The quark model gives not only a simple explanation but also a definite prediction for both the magnitude and the slope of this forward peak in the differential cross section. In the quark model this process can go only via double scattering. If we can find the single scattering amplitudes from other processes we obtain a prediction. Remembering the quark structure of the following mesons and nucleons: $K^-(\bar{p}\lambda)$, $\bar{K}^0(\bar{n}\lambda)$, $\pi^-(n\bar{p})$, $\pi^+(\bar{n}p)$, $P(ppn)$, $N(pnn)$, $\Sigma^+(pp\lambda)$, $\Sigma^-(nn\lambda)$ we see that reaction (VI.1) is given by

$$\begin{array}{ccc} \bar{p}\lambda & & \bar{n}p \\ & \longrightarrow & \\ ppn & & nn\lambda \end{array} \qquad \text{(VI.2)}$$

or by the second order amplitude which is the product of two single scatterings

$$F(K^-P \to \pi^+\Sigma^-;\underline{\Delta}) = \frac{2i}{4\pi}\int 2f(\bar{p}p \to \bar{n}n;\underline{\Delta}_1)\, f(\lambda p \to p\lambda);\underline{\Delta}_2)\, d^2\Delta_1\, d^2\Delta_2$$

$$\cdot\, \delta(\underline{\Delta} - \underline{\Delta}_1 - \underline{\Delta}_2)\, S_{K\pi}(\underline{\Delta}_1, \underline{\Delta}_2)\, S_{P\Sigma}(\underline{\Delta}_1, \underline{\Delta}_2) \qquad \text{(VI.3)}$$

$S_{K\pi}$ is the $K\pi$ overlapp function and $S_{P\Sigma}$ the one for $P\Sigma$. Here we make the strong binding assumption and set both equal one

$$S_{K\pi} = S_{P\Sigma} = 1 \qquad \text{(VI.4)}$$

To obtain the quark-quark amplitudes in (VI.3) we study the single charge exchange processes

$$F(K^-P \to \bar{K}^0N) \sim 2f(\bar{p}p \to \bar{n}n, \underline{\Delta})$$

$$F(K^-P \to \pi^-\Sigma^+) \sim f(\lambda n \to n\lambda, \underline{\Delta}) \qquad \text{(VI.5)}$$

which are just given by the two amplitudes we need and can be obtained from experiment. We use again Gaussian parametrization

$$f(\bar{p}p \to \bar{n}n, \underline{\Delta}) = f(0)\, e^{-A\,\underline{\Delta}^2/2}$$

$$f(\lambda n \to n\lambda, \underline{\Delta}) = g(0)\, e^{-B\,\underline{\Delta}^2/2} \qquad \text{(VI.6)}$$

giving

$$F(K^-P \to \pi^+\Sigma^-;\underline{\Delta}) = \frac{2i}{A+B}\, f(0)\, g(0)\, \exp\left[-\tfrac{1}{2}\,\frac{AB}{A+B}\,\underline{\Delta}^2\right]. \qquad \text{(VI.7)}$$

We can now compare the differential cross sections for the three processes

$$\frac{d\sigma}{dt}(K^-P \to K^0N) = \pi|f(0)|^2\, e^{-A\Delta^2}$$

$$\frac{d\sigma}{dt}(K^-P \to \pi^4\Sigma^+) = \pi|g(0)|^2\, e^{-B\Delta^2}$$

$$\frac{d\sigma}{dt}(K^-P \to \pi^+\Sigma^-) = \pi\frac{|f(0)\, g(0)|^2}{(A+B)^2}\, e^{-\frac{AB}{A+B}\Delta^2} \qquad \text{(VI.8)}$$

If we assume that the slopes of the single charge exchange cross sections are approximately the same $A \sim B$, which is born out experimentally, we see that near the forward direction the double charge exchange differential cross section has just half the slope

$$\frac{AB}{A+B} = \frac{A}{2} = \frac{B}{2} \qquad \text{(VI.9)}$$

From the scarce experimental data one finds, we quote N. Dean[5] who has just carried out a more detailed analysis of these processes,

$$\frac{d\sigma}{dt}(K^-P \to \bar{K}^0N) = .85\, e^{-1.5\Delta^2}\ \frac{mb}{(GeV/c)^2}$$

$$\frac{d\sigma}{dt}(K^-P \to \pi^-\Sigma^+) = .30\, e^{-1.5\Delta^2}\ \frac{mb}{(GeV/c)^2} \qquad \text{(VI.10)}$$

and predicts for the double charge exchange process

$$\frac{d\sigma}{dt}(K^-P \to \pi^+\Sigma^-) \cong .02\ \ e^{-.75\Delta^2}\ \frac{mb}{(GeV/c)^2} \qquad \text{(VI.11)}$$

This is confirmed quite well by the sparse experimental data.[5]

VII. Discussion

By assuming that nucleons, anti-nucleons and pions are composed of three quarks, three anti-quarks, and a quark-anti-quark pair, respectively, and including contributions from single-, double-, and triple-scattering, we have been able to obtain rather satisfactory fits to the pp, $\bar{p}p$ and πp differential cross sections over a wide range of momentum transfer. Our model is phenomenological in that it contains 10 parameters, but 7 of these can be determined from the total sections, the ratio of real to imaginary part of the forward amplitudes and the initial slopes of the differential cross sections. The complete structures of the three differential cross sections away from the forward direction are therefore fitted and predicted by three free parameters. A glance at Figs. 6, 8 and 9 should convince you that this is not a trivial result.

We recall that there are two main components which enter the Glauber model. The first one is the impact parameter representation of the amplitudes. This can be considered as a mere Fourier transform and its validity is therefore quite general. The second is the assumption of additivity of the elementary phase shifts. This is a physical assumption and has to be tested. Fraunhofer scattering suggests that it may be a good approximation. Tests in high energy scattering in nuclear physics show excellent results. Our, somewhat arbitrary, parametrization by means of Gaussians, which was introduced to some extent to get analytically do-able integrals, may have influenced the final values of the parameters but should not be crucial for the basic idea. Our most hypothetical input has been the quark model.

While a detailed analysis of the differential cross sections may well be the most promising way to convince us of the existence of the quarks, the present results do not prove of course that the quark model is the correct one. There are many alternative explanations[32] for the structure in the differential cross sections but most of these other models address themselves to a single structural detail. Our general conclusion is that the quark model with multiple scattering provides a very simple explanation for the structure of high energy differential cross sections, and thus far we have run into no serious contradictions with experiment. If further work proves as successful the quark model must be regarded as a very convenient device for correlating a large amount of experimental information.

We should emphasize that the present model is in no contradiction at all to some of these alternate approaches to explain the structure of the differential cross sections, and indeed it might provide a more solid base for some of them. Firstly we recall that our model provides a natural explanation to Van Hove's observation that the cross sections, for small $|t|$, behave like the fourth power of the form factors. Attempts to explain the dips in the differential cross sections with the help of interfering Regge trajectories have been qualitatively successful. But we should not forget that the Regge pole model remains completely phenomenological unless we have a definite bound state model to create the trajectories. A Regge pole model could be used to produce the diffraction peak . It would give its energy dependence on which we have nothing to say. Combined with our model one would then get the entire wide angle area. We are studying these problems at present and also how sensitive our results are to the number of "quarks" that make up a nucleon.

Finally we would like to conclude with a plea to the experimental physicists to get data on elastic pp scattering and on double charge exchange covering a wide range of momentum transfer at one

single high energy. Care should be taken to cover the region of the breaks, especially the first one, carefully, and to assure that small angle and large angle measurements have consistent normalization by covering overlapping regions. With such experimental data the quark model for the structure of the nucleons could be tested more rigorously and the parameters determined more precisely.

References

1. D. Harrington and A. Pagnamenta, Phys. Rev. Letters 18, 1147 (1967) and Phys. Rev. 173, 1599 (1968).
2. R. J. Glauber, Lectures in Theoretical Physics (Interscience Publishers, Inc., New York, 1959), ed. W. E. Brittin, p. 315. See also the references given in I. and V. Franco and R. J. Glauber, Phys. Rev. 142, 1195 (1966).
3. V. Franco, Phys. Rev. Letters 18, 1159 (1967).
4. A. D. Krisch, Phys. Rev. Letters 19, 1149 (1967).
5. N. W. Dean, Nucl. Phys. B7, 311 (1968).