Principles and problems of phase-shift analysis

1975 Rep. Prog. Phys. 38 1099

(http://iopscience.iop.org/0034-4885/38/9/002)

Principles and problems of phase-shift analysis

J E BOWCOCK AND HUGH BURKHARDT

Department of Mathematical Physics, University of Birmingham, Birmingham B15 2TT

Abstract

Phase-shift analysis is the extraction of the scattering amplitude from the scattering cross section and other experimentally observable quantities such as polarizations. When only elastic scattering is energetically allowed, unitarity determines the unobservable angle-dependent complex phase of the scattering amplitude with, at most, only a few discrete alternative solutions. Above the inelastic threshold the unitarity constraint on a scattering amplitude is only an inequality and a continuum of different amplitudes will correspond to exactly the same observables. In practical cases these differences can be important. Extra theoretical input of a dynamical nature can, in principle, remove the continuum ambiguity but, because numerical analytic continuation is always involved, data of absurd accuracy are required. Thus unique answers can, in practice, only be found by introducing further model-dependent assumptions; it is important to recognize this and ensure that these assumptions are as dynamically plausible as possible. Recent results using the structure of the amplitude in both kinematic variables suggest that fixed-t dispersion relations might form a sound basis for an inelastic phase-shift analysis.

This review was completed in July 1975.

Contents

1. Introduction

Phase-shift analysis of scattering data in high-energy physics has made a very important contribution to the picture we have developed of hadrons and their strong interactions. In particular the detailed spectrum of hadron states and the pattern of their decay rates, which fits in so well with that expected from the quark model, was largely derived from phase-shift analysis. The results of this work and the theoretical ideas that it has confirmed have been extensively reviewed, for example by Donnachie (1973). Phase-shift analysis is also extensively employed in other areas of nuclear physics.

Here we shall be concerned with understanding the process of phase-shift analysis itself, the extraction of the scattering amplitude from the experimental data. We shall see that, once inelastic processes are energetically allowed, phase-shift analysis is far from being a straightforward, mathematically reliable process; indeed, there is no compelling reason to believe that the resultant amplitudes are correct. The circumstantial evidence discussed above suggests that they are roughly right, however, and we shall comment on the dynamical implications of this success. It is worth emphasizing that a totally reliable way of finding the amplitude from the data is of immense value, since the amplitude is much easier to study from a dynamical point of view than the observables, which all contain complicated interference effects; for example, the dispersion relation approach to dynamics is based on the analytic properties of the amplitudes.

Phase-shift analysis is generally used to mean the process of obtaining the (complex) scattering amplitude from the (real) scattering cross sections, differential or total. If the particles involved in the scattering have spin, there will be more than one amplitude and, correspondingly, other experimental observables such as polarizations which are sufficient to determine the magnitudes and *relative* phases of the various amplitudes. Thus the most important problem, namely the determination of the absolute, angle-dependent phase of the scattering amplitudes, is essentially the same as in the spinless case, so we shall leave the complications of spin until §7.

Since the scattering amplitude f is a complex number and the differential cross section

$$\mathrm{d}\sigma/\mathrm{d}\Omega = |f|^2 \tag{1.1}$$

is real, it is not obvious that the information exists to fix f. We shall examine what theoretical assumptions are usually made and how far they leave ambiguities in the resultant amplitude. From (1.1) it is clear that it is the phase of f which is undetermined, so that for any function ϕ of energy and angle which is real in the whole physical region

$$f' = \mathrm{e}^{\mathrm{i}\phi} f \tag{1.2}$$

gives exactly the same cross section.

It should be remarked that this phase uncertainty $\phi(s)$ has nothing to do with the unobservable overall phase of wavefunctions in quantum mechanics; the asymptotic wavefunction at large distances from the scattering centre may be written

$$\psi(\boldsymbol{x}) \approx \mathrm{e}^{\mathrm{i}\boldsymbol{k}.\boldsymbol{x}} + f(\theta)\frac{\mathrm{e}^{\mathrm{i}kr}}{r}, \qquad r \to \infty$$

so the phase of the scattering amplitude is the *relative* phase of the incident and scattered waves. This phase has observable consequences in situations where multiple scattering occurs, such as scattering from deuterons or complex nuclei. Unfortunately, the internal dynamics of the scattering systems is not understood sufficiently well to make the extraction of the phase possible in practice, except within the framework of approximate models such as that of Glauber (1959).

Goldberger *et al* (1963) discuss the possibility of scattering correlation measurements, using a phase reconstruction method like that devised by Hanbury-Brown and Twiss (1956) in radio astronomy, but it is not a practical possibility. So we shall here regard the overall phase of the amplitude as unmeasurable and consider in detail how far it is fixed by the theoretical constraints on the amplitude and the measurement of its modulus—this is the problem of phase-shift analysis.

1.1. Variables and amplitudes

First we shall briefly define our notation for the kinematic variables and scattering amplitudes—more detailed derivations can be found in many places, including Burkhardt (1969). We shall be largely concerned with the quantum scattering of two spinless particles with a total energy E and momenta q in the centre-of-mass system. They scatter through the angle θ from a direction $\boldsymbol{q}$ to a direction $\boldsymbol{q}'$, and we write

$$z = \cos\theta = \boldsymbol{q}.\boldsymbol{q}'/qq'. \tag{1.3}$$

However, we shall also need to discuss those general properties of scattering amplitudes which are most simply expressed in terms of the relativistically invariant Mandelstam variables s, t and u; these are just the squares of the three possible sums of pairs of 4-momenta of the scattering particles in figure 1, taking momentum conservation into account and counting *incoming* momenta as positive.

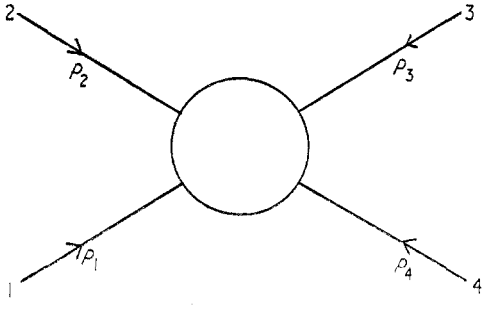

Figure 1. Kinematics of a scattering process—note signs chosen.

In the simple case of all equal masses they are related to the centre-of-mass variables by

$$\begin{aligned} s &\equiv (p_1+p_2)^2 = (-p_3-p_4)^2 = 4m^2+4q^2 = 4m^2+4q'^2 \\ t &\equiv [p_1-(-p_4)]^2 = (p_2+p_3)^2 = -2q^2(1-z) \\ u &= [p_1-(-p_3)]^2 = (p_2+p_4)^2 = -2q^2(1+z) \end{aligned} \tag{1.4}$$

so that

$$q^2 = \tfrac{1}{2}(s-4m^2)^{1/2} \tag{1.5}$$

$$z = 1+\frac{t}{2q^2} = -1-\frac{u}{2q^2} = \frac{t-u}{4q^2}. \tag{1.6}$$

For general mass values one finds

$$\begin{aligned} q^2 &\equiv q_{12}{}^2 = [s-(m_1+m_2)^2][s-(m_1-m_2)^2]/4s \\ q'^2 &\equiv q_{34}{}^2 = [s-(m_3+m_4)^2][s-(m_3-m_4)^2]/4s \\ z &= \frac{s(t-u)+(m_1{}^2-m_2{}^2)(m_3{}^2-m_4{}^2)}{4sqq'}. \end{aligned} \tag{1.7}$$

Of course, all the masses can only be different in an inelastic scattering process. Only two of the kinematic variables can be independent and, using the mass condition $p_i{}^2 = m_i{}^2$, we find

$$s+t+u = \sum_{i=1}^{4} m_i{}^2. \tag{1.8}$$

The variables, with this constraint, are plotted for equal masses in figure 2 in triangular coordinates; the value of each variable is its perpendicular distance from its axis. The physical regions are shown shaded.

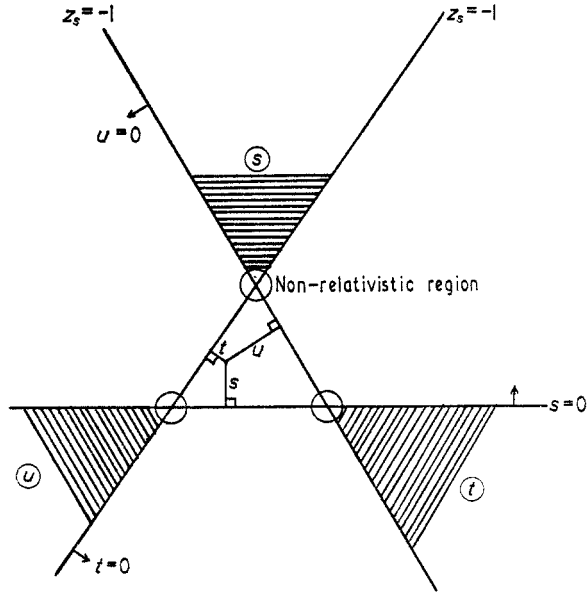

Figure 2. Physical regions of crossing-related processes in triangular coordinates.

The crossing-symmetric amplitude $A(s, t, u)$ describes each of the following scattering processes in its appropriate physical region, which for equal masses are

$$\begin{aligned} &1+2 \to 4+3 \qquad s>4m^2, \quad t \text{ and } u<0 \\ &1+\bar{4} \to \bar{2}+3 \qquad t>4m^2, \quad s \text{ and } u<0 \\ &1+\bar{3} \to \bar{2}+4 \qquad u>4m^2, \quad s \text{ and } t<0. \end{aligned} \tag{1.9}$$

The time-reversed processes, and also those where each particle N is replaced by its own antiparticle $\bar{N}$, are also given by the same amplitude in the same region. This amplitude is related to the differential cross section for $1+2 \to 4+3$ by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{q'}{q}\left|\frac{2A}{\sqrt{s}}\right|^2 \tag{1.10}$$

and so to the usual non-relativistic amplitude f by

$$f=\frac{F}{q}=\frac{2\mathrm{A}}{\sqrt{s}}. \tag{1.11}$$

The quantity F is introduced for convenience because of its simple unitarity properties.

1.2. Ideal phase-shift analysis

Let us look first at the theoretical assumptions that lie behind an ideal phase-shift analysis and try to see how they may enable us to find a complex number (the amplitude) from a real one (the cross section) at each centre-of-mass energy E and scattering angle θ. First, for completeness, we must mention Lorentz invariance (L) under translations in space and time, rotations and boosts of velocity, which tells us that the amplitude for the scattering of two particles will depend on only two variables, such as energy and scattering angle, and not on the position, orientation or velocity of the whole system. Secondly, one explicitly uses the short-range nature of the forces in strong interactions of hadrons (R). Normally we expand the amplitude in 'partial waves' of definite angular momentum l

$$f=\sum_{l=0}^{\infty}(2l+1)f_l(E)P_l(z) \tag{1.12}$$

where $z=\cos\theta$. The existence of a lightest particle, mass μ, which can be 'exchanged' in scattering gives rise to a Yukawa potential contribution of the form

$$V(r)=\lambda\frac{\mathrm{e}^{-\mu r}}{r}. \tag{1.13}$$

Since the classical angular momentum l at radius r is $l=qr$, where q is the relative momentum, it is not surprising that for large l

$$f_l\sim\exp\left(-\frac{\mu}{q}l\right), \qquad l\to\infty \tag{1.14}$$

so that the high partial waves are very small. The exponential behaviour is rigorously true, though the argument is not!

In practical phase-shift analyses the partial-wave amplitudes f_l are usually assumed to be zero above some cut-off $l=L$

$$f_l\equiv 0, \qquad l>L. \tag{1.15}$$

As we shall see, this apparently innocent assumption conceals the true ambiguity situation since (1.2) implies an infinite number of partial waves, consistent with the true condition (1.14) but not with (1.15).

Mathematically, the exchanged systems which give rise to scattering appear as singularities of the scattering amplitude in the complex z plane at the unphysical values

$$z=1+\frac{t_0}{2q^2} \tag{1.16}$$

where $t=t_0$ is the square of the mass of the exchanged system. A discrete particle exchange gives an isolated pole, while a multi-particle system, which can have a continuous range of masses, gives a cut stretching over the corresponding range of z, given by (1.16). The cuts are as shown in figure 3.

The word 'exchange' is used in two different senses here. All forces in quantum scattering may be ascribed to the exchange of particles—in the simplest case the Coulomb interaction is photon exchange; this is the sense we have used so far, figure 4(*a*). Nuclear physicists use the phrase 'exchange forces' to describe interactions in which the two scattered particles exchange their identities in the final state, so that for example, figure 4(*b*), forward nucleon exchange in pion–nucleon scattering looks like

Figure 3. Singularities of the scattering amplitude in $z = \cos\theta$.

backward scattering; such processes give rise to singularities at positive u and so for real $z < -1$ from equation (1.6), but their effects on the phase-shift ambiguity situation are qualitatively similar to the t singularities so we shall not discuss them further explicitly.

The third important theoretical constraint is that provided by unitarity (U), which is the statement of probability flux conservation. It is not surprising that it is much more powerful in the elastic region, where only elastic scattering is allowed and can be fully observed, than at higher energies where some of the probability flux goes in inelastic channels. Elastic unitarity requires

$$f_l = \frac{F_l}{q} = \frac{\exp(i\delta_l)\sin\delta_l(E)}{q} = \frac{\exp(2i\delta_l)-1}{2iq} \tag{1.17}$$

so that for each partial wave the complex amplitude f_l is, in fact, determined by a single *real* number δ_l; it is plausible that the real cross section contains enough information to fix the δ_l. In fact, we shall see in the next section that there may still be a limited number of discrete alternative amplitudes. In the inelastic region unitarity is much less powerful and simply requires

$$f_l = \frac{\eta_l \exp(2i\delta_l)-1}{2iq}, \qquad 0 \leqslant \eta_l \leqslant 1 \tag{1.18}$$

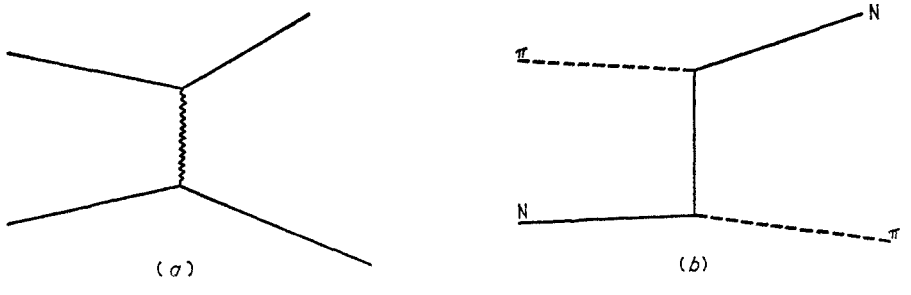

Figure 4. Single-particle exchange and 'exchange' forces.

so that there are indeed two real parameters for each partial wave, subject only to the inequality constraints on the η_l. These simply require each partial-wave amplitude to lie inside the circle in the complex z plane shown in figure 5(a); elastic unitarity says that $\eta_l = 1$ so that the amplitude must lie on the boundary circle. For an inelastic scattering process the -1 in the numerators of equations (1.17) and (1.18), which arise from the incident plane wave in the scattering wavefunction, are missing so the unitary circles are centred at $F_l = 0$ rather than $F_l = \frac{1}{2}i$, as shown in figure 5(b).

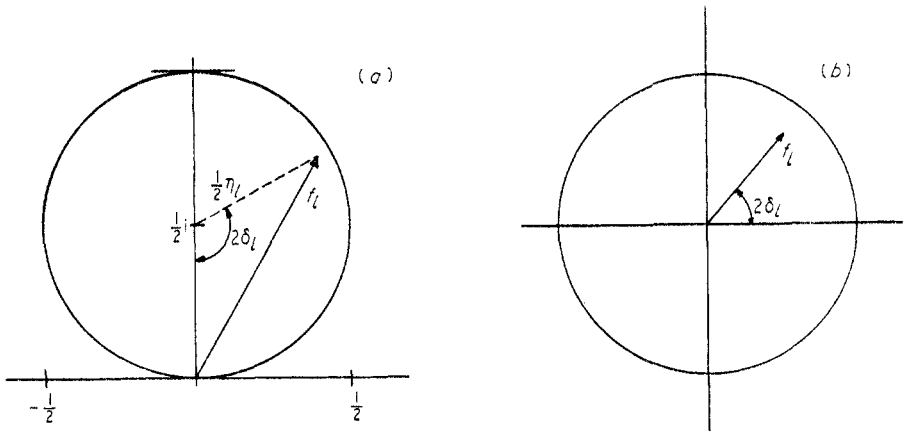

Figure 5. The unitary circle. (a) Elastic, (b) inelastic scattering amplitudes.

The weaker inequality constraints allow a continuous set of amplitudes which give exactly the same cross section. This continuum ambiguity is potentially far more serious than a few discrete alternatives and we shall be largely concerned with describing how big it is and what might be done to cure it by including extra theoretical input. We shall see that, in principle, one can remove the ambiguity by using the analytic properties of the amplitude but that in practice, useful results require data of unachievable precision.

The general form of the unitarity relation for the scattering from a state i to a state f can be written

$$\text{Im } A_{fi} = 8\pi \sum_j \int A_{fj}^* A_{ji} \, \mathrm{d}\Phi_j \tag{1.19}$$

where the integral is over the invariant phase space of the intermediate state

$$\mathrm{d}\Phi_j = \prod_{n=1}^{N_j} \frac{\mathrm{d}^3 p_n}{2E_n(2\pi)^3} (2\pi)^4 \delta^{(4)} \left(\sum^{N_j} p_n - P^i \right). \tag{1.20}$$

The sum is over all accessible intermediate states, j. An immediate consequence of this is the so-called optical theorem, which is found by making the states i and f identical. The right-hand side of (1.19) is then essentially the total cross section and we have for the forward elastic amplitude

$$\text{Im } A(s, \theta = 0) = \tfrac{1}{2}\sqrt{s} \, \text{Im} f(E, 0) = \frac{q\sqrt{s}}{8\pi} \sigma_{\text{tot}}. \tag{1.21}$$

This result provides a further constraint on the forward scattering, fixing the forward amplitude except for a sign uncertainty of the real part.

Information on the sign of the real part can often be found experimentally by observing the interference very close to the forward direction between the hadronic scattering amplitude and the Coulomb scattering, which is known theoretically. The latter is generally much weaker but, being peaked sharply forward, is visible close to the forward direction.

1.3. Practical phase-shift analysis

This introduction would not be complete without a brief summary of the method and successes of practical phase-shift analysis, as we shall be dwelling at length on the uncertainties of these results. In practice, the scattering amplitude is usually calculated in terms of the phase shifts and elasticities in each partial wave up to the cut-off L, and these are varied until the reusltant cross sections and polarizations are a best fit to the data at a particular energy. This process typically yields a large number of solutions; often several hundred random searches are made, each giving a different 'best fit'. The alternative method often employed is to make a Legendre expansion of the cross section and to fit the coefficients from this to those predicted from the amplitude. Some sort of energy-smoothing criterion is then used to pick out of all these solutions, one which joins smoothly to lower energies. We shall discuss these methods in §6.

Various phase-shift groups have found pictures of the resonance situation in π–N scattering that are fairly, but not perfectly, consistent. Their amplitudes as a function of angle tend to be less in agreement. This unanimity seems to provide something of a puzzle, if there really is a serious continuum ambiguity. We shall find plausible reasons for the amount of unanimity involved, which give no confidence that it implies reliable amplitudes at high energies or in other processes.

2. Elastic phase-shift analysis

In most scattering processes there is an energy region where the only energetically allowed strong scattering process is the elastic scattering of the two incident particles. It was in this elastic region that phase-shift analysis was originally applied, when accelerator energies were generally still too low to create additional particles. The unitarity condition is, as we have already remarked, particularly simple and powerful here since it is an equality involving only the elastic amplitude itself. The mathematical problem of determining the amplitude from the cross section is well defined. While it is still not completely solved, even for the case of spinless scattering, it is very largely understood and the results are sufficient to give complete confidence in the reliability of phase-shift analysis in the elastic region. The main results for perfect data with zero errors are essentially these:

(i) The ambiguity is discrete with, at most, a finite number of discrete solutions.

(ii) Provided that all partial waves, except the s wave, are not too large the amplitude is unique, apart from the so-called trivial ambiguity

$$\tilde{F}(z) = -F^*(z) \tag{2.1}$$

which corresponds to changing the sign of all phase shifts.

(iii) If the partial waves become larger, then in a number of situations there is an additional twofold ambiguity first noted by Crichton (1966). For a variety of artificial cases it has been shown that this is the only ambiguity that can exist; it is plausible that these results are universally true in the elastic region.

We shall now go on to consider these points in detail.

A good deal of attention has been devoted to the conditions for the *existence* of a unitary elastic amplitude for a given cross section. It is clear that not all cross sections which could be written down are physically possible—for example, they must not vanish in the forward direction since the optical theorem (1.21) requires the forward amplitude to have at least an imaginary part, if there is any scattering at all. Martin (1969) has given a more sophisticated example. Clearly, the existence problem is not important in practice since we know that the data *do* come from a unitary amplitude, but it is of some mathematical interest. Conversely, the *uniqueness* problem is of great practical importance, since the results of a phase-shift analysis are assumed to be *the* amplitude for the process, which may not be true if there are alternative solutions.

2.1. The unitarity condition

It is convenient to write the unitarity relation in its integral form, which in the elastic region is simply

$$\operatorname{Im} F_{12}=\frac{1}{4\pi}\int F_{13}F_{32}\,\mathrm{d}\Omega_3 \tag{2.2}$$

where $F_{12}=F(\cos\theta_{12})$ is the scattering amplitude for a direction 1 to a direction 2. From this form it is clear that the trivial ambiguity, which reverses all real parts, will always be present since it changes neither (2.2) nor the modulus of the amplitude. The addition theorem for spherical harmonics:

$$\int \frac{\mathrm{d}\Omega_3}{4\pi} P_{l_1}(z_{13})P_{l_1}(z_{32})=\frac{\delta_{l_1l_2}}{2l_1+1}P_{l_1}(z_{13})$$

and (1.12) in (2.2) yields the partial-wave unitarity relation (1.17). To combine the knowledge of the differential cross section, or equivalently of $|F(\cos\theta)|$, with the unitarity condition (2.2) it is convenient to write

$$F(\cos\theta)=|F(\cos\theta)|\exp[\mathrm{i}\psi(\cos\theta)] \tag{2.3}$$

when (2.2) becomes

$$\sin\psi_{12}=\frac{1}{4\pi}\frac{\int \mathrm{d}\Omega_3|F_{13}||F_{32}|\cos(\psi_{13}-\psi_{32})}{|F_{12}|} \tag{2.4}$$

which is a nonlinear equation for the phase function ψ.

2.2. The existence question

A condition for the existence of a unitary amplitude which yields a given cross section can be found by applying the contraction mapping principle to equation (2.4), which is regarded as a mapping from a function space containing $\psi(z)$ onto itself—any function $\psi(z)$ substituted in the right-hand side of (2.4) will yield another function $\psi'(z)$ on the left-hand side, and each of these functions is a point in a suitable space—so that

$$\psi'=\mathrm{O}(\psi). \tag{2.5}$$

If the output region of the function space $\psi'(z)$ produced by the mapping lies entirely within the input region for any choice of ψ, and if any pair of points are brought closer together by the mapping, then the mapping is called a contraction mapping; iteration of the mapping will then give successively smaller regions and it may be proved that there always exists a fixed point, which is mapped onto itself.

Such fixed-point theorems are powerful weapons in nonlinear analysis. They were first applied to these problems by Klepikov (1965), Newton (1968), Martin (1969) and Atkinson *et al* (1972, 1973b); the fixed-point value of the phase function $\psi(z)$ is, of course, the unitary solution which we seek.

In order that the mapping (2.4) has at least one fixed point, it is plausible that some limit will be needed on

$$\sin \mu \equiv \max \int \frac{d\Omega_3}{4\pi} \frac{|F_{13}||F_{32}|}{|F_{12}|} \tag{2.6}$$

where the maximum is over all possible values of the angle (12). If the phase is slowly varying then (2.4) is indeed roughly equal to the maximum value of $\sin \mu$. For this to be physical we need

$$\sin \mu < 1.$$

One can then show, using standard mathematical techniques, that this condition is sufficient to ensure that (2.4) is indeed a contraction mapping and so must have at least one fixed point, whose phase ensures that elastic unitarity is satisfied. The iteration procedure implied indeed converges and can be carried out on a computer to give the solution. This result is interesting but the condition $\sin \mu < 1$ is, in fact, very restrictive; for instance, due to the denominator of F_{12} in (2.6) it excludes differential cross sections with deep dips.

If we remove the trivial ambiguity by requiring that the real part be positive, then

$$0 < \psi(\cos\theta) < \mu < \tfrac{1}{2}\pi$$

which in turn implies that

$$\text{Re } F(\cos\theta) > 0, \qquad \text{Im } F(\cos\theta) > 0 \tag{2.7}$$

so there are no sign changes. Furthermore one can argue for $l \geqslant 1$

$$\text{Re } F_0 \pm \text{Re } F_l = \int_{-1}^{1} \tfrac{1}{2}\, dx\, \text{Re } F(x)[1 \pm P_l(x)] > 0 \tag{2.8}$$

since $|P_l(x)| < 1$ and similarly for the Im $(F_0 \pm F_l)$. Since all partial waves are on their unitary circles these inequalities imply Im $F_l < \frac{1}{2}$ or $|\delta_l| < \frac{1}{2}\pi$. This bound has been improved to $\delta_l < \frac{1}{6}\pi$, $l \geqslant 1$. So all the waves except the s wave must be fairly small, and certainly non-resonant for these results to hold; they do not therefore apply to the regions most interesting experimentally. However, the existence problem is not important in practice, because we are confident that our cross sections do arise from unitary amplitudes.

2.3. The uniqueness problem

A stronger condition than $\sin \mu < 1$ is likely to be needed to ensure that the solution is unique; various such conditions have been derived and the best so far is

$$\sin \mu < 0{\cdot}79 \tag{2.9}$$

due to Martin (1969). We shall give a simple proof, also due to Martin, that the solution is unique as long as $\sin \mu < (2)^{-1/2}$. It employs the standard technique of assuming two solutions, F and G, and showing that their difference is zero.

It follows from these assumptions that we can write

$$\operatorname{Im} F_{12} - \operatorname{Im} G_{12} = \frac{1}{4\pi} \int d\Omega_3 [(\operatorname{Re} F - \operatorname{Re} G)_{13} (\operatorname{Re} F + \operatorname{Re} G)_{32} + (\operatorname{Im} F - \operatorname{Im} G)_{13} (\operatorname{Im} F + \operatorname{Im} G)_{32}] \quad (2.10)$$

but since $|F| = |G|$

$$(\operatorname{Im} F + \operatorname{Im} G)(\operatorname{Im} F - \operatorname{Im} G) + (\operatorname{Re} F + \operatorname{Re} G)(\operatorname{Re} F - \operatorname{Re} G) = 0 \quad (2.11)$$

giving

$$(\operatorname{Im} F - \operatorname{Im} G)_{12} = \int \frac{d\Omega_3}{4\pi} (\operatorname{Im} F - \operatorname{Im} G)_{13} \times \left[\left(\frac{\operatorname{Im} F + \operatorname{Im} G}{\operatorname{Re} F + \operatorname{Re} G} \right)_{32} - \left(\frac{\operatorname{Im} F + \operatorname{Im} G}{\operatorname{Re} F + \operatorname{Re} G} \right)_{13} \right] (\operatorname{Re} F + \operatorname{Re} G)_{23}. \quad (2.12)$$

Now we have seen that $(\operatorname{Re} F + \operatorname{Re} G)$ may be chosen to be always positive, so it follows that

$$|\operatorname{Im} F - \operatorname{Im} G| \leqslant \max |\operatorname{Im} F - \operatorname{Im} G| \tan \mu \int \frac{d\Omega}{4\pi} (\operatorname{Re} F + \operatorname{Re} G). \quad (2.13)$$

But $\int (d\Omega/4\pi) \operatorname{Re} F$ is the real part of the s wave, which must always be less than $\frac{1}{2}$ in the modulus so, finally,

$$|\operatorname{Im} F - \operatorname{Im} G| \leqslant \max |\operatorname{Im} F - \operatorname{Im} G| \tan \mu. \quad (2.14)$$

If $\tan \mu < 1$ this can only be true if $\operatorname{Im} F = \operatorname{Im} G$, which combines with their equal moduli to give $F = G$ (the arbitrary sign of the real part is the trivial ambiguity again). Hence for $\sin \mu < 1/\sqrt{2}$ the solution is unique.

Somewhat more powerful arguments give the result (2.9), but even these involve using some upper bounds that are far from tight and Martin has conjectured that uniqueness probably holds for $\sin \mu < 1$. We have already emphasized that this is a very restrictive condition that excludes most interesting scattering situations, but it does cover one most important case.

2.4. The threshold region

In most processes there is a region near threshold in which the s wave is dominant, since it can be shown that the higher partial waves

$$F_l \sim q^{2l+1} \quad (2.15)$$

near $q = 0$. This follows essentially because the coefficient of z^l in the expansion of the longest-range part of the scattering

$$F(z) \sim \frac{q}{E} \frac{\lambda}{\mu^2 - t} = \frac{q}{E} \frac{\lambda}{\mu^2 + 2q^2(1-z)} \quad (2.16)$$

in powers of z will always contain the coefficient q^{2l} because it is the momentum transfer $t = 2q^2(1-z)$ that always appears in the scattering amplitude. Thus, unless for some special reason the s-wave scattering length is zero, there will always be an energy

region in which the s wave dominates, when

$$\sin \mu \doteqdot \sin \delta_0 \leqslant 1. \tag{2.17}$$

In this region the phase-shift analysis will be unique.

2.5. Artificial cases—polynomial or entire amplitudes

A much fuller understanding of the uniqueness problem and of the nature of the ambiguities has been obtained under certain artificial assumptions, though even here the picture is incomplete. We shall first discuss situations in which there are only a finite number of partial waves, so that the amplitude is a polynomial in z. This is an unphysical assumption since the tail of the scattering interaction affects all partial waves, as shown in (1.14), and it is incompatible with general principles, particularly crossing symmetry, which says that the same function of the kinematic variables describes the different scattering processes obtained by reversing external lines of the scattering picture, and at the same time changing the corresponding particles to their antiparticles. However, it is also true that if we take into account a large enough number of partial waves, any reasonable angular distribution can be fitted to any chosen precision by a finite number of partial waves, as long as we stay within the physical region. In practice, most analyses are done by fitting with a finite cut-off L in angular momentum. We shall see later that this reasonable approximation is not as innocent as it may seem.

If the amplitude is a polynomial of degree L, it may be written

$$F(z)=\sum_{l=0}^{L}(2l+1)F_l P_l(z) \tag{2.18}$$

and the cross section will have the form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}=\sum_{n=0}^{2L} c_n(2n+1)P_n(z). \tag{2.19}$$

The expansion coefficients c_n will then be bi-linear combinations of f_l and $f_{l'}{}^*$ of the form

$$c_n=\sum_{l,l'}(C_{l0,l'0}{}^{n0})^2 f_l f_{l'}{}^*\left(\frac{(2l+1)(2l'+1)}{(2n+1)}\right)=\sum_{l,l'}\bar{C}_{l,l'}{}^n f_l f_{l'}{}^* \tag{2.20}$$

where the $C_{lm,lm'}{}^{nM}$ are the usual Clebsch–Gordan coefficients. To determine the partial-wave amplitudes f_l from the coefficients c_n, we can work down from the top (this is a highly artificial thing to do since in practice it is the high coefficients in the Legendre expansion of the cross section that will be least well determined, but in the cut-off model it is exact). We have

$$\begin{aligned} c_{2L}&=\bar{C}_{L,L}{}^{2L}|f_L|^2 \\ c_{2L-1}&=\bar{C}_{L,L-1}{}^{2L-1}\,2\mathrm{Re}\,f_L f_{L-1}{}^* \\ c_{2L-2}&=\bar{C}_{L-1,L-1}{}^{2L-2}|f_{L-1}|^2+\bar{C}_{L,L-2}{}^{2L-2}\,2\mathrm{Re}\,f_L f_{L-2}{}^* \end{aligned} \tag{2.21}$$

so clearly c_{2L} determines $|f_L|^2$. Since f_L must lie on the unitary circle, there are two possibilities, shown in figure 6. Remembering the trivial ambiguity, we may choose $\mathrm{Re}\,f_L>0$. Going on to the next coefficient c_{2L-1}, we can determine $\mathrm{Re}\,f_{L-1}f_L{}^*$, which

gives us the projection of f_{L-1} on the direction of f_L—again, this allows two f_{L-1} possibilities, as in figure 6. For each of these, we go on to c_{2L-2} and determine $\mathrm{Re}\, f_{L-2} f_L^*$, which again will give two solutions for f_{L-2}. So working down to c_L we find a set of 2^L possible solutions.

The remaining L coefficients $c_0, c_1, \ldots, c_{L-1}$ provide additional constraints on these solutions and one might expect that they would select an unique solution (real numbers contain infinitely more than the L bits of information required, *if* they are

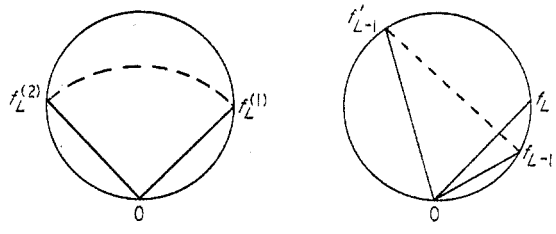

Figure 6. The deduction of a polynomial amplitude from the cross section.

truly independent). However, Crichton (1966) constructed an example with $L=2$ which has a twofold ambiguity. He found that the two sets of phase shifts

$$\begin{aligned} &\delta_0 = -23^\circ\,20' \qquad \delta_1 = -43^\circ\,27' \qquad \delta_2 = 20^\circ \\ &\delta_0 = 98^\circ\,50' \qquad \delta_1 = -26^\circ\,33' \qquad \delta_2 = 20^\circ \end{aligned} \tag{2.22a}$$

give identical cross sections. These amplitudes do not violate the Martin uniqueness theorems because Re F changes sign and $\sin\mu > 1$.

2.6. Crichton's ambiguity and the counting argument

There are grounds for hoping that the twofold ambiguity, which Crichton (1966) illustrated, may be the only sort of ambiguity which is present in the elastic unitarity situation. Martin (1969) showed that this was so for $L=2$, and elucidated the relation between the two solutions, which have the form

$$\frac{15}{2}\exp(\mathrm{i}\delta_2)\sin\delta_2\,[\cos\theta - x_1(\delta_2) \pm \mathrm{i}y_1(\delta_2)]\,(\cos\theta - \tfrac{4}{5} + \tfrac{1}{5}\mathrm{i}\cot\delta_2). \tag{2.22b}$$

These results have been extended by Atkinson *et al* (1973a, 1974) and by Cornille and Drouffe (1974) to the case of $L=4$ and 5, while Itzykson and Martin (1973) have found the same result for entire functions which are not polynomials by the powerful use of the substantial mathematical understanding of such functions that has been built up. There is also no known example of an elastic unitary amplitude with more than a twofold ambiguity, always in addition to the trivial ambiguity (2.1).

Furthermore, there is a general argument due to Berends and Ruysenaars (1973) which, though not conclusive, lends support to the idea that there can be, at most, a twofold ambiguity. It is interesting enough to be worth summarizing here. We choose to write the amplitude as a product over its zeros

$$F(z) = \lambda \prod_{i=1}^{L} (z - z_i) \tag{2.23}$$

the overall coefficient of z^L is clearly proportional to F_L. In fact

$$\lambda = \exp(i\delta_L) \sin\delta_L (2L+1) \frac{2L!}{2^L (L!)^2}. \tag{2.24}$$

Now the cross section has the form

$$\frac{d\sigma}{d\Omega} = \frac{1}{q^2} |\lambda|^2 \prod_{i=1}^{L} (z - z_i)(z - z_i^*) \tag{2.25}$$

and will be unaffected if we replace any of the roots z_n by its complex conjugate (this was first noted by Gersten 1969). We can make these changes in 2^L possible ways, as long as we do not bother about unitarity, except for the highest partial wave where it fixes $\sin^2\delta_L$. Let us choose this time to remove the trivial ambiguity by choosing $\sin\delta_L > 0$. Now let us impose unitarity.

Let us assume that there are N unitary amplitudes, out of the possible 2^L. The amplitudes are functions of $2L+1$ variables—δ_L, the real parts of the L zeros, and the moduli of their imaginary parts—each amplitude being a different function of those variables because of the sign changes from the complex conjugations. Now the unitarity equations

$$\text{Im } F_l^k = |F_l^k|^2 \qquad k = 1, 2, \ldots, N, \quad l = 0, 1, 2, \ldots, L-1 \tag{2.26}$$

for each of the solutions provide a total of NL constraints; *if these equations are independent*, and there is to be a solution there must not be more equations than variables so

$$NL \leqslant 2L+1 \tag{2.27}$$

which requires $N \leqslant 2$. Since the independence has effectively been proved for polynomials with $L \leqslant 5$ and for entire functions, it is not unreasonable to suppose it holds more generally.

In summary, it seems fairly likely that there is at most a twofold non-trivial ambiguity in the elastic case. Notice that the complex conjugation of the zeros in (2.23) applies equally well to polynomial amplitudes in the inelastic region (Gersten 1969), and shows that the ambiguity there is again discrete and of order $\leqslant 2^L$—this result is not true except in that artificial case.

2.7. Removal of the trivial ambiguity

We have seen that the trivial ambiguity, which corresponds to reversing the sign of all real parts and thus of all phase shifts, is always present under the normal assumptions of phase-shift analysis. For charged particles this can, in principle, be removed by observation of Coulomb interference. It can also be removed by the inclusion of extra dynamical constraints on the amplitude, but we should want them always to be of a most reliable kind, and not merely based on an uncertain dynamical model. The forward dispersion relations provide us with such a weapon, since they directly express the forward real part of the amplitude in terms of integrals over the forward imaginary part, which is in turn related to the total cross section by the optical theorem (1.21). Since they relate amplitudes at all energies this moves us out of the domain of strict single-energy analysis and looks towards the multi-energy procedures we shall discuss in detail in §6. The argument is not quite as simple as it might be because the relations

contain subtraction terms and have the form, for example for pion–nucleon scattering

$$A(E,\,\theta=0)=A+BE+\frac{(E^2-E_0^2)}{\pi}\int_{\mu}^{\infty}\mathrm{d}E'\,\frac{q\sigma_{\mathrm{tot}}(E')}{(E'^2-E_0^2)(E'^2-E^2)} \tag{2.28}$$

but it can be shown that, except in pathological circumstances, only one of the two solutions will be compatible with them. The argument notes that the optical theorem gives Im A and the data $|A|^2$ on the physical cut of the integral. So Re A is fixed, apart from a sign, and if there are two different amplitudes they can only differ in this way. Both amplitudes, $\tilde{A}$ and A, satisfy the dispersion relation with the same integral; they can only differ in the subtraction terms, so

$$\tilde{A}-A=\alpha+\beta E \tag{2.29}$$

but $\mathrm{Im}\,\tilde{A}=\mathrm{Im}\,A$, and $\mathrm{Re}\,\tilde{A}=-\mathrm{Re}\,A$ so that $\mathrm{Re}\,A=\frac{1}{2}(\alpha+\beta E)$. It follows that $A-\frac{1}{2}(\alpha+\beta E)$ is real below cuts, imaginary on cuts and bounded by $E\ln E$. It must therefore have the form

$$\sim \text{constant}\,[(E-E_0)(E+E_0)]^{1/2}$$

which we know is nonsense, so the trivial ambiguity is removed.

2.8. Summary on the elastic region

In summary, we have shown that the ambiguity is, at most, discrete in the elastic region, and that it is no more than twofold in all known cases, apart from the trivial ambiguity. Most important, there is a region near threshold where the solution is unique. There is little doubt that extra dynamical input, such as dispersion relations, can reliably handle the problem of sorting out these discrete alternatives and finding the right amplitude from the cross section.

The situation is very different in the inelastic region.

3. The inelastic region

When we move to the energy region where inelastic processes occur, the ambiguity situation alters completely. Unitarity, which in the elastic region directly relates real and imaginary parts of the amplitude for each partial wave, now only provides an *inequality* constraint between them, requiring each partial-wave amplitude to lie inside or upon its unitary circle. It is now almost self-evident that the standard constraints L, R and U allow a continuum of solutions. If all partial waves lie inside their circles and a finite distance from the edge, there will be a whole family of phase functions $\phi(z)$ in (1.2), of limited magnitude but of infinite variety of functional form, which will not move any wave outside its circle. This is almost self-evident and a formal proof has been given by Atkinson *et al* (1973a). All these amplitudes f will thus satisfy L, R and U. The only tricky point concerns the high partial waves of elastic scattering processes where we know from (1.14) that the amplitudes are very small and thus lie very close to the edge of the unitary circle, figure 5(*a*). We must make sure that the transformation ϕ keeps these waves *inside* their circles. Since the high partial waves are controlled by the long-range forces, that is, by the tip of the cut near $z=z_0$, this requirement simply constrains $\phi(z)$ at that point. We have seen that the assumption R ensures that ϕ has no cut nearer to the physical region than z_0.

The sharp change in the ambiguity situation at the inelastic threshold is marked in practical terms by the doubling of the number of parameters in the partial-wave analysis—for each partial wave the phase shift δ_l is joined by an inelasticity parameter η_l as in (1.18). Although the total inelastic cross section is measurable, the η_l are not and they effectively double the amount of information which must be found from the cross section, a demand it cannot meet.

If all the coupled inelastic channels could be fully measured, the unitarity relation would again become an equality, with only one real parameter per process for each partial wave; the situation would be akin to the elastic case, though it is more complicated and is far from being completely understood. However, apart from charge

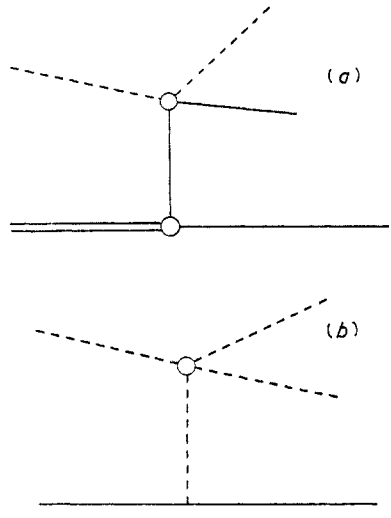

Figure 7. Models of scattering from composite targets.

exchange processes, it is never possible in practice to measure all coupled processes, because only proton targets are available so, for example, $\overline{\mathrm{K}}\mathrm{p} \to \overline{\mathrm{K}}\mathrm{p}$, $\pi\Lambda$, $\pi\Sigma$ and $\eta\Sigma$ are measurable but $\pi\Lambda \to \pi\Lambda$ etc are not. (By using a theoretical model, illustrated in figure 7, for scattering from composite systems like the deuteron (for neutron targets) or the proton (for π, K or Λ targets) the range of processes could be extended at the price of greatly reduced accuracy and some uncertainties in the model itself, but no-one has yet used this approach to measure a complete range of processes.)

Once three-body channels open up, this extension of elastic arguments becomes impossible, even in principle. We shall therefore discuss the problem of the analysis of one-process measurements only. To summarize, the fact that unitarity is only an inequality constraint introduces a continuously infinite set of solutions.

3.1. The size of the ambiguity continuum

Having shown that there is a continuum of scattering amplitudes for each set of perfect data measurements, corresponding to different functions $\phi(z)$ in (1.2), we now want to discuss its size in practical terms. The ambiguity continuum will correspond to an area, or areas, on each partial-wave Argand plot, each point of which is linked to a point in a similar area for each partial wave. The continuum is limited by unitarity,

which requires that these points lie inside the unitary circle for *all* partial waves. In practice, if these areas, or *islands of ambiguity*, are small, the ambiguity will not be serious—even though there is an infinite set of solutions, they are all similar. If, however, the islands of ambiguity cover a significant fraction of the Argand circle for some waves, it is clearly serious; we shall see that this is the usual situation.

The ambiguity continuum is a functional of the phase function $\phi(z)$ and to find its true boundaries is a very difficult problem on which no progress has been made—it involves exploring the whole of an infinite dimensional function space. However, some explorations have been made by choosing particular directions in the function space and looking in those; such investigations will at least find a lower limit on the size of the ambiguity continuum. There have been two main approaches, both involving extensive numerical work, for nobody has thought of a simple analytical way of approaching the problem quantitatively.

The first approach involves choosing particular forms for $\phi(z)$

$$\phi(z)=\sum a_i\phi_i(z) \tag{3.1}$$

and varying their magnitudes a_i until the unitarity limit is reached in some partial wave. This was first used by Bowcock *et al* (1971, summary given in Burkhardt 1973) who made a very limited exploration of this phase ambiguity in pion–nucleon scattering. Even so, they found that substantial changes could be made in the original Saclay phase shifts, involving significant changes in the parameters of some of the resonances. It has since been extended by Pietarinen (1973a,b).

An illustration of the kind of results which emerge is provided in a simple example worked out on a model amlitpude (Bowcock and Hodgson 1972). Taking

$$f(z)=\frac{c}{(a^2+b^2-2abz)^{1/2}} \tag{3.2}$$

and

$$\phi(z)=\lambda[(a^2+b^2-2abz)^{1/2}-(a^2+b^2-2ab)^{1/2}] \tag{3.3}$$

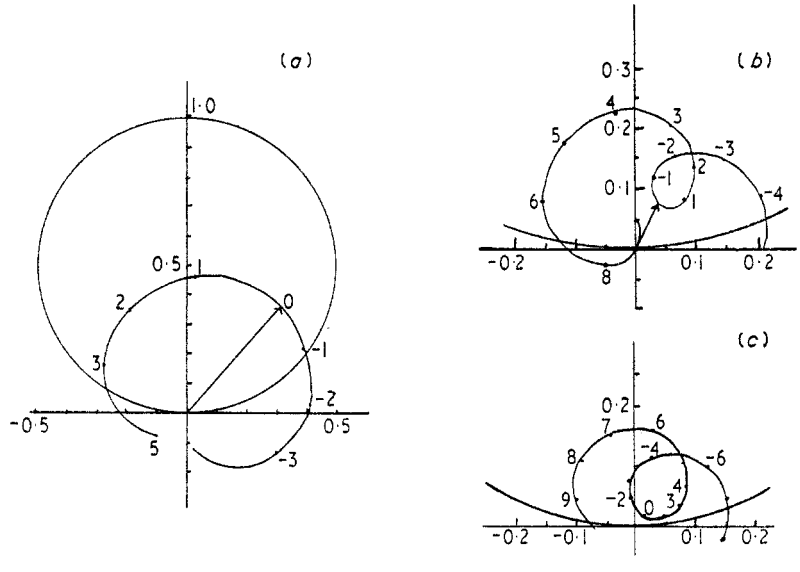

Figure 8. A model one-dimensional continuum ambiguity. (*a*) s wave, (*b*) p wave, (*c*) d wave.

the behaviour of individual partial waves may be plotted as the phase function $\phi(z)$ is altered by modifying the strength of λ. The behaviour of the high partial waves in this case can be investigated analytically and controlled so that they all satisfy the unitarity bounds. Figure 8 shows how the s, p and d waves change as λ is varied in both positive and negative directions.

For values of λ between -2 and 1 all waves lie inside the unitarity circles and corresponding points on the arcs represent allowed partial waves. Two things are clear. One, that there can appear, even in a very simple example like this, a considerable range of ambiguity and two, the fact the partial waves may re-enter the circles indicates the possibility of discrete patches of ambiguity. It should be stressed that the ambiguities form lines only because the phase function $\phi(z)$ has one parameter; generally they will form areas.

3.2. The partial-wave method

A second, more systematic approach has been developed by Atkinson and various co-workers (Atkinson *et al* 1974). It rests on the fact that there is no continuum ambiguity in the elastic region, where the inelasticities are fixed and uses the latter as variables with which to parametrize the continuum—this seems a sensible thing to do. The method is interesting mathematically so we shall describe it in some detail. The essential ideas and their mathematical expression are these:

(i) The partial-wave unitarity equation for an elastic process in the inelastic region can be written

$$S_l = A_l - D_l{}^2 - A_l{}^2 - I_l = 0 \tag{3.4}$$

where the physical scattering amplitude $F_l = A_l + \mathrm{i}D_l$, and I_l represents the inelastic contribution to the absorbtive part, A_l, of the wave; below the inelastic threshold $I_l \equiv 0$. The method parametrizes the ambiguity through the I_l but, as we shall see, there are constraints on them which complicate the situation.

(ii) The differential cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(z) = D^2(z) + A^2(z) \tag{3.5}$$

so that if we know the A_l we can calculate $A(z)$ and so can find

$$R(z) \equiv D^2(z) = \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(z) - A^2(z). \tag{3.6}$$

(iii) We project out the D_l using the complex partial-wave projection

$$D_l = \int_\epsilon \frac{\mathrm{d}z}{2\pi\mathrm{i}} R^{1/2}(z) Q_l(z) \tag{3.7}$$

where the $Q_l(z)$ are the Legendre functions of the second kind; the integral is around an ellipse ϵ in the complex z plane which surrounds the physical region. Provided $R^{1/2}(z)$ is analytic inside ϵ, the ellipse can be shrunk down onto the physical region $-1 < z < 1$. Then the integral only involves

$$Q_l(z + \mathrm{i}\epsilon) - Q_l(z - \mathrm{i}\epsilon) = \mathrm{i}\pi P_l(z) \tag{3.8}$$

and the usual real partial-wave projection emerges from (3.7). There are two reasons for preferring the complex form, one concerns the zeros of $R(z)$, which are the main

source of complication in this method (we shall return to them in a moment). The other, more important reason is that (3.7) automatically preserves the large l behaviour (1.14).

These equations provide the basis for an iterative scheme, in which the I_l are chosen as variables specifying the inelasticity and the A_l and D_l are determined self-consistently by equations (3.4) to (3.7).

Since $R(z)=D(z)^2$ is the square of a function analytic in the ellipse, its zeros should all be even-order zeros (in practice, usually double). However, there is nothing in equation (3.6) to ensure that this will always be so; we therefore impose this constraint at the zeros of R, in the form

$$\left.\begin{aligned} R(A, z_j)&=0 \qquad (3.9a)\\ \frac{\mathrm{d}R}{\mathrm{d}z}(A, z_j)&=0 \qquad (3.9b)\end{aligned}\right\} \quad j=1, 2, \ldots, N$$

where we assume there are N such zeros. The effect of these $2N$ extra constraints, in addition to the

$$S_l(A, I)=0 \qquad (3.10)$$

required by (3.4), is to fix not only the positions of the N zeros z_j but also N of the inelasticities $I_l \simeq \bar{I}_l$, in addition to the A_l of the unconstrained problem. We may choose the $\bar{I}_l$ in various ways, but let us for simplicity assume they are the first N of I_l, $l=0, 1, \ldots, N-1$. Thus as free input parameters with which to explore the ambiguity, we are left with the I_l for $l \geqslant N$.

Notice that, compared with the first method, the space of functions has been reduced to a denumerably infinite set, and that the inelastic unitarity condition is automatically satisfied if we choose our $I_l \leqslant \frac{1}{4}$, except for the $\bar{I}_l$ which must be checked by hand. Notice also that for a polynomial amplitude, where $F_l=0$ for $l>L$, there will be just L zeros and no cuts in the z plane, so that we may choose the ellipse ϵ to include all the zeros and the L I_l are determined. The problem is thus completely constrained and no continuum ambiguity exists, as we should expect.

The method which Atkinson *et al* (1974) have used to solve these nonlinear equations is to linearize them, by expanding to first order in δA_l, δI_l and δz_j and, having chosen the free I_l, to iterate the linearized equations so that, after n iterations,

$$\begin{aligned} &S_l(A^{(n)}, I^{(0)}+\delta I)+\sum_{m=0}^{\infty}\frac{\partial S_l}{\partial A_m}(A^{(0)}, I^0+\delta I)\delta A_m{}^{(n+1)}+\sum_{k=1}^{N}\delta_{l,l_k}\delta I_{l_k}{}^{(n+1)}=0\\ &R(A^{(n)}, z_j{}^{(n)})+\sum_{m=0}^{\infty}\frac{\partial R}{\partial A_m}(A^{(0)}, z_j{}^{(0)})\delta A_m{}^{(n+1)}=0 \qquad (3.11)\\ &R'(A^{(n)}, z_j{}^{(n)})+\sum_{m=0}^{\infty}\frac{\partial R'}{\partial A_m}(A^0, z_j{}^{(0)})\delta A_m{}^{(n+1)}+R''(A^{(0)}, z_j{}^{(0)})\delta z_j{}^{(n+1)}=0\end{aligned}$$

where

$$\delta A_m{}^{(n+1)}=A_m{}^{(n+1)}-A_m{}^{(n)}.$$

Convergence of this Newton–Kantorovich system is both proved and found to result in practice. The proof is needed because the system of equations is infinite, since l goes

up to infinity, and shows that the kernel reduces to unity plus a compact operator on a Banach space, so that an inverse almost always exists. The equations are solved on a computer—practical computing problems are largely concerned with handling the zeros of R, in particular new zeros which may cross the boundary of the ellipse.

The results are of great interest. The first application was to the scattering of alpha particles on alpha particles at 35 MeV, chosen because a good phase-shift analysis existed (Darriulat *et al* 1965). In order to work with perfect data, the differential cross section used was that given exactly by the square of the modulus of the amplitude found in the phase-shift analysis. The object was then to find a set of other amplitudes which give exactly the same cross section. Results are shown in figure 9, each letter corre-

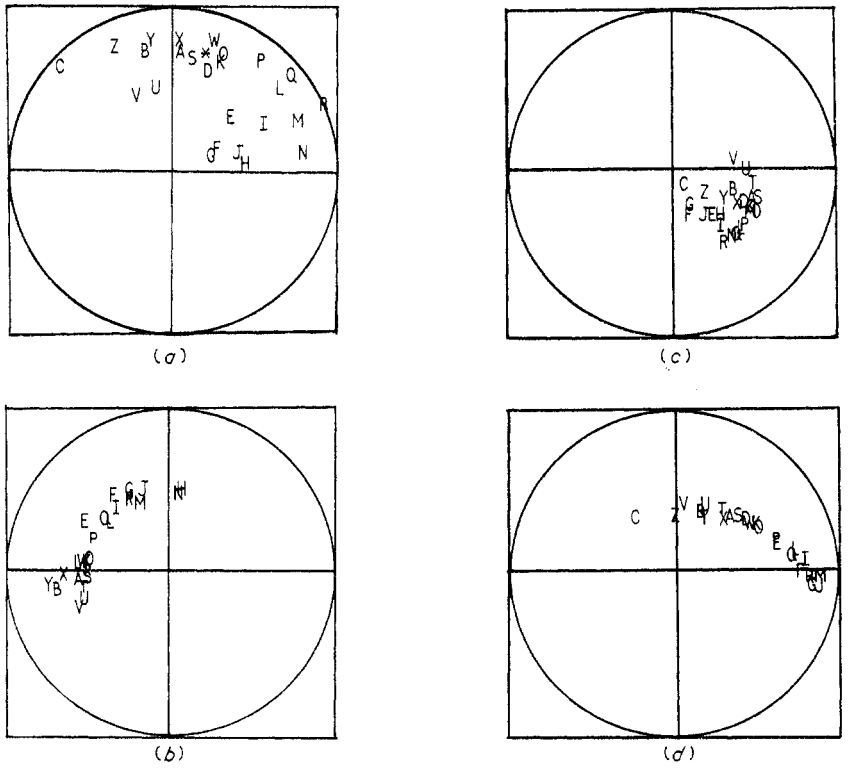

Figure 9. Points in the ambiguity continuum for $\alpha\alpha$ scattering at 35 MeV. (*a*) $L=0$, s wave; (*b*) $L=2$, d wave; (*c*) $L=4$, g wave; (*d*) $L=6$, i wave.

sponding to a different choice of the free I_l. The method yields a discrete set of alternative amplitudes, one for each specific choice of these I_l, but since the I_l can take on a continuous range of values these points are just points in a continuum. The envelope of these points gives a lower bound to the size of the islands of ambiguity in this case. Notice that for the lower partial waves it covers a significant fraction of the Argand circle.

Recently, they have extended this work to the more exciting situation of pion–nucleon scattering. There are some complications which arise from the inclusion of spin in the scattering formalism and we shall discuss these in §7, but the nature of the continuum ambiguity remains the same, so we shall describe the results here. Figure 10 shows the results for various partial waves in π^+p scattering. The starting points at

(*a*)

Point	W (MeV)	T_L
1	1390	411
2	1500	580
3	1572	689
4	1600	745
5	1615	771
6	1629	795
7	1658	846
8	1738	991
9	1900	1305
10	2043	1606
11	2108	1750
12	2150	1845
13	2277	2145

(*b*)

Point	W (MeV)	T_L
1	1160	98
2	1201	150
3	1232	190
4	1258	225
5	1275	247
6	1292	270
7	1320	310
8	1390	411
9	1444	492
10	1470	533
11	1512	599
12	1531	630
13	1658	846
14	1716	950
15	1768	1047
16	1860	1225
17	1900	1305
18	1932	1370
19	1997	1506
20	2235	2043
21	2388	2420

(*c*)

Point	W (MeV)	T_L
1	1768	1047
2	1860	1225
3	1900	1305
4	1932	1370
5	1968	1445
6	2022	1560
7	2043	1606
8	2108	1750
9	2150	1845
10	2235	2042
11	2320	2250
12	2388	2420

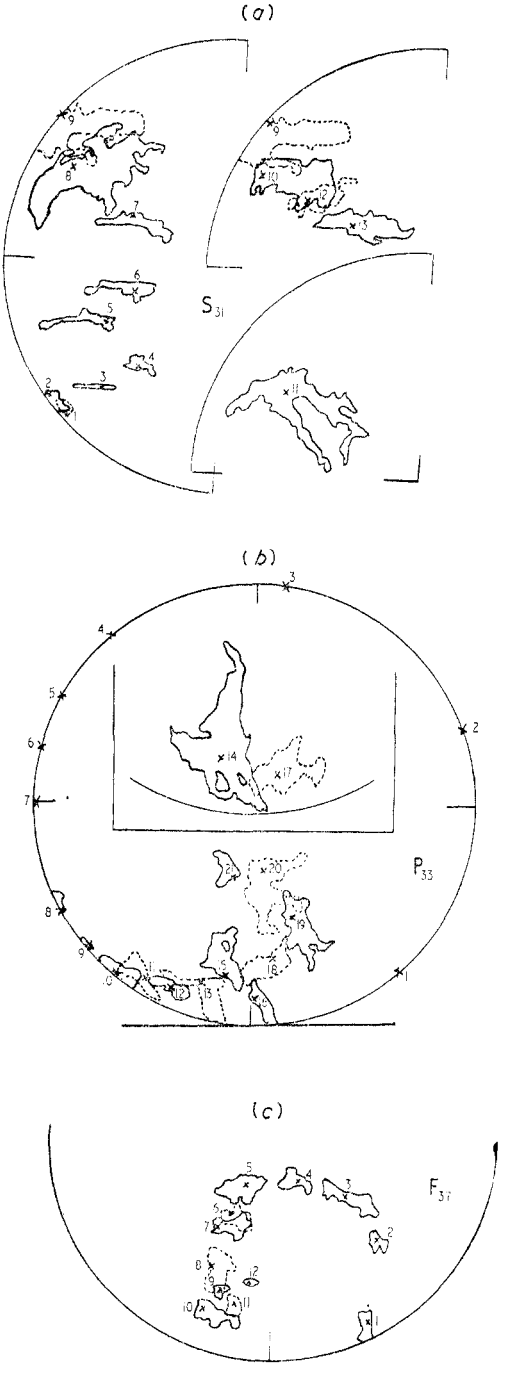

Figure 10. Islands of ambiguity in π^+p scattering in three typical partial waves. The energy of each island is given in the table, and adjacent energies have sometimes been drawn on separate circles to reduce confusion.

each energy are the results of the Saclay phase-shift analysis and the islands of ambiguity indicate the range of different amplitudes which are found to give exactly the same data. The islands are, generally speaking, smaller than in the $\alpha\alpha$ case, and are in some cases remarkably small. However, in some cases where small resonance circles

Table 1

E_R (MeV)	Γ (MeV)	Normal rating	Partial wave, L_{IJ}	Status
1232	115	****	P_{33}	Solid
1650+40	170	****	S_{31}	No significant change
1670+30	230	***	D_{33}	No significant change
1690	?	*	P_{33}	Uncertain
1890+40	250+100	***	F_{35}	No significant change
1900	?	*	S_{31}	Can be removed
1910+80	270+70	***	P_{31}	Big changes possible
1950+20	230+50	****	F_{37}	No significant change

are involved there is the real possibility of suppressing the resonance. Table 1 gives their view on the status of the resonances which have been suggested, and it is only these which were previously uncertain that are in doubt. This is no surprise since it is likely that different phase-shift groups, although they generally use similar methods, will find results in different points of the ambiguity continuum, and so will find less unanimity when the ambiguity is serious. The details are given in Atkinson *et al* (1975).

While most of the resonances seem to survive this investigation in π^+p, in K^+p the situation is more ambiguous. Van Driel (1975) finds large islands of ambiguity which can vary the speed on the Argand plot so that none of the usual structures need be taken seriously as resonances. These results are still preliminary in the sense that further explorations are taking place. It should be stressed that all these explorations are only partial—the local continuum could be larger and there could be other islands of ambiguity as well. Also, the variation of the cross sections allowed by finite errors will increase the size of the ambiguity. The marked variations in size and shape of the islands found at neighbouring energies suggests that the exact form of the cross section may be restrictive and the effects of data errors could be large.

We believe that the reason the ambiguity is less serious in π^+p may be that the strong 'peripheral' resonances in high angular momentum states give a characteristic signature to the angular distribution which is hard to imitate; this is, however, speculation.

4. The search for uniqueness

We have stressed earlier the usefulness of a unique phase-shift analysis procedure. Now we want to see how one might remove the ambiguities inherent in the inelastic region by including more theoretical input information in the analysis. The approaches may be divided into two classes. Most methods used in practice have been based on imposing smooth behaviour on the amplitude as a function of energy—we shall examine the reliability of these later. First, we shall look at how one might hope to obtain a unique single-energy phase-shift analysis.

It is important to recognize that there are many unsound ways to find a unique

solution in a practical analysis. It is only necessary to over-restrict the parametrization employed. For example, the amplitude

$$f(E, z)=\frac{\frac{1}{2}\Gamma}{E_R-E-\frac{1}{2}i\Gamma}\exp(i\phi(z))$$

gives a flat angular distribution for any finite ϕ which is real in the physical region. However, the 'reasonable' assumption that the scattering is 'dominantly s wave' requires $\phi(z)=0$ and removes the continuum ambiguity. The cut-off at $l=L$ in practical analyses will similarly pick a discrete set of points out of the ambiguity continuum (1.2) on an equally doubtful basis. Now we shall look for a well-founded route to uniqueness.

Single-energy analysis is concerned with the behaviour of the amplitude as a function of the cosine of the scattering angle, $\cos\theta=z$, or equivalently as a function of the momentum transfer, $t=-2q^2(1-z)$. In fact, we know quite a lot about this structure; it is usually discussed in terms of the analytic structure of the amplitude as a function of the complex variable z, as in figure 3. We may hope to exploit this information.

For a long time it has been recognized that a knowledge of the long-range forces can help a phase-shift analysis, because these forces dominate the high partial waves. These forces correspond to the nearest singularities in z to the physical region. It is thus worth examining what further information about the z structure of the amplitude might be included, its reliability and whether it would remove the continuum ambiguity. Two sorts of information are worth distinguishing between. The general analytic structure in z, including the position of the singularities, is known reliably; it depends only on kinematic calculations and some general assumptions. The strength of singularities, on the other hand, is only calculable in dynamical models, and these are not generally reliable. The poles, due to particle exchange, are best known; of the cuts only the nearby tips, where few particles can contribute, are at all well understood.

Now to answer these questions for phase-shift analysis. It is fairly clear that the location of singularities alone does not remove the continuum ambiguity. In (1.2) it is only necessary that ϕ be analytic apart from the same cuts, which certainly does not require that $\phi\equiv 0$. Burkhardt (1972) showed that further general constraints, while reducing the allowed $\phi(z)$, do not give uniqueness. If, however, the *strength* of a piece of cut is given, a unique solution follows. The argument illustrates the great strength of the assumption of analyticity; in principle, when an analytic function is known over any finite line, however small, it is known everywhere. Later we shall see that in practice, as commonsense suggests, this result is not nearly so powerful, because very small errors in input can become exponentially magnified in the extrapolation.

4.1. The cut discontinuity

In a region like the physical region $-1<z<1$, with no cut in the amplitude on the real axis, we may write the cross section

$$\sigma(z)\equiv(d\sigma/d\Omega)(z)=f(z)f^*(z^*) \tag{4.1}$$

which, because the complex conjugation of z is double, is an analytic function of z, which we can thus continue, or extrapolate in z. Let us assume, as can be shown, that

we can write on the two sides of the real axis

$$f = f_R + \mathrm{i}f_s \pm \mathrm{i}f_t \tag{4.2}$$

where f_t is the cut strength, or discontinuity, which we assume is known theoretically. Substituting in (4.1) and continuing to just above the cut we find

$$\begin{aligned}\sigma(s) &= (f_R + \mathrm{i}f_s + \mathrm{i}f_t)(f_R - \mathrm{i}f_s + \mathrm{i}f_t) \\ &= f_R^2 + f_s^2 - f_t^2 + 2\mathrm{i}f_R f_t. \end{aligned} \tag{4.3}$$

Knowing f_t, the imaginary part of $\sigma(s)$ fixes f_R and hence the real part fixes f_s, apart from a sign. Thus we know f itself on the piece of cut where we know f_t, and thus in principle everywhere; again an analytic continuation is implied.

Notice that knowing a pole is not enough, since an analytic function is *not* defined by its value at a point. We shall see later that this hopeful route to uniqueness is, in fact, vitiated in practice by the instability of the numerical analytic continuations it implies.

4.2. Other ideas

Other suggestions for including long-range force contributions are based on fixing a few high partial waves. Now if all the f_l for $l > L$ are exactly known they define a cut discontinuity and the above result gives uniqueness. In practice, however, an approximate calculation is made of only the next few waves above the cut-off L. It is clear from Atkinson's argument that an infinite number of I_l are still unconstrained and the continuum exists. How big it is is again a matter for numerical investigation. Present indications are that the reduction is not large.

Alcock and Cottingham (1973) made dynamical calculations of a few higher partial waves, based on a two-particle exchange model. They found that adding them to an existing phase-shift analysis improved the fit. Though it is clear that it is not the fixing of the high partial waves which produced the unique fit, since the analysis was already stable, it is circumstantial support for the result.

It is already clear that, in order to understand the practical possibilities of using analytic structure to eliminate the continuum ambiguity in this way, we must look further at numerical analytic continuation and the way errors in the input from data affect the results. We shall digress to do this before returning to look at how the properties of the amplitude as a function of both energy and angle have been and might be used to reduce ambiguities.

5. Practical difficulties of analytic continuation

It was shown in the last section that the analytic structure of amplitudes as functions of the kinematic variables could provide constraints which, in principle, might remove the ambiguities. If this is so, why has wider use not been made of these constraints to reduce the number of solutions? The answer lies in the fact that the application of the constraints involves in some form the analytic continuation of a function whose values are given only approximately, and that such extrapolations are highly unstable. In this section we want to illustrate this instability and indicate how one can estimate when such extrapolations can be profitably made.

The essence of the problem is that you can find functions with a chosen analytic structure, which are arbitrarily small in one region and arbitrarily large in another, for example the function

$$f(z)=10^9 \exp\left[-10^9(1{\cdot}5-z)^{1/2}\right] \tag{5.1}$$

is totally negligible in the physical region $|z|\leqslant 1$ but very large on the cut $z>1{\cdot}5$ where the square root is pure imaginary. So it is easy to see that, even if one knows the analytic structure of an amplitude, a very small uncertainty in its value in one region can build up rapidly on extrapolation. It is not, of course, an accident that $f(z)$ in (5.1) oscillates rapidly on the cut.

Let us now make this argument more general. We shall choose to take the region on which we measure the function, the known region K, as the real axis $z=x$, and assume we wish to continue analytically in the complex plane up to the unknown region U the line $z=x+\mathrm{i}y$. We assume that the function is analytic in this strip; any region of analyticity can be mapped into such a strip by a suitable change of variable. Now under certain reasonable conditions we can write the function as a Fourier transform

$$f(z)=\int_{-\infty}^{\infty} \tilde{f}(\omega) \exp(-\mathrm{i}\omega z)\, \mathrm{d}\omega \tag{5.2}$$

where

$$\tilde{f}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x) \exp(\mathrm{i}\omega x)\, \mathrm{d}x \tag{5.3}$$

is defined by its value on the real axis. Now (5.2) defines $f(z)$ for complex z as long as the integral converges, so

$$f(x+\mathrm{i}y)=\int_{-\infty}^{\infty} \tilde{f}(\omega) \exp(\omega y) \exp(-\mathrm{i}\omega x)\, \mathrm{d}\omega. \tag{5.4}$$

Note that the high Fourier components are exponentially magnified. Now let us look at the effect of errors on the known region K; these arise in two ways. First, the measurements of $f(x)$ have a finite precision, with errors which are non-zero and, in the case of hadron scattering experiments, rarely less than 1%. Secondly, these measurements only exist at a set of points x_i, which also have errors in their position; this means that $f(x)$ in (5.3) must be defined by interpolation and there will be a largest frequency for which the Fourier component $\tilde{f}(\omega)$ is well determined. We note in (5.4) that it is precisely large ω values that the error builds up fastest on extrapolation.

Now there are two different ways of controlling this problem which have been suggested. The first is simply to require that these exponentially growing factors do not cause trouble by imposing a bound, M say, on $|f(z)|$, or some measure of it, throughout the region of analyticity assumed for the extrapolation. Unfortunately except in a very few special cases such bounds can only be obtained from particular dynamical models and are therefore unreliable. We shall not discuss them further. The second possibility is more subtle; it is aimed at using the fact that the worst behaviour on extrapolation comes from the high Fourier components in (5.2). These correspond to short-distance fluctuations in x, which we do not measure but which also, on general dynamical grounds, we do not believe to be important in hadron phenomena. (The recent discovery of the narrow high-mass mesons called J or ψ particles is a reminder that really new phenomena can upset the most sensible assumptions!) So if we include a cut-off function $\tilde{g}_{\omega_0}(\omega)$ in (5.2) to eliminate the effect of components with $\omega>\omega_0$, and use the convolution theorem for Fourier transforms, we are not extrapolating $f(z)$ but

$$\int_{-\infty}^{\infty} g(z-z')f(z')\, \mathrm{d}z'=\int_{-\infty}^{\infty} \tilde{g}_{\omega_0}(\omega)\tilde{f}(\omega) \exp(-\mathrm{i}\omega z)\, \mathrm{d}\omega. \tag{5.5}$$

Now because of finite resolution in z, this is the sort of function which is measured. Because of the smoothing by g, it contains, of course, less information than $f(z)$ itself but in a controlled and model-independent way. We know then that in an extrapolation to y the maximum error multiplication can be approximately exp $(\omega_0 y)$ so that the errors δf satisfy

$$|\delta f_U| < \exp(\omega_0 y)|\delta f_K|. \tag{5.6}$$

It is clear that all information about the unknown region is lost if the relative error exceeds unity. We therefore have the relation

$$\omega_0 y < \ln(\text{relative experimental error}) = \ln \delta \tag{5.7}$$

so that the minimum smoothing range for an extrapolation to y is

$$\delta x \sim \frac{1}{\omega_0} \gtrsim \frac{y}{\ln \delta}. \tag{5.8}$$

Unfortunately, with the usual errors of a few per cent it turns out that the sort of extrapolations we need imply smoothing over several hundred MeV in energy or distances in cos θ comparable to the size of the physical region. This is clearly useless. The logarithmic dependence on accuracy means that quite unachievable precision is needed to get the required order of magnitude improvement, so that the smoothing distance will be of the order of the energy resolution or a small fraction of the angular range.

We shall now go on to illustrate these points in specific cases. A more thorough general discussion of the methods we shall describe has been given by Ciulli (1972).

5.1. Continuation to average values by weighted dispersion relations

We now want to give a very specific illustration of these ideas using a weighted dispersion relation. The mapping of the usual cut plane (figure 11*a*) in cos $\theta = \zeta$ into the strip in the z plane (figure 11*b*) is achieved by the change of variable

$$\cosh z = (2\zeta - z_0)/z_0. \tag{5.9}$$

Analyticity in the cut ζ plane of any function becomes analyticity in the strip. A weighted dispersion relation consists of applying the convolution integral (5.5) to our function $f(z)$. $g(z)$ is chosen also to be analytic in the strip so that, integrating around the edge of the strip, Cauchy's theorem gives us

$$\int_{C_1} g(z-z')f(z')\,dz' = \int_{C_2} g(z-z')f(z')\,dz' \tag{5.10}$$

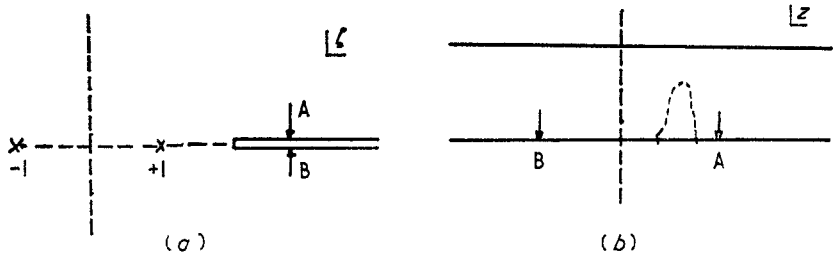

Figure 11. The complex mapping (5.9).

provided $g(z-z')$ goes sufficiently rapidly to 0 at the ends of the strip. To determine an average value of the function in the unknown region, we can choose g to be sharply peaked and positive there, for example

$$g(z-z')=\frac{1}{\alpha\sqrt{\pi}}\exp\left(\frac{-(z'-z)^2}{\alpha^2}\right) \tag{5.11}$$

could be used where the real number α gives the width of the positive peak. From equation (5.10) this average value is given as an integral over the known region of the form

$$\frac{1}{\alpha\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left[-\left(\frac{z'-x}{\alpha}\right)^2+\frac{4\pi^2}{\alpha^2}+4\pi\mathrm{i}\,\frac{(x'-x)}{\alpha}\right]f(x')\,\mathrm{d}x'. \tag{5.12}$$

Notice its peak value has been increased and it has become oscillatory because of the factor $\exp\{[4\pi\mathrm{i}(x'-x)]/\alpha\}$. This rapidly oscillating behaviour emphasizes the high Fourier components, as we have already discussed. So again we see from a slightly different point of view that the extrapolation is highly sensitive to small errors in the data and to the finite distance between the measurement points. The shorter the interval on U over which the average value is required the smaller α must be made and the more accurately the experimental points must be given.

This 'complementarity' can be used in a quantitative way. Going back to the $\cos\theta$ plane, let us suppose the information required is the average value of the function over the nearest part of the cut. If, for example, $z_0=1{\cdot}5$ and we want the average value over the interval of the cut from $\theta=1{\cdot}5$ to $1{\cdot}6$ then equation (5.12) says that the data need to be known to better than 1 part in 10^{32}! Even if the average is taken over a broader interval from, say, $1{\cdot}5$ to $2{\cdot}0$ the data are still required to 1 part in 10^{14}.

The arguments just given were based on a particular choice of weight function. However, it can be shown that any weight function chosen to be positive on one side of the strip must oscillate on the other edge and the narrower the weight function the more rapid the oscillations. Pisut and Presnajder (1970) have found a best weight function optimized with respect to some specific conditions. Using their weight functions instead of the simple form chosen for illustration reduces the accuracy required on the experimental data by a few orders of magnitude but it is still well outside anything that could possibly be measured, unless the smoothing region is made uselessly large. They have also treated the case of continuation from boundary value to boundary value, ie where both K and U are on the cuts, and again found an optimum weight function. The results are similar, oscillations and the resulting high accuracy required appearing again.

5.2. Continuation by truncated polynomial expansions

Our final illustration is chosen to show how truncated polynomial expansions are used in practice to make analytic continuations. The experimental data on the differential cross section will normally be given in the form of the value of the cross section, together with an error bar at a discrete set of angles, θ_i. The size of the error bar will depend on the statistical error and also on geometry and systematic errors. Analytic continuation from this discrete set of points is, of course, impossible even if there were no errors attached and it is necessary to make a smooth interpolation in order to proceed. For the simplest power series expansion it is an advantage to use a variable in which the cut plane is mapped to the inside of the unit circle so that all

points on the cut lie on the circle of convergence of the expansion. This mapping is illustrated in figure 12 and is achieved by the function

$$w=\frac{b-(z_0-z)^{1/2}}{b+(z_0-z)^{1/2}} \qquad \text{where } b\geqslant 1. \tag{5.13}$$

A function analytic in the cut z plane is now analytic inside the circle in the w plane and may be expanded as a power series

$$\sum_{n=0}^{\infty} a_n w^n. \tag{5.14}$$

The data now give values of the function at n points w_i between A and B. Straightforward interpolation of the data may be made by using these n values to determine the coefficients in a truncated series $\Sigma_{n=0}^{N} a_n w^n$. In other words, we are prescribing a

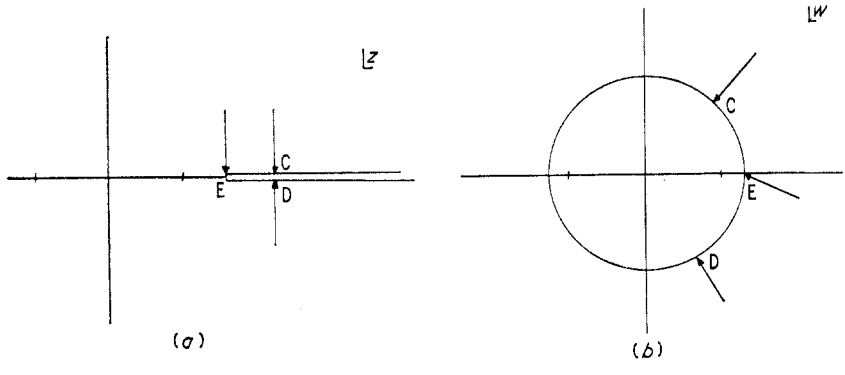

Figure 12. The complex mapping (5.13).

continuous function which passes through the data points, which is smooth in the sense that all derivatives of order N or greater are 0. This prescription is to some extent arbitrary; one can define other mappings with rather different expansions but the principle of truncation remains the same.

The discontinuity across the cross in the z plane at point C is now the difference of the values of the function at CD in the w plane

$$\sum_{n=1}^{N} a_n \exp(\mathrm{i}\chi)-\sum_{n=1}^{N} a_n \exp(-\mathrm{i}\chi)=\sum_{n=1}^{N} a_n 2\mathrm{i}\sin n\chi. \tag{5.15}$$

Thus interpolating the data points by a truncated power series leads to the representation of the cut discontinuities as a truncated Fourier series. The vital question which now arises is the relation between the discontinuity obtained in this way and the true discontinuity, ie between

$$d_N(\chi)=\sum_{n=1}^{N} a_n \sin n\chi \tag{5.16a}$$

and

$$d(\chi)=\sum_{n=1}^{\infty} a_n \sin n\chi. \tag{5.16b}$$

Since the Fourier coefficient a_n is given by

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{d(\chi')}{2\mathrm{i}}\sin n\chi'\,\mathrm{d}\chi' \tag{5.17}$$

we get

$$d_N(\chi) = \int_{-\pi}^{\pi} \sum_{n=1}^{N} \frac{1}{\pi} d(\chi') \sin n\chi' \sin n\chi \, d\chi'$$

$$= \int_{-\pi}^{\pi} \frac{\sin (N+\frac{1}{2})(\chi-\chi')}{2 \sin \frac{1}{2}(\chi-\chi')} d(\chi') \, d\chi' \tag{5.18}$$

$$= \int_{-\pi}^{\pi} K_N(\chi-\chi') d(\chi') \, d\chi'$$

so what we determine using this truncated interpolation is a weighted integral of the true discontinuity where the weight function depends on the number of terms included. This is illustrated in figure 13. The relation to our previous discussion of extrapolation is obvious. As $N \to \infty$ the function tends towards a δ function and the extrapolation becomes exact. However, for the small values of N which are used in practice $K_N(x-x')$ is quite a broad oscillatory function as shown for $N=5$ in figure 13. Thus it is seen that there is no possibility of obtaining the actual value of discontinuity at a point on the cut, only information on an integral over it.

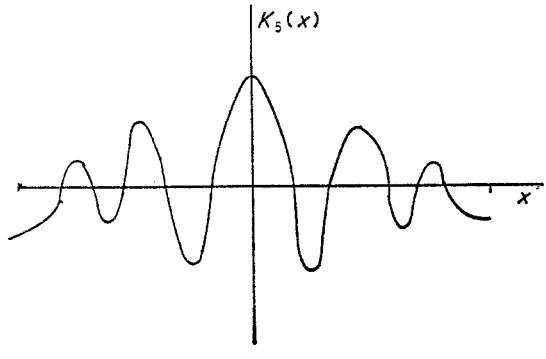

Figure 13. The effect of truncation of an expansion for a kernel. $K_5(x)$ is shown and $K_\infty(x) = \delta(x)$.

Of course, the introduction of error bars to the data only makes matters worse. The argument used here is not confined to a truncated Fourier series; it has more general validity. It can be shown (Cutkosky 1969) that the rate of convergence of the series to be fitted to the data can be improved by mapping the cut plane into an ellipse instead of a circle and the expansion is then made in terms of a series of polynomials in the appropriate variable. However, similar analysis shows that the value of the discontinuity calculated is again a weighted integral over the true discontinuity.

5.3. Summary for analytic extrapolation

It seems, therefore, that any attempt to make practical use of the constraints of analyticity to remove ambiguities in amplitudes by the analytic continuation of numerical information will fail because of its instability, unless model-dependent assumptions are introduced.

6. Energy smoothing and the search for uniqueness

We now return to the search for a sound, reliable method of finding the scattering amplitude from the experimental data. We have seen that, for single-energy analysis, there is a continuum ambiguity and that while, in principle, this may be removed by specifying the discontinuity on a piece of the cut, in practice this is of no use because the numerical analytic continuation involved is not adequately stable, except with data of incredible precision. Most practical analyses find this ambiguity in single-energy fits and resolve it by requiring the amplitude to *vary smoothly with energy*; this is, of course, an unexceptionable requirement, but we must see if it is indeed *enough* to give a unique solution. It is fairly obvious that it is not, as we shall now show.

We have seen that at each energy the phase ambiguity allows a continuum of amplitudes for a given cross section; the true amplitude is a point in this continuum at each energy, giving a point in the island of ambiguity on each Argand plot. This true point moves, of course, smoothly with energy. However, it is self-evident that there is an infinite number of other smooth paths, lying entirely within the islands of ambiguity, which diverge in their different ways from the true amplitude. So smoothness is not enough; one must impose more stringent constraints in order to achieve more. Three methods have been commonly used in practice: the 'shortest-path method', specific analytic parametrizations, and methods based on dispersion relations of various kinds. We shall now look at the soundness of each of them.

6.1. The shortest-path method

This approach has been the most popular in the past (Johnson 1967). As its name implies, it seeks the sequence of single-energy amplitudes which sweep out the shortest path in a space defined by the real and imaginary parts of all the partial-wave amplitudes used. The path length is defined by

$$P=\min_{j_{n+1},j_n}\sum_n\sum_{l=0}^{L} w_l\left|f_l(E_{n+1},j_{n+1})-f_l(E_n,j_n)\right|^2 \qquad (6.1)$$

where $f_l(E_n, j_n)$ is the jth possible solution in the single-energy analysis at energy E_n. The weights w_l may be chosen to discourage fast movements in the small higher partial waves more than in the lower waves.

For a situation with a discrete set of not too many alternative solutions at each energy, this method is an obviously sensible one for ensuring continuity with energy. However, once one recognizes that the discrete set is a fortuitous result of the sharp cut-off at L, which is statistically acceptable because of the finite data errors, and that the underlying continuum ambiguity is continuous, it seems a sweeping assumption that nature always chooses the shortest path. Clearly a shortest path will exist and it will define a unique solution, but there is no reason why it should have any relation to the right amplitude.

Even as a frank dynamical assumption, the shortest path seems dubious. We know of a number of situations where amplitudes have a form which does not correspond to the shortest path in this sense. For example, the Regge pole amplitude

$$A(s,t)=\beta(t)\left(\frac{s}{s_0}\right)^{\alpha(t)}\exp\left(\mathrm{i}\pi\alpha(t)\right) \qquad (6.2)$$

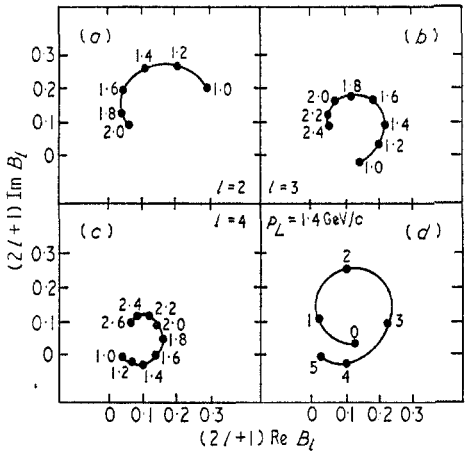

Figure 14. Regge exchange. An example of phase behaviour which is *not* of the shortest-path type.

is a dynamical model, which almost certainly describes quite well some processes at moderate and high energies. It corresponds to partial-wave amplitudes which move with energy in a coherent circular way, as shown in figure 14. The movement is vaguely reminiscent of resonance motion. It is manifest that this is not shortest-path behaviour—if the phase factor exp ($i\pi\alpha(t)$) were replaced by a constant, the movement would be less—but there are good dynamical reasons, based on the analytic structure of the amplitude as a function of energy, for believing that this factor must be there.

This example simply makes the point that a richer phase structure can occur, which increases path length without changing the modulus of the amplitude. It may be argued (it may indeed even be true) that the consistency of the present view of the hadron spectrum shows that the phase structure of the amplitude is as simple as is consistent with general principles. This in turn minimizes the number of resonances.

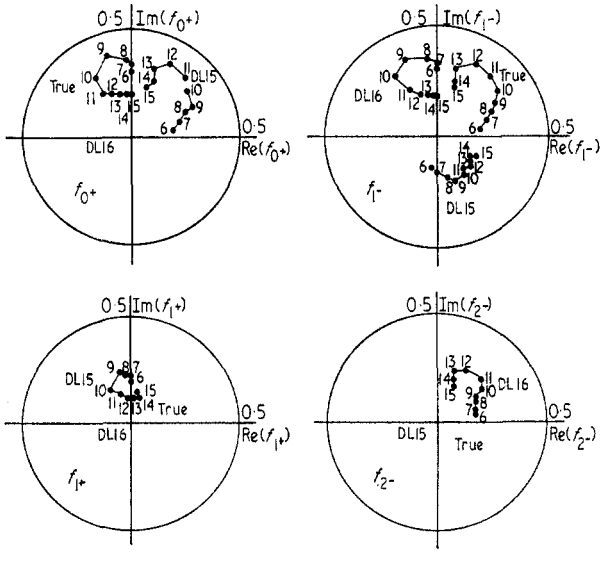

Figure 15. Failure of the shortest-path method in a simulated phase-shift analysis.

However, it is a sophisticated view of the present hadron spectrum to say that it represents a minimum number of resonances; it is also clearly a dynamical assumption and not of the reliability that we traditionally associate with phase-shift analysis.

Recently Dean *et al* (1975) have shown explicitly that this is so by using the shortest-path method in a phase-shift analysis of simulated data generated from a known amplitude; the shortest-path solutions are *not* the correct ones. Figure 15 shows two shortest-path solutions quite different from the true amplitude.

6.2. Explicit analytic parametrizations in energy

Other approaches tend to work by assuming an analytic structure for the amplitude as a function of energy. The simplest of these takes a specific functional form, with a number of free parameters, usually for the partial-wave amplitudes.

An early favourite was to assume that each wave consisted of a slowly varying polynomial background, with or without some Breit–Wigner resonance terms, so that, for example,

$$f_l(E)=q^{2l}\left(A_l+B_lE+\frac{C_l}{E_l-E-\mathrm{i}G_l}\right). \tag{6.3}$$

This parametrization has some dynamical basis in that resonances do have Breit–Wigner shapes in simple cases and, in the absence of any sound neutral method, there is something to be said for it. Unfortunately, in order to get a unique fit in practice, the number of parameters must be restricted beyond what is manifestly sensible, only a few resonances being allowed in each energy range, with tightly restricted backgrounds. In addition, the Breit–Wigner form is reliable only for narrow, isolated resonances and is distorted in different ways by the background; for broader resonances the resonance parameters show energy dependence.

It is nowadays more fashionable to parametrize the partial-wave amplitudes with functions that have a less obvious physical interpretation, in order to avoid making assumptions about which waves are likely to resonate in a particular region. For example, the phase shifts $\delta_l(E)$ and inelasticities $\eta_l(E)$ may be expanded as polynomials in energy, usually of fairly low order. Clearly this sort of approach can remove the ambiguity problems by restricting the parametrization until a unique result is found. The fit will be good as long as the parametrizations stay reasonably close to a point in the ambiguity continuum, but there is no reason why it should be the true point. A neutral parametrization is as much a model for the amplitude as any other; it does not seem a virtue that it is a model without dynamical motivation.

It seems to us that specific parametrizations must gain some benefit from including sound dynamical assumptions in their parametrizations, rather than using *ad hoc* mathematical forms. This result is obvious, once one recognizes that neutral methods cannot remove the continuum ambiguity, which can only be resolved in the context of a particular model.

6.3. Discrete ambiguities and energy dependence

Much has been made in some practical analyses of the resolution of discrete ambiguities of the Gersten (1969) type by imposing continuity requirements on the zeros of the amplitude or otherwise. Barrelet (1972) and co-workers, in particular, have developed procedures for checking the smoothness of energy behaviour in this way. These procedures are useful in helping the analyser to avoid taking the wrong turning,

particularly when a zero of the amplitude comes close to or crosses the physical region, but because they ignore the fact that the true ambiguity is a continuum, they are more relevant to the artificial problem of polynomial amplitudes than to the real situation. However, some of the techniques are of interest and we shall discuss this briefly.

We have seen that the zeros of amplitudes in the complex z ($=\cos\theta$) plane seem to be intrinsically involved in the question of ambiguities. If we knew with certainty the number and location of the zeros this would clearly restrict the forms and multiplicity of allowed amplitudes. It is, however, apparent that this knowledge alone, even if it were available, could not give uniqueness. For if we take an amplitude $f(z)$ which fits the data and has a given number of zeros then $\exp(i\phi(z))f(z)$ has the same zeros, provided that $\phi(z)$ is analytic in the same region as f. Indeed, some of the amplitudes showing continuum ambiguities which have already been constructed possess identical zeros.

Nonetheless, there are ambiguities, such as those suggested by Gersten (1969), or in a more general form by Barrelet (1972), in which one or more complex zeros in the amplitude are replaced by zeros at the complex conjugate points. Such zeros may be useful to consider in trying to follow the behaviour of an amplitude over a range of energies.

Since we believe the amplitude to be analytic in both z and the energy variable, a zero in the complex z plane should vary smoothly as the energy changes. Thus a phase-shift analysis amplitude at one energy may be linked onto one at a neighbouring energy by requiring the positions of the zeros to change only slightly.

In any attempt to make use of this continuity of zeros a distinction can be made between two groups of zeros: the set of zeros which lie rather close to the physical region and those sited further away. The first group of zeros close to the physical region are closely tied to the minima of the differential cross section; if there are reasonably sharp dips the zeros will be nearby in the complex plane. The second set, called 'statistical' zeros by Barrelet are different in that they depend strongly on experimental errors and the number of terms used in the polynomial expansion of the amplitude. We ought, therefore, to concentrate on those stable zeros closer to the physical region.

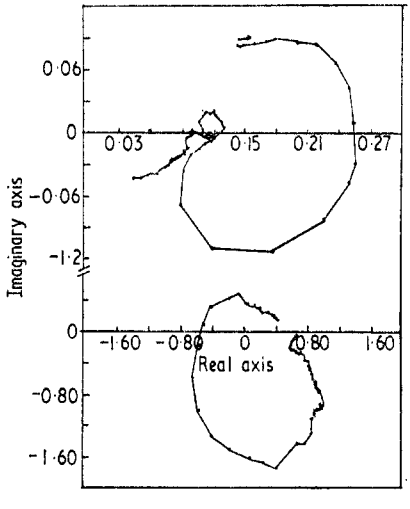

Figure 16. Zeros in the complex plane for $\alpha\alpha$ scattering from 17·6 to 35 MeV.

To get an indication of what the trajectories of such zeros may look like figure 16 shows the trajectories of two of the zeros resulting from phase-shift analysis of $\alpha\alpha$ scattering between 17·6 and 33 MeV. (Due to the symmetry the amplitude is a function only of z^2 and the zeros are therefore plotted in the z^2 plane.)

However, de Roo (1974) found that by conjugating either of these zeros over a range of energy equally acceptable unitary amplitudes resulted. An example of what can happen is shown in figure 17.

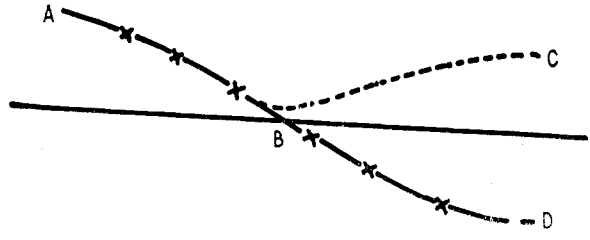

Figure 17. An example of how a zero track may divide.

Over the energy range covered between A and B only one amplitude may be allowed by unitarity but then over BC(BD) zero-conjugation also gives rise to two acceptable unitary amplitudes. In phase-shifting the cross section, if we find that we have suddenly jumped from P to Q at neighbouring energies we would realize that a zero has been conjugated and switch back to the top trajectory.

There are a number of points which should lead us to be cautious with regard to using zeros. One is the method by which they are treated. This is normally done by first performing a PSA and then numerically identifying the zeros of the resulting polynomial. In so far as the coefficients in the phase-shift amplitude are sensitive to the number of partial waves kept, so the location of the zeros will be uncertain. Even the existence of a dip in the differential cross section is not necessarily a guarantee of a zero nearby. In locating the zeros in pion–pion scattering Pennington and Protopopescu (1972) used only a quadratic form for the amplitude even at moderate energies. One can easily verify that the addition of small higher-order terms markedly changes the positions of the zeros (Lutterodt 1973).

As continuity of the zeros is closely related to energy smoothing another important point is how one gets onto a particular zero trajectory to start with. If, as in figure 17, the trajectory breaks smoothly into two branches when the imaginary part of the position of the complex zero is small the choice of up or down cannot be made on an objective basis.

Parametrization of the amplitude directly in terms of a product of zeros has advantages if one believes them to be definitive in specifying the amplitude. Unfortunately, this has the disadvantage of making unitary bounds in individual partial waves difficult to maintain.

In conclusion we should re-emphasize that the tracking of zeros in no way helps the central problem of the continuum ambiguity.

6.4. Energy parametrizations using dispersion relations

Finally, there is a whole class of energy-smoothing procedures based on imposing analytic behaviour in the energy by using dispersion relations of various kinds.

Dispersion relations have been used to define a smooth path through the sets of phase-shift solutions at each energy ever since phase-shift analysis was boldly extended into the inelastic region in the mid-1960's. One of the first such analyses, by Donnachie *et al* (1968), used partial-wave dispersion relations in this way. At that time there was no conscious awareness of the continuum phase ambiguity (the many single-energy solutions found were generally ascribed to defects in the data), and so the problem of finding wrong, alternative solutions was ignored.

We should now require a justification of any given procedure as at least being capable of giving a unique correct solution with ideal data. No such proof exists for any of the currently used procedures. In general, it is hoped that by including sufficient dynamical constraints of a sensible kind the true solution will emerge and, if it is accepted that phase-shift analysis is inevitably model-dependent, this is a sensible type of model to use. If, however, we reserve the title 'phase-shift analysis' for procedures which do not depend on detailed dynamical assumptions, we must be more careful. In particular, we must bear in mind the problems of numerical analytic continuation, discussed in the last section, and not rely on consequences of analytic behaviour which imply the analytic continuation of numerical information.

Very recently, however, Burkhardt and Martin (1975) have tried another approach to the use of analytic properties, which seems to offer hope of a justification of one of the phase-shift procedures—that based on fixed-momentum-transfer dispersion relations. The aim is to get uniqueness from the general analytic structure of the amplitude without any specific analytic extrapolation of numerical information from one region to another. For example, if the amplitude were a polynomial in z, we know that the ambiguity would be at most a discrete one and easy to eliminate in practice. Unfortunately, perhaps, it is *not* a polynomial. However, we know many properties which the amplitude does have which are not built into phase-shift analyses. The aim is to find a set of such assumptions which ensures uniqueness, first with perfect data and then in a practical analysis with error stability taken into account. Burkhardt and Martin looked for such assumptions for perfect data but with a view to numerical stability. In particular, they avoided using elastic unitarity as far as possible because the information it contains, though ensuring uniqueness with perfect data, is rapidly lost under continuation to higher energies. Instead, they aimed at a maximum use of analyticity. Burkhardt (1972) had already shown that analyticity in one variable, z, was not enough (numerical information on the cut discontinuity was needed), so they looked at the amplitude as a function of two variables. In practice, this implies fitting data at all energies and angles at the same time (which is a daunting computational prospect, but one which is already being faced).

Their results are roughly these. First, they show that imposing the full analytic structure in both energy and momentum transfer, together with the modulus of the amplitude throughout the physical regions of all three crossed channels (see figure 2), still allows a continuum of amplitudes $\tilde{A}$, whose ratio, R, to the true amplitude has the form, for equal masses of the scattered particles,

$$R \equiv \frac{\tilde{A}}{A} = \pm \frac{M(s,t,u) + [(4m^2-s)(4m^2-t)(4m^2-u)]^{1/2}}{M(s,t,u) - [(4m^2-s)(4m^2-t)(4m^2-u)]^{1/2}} \tag{6.4}$$

where $M(s, t, u)$ is an arbitrary real meromorphic function (it only has poles as singularities in the product of the finite complex planes of the two independent variables, but is otherwise free). Notice that the ratio is such that, because the square root is

imaginary in the physical regions and real outside, $|R|=1$ in the physical regions. It also ensures that the amplitudes continue to satisfy the reality condition

$$A^*(s^*, t^*, u^*)=A(s, t, u) \tag{6.5}$$

as long as M does, as well as the analytic assumptions. In short, this result shows that a little bit of analyticity is not enough and even the fullest realistic restriction on the location of singularities leaves one far from uniqueness.

Secondly, they imposed, in addition to the analytic properties, positivity of the imaginary part of the amplitude in the 'strip' beyond the forward direction $0<t<4\mu^2$, where 2μ is the lightest exchanged mass, as in pion–nucleon scattering. To cover this case they only assumed data, and thus $|R|=1$, in the s- and u-channel physical regions. Under these circumstances the general form for M is

$$M=\frac{A(t)+sB(t)+s^2C(t)}{\alpha(t)+s\beta(t)} \tag{6.6}$$

which contains five arbitrary functions of the variable t, analytic in the full domain of t at fixed s. Positivity follows from the unitary bound on the partial-wave amplitudes, but is a weaker assumption than that. It has the practical advantage of being a linear property, unlike unitarity. It is powerful because it restricts the number of zeros that the amplitude can have, and the phase ambiguity is closely linked to the zeros of the amplitude. If there are no zeros of A, you can write a dispersion relation for $\ln A$, whose real part is the modulus and whose imaginary part is the phase of A, and which determines one in terms of the other.

Still, you see that something further is needed to give $R=1$ as the general form and uniqueness. Let us mention two cases. First, in the case with complete crossing symmetry so that the scattering in all three channels is the same, (6.6) reduces to

$$M=a+b(s^2+t^2+u^2)+cstu$$

which has just three real constants to be found. $\pi^0\pi^0\to\pi^0\pi^0$ is an example of such a simple system. The simplest way to fix these is to include the data on the total cross section and to use the forward dispersion relation (2.28) and the optical theorem. These give you the phase in the forward direction, requiring $a=b=c=0$ and $R=1$.

In the more general case of pion–nucleon scattering you need something more. This can be supplied by the general constraint that elastic unitarity (here it is at last) imposes at threshold, which requires the imaginary part of the amplitude to be smaller than the real part in a particular way. Since the unknown functions in (6.6) are independent of the energy, this constraint at the s- and u-channel thresholds proves enough to eliminate the ambiguity.

The first impression of these results is negative. The range of restrictions and the practical implications are both forbidding. The construction of a formalism which embodies these requirements will not be easy. The error analysis implied in checking its stability with respect to finite errors in the input data, its extrapolation to high energies, and the truncation of the expansions involved, may be possible only by simulation techniques. The simultaneous fitting of data at different angles and at all energies will provide massive computational problems. Nonetheless the situation is not altogether without hope, as we shall see later. It is, of course, possible that a different set of less forbidding conditions will be found in the future.

6.5. Partial-wave dispersion relations

The earliest use of dispersion relations in phase-shift analysis was the inclusion of partial-wave dispersion relations in the analysis of Donnachie *et al* (1968). It followed on extensive phenomenological applications by Hamilton and various collaborators which are reviewed in Hamilton (1967). The cut structure of the partial-wave amplitude $f_l(s)/q^{2l}$ is shown in figure 18. The real and imaginary parts on the right-hand physical cut can be taken from phase-shift analysis up to its highest energy, and above that a model discontinuity is used. The left-hand cuts correspond to crossed channel effects due to the exchange of particles. The nearby cuts can be approximately calculated, but the rest of the cut must be parametrized. The dispersion relations are evaluated at points in the phase-shift region and compared with the real parts from the phase-shift analysis. The threshold factor q^{2l} has the effect of reducing the contributions from the distant parts of the integral in the high partial waves. Those sets of

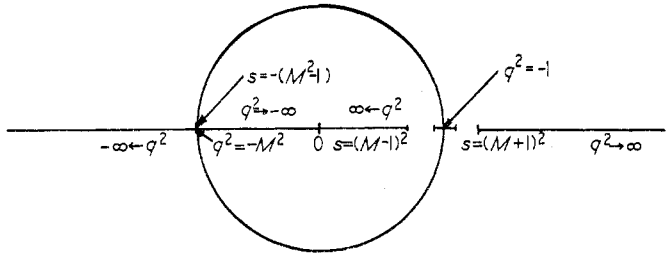

Figure 18. Singularity structure of a partial-wave amplitude.

solutions at each energy were selected which satisfied the dispersion relations mos nearly. The method worked to produce some of the first information on the hadron resonances, which now forms such an important part of our picture of elementary particle physics.

It was clear from the beginning, however, that there were problems of a deep as well as a practical kind. Practically the method was extremely laborious, since it involved, in principle, testing all possible combinations of alternative solutions at every energy used in the analysis. Later, the CERN group abandoned it as a method of primary smoothing and used the shortest-path method, which took into account only two energies at once, when they found that this gave much the same results; they retained the dispersion relations as a way of producing an analytically smoothed amplitude. On a deeper level, it did not prove possible to find solutions to the dispersion relations which gave a reasonable fit to the data if the number of parameters was small enough for the fit to be unique. This was partly due to inconsistencies in the data but is exactly what one expects to find in a continuum ambiguity situation, where the parametrization must be overconstrained to get uniqueness.

The real difficulty with the partial-wave dispersion relations lies in their left-hand cuts, which involve integrals over the amplitude in the spectral regions, where two kinematic variables are positive. Once one gets to high energies, and thus far from the physical regions, these left-hand cuts can only be predicted from models or parametrized in a general way, which in practice has to be too tightly restricted. The early hope was that only the nearby singularities, which can be calculated at least roughly, were important but this has been shown to be untrue by the failure of models based on

that hope to fit the data, and by the discovery of the rise of the Regge trajectories of particles to high angular momentum values.

6.6. Fixed-momentum-transfer dispersion relations

Finally, we come to the use of fixed-t dispersion relations in phase-shift analysis of data in the s and u channels. Such dispersion relations have long been used as part of the S-matrix approach to dynamical calculations and, indeed, were combined with the assumption of Regge-pole dominance of asymptotic behaviour (as in Phillips 1966) to give the so-called finite energy sum rules which formed the basis of the self-consistent approach to hadron dynamics called duality.

Their use in phenomenology is also well established. They have the great advantage that, for small negative t at least, the line of integration lies entirely within the physical region, except for a small region between threshold and the lowest energy for which the t value is physical in the s and u channels. In this region the imaginary part is calculated by extrapolating the truncated partial-wave expansion of the low-energy phase-shift analysis. More recently, Höhler and Jakob (1974), for example, have used such relations as a constraint to distinguish between different phase-shift solutions. However, in none of these cases is the dispersion relation used to try to remove the continuum ambiguity, since in each case the phase-shift analysis has already found a specific result. The dispersion relation is then simply a test of the consistency of these results.

In the last few years there has been an effort, particularly by Pietarinen (1972), aimed at developing a procedure for finding the amplitude from the data, in which fixed-t dispersion relations are the means of ensuring smooth energy variation based on analyticity. In essence, the forward amplitude is fixed by the forward ($t=0$) dispersion relation and the analysis extended in small steps to negative t. At each step, the modulus of the amplitude is fixed by the data and the dispersion relation constraint is applied. The continuation in t is by means of a truncated partial-wave expansion for Im A. Thus data at all energies and a range of t are fitted simultaneously.

This method can be related directly to the result of Burkhardt and Martin (1975), since Martin (1966) has shown that if an amplitude satisfies a fixed-t dispersion relation for a range of negative t, $-T<t<0$, and if the partial-wave expansion in z_s implied converges and satisfies both positivity and a bound $|f_l|<1$, then the amplitude is analytic in s in the strip $-T<t<4m^2$, and further satisfies the Froissart bound at high energies $\sigma_{\text{tot}}<c\ln^2 s$. With the addition of the threshold constraint, these are sufficient conditions for the uniqueness theorem to hold.

More analysis of Pietarinen's procedure is necessary before one can decide whether it really conforms to all these constraints. The analytic continuation in z_t is truncated—one must show that the amplitude errors resulting are stable and sufficiently small under this truncation. Bowcock and Ng (1970) have shown that in certain circumstances instabilities are a problem. The data on differential and total cross sections cannot be measured up to arbitrarily high energies; the stability under errors in these extrapolations must be checked. This could be a source of difficulty since the restricted range of t used so far, $|t|<1$ GeV^2, does not specifically include the quite rich structure of secondary peaks in the high-energy data but which are clearly part of the two-variable information that the uniqueness proof implies. Pietarinen has begun an error analysis of his method, which is also discussed in detail by Hamilton and Petersen (1975) in their review of modern developments along these lines. It may be that it can

be completed analytically but one may need to fall back on testing the method on simulated data, along the lines developed by Burkhardt *et al* (1974). This in itself would be a massive task to do properly and it must be stated, once again and with regret, that we are still without an authenticated method of inelastic phase-shift analysis.

7. The complications of spin

It is with the process of deducing amplitudes from data relating to elementary particles that we are mainly concerned and since the normal target is a nucleon the additional complications due to spin must be looked at. So in this section we shall extend the results of the previous sections to the case of spin 0–spin $\frac{1}{2}$ scattering. This will therefore cover the πN and KN systems from which phase-shift analyses have extracted many of the well known particle resonances. The nucleon–nucleon system is far more involved, with five independent amplitudes for each charge state and will not be treated here.

Lorentz invariance implies that a spin 0–spin $\frac{1}{2}$ system may be described by two complex-valued functions of energy $f_1(s, z)$ and $f_2(s, z)$. However, although we have more amplitudes there are more experimental quantities to be measured. We may list these as the differential cross section

$$\frac{d\sigma}{d\Omega} = |f_1 + zf_2|^2 + |f_2|^2(1-z^2),$$

the polarization

$$\frac{d\sigma}{d\Omega} P = (1-z^2)^{1/2}\, 2\mathrm{Im}\,[(f_1 + zf_2)f_2^*],$$

and double-scattering experiments also determine the quantities

$$R = 2(1-z^2)^{1/2}\,\mathrm{Re}\,[(f_1 + zf_2)f_2^*]$$
$$A = |f_1 + zf_2|^2 - |f_2|^2(1-z^2).$$

It is worth noting here that it is possible to use an elegant formulation, due to Barrelet, to combine the two amplitudes f_1 and f_2 into a single analytic function. Although this is convenient for investigating continuum ambiguities when only the transformation (T) is considered, there are difficulties when alternative transformations (X) are to be discussed. We shall therefore treat f_1 and f_2 as separate functions.

Since each of these measurements is a bi-linear form of the amplitude it is clear that by themselves they can only determine f_1 and f_2 up to a common phase factor which may be a function of both s and z, ie the replacements

$$f_1 \to f_1 \exp(i\phi(s, z)), \qquad f_2 \to f_2 \exp(i\phi(s, z)) \quad \text{(T)} \tag{7.1}$$

leave the experimental quantities unchanged.

Unitarity may be imposed through partial-wave decompositions

$$f_1 = \sum (f_{l+1/2}P_{l+1}'(z) - f_{l-1/2}P_{l-1}'(z))$$
$$f_2 = \sum (f_{l-1/2} - f_{l+1/2})P_l'(z)$$

where

$$f_{l\pm1/2} = \frac{1}{2iq}\,[\eta_{l\pm1/2} \exp(2i\delta_{l\pm1/2}) - 1]$$

with $\eta_{l\pm1/2} = 1$ in the elastic region and $0 \leqslant \eta_{l\pm1/2} \leqslant 1$ above the inelastic threshold.

For the elastic region a generalization of the contraction mapping method used in the spinless case leads to a similar result. If $d\sigma/d\Omega$ and P satisfy certain inequalities then it may be shown that there exist unique amplitudes $f_1(z)$ and $f_2(z)$ which give rise to the measured values of $d\sigma/d\Omega$, P, A and R. So, at energies below the inelastic threshold, and in particular at very low kinetic energies where these inequalities are met, we can be confident that accurate experiments lead to the determination of reliable amplitudes.

For energies above the inelastic threshold there is again a continuum set of amplitudes corresponding to the same measurable quantities. The nature and extent of these depend on whether all the possible experimental quantities are measured or, as is more common, just $d\sigma/d\Omega$ and P. If all are measured and used in the amplitude analysis then as in the spinless case, apart from the trivial ambiguity, the most general type is that given in (7.1).

However, if no double-scattering experiments are performed, so that only $d\sigma/d\Omega$ and P are used it can be seen that the data are also invariant under the transformation

$$[\mathrm{Re}\, f_1 + \exp(i\theta)\, \mathrm{Re}\, f_2] \to \exp(i\chi(z))\, [\mathrm{Re}\, f_1 + \exp(i\theta)\, \mathrm{Re}\, f_2]$$

and

$$[\mathrm{Im}\, f_1 + \exp(i\theta)\, \mathrm{Im}\, f_2] \to \exp(i\chi(z))\, [\mathrm{Im}\, f_1 + \exp(i\theta)\, \mathrm{Im}\, f_2].$$

Thus the allowed amplitudes will be even more numerous and greater ambiguities will occur. This transformation is a generalization of those that can be constructed by taking a product of any two of the following transformations:

$f_1(z) \to f_1^*(z)$, $\quad f_2(z) \to -f_2^*(z)$: reflection

$f_1(z) \to z f_1 - (1-z^2)^{1/2} f_2(z)$, $\quad f_2(z) \to -(1-z^2)^{1/2} f_1(z) - z f_2(z)$: Minami

$f_1(z) \to f_1(z)$, $\quad f_2(z) \to -f_2(z)$: Yang.

Each of these keeps the differential cross section unchanged but changes the polarization. A product of any two therefore leaves $d\sigma/d\Omega$ and P unchanged. It can easily be verified that particular choices of constant $\chi(z)$ give rise to these transformations (Bowcock *et al* 1971).

Calculations on the extent of the second type of ambiguity where only $d\sigma/d\Omega$ and P are used indicate that the ambiguity lines in this case are somewhat different in character, tending not to lie on such circular arcs as in the simple phase transformation (7.1). In fact, these results question the existence of the resonances that are not so well established.

8. Conclusions

We hope we have shown that while phase-shift analysis in the elastic region is a well defined and reliable procedure which will yield the scattering amplitude from the experimental data with only a few discrete alternative solutions at most, no reliable inelastic phase-shift analysis has been convincingly justified. Numerical studies show that the uncertainties in the amplitude produced by the continuum ambiguity are serious, though they are less dramatic in the crucial case of pion–nucleon scattering than elsewhere, in so far as they have been explored. While many of the methods of energy smoothing which have been used to find a unique amplitude are quite arbitrary, there now exists a possible basis for a justification of the method of multi-energy

analysis based on fixed-momentum-transfer dispersion relations as a sound route to the true amplitude.

The beautifully consistent picture of the baryon spectrum which has emerged from the phase-shift analyses of the last ten years remains something of a surprise, in view of the continuum ambiguity problem. It lends support to the numerical observation just mentioned, that the continua are less serious in this process in the energy range so far covered. There seems little doubt, however, that any detailed study of weaker resonances, daughter states and other subtler features of the scattering amplitude should await a more clearly reliable procedure than has been used in the past.

Acknowledgments

We should like to acknowledge many helpful discussions with André Martin and David Atkinson.

References

ALCOCK J and COTTINGHAM W N 1973 *Nucl. Phys.* B **62** 533–8
ATKINSON D, HEEMSKERK A C and SWIERSTRA S D 1975 *Preprint* University of Groningen
ATKINSON D, JOHNSON P W, MEHTA N and DE ROO M 1973a *Nucl. Phys.* B **55** 125–42
ATKINSON D, JOHNSON P W and WARNOCK R L 1972 *Commun. Math. Phys.* **28** 133–58
—— 1973b *Commun. Math. Phys.* **33** 221–42
ATKINSON D and KOK L P 1974 *Z. Phys.* **266** 373–81
ATKINSON D, KOK L P, DE ROO M and JOHNSON P W 1974 *Nucl. Phys.* B **77** 109–33
BARRELET E 1972 *Nuovo Cim.* A **8** 331–71
BERENDS F A and RUYSENAARS S N M 1973 *Nucl. Phys.* B **56** 507–24
BOWCOCK J E, COTTINGHAM W N and NG P 1971 *Proc. Int. Conf. on Elementary Particles, Amsterdam* (Amsterdam: North-Holland) Contributed Paper, summarized in Burkhardt (1973)
BOWCOCK J E and HODGSON D C 1972 *University of Birmingham Report*
BOWCOCK J E and NG P H 1970 *Nucl. Phys.* B **25** 93–108
BURKHARDT H 1969 *Dispersion Relation Dynamics* (Amsterdam: North-Holland) pp66–79, 92–107
—— 1972 *Nuovo Cim.* A **10** 379–88
—— *Proc. Triangle Seminar* ed D Krupa (Bratislava: Slovak Academy of Sciences); *CERN Preprint* TH 1784
BURKHARDT H, MCCAULEY G P, SCHONFELDER L and YORK M J 1974 *University of Birmingham Report*
BURKHARDT H and MARTIN A 1975 *CERN Preprint* TH 2017
CIULLI S 1972 *Strong Interaction Physics, Lecture Notes in Physics* vol 17, ed W Rühl and A Vancura (Berlin, Heidelberg, New York: Springer-Verlag), see also 1975 *Phys. Lett.* **17C** 133
CORNILLE H and DROUFFE J M 1974 *Nuovo Cim.* A **20** 401–37
CRICHTON J H 1966 *Nuovo Cim.* A **45** 256–8
CUTKOSKY R E 1969 *Ann. Phys., NY* **54** 350–78
DARRIULAT *et al* 1965 *Phys. Rev.* **137** B315–25
DEAN N W, JENSEN W F and LONG W F 1975 *Phys. Lett.* **53B** 462–6
DONNACHIE A 1973 *Rep. Prog. Phys.* **38** 695–753
DONNACHIE A, KIRSOPP R and LOVELACE C 1968 *Phys. Lett.* **26B** 161–5
GERSTEN A 1969 *Nucl. Phys.* B **12** 537–48
GLAUBER R J 1959 *Lectures in Theoretical Physics* vol 1, ed W E Brittin and L G Dunham (New York: Interscience)
GOLDBERGER M L, LEWIS H W and WATSON K M 1963 *Phys. Rev.* **132** 2764–87

HAMILTON J 1967 *High Energy Physics* vol 1, ed E H S Burhop (London: Academic Press)
HAMILTON J and PETERSEN J L 1975 *New Developments in Dispersion Theory, Lectures at Nordita, Copenhagen*
HANBURY-BROWN R and TWISS R Q 1956 *Nature, Lond.* **177** 27–30
HÖHLER G and JAKOB H P 1974 *Z. Phys.* **268** 75–84
ITZYKSON C and MARTIN A 1973 *Nuovo Cim.* A **17** 245–87
JOHNSON C H 1967 *PhD Thesis, Lawrence Radiation Laboratory Report* UCRL-17683
KLEPIKOV N P 1965 *Sov. Phys.–JETP* **14** 846–50
LUTTERODT C H 1973 *Thesis* University of Birmingham
MARTIN A 1966 *Nuovo Cim.* A **42** 930–53
—— 1969 *Nuovo Cim.* A **59** 131–52
NEWTON R G 1968 *J. Math. Phys.* **9** 2050–5
PENNINGTON M R and PROTOPOPESCU S D 1972 *Phys. Lett.* **40B** 105–8
PHILLIPS R J N 1966 *Nuovo Cim.* A **45** 245–8
PIETARINEN E 1972 *Nuovo Cim.* A **12** 522–31
—— 1973a *Nucl. Phys.* B **55** 541–50
—— 1973b *Proc. Triangle Seminar* ed D Krupa (Bratislava: Slovak Academy of Sciences)
PISUT J and PRESNAJDER P 1970 *Nucl. Phys.* B **22** 365–84
DE ROO M 1974 *Thesis* University of Groningen
VAN DRIEL 1975 *University of Groningen Report*