Rep. Prog. Phys., Vol. 41, 1978. Printed in Great Britain

# Magnetic monopoles in gauge field theories

P GODDARD† and D I OLIVE‡§

† Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK
‡ CERN, Geneva, Switzerland

**Abstract**

An account is given of the new insight into the theory of magnetic monopoles originating from the work of 't Hooft and Polyakov. Their magnetic monopole, associated with the conventional electromagnetic gauge group U(1), occurs as a finite-energy smooth soliton solution to an SU(2) gauge theory. A precise picture of its internal structure, the values of its magnetic charge and its mass are obtained. These new developments bring together previously unrelated fields of study, namely the Dirac monopole (with point structure) and the Sine-Gordon soliton in two-dimensional space–time.

Properties of more general monopoles, associated with large gauge groups now thought to be relevant in physics, are discussed. Particular attention is paid to topological properties. Based on this new viewpoint, conjectures can be made about a future quantum theory of monopoles.

This review was received in January 1978.

§ Present address: Department of Physics, Imperial College of Science and Technology, Prince Consort Road, London SW7 2AZ, UK.

0034–4885/78/0009–1357 $05.00 

**Contents**

## 1. Introduction

Two reasons can be given for the recent upsurge of interest in the theory of magnetic monopoles. One was the possible experimental observation of such an object which was unfortunately not confirmed by closer analysis. The second and more profound reason is that recent developments in the understanding of monopoles has brought together a number of apparently unrelated ideas. Such a synthesis is always attractive to theoreticians, and the new insights gained hold promise of further progress which could be of enormous importance in, for example, understanding the inter-relationship between the strong and weak interactions. Even if this last hope is not fulfilled, one is learning more about the structure of non-Abelian gauge theories and how they should be interpreted, at least at the classical level. Although many of the more important effects which are anticipated are thought to be intrinsically quantum-mechanical (e.g. colour confinement), it is reasonable to expect that the experience gained at the classical level will be an important aid in understanding the quantum mechanics.

The organisation of our review is as follows. In §2 we go back to the beginning of the modern theory of magnetic monopoles by giving an account of the way they were introduced by Dirac (1931). Our treatment will be more detailed in some respects; we will pay particular attention to some interesting lessons that can be learnt from his work, namely (*a*) that a gauge-invariant field theory can accommodate line singularities in the vector potential quite consistently, a point which has been largely ignored in quantum field theory; (*b*) that in the presence of a monopole a term has to be added to the orbital angular momentum in order to obtain a conserved quantity; and (*c*) that the possibility of monopoles effectively modifies the global structure of the electromagnetic gauge group so that it is a compact group, U(1). This last statement, which is just equivalent to the quantisation of electric charge, could also be understood if the electric charge operator, $Q$, were a generator of a non-Abelian compact gauge group which has been spontaneously broken by an asymmetric vacuum to the electromagnetic U(1) subgroup generated by $Q$.

In §3, we start a superficially unrelated line of discussion by reviewing briefly the existence of soliton solutions to the Sine-Gordon equation in a two-dimensional space–time, and the intriguing relationship of this theory to the Thirring model at the quantum-mechanical level. In attempting to generalise this phenomenon to four-dimensional space–time one is led quite naturally to spontaneously broken gauge theories once again.

Thus, remarkably, these two independent discussions converge upon the class of theories, currently believed to be relevant to the strong, weak and electromagnetic interactions of elementary particles. The simplest theory which illustrates how the ideas explained in §§2 and 3 have been synthesised is that originally suggested by 't Hooft (1974) and Polyakov (1974). In §4 we discuss this theory and, in particular, the solution to which 't Hooft and Polyakov drew attention. This has a natural interpretation as an extended object, which at large distances provides an explicit model of a Dirac monopole, but whose short-distance structure has been modified so that it has finite energy. We explain carefully how electromagnetism is embedded in the theory and the way topological properties of the asymptotic values of the Higgs

field lead to the Dirac quantisation condition for the magnetic charge. We derive the lower bound on the mass of any solution to this theory in terms of its electric and magnetic charges, first found by Bogomolny, and discuss the exact solution found by Prasad and Sommerfield in the limit in which the self-interaction of the Higgs field vanishes. Dyons and candidates for the magnetic current inside the monopole are also discussed.

Armed with the insight gained from the 't Hooft–Polyakov monopole we begin to consider its possible generalisations in which, for example, the exact symmetry group $H$, remaining after spontaneous symmetry breaking, may itself be non-Abelian. First we briefly review the formalism for a theory with a general gauge symmetry group, $G$, assumed compact and connected. The rest of §5 is devoted to studying the large-scale, or *macroscopic*, properties of monopoles. Here it is $H$ rather than $G$ that plays the crucial role. We begin by describing these properties in terms of the way the Higgs field realises its boundary conditions asymptotically, i.e. outside the monopole. In any finite-energy solution the Higgs field must approach a minimum of the potential function, $V$, describing its self-interaction, as we approach infinity along any radial direction. In this way we associate with each direction in space a minimum of $V$, and we may think of this association as defining a map from the unit sphere in three-dimensional space to the set of minima of $V$. The topological characteristics of this map define topological conservation laws for these solutions generalising those we met in §3. The appropriate topological concept is that of homotopy; relevant aspects of homotopy theory are outlined in this section and summarised more precisely in appendix 1. Results from homotopy theory enable us to take the description of the topological conservation laws in terms of the way the Higgs field approaches the set of minima of $V$ and rephrase it in terms of topological properties of $H$. Using a non-Abelian version of Stokes' theorem we then go further and re-express the topological quantum number entirely in terms of $H$ gauge fields. This expression is then used to derive generalisations of Dirac's quantisation condition firstly assuming that $H$ has, at least locally, a U(1) factor which may be identified with electromagnetism and, secondly, assuming a generalised inverse square law form for the field tensor.

In a sense, the theme of §5 is the progressive relegation of the Higgs field and broken symmetry group $G$ in the discussion of the macroscopic properties of solutions. We end with everything expressed in terms of $H$ and we are then able to relate the Higgs field formalism we have used to that of Wu and Yang, in which only the exact symmetry group, $H$, appears.

In §6 we consider the *microscopic* properties of monopoles, their internal structure. Here the statements we can make are less complete though it seems that the Higgs field and full symmetry group $G$ now play a crucial role in obtaining finite-energy solutions. As in the 't Hooft–Polyakov case the solutions that have been found and analysed tend to possess certain spherical symmetry properties which we discuss. We review the corresponding possibilities for $G=\mathrm{SU}(3)$ discussing both $H=\mathrm{U}(2)$ and $H=\mathrm{SO}(3)$. The features of these SU(3) solutions are set in a more general context by quantisation conditions for spherically symmetric monopoles which we derive. In particular we discuss the relationship of these results to the analysis of Wilkinson and Goldhaber.

Thus §§3–6 attempt to survey what is known about classical soliton (monopole-like) solutions to field theories in three space and one time dimensions. The last section deals with two remaining questions. Firstly, what other sort of soliton solu-

tions can be constructed in other dimensions of (flat) space–time? Secondly, and this is probably the most important question posed by consideration of this subject, what is the quantum theory of monopoles? Little can be said despite an extensive literature, but there exist some speculations based on the lesson of the Sine-Gordon theory described in §3. These at least serve to show how interesting the answer might be.

## 2. The Dirac monopole

### *2.1. The duality of electricity and magnetism*

The equations governing the electromagnetic field, Maxwell's equations:

$$\nabla.\boldsymbol{E}=\rho \qquad \nabla\wedge\boldsymbol{B}-\dot{\boldsymbol{E}}=\boldsymbol{j} \tag{2.1}$$

$$\nabla.\boldsymbol{B}=0 \qquad \nabla\wedge\boldsymbol{E}+\dot{\boldsymbol{B}}=0 \tag{2.2}$$

may be written in the compact relativistic notation:

$$\partial_\nu F^{\mu\nu}=-j^\mu \tag{2.3}$$

$$\partial_\nu {}^*F^{\mu\nu}=0 \tag{2.4}$$

where $F^{\mu\nu}$ is the electromagnetic field tensor:

$$(F^{\mu\nu})=\begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B^3 & B^2 \\ E^2 & B^3 & 0 & -B^1 \\ E^3 & -B^2 & B^1 & 0 \end{pmatrix} \tag{2.5}$$

i.e.

$$F^{0i}=-E^i \qquad F^{ij}=-\epsilon_{ijk}B^k \tag{2.6}$$

$(j^\mu)=(\rho,\boldsymbol{j})$ is the electric current four-vector and ${}^*F^{\mu\nu}$ is the dual tensor of $F^{\mu\nu}$:

$${}^*F^{\mu\nu}=\tfrac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma} \tag{2.7}$$

which may be obtained formally from $F^{\mu\nu}$ by replacing $\boldsymbol{E}$ by $\boldsymbol{B}$ and $\boldsymbol{B}$ by $-\boldsymbol{E}$. (We use the conventions that $\epsilon^{\lambda\mu\nu\rho}$ is totally antisymmetric with $\epsilon^{0123}=1$, and Greek indices take the values 0, 1, 2, 3, whilst Latin indices only take the values 1, 2, 3.)

In vacua, where $j^\mu$ vanishes, the Maxwell equations are symmetric under the 'duality' transformation:

$$F^{\mu\nu}\rightarrow {}^*F^{\mu\nu} \qquad {}^*F^{\mu\nu}\rightarrow -F^{\mu\nu} \tag{2.8}$$

or, equivalently, $\boldsymbol{E}\rightarrow\boldsymbol{B}$ and $\boldsymbol{B}\rightarrow-\boldsymbol{E}$, which, roughly speaking, interchanges electricity with magnetism. Could such a symmetry be valid even in the presence of matter? In such a theory we would have to introduce a magnetic current $(k^\mu)=(\sigma,\boldsymbol{k})$, on the right-hand side of equation (2.2) and (2.4), giving the new field equations:

$$\partial_\nu F^{\mu\nu}=-j^\mu \qquad \partial_\nu {}^*F^{\mu\nu}=-k^\mu. \tag{2.9}$$

Equations (2.9) are symmetric under the duality transformation of equations (2.8) augmented by:

$$j^\mu\rightarrow k^\mu \qquad k^\mu\rightarrow -j^\mu. \tag{2.10}$$

If the electric and magnetic currents result from point particles at space–time points $x_i$, as we shall suppose:

$$j^\mu(x) = \sum_i q_i \int \mathrm{d}x_i{}^\mu \delta_4(x - x_i) \tag{2.11(a)}$$

$$k^\mu(x) = \sum_i g_i \int \mathrm{d}x_i{}^\mu \delta_4(x - x_i) \tag{2.11(b)}$$

where the integral over $x_i$ is taken along the world line of the $i$th particle whose electric and magnetic charges are $q_i$ and $g_i$, respectively. In conventional electrodynamics the Lorentz force law for a particle of (electric) charge $q$ and rest mass $m$ leads to the equation of motion:

$$m \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = qF^{\mu\nu} \frac{\mathrm{d}x_\nu}{\mathrm{d}\tau}. \tag{2.12}$$

In a symmetric theory this equation would be generalised to:

$$m \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = (qF^{\mu\nu} + g{}^*F^{\mu\nu}) \frac{\mathrm{d}x_\nu}{\mathrm{d}\tau} \tag{2.13}$$

where $g$ is the particle's magnetic charge. Equations (2.9), (2.11) and (2.13) completely specify the dynamics of a classical (i.e. non-quantum-mechanical) system of electrically and magnetically charged particles interacting with the electromagnetic field in such a way that it possesses the dual symmetry of equations (2.8) and (2.10).

In discussing further whether nature might indeed possess such a duality it is natural to ask at this point whether it is consistent with quantum theory. Actually Dirac (1931) was led naturally to a theory possessing this symmetry by considering a quantum mechanics in which the wavefunction had a non-integrable (or path-dependent) phase factor. (This formalism has been exploited by other authors more recently; see, for example, Mandelstam (1962, 1968) and Christ (1975).) Dirac's work pointed out the profound theoretical consequences of the existence of magnetic monopoles at the quantum level. One can see immediately that quantisation may not be straightforward since this procedure usually exploits the canonical (Hamiltonian) formalism. Now the canonical variables for the electromagnetic field are not the components of $F^{\mu\nu}$ but rather the components of the four-vector potential $(A^\mu) = (\phi, \boldsymbol{A})$, whose defining property is:

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.14}$$

This equation itself implies the vanishing of $\partial_\nu {}^*F^{\mu\nu}$ and, consequently, the magnetic current, $k^\mu$, destroying the dual symmetry.

Dirac was able to circumvent this difficulty, showing that a dually symmetric electromagnetic theory could be quantised, provided that for any electric charge $q$ and magnetic charge $g$ in theory, the condition:

$$\frac{qg}{4\pi\hbar} = \tfrac{1}{2}n \qquad n \text{ an integer} \tag{2.15}$$

was satisfied. This is the celebrated *Dirac quantisation condition*. The occurrence of the modified Planck constant, $\hbar$, emphasises that, in Dirac's approach, it is quantum-mechanical in origin. Much of the rest of this review will be devoted to deriving and re-deriving this condition and its generalisations with progressive degrees of sophistication. Dirac's approach assumed that a particle had either electric or magnetic charge but not both; we will also assume this henceforth unless it is otherwise stated.

Finally in this section we comment on the units we have employed. They are the units conventionally used in modern quantum field theory with, in particular, the velocity of light $c=1$. In this we differ from Dirac (1931, 1948). To obtain his quantities we must multiply our field tensor by $\sqrt{4\pi}$ and divide our charges and currents by $\sqrt{4\pi}$. Thus his Maxwell equations involve a factor of $4\pi$ on the right-hand side but the Lorentz force law is unaltered. Further, in Dirac's convention the factor of $4\pi$ in condition (2.15) disappears. In recent literature on monopoles an intermediate convention has been adopted implicitly which differs from that given here only in that the unit of magnetic charge is greater by a factor of $4\pi$. This permits the retention of both the conventions of modern field theory and Dirac's original form of equation (2.15) but is unsatisfactory because of its asymmetry.

Previous reviews and surveys of literature on Dirac monopoles can be found in Goldhaber (1965), Zumino (1966), Amaldi and Cabibbo (1972) and Goldhaber and Smith (1975). Amaldi and Cabibbo (1972) give a survey of experimental evidence.

### 2.2. *The motion of an electrically charged particle in a radial magnetic field*

Our first derivation of Dirac's quantisation condition will be the most naive. As is usual in physics, we begin by examining the simplest possible solution, that of a particle of mass $m$ and electric charge $q$ moving in the field of a magnetic monopole of strength $g$ fixed at the origin:

$$\boldsymbol{B}=\frac{g}{4\pi r^3}\boldsymbol{r}. \tag{2.16}$$

The equation of motion of the particle is:

$$m\ddot{\boldsymbol{r}}=q\dot{\boldsymbol{r}}\wedge\boldsymbol{B}. \tag{2.17}$$

The magnetic field of equation (2.16) is spherically symmetric and one therefore expects something like the conservation of angular momentum. However, it is not quite the orbital momentum that is conserved because equation (2.17) is not a central force (i.e. not directed towards the origin). In fact, the rate of change of orbital angular momentum:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}\wedge m\dot{\boldsymbol{r}})&=\boldsymbol{r}\wedge m\ddot{\boldsymbol{r}}\\&=\frac{qg}{4\pi r^3}\boldsymbol{r}\wedge(\dot{\boldsymbol{r}}\wedge\boldsymbol{r})\\&=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{qg}{4\pi}\hat{\boldsymbol{r}}\right)\end{aligned}$$

where $\hat{\boldsymbol{r}}=\boldsymbol{r}/r$. These results, first due to Poincaré (1896), suggest that we should define the *total* angular momentum to be:

$$\boldsymbol{J}=\boldsymbol{r}\wedge m\dot{\boldsymbol{r}}-\frac{qg}{4\pi}\hat{\boldsymbol{r}} \tag{2.18}$$

and then it will be conserved. To give a physical interpretation to the second term in this equation we must consider the only other possible source of angular momentum, the electromagnetic field. Classically the angular momentum of the electromagnetic

field is obtained by integrating the moment of the Poynting vector, $\boldsymbol{E}\wedge\boldsymbol{B}$, over all space:

$$\boldsymbol{J}_{\rm em}=\int \mathrm{d}^3x\boldsymbol{x}\wedge(\boldsymbol{E}\wedge\boldsymbol{B}).$$

Here $\boldsymbol{B}$ is the radial field given by equation (2.16) and $\boldsymbol{E}$ is the field due to the electric pole $q$ at $\boldsymbol{r}$. Thus:

$$\begin{aligned} J_{\rm em}{}^i &= \int \mathrm{d}^3x E^j(\delta_{ij}-\hat{x}^i\hat{x}^j)\frac{g}{4\pi x} \\ &= \int \mathrm{d}^3x E^j \frac{\partial}{\partial x^j}\left(\frac{g\hat{x}^i}{4\pi}\right) \\ &= -\int \mathrm{d}^3x \nabla.\boldsymbol{E}\frac{g}{4\pi}\hat{x}^i \end{aligned}$$

giving

$$J_{\rm em}=-\frac{qg}{4\pi}\hat{\boldsymbol{r}}$$

since $\nabla.\boldsymbol{E}=q\delta(\boldsymbol{x}-\boldsymbol{r})$. Thus the total angular momentum which is conserved is indeed the sum of the orbital angular momentum of the particle and the angular momentum of the electromagnetic field.

From equation (2.18) we see that the radial component of $\boldsymbol{J}$ is:

$$\hat{\boldsymbol{r}}.\boldsymbol{J}=-\frac{qg}{4\pi}. \tag{2.19}$$

Since $\boldsymbol{J}$ is a constant of the motion this means that the particle moves on a cone with semi-vertical angle $\cos^{-1}(qg/4\pi J)$ and axis $-\boldsymbol{J}$ with its apex at the monopole. The charges $q$ and $g$ behave rather as if repelled by one another.

So far our discussion has been in the context of classical mechanics (and, indeed, non-relativistic electromagnetic theory in that we have neglected radiation effects). In quantum mechanics we would expect the components of $\boldsymbol{J}$ to satisfy the usual angular momentum commutation relations and hence to have eigenvalues which are integral multiples of $\frac{1}{2}\hbar$. Since one might suppose that the orbital part of $\boldsymbol{J}$ has eigenvalues which are integral multiples of $\hbar$ we are left with (Saha 1936, 1949; see also Zumino 1966):

$$\frac{qg}{4\pi}=\tfrac{1}{2}n\hbar \qquad n \text{ an integer}$$

the Dirac quantisation condition, equation (2.15). This argument is plausible but too vague to be convincing as it stands. For example, since no fermions are involved it might seem more reasonable to have the components of $\boldsymbol{J}$ quantised in integral, rather than half-integral, multiples of $\hbar$. Before giving a rigorous derivation of equation (2.15) we will discuss the reasons why it is so significant.

### *2.3. The fascination of the Dirac quantisation condition*

Consider a world in which particles may carry either electric or magnetic charge but not both, the possible values of the electric and magnetic charges being $q_i$ and $g_i$, respectively. In such a situation the form the Dirac quantisation condition takes is:

$$\frac{q_ig_j}{4\pi}=\tfrac{1}{2}n_{ij}\hbar \qquad n_{ij} \text{ an integer} \tag{2.20}$$

and we shall see that this appears to be a necessary and sufficient condition for the existence of a quantum theory with these charges.

Taking *any* fixed magnetic charge $g_j$, *all* electric charges $q_i$ must be integral multiples of $2\pi\hbar/g_j$ and we may construct the highest common factor $n_{0j}$ of the integers $n_{ij}$. All electric charges are multiples of $q_0 = n_{0j} 2\pi\hbar/g_j$:

$$q_i = n_i q_0. \tag{2.21}$$

But further (by Euclid's algorithm in number theory) $n_{0j}$ must be a linear combination of the $n_{ij}$ with integer coefficients. Consequently $q_0$ is a linear combination of the $q_i$ with integer coefficients (which may not all be positive). Since combining states yields a state with the sum of the charges of the states combined, and charge conjugation yields a state with the opposite charge, $q_0$ must be the charge of a state which can be physically realised. Similar considerations lead to the conclusion that the magnetic charges, $g_j$, are all integral multiples of a physically realisable unit of magnetic charge, $g_0$. Clearly $q_0$ and $g_0$ are uniquely determined up to sign and must satisfy the Dirac quantisation condition themselves:

$$\frac{q_0 g_0}{4\pi} = \tfrac{1}{2} n_0 \hbar \qquad n_0 \text{ an integer.} \tag{2.22}$$

This conclusion of Dirac (1931), that the mere existence of an isolated magnetic charge implies the quantisation of electric charge, is very powerful and striking. The reason for its impact is that no other reason for the quantisation of electric charge had been discerned and, experimentally, it seems to be true in nature with $q_0$ being (plus or minus) the charge on the electron, usually denoted by $-e$. More recently the possibility of fractionally charged quarks has slightly altered the situation to the extent of possibly revising the basic unit, but we shall return to this point later.

Two electric charges of strength $q_0$, a distance $r$ apart, repel each other with a force of strength $q_0^2/4\pi r^2$. There would be a similar force between two magnetic charges, and if their strength were $g_0$, the magnetic force would be, as a fraction of the force for similarly situated electric charges:

$$\frac{g_0^2}{q_0^2} = \frac{n_0^2}{4}\left(\frac{q_0^2}{4\pi\hbar}\right)^{-2}. \tag{2.23}$$

If $q_0$ is the charge on the electron $q_0^2/4\pi\hbar$ is the fine-structure constant, approximately 1/137, giving a value for the ratio of forces of $5 \times 10^3 n_0^2$, which is large even if $n_0$ is one. So, although there is a theoretical symmetry between electricity and magnetism once magnetic monopoles are introduced, using experimental knowledge of the value of the fine-structure constant we see that the magnetic force is much stronger and there is a consequent practical asymmetry. As Dirac (1931) pointed out this should mean that magnetic monopoles are much more difficult to pair-produce than electrically charged particles and also much heavier. (In §4.6 we shall find a precise estimate for the mass in the context of a particular theoretical model.) Dirac had hoped that his theoretical considerations would lead to an understanding of the value of the fine-structure constant rather than an understanding of the possible relation between electric and magnetic forces given that value. He saw these considerations, leading to the quantisation of electric charge through the existence of magnetic monopoles, as being in many ways comparable to his arguments leading to the relativistic wave equation from which the existence of positrons was deduced.

A very peculiar situation results if it is possible to make a system, composed of an

electric and a magnetic pole, for which $qg/4\pi\hbar$ is half an odd integer. Indeed at the level of argument used in the last subsection we might be tempted to discard this possibility but our more rigorous arguments given in §2.5 do admit it. In such a situation there is a half-odd-integral angular momentum contribution from the electromagnetic field and the composite system has angular momentum which is not an integral multiple of $\hbar$. Thus, accepting the connection between spin and statistics, the composite system is a fermion and it has been constructed out of bosons. This has been a long standing paradox. Recently Goldhaber (1976) has given a subtle explanation of why this is possible (see also Jackiw and Rebbi 1976, Hasenfratz and 't Hooft 1976).

Schwinger (1966a, b, 1968, 1969) has presented arguments that the integer $n$ in the Dirac condition (2.15) should be an even integer or even a multiple of four, but these arguments are not generally accepted and he has since withdrawn his stand on the latter stronger restriction (Schwinger 1976).

### *2.4. The canonical formalism and quantisation*

Having appreciated the significance of the Dirac quantisation condition we proceed towards a rigorous derivation of it. To this end we discuss the quantisation of the motion of a particle in a given electromagnetic field. The conventional way to do this is to set up the Hamiltonian formalism of mechanics and replace the Poisson brackets by commutators. The non-relativistic equations of motion for a particle of mass $m$, electric charge $q$ (and no magnetic charge) can be derived from the Lagrangian:

$$L=\tfrac{1}{2}m\dot{\boldsymbol{r}}^2+q\dot{\boldsymbol{r}}.\boldsymbol{A}-q\phi \tag{2.24}$$

where $(A^\mu)=(\phi, \boldsymbol{A})$ is the electromagnetic four-potential, introduced in equation (2.14), describing the given electromagnetic field. This Lagrangian leads to the non-relativistic limit of the Lorentz force law of equation (2.12). The canonical momentum $\partial L/\partial\dot{\boldsymbol{r}}$ is:

$$\boldsymbol{p}=m\dot{\boldsymbol{r}}+q\boldsymbol{A} \tag{2.25}$$

and the Hamiltonian

$$H=\boldsymbol{p}.\dot{\boldsymbol{r}}-L=\frac{1}{2m}(\boldsymbol{p}-q\boldsymbol{A})^2+q\phi \tag{2.26}$$

which is indeed the energy, the sum of the kinetic and the potential energy. It may be obtained from the Hamiltonian for a free particle by the substitution $p^\mu\rightarrow p^\mu-qA^\mu$ in four-vector notation.

Notice that this formalism depends heavily on the existence of the electromagnetic potential $A^\mu$ and we have already remarked in §2.1 that this assumption is false in the presence of magnetic charge. For the moment we ignore this difficulty.

The Poisson bracket of two dynamical variables $\alpha$ and $\beta$ is defined by:

$$\{\alpha, \beta\}=\sum_{i=1}^{3}\left(\frac{\partial\alpha}{\partial r^i}\frac{\partial\beta}{\partial p^i}-\frac{\partial\alpha}{\partial p^i}\frac{\partial\beta}{\partial r^i}\right)$$

so that

$$\{r^i, r^j\}=\{p^i, p^j\}=0 \qquad \{r^i, p^j\}=\delta_{ij}. \tag{2.27}$$

All the quantities occurring in the field equations (2.3) and (2.4) and the equations

of motion (2.12) are *gauge-invariant*, i.e. they are unchanged if we perform the following *gauge transformation* on the vector potential:

$$A_\mu \to A_\mu' = A_\mu + \partial_\mu \chi \tag{2.28}$$

where $\chi$ is an arbitrary (suitably smooth single-valued) function of position and time. But, unlike $\boldsymbol{r}$ and $F^{\mu\nu}$, $p^\mu$ is not gauge-invariant because $A^\mu$ itself enters into equations (2.25) and (2.26). For this reason it is more convenient and aesthetic to recast equations (2.27) in terms of $m\dot{\boldsymbol{r}}$ rather than $\boldsymbol{p}$:

$$\begin{aligned} \{r^i, r^j\} &= 0 \\ \{r^i, m\dot{r}^j\} &= \delta_{ij} \\ \{m\dot{r}^i, m\dot{r}^j\} &= -q(\partial^i A^j - \partial^j A^i) = q\epsilon_{ijk}B^k. \end{aligned} \tag{2.29}$$

From these Poisson brackets we may compute those of $\boldsymbol{r}$ and $m\dot{\boldsymbol{r}} = \boldsymbol{p} - q\boldsymbol{A}$ with the orbital angular momentum $\boldsymbol{L} = \boldsymbol{r} \wedge m\dot{\boldsymbol{r}}$:

$$\begin{aligned} \{L^i, r^j\} &= \epsilon_{ijk} r^k \\ \{L^i, m\dot{r}^j\} &= \epsilon_{ijk} m\dot{r}^k + q(\delta_{ij}\boldsymbol{r}.\boldsymbol{B} - B^i r^j). \end{aligned} \tag{2.30}$$

Thus $m\dot{\boldsymbol{r}}$ does not transform as a vector with respect to the orbital angular momentum. Now suppose that $\boldsymbol{B}$ has the specific radial form of equation (2.16). For such a $\boldsymbol{B}$ we cannot construct a single-valued vector potential everywhere but we may inside some suitable region not containing the origin. The last term in equation (2.30) reads:

$$\frac{qg}{4\pi r}(\delta_{ij} - \hat{r}^i \hat{r}^j) = \left\{ \frac{qg}{4\pi} \hat{r}^i, m\dot{r}^j \right\}.$$

Hence, introducing the total angular momentum $\boldsymbol{J} = \boldsymbol{L} - gq\hat{\boldsymbol{r}}/4\pi$, we obtain:

$$\begin{aligned} \{J^i, m\dot{r}^j\} &= \epsilon_{ijk} m\dot{r}^k \\ \{J^i, r^j\} &= \epsilon_{ijk} r^k \\ \{J^i, J^j\} &= \epsilon_{ijk} J^k. \end{aligned} \tag{2.31}$$

In particular as for this $\boldsymbol{B}$, $H = \frac{1}{2}m\dot{\boldsymbol{r}}^2$, we see that the Poisson brackets of $\boldsymbol{J}$ with the Hamiltonian vanish and thus we have rederived its conservation, first obtained in §2.2. The Poisson brackets of the components of $\boldsymbol{J}$ yield the familiar angular momentum algebra.

The canonical procedure of quantising the motion of the charged particle (treating the electromagnetic field classically) is to replace Poisson brackets by commutators:

$$\{\alpha, \beta\} \to \frac{1}{i\hbar}[\alpha, \beta]. \tag{2.32}$$

A representation of the canonical commutation relations resulting from equations (2.27) is provided by:

$$\boldsymbol{p} = -i\hbar\boldsymbol{\nabla}$$

so that

$$m\dot{r}^i = i\hbar\, \mathscr{D}^i = i\hbar\, \partial^i - qA^i \tag{2.33}$$

where

$$\mathscr{D}^\mu = \partial^\mu + ieA^\mu$$

is called the covariant derivative and we have introduced $e = q/\hbar$. (In our units with

$c=1$, $q$ and $g$ have the units of the square root of action, i.e. (mass × length)$^{1/2}$, so that $e$ has the units of inverse charge.) Consistently with equation (2.29):

$$[\mathscr{D}^\mu,\ \mathscr{D}^\nu]=\mathrm{i}eF^{\mu\nu}. \tag{2.34}$$

This equation and its generalisations are of fundamental importance in further developments.

The Schrödinger equation for the wavefunction of the charged particle is:

$$-\frac{\hbar^2}{2m}\,\mathscr{D}^2\psi+q\phi\psi=\mathrm{i}\hbar\,\frac{\partial\psi}{\partial t} \tag{2.35}$$

and consequently not gauge-invariant. However, we may show the equivalence of Schrödinger equations for different gauges and make equation (2.35) gauge-*covariant* if we specify that under the gauge transformation of equation (2.28):

$$\psi\rightarrow\psi'=\exp(-\mathrm{i}e\chi)\,\psi. \tag{2.36}$$

Thus in quantum mechanics it is not merely the electromagnetic potential that changes in a gauge transformation but also the wavefunction of a charged particle and the way it transforms depends on the electric charge $q=\hbar e$ of the particle it describes.

### *2.5. The vector potential for a monopole: the Dirac string*

In the last two subsections we have seen the crucial role played by the vector potential in the Hamiltonian mechanics and canonical quantisation of the motion of an electrically charged particle in a magnetic field. But as we remarked in §2.1 such a potential cannot exist everywhere if there are isolated magnetic charges. In particular if we again consider the radial magnetic field of equation (2.16), for any closed surface, S, containing the origin:

$$g=\int_S \boldsymbol{B}.\mathrm{d}\boldsymbol{S} \tag{2.37}$$

but if $\boldsymbol{B}=\nabla\wedge\boldsymbol{A}$ this integral would have to vanish. Thus $\boldsymbol{A}$ cannot exist everywhere on S, even though $\nabla.\boldsymbol{B}$ is only non-zero at the origin, and the best we can do is to find an $\boldsymbol{A}$ defined everywhere except on a line joining the origin to infinity, such that $\boldsymbol{B}=\nabla\wedge\boldsymbol{A}$. To see that this is possible consider the field due to an infinitely long and thin solenoid placed along the negative $z$ axis with its positive pole which has strength $g$ at the origin. Its magnetic field would be:

$$\boldsymbol{B}_{\mathrm{sol}}=\frac{g}{4\pi r^2}\,\hat{\boldsymbol{r}}+g\theta(-z)\,\delta(x)\,\delta(y)\,\hat{\boldsymbol{z}}$$

where $\hat{\boldsymbol{z}}$ is a unit vector in the $z$ direction and $\theta(\xi)=0$ if $\xi<0$, $\theta(\xi)=1$ if $\xi>0$. This magnetic field differs from $\boldsymbol{B}$ only by the singular magnetic flux along the solenoid but it is clearly source-free; $\nabla.\boldsymbol{B}_{\mathrm{sol}}$ vanishes, even at the origin. Thus it may be represented by a vector potential, $A$ say, everywhere and we may write:

$$\frac{g}{4\pi r^2}\,\hat{\boldsymbol{r}}=\nabla\wedge\boldsymbol{A}-g\theta(-z)\,\delta(x)\,\delta(y)\,\hat{\boldsymbol{z}}. \tag{2.38}$$

The line occupied by the solenoid is called the 'Dirac string'. The effect of equation (2.38) is graphically represented in figure 1. We should think of the field $\boldsymbol{B}$ as being represented not just by $\boldsymbol{A}$ but by $\boldsymbol{A}$ together with a string $\mathscr{S}$ on which it is singular.

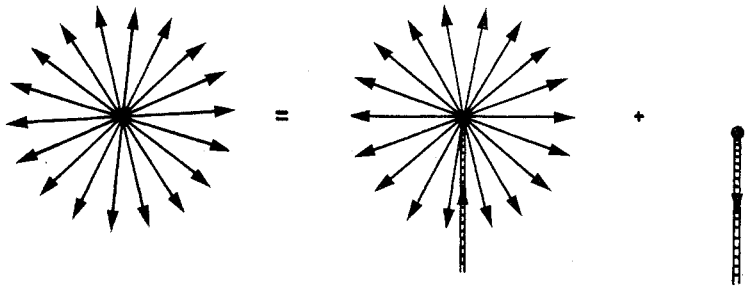

**Figure 1.**

Given our choice of the position of the string as the negative $z$ axis we can easily calculate an explicit form for $\boldsymbol{A}$ by exploiting the axial symmetry. Using spherical polar coordinates $(r, \theta, \phi)$ we expect by symmetry to be able to find a vector potential $\boldsymbol{A}(\boldsymbol{r})=A(r, \theta)\,\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\phi}}$ being a unit vector in the $\phi$ direction. The magnetic flux through a circle, $C$, corresponding to fixed values of $r$ and $\theta$, and $\phi$ ranging over the values 0 to $2\pi$, is given by the solid angle subtended by $C$ at the origin multiplied by $g/4\pi$, namely $\frac{1}{2}g(1-\cos\theta)$. Consequently:

$$\tfrac{1}{2}g(1-\cos\theta)=\int \boldsymbol{B}.\mathrm{d}\boldsymbol{S}=2\pi A(r, \theta)\, r\sin\theta$$

and

$$\boldsymbol{A}(\boldsymbol{r})=\frac{g}{4\pi r}\frac{(1-\cos\theta)}{\sin\theta}\hat{\boldsymbol{\phi}} \tag{2.39}$$

showing the anticipated singularity on the negative $z$ axis.

Dirac (1931) suggested that the vector potential obtained in this way be used to set up the Schrödinger equation for the motion of a charged particle in the field $\boldsymbol{B}$. To show consistency it is necessary to show that the resulting equations are equivalent for different choices of the position of the Dirac string, $\mathscr{S}$. In the next subsection we will show that this is consistent if the Dirac quantisation condition is satisfied. For the moment we will remark that using equation (2.39) in the canonical quantisation procedure we find, from equation (2.18):

$$\boldsymbol{J}=\boldsymbol{r}\wedge\boldsymbol{p}-\frac{qg}{4\pi}(\hat{\boldsymbol{r}}+\hat{\boldsymbol{z}})(1+\cos\theta)^{-1} \tag{2.40}$$

and thus:

$$J_z=-\mathrm{i}\hbar\frac{\partial}{\partial\phi}-\frac{qg}{4\pi}. \tag{2.41}$$

This is not itself singular on the $z$ axis but for the wavefunctions to be single-valued the eigenvalues of $J_z$ must be $n\hbar-qg/4\pi$ for integral $n$. But the angular momentum commutation relations satisfied by $\boldsymbol{J}$ ensure that these must be of the form $\frac{1}{2}N\hbar$ where $N$ is integral. This strengthens the argument following equation (2.19) leading to the Dirac quantisation condition.

### *2.6. Generalised gauge transformations and a rigorous derivation of the Dirac condition*

The vector potential of equation (2.39) is not unique, of course. Firstly, we make a non-singular gauge transformation, as in equation (2.28), $\boldsymbol{A}\rightarrow\boldsymbol{A}+\nabla\chi$, where $\chi$ is a non-singular, single-valued function of position. The term $\nabla\wedge\boldsymbol{A}$ in equation (2.38)

will remain unchanged and so, therefore, must the Dirac string term. But the position of the string is arbitrary and so we must find the relationship between the singular potentials corresponding to different positions of the string, which need not even be straight. We will need to extend the concept of gauge transformation in order to be able to move the string. We rewrite equation (2.38) as:

$$\boldsymbol{B}(\boldsymbol{r})=\nabla\wedge\boldsymbol{A}+\boldsymbol{h}(\mathscr{S},\boldsymbol{r}) \tag{2.42}$$

where $\boldsymbol{h}(\mathscr{S},\boldsymbol{r})$ represents the contribution of the Dirac string $\mathscr{S}$, a flux of strength $g$ passing out along the string $\mathscr{S}$ from the origin to infinity:

$$\boldsymbol{h}(\mathscr{S},\boldsymbol{r})=g\int_{\mathscr{S}}\mathrm{d}\boldsymbol{x}\,\delta_3(\boldsymbol{r}-\boldsymbol{x}). \tag{2.43}$$

As defined in §2.5 the string $\mathscr{S}$ runs along the negative $\mathscr{S}$ axis. Consider another string $\mathscr{S}'$ running from the origin to infinity. Let $\Gamma$ denote the curve $-\mathscr{S}'$ ($\mathscr{S}'$ taken in the reverse direction) followed by $\mathscr{S}$. We may treat this as a closed curve, either by making suitable assumptions about what happens at infinity or by assuming that $\mathscr{S}'$ differs from $\mathscr{S}$ only over a finite range. Let $\Omega(\boldsymbol{r})$ denote the solid angle subtended at $\boldsymbol{r}$ by some particular surface spanning $\Gamma$. Various choices of spanning surface will lead to values of $\Omega$ differing by multiples of $4\pi$ but will yield the same value for $\nabla\Omega$, except on $\Gamma$ itself where $\Omega$ and $\nabla\Omega$ are always ill-defined. So consider the extended gauge transformation defined by

$$\boldsymbol{A}\rightarrow\boldsymbol{A}'=\boldsymbol{A}-\frac{g}{4\pi}\nabla\Omega \tag{2.44}$$

when $\boldsymbol{r}$ is not on $\Gamma$. Then $\nabla\wedge\boldsymbol{A}'=\nabla\wedge\boldsymbol{A}=\boldsymbol{B}$ except on the two strings. Applying Stokes' theorem to a small loop encircling $\Gamma$ we see that the flux of $\nabla\wedge(\boldsymbol{A}'-\boldsymbol{A})$ along $\Gamma$ is $g$, and so:

$$\nabla\wedge(\boldsymbol{A}'-\boldsymbol{A})=\{\boldsymbol{h}(\mathscr{S},\boldsymbol{r})-\boldsymbol{h}(\mathscr{S}',\boldsymbol{r})\}$$

$$\boldsymbol{B}=\nabla\wedge\boldsymbol{A}+\boldsymbol{h}(\mathscr{S},\boldsymbol{r})=\nabla\wedge\boldsymbol{A}'+\boldsymbol{h}(\mathscr{S}',\boldsymbol{r}). \tag{2.45}$$

Thus a gauge transformation of the form of equation (2.44) shifts the Dirac string, and using such multivalued gauge transformations we may relate any pair of Dirac potentials for a monopole.

If we have a magnetic field with a number of magnetic monopoles with charges $g_i$ we will need a string for each. The general gauge transformation will then take the form:

$$\chi(\boldsymbol{r})=\chi_0(\boldsymbol{r})+\sum_i\frac{g_i}{4\pi}\,\Omega_i(\boldsymbol{r}) \tag{2.46}$$

where $\chi_0$ is single-valued and $\Omega_i(\boldsymbol{r})$ is the angle subtended at $\boldsymbol{r}$ by the reversed final string followed by the initial string attached to the $i$th monopole.

The crucial consistency condition is that the general gauge transformation should yield an equivalent quantum mechanics. This will be so if the effect of the gauge transformation on the wavefunction:

$$\psi\rightarrow\psi'=\exp(-ie\chi)\,\psi$$

is not to produce a multivalued result. Since there is an ambiguity of $4\pi$ in the $\Omega_i$ in

equation (2.46) we need:

$$\frac{eg_i}{4\pi}=\tfrac{1}{2}n_i \qquad n_i \text{ an integer}$$

which, since $q=\hbar e$, is Dirac's quantisation condition.

Strictly speaking, to show the existence of the quantum mechanics it is necessary to show that the singular Hamiltonian is self-adjoint. The sufficiency of the Dirac condition for this purpose was demonstrated by Hurst (1968).

### *2.7. Further comments on the Dirac monopole*

*2.7.1. The Aharonov–Bohm effect.* The increased significance of the vector potential at the quantum level was greatly emphasised by the work of Aharonov and Bohm (1959, 1961). They showed that there would be quantum interference effects between two parts of a beam of charged particles, charge $q$ say, which had passed either side of a region through which passes a magnetic field confined to a narrow tube in a direction transverse to the beam. This effect was subsequently experimentally verified (Mollenstedt and Bayh 1962). The magnetic field produces an interference effect even though the beam only passes through regions of zero field. The condition for the absence of the Aharonov–Bohm effect is that $q\Phi/2\pi\hbar$ be an integer where $\Phi$ is the total magnetic flux. The Dirac quantisation condition is thus precisely the condition that the flux along the Dirac string should give rise to no Aharonov–Bohm effect. The magnetic field of the monopole then differs from that of the infinitely long solenoid used to define the vector potential by an unobservable tube of flux and we may regard replacing $\boldsymbol{B}$ by $\boldsymbol{\nabla}\wedge\boldsymbol{A}$ as changing the field in an unobservable way.

*2.7.2. Dyons.* So far we have considered particles with either electric or magnetic charge but not both. Dirac (1948) was uncertain whether they could exist. Schwinger (1969) has called them dyons. Consider a dyon with charges $(q_1, g_1)$ fixed at the origin with another with charges $(q_2, g_2)$ orbiting about it. The naive angular momentum analysis of §2.2 can be repeated. The contribution of the electromagnetic field to the angular momentum is now:

$$(q_1g_2-q_2g_1)\,\hat{\boldsymbol{r}}/4\pi$$

and the quantisation condition becomes:

$$\frac{q_1g_2-q_2g_1}{4\pi}=\tfrac{1}{2}n_{12}\hbar \qquad n_{12} \text{ an integer.}$$

This condition is invariant under a rotation in the $(q, g)$-plane generalising the symmetry under the duality transformation of equation (2.10) (Schrödinger 1935, Schwinger 1968), which is just a rotation through $\frac{1}{2}\pi$.

It has been suggested that quarks are dyons. For example, in the magnetic string model of hadrons, the quarks lie at the end points of the string and must have magnetic charges which are the sources of the flux along the string, as well as the ordinary electric charges (Schwinger 1968, 1969, Artru 1977).

### *2.8. The moral of the monopole*

The discussion of this section can be regarded as a progressive development and

elevation of the concept of gauge invariance. An important consequence of the quantisation of electric charge is that the group of possible gauge transformations, at a given point, can be thought of as a compact group, U(1), the group of complex numbers of unit modulus or, equivalently, of displacements round a circle, rather than the non-compact group, $\mathbb{R}$, of real numbers under addition. The action of a gauge transformation on a wavefunction is to multiply it by a phase factor:

$$\exp(-iq\chi/\hbar) = [\exp(-iq_0\chi/\hbar)]^n$$

where $q = nq_0$, $n$ an integer and $q_0$ is the basic unit of electric charge. We can think of $\exp(-iq_0\chi/\hbar)$ as the basic gauge transformation and all other gauge transformations on wavefunctions as representations of it; if $\exp(-iq_0\chi/\hbar) = 1$, the gauge transformation has no effect on any wavefunction. Thus we see that the compactness of the gauge group, being a consequence of charge quantisation, follows from the existence of a magnetic monopole (Yang 1970). In the later sections we shall see how the global topology of a gauge group is intimately connected with the existence of monopoles.

In his original paper Dirac (1931) only treated the case, discussed here, of an electrically charged particle moving in a fixed magnetic monopole field. Dirac (1948) proceeded further and developed the relativistic classical and quantum dynamics of a system of moving magnetic monopoles and electric charges in interaction, based on an action principle. This is an impressive achievement because of the difficulties presented by the strings. Some problems remained. There were the familiar ones which occur when there are point particles with divergent self-energy. In order to derive the equations of motion from an action principle it was necessary to postulate that a charged particle must never pass through a string, the 'Dirac veto'. Recently, Brandt and Primack (1977a, b) have discussed how the Dirac veto may be avoided. Despite the difficulties, a quantum field theory of electric and magnetic charges has been developed (Weinberg 1965, Zwanziger 1965, Schwinger 1966a, b, c).

We shall not follow these developments here because there is another, more recent, line of approach which it is the central purpose of this review to expound. In our discussion the concept of gauge invariance has played a fundamental role and it is natural to enquire about its possible generalisations from U(1) to other compact gauge groups such as SU(2). The crucial step in this direction was taken by Yang and Mills (1954) and Shaw (1955). If the electric charge operator is one of the generators of a compact gauge group such as SU(2), its eigenvalues, which are proportional to the possible electric charges of the particles in theory, would be quantised by the familiar arguments used in angular momentum theory. At first sight, nature does not seem to possess such a wide gauge invariance; for each generator of the gauge group we would expect a massless vector meson, but the only spin-one particle with zero or near zero mass is the photon. For this reason gauge groups larger than U(1) seemed to lack physical interest. The situation changed with the work of Higgs (1964a, b, 1966) (see also Englert and Brout 1964, Guralnik *et al* 1964, Kibble 1967); if the vacuum state corresponds to a non-zero value of some scalar (Higgs) field, which is an SU(2) vector, it can break the SU(2) symmetry down to U(1). The electric charge operator would be the generator of this compact subgroup. Obtaining the electromagnetic gauge group as a necessarily compact subgroup of a compact gauge group seemed to provide an alternative explanation of charge quantisation which did not rely on the existence of magnetic monopoles. But this appearance is superficial for it is in exactly these Higgs models that 't Hooft (1974) and Polyakov (1974) found solutions corresponding to magnetic monopoles. However, these models do have an

advantage over the original theory of Dirac: the solutions no longer require point sources to be put in by hand and the magnetic charge is smoothed out, with a finite mass due to self-energy. Further the same Lagrangian describes any number of interacting monopoles.

Before studying these developments we shall follow in the next section an apparently different line of thought which again leads to the same conclusion: the interest of a non-Abelian gauge theory in which the symmetry is spontaneously broken by the vacuum.

## 3. An unusual relationship: the Sine-Gordon and Thirring models

### *3.1. The Sine-Gordon model*

In this section we shall review some interesting and intriguing properties of the Sine-Gordon equation in one space and one time dimension. This section is less detailed than the others since this is a subject in itself dealt with in other reviews (Scott *et al* 1973, Coleman 1975b). It is when we try to generalise the results described here to a physical space–time of three space and one time dimensions that we shall be led again to a non-Abelian gauge theory with the vacuum spontaneously breaking the symmetry.

The Sine-Gordon equation:

$$\frac{\partial^2\phi}{\partial t^2}-\frac{\partial^2\phi}{\partial x^2}=-\alpha\sin\beta\phi \qquad [\simeq -\alpha\beta\phi+\mathrm{O}(\phi^2)] \tag{3.1}$$

seems to describe a scalar field with 'mass' $m=\sqrt{\alpha\beta}$ and a non-polynomial self-interaction. The equation may be derived from the Lagrangian density:

$$\mathscr{L}=\frac{1}{2}\left\{\left(\frac{\partial\phi}{\partial t}\right)^2-\left(\frac{\partial\phi}{\partial x}\right)^2\right\}-V(\phi) \tag{3.2}$$

where

$$V(\phi)=\frac{\alpha}{\beta}(1-\cos\beta\phi). \tag{3.3}$$

It is very easy to visualise a physical system which would be described by such an equation. Consider a long straight, horizontal 'clothesline' with identical pegs attached at equal distances along its length. Adjacent clothes pegs are connected by equal springs and each peg is acted on by gravity. If $\beta\phi(x, t)$ is the angle between the peg, at the point $x$, and the downward vertical at time $t$ then $\mathscr{L}$ is indeed the appropriate Lagrangian density in the continuum limit (in which the distance between the pegs tends to zero). The terms in $\mathscr{L}$ represent the kinetic energy, the potential energy stored in the springs and the gravitational potential energy, respectively, provided units are suitably chosen.

Obviously the 'ground state' of the system is when all pegs hang down motionless. The constant in equation (3.3) has been chosen so that the Hamiltonian constructed from equation (3.2) has zero as its lower bound, which is attained in this ground state. Consider a situation in which the pegs hang down everywhere except in a finite central region in which $\beta\phi$ increases by $2\pi$, so that there is a twist ('kink' or 'soliton') where the pegs flip over. This configuration could never decay into the previously described ground state since it would require an infinite amount of energy to flip

over all the pegs to the left (or right) of the kink in a finite time. Therefore one would expect a stable motionless state corresponding to this kink to exist as a solution to equation (3.1) and this may be constructed by direct integration under the assumption of time independence.

This state resembles a particle, with structure, in several respects: its energy density is concentrated in a finite region; it can be made to move with any velocity less than unity (the velocity of light) since equation (3.1) is Lorentz-invariant (so that any solution will remain a solution after a Lorentz boost has been applied). Further, one can consider solutions in which several kinks move with different velocities, eventually colliding and scattering. Such solutions can be constructed explicitly, and using the mechanical model described above and other physical systems described by equation (3.1) they may be studied experimentally.

The mass of the kink is the energy of the time-independent solution in which $\beta\phi$ goes through $2\pi$ as we go along the line:

$$M=\int_{-\infty}^{\infty}\left[\frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2+V(\phi)\right]\mathrm{d}x.$$

For a time-independent solution the first integral of equation (3.1) is:

$$\frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2-V(\phi)=0$$

using the boundary condition that the pegs hang down at large distances. Hence the mass can be written:

$$\begin{aligned}M&=\textstyle\int_0^{2\pi/\beta}[V(\phi)]^{1/2}\,\mathrm{d}\phi\\&=8m/\beta^2.\end{aligned}\tag{3.4}$$

Thus there is an unexpected particle-like solution to the Sine-Gordon equation, with a mass which increases as the coupling strength $\beta$ decreases. It has little to do with the excitations of the $\phi$ field which correspond to particles of mass $m\hbar$ *after we have passed to quantum mechanics*. It is a classical object which can be given both a specific position and a specific momentum.

### *3.2. Topological versus Noether conservation laws*

With a view to generalising to three space and one time dimensions, we shall try to understand and isolate the peculiar features of the Sine-Gordon theory which are responsible for its interesting behaviour.

The potential $V(\phi)$ of equation (3.3) has the form illustrated in figure 2 and so has an infinite set of degenerate minima:

$$\mathcal{M}_0=\{2n\pi/\beta\colon n=0,\ \pm1,\ \pm2,\ldots\}.\tag{3.5}$$

Physically the various elements of this set, $\mathcal{M}_0$, of minima are equivalent. Since $\beta\phi$ is an angular variable they all correspond to hanging down. Mathematically the fact that $\beta\phi$ is an angular variable manifests itself in the fact that $\mathscr{L}$ is invariant with respect to the discrete transformations:

$$\phi\rightarrow\phi'=\phi+2n\pi/\beta\qquad n=0,\ \pm1,\ \pm2,\ldots.\tag{3.6}$$

These transformations form a discrete group, the group of integers under addition, which we will denote by $\mathbb{Z}$. Any two elements of $\mathcal{M}_0$ can be related by an element of

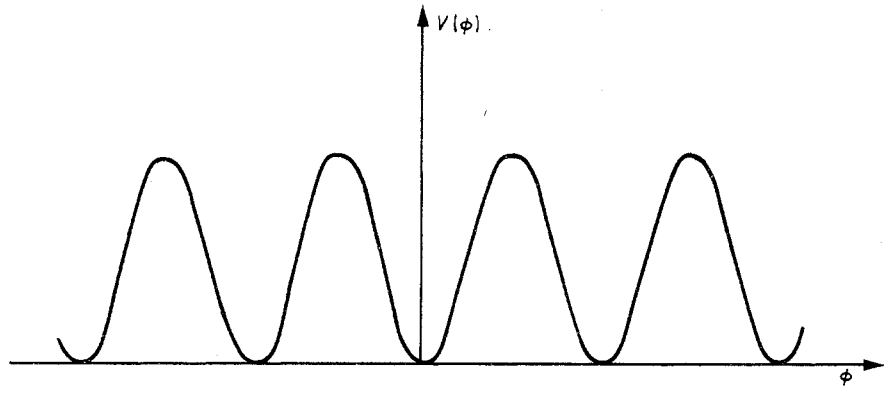


**Figure 2.**

this group; that is, any two degenerate ground states can be related by a symmetry of the theory.

In order for a solution to equation (3.1) to have finite energy $\phi$ must tend to limits in $\mathscr{M}_0$ as $x$ tends to $\pm\infty$, respectively; that is, the pegs must hang down at infinity. So:

$$\phi(\infty)-\phi(-\infty)=2\pi N/\beta \qquad \text{for some } N\in\mathbb{Z} \tag{3.7}$$

always. This integer can be regarded as a 'quantum number' characterising the state of the system. It takes the value 0 for the ground state, $+1$ for a kink, $-1$ for an antikink (in which $\beta\phi$ flips through $-2\pi$), and so on. It is conserved as the system evolves in time since the pegs would need an infinite amount of energy to vary their position at positive or negative infinity. Given this conserved quantity one may seek a corresponding conserved current. Consider

$$j^\mu=\frac{\beta}{2\pi}\,\epsilon^{\mu\nu}\,\partial_\nu\phi \tag{3.8}$$

where $\epsilon^{\mu\nu}$ is antisymmetric with $\epsilon^{01}=1$. Then $j^\mu$ is automatically conserved (i.e. $\partial_\mu j^\mu=0$) and the corresponding charge:

$$Q=\int_{-\infty}^{\infty} j^0\,\mathrm{d}x=\frac{\beta}{2\pi}\int_{-\infty}^{\infty}\frac{\partial\phi}{\partial x}\,\mathrm{d}x$$

$$=\frac{\beta}{2\pi}\,[\phi(\infty)-\phi(-\infty)]=N. \tag{3.9}$$

Thus we have constructed a conserved current corresponding to $N$. However, it differs from the usual sorts of currents derived from symmetries; we shall call it topological for reasons that will become clearer later. It has the following distinctive features.

(i) It is conserved independently of the equations of motion.

(ii) $j^0$ contains only canonical coordinates and no momenta. Therefore its Poisson brackets with coordinates vanish and it generates no symmetry.

(iii) $j^0$ is a spatial divergence. Its spatial integral is non-zero only if the field satisfies the boundary conditions, $\phi(\pm\infty)\in\mathscr{M}_0$, in a (topologically) non-trivial way.

These features distinguish a topological current from a Noether current associated

with a continuous symmetry of the Lagrangian (indeed equation (3.2) has none) which has the following properties.

(i) Its conservation follows from the equations of motion.

(ii) $j^0$ contains canonical momenta.

These comments suggest the conclusion that when there are degenerate vacua there can exist a new sort of conserved current, a topological current, transporting a conserved quantity which, unlike a Noether current or charge, is not associated with any manifest symmetry of the Lagrangian.

### *3.3. The Thirring model*

We shall now see that the conclusion of the last subsection is, in a sense, false, but in a very subtle and interesting way.

One strange aspect of the analysis of §3.2 is that we obtained a quantum number, i.e. a conserved quantity which takes only integer values, but in a classical rather than a quantum-mechanical system. Usually the existence of such a quantity is a quantum phenomenon, the discrete values crowding together to form a continuum in the classical limit. Consider, for example, the Thirring model (Thirring 1958) given by the Lagrangian density:

$$\mathscr{L} = \tfrac{1}{2}\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - \tfrac{1}{2}g\bar{\psi}\gamma^\mu\psi\bar{\psi}\gamma_\mu\psi. \tag{3.10}$$

(Like the Sine-Gordon model this would define a renormalisable quantum field theory in one space and one time dimension.) This theory has a continuous (global) U(1) symmetry:

$$\psi \rightarrow \psi' = \exp(\mathrm{i}\alpha)\,\psi \tag{3.11}$$

which begets the conserved Noether current:

$$J^\mu = \bar{\psi}\gamma^\mu\psi \tag{3.12}$$

and the associated charge:

$$Q = \int_{-\infty}^{\infty} J^0 \,\mathrm{d}x.$$

This charge has a spectrum which is quantised in integral multiples of $\hbar$, in the quantum field theory of the Thirring model (where $\psi$ is a Fermi field), because of the commutator:

$$[Q, \psi(x)] = -\hbar\psi(x). \tag{3.13}$$

Up to the factor of $\hbar$, this equation says that the field $\psi$ carries unit quantum number, and so $Q$ can be thought of as a sort of baryon number associated with the elementary field of the theory.

The two currents $J^\mu/\hbar = \bar{\psi}\gamma^\mu\psi/\hbar$ and $j^\mu = \beta\epsilon^{\mu\nu}\partial_\nu\phi/2\pi$ each have charges whose spectrum is just $\mathbb{Z}$, the integers, yet one is a Noether current whilst the other is topological. Surprisingly they are the same if the Sine-Gordon model is quantised,

$$\frac{\beta^2\hbar}{4\pi} = \frac{1}{1+g\hbar/\pi} \tag{3.14}$$

and other parameters suitably adjusted.

The equivalence of these currents and, more generally, of the Sine-Gordon and Thirring models, has a long history, originating in the pioneering work of Skyrme (1958, 1959, 1961a,b) and culminating in the work of Coleman (1975a) and Mandelstam (1975), which again brought it to the attention of particle physicists. It is a new

sort of equivalence theorem because quantum mechanics is an essential feature and because the relationship between $\phi$ and $\psi$ is non-local.

In the quantum Sine-Gordon model, $j^\mu = \beta \epsilon^{\mu\nu} \partial_\nu \phi / 2\pi$ is linear in the quantum fields and there is no ambiguity in the quantum operator or in calculating its equal-time commutators:

$$[j^0(x, t), j^0(y, t)] = [j^1(x, t), j^1(y, t)] = 0 \qquad (3.15(a))$$

$$[j^0(x, t), j^1(y, t)] = \mathrm{i}C \frac{\partial}{\partial x} \delta(x-y) \qquad (3.15(b))$$

with $C = \beta^2 \hbar / 4\pi^2$. The corresponding commutators for the Thirring model are more subtle to calculate but are known to be of the same form with $C = (\pi + g\hbar)^{-1}$. Hence the equal-time algebra generated by the currents coincides if equation (3.14) is satisfied and consequently this is a necessary condition for the identification of the two currents.

Now the energy-momentum tensors in the two theories may be written in the Sugawara form (Sugawara 1968, Callan *et al* 1968, Dell'Antonio *et al* 1972):

$$\theta^{\mu\nu} = \frac{\hbar}{C} (j^\mu j^\nu - \tfrac{1}{2} g^{\mu\nu} j^2) + \sigma g^{\mu\nu}$$

with $\sigma = (\alpha/\beta)(1 - \cos \beta\phi)$ and $m\bar{\psi}\psi$, respectively. If the parameters are suitably adjusted, the equal-time algebra generated by $j^\mu = J^\mu/\hbar$ and $\sigma$ coincides in the two theories. This is sufficient to guarantee the identity of the Green functions for currents in the two theories.

More details and the explicit relationship between $\phi$ and $\psi$ can be found in the literature (Coleman 1975a, b, Mandelstam 1975). $\psi$ is the quantum field operator for the kink (or soliton) of the Sine-Gordon theory since it is a local field operator creating unit topological quantum number. Further it is fermionic though the distinction between bosons and fermions is less clear in one space and one time dimension.

### *3.4. Generalising to four-dimensional space–time*

The structure we have described so far in this section is very interesting in itself but this interest would be enhanced if it could be generalised from two- to four-dimensional space–time. We shall find that in order to have topological quantum numbers resulting from the non-trivial boundary conditions asymptotically satisfied by the scalar field, long-range fields of magnetic type must be present. As the analogues of the kinks of the Sine-Gordon model we shall be led again to magnetic monopoles, but with some further insights. The analogue of the Thirring model, i.e. the field theory of monopoles, has not yet been discovered (but see Montonen and Olive (1977) for a recent conjecture).

We consider first finite-energy solutions to a scalar field theory:

$$\mathscr{L} = \tfrac{1}{2}(\partial\phi)^2 - V(\phi) \qquad V(\phi) \geqslant 0. \qquad (3.16)$$

We use $\mathscr{M}_0$ to denote the set of values of $\phi$ which minimise the potential function $V(\phi)$ describing the self-interaction of $\phi$:

$$\mathscr{M}_0 = \{\phi : V(\phi) = 0\}. \qquad (3.17)$$

We shall assume that these values are related by elements of the symmetry group, $G$, of $\mathscr{L}$, just as in the Sine-Gordon theory.

There are difficulties in obtaining finite-energy solutions to the theory specified by equation (3.16) if the dimension of space–time is bigger than two. The first arguments in this direction were presented by Derrick (1964) who demonstrated the absence of static solutions of this type by exploiting scale transformations. We will review the conclusions which can be reached in this way in §7.1. Here we will present a more general argument against the existence of topologically stable finite-energy solutions to the theory of equation (3.16).

To understand the influence of the dimension of space–time consider it to have $D$ space and one time dimension. If we have a finite-energy solution to the theory of equation (3.16), $\phi$ must tend to a point of $\mathscr{M}_0$ as we go to infinity in any direction. The possible directions in which we may go to infinity are labelled by the unit vectors in $D$-dimensional space:

$$S^{D-1}=\{\hat{\boldsymbol{r}}:\hat{\boldsymbol{r}}^2=1\}. \tag{3.18}$$

Finite energy then implies:

$$\phi_\infty(\hat{\boldsymbol{r}})=\lim_{R\to\infty}\phi(R\hat{\boldsymbol{r}})\in\mathscr{M}_0 \qquad \hat{\boldsymbol{r}}\in S^{D-1}. \tag{3.19}$$

This is the analogue of the statement that eventually the pegs hang downwards at large distances in the Sine-Gordon theory. (Finite energy also requires $\partial\phi/\partial r\to 0$ faster than $r^{-D/2}$; so that the limit in equation (3.19) exists at least for $D\geqslant 3$.) In the Sine-Gordon theory $D=1$, and $S^0$ is a discrete set consisting of two points, $\pm 1$. For $D\geqslant 2$, $S^{D-1}$ is a connected set and this is a very crucial difference. For suppose that $D\geqslant 2$ and $\mathscr{M}_0$ is discrete. Assuming $\phi_\infty$ to be continuous, it would have to be constant and so topologically trivial. To get an interesting situation $\mathscr{M}_0$ must be some sort of manifold of non-zero dimension. To achieve this the scalar field must have several components, taking its values in some representation space of $G$.

A good example of this is to take $G=\mathrm{SO}(3)$, the three-dimensional rotation group acting in an internal space, which we will refer to as isospin figuratively. Then $\phi=(\phi_1,\phi_2,\phi_3)$ and

$$\mathscr{L}=\tfrac{1}{2}\sum_{i=1}^{3}(\partial\phi_i)^2-V(\phi) \tag{3.20}$$

with

$$V(\phi)=\tfrac{1}{4}\lambda(\phi_1{}^2+\phi_2{}^2+\phi_3{}^2-a^2)^2. \tag{3.21}$$

Then $\mathscr{M}_0$, as defined by (3.17), is given by:

$$\mathscr{M}_0=\{\phi:\phi_1{}^2+\phi_2{}^2+\phi_3{}^2=a^2\} \tag{3.22}$$

that is a sphere in three dimensions (essentially $S^2$), the sort of manifold required by the arguments we have just given. Note further that any two points of $\mathscr{M}_0$ can be related by an element of SO(3). This example will constantly be used as a paradigm in the rest of this review.

We shall now argue that it is necessary to extend this model further in order to get topologically stable finite-energy solutions. The energy of a given configuration is:

$$\begin{aligned} H&=\int \mathrm{d}^Dx[\tfrac{1}{2}\dot{\phi}^2+\tfrac{1}{2}(\nabla\phi)^2+V(\phi)]\\ &\geqslant\int \mathrm{d}^Dx[\tfrac{1}{2}(\nabla\phi)^2+V(\phi)]. \end{aligned} \tag{3.23}$$

(No assumption has been made about time independence and the inequality (3.23)

will remain true even if the model is modified by the addition of further fields, provided that the effect of these fields on $H$ is the addition of further terms which are, by themselves, positive.) Now we may write $(\nabla\phi)^2$ as the sum of a radial and a transverse term, which for $D=3$ take the forms:

$$(\nabla\phi)^2=\left(\frac{\partial\phi}{\partial r}\right)^2+(\hat{r}\wedge\nabla\phi)^2 \tag{3.24}$$

and if $\phi_\infty$ is not constant this latter term is of order $r^{-2}$ as $r\to\infty$. For $D\geqslant 2$ this contribution is sufficient to ensure the divergence of (3.23), and the finite-energy requirement forces $\phi$ to be topologically trivial at infinity.

At the quantum-mechanical level this model would have some undesirable features: there would be Goldstone bosons, massless particles associated with excitations in the field components tangential to $\mathcal{M}_0$. This problem can be avoided with the Higgs mechanism where $G$ is changed into a gauge symmetry and the massless gauge particles 'eat up' the scalar Goldstone bosons, becoming massive. The massless particles which remain are the gauge mesons associated with the generators of the little group, $H$, of $\phi_\infty$. (This is discussed in greater detail later on; for a general review of gauge theories see Abers and Lee (1973) and Taylor (1976).)

Replacing the global symmetry by a gauge symmetry also circumvents the difficulties presented by the inequality (3.23), since we must replace $\partial^\mu\phi$ by:

$$\mathscr{D}^\mu\phi=\partial^\mu\phi+ieW_a{}^\mu D(T^a)\,\phi \tag{3.25}$$

where the covariant derivative, $\mathscr{D}^\mu$, involves the gauge potentials, $W_a{}^\mu$, and the representatives, $D(T^a)$, of the generators, $T^a$, of $G$ in the representation, $D$, of $G$ which acts on $\phi$. $(\nabla\phi)^2$ no longer occurs as a separate positive contribution to the energy. It is possible to have $\mathscr{D}^i\phi$ decreasing like $r^{-2}$ whilst $W_a{}^i$ and $\nabla\phi$ both decrease like $r^{-1}$. In this way we may have both finite-energy and non-trivial behaviour at infinity for $D=3$ but a subtle cancellation has to be arranged. Note that since the potentials decrease like $r^{-1}$, we expect the field strengths to decrease like $r^{-2}$, an inverse square law. We shall see this explicitly in the next subsection for $D=3$. For $D=2$ this leads to the vortex line solutions of Nielsen and Olesen (1973), which are discussed further in §7.

Finally we should mention that there are other interesting ways of generalising the features of the Sine-Gordon theory to three space dimensions. One approach is to regard the angular field variable, $\phi$, as taking its values on the circle, $S^1$; the topological quantum number is the number of times $\phi$ covers the circle as $x$ covers space, identifying positive and negative infinity. In this approach one-dimensional space has been compactified to make a circle, and the topological quantum number is the 'winding number' of the map $\phi$ defines from $S^1$ to $S^1$. Skyrme (1958, 1959, 1961a, b) generalised this to three-dimensional space by considering the non-linear $\sigma$ model in which $\phi$ takes its values in $S^3$. Assuming $\phi$ tends to a constant at infinity space may be compactified to form another three-dimensional sphere and a topological quantum number is obtained from the winding number of the map $\phi\colon S^3\to S^3$, the number of times it covers the image space. It is also possible to produce models in which there are time-dependent extended solutions which owe their stability to conventional Noether conservation laws (Lee 1976, Friedberg *et al* 1976a, b, c; see also Coleman 1975a, b).

## 4. The 't Hooft–Polyakov model of a monopole

### *4.1. The model*

Each of the apparently independent lines of thought followed in the two preceding sections has led us to consider a gauge theory with the symmetry group, $G$, spontaneously broken by the vacuum to a subgroup $H$. From §2 we expect to find electric charge quantisation (at least if the electric charge, $Q$, generates $H=\mathrm{U}(1)$) and magnetic monopoles, whilst according to §3 we expect to find extended 'soliton' solutions to such theories with long-range magnetic fields. In this way we are led to anticipate a non-singular extended solution which at large distances looks like a Dirac monopole, and further that its magnetic charge will be related to the topological quantum number specified by the boundary conditions on the scalar fields. This section will show how these expectations are indeed realised.

The sort of theory which we have been led to consider is of considerable interest in its own right. Such theories are currently thought to provide a framework for the unification of weak and electromagnetic interactions. The simplest example, the one which we will discuss in this section, is known as the Georgi–Glashow model (Georgi and Glashow 1972). It consists of an SO(3) gauge field interacting with an isovector Higgs field $\boldsymbol{\phi}$, and the Lagrangian is:

$$\mathscr{L}=-\tfrac{1}{4}G_a{}^{\mu\nu}G_{a\mu\nu}+\tfrac{1}{2}\mathscr{D}^\mu\boldsymbol{\phi}\cdot\mathscr{D}_\mu\boldsymbol{\phi}-V(\boldsymbol{\phi}) \tag{4.1}$$

with $V(\phi)$ given by equation (3.21). $G_a{}^{\mu\nu}$ is the gauge field strength:

$$G_a{}^{\mu\nu}=\partial^\mu W_a{}^\nu-\partial^\nu W_a{}^\mu-e\epsilon_{abc}W_b{}^\mu W_c{}^\nu \qquad (a=1,2,3) \tag{4.2}$$

and $W_a{}^\mu$ is the gauge potential. The covariant derivative, $\mathscr{D}^\mu\boldsymbol{\phi}$, of $\boldsymbol{\phi}$ is given by:

$$(\mathscr{D}^\mu\phi)_a=\partial^\mu\phi_a-e\epsilon_{abc}W_b{}^\mu\phi_c.$$

The quantities $\phi_a$, $G_a{}^{\mu\nu}$ and $(\mathscr{D}^\mu\phi)_a$ all transform as vectors with respect to local SO(3) rotations. (These statements are elaborated in §5.2 when we discuss the general case.) The equations of motion are:

$$(\mathscr{D}_\nu G^{\mu\nu})_a=-e\epsilon_{abc}\phi_b(\mathscr{D}^\mu\phi)_c \tag{4.3}$$

$$(\mathscr{D}^\mu\,\mathscr{D}_\mu\phi)_a=-\lambda\phi_a(\phi^2-a^2). \tag{4.4}$$

Further we have the Bianchi identities:

$$\mathscr{D}_\mu{}^*\boldsymbol{G}^{\mu\nu}=0. \tag{4.5}$$

The energy density corresponding to the Lagrangian of equation (4.1) is:

$$\theta_{00}=\tfrac{1}{2}\{(\mathscr{E}_a{}^i)^2+(\mathscr{B}_a{}^i)^2+(\Pi_a)^2+[(\mathscr{D}^i\phi)_a]^2\}+V(\boldsymbol{\phi}) \tag{4.6}$$

where

$$G_a{}^{0i}=-\mathscr{E}_a{}^i \qquad G_a{}^{ij}=-\epsilon_{ijk}\mathscr{B}_a{}^k \tag{4.7}$$

in analogy with equations (2.6), and $\Pi_a=(\mathscr{D}^0\phi)_a$. Notice that $\theta_{00}\geqslant 0$ and vanishes if, and only if:

$$G_a{}^{\mu\nu}=0 \tag{4.8}$$

$$(\mathscr{D}^\mu\phi)_a=0 \tag{4.9}$$

$$V(\phi)=\tfrac{1}{4}\lambda(\phi^2-a^2)^2=0. \tag{4.10}$$

A field configuration which satisfies $\theta_{00}=0$ everywhere, and thus equations (4.8), (4.9) and (4.10), is what we shall call a vacuum configuration. An example is:

$$\phi_a = a\delta_{a3} \qquad W_a{}^\mu = 0. \tag{4.11}$$

Since $\theta_{00}=0$ is a gauge-invariant condition, any gauge transform of equation (4.11) will also provide a vacuum configuration.

An important concept in what follows is that of the *Higgs vacuum.* We shall say that the fields in a certain region of space–time are in the Higgs vacuum if equations (4.9) and (4.10), but not necessarily equation (4.8), are satisfied. We have seen that the condition of finite energy enforces these equations asymptotically at large distances. In particular this requires $\boldsymbol{\phi} \in \mathcal{M}_0$, the set of $\boldsymbol{\phi}$ which minimise $V(\boldsymbol{\phi})$, which we introduced in equations (3.17) and (3.22). $\mathcal{M}_0$ is a two-dimensional sphere of radius $a$ in isotopic space. The little group $H_\phi$, of $\boldsymbol{\phi} \in \mathcal{M}_0$, consisting of those elements in $G=\mathrm{SO}(3)$ which leave the given $\boldsymbol{\phi}$ invariant, is just the group of rotations about the $\phi$ axis and so is isomorphic to SO(2) or, equivalently, U(1). Since the $H_\phi$ for $\boldsymbol{\phi} \in \mathcal{M}_0$ are all isomorphic we shall denote any one of them by $H$. Physically $H$ is of prime importance; it is the exact symmetry group of the theory. The original symmetry $G$ is spontaneously broken down to $H$ by $\boldsymbol{\phi}$.

Thus after symmetry breaking we are left with a U(1) gauge theory which, consequently, has all the characteristics of Maxwell's electromagnetic theory. It is reasonable to identify the U(1) with the electromagnetic gauge group. If $T^a$ $(a=1, 2, 3)$ are the SO(3) generators, this U(1) is generated by $\boldsymbol{\phi}.\boldsymbol{T}/a$ which will therefore be proportional to the electric charge $Q$. This precise correspondence with Maxwell's theory only holds in the Higgs vacuum. When equations (4.9) and (4.10) fail to hold new phenomena can occur, amongst them, as we shall see, magnetic monopoles.

We must emphasise that in seeking solutions with particle attributes we shall be treating the equations purely classically. We shall never discuss quantising them, although we should like to do this, because as yet it is not understood how to do this properly. Thus we will not have to face the difficulties of renormalisation. Nevertheless the language provided by the particle interpretation, which fields acquire on quantisation, is ingrained and we can use it to obtain heuristically expectations of what might happen if one could quantise. The spectrum that would be expected conventionally in the theory specified by equation (4.1) is shown below, the physical properties being related to the parameters of the Lagrangian to lowest order.

| | Mass | Spin | Electric charge |
|---|---|---|---|
| Higgs particle | $\mu = a(2\lambda)^{1/2}\,\hbar$ | 0 | 0 |
| Photon | 0 | $\hbar$ | 0 |
| Massive gauge particles | $M = ae\hbar = aq$ | $\hbar$ | $\pm q = \pm e\hbar$ |

The masses are calculated from the Lagrangian in the usual way by recognising that when we expand about the vacuum the coefficient of the quadratic term in the boson fields is the square of the mass divided by $2\hbar^2$. Because $|\boldsymbol{\phi}| = a$ in the vacuum, the Higgs mechanism (Higgs 1964a, b, 1966, Englert and Brout 1964, Guralnik *et al* 1964, Kibble 1967; for reviews see Coleman 1973, O'Raifeartaigh 1979) operates; i.e. charged components of the Higgs field are absorbed into the charged components of the gauge field giving it the specified mass and leaving a single massless vector

field as well as the remaining massive Higgs scalar particle. The mass divided by $\hbar$ is the inverse of the Compton wavelength (of the particle concerned) which will play an essential role in determining the scale of the finite-energy solutions.

The electric charge is obtained by comparing the SO(3) covariant derivative:

$$\partial^\mu + ieW_a^\mu T_a \qquad \text{where} \qquad (T_a)^{ij} = -i\epsilon_{aij}$$

with the electromagnetic covariant derivative of equation (2.33), $\partial^\mu + iQA^\mu/\hbar$. Identifying $\boldsymbol{\phi}.\boldsymbol{W}^\mu/a$ with $A^\mu$ we find:

$$Q = e\boldsymbol{\phi}.\boldsymbol{T}\hbar/a \tag{4.12}$$

which yields the results given above.

The expression for $Q$ given in equation (4.12) is valid in any representation. If extra fields are added to the model, $T_a$ may have any eigenvalue which is half an integer. Thus the possible eigenvalues of $Q$ are multiples of $\frac{1}{2}e\hbar$ and electric charge is quantised in these units.

The factors of $\hbar$ in the expressions we have given for masses and charges makes their quantum nature explicit.

### *4.2. Search for solutions using a simplifying ansatz*

We are seeking finite-energy non-singular classical solutions to the equations (4.3) and (4.4) which will be simple, but unlike the vacuum solution of equations (4.9)–(4.11), not constant. Since the equations of motion are complicated and non-linear some sort of strategy is required. As we argued in §§3.4 and 4.1 the finite-energy requirement forces the fields to be in the Higgs vacuum asymptotically at large distances; when this has been reached only the electromagnetic characteristics survive.

Physically one would expect the solution with lowest non-zero energy to be time-independent and to have a high degree of symmetry. We shall now seek to make this expectation more concrete by using it to derive a simplifying ansatz for the fields. *The reader who is prepared to accept this ansatz without further rationalisation may proceed immediately to the next subsection.* It was originally made by 't Hooft (1974) and Polyakov (1974) (but see Wu and Yang 1969).

The symmetries of a given solution form a group, $\mathscr{G}_0$ say, which is a subgroup of the full group, $\mathscr{G}$, of symmetries of the equations of motion. $\mathscr{G}_0$ cannot be the full group $\mathscr{G}$ since only the trivial zero solution has this symmetry. Let us consider what $\mathscr{G}_0$ might be for the solution of lowest non-zero energy. Time independence presupposes a choice of Lorentz frame in which the fields are at rest. Given this choice the equations of motion have an SO(3) spatial rotational symmetry and a translational symmetry. Further there is an internal SO(3) symmetry associated with spatially constant gauge transformations. Finally there are certain discrete symmetries:

$$P\colon \phi_a(\boldsymbol{r}) \to \phi_a(-\boldsymbol{r}) \qquad W_a^i(\boldsymbol{r}) \to -W_a^i(-\boldsymbol{r}) \qquad W_a^0(\boldsymbol{r}) \to W_a^0(-\boldsymbol{r}) \tag{4.13}$$

$$Z\colon \phi_a(\boldsymbol{r}) \to -\phi_a(\boldsymbol{r}) \qquad W_a^\mu(\boldsymbol{r}) \to W_a^\mu(\boldsymbol{r}). \tag{4.14}$$

(which generate $\mathbb{Z}_2 \times \mathbb{Z}_2$, where $\mathbb{Z}_N$ denotes the cyclic group of order $N$). The action of $P$ and $Z$ on $F^{\mu\nu} = \boldsymbol{\phi}.\boldsymbol{G}^{\mu\nu}/a$, which we identify as the electromagnetic field tensor in the Higgs vacuum, is given by:

$$P\colon F^{ij}(\boldsymbol{r}) \to F^{ij}(-\boldsymbol{r}) \qquad F^{i0}(\boldsymbol{r}) \to -F^{i0}(-\boldsymbol{r})$$

$$Z\colon F^{ij}(\boldsymbol{r}) \to -F^{ij}(\boldsymbol{r}) \qquad F^{i0}(\boldsymbol{r}) \to -F^{i0}(\boldsymbol{r}).$$

We may identify $P$ with the parity transformation since it has the appropriate action on the electromagnetic field. Both $P$ and $Z$ reverse the sign of $\nabla . \boldsymbol{B}$ and, consequently, the magnetic charge of any solution. If either $P$ or $Z$ is included in $\mathscr{G}_0$ the solution must have zero magnetic charge. This objection does not apply to $PZ$ and we will seek to include this in $\mathscr{G}_0$.

Let us turn to the possible continuous symmetries. The solution will be localised and this breaks translation invariance. This leaves us with SO(3) × SO(3), the product of spatial and isotopic rotations, but this group is itself too big. It has various covariant SO(3) subgroups: those consisting of either spatial or isotopic rotations alone, and the diagonal subgroup consisting of simultaneous and equal rotations in real and isotopic space. Invariance with respect to spatial rotations forces $\boldsymbol{\phi}$ to be constant asymptotically, leaving the boundary conditions satisfied in a topologically trivial way. Isotopic invariance forces $\boldsymbol{\phi}$ to vanish everywhere and the boundary conditions are not satisfied at all. The general ansatz invariant with respect to the diagonal group, which has generators $-\mathrm{i}\boldsymbol{r} \wedge \nabla + \boldsymbol{T}$, is:

$$\phi_a(\boldsymbol{r}) = H(aer)\, r^a/er^2 \qquad W_a{}^0(\boldsymbol{r}) = J(aer)\, r^a/er^2 \tag{4.15}$$

$$W_a{}^i(\boldsymbol{r}) = -\epsilon_{aij}\frac{r^j}{er^2}[1-K(aer)] + \frac{r^2\delta_{ai} - r^i r^a}{er^3} B(aer) + \frac{r^i r^a}{er^3} C(aer). \tag{4.16}$$

Invariance with respect to $PZ$ forces $B = C = 0$. The solution then has an invariance group $\mathscr{G}_0 \simeq \mathrm{SO}(3) \times \mathbb{Z}_2$, the SO(3) being the diagonal group in the product of spatial and isotopic rotations and the $\mathbb{Z}_2$ being the group generated by $PZ$.

Another argument leading to this ansatz has been given by Corrigan *et al* (1976). Their argument starts from the assumption of invariance under simultaneous isotopic and spatial rotations but does not implement $PZ$ invariance. The assumed invariance leads to equations (4.15) and (4.16). These are gauge-dependent statements and, to that extent, the assumptions imply the use of a conventionally chosen gauge. Further, they do not fix the gauge in that local gauge transformations generated by $\boldsymbol{r} . \boldsymbol{T}$ respect the assumed invariance. Gauge transformations of this sort may always be used to set $B$ to zero. The vanishing of $C$ then follows from the equations of motion.

Finally we note that it is possible to anticipate that the solution is a magnetic monopole with charge $-4\pi/e$. The ansatz we have obtained is spherically symmetric in that any spatial rotation may be compensated for by a global gauge transformation. Thus we may regard $-\mathrm{i}\boldsymbol{r} \wedge \nabla + \boldsymbol{T}$ as a sort of generalised rotation generator and if we identify this with $\boldsymbol{J}/\hbar$, where $\boldsymbol{J}$ is the angular momentum of a charged particle moving in the given field as in §2.2, we obtain:

$$\hat{\boldsymbol{r}} . \boldsymbol{J} = \hat{\boldsymbol{r}} . \boldsymbol{T}\hbar = Q/e.$$

Comparing with equation (2.19) we see this equals $-Qg/4\pi$ and thus the magnetic charge, $g$, has the stated value, $-4\pi/e$. We shall develop these considerations in a less heuristic fashion in §6.

### *4.3. The monopole solution*

We shall now consider the ansatz obtained in the last subsection, with $J = 0$ for the time being:

$$\phi_a = \frac{r^a}{er^2} H(aer) \qquad W_a{}^i = -\epsilon_{aij}\frac{r^j}{er^2}[1-K(aer)] \qquad W_a{}^0 = 0. \tag{4.17}$$

The asymptotic condition (4.10) implies that:

$$\phi_{\infty a}(\boldsymbol{r}) = \lim_{r\to\infty} \phi_a(r\hat{\boldsymbol{r}}) = a\hat{r}^a. \tag{4.18}$$

Thus $\phi_\infty$ maps each point on the sphere, $S^2$, of possible directions in which $\boldsymbol{r}$ may go to infinity, to the corresponding point on $\mathcal{M}_0$, which is also essentially $S^2$. It is not possible to continuously deform this into the constant map given by the vacuum configuration of equation (4.11). This, in itself, means that the monopole is stable against decay into the vacuum state.

For the ansatz of equation (4.17) the energy takes the form:

$$E = -\int \mathscr{L}\, \mathrm{d}^3\boldsymbol{r}$$
$$= \frac{4\pi a}{e}\int_0^\infty \frac{\mathrm{d}\xi}{\xi^2}\left[\xi^2\left(\frac{\mathrm{d}K}{\mathrm{d}\xi}\right)^2 + \tfrac{1}{2}\left(\xi\frac{\mathrm{d}H}{\mathrm{d}\xi} - H\right)^2 + \tfrac{1}{2}(K^2-1)^2 + K^2H^2 + \frac{\lambda}{4e^2}(H^2-\xi^2)^2\right] \tag{4.19}$$

using $\xi$ for the argument, $aer$, of $H$ and $K$. The conditions for $E$ to be stationary with respect to variations of $H$ and $K$ are:

$$\xi^2\frac{\mathrm{d}^2K}{\mathrm{d}\xi^2} = KH^2 + K(K^2-1) \tag{4.20}$$

$$\xi^2\frac{\mathrm{d}^2H}{\mathrm{d}\xi^2} = 2K^2H + \frac{\lambda}{e^2}H(H^2-\xi^2). \tag{4.21}$$

In fact, equations (4.20) and (4.21) are the equations of motion for the ansatz (4.17). This may be directly verified by laboriously substituting from equations (4.17) into equations (4.4) and (4.5). However, there is a general principle given by Faddeev (1976a) and Coleman (1975b) which obviates the necessity of doing this. Roughly (that is, without the necessary smoothness assumptions) it may be described as follows. Suppose we wish to find the stationary points of some function $F$ (here a functional) whose argument ranges over some set $X$. Let $\mathscr{G}_0$ be a group acting on $X$ which consists of symmetries of $F$. If $X_0$ denotes those points of $X$ which are left fixed by all elements of $\mathscr{G}_0$, then a stationary point of $F$ restricted to $X_0$ is also a stationary point of $F$ over $X$. Here $X$ is the set of all possible field configurations, $\mathscr{G}_0$ is the $\mathrm{SO}(3)\times\mathbb{Z}_2$ group described in §4.2 and $X_0$ are those configurations given by equations (4.17).

The appropriate boundary conditions for a finite-energy solution are

$$K-1 \leqslant \mathrm{O}(\xi) \qquad H \leqslant \mathrm{O}(\xi) \text{ as } \xi\to 0 \tag{4.22}$$

$$K\to 0 \qquad H\sim\xi \text{ sufficiently fast as } \xi\to\infty. \tag{4.23}$$

That the system of equations defined by equations (4.20)–(4.23) has a solution was first indicated by computation. A simple argument that $E$ has a lower bound of $4\pi a/e$ will be given in §4.6. That solutions do indeed exist is physically plausible and a rigorous proof has been given by Schwarz (1976). The forms of $H$ and $K$ are sketched in figure 3.

The total energy of the solution, which will be interpreted as the classical mass, is given by equation (4.19) where the integral is a function, $f(\lambda/e^2)$ say, of $\lambda/e^2$, so:

$$\text{Mass} = \frac{4\pi a}{e} f(\lambda/e^2).$$

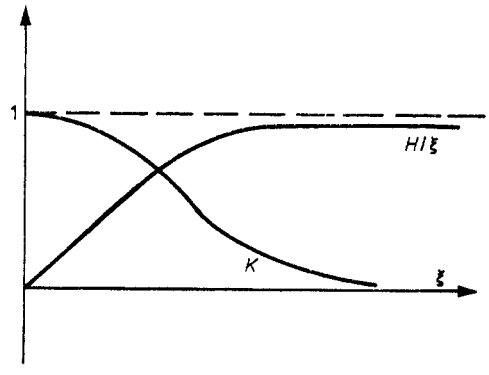


**Figure 3.**

The following values for $f$ have been calculated:

| | |
|---|---|
| $f(0)=1$ | (Prasad and Sommerfield 1975) |
| $f(0{\cdot}1)=1{\cdot}1$ | ('t Hooft 1974) |
| $f(0{\cdot}5)=1{\cdot}42$ | (Julia and Zee 1975) |
| $f(10)=1{\cdot}44$ | ('t Hooft 1974). |

For a detailed numerical study of the monopole solution (and the dyon solutions discussed in §4.8) see Bais and Primack (1976). Numerical work on the monopole mass function, $f$, is also reported in Bogomolny and Marinov (1976).

Using the conditions (4.23) we find that, to leading order at large distances:

$$G_a{}^{ij} \sim \frac{1}{er^4}\,\epsilon_{ijk}r^a r^k \sim \frac{1}{aer^3}\,\epsilon_{ijk}r^k\phi_a.$$

Only the component of $G_a{}^{ij}$ associated with the neutral vector meson, the photon, survives yielding the magnetic field:

$$B^i = -\frac{1}{e}\frac{r^i}{r^3} \tag{4.24}$$

and, on comparing with equation (2.16), we see that we have indeed a magnetic charge of magnitude:

$$g = -4\pi/e.$$

The question of how this result relates to Dirac's quantisation condition immediately presents itself, particularly in view of the fact that the 't Hooft–Polyakov monopole has been constructed within an entirely classical framework and Planck's constant nowhere enters into consideration. Now the constant $e$ is a coupling constant not an electric charge; it has the inverse dimensions to those of charge and it is related to the charge $q$ of the charged vector bosons by $q=\pm e\hbar$, as explained in §4.1. Thus $qg=-4\pi\hbar$, consistent with Dirac's condition (2.15). However, $q$ is not the smallest possible charge which might enter the theory, i.e. $q_0=\frac{1}{2}q$ because equation (4.12) relates the possible electric charges to the eigenvalues of the isotopic spin generators. This gives:

$$\frac{q_0 g}{4\pi\hbar} = -\frac{1}{2} \tag{4.25}$$

and $g$ assumes the lowest value compatible with the Dirac condition. Thus we say that the 't Hooft–Polyakov monopole carries one Dirac unit of magnetic charge. A solution with the opposite charge is obtained by applying either of the transformations $P$ and $Z$ (see (4.13) and (4.14)) to the solution we have described. Note that the 't Hooft–Polyakov solution is not invariant under the operation, $P$, although the original equations were.

Finally we consider how the asymptotic values (4.23) are approached as $\xi \to \infty$. Asymptotically the radial equations (4.20) and (4.21) become:

$$\frac{d^2K}{d\xi^2} = K \qquad \frac{d^2h}{d\xi^2} - \frac{2\lambda}{e^2}h = 0$$

where $H = h + \xi$. Consequently we deduce:

$$\begin{aligned} K &= O[\exp(-\xi)] = O[\exp(-Mr/\hbar)] \\ H - \xi &= O[\exp(-\mu\xi/M)] = O[\exp(-\mu r/\hbar)] \end{aligned} \tag{4.26}$$

where we introduced $\mu = (2\lambda)^{1/2} a\hbar$ and $M = ae\hbar$ in §4.1. The approach to the asymptotic form is thus given by the Compton wavelength of the massive particle associated in the field in question. This means that we can think of the 't Hooft–Polyakov monopole as having a definite size determined by these Compton wavelengths, inside which the massive fields play a role in providing a smooth structure and outside which they rapidly vanish, leaving a field configuration indistinguishable from that of the Dirac monopole.

### *4.4. Magnetic charge and topology*

It follows from equations (4.26) that the 't Hooft–Polyakov monopole has a finite radius, $R_0$ say, determined by the Compton wavelengths $\hbar/M$ and $\hbar/\mu$, of the heavy particles of the theory, such that outside the radius $R_0$ the field configuration is exponentially close to a Higgs vacuum; that is:

$$\mathscr{D}^\mu \boldsymbol{\phi} \equiv \partial^\mu \boldsymbol{\phi} - e\boldsymbol{W}^\mu \wedge \boldsymbol{\phi} = 0 \qquad \phi^2 = a^2 \qquad \text{for } r \gg R_0 \tag{4.27}$$

with an error of order $\exp(-r/R_0)$.

We shall now assume that *any* finite-energy solution satisfies equations (4.27) very closely, except in a finite number of compact localised regions in space corresponding to monopoles, even if the solution is time-dependent. As yet no proof of this statement exists. We shall now analyse equations (4.27) with a view to clarifying our previous remarks about the Higgs vacuum.

Given $\boldsymbol{\phi}$ outside the localised regions corresponding to the monopoles the general form of $\boldsymbol{W}^\mu$ satisfying equation (4.27) is (Corrigan *et al* 1976):

$$\boldsymbol{W}^\mu = \frac{1}{a^2 e}\boldsymbol{\phi} \wedge \partial^\mu \boldsymbol{\phi} + \frac{1}{a}\boldsymbol{\phi} A^\mu \tag{4.28}$$

where $A^\mu$ is arbitrary. It follows that:

$$\boldsymbol{G}^{\mu\nu} = \frac{1}{a}\boldsymbol{\phi} F^{\mu\nu} \tag{4.29}$$

where

$$F^{\mu\nu} = \frac{1}{a^3 e}\boldsymbol{\phi} \cdot (\partial^\mu \boldsymbol{\phi} \wedge \partial^\nu \boldsymbol{\phi}) + \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{4.30}$$

Further it follows from equation (4.27) and the field equation (4.4) that:

$$\partial_\nu F^{\mu\nu}=0 \qquad \text{and} \qquad \partial_\nu {}^*F^{\mu\nu}=0$$

which are precisely the Maxwell equations (2.3) and (2.4). We have reached the important conclusion that in the Higgs vacuum (4.27) the only non-zero component of the gauge field tensor is the component associated with the U(1) group of rotations about $\boldsymbol{\phi}$, $F^{\mu\nu}$, which satisfies Maxwell's equations. In this sense, outside the regions of the monopole, the SO(3) gauge theory is locally indistinguishable from conventional electromagnetic theory. This conclusion is unaltered by the introduction of additional charged fields.

Now we shall consider the global attributes of the Higgs vacuum by studying the magnetic flux, $g_\Sigma$, through the closed surface $\Sigma$. By Maxwell's equations $g_\Sigma$ will be non-zero only if $\Sigma$ surrounds a region in which equations (4.27) fails, a potential monopole:

$$\begin{aligned} g_\Sigma &= \int_\Sigma \boldsymbol{B}.\mathrm{d}\boldsymbol{S} \\ &= -\frac{1}{2ea^3}\int_\Sigma \epsilon_{ijk}\boldsymbol{\phi}.(\partial^j\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi})\,\mathrm{d}S^i \end{aligned} \tag{4.31}$$

using equation (4.30) and the fact that the contribution of $A^\mu$ vanishes by Stokes' theorem. Notice that the derivatives $\partial^j\boldsymbol{\phi}$ occurring in equation (4.31) are those tangential to $\Sigma$, so that the magnetic charge within $\Sigma$ depends only on the values of the Higgs field on $\Sigma$. In fact, it depends on less, for if we consider a slightly different Higgs field satisfying equation (4.27):

$$\boldsymbol{\phi}'=\boldsymbol{\phi}+\delta\boldsymbol{\phi} \qquad \boldsymbol{\phi}.\delta\boldsymbol{\phi}=0. \tag{4.32}$$

Then

$$\delta[\boldsymbol{\phi}.(\partial^j\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi})]=3\delta\boldsymbol{\phi}.(\partial^j\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi})+\partial^j[\boldsymbol{\phi}.(\delta\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi})]-\partial^k[\boldsymbol{\phi}.(\delta\boldsymbol{\phi}\wedge\partial^j\boldsymbol{\phi})].$$

The integral of the last two terms in this expression vanishes by Stokes' theorem. Further, since $\partial^j\boldsymbol{\phi}$ is perpendicular to $\boldsymbol{\phi}$, $\partial^j\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi}$ is parallel to $\boldsymbol{\phi}$ and so the remaining term vanishes by equation (4.32). Consequently a small variation in the Higgs field $\boldsymbol{\phi}$, subject to equations (4.27), produces no change in the flux, $g$. This is a fundamental result. It extends to any change in $\boldsymbol{\phi}$ which can be built up by small deformations. Such a deformation is called a *homotopy*. Examples of homotopies in the physical context under discussion are: (i) the time development of $\boldsymbol{\phi}$, (ii) the change in $\boldsymbol{\phi}$ under a continuous gauge transformation, and (iii) the change induced by altering $\Sigma$ continuously within the Higgs vacuum. Consequently $g_\Sigma$ is time-independent, gauge-invariant, and unchanged under any continuous deformation of the surface $\Sigma$ containing the monopole or monopoles.

In particular with reference to figure 4, in which the unshaded regions are close to the Higgs vacuum and the shaded regions are the monopoles:

$$g_{\Sigma_{12}}=g_{\Sigma_1}+g_{\Sigma_2}. \tag{4.33}$$

Thus $g$ is an additive 'quantum number'; to see that it is quantised note that we may write $g_\Sigma=-4\pi N/e$ where:

$$N=\frac{1}{4\pi a^3}\int_\Sigma \mathrm{d}S^i\tfrac{1}{2}\epsilon_{ijk}\boldsymbol{\phi}.(\partial^j\boldsymbol{\phi}\wedge\partial^k\boldsymbol{\phi}). \tag{4.34}$$

$N$ has the geometrical interpretation (Arafune *et al* 1975) of being the number of

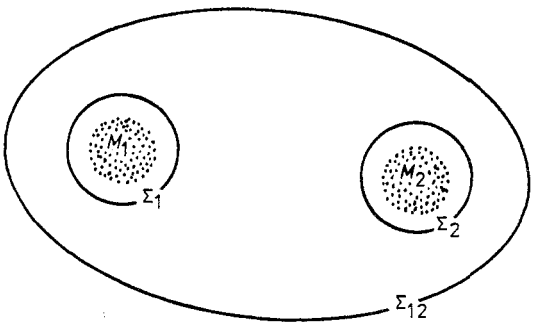


**Figure 4.**

times $\boldsymbol{\phi}(\boldsymbol{r})$ covers the sphere $\mathcal{M}_0$ as $\boldsymbol{r}$ covers $\Sigma$ once (the number of times $\Sigma$ is wrapped about $\mathcal{M}_0$ by the map $\phi: \Sigma \to \mathcal{M}_0$), since the integrand is the Jacobian of $\boldsymbol{\phi}$. Thus $N$ must be an integer; it is called by mathematicians the Brouwer degree or Poincaré–Hopf index of the map.

To show that every integer $N$ may be realised for suitable $\boldsymbol{\phi}$ consider:

$$\boldsymbol{\phi}_N(\boldsymbol{r}) = a(\cos N\chi \sin\theta, \sin N\chi \sin\theta, \cos\theta) \tag{4.35}$$

where $(r, \theta, \chi)$ are spherical polar coordinates. This covers $\mathcal{M}_0$ $N$ times as $\hat{\boldsymbol{r}}$ covers $S^2$ once, and yields $N$ in equation (4.34).

The maps $\boldsymbol{\phi}: S^2 \to \mathcal{M}_0$ can be divided into equivalence classes under homotopy; two maps are in the same homotopy class if and only if they are homotopic. It is a result in homotopy theory that if $\mathcal{M}_0$ is a sphere, $S^2$, the integer $N$ in equation (4.34) completely determines the homotopy class. Thus the magnetic charge $g_\Sigma$ depends only on the homotopy class of the map $\boldsymbol{\phi}: S^2 \to \mathcal{M}_0$.

We have seen that the magnetic charge is topologically conserved and quantised in units of $4\pi/e$ for topological reasons. This is very reminiscent of the Sine-Gordon theory described in §3. Further, since the smallest electric charge that we expect on quantising the theory is $q_0 = \frac{1}{2}e\hbar$ we have obtained Dirac's quantisation condition:

$$\frac{gq_0}{4\pi\hbar} = -\tfrac{1}{2}N$$

exactly the same as in §2, but in a different context and by topological methods.

It is believed that the homotopy classes of the Higgs field are separated by infinite potential barriers which prevent quantum transitions between them. In this case magnetic charge will be conserved quantum mechanically as well as classically.

### *4.5. The gauge relation between the Dirac string and the Higgs field*

The Dirac and 't Hooft–Polyakov monopoles differ in their internal structure. The Dirac monopole has a point singularity for which a source has to be 'put in by hand' whilst the 't Hooft–Polyakov monopole has a smooth internal structure satisfying the SO(3) gauge theory equations without the need for external sources. Outside the

structure there is just one physical degree of freedom, the electromagnetic field, which satisfies Maxwell's equations and yields a non-zero magnetic flux in each of the two cases. The only difference in this outer region is a technical one; the electromagnetic field tensor is expressed in terms of the vector potential $A^\mu$ by:

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu + (\text{extra term})$$

where in the one case the extra term is singular and involves the Dirac string (equation (2.38)) or, in the other case, is smooth and involves the Higgs field (equation (4.30)). (In the former case the string singularity cancels between the two terms, of course.) We shall now show that the 't Hooft–Polyakov form of equation (4.30) can be put into the Dirac form of equation (2.38) by a gauge transformation.

We saw in §2 that the Dirac string could be moved by a gauge transformation singular on the initial and final strings. We shall now see that the SO(3) gauge transformation which makes the Higgs field constant, thus formally putting $\boldsymbol{\phi}.(\partial^\mu\boldsymbol{\phi} \wedge \partial^\nu\boldsymbol{\phi})$ to zero in equation (4.30), is necessarily singular and automatically creates a Dirac string along its line of singularity. Conversely one can say that the 't Hooft–Polyakov approach succeeds in 'smoothing out' the Dirac string into the other SO(3) directions and we shall see how this stops the Bianchi identities preventing a net flux. A calculation similar to that now to be presented has been given by Boulware *et al* (1976), who also considered scattering on the 't Hooft–Polyakov monopole in some detail, showing that deviations from the Dirac theory occur only in deep scattering.

Consider rotating the directions of $\boldsymbol{\phi}$ and $\boldsymbol{G}^{\mu\nu}$, which are parallel at each point to $\hat{\boldsymbol{r}}$ in the Higgs vacuum, so that everywhere they point in the same direction, that of the $z$ axis, say. This cannot be done continuously throughout all space; indeed it cannot be done continuously on any sphere containing the origin. For it is impossible to find a rotation defined continuously over the unit sphere $S^2$ which rotates $\hat{\boldsymbol{r}}$ to the fixed direction $\hat{\boldsymbol{z}}$. But it can be done throughout the whole of space outside a cone with arbitrary small semi-vertical angle surrounding the negative $z$ axis. In the limit as the solid angle contained by the cone tends to zero we regain the Dirac potential of equation (2.39) and the expression for the radial magnetic field is just that of equation (2.38), complete with the Dirac string. We shall see in detail how this comes about, starting with asymptotic forms of the fields, valid in the Higgs vacuum, obtained in §4.3. (We use $\boldsymbol{\sigma}$ for the Pauli matrices; for further details of the formalism of gauge theories see §5.2.)

Let us define, in the Higgs vacuum:

$$\underline{\boldsymbol{\phi}} = \tfrac{1}{2}a\hat{\boldsymbol{r}}.\boldsymbol{\sigma} \qquad \mathbf{W}^\mu = \tfrac{1}{2}W_a{}^\mu\sigma^a = \frac{1}{2e}\,\boldsymbol{\sigma}.(\hat{\boldsymbol{r}} \wedge \partial^\mu\hat{\boldsymbol{r}})$$

and

$$\mathbf{G}^{ij} = \frac{1}{2er^2}\,\hat{\boldsymbol{r}}.\hat{\boldsymbol{\sigma}}\epsilon_{ijk}\hat{\boldsymbol{r}}^k.$$

Under a gauge transformation,

$$u = \cos\tfrac{1}{2}\psi + \mathrm{i}\sin\tfrac{1}{2}\psi\,\boldsymbol{k}.\boldsymbol{\sigma} \in \mathrm{SU}(2) \qquad \text{where} \qquad \boldsymbol{k}^2 = 1$$

$$\mathbf{W}^\mu \to u\mathbf{W}^\mu u^{-1} + \frac{\mathrm{i}}{e}(\partial^\mu u)\,u^{-1} \tag{4.36}$$

$$\mathbf{G}^{\mu\nu} \to u\mathbf{G}^{\mu\nu}u^{-1}.$$

Now

$$u\boldsymbol{\sigma}u^{-1}=\boldsymbol{\sigma}\cos\psi+\boldsymbol{k}\wedge\boldsymbol{\sigma}\sin\psi+(1-\cos\psi)\,\boldsymbol{k}(\boldsymbol{k}.\boldsymbol{\sigma}).$$

So that if we choose:

$$\boldsymbol{k}=\hat{\boldsymbol{\chi}}=\hat{\boldsymbol{z}}\wedge\hat{\boldsymbol{r}}/\sin\theta$$

$$\boldsymbol{u}\boldsymbol{r}.\boldsymbol{\sigma}\boldsymbol{u}^{-1}=\boldsymbol{r}.\boldsymbol{\sigma}\frac{\sin(\theta-\psi)}{\sin\theta}+r\hat{\boldsymbol{z}}.\boldsymbol{\sigma}\frac{\sin\psi}{\sin\theta} \tag{4.37}$$

where $\hat{\boldsymbol{z}}.\hat{\boldsymbol{r}}=\cos\theta$. We now choose $\psi(\theta)$ to be a suitable differentiable function of $\theta$ with $\psi(\pi)=0$ and $\psi(\theta)=\theta$ for $0\leqslant\theta\leqslant\pi-\epsilon$, and consider a sequence of such functions with $\epsilon\downarrow 0$ so that $\psi(\theta)\uparrow\theta$ (see figure 5). In the limit $G_a{}^{ij}\to\delta_{a3}\epsilon_{ijk}r^k/er^3$.

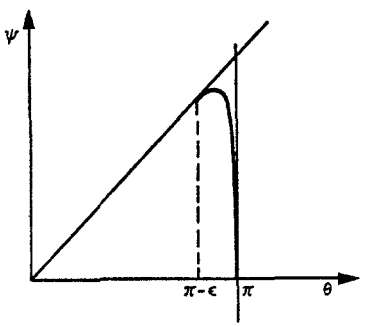


**Figure 5.**

Under the gauge transformation $u$:

$$\mathbf{W}_r\to 0 \qquad \mathbf{W}_\theta\to-\frac{1}{2er}(1-\psi')\,\hat{\boldsymbol{\chi}}.\boldsymbol{\sigma}$$

and

$$\mathbf{W}_\chi\to\frac{1}{2er}\frac{\sin(\theta-\psi)}{\sin\theta}\,\hat{\boldsymbol{\theta}}.\boldsymbol{\sigma}-\frac{1}{2er}\frac{(1-\cos\psi)}{\sin\theta}\,\hat{\boldsymbol{z}}.\boldsymbol{\sigma}.$$

Let us introduce the potential $A^\mu=\boldsymbol{\phi}.\boldsymbol{W}^\mu/a$. In the limit of $\psi\uparrow\theta$, $A^\mu$ becomes the Dirac potential of equation (2.39):

$$\boldsymbol{A}=-\frac{1}{er}\frac{(1-\cos\theta)}{\sin\theta}\,\hat{\boldsymbol{\chi}}.$$

The field tensor:

$$F^{\mu\nu}=\boldsymbol{\phi}.\boldsymbol{G}^{\mu\nu}/a$$

$$=\partial^\mu A^\nu-\partial^\nu A^\mu+\frac{e}{a}\boldsymbol{\phi}.(\boldsymbol{W}^\mu\wedge\boldsymbol{W}^\nu).$$

In the limit in which $\psi\uparrow\theta$:

$$\frac{e}{a}\boldsymbol{\phi}.(\boldsymbol{W}^i\wedge\boldsymbol{W}^j)\sim\frac{1}{er^2}(1+\psi')\frac{\sin(\theta-\psi)}{\sin\theta}\epsilon_{ijk}\hat{r}^k$$

$$\to-\frac{4\pi}{e}\epsilon_{ij3}\,\delta(x)\,\delta(y)\,\theta(-z)$$

yielding the Dirac representation of the magnetic field complete with the string, which cancels between the two terms in the expression for $F^{\mu\nu}$.

### *4.6. The Bogomolny bound on the monopole mass*

An important feature of a monopole solution with a smooth internal structure is that its mass is calculable, in contrast to the Dirac monopole which, because an external source has to be supplied in an *ad hoc* fashion, has an arbitrary mass. Further this arbitrary mass suffers an infinite renormalisation due to the fact that the self mass of the inverse square law magnetic field diverges at its origin.

In the Higgs vacuum the electromagnetic tensor $F^{\mu\nu}=\boldsymbol{\phi}.\boldsymbol{G}^{\mu\nu}/a$. For any solution the magnetic charge:

$$\begin{aligned} g&=\int \boldsymbol{B}.\mathrm{d}\boldsymbol{S}=\frac{1}{a}\int \mathscr{B}_a{}^k\phi_a\,\mathrm{d}S^k \\ &=\frac{1}{a}\int \mathscr{B}_a{}^k(\mathscr{D}^k\phi)_a\,\mathrm{d}^3r \end{aligned} \tag{4.38}$$

where the surface integral is to be understood as taken in a limiting sense over the sphere at infinity. $\mathscr{B}_a{}^k$ is as defined in equation (4.7) and we have used the Bianchi identities of equation (4.5) which give $\mathscr{D}^k\mathscr{B}_a{}^k=0$. Similarly, employing the equations of motion (4.3), the electric charge:

$$q=\int \boldsymbol{E}.\mathrm{d}\boldsymbol{S}=\frac{1}{a}\int \mathscr{E}_a{}^k(\mathscr{D}^k\phi)_a\,\mathrm{d}^3r. \tag{4.39}$$

Consider the centre-of-mass frame of the monopole. Its mass is given by:

$$\begin{aligned} M&=\int \mathrm{d}^3r\{\tfrac{1}{2}[(\mathscr{E}_a{}^k)^2+(\mathscr{B}_a{}^k)^2+(\mathscr{D}^0\boldsymbol{\phi})^2+(\mathscr{D}^i\boldsymbol{\phi})^2]+V(\phi)\} \\ &\geqslant \int \mathrm{d}^3r\tfrac{1}{2}\{(\mathscr{E}_a{}^k)^2+(\mathscr{B}_a{}^k)^2+(\mathscr{D}^i\boldsymbol{\phi})^2\} \\ &=\tfrac{1}{2}\int \mathrm{d}^3r\{\mathscr{E}_a{}^k-(\mathscr{D}^k\phi)_a\sin\theta\}^2+\tfrac{1}{2}\int \mathrm{d}^3r\{\mathscr{B}_a{}^k-(\mathscr{D}^k\phi)_a\cos\theta\}^2+a(q\sin\theta+g\cos\theta) \\ &\geqslant a(q\sin\theta+g\cos\theta) \end{aligned} \tag{4.40}$$

for any real angle $\theta$. Choosing $\theta$ to obtain the most stringent inequality:

$$M\geqslant a(q^2+g^2)^{1/2}. \tag{4.41}$$

For the 't Hooft–Polyakov monopole:

$$M\geqslant a|g|. \tag{4.42}$$

These bounds were first obtained by Bogomolny (1976) and Faddeev (1976a, b). Here we have followed the treatment of Coleman *et al* (1977).

The existence of a lower bound on the mass in a given sector does not necessarily guarantee the existence of a time-independent solution in that sector. For example, the sector with $g=2(-4\pi/e)$ and $q=0$ presumably contains solutions describing the interaction of two like monopoles whose energy can always be reduced by further separation. Only in the single monopole sectors with $g=\pm 4\pi/e$ have time-independent solutions been found.

Using the value of the magnetic charge $|g|=4\pi/e$ we can relate the monopole

mass $M_g$ to the mass of the heavy gauge boson $M_q = ae\hbar = qa$:

$$M_g \geqslant \frac{4\pi\hbar}{q^2} M_q = \frac{\nu}{\alpha} M_q$$

where $\alpha$ is the fine-structure constant and $\nu = 1$ or $\frac{1}{4}$ depending on whether the charge on the electron is $q$ or $\frac{1}{2}q$. $M_g$ is thus much larger than $M_q$, which itself would be very large if we estimate its value at that usually assigned to the intermediate vector boson in unified theories. This puts the monopole well outside the range of current observations.

### *4.7. The Bogomolny–Prasad–Sommerfield monopole*

We shall now consider whether it is possible to saturate the bound (4.42) and obtain a solution with $M = a|g|$. From the analysis of §4.6 this clearly requires the following to hold throughout space:

$$\mathscr{D}^0\boldsymbol{\phi} = 0 \qquad \mathscr{E}_a{}^i = 0 \tag{4.43}$$

$$\mathscr{B}_a{}^i = \pm(\mathscr{D}^i\boldsymbol{\phi})_a \qquad \text{as} \qquad g \gtrless 0 \tag{4.44}$$

$$V(\boldsymbol{\phi}) = 0. \tag{4.45}$$

The last condition can only be realised if the coupling constant $\lambda$ vanishes. However, we will understand this condition in the limiting sense $\lambda \downarrow 0$ and thus retain, as a vestige of $V$, the boundary condition:

$$|\boldsymbol{\phi}| \to a \qquad \text{as} \qquad r \to \infty. \tag{4.46}$$

This guarantees that the charges are still defined and quantised as in §§4.4 and 4.6. It is easy to show that equations (4.43) and (4.44), together with the Bianchi identities, $(\mathscr{D}^i\mathscr{B}^i)_a = 0$, imply the equations of motion (4.3) and (4.4) with $\lambda = 0$:

$$\mathscr{D}_\nu \boldsymbol{G}^{\mu\nu} = e\boldsymbol{\phi} \wedge \mathscr{D}^\mu\boldsymbol{\phi} \qquad \mathscr{D}_\mu \mathscr{D}^\mu\boldsymbol{\phi} = 0. \tag{4.47}$$

Equation (4.44) has the virtue of being a first-order differential equation (Bogomolny 1976, Coleman *et al* 1977). Substituting the ansatz of equation (4.17) into this equation yields:

$$\xi \frac{\mathrm{d}K}{\mathrm{d}\xi} = -KH \qquad \xi \frac{\mathrm{d}H}{\mathrm{d}\xi} = H - (K^2 - 1) \tag{4.48}$$

which naturally imply equations (4.20) and (4.21). The change of variables $-H = 1 + \xi h$ and $K = \xi k$ leads to $k' = hk$ and $h' = k^2$. Using the asymptotic boundary condition (4.23) we obtain $h^2 - k^2 = 1$, leading to the solution

$$H = H_0(\xi) = \xi \coth \xi - 1 \qquad K = K_0(\xi) = \xi/\sinh \xi. \tag{4.49}$$

This remarkable and useful solution in terms of elementary functions was first obtained by Prasad and Sommerfield (1975) by guesswork during some numerical investigations.

As $\xi \to \infty$, $H$ approaches its asymptotic form rather slowly:

$$H - \xi = 1 + \mathrm{O}[\exp(-\xi)]$$

much slower than the exponential behaviour given by equations (4.26) in general. There is no contradiction however since here $\mu = 0$. The Higgs field is now massless

and long range like the photon. Further, because of equation (4.44) the contributions of these two fields to the mass density are equal, giving a density in the tail which is double that of the Dirac or 't Hooft–Polyakov monopole (for $\lambda > 0$). This would have observable consequences within the context of the model, e.g. by gravitational interactions. Further the long-range force exerted by the Higgs field is always attractive and is found to be (Manton 1977) equal in magnitude, for static BPS monopoles, to the inverse square law magnetic force. For oppositely charged monopoles these effects reinforce one another but for equally charged monopoles they exactly cancel. Thus the Bogomolny–Prasad–Sommerfield (BPS) monopole, unlike the general 't Hooft–Polyakov monopole, differs in essential features from the Dirac monopole even at large distances and is not localised in the same way.

Note that for the BPS monopole the mass density is simply:

$$(\mathscr{D}^i\boldsymbol{\phi})^2 = \partial^i(\boldsymbol{\phi}\,.\,\mathscr{D}^i\boldsymbol{\phi})$$

$$= \nabla^2(\phi^2) = \frac{1}{r}\frac{\mathrm{d}^2}{\mathrm{d}r^2}\left(\frac{H^2}{e^2 r}\right).$$

But $H(\xi) = \frac{1}{6}\xi^2 + \mathrm{O}(\xi^2)$ for small $\xi$ and so the mass density at the origin is finite, not merely integrable.

### *4.8. Dyons*

The monopole solution of 't Hooft and Polyakov obtained in §4.3 was electrically neutral because the condition $W_a{}^0 = 0$ was imposed in equation (4.17). This is not a necessary consequence of spherical symmetry in the sense of §4.2. Julia and Zee (1975) obtained spherically symmetric solutions with $W_a{}^0 = J(aer)\, r^a/er^2$ as in equation (4.15). Because of the arguments of §4.4 the magnetic charge is quantised and further because of the symmetry is $g = -4\pi/e$. But the electric charge is arbitrary, at least classically. Following Schwinger (1969) such solutions are called dyons (see §2.7). The expression for the energy now takes the form:

$$E = \frac{4\pi a}{e}\int_0^\infty \frac{\mathrm{d}\xi}{\xi^2}\left[\xi^2\left(\frac{\mathrm{d}K}{\mathrm{d}\xi}\right)^2 + \frac{1}{2}\left(\xi\frac{\mathrm{d}H}{\mathrm{d}\xi} - H\right)^2 + \frac{1}{2}\left(\xi\frac{\mathrm{d}J}{\mathrm{d}\xi} - J\right)^2 + \tfrac{1}{2}(K^2-1)^2 \right.$$
$$\left. + K^2(H^2 - J^2) + \frac{\lambda}{4e}(H^2 - \xi^2)^2\right]. \qquad (4.50)$$

Using again the argument of Faddeev (1976a) and Coleman (1975b) we can obtain the equations of motion by applying the variation principle directly to equation (4.50). The resulting equations:

$$\xi^2\frac{\mathrm{d}^2K}{\mathrm{d}\xi^2} = K(H^2 - J^2) + K(K^2 - 1)$$

$$\xi^2\frac{\mathrm{d}^2H}{\mathrm{d}\xi^2} = 2K^2H + \frac{2\lambda}{e^2}(H^2 - \xi^2)\,H \qquad (4.51)$$

$$\xi^2\frac{\mathrm{d}^2J}{\mathrm{d}\xi^2} = 2K^2J.$$

The appropriate boundary conditions for finite-energy solutions are:

$$\begin{aligned} &K-1\leqslant O(\xi),\ H\leqslant O(\xi),\ J\leqslant O(\xi) \qquad \text{as} \qquad \xi\to 0 \\ &K\to 0,\ H\sim\xi \text{ sufficiently fast},\ J\leqslant O(\xi) \qquad \text{as} \qquad \xi\to\infty. \end{aligned} \tag{4.52}$$

In integrating these equations there arises an arbitrary constant related to the electric charge, $q$. The bound (4.41) applied to the resultant mass and may be saturated by taking $\lambda=0$:

$$\mathscr{D}^0\boldsymbol{\phi}=0 \qquad \mathscr{E}_a{}^k=(\mathscr{D}^k\boldsymbol{\phi})_a\sin\theta \qquad \mathscr{B}_a{}^k=(\mathscr{D}^k\boldsymbol{\phi})_a\cos\theta. \tag{4.53}$$

Using the spherically symmetric ansatz in these equations leads to (Prasad and Sommerfield 1975, Bogomolny 1976):

$$\begin{aligned} H(\xi)&=H_0(\xi\cos\theta)/\cos\theta \\ J(\xi)&=H_0(\xi\cos\theta)/\tan\theta \\ K(\theta)&=K_0(\xi\cos\theta) \end{aligned}$$

where $H_0$ and $K_0$ are the functions given in equations (4.49). From equations (4.38), (4.39) and (4.53) we see that the electric charge of the dyon is related to its magnetic charge by:

$$q=g\tan\theta=-\frac{4\pi}{e}\tan\theta.$$

Semiclassical arguments have been used to argue that in a proper quantum-mechanical treatment $q$ must be quantised (Tomboulis and Woo 1976, Gervais *et al* 1976):

$$q=n\hbar e \qquad n=0,\ \pm 1,\ \pm 2,\ \ldots. \tag{4.54}$$

It is interesting that $n=\pm\frac{1}{2}$ is excluded even though allowed by the Dirac condition.

### *4.9. Candidates for the magnetic current*

It has been argued in §§4.1 and 4.4 that in the Higgs vacuum (specified by $\mathscr{D}^\mu\boldsymbol{\phi}=0$, $\phi^2=a^2$) the electromagnetic tensor $F^{\mu\nu}$ should be identified with the component of $\boldsymbol{G}^{\mu\nu}$ in the direction of $\boldsymbol{\phi}$, which corresponds to the unbroken symmetry. Further we saw that this was the only remaining component, $\boldsymbol{G}^{\mu\nu}=\boldsymbol{\phi}F^{\mu\nu}/a$. We shall now discuss the extent to which it is possible to identify $F^{\mu\nu}$ inside a monopole. From the equation:

$$\partial_\nu{}^*F^{\mu\nu}=-k^\mu$$

we see that this is equivalent to determining the magnetic charge density throughout the monopole.

It has been made clear by Coleman (1975b) that there is no unique prescription for $F^{\mu\nu}$ outside the Higgs vacuum. If we could probe the interior of the monopole with a magnetometer what it would measure would depend on its detailed mechanism and how it responded to the other degrees of freedom such as the $\boldsymbol{\phi}$ field. We will consider two proposals that have been made.

Originally 't Hooft (1974) suggested:

$$F^{\mu\nu}=\hat{\boldsymbol{\phi}}\cdot\boldsymbol{G}^{\mu\nu}+\frac{1}{e}\hat{\boldsymbol{\phi}}\cdot(\mathscr{D}^\mu\hat{\boldsymbol{\phi}}\wedge\mathscr{D}^\nu\hat{\boldsymbol{\phi}}) \qquad \hat{\boldsymbol{\phi}}=\boldsymbol{\phi}/|\boldsymbol{\phi}|. \tag{4.55}$$

This is gauge-invariant and reduces to the desired form in the Higgs vacuum. Further it has the interesting property:

$$\partial_\nu {}^*F^{\mu\nu}=0 \qquad \text{if} \qquad \boldsymbol{\phi}\neq 0. \tag{4.56}$$

To see this note that it is always possible to gauge transform $\hat{\boldsymbol{\phi}}$ to a constant in a neighbourhood of any point at which $\boldsymbol{\phi}\neq 0$. In that gauge:

$$F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu \qquad \text{for} \qquad A^\mu=\hat{\boldsymbol{\phi}}.\boldsymbol{W}^\mu$$

and equation (4.56) follows in that gauge and, because it is gauge-invariant, in any gauge. Thus with this definition of $F^{\mu\nu}$, magnetic charge can only reside at zeros of the Higgs field. For the 't Hooft–Polyakov solution it must be all concentrated at the origin. The definition of equation (4.55) seems to us unsatisfactory for the reason that it leads to point singularities when the essential difference between the 't Hooft–Polyakov and Dirac monopoles seems to be that in the former case singularities have been smoothed out.

Another proposal (Bogomolny 1976, Faddeev 1976a, b) is simply:

$$F^{\mu\nu}=\boldsymbol{\phi}.\boldsymbol{G}^{\mu\nu}/a. \tag{4.57}$$

The corresponding magnetic current is:

$$\begin{aligned} k^\mu &= \partial^\nu(\boldsymbol{\phi}.{}^*\boldsymbol{G}^{\mu\nu})/a \\ &= (\mathscr{D}^\nu\boldsymbol{\phi}).{}^*\boldsymbol{G}^{\mu\nu}/a \end{aligned} \tag{4.58}$$

using the Bianchi identities of equation (4.5). The conservation of the magnetic current follows, without use of the equations of motion, from its definition as the divergence of an antisymmetric tensor. Further, $k^0$ involves no canonical momenta and is the spatial divergence of the magnetic field $\phi_a\mathscr{B}_a{}^i/a$, as in equation (4.38). Thus it has much in common with the topological current (3.8) of the Sine-Gordon theory, in particular the properties we listed in §3.2. Finally we note that for the BPS monopole this charge density is everywhere proportional to the mass density and so completely smooth.

## 5. Macroscopic properties of generalised monopoles

### *5.1. Larger gauge groups*

The 't Hooft–Polyakov monopole, described in the previous section, gives a clear picture of a magnetic monopole associated with an exact electromagnetic U(1) gauge group that has a definite internal structure and a calculable mass, even though it is too heavy to be directly relevant to contemporary physics.

Since the time of Dirac's original work on monopoles it has become increasingly apparent that gauge symmetry groups play an important and perhaps universal role in the theory of elementary particle interactions. The exact gauge symmetry group of nature may very well be larger than the U(1) of electromagnetism. There might be an exact strong gauge symmetry group, the SU(3) of colour, and a weak gauge group, exact in some approximation. It is therefore worthwhile asking what sort of generalised monopoles can occur for these gauge groups, and to investigate their properties. The groups which appear to be of physical interest at the moment seem

to be sufficiently complicated and uncertain that it appears to be sensible to work with an arbitrary exact gauge group, $H$, requiring only that it be compact. We might even gain information about possible interrelations between strong and weak interactions, which would be of the greatest interest.

In order to classify the way in which the Higgs field realises its possible boundary conditions we shall have to go further into homotopy theory (which we mentioned in §4.4). This sharpening of our mathematical equipment will enable us to relate the homotopy class to a more physical concept, a generalised magnetic charge. The present section is devoted to this analysis.

The properties discussed in this section may be thought of as *macroscopic* in that they are large-scale properties, which may be characterised by calculating generalised flux integrals in the Higgs vacuum, without probing the internal structure of the monopole. This internal structure is largely determined by the gauge group $G$ within which $H$ is embedded and which is spontaneously broken down to $H$ by the vacuum. In this section we shall always assume that $G$ is compact and connected. We shall return to the study of internal structure in §6 when we discuss smooth solutions to theories with gauge group $G$ larger than SU(2). There $G$ will play an important role but in this section we shall see how its importance can be suppressed and the exact gauge group, $H$, brought to the fore.

We shall describe these macroscopic properties first in terms of the values of the Higgs field outside the monopole. In §5.4 this leads us to topological quantum numbers, generalising the soliton number of the Sine-Gordon theory which, under suitable assumptions, satisfy Abelian combination laws. Using the structure of the Higgs vacuum, analysed in §5.3, and certain results on homotopy theory, it will be argued in §5.5 that the structure of the topological conservation laws depends essentially on the global properties of the exact gauge group $H$. This suggests that it should be possible to dispense with the Higgs fields and reformulate the topological quantum number entirely in terms of the $H$ gauge fields. This construction will be performed in §5.6, using a sort of non-Abelian version of Stokes' theorem. It is interesting in itself and also has important applications. In §5.7 we shall use it to derive a generalisation of Dirac's quantisation condition in the case where $H$ has the structure $\mathrm{U}(1)\times\mathrm{K}$, at least locally; we may think of U(1) as the electromagnetic group and K as the colour group.

One difficulty with attaching physical significance to these topological quantum numbers is that they appear to be always Abelian. However, there are indications that this may not be the whole truth. It is possible that a non-Abelian group structure may play some role but that only vestiges of it remain at the classical level. Using again the analysis of §5.6 to obtain a rather general quantisation condition, we shall outline in §5.8 the possibility that a hidden 'dual' gauge group, $H^{\mathrm{v}}$, may play a role in classifying monopole states.

Having related the macroscopic properties to $H$ we shall be able to relate the formalism used here, based on the Higgs mechanism, with that of Wu and Yang (1975, 1976) in §5.9.

### *5.2. Review of general gauge theory formalism*

We discussed the formalism of gauge theories for the U(1) of electromagnetism in §2 and, briefly, for SO(3) in §4 (Yang and Mills 1954, Shaw 1955). We now wish to give a more detailed treatment of a general gauge group $G$ (Utiyama 1956, Glashow

and Gell-Mann 1961) assumed compact and connected, following Coleman (1973, 1975b).

We can regard $G$ as a group of matrices by taking any faithful (i.e. one to one) representation of $G$. Suppose $\{T^a\}$ is a set of Hermitian generators of $G$, i.e. a basis for the Lie algebra, $L(G)$, of $G$. Let $\phi$ be a Lorentz scalar field transforming under a (real or complex) representation, $D$, of $G$:

$$\phi \to D(g)\,\phi. \tag{5.1}$$

Local gauge transformations are defined by taking $g$ in equation (5.1) to be a function over space–time: $g \equiv g(x)$. Under such transformations:

$$\partial^\mu \phi \to D(g)\,\partial^\mu \phi + \partial^\mu D(g)\,\phi.$$

To remove the unwanted second term in the result of this transformation we introduce gauge fields $W_a{}^\mu$ and associate with them a matrix in the Lie algebra of $G$:

$$\mathbf{W}^\mu = W_a{}^\mu T^a \in L(G) \tag{5.2}$$

using the summation convention. If we specify that under a gauge transformation:

$$\mathbf{W}^\mu \to g\mathbf{W}^\mu g^{-1} + \frac{\mathrm{i}}{e}(\partial^\mu g)\,g^{-1} \tag{5.3}$$

then the modified derivative:

$$\begin{aligned} \mathscr{D}^\mu \phi &= \partial^\mu \phi + \mathrm{i}eD(\mathbf{W}^\mu)\,\phi \\ &\to D(g)\{\partial^\mu + \mathrm{i}eD(\mathbf{W}^\mu)\}\,\phi + \{\partial^\mu D(g)\,D(g^{-1}) - D(\partial^\mu g g^{-1})\}\,D(g)\,\phi \end{aligned} \tag{5.4}$$

$$= D(g)\;\mathscr{D}^\mu \phi \tag{5.5}$$

and so transforms covariantly; it is called the covariant derivative.

To define the field tensor for non-Abelian gauge fields in a way which makes its transformation properties clear, consider:

$$\begin{aligned} [\mathscr{D}^\mu,\ \mathscr{D}^\nu]\,\phi &= \mathscr{D}^\mu(\mathscr{D}^\nu \phi) - \mathscr{D}^\nu(\mathscr{D}^\mu \phi) \\ &= [\partial^\mu + \mathrm{i}eD(\mathbf{W}^\mu),\ \partial^\nu + \mathrm{i}eD(\mathbf{W}^\nu)]\,\phi \\ &= \mathrm{i}e\{D(\partial^\mu \mathbf{W}^\nu) - D(\partial^\nu \mathbf{W}^\mu) + \mathrm{i}eD([\mathbf{W}^\mu,\ \mathbf{W}^\nu])\}\,\phi. \end{aligned}$$

Consequently, if we define the antisymmetric gauge field tensor by:

$$\mathbf{G}^{\mu\nu} = G_a{}^{\mu\nu} T^a = \partial^\mu \mathbf{W}^\nu - \partial^\nu \mathbf{W}^\mu + \mathrm{i}e[\mathbf{W}^\mu,\ \mathbf{W}^\nu] \tag{5.6}$$

we have:

$$[\,\mathscr{D}^\mu,\ \mathscr{D}^\nu]\,\phi = \mathrm{i}eD(\mathbf{G}^{\mu\nu})\,\phi. \tag{5.7}$$

From this equation, which holds for any $\phi$, we may deduce the effect of a gauge transformation on $G_a{}^{\mu\nu}$. From equation (5.5) we see that under a gauge transformation, $\mathscr{D}^\mu\,\mathscr{D}^\nu \phi \to D(g)\;\mathscr{D}^\mu\,\mathscr{D}^\nu \phi$ and, consequently, from equation (5.7):

$$\mathbf{G}^{\mu\nu} \to g\mathbf{G}^{\mu\nu} g^{-1}. \tag{5.8}$$

Notice that $\mathbf{G}^{\mu\nu}$ transforms *covariantly* according to the adjoint representation of the group; it is only invariant for an Abelian group. The adjoint representation of a Lie group is the representation of the same dimension as the group and defined by:

$$\xi_a \to \xi_a' = D_{ab}(g)\,\xi_b \qquad \text{where} \qquad \boldsymbol{\xi}' = \xi_a' T^a = g\boldsymbol{\xi} g^{-1}. \tag{5.9}$$

The generalisation of the homogeneous Maxwell equations of equations (2.2) and (2.4) follows from the Jacobi identity for the differential operators $\mathscr{D}^\lambda$. The Jacobi identity reads:

$$[\mathscr{D}^\lambda, [\mathscr{D}^\mu, \mathscr{D}^\nu]]+[\mathscr{D}^\mu, [\mathscr{D}^\nu, \mathscr{D}^\lambda]]+[\mathscr{D}^\nu, [\mathscr{D}^\lambda, \mathscr{D}^\mu]]=0.$$

We can apply this to any $\phi$ using equation (5.7):

$$\begin{aligned}[\mathscr{D}^\lambda, [\mathscr{D}^\mu, \mathscr{D}^\nu]]\,\phi &= ie[\mathscr{D}^\lambda, D(\mathbf{G}^{\mu\nu})]\,\phi \\ &= ieD(\mathscr{D}^\lambda \mathbf{G}^{\mu\nu})\,\phi\end{aligned}$$

where

$$\mathscr{D}^\lambda \mathbf{G}^{\mu\nu} = \partial^\lambda \mathbf{G}^{\mu\nu} + ie[\mathbf{W}^\lambda, \mathbf{G}^{\mu\nu}] \tag{5.10}$$

the appropriate form of the covariant derivative for the adjoint representation. Thus the Jacobi identity yields:

$$\mathscr{D}^\lambda \mathbf{G}^{\mu\nu} + \mathscr{D}^\mu \mathbf{G}^{\nu\lambda} + \mathscr{D}^\nu \mathbf{G}^{\lambda\mu} = 0. \tag{5.11}$$

We can rewrite these Bianchi identities in a concise form by using the dual field tensor:

$$^*\mathbf{G}^{\lambda\mu} = \tfrac{1}{2}\epsilon^{\lambda\mu\nu\rho}\mathbf{G}_{\nu\rho}. \tag{5.12}$$

Then equation (5.11) becomes:

$$\mathscr{D}_\nu {}^*\mathbf{G}^{\mu\nu} = 0 \tag{5.13}$$

in exact analogy with equation (2.4).

This completes the discussion of the kinematics of gauge fields, but before we proceed to discuss the dynamics following from a gauge-invariant Lagrangian, we give the form of infinitesimal gauge transformations for the sake of completeness. Under the infinitesimal gauge transformation:

$$g(x) = 1 - iT^a \epsilon_a(x) \tag{5.14}$$

$\phi \to \phi + \delta\phi$, etc, where:

$$\delta\phi = -i\epsilon_a D(T^a)\,\phi \tag{5.15}$$

$$\delta W_a{}^\mu = C^{bc}{}_a \epsilon_b W_c{}^\mu + \frac{1}{e}\partial^\mu \epsilon_a \tag{5.16}$$

$$\delta G_a{}^{\mu\nu} = C^{bc}{}_a \epsilon_b G_c{}^{\mu\nu}. \tag{5.17}$$

In these equations $C^{ab}{}_c$ are the structure constants of the group (corresponding to the basis $\{T^a\}$ of the Lie algebra):

$$[T^a, T^b] = iC^{ab}{}_c T^c. \tag{5.18}$$

Since the group $G$ is assumed to be compact we can always arrange that

$$\mathrm{Tr}\,(T^a T^b) = \kappa\delta^{ab} \tag{5.19}$$

and then $C^{ab}{}_c$ is totally antisymmetric in $a$, $b$ and $c$.

The Lagrangian density:

$$\mathscr{L} = -\tfrac{1}{4}G_a{}^{\mu\nu}G_{a\mu\nu} + (\mathscr{D}^\mu\phi)^\dagger\,\mathscr{D}_\mu\phi - V(\phi) \tag{5.20}$$

where $V(\phi) \geqslant 0$, is gauge-invariant provided that $V$ is symmetric under $G$; i.e.:

$$V(D(g)\,\phi) = V(\phi) \tag{5.21}$$

and we choose the basis $\{T^a\}$ so that equation (5.19) holds. The invariance of the field tensor term follows from:

$$G_a{}^{\mu\nu}G_{a\mu\nu} = \frac{1}{\kappa}\,\mathrm{Tr}\,(\mathbf{G}^{\mu\nu}\mathbf{G}_{\mu\nu})$$

$$\to \frac{1}{\kappa}\,\mathrm{Tr}\,(g\mathbf{G}^{\mu\nu}g^{-1}g\mathbf{G}_{\mu\nu}g^{-1})$$

$$= \frac{1}{\kappa}\,\mathrm{Tr}\,(\mathbf{G}^{\mu\nu}\mathbf{G}_{\mu\nu}). \tag{5.22}$$

The Lagrangian density of equation (5.20) leads to the equations of motion:

$$(\mathscr{D}^\mu\mathscr{D}_\mu\phi)_a = -\partial V/\partial\phi_a \tag{5.23}$$

$$\mathscr{D}_\nu\mathbf{G}^{\mu\nu} = -\mathbf{j}^\mu \tag{5.24}$$

where

$$j_a{}^\mu = \mathrm{i}e\phi^\dagger D(T^a)\ \mathscr{D}^\mu\phi - \mathrm{i}e(\mathscr{D}^\mu\phi)^\dagger D(T^a)\,\phi. \tag{5.25}$$

(In the case of a real scalar field the Lagrangian (5.20) is conventionally replaced by

$$\mathscr{L} = -\tfrac{1}{4}G_a{}^{\mu\nu}G_{a\mu\nu} + \tfrac{1}{2}(\mathscr{D}^\mu\phi)^{\mathrm{T}}\,\mathscr{D}_\mu\phi - V(\phi) \tag{5.26}$$

leading to equations of motion given by equations (5.24) and (5.25) with:

$$j_a{}^\mu = \mathrm{i}e\phi^{\mathrm{T}}D(T^a)\,\mathscr{D}^\mu\phi.) \tag{5.27}$$

The symmetric energy momentum tensor corresponding to the Lagrangian of equation (5.20) is:

$$\theta^{\mu\nu} = -G_a{}^{\mu\nu}G_a{}^\nu{}_\lambda + \tfrac{1}{2}(\mathscr{D}^\mu\phi)^\dagger\ \mathscr{D}^\nu\phi + \tfrac{1}{2}(\mathscr{D}^\nu\phi)^\dagger\,\mathscr{D}^\mu\phi - g^{\mu\nu}\mathscr{L}. \tag{5.28}$$

### *5.3. The structure of the Higgs vacuum*

We shall be examining finite-energy, though not necessarily time-independent, solutions of the theory defined by the Lagrangian of equations (5.20) or (5.26). At any given time we shall expect them to satisfy the simpler equations:

$$V(\phi) = 0 \tag{5.29}$$

$$\mathscr{D}^\mu\phi = 0 \tag{5.30}$$

to a very good approximation everywhere in space apart from a finite number of compact regions which we shall call monopoles. (We are assuming the zero level of energy has been chosen to coincide with the minimum value of $V$, as it was in §4.) The expectation just stated was true for the explicit solutions discussed in the last section but, since no general proof exists as yet, we must treat it as an assumption.

As before, in any region of space where the fields satisfy equations (5.29) and (5.30) we shall say that they are in the *Higgs vacuum*. We now proceed to discuss the general features of the Higgs vacuum. Because $V(\phi)$ is invariant under the action of $G$, if $\phi_0$ satisfies (5.29) so does $D(g)\,\phi_0$ for any $g \in G$. So, defining the vacuum manifold as in equation (3.17):

$$\mathscr{M}_0 = \{\phi : V(\phi) = 0\} \tag{5.31}$$

we see that $G$ acts on $\mathscr{M}_0$, i.e. every $g \in G$ takes each point of $\mathscr{M}_0$ to another point

of $\mathcal{M}_0$. The interesting cases are those in which $\mathcal{M}_0$ is non-trivial in the sense of consisting of more than one point; this is equivalent to saying that $\phi$ is a Higgs field in the sense of having a non-vanishing vacuum expectation value.

Two points $\phi_1$, $\phi_2$ which can be related by an element $g \in G$:

$$\phi_1 = D(g)\,\phi_2 \tag{5.32}$$

are said to be on the same orbit. In what follows we shall, in general, make an additional assumption, namely that $\mathcal{M}_0$ consists of a single orbit of the gauge group $G$. Another way to express this is to say that $G$ acts *transitively* on $\mathcal{M}_0$; that is, given:

$$\phi_1, \phi_2 \in \mathcal{M}_0, \exists g_{12} \in G \qquad \text{such that} \qquad \phi_1 = D(g_{12})\,\phi_2. \tag{5.33}$$

This sounds like a rather technical assumption but we shall see that it is desirable physically. It is roughly the same as saying that all of the vacuum degeneracy is a consequence of the gauge symmetry group $G$ and not of any accidental discrete or continuous global (as opposed to gauge) symmetry of $V$ (Coleman 1975b). In the Georgi–Glashow model which contains the 't Hooft–Polyakov monopole the action of the gauge group SO(3) on the vacuum manifold, a sphere in three-dimensional space, is clearly transitive because any given point of a sphere can be rotated to any other. An example of a non-transitive action, the sort we are excluding by assumption, is supplied by taking $\phi$ to be an octet in an SU(3) gauge theory and

$$V(\phi) = \tfrac{1}{2}\lambda(\phi^2 - a^2)^2 \tag{5.34}$$

$\mathcal{M}_0$ is a sphere, $S^7$, in eight-dimensional space. It is not difficult to see that the action of the eight-dimensional group SU(3) cannot be transitive on this seven-dimensional manifold.

A fundamental, physically important, concept is that of the little group $H_\phi$ of a point $\phi \in \mathcal{M}_0$. For a given $\phi \in \mathcal{M}_0$ we define:

$$H_\phi = \{h \in G : D(h)\,\phi = \phi\}. \tag{5.35}$$

As $\phi$ varies within $\mathcal{M}_0$, $H_\phi$ varies within $G$ but in a very convenient way, provided that the action of $G$ on $\mathcal{M}_0$ is transitive. For if $\phi_1, \phi_2 \in \mathcal{M}_0$ are related as in (5.33):

$$H_{\phi_1} = g_{12}^{-1} H_{\phi_2} g_{12}. \tag{5.36}$$

Thus $H_\phi$ varies within $G$ by conjugation and, consequently, is isomorphic for different $\phi$. Since $H_\phi$ is the exact, and therefore the directly observable, gauge symmetry group it is highly desirable that its structure should be independent of $\phi$. In particular, its dimension is the number of massless gauge particles and the eigenvalues of its generators determine the possible values of the various physical charges, electric, etc. There would be problems of interpretation if these varied with $\phi$. This undesirable situation could obtain if the action of $G$ on $\mathcal{M}_0$ were not transitive. In the example of equation (5.34), with $G = \mathrm{SU}(3)$, $H_\phi$ could be U(2) or $\mathrm{U}(1) \times \mathrm{U}(1)$, depending on the particular $\phi \in \mathcal{M}_0$ chosen.

The assumption that the action of $G$ is transitive means that the structure of $\mathcal{M}_0$ is determined by $G$ and $H = H_{\phi_0}$ for any given $\phi_0 \in \mathcal{M}_0$. In fact:

$$\mathcal{M}_0 = G/H \tag{5.37}$$

the space of right cosets of $H$ in $G$. ($g_1, g_2 \in G$ are said to be in the same right coset of $H$ in $G$ if and only if there exists an $h \in H$ such that $g_1 = g_2 h$. This defines an eqvivalence relation on $G$ and the equivalence classes are the right cosets.) To see that

$\mathcal{M}_0$ does indeed have the structure specified in equation (5.37), associate with $g \in G$ the point:

$$\phi = D(g)\,\phi_0 \tag{5.38}$$

in $\mathcal{M}_0$. The elements $g_1, g_2 \in G$ will be associated with the same $\phi \in \mathcal{M}_0$ if and only if:

$$D(g_1^{-1}g_2)\,\phi_0 = \phi_0$$

that is, if and only if $g_1^{-1}g_2 \in H$ or, equivalently, $g_1$ and $g_2$ belong to the same right coset of $H$ in $G$. Thus equation (5.38) associates points of $\mathcal{M}_0$ with right cosets of $H$ in $G$ in a one-to-one fashion and, since transitivity implies that every point of $\mathcal{M}_0$ is associated with some coset, we may identify $\mathcal{M}_0$ with the right coset space $G/H$. The result (5.37) means that once $H$ has been determined the other details associated with the Higgs field may be ignored, at least as far as the structure of $\mathcal{M}_0$ is concerned.

Now let us turn to the implication of equation (5.30). Using equation (5.7) we deduce that:

$$D(\mathbf{G}^{\mu\nu})\,\phi = 0 \tag{5.39}$$

in the Higgs vacuum. Since, from the definition of equation (5.35), the generators of $H_\phi$ are those which annihilate $\phi$, the only non-zero components of the gauge field tensor are those corresponding to $H_\phi$. Thus only the $H$ gauge fields permeate the Higgs vacuum, the region outside the monopoles; the other components of $\mathbf{G}^{\mu\nu}$ are unexcited. (Note that in the 't Hooft–Polyakov case equation (5.39) implies equation (4.29), a result we had obtained by direct calculation.)

Later on, in §5.6, we shall demonstrate that equation (5.30) implies that if $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are two points outside the monopole regions then $\phi(\boldsymbol{r}_1)$ and $\phi(\boldsymbol{r}_2)$ are necessarily on the same orbit of $G$ in $\mathcal{M}_0$, irrespective of the assumption of transitivity. Of course, completely different finite-energy solutions could select different orbits of $\phi$ in $\mathcal{M}_0$, but such solutions could not be fitted together in the same universe unless the orbit was the same. To this extent the transitivity assumption may be irrelevant.

### *5.4. Topological quantum numbers and the Higgs field*

As before consider a field configuration, at a given time, consisting of several extended monopoles occupying compact regions $M_1, M_2, \ldots, M_N$ surrounded by a region, $\mathcal{H}$, in which the equations (5.29) and (5.30), defining the Higgs vacuum, hold to a good approximation. To this approximation:

$$\phi(\boldsymbol{r}) \in \mathcal{M}_0 \qquad \text{if} \qquad \boldsymbol{r} \in \mathcal{H}. \tag{5.40}$$

Consider a closed surface $\Sigma$, lying within $\mathcal{H}$ and enclosing $M_1$ once. Then $\phi$ defines a continuous map from $\Sigma$ to $\mathcal{M}_0$.

As time evolves the monopoles may move and change shape. As long as they do not intersect $\Sigma$, $\boldsymbol{r} \to \phi(\boldsymbol{r}, t)$ defines a map $\Sigma \to \mathcal{M}_0$ which varies continuously with time. (We are implicitly assuming continuity as a consequence of the classical field equations.) As we mentioned in §4.4, such a change is called a homotopy and $\phi(\boldsymbol{r}, t_1)$, $\phi(\boldsymbol{r}, t_2)$ are said to define homotopic maps $\Sigma \to \mathcal{M}_0$. Homotopy defines an equivalence relation and the resulting homotopy classes provide a classification of monopoles according to the way the Higgs field realises its boundary conditions. For the 't Hooft–Polyakov monopole we saw in §4.4 that $\phi: \Sigma \to \mathcal{M}_0$ was essentially just a mapping between spheres, $S^2$, and the homotopy classes were labelled by the winding or

wrapping number of the map. Further the magnetic charge was just $-4\pi/e$ times this wrapping number $N$, defined by equation (4.34).

To avoid pitfalls whilst developing these ideas further we must provide more precise definitions of the mathematical concepts. Two maps, between topological spaces $X$ and $Y$, $f_1, f_2: X \to Y$ are said to be *homotopic* if there exists a continuous map $F$ sending $(x, t) \to F(x, t) \in Y$, where $0 \leqslant t \leqslant 1$ and $x \in X$, such that:

$$F(x, 0) = f_1(x) \qquad \text{and} \qquad F(x, 1) = f_2(x). \tag{5.41}$$

Thus $F$ maps $X \times [0, 1] \to Y$ and constitutes a continuous deformation of the map $f_1$ into the map $f_2$. It is called a *homotopy*.

Frequently $X$ will be taken to be an $n$-dimensional sphere, $S^n$, and the homotopy classes of maps $S^n \to Y$ will be denoted by $\tilde{\Pi}_n(Y)$. Since, in homotopy theory, maps related by continuous deformations are equivalent, it is adequate, and it is frequently convenient, to regard $S^n$ as a unit cube in $n$-dimensional Euclidean space with all points of its surface identified (as a single point). Thus $S^2$ may be regarded as the unit square with its perimeter identified. In particular, this construction provides us with a coordinate system on the sphere. It is singular at the point to which the whole of the perimeter of the square has been identified, but every coordinate system on the sphere must have at least one singularity.

The surface $\Sigma$ surrounding the monopole and the sphere $S^2$ are equivalent for the purposes of homotopy theory (indeed they are homeomorphic). Thus $\phi: \Sigma \to \mathcal{M}_0$ defines an element of $\tilde{\Pi}_2(\mathcal{M}_0)$. Further, this element is gauge-independent. For any gauge transformation defines a continuous map $\Sigma \to G$ which hence defines an element of $\tilde{\Pi}_2(G)$. But it is a celebrated result of E Cartan that every map $S^2 \to G$ is homotopic to a constant map, and since we are assuming $G$ to be connected this constant may be taken to be the identity element of $G$. This homotopy demonstrates that $\phi$ and its gauge transform define the same element of $\tilde{\Pi}_2(\mathcal{M}_0)$.

Thus the homotopy class in $\tilde{\Pi}_2(\mathcal{M}_0)$, defined by the map $\Sigma \to \mathcal{M}_0$ provided by the Higgs field, gives a classification of monopoles which is gauge-invariant, conserved in time and independent of the particular surface $\Sigma$ used to enclose the monopole, provided that it only does it once. (The orientation of $\Sigma$ must also be specified to avoid sign ambiguities.) In other words the homotopy class associated with a monopole is a topological 'quantum number', appearing at the classical level, similar to the soliton number (3.9) of the Sine-Gordon theory and generalising the magnetic charge (4.31) of the 't Hooft–Polyakov model. In contrast to those cases we have not found an integral formula which characterises the quantum number in terms of the Higgs field and, as far as we are aware, no such formula exists at present. However, for many purposes this does not matter.

The relevance of homotopy classes in classifying generalised 't Hooft–Polyakov monopoles was pointed out by Tyupkin *et al* (1975) and Monastyrskii and Perelomov (1975). Amongst other discussions and developments of homotopy in this context are Arafune *et al* (1975), Patrascioiu (1975) and, particularly, Coleman (1975b) and Goldstone (1976). Earlier work on the relevance of homotopy to the classification of solutions to field theories is to be found in Finkelstein and Misner (1959), Enz (1963), Lubkin (1963) and Finkelstein (1966).

Now a physical question of crucial importance arises, namely, how do the topological quantum numbers combine? Since information about the structure of particles is gained from scattering experiments, a quantum number is of little use unless we know the answer. Put more precisely, if the Higgs field $\phi$ defines homotopy

classes $C_1$, $C_2$ and $C_{12}$ by its values on the surfaces $\Sigma_1$, $\Sigma_2$ and $\Sigma_{12}$ respectively in figure 4, how (if at all) can one determine $C_{12}$ from knowledge of $C_1$ and $C_2$ alone? This question has been discussed extensively by Coleman (1975b) and Goldstone (1976). In general, $C_{12}$ is not uniquely determined by $C_1$ and $C_2$, but a sufficient condition for it to be so determined is that $\mathscr{M}_0$ be simply connected. This condition is just that $\tilde{\Pi}_1(\mathscr{M}_0)$ be trivial in the sense that all paths in $\mathscr{M}_0$ are homotopic to each other (Steenrod 1951). (A closed path is just a continuous map from the circle $S^1$.) Given this condition the combination law corresponds to the group operation which plays a central role in homotopy theory, as we shall now describe.

The homotopy classes, $\tilde{\Pi}_n(Y)$, which we have defined are called absolute homotopy classes. This concept, which is the physically relevant one, is not the one which is most convenient mathematically. In homotopy theory the more important concept is that of the relative homotopy classes, $\Pi_n(Y)$. To define these relative homotopy classes one considers only maps in which the image of one particular point of the sphere is kept fixed at a certain base point $y_0 \in Y$. (This constraint applies to the homotopies as well as the maps being divided into classes.) If the space $Y$ is connected (as indeed $\mathscr{M}_0$ is, as a result of transitivity and the connectedness of $G$) $\Pi_n(Y)$ is essentially independent of the base point, $y_0$. The reason for fixing a base point is that this makes it possible to define a binary operation on $\Pi_n(Y)$ turning it into a group, the *nth homotopy group* of $Y$. For a summary of this and other aspects of homotopy theory see appendix 1. Two maps which are relatively homotopic are certainly absolutely homotopic. So $\Pi_n(Y)$ provides a finer classification than $\tilde{\Pi}_n(Y)$, in general. However, if $Y$ is simply connected, the distinction between relative and absolute homotopy disappears (Steenrod 1951, p86). Consequently, if we assume that the first homotopy group of $\mathscr{M}_0$ is trivial:

$$\Pi_1(\mathscr{M}_0)=0 \tag{5.42}$$

the values of the Higgs field $\phi$ on $\Sigma$ define an element of $\Pi_2(\mathscr{M}_0)$. It is possible to construct models in which equation (5.42) fails (Coleman 1975b, Goldstone 1976), but these are rather abstruse, and we relegate a discussion of these to appendix 2 and henceforth assume equation (5.42).

The group operation on $\Pi_2(\mathscr{M}_0)$ corresponds precisely to the combination law for monopoles, which is well-defined given equation (5.42) (Coleman 1975b). For $n \geqslant 2$, $\Pi_n(Y)$ is always Abelian and the topological quantum numbers may always be thought of as additive:

$$C_{12}=C_1+C_2. \tag{5.43}$$

Typically $\Pi_2(\mathscr{M}_0)$ might be the integers, $\mathbb{Z}$, the integers modulo some integer $N$, $\mathbb{Z}_N$, or some product of such groups. The structure of $\mathscr{M}_0$ as a coset space provides us with much information about $\Pi_2(\mathscr{M}_0)$ which we shall discuss in the next section. Then, in subsequent parts of this section, we shall show how the topological quantum number can be calculated in suitable circumstances.

### *5.5. Topological quantum numbers and the structure of H*

We shall now see how the structure of $\mathscr{M}_0$, which we analysed in §5.3, enables us to relate the group of topological quantum numbers, seen in §5.4 to be isomorphic to $\Pi_2(\mathscr{M}_0)$, to properties of $H$.

In the last subsection we outlined the definition of the homotopy groups, $\Pi_n(Y)$, $n \geqslant 1$. (For a more precise summary see appendix 1.) For various reasons it is both convenient and natural to define $\Pi_0(Y)$ to be the set of path components of $Y$. (Two elements of $Y$ are in the same path component if they can be joined by a continuous path in $Y$.) When considering a group $\Gamma$, $\Pi_0(\Gamma)$ can be made into a group by defining the product of two components to be the component containing the product of any two elements, one taken from each of the two components. Then:

$$\Pi_0(\Gamma) \simeq \Gamma/\Gamma_0 \tag{5.44}$$

where $\Gamma_0$ is the path component of the identity, the largest connected subgroup of $\Gamma$ containing the identity.

We now seek to use the knowledge that $\mathscr{M}_0$ can be identified with the right coset space $G/H$. Let us assume for the moment that $G$ is simply connected as well as connected; these statements may be summarised as:

$$\Pi_0(G) = 0 \qquad \Pi_1(G) = 0. \tag{5.45}$$

A theorem in homotopy theory then tells us that:

$$\Pi_1(G/H) \simeq \Pi_0(H) \tag{5.46}$$

and

$$\Pi_2(G/H) \simeq \Pi_1(H). \tag{5.47}$$

The isomorphism (5.46) tells us that the assumption that $\mathscr{M}_0$ is simply connected is equivalent to assuming that $H$ is connected. We made this assumption in the last subsection to ensure that the combination law for topological quantum numbers is well-defined. The isomorphism (5.47), first employed in this context by Tyupkin *et al* (1975), Monastyrskii and Perelomov (1975) and Coleman (1975b), is very important physically because it expresses the structure of topological quantum numbers in terms of $H$ only.

The second isomorphism will be discussed and explicitly constructed, in the context of the Higgs mechanism, in the next subsection and both isomorphisms are outlined from a more abstract viewpoint in appendix 1. In the remainder of this subsection we shall exemplify (5.47) and discuss its significance.

The assumption that $G$ is simply connected can be removed if we replace $\Pi_1(H)$ in (5.47) by the subgroup of closed paths in $\Pi_1(H)$ which are trivial (i.e. may be contracted to a point) in $G$. If we denote this by $\Pi_1(H)_G$ we may replace (5.47) by:

$$\Pi_2(G/H) \simeq \Pi_1(H)_G \tag{5.48}$$

which is true independently of equations (5.45). The assumption that $G$ is simply connected may not be too severe since given any Lie group $G$ we may replace it by an essentially unique simply connected group, $\tilde{G}$ (the universal covering group of $G$), just as we may always replace SO(3) by SU(2). If $G$ is compact and semi-simple (i.e. has no local U(1) factors), $\tilde{G}$ will also be compact.

It is clear that for solutions to gauge field theories of the class so far considered (namely without Dirac strings) (5.47) and (5.48) provide the same information. That is, for a theory with gauge group $G$ and with the little group of the Higgs field in $\mathscr{M}_0$ being $H$, we could instead use $\tilde{G}$ and obtain a little group $\hat{H}$ of the Higgs field. Then $\tilde{G}/\hat{H} = G/H$ and $\Pi_1(H)_G \simeq \Pi_1(\hat{H})$. On the other hand, if Dirac strings are allowed all the elements of $\Pi_1(H)$ may be realised independently of $G$, and then there is some difference between the approaches.

Let us illustrate these observations in the context of the simplest example: the Georgi–Glashow model which contains the 't Hooft–Polyakov monopole. Here the gauge group $G=\mathrm{SO}(3)$, though we could replace it by SU(2) to obtain a simply connected group. The homotopically distinct closed paths in $H$ are:

$$h(t)=\exp(\mathrm{i}\boldsymbol{\phi}.\boldsymbol{T}4\pi Nt/a) \qquad 0\leqslant t\leqslant 1 \tag{5.49}$$

where $(T_a)_{ij}=-\mathrm{i}\epsilon_{aij}$ for $G=\mathrm{SO}(3)$ and $\boldsymbol{T}=\frac{1}{2}\boldsymbol{\sigma}$ for $G=\mathrm{SU}(2)$. For $G=\mathrm{SU}(2)$, $H=\mathrm{U}(1)$ and $N$ can be any integer. For $G=\mathrm{SO}(3)$, $H=\mathrm{SO}(2)$, which is isomorphic to U(1) and $N$ can be any half-integer (as the eigenvalues of $T_a$ are now integral rather than half-integral). In the latter case only those paths for which $N$ is an integer are trivial in $G=\mathrm{SO}(3)$. The path with $N=\frac{1}{2}$, for example, is just a rotation through $2\pi$ and it is the familiar fact that this is a non-trivial closed path that allows spinor wavefunctions to change sign under such a rotation. So we see that $\Pi_1(\mathrm{SO}(2))_{\mathrm{SO}(3)}$ differs from $\Pi_1(\mathrm{SO}(2))$; it contains just half the paths in the latter, although both are isomorphic to the additive group of integers, $\mathbb{Z}$.

We can see that both approaches yield the same information if we use the fact that, for both SU(2) and SO(3), the magnetic charge is related to the $N$ occurring in equation (5.49) by:

$$g=4\pi N/e. \tag{5.50}$$

In each case $N$ must be an integer. Equation (5.50) will be established in the next subsection. For an SU(2) theory the unit of electric charge is $q_0=\frac{1}{2}e\hbar$ and equation (5.50) is just the Dirac quantisation condition. For a genuine SO(3) the unit of electric charge would be $q_0=e\hbar$ and to obtain all the possibilities allowed by Dirac's condition we would have to consider solutions with strings. This would permit half-integral values of $N$ in equation (5.50).

A more drastic example of the restriction implied by the isomorphism (5.48) occurs in the Salam–Weinberg model (Salam 1968, Weinberg 1967) where $G\simeq \mathrm{SU}(2)\times\mathrm{U}(1)$ and $H\simeq\mathrm{U}(1)$ with electric charge, $Q$, being a sum of an SU(2) and the U(1) generator, so that $H$ does not lie entirely within the U(1) factor. Any closed path in $H$ may be deformed in $G$ to lie completely inside the U(1) factor and unless it is trivial there it cannot be deformed to a point in $G$. Thus, although $\Pi_1(H)\simeq\mathbb{Z}$, $\Pi_1(H)_G=0$ and any topologically non-trivial monopole must have a string which cannot be gauged away in $G$ ('t Hooft 1974).

In this subsection we have seen that the structure of the topological quantum numbers can be characterised in terms of the fundamental group of $H$, $\Pi_1(H)$. This is the maximum possible structure, and if the original broken symmetry group $G$ is not simply connected it will require solutions with string singularities to obtain all the possibilities corresponding to $\Pi_1(H)$. On the other hand, if $\Pi_1(H)=0$ no solutions with non-trivial topological quantum numbers are possible. Thus for $H=\mathrm{SU}(2)$ there are no topological quantum numbers, whilst for $H=\mathrm{SO}(3)$, $\Pi_1(\mathrm{SO}(3))=\mathbb{Z}_2$, the cyclic group with two elements. In general, for any semi-simple compact group $H$ we may write $H\simeq\tilde{H}/k(H)$ where $k(H)$ is a subgroup of $Z(\tilde{H})$, the centre of $\tilde{H}$ (the finite Abelian group consisting of those elements of $\tilde{H}$ which commute with all others). It is not difficult to show that:

$$\Pi_1(H)\simeq k(H) \tag{5.51}$$

which is a finite Abelian group and this makes the structure of such topological conservation laws rather unusual.

*5.6. Topological quantum numbers and the H gauge fields*

So far in this section we have progressed from having the topological quantum number defined in terms of the way the Higgs field realises its boundary conditions to seeing how the structure of such quantum numbers is determined by the global properties of the exact symmetry group, $H$. Now we will go further and show how to formulate the topological quantum number in terms of the $H$ gauge fields, dispensing with the Higgs field. Physically this is eminently reasonable since, as we argued in §5.3, only the $H$ gauge fields survive outside the monopole and thus they carry the long-range characteristics of the monopole.

This construction is interesting in itself since it involves 'path-dependent' or 'non-integrable' phase factors which were originally introduced by Dirac (1931) in the electromagnetic context as we discussed in §2.1. They are, roughly speaking, exponentiated magnetic fluxes and, thus, enable us to relate topological quantum numbers to generalised magnetic charges in various situations. An important tool in this is a non-Abelian generalisation of Stokes' theorem which we shall establish in this subsection and apply in the next two. The construction of the topological quantum number in terms of $H$ fields was foreshadowed in the work of Lubkin (1963).

Consider again the situation which we studied in §5.4 with a closed surface $\Sigma$ surrounding a monopole. On and around $\Sigma$, in the Higgs vacuum, equation (5.30) holds:

$$\mathscr{D}^{\mu}\phi = 0 \tag{5.52}$$

and this is the equation which will enable us to express the topological quantum number in terms of the $H$ gauge fields. As we explained in §5.4, $\Sigma$ is topologically equivalent to a sphere. We may parametrise $\Sigma$ in the way we said it was often convenient to parametrise $S^2$, i.e. by regarding it as the unit square with its perimeter identified to a single point. Thus:

$$\Sigma = \{\boldsymbol{r}(s, t): 0 \leqslant s \leqslant 1,\ 0 \leqslant t \leqslant 1\} \tag{5.53}$$

where $(s, t) \rightarrow \boldsymbol{r}(s, t)$ is one to one, save that the whole of the perimeter is mapped to a single point $\boldsymbol{r}_0 \in \Sigma$. For each fixed $s$, $\boldsymbol{r}(s, t)$ describes a loop on $\Sigma$, starting and finishing at $\boldsymbol{r}_0$ as $t$ varies from 0 to 1. Then, as $s$ varies from 0 to 1, these loops trace out the whole of $\Sigma$, starting and finishing with trivial loops consisting of a single point at $\boldsymbol{r}_0$. Apart from $\boldsymbol{r}_0$, each point of $\Sigma$ lies on precisely one of these loops.

The topological quantum number of the monopole within $\Sigma$ is determined by the map $\phi: \Sigma \rightarrow \mathscr{M}_0$. If we write $\phi_0 = \phi(\boldsymbol{r}_0)$ and

$$\mathscr{D}_t = \frac{\partial r^i}{\partial t}\, \mathscr{D}_i \tag{5.54}$$

this map can be defined by the partial differential equation:

$$\mathscr{D}_t \phi = 0 \tag{5.55}$$

subject to the boundary condition:

$$\phi(s, 0) = \phi_0 \qquad 0 \leqslant s \leqslant 1 \tag{5.56}$$

where $\phi(s, t) \equiv \phi(\boldsymbol{r}(s, t))$. Equation (5.55) can be written in the same form as the Schrödinger equation:

$$\frac{\partial \phi}{\partial t} = ieD(\mathbf{W}^i)\, \phi\, \frac{\partial r^i}{\partial t}$$

and our approach to solving it is to introduce the analogue of the time evolution operator. It is an element of the group, defined by the equation:

$$\mathscr{D}_t g(s, t) = 0 \qquad \text{subject to} \qquad g(s, 0) = 1. \tag{5.57}$$

This is solved by Dyson's formula:

$$g(s, t) = \mathscr{T}\left[\exp\left(ie \int_0^t \mathbf{W}^i \frac{\partial r^i}{\partial t}\, dt\right)\right] \tag{5.58}$$

where the $\mathscr{T}$ operation indicates that the exponential is symbolic; to evaluate the right-hand side of equation (5.58) the exponential must be expanded and the factors of $\mathbf{W}^i$ ordered with larger values of $t$ occurring to the left of smaller ones before the integrations are performed.

Equation (5.58) provides us with the unique solution to equation (5.57) but it is important to remember that it is well-defined over the unit square rather than $\Sigma$. For $g$ to be defined as a function on $\Sigma$, we would need it to have the same value on the whole of the perimeter of the unit square. In fact, $g = 1$ on three sides of it, because of the boundary conditions in equation (5.57) and because $\partial r^i / \partial t = 0$ on $s = 0$ and 1. But on $t = 1$, $g(s, 1)$ will not be equal to 1 in general and we define:

$$h(s) = g(s, 1) = \mathscr{T}\left[\exp\left(ie \int_0^1 \mathbf{W}^i \frac{\partial r^i}{\partial t}\, dt\right)\right]. \tag{5.59}$$

The quantity $h(s)$ is the path-dependent phase factor associated with the closed loop $\boldsymbol{r}(s, t)$, $0 \leqslant t \leqslant 1$, $s$ fixed. Further, as $s$ varies from 0 to 1, $h(s)$ describes a closed loop in the group, since:

$$h(0) = h(1) = 1. \tag{5.60}$$

The object in constructing $g$ was to find $\phi: \Sigma \to \mathscr{M}_0$. Indeed $\phi$ is given by:

$$\phi(s, t) = D(g(s, t))\, \phi_0 \tag{5.61}$$

as it follows from equation (5.57) that $\phi$ then satisfies equations (5.55) and (5.56). Further it provides an expression for $\phi: \Sigma \to \mathscr{M}_0$ entirely in terms of the gauge fields, apart from $\phi_0$, which may be varied without changing the topological quantum number, assuming only that $\mathscr{M}_0$ is connected. (It also demonstrates the truth of the comment made in §5.3 that if $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are any two points in the Higgs vacuum $\mathscr{H}$, assumed connected, $\phi(\boldsymbol{r}_1)$ and $\phi(\boldsymbol{r}_2)$ are necessarily on the same orbit of $G$ in $\mathscr{M}_0$, irrespective of any assumption of transitivity.)

Equation (5.61) may be interpreted another way by inverting $D(g)$:

$$D(g(s, t)^{-1})\, \phi(s, t) = \phi_0. \tag{5.62}$$

This says that $D(g(s, t)^{-1})$ is the gauge transformation which rotates $\phi(\boldsymbol{r})$ to the fixed direction $\phi_0$ for $\boldsymbol{r} \in \Sigma$. Because $h(s) \neq 1$ this gauge transformation is singular at $\boldsymbol{r} = \boldsymbol{r}_0$ in general. If such a singularity were not forced on us in general all topological quantum numbers would be trivial since they would be related by gauge transformations. This singularity may be regarded as occurring at a point at which a Dirac string crosses $\Sigma$; $g(s, t)$ is just a generalisation of the gauge transformation explicitly constructed in §4.5. Although $h(s) \neq 1$ in general, it is restricted, because $\phi(s, 1) = \phi_0$ so that from equation (5.61) we see that $h(s) \in H$. So using equation (5.60) we see

that $h(s)$ defines a closed loop in $H$. This closed loop in turn defines an element of $\Pi_1(H)$. It is this element which corresponds to the element of $\Pi_2(G/H)$, defined by $\phi: \Sigma \to \mathscr{M}_0$ under the isomorphism (5.47) or, more generally, (5.48).

We saw that $g$ was uniquely determined by equation (5.57) given $\mathbf{W}^i$. Now we wish to argue a somewhat different point, namely that, although the construction of equation (5.59) only involves $\mathbf{W}^i$, the homotopy class of $h(s)$ only depends on $\phi: \Sigma \to \mathscr{M}_0$. What makes this possible is the intimate relation between $\phi$ and $\mathbf{W}^i$ brought about by equation (5.55). Suppose that both $g = g_1(s, t)$ and $g = g_2(s, t)$ satisfy equation (5.61) and the conditions:

$$g(s, 0) = g(0, t) = g(1, t) = 1. \tag{5.63}$$

Then:

$$D(g_1(s, t))\, \phi_0 = D(g_2(s, t))\, \phi_0$$

so that $g_1(s, t)^{-1} g_2(s, t) \in H$. Now, writing $h_i(s) = g_i(s, 1)$, consider:

$$\eta(s, t) = h_1(s)\, g_1(s, t)^{-1} g_2(s, t) \in H.$$

It is continuous as a function of $s$ and $t$ and $\eta(s, 0) = h_1(s)$ whilst $\eta(s, 1) = h_2(s)$. This shows that $h_1$ and $h_2$ are homotopic, defining the same element of $\Pi_1(H)$, and establishing that $\phi$ determines the homotopy class of $h$. It is possible to argue further along these lines to show that this homotopy class depends only on the homotopy class of $\phi$, the element of $\Pi_2(G/H)$ that it defines and that the map $\Pi_2(G/H) \to \Pi_1(H)$ constructed in this way is a group homomorphism.

The loops in $H$ obtained in this way are clearly trivial in $G$ since $g(s, t)$ itself defines a homotopy in $G$ between the trivial path $g(s, 0) = 1$ and $h(s) = g(s, 1)$. Finally, to establish (5.47) and (5.48) it is necessary to show that the map $\Pi_2(G/H) \to \Pi_1(H)$ is one-to-one. This depends on the result of Cartan that $\Pi_2(G) = 0$ and, together with the other assertions we have made, is discussed further in appendix 1. In this way we arrive at a proof of the theorem quoted in §5.5 but with a very useful explicit expression for the element of $\Pi_1(H)$ in terms of the gauge potentials.

We will now develop equation (5.59) further to obtain an expression for $h(s)$ in terms of the field tensor $\mathbf{G}^{ij}$. In equation (5.64) we view both $\mathscr{D}_t$ and $g(s, t)$ as acting on some further vector or matrix and replace equation (5.57) by:

$$\mathscr{D}_t g = g\, \partial_t \qquad \text{where} \qquad \partial_t = -\frac{\partial r^i}{\partial t}\, \partial^i. \tag{5.64}$$

Consequently:

$$g^{-1}\, \mathscr{D}_t = \partial_t g^{-1}$$

and, defining $\mathscr{D}_s$ by an equation similar to equation (5.54):

$$\begin{aligned} \partial_t(g^{-1}\, \mathscr{D}_s g) &= g^{-1}\, \mathscr{D}_t\, \mathscr{D}_s g \\ &= g^{-1}[\mathscr{D}_t, \mathscr{D}_s]\, g \qquad \text{by equation (5.57)} \\ &= \mathrm{i} e g^{-1} \mathbf{G}_{ij} g\, \frac{\partial r^i}{\partial t}\, \frac{\partial r^j}{\partial s}. \end{aligned}$$

Now integrating with respect to $t$ from 0 to 1 and using the facts that:

$$g^{-1}\, \mathscr{D}_s g = 0,\ t = 0 \qquad g^{-1}\, \mathscr{D}_s g = h^{-1} \frac{\mathrm{d}h}{\mathrm{d}s},\ t = 1$$

we obtain

$$h^{-1}\frac{\mathrm{d}h}{\mathrm{d}s}=\mathrm{i}e\int_0^1 g^{-1}\mathbf{G}_{ij}g\,\frac{\partial r^i}{\partial t}\,\frac{\partial r^j}{\partial s}\,\mathrm{d}t. \tag{5.65}$$

This expression was given by Goldstone (1976), but there is an earlier version by Christ (1975), and it is possible that there are others. It provides a sort of non-Abelian Stokes' theorem.

Equation (5.65) gives an expression for $h(s)$ in terms of $H$ gauge fields. (This is clear from the left-hand side of that equation and may be verified for the right-hand side using equations (5.39) and (5.61).) For $H=\mathrm{U}(1)$ it reduces to:

$$h(s)=\exp\left(\mathrm{i}e\int_\Sigma \boldsymbol{B}.\mathrm{d}\boldsymbol{S}\right)$$

and $h(1)=1$ becomes the Dirac quantisation condition, $eg=2n\pi$, $n\in\mathbb{Z}$.

### *5.7. Quantisation of charge in the presence of a colour gauge group*

An interesting possibility to consider is that $H$ consists, at least locally, of two factors: U(1), which may be thought of as the electromagnetic gauge group, and $K$, say, which may be thought of as a colour gauge group and is such that the electric charge $Q$, which generates the U(1), is a colour singlet. With $K=\mathrm{SU}(3)$ this could well be the exact gauge symmetry of nature. We now consider how the colour gauge symmetry influences the electromagnetic charge quantisation conditions and how this might relate to the fractional charges of the quarks.

The local information given so far fixes the Lie algebra of $H$ but not its global structure. We shall see that this relates both to the possible monopoles which may exist and the specific electric charge assignments of the irreducible representations of $K$.

A natural way of realising the situation we have just described is to have the Higgs field in the adjoint representation of $G$, the full gauge symmetry group, before symmetry breaking. The condition that $M\in L(G)$, the Lie algebra of $G$, be a generator of $H$, the little group of $\phi$, is that $M\phi$ vanishes, which may be rephrased:

$$[M,\underline{\Phi}]=0 \tag{5.66}$$

where $\underline{\Phi}=\phi_a T^a$. Evidently $\underline{\Phi}\in L(H)$ and, further, it commutes with all other generators of $H$. Thus $\underline{\Phi}$ generates an invariant U(1) subgroup of $H$, which we identify as the electromagnetic gauge group. Because every $M\in L(H)$ satisfies equation (5.66) we may write $L(H)=u(1)\oplus L(K)$ where $u(1)=\{Q\}=L(\mathrm{U}(1))$ and $K$ is the colour group, defined as being generated by those generators of $H$ orthogonal to $\underline{\Phi}$:

$$L(K)=\{M\in L(H):\mathrm{Tr}\,(\underline{\Phi}M)=0\}. \tag{5.67}$$

$\mathbf{W}^\mu$ has the expansion:

$$\mathbf{W}^\mu=A^\mu\underline{\Phi}/a+\mathbf{X}^\mu \tag{5.68}$$

where $a$ is the length of $\phi$ in $\mathcal{M}_0$ and $\mathrm{Tr}\,(\underline{\Phi}\mathbf{X}^\mu)=0$. Comparing the U(1) covariant derivative $\partial^\mu+\mathrm{i}QA^\mu/\hbar$ with the $G$ covariant derivative $\partial^\mu+\mathrm{i}e\mathbf{W}^\mu$, just as in equation (4.12), we see that the electric charge operator:

$$Q=\frac{e\hbar}{a}\,\underline{\Phi}. \tag{5.69}$$

't Hooft (1976) has constructed a model with $H$ having the structure just described

but in which the Higgs field is not in the adjoint representation. The electromagnetic direction is picked out by a vector $\chi$, in the adjoint representation, constructed out of the Higgs field in such a way that it is covariantly constant when the Higgs field is. Our analysis applies equally well to such a situation, replacing $\underline{\Phi}$ with $\underline{\chi}$ in equations (5.68) and (5.69).

Now consider the situation described in the previous subsection, in which $\phi$ is covariantly constant on a surface $\Sigma$ surrounding a possible region of magnetic charge. Since $h^{-1}(\mathrm{d}h/\mathrm{d}s)$ is a generator of $H$, we may write:

$$h^{-1}\frac{\mathrm{d}h}{\mathrm{d}s}=\frac{ie}{a}\alpha(s)\,\underline{\Phi}_0+i\beta_a(s)\,K^a \tag{5.70}$$

where $\{K^a\}$ is a basis for $L(K)$, for suitable coefficients $\alpha(s)$ and $\beta_a(s)$. Using the non-Abelian Stokes' theorem (5.65) we shall show that the coefficient $\alpha(s)$ has a simple physical interpretation: it is the derivative of the U(1) magnetic flux, $\Phi(s)$ through a surface spanning the loop $\Gamma_s$, defined by $s=$ constant. By equations (5.65) and (5.70), using $\mathrm{Tr}\,(T^aT^b)=\kappa\delta_{ab}$:

$$\begin{aligned}\alpha(s)&=-\frac{\mathrm{i}}{ae\kappa}\,\mathrm{Tr}\left(\underline{\Phi}_0 h^{-1}\frac{\mathrm{d}h}{\mathrm{d}s}\right)\\&=\frac{1}{a\kappa}\,\mathrm{Tr}\left(\underline{\Phi}_0\int_0^1 g(\boldsymbol{r})^{-1}\,\mathbf{G}_{ij}(\boldsymbol{r})\,g(\boldsymbol{r})\,\frac{\partial r^i}{\partial t}\,\frac{\partial r^j}{\partial s}\,\mathrm{d}t\right).\end{aligned}$$

Since $\phi$ is covariantly constant, rephrasing equation (5.61):

$$g(\boldsymbol{r})\,\underline{\Phi}_0 g(\boldsymbol{r})^{-1}=\underline{\Phi}(\boldsymbol{r})$$

so that

$$\alpha(s)=\frac{1}{a\kappa}\int_0^1 \mathrm{Tr}\,\{\underline{\Phi}(\boldsymbol{r})\,\mathbf{G}_{ij}(\boldsymbol{r})\}\,\frac{\partial r^i}{\partial t}\,\frac{\partial r^j}{\partial s}\,\mathrm{d}t.$$

We identify the electromagnetic tensor $F^{\mu\nu}$ with the component of $\mathbf{G}^{\mu\nu}$ in the direction of $\underline{\Phi}$:

$$F^{\mu\nu}=\frac{1}{a\kappa}\,\mathrm{Tr}\,(\underline{\Phi}\mathbf{G}^{\mu\nu})$$

(which satisfies the homogeneous Maxwell equations in the Higgs vacuum). Then:

$$\alpha(s)=\int_0^1 F_{ij}\,\frac{\partial r^i}{\partial t}\,\frac{\partial r^j}{\partial s}\,\mathrm{d}t=\frac{\mathrm{d}\Phi}{\mathrm{d}s}$$

as claimed. Integrating equation (5.70):

$$\begin{aligned}h(s)&=k(s)\,\exp\left(\frac{ie}{a}\,\Phi(s)\,\underline{\Phi}_0\right)\\&=k(s)\,\exp\,(\mathrm{i}Q\Phi(s)/\hbar)\end{aligned}$$

using equation (5.69), where $k(s)\in K$, the colour group, is obtained by integration. Now $h(1)=1$ and $\Phi(1)=g$, the total U(1) magnetic charge enclosed within $\Sigma$. Hence we obtain the quantisation condition (Corrigan and Olive 1976):

$$\exp\,(\mathrm{i}gQ/\hbar)=k\in K. \tag{5.71}$$

This generalises the Dirac condition of previous sections, which corresponds to the

case where $K$ is trivial and so $k=1$. Note that this condition has been derived without reference to the equations of motion.

We shall now proceed to discuss the consequences of equation (5.71) in some detail. A feature which is, at first sight, peculiar is that it relates the left-hand side, which is an element of the electromagnetic U(1) group, to an element of the colour group $K$ on the right-hand side. If $H$ were precisely $\mathrm{U}(1)\times K$, the direct product of the electromagnetic and colour groups, equation (5.71) could only be satisfied by both sides equalling the identity. However, all that we have assumed is that $Q$ is a colour singlet which implies that $H$ is *locally* a direct product, that $L(H)=u(1)\oplus L(K)$. In general, this need not integrate to a global property, as we shall now see.

Since $Q$ is a colour singlet, $k$, as defined by equation (5.71), must commute with the whole of the colour group, $K$, i.e. it lies in the centre, $Z(K)$, of $K$. Now if $K$ is semi-simple, as we shall now assume, its centre is finite. For definiteness let us specialise to the case of $K=\mathrm{SU}(N)$. Any element of the centre of $\mathrm{SU}(N)$ must be a multiple of the identity matrix, $1_N$:

$$k=\lambda 1_N.$$

The restriction that $k$ have unit determinant yields the possibilities:

$$\lambda=\exp\left(\frac{2\pi i n}{N}\right) \qquad n=1, 2, \ldots, N.$$

Thus $Z(\mathrm{SU}(N))$ is isomorphic to $\mathbb{Z}_N$, the group consisting of the complex $N$th roots of unity. If $K=\mathrm{SU}(N)$, it is possible to arrange that:

$$\mathrm{U}(1)\cap K=\mathbb{Z}_N$$

or any subgroup of $\mathbb{Z}_N$. Thus there may be a finite number of different ways of satisfying equation (5.71) apart from the trivial one of having both sides unity.

The non-trivial ways of satisfying equation (5.71) are physically interesting because they relate the electric charges of particles to their colour transformation properties. For example, if $|s\rangle$ denotes a colour singlet state:

$$k|s\rangle=|s\rangle$$

and so, by equation (5.71):

$$\exp(igq_s/\hbar)=1 \tag{5.71′}$$

where $q_s$ is the electric charge of any colour singlet state and $g$ is the U(1) magnetic charge of any monopole. Equation (5.71′) is just the usual Dirac quantisation condition, but in this context it is only for colour singlet particles. Let $q_0$ be the unit of electric charge for colour singlet states, so that the possible values of $q$ are:

$$q_s=nq_0 \qquad n\in\mathbb{Z}. \tag{5.72}$$

Then the possible values of $g$ are:

$$g=mg_0 \qquad m\in\mathbb{Z} \tag{5.73}$$

where $q_0g_0=2\pi\hbar$.

Let us suppose that there is a monopole with charge $g_0$ that does *not* satisfy equation (5.71) in the trivial way of having $k=1$. Then the monopole must emanate a 'colour magnetic' flux. The question of confinement does not arise since we are treating the monopoles classically; they are extended solutions and such solutions with $k\neq 1$ are know to exist, at least if $K=\mathrm{SU}(2)$ (Corrigan *et al* 1976). It is the electri-

cally charged particle states that are being treated quantum mechanically, moving in the given classical field of the monopole.

The particle states form representations ($c$) of the colour group, $K=\mathrm{SU}(N)$, and to these representations may be attached an integer $t(c)$, defined modulo $N$:

$$k|c\rangle=\exp(\mathrm{i}g_0Q/\hbar)|c\rangle=\exp(2\pi\mathrm{i}t(c)/N)|c\rangle \tag{5.74}$$

(using Schur's lemma and $k^N=1$). $t(c)$ is an additive quantum number in the sense that:

$$t(c_3)=t(c_1)+t(c_2) \qquad \mathrm{mod}\ (N) \qquad \text{if} \qquad |c_3\rangle=|c_1\rangle\otimes|c_2\rangle.$$

For $N=3$, $t(c)$ is familiar as the triality of the representation ($c$). Now it follows from equation (5.74) that:

$$q_c=q_0(m+t(c)/N) \qquad \text{for some } m\in\mathbb{Z} \tag{5.75}$$

where $q_c$ is the electric charge of any coloured state with the given value of $t(c)$, the generalised colour triality. So within this framework fractional charge arises quite naturally, with the fractional part related to the generalised colour triality. This is indeed what is observed in nature with $K=\mathrm{SU}(3)$. Note in particular that what is involved in equation (5.75) has to be the colour triality, not the flavour triality, as only the colour group is an exact gauge symmetry group. This is in agreement with the fractionally charged charmed quark being an SU(3) colour triplet but an SU(3) flavour singlet.

Of course, the situation just described is crucially dependent on having $H=\mathrm{U}(3)$ rather than $\mathrm{U}(1)\times\mathrm{SU}(3)$, and the former structure would be implied by the existence of a coloured monopole with magnetic charge $g_0$. Once again we see that the global topological properties of the gauge group relate to the existence and properties of monopoles. Similar comments could be made for any simple colour group $K$.

The possibility of a connection between magnetic monopoles and the fractional charges and confinement of the quarks has been raised a number of times, though there have been differences, some subtle, between the proposals. An early version of the ideas, outside the context of gauge theories, was given by Schiff (1966, 1967). For more recent discussions see 't Hooft (1976) and Corrigan and Olive (1976).

Finally let us mention that there is a finer structure for $g$ analogous to that in equation (5.75) for $q$. If $g$ is an integral multiple of $Ng_0$ the monopole is colourless by equation (5.71). This illustrates what may be a very fundamental feature of monopole theory, a symmetry between electric and magnetic properties. We shall discuss this further in the next section.

### *5.8. Non-Abelian magnetic charge and its quantisation*

Most of the detailed results we have obtained so far refer specifically to Abelian and, in particular, U(1), magnetic charge, but in particle physics non-Abelian gauge groups $H$, such as the SU(3) of colour, play an important role. We have seen that when the exact symmetry group, $H$, is embedded in a larger spontaneously broken gauge group, $G$, the possible magnetic monopoles possess topological quantum numbers with the structure of $\Pi_1(H)$. If this were the whole story, it would mean that for $H=\mathrm{SU}(3)$, which is simply connected, there would be no topological quantum numbers. However, there are hints that there may be a finer structure to monopole solutions which will only be revealed when the theory is properly quantised. At the

present this has the status of a tentative conjecture which we shall attempt to explain in this subsection after deriving quantisation conditions on non-Abelian magnetic charge using the generalised Stokes' theorem of §5.6.

Consider a static monopole solution in the gauge in which:

$$\boldsymbol{r}.\boldsymbol{W}_a=0. \tag{5.76}$$

(For an argument that it is legitimate to impose this gauge condition see Coleman (1975b).) Then it follows from the fact that asymptotically its covariant derivative vanishes, equation (5.52), that $\phi$ is a function of direction, $\hat{\boldsymbol{r}}$, only, asymptotically and that, thus:

$$|\nabla\phi|=O(r^{-1}).$$

It then seems reasonable to suppose that the gauge potential has the form:

$$\mathbf{W}^i(\boldsymbol{r})=\frac{1}{r}\,\mathbf{Y}^i(\hat{\boldsymbol{r}})+O(r^{-1-\delta}) \qquad \text{for some } \delta>0. \tag{5.77}$$

Since, in the chosen gauge:

$$r^i\mathbf{G}_{ij}=\frac{\partial}{\partial r}(r\mathbf{W}_j)$$

it follows that, to leading order in the radius, the magnetic field is radial and proportional to the inverse square of the radius:

$$\mathbf{G}_{ij}=\frac{1}{4\pi r^2}\,\epsilon_{ijk}\hat{r}^k\mathbf{G}(\hat{\boldsymbol{r}})+O(r^{-2-\delta}). \tag{5.78}$$

All known monopole solutions, including dyon solutions, enjoy the property (5.78) with, in addition $\mathbf{G}(\boldsymbol{r})$ being covariantly constant:

$$\mathscr{D}_i\mathbf{G}=0. \tag{5.79}$$

The radial component of equation (5.79) is an immediate consequence of equations (5.76) and (5.78). On the other hand, the transverse component of equation (5.79) is a new statement which would follow from the spatial components of the equations of motion (5.24) if, for example, $\mathbf{G}_{i0}$ were known to be radial and the terms involving the Higgs field vanished (as seems to happen in practice in the known solutions).

Taken together equations (5.78) and (5.79) constitute a *generalised inverse square law*, for the (generalised) magnetic field, which enables us to evaluate the path-dependent phase factor of equation (5.59) exactly for large closed loops (Goddard *et al* 1977).

It follows from equation (5.57) and (5.79) that:

$$\mathbf{G}(\hat{\boldsymbol{r}}(s,t))=g(s,t)\,\mathbf{G}_{(0)}g(s,t)^{-1} \qquad \text{with} \qquad \mathbf{G}_{(0)}=\mathbf{G}(\hat{\boldsymbol{r}}_0)$$

and so using the non-Abelian Stokes' theorem of equation (5.65):

$$\begin{aligned} h^{-1}\frac{\mathrm{d}h}{\mathrm{d}s}&=\frac{ie}{4\pi}\,\mathbf{G}_{(0)}\int_0^1\frac{\partial r^i}{\partial t}\frac{\partial r^j}{\partial s}\,\epsilon_{ijk}\frac{\hat{r}^k}{r^2}\,\mathrm{d}t \\ &=\frac{ie}{4\pi}\,\mathbf{G}_{(0)}\frac{\mathrm{d}\Omega}{\mathrm{d}s} \end{aligned} \tag{5.80}$$

where $\Omega(s)$ is the solid angle subtended by the loop $\Gamma_s$ (defined by $s$ constant) at the

origin. Equation (5.80) can be integrated immediately to give:

$$h(s)=\exp\left(\frac{ie}{4\pi}\mathbf{G}_{(0)}\Omega(s)\right). \tag{5.81}$$

Hence, using $h(1)=1$ and $\Omega(1)=4\pi$ we derive the quantisation condition:

$$\exp(ie\mathbf{G}_{(0)})=1 \tag{5.82}$$

(Goddard *et al* (1977). Earlier derivations in rather different formalisms are given by Klimo and Dowker (1973) and Englert and Windey (1976).) The path of equation (5.81) is clearly homotopic to:

$$h(s)=\exp(ie\mathbf{G}_{(0)}s) \tag{5.83}$$

and this shows that all the information about the topological quantum number is contained in 'generalised magnetic charge', $\mathbf{G}_{(0)}$, defined by the asymptotic (generalised) magnetic field.

Given the assumed form for the field of equation (5.78) and equation (5.79) these results are manifestly independent of the parametrisation, $(s, t)$, and the choice of surface, $\Sigma$, and it is not difficult to see that they vary only by conjugation on changing $\hat{\boldsymbol{r}}_0$ or the gauge. So far what we have said is valid for any compact group, including the familiar case of $H=\mathrm{U}(1)$. We will now investigate the particular features of the non-Abelian case.

In the non-Abelian case $\mathbf{G}_{(0)}$ will transform under a gauge transformation $\gamma(\boldsymbol{r})$:

$$\mathbf{G}_{(0)}\to\gamma_0\mathbf{G}_{(0)}\gamma_0^{-1} \qquad \text{where} \qquad \gamma_0=\gamma(\boldsymbol{r}_0).$$

We should like to extract from $\mathbf{G}_{(0)}$ a gauge-invariant structure subject to the condition of equation (5.82). Consider $H=\mathrm{SU}(2)$ or $\mathrm{SO}(3)$ for definiteness. Then $\mathbf{G}_{(0)}$ is a linear combination of the three generators $T^1$, $T^2$, $T^3$ and we can remove most of the arbitrariness by using the gauge freedom to rotate $\mathbf{G}_{(0)}$ till it is parallel to $T^3$:

$$\mathbf{G}_{(0)}\to\beta T^3.$$

Clearly the magnitude of $\beta$ is determined by the length of $\mathbf{G}_{(0)}$, but there is a sign ambiguity (corresponding to the possibility of rotating through $\pi$ about the second axis, say). Otherwise $\beta$ contains precisely the gauge-invariant information available in $\mathbf{G}_{(0)}$. The quantisation condition of equation (5.82) becomes:

$$\exp(ie\beta T^3)=1. \tag{5.84}$$

If $H=\mathrm{SO}(3)$ and $T^3$ is chosen as in §4.1, it has integral eigenvalues (called in this context the weights of SO(3)). Equation (5.84) then yields:

$$\beta=\frac{4\pi}{e}\frac{n}{2} \qquad \text{for some } n\in\mathbb{Z}. \tag{5.85}$$

If $H=\mathrm{SU}(2)$ and $T^3=\frac{1}{2}\sigma^3$, it has half-integral eigenvalues (the weights of SU(2)). Equation (5.84) then yields:

$$\beta=\frac{4\pi}{e}n \qquad \text{for some } n\in\mathbb{Z}. \tag{5.86}$$

Thus we note that if $H=\mathrm{SU}(2)$, $\beta$ is $4\pi/e$ times a weight of SO(3) and, conversely, if $H=\mathrm{SO}(3)$, $\beta$ is $4\pi/e$ times a weight of SU(2): the two groups SO(3) and SU(2) appear in a dual relationship.

The quantised structure of the 'magnetic weights' $e\beta/4\pi$ provides a finer gauge-invariant classification of the monopoles than that given by the topological quantum numbers (provided we regard $\beta$ and $-\beta$ as equivalent). For $H=\mathrm{SO}(3)$, the topological quantum numbers have the structure of $\Pi_1(\mathrm{SO}(3))=\mathbb{Z}_2$. We see from equation (5.83) that the integral values of $e\beta/4\pi$ are all equivalent topologically to the trivial path, whilst the non-integral values are all equivalent to the only homotopically distinct non-trivial path, a rotation through $2\pi$. Thus the topological quantum number is the integer $n$ in equation (5.85) taken modulo 2. For $H=\mathrm{SU}(2)$ all of the values of $e\beta/4\pi$ correspond to topologically trivial paths since $\Pi_1(\mathrm{SU}(2))=0$.

The first example of a non-Abelian monopole presented in the literature was for an SO(3) gauge theory (Wu and Yang 1969). It corresponds to the topologically trivial case $n=2$ in equation (5.85). Indeed the fact that it was written down without the aid of Higgs fields or Dirac strings makes the absence of a topological quantum number obvious.

Although it is topologically trivial it is not clear that such monopoles are trivial from a physical point of view. It has been conjectured (Goddard *et al* 1977) that in a complete quantum field theory the monopoles would behave as multiplets of the dual group (that is, SU(2) if $H=\mathrm{SO}(3)$ and vice versa). Then the combination law for monopoles would be given by the Clebsch–Gordan series for combining representations for the dual group. Thus the total generalised magnetic charge would not be determined uniquely by the charges of its constituents since the Clebsch–Gordan series for the product of two irreducible representations contains a number of different irreducible representations. The sign ambiguity in $\beta$ mentioned above provides a mechanism for introducing ambiguities in combining monopoles, even classically.

The concept of the dual group, $H^{\mathrm{v}}$, of a given group $H$ extends to an arbitrary compact connected gauge group (Englert and Windey 1976, Goddard *et al* 1977). (The global structure of $H^{\mathrm{v}}$ is determined by $H$, and the Lie algebras differ in the general case.) An important distinction between non-Abelian and Abelian monopoles is the discrete ambiguity in defining the magnetic weights $\beta$, suitably generalised. (This discrete ambiguity generalises to the magnetic weights being defined modulo the Weyl group.) It is clear that the appropriate group for classifying states is $H\times H^{\mathrm{v}}$ rather than just $H$ if $H=\mathrm{U}(1)$, for which $H^{\mathrm{v}}=\mathrm{U}(1)$, since the states must be classified by both the electric and magnetic charge.

Much more could be said about these conjectures (Goddard *et al* 1977, Montonen and Olive 1977) but proofs would require a much deeper understanding of the quantum field theory of monopoles than is available at present. We will make some further comments in §7.2.

### *5.9. The relationship to the Wu–Yang formulation*

In this section, in which we have tried to classify monopoles by their long-range, or macroscopic, structure, the significance of the Higgs field has progressively declined in favour of the exact symmetry group, $H$, and its gauge fields until, in the last subsection, it has completely disappeared from consideration. On the other hand, let us emphasise that the Higgs field appears to be an essential ingredient in the microscopic structure of the monopole if it is to have finite energy. There is an alternative formulation of generalised monopoles due to Wu and Yang making no reference to Higgs fields at all, which we shall now discuss and relate to the analysis of the earlier subsections.

Wu and Yang (1975, 1976) formulated their approach for a Dirac monopole and then generalised it to the case of a non-Abelian group $H$, discussing SU(2) and SO(3) in particular. In this formulation the group $H$, assumed compact and connected, is the only gauge symmetry; in our previous language $G=H$ since there is no Higgs field. In consequence there must, in general, be strings if we try to describe the whole solution in a single gauge.

Let us describe the region around a monopole by spherical polar coordinates $(r, \theta, \chi)$. In the Wu–Yang formulation the monopole has no internal structure and may be considered to be located at the origin. In the Higgs field formulation we may consider the monopole to be located effectively inside some finite radius, $a_0$, say, as an abbreviation for the fact that its internal structure decays exponentially at large distance. We divide the region outside the monopole, i.e. $r>a_0$ (with $a_0=0$ in the Wu–Yang case), into two overlapping regions: an upper region $R_+$, for which $0\leqslant\theta\leqslant\frac{1}{2}\pi+\epsilon$, and a lower region $R_-$, for which $\frac{1}{2}\pi-\epsilon\leqslant\theta\leqslant\pi$, where $\epsilon$ is a fixed positive constant.

The Wu–Yang formulation involves only non-singular $H$ gauge fields written in two different gauges in the two overlapping regions $R_+$ and $R_-$. Although there may not be any gauge in which the $H$ gauge fields would be non-singular everywhere, the fundamental requirement of Wu and Yang is that there exists a non-singular gauge transformation $\eta$, defined through the overlap region $R_+\cap R_-$, which takes us from the $R_-$ gauge to the $R_+$ gauge. In particular, it is crucial that $\eta(r, \frac{1}{2}\pi, \chi)$ be single-valued as a function of the azimuthal angle, $\chi$. Then, for any fixed value of $r$:

$$\eta\equiv\eta(r, \tfrac{1}{2}\pi, 2\pi s)\in H \qquad 0\leqslant s\leqslant 1 \tag{5.87}$$

defines a closed path in $H$ and, hence, an element of $\Pi_1(H)$. Clearly this homotopy class is independent of $r$ and, further, would be the same if we replaced $(r, \frac{1}{2}\pi, 2\pi s)$, $0\leqslant s\leqslant 1$, by any homotopic path in $R_+\cap R_-$. It is also clear that the class will be gauge-invariant, assuming $H$ to be connected, and so characterises some intrinsic property of the monopole.

For $H=\mathrm{U}(1)$, the electromagnetic gauge group, Wu and Yang (1975) showed that the single valuedness of $\eta$ leads to the Dirac quantisation condition and that the homotopy class of $\eta$ is labelled by the magnetic charge. The relationship between the work of Wu and Yang and Dirac has been discussed by Brandt and Primack (1977a).

Returning to the Higgs field formulation, in a region, like $r>a$, in which the field configuration is in the Higgs vacuum, we may pass to the Wu–Yang formulation as follows. (Of course, it is essential that all the degrees of freedom are contained in the $H$ gauge fields for us to be able to do this.) We find gauge transformations $g_+(\boldsymbol{r})$ and $g_-(\boldsymbol{r})$ defined for $\boldsymbol{r}\in R_+$ and $\boldsymbol{r}\in R_-$ respectively such that:

$$\phi(\boldsymbol{r})=g_\pm(\boldsymbol{r})\,\phi_0 \qquad \text{for} \qquad \boldsymbol{r}\in R_\pm.$$

We apply $g_\pm(\boldsymbol{r})^{-1}$ to obtain $H=H_{\phi_0}$ gauge fields in the two regions $R_\pm$. Then $\eta$ is defined by:

$$\eta(\boldsymbol{r})=g_+(\boldsymbol{r})^{-1}\,g_-(\boldsymbol{r})\in H \tag{5.88}$$

for $\boldsymbol{r}\in R_+\cap R_-$.

In §5.5, we saw how to associate another path $h(s)$ in $H$ or, more properly, an element of $\Pi_1(H)$, with a monopole in the Higgs field formulation. What we wish to show now is that:

(*a*) The construction of $h$ can be extended to the Wu–Yang formulation, given their single-valuedness condition.

(*b*) In either formulation, $h$ and $\eta$ define the same element of $\Pi_1(H)$, i.e. are homotopic.

It will then follow that $\eta$ provides an alternative and more direct formulation of the topological quantum number, which we have discussed in §§5.4 and 5.8, which generalises to situations in which the Higgs field is absent.

To demonstrate the validity of proposition (*a*) we need to use again the path-dependent phase factors introduced in equation (5.58), defining:

$$\zeta(\mathscr{C}) = \mathscr{T}\left[\exp\left(ie\int_0^1 \mathbf{W}^i \frac{\partial r^i}{\partial t}\,dt\right)\right] \in H \tag{5.89}$$

the path-dependent phase factor associated with the path, $\mathscr{C}$: $\boldsymbol{r}(t)$, $0 \leqslant t \leqslant 1$. We shall consider paths on a given sphere, $S$, $r = a > a_0$. Let $\zeta^+(\theta, \phi)$ denote the path-dependent phase factor associated with the path $\boldsymbol{r}(a, \theta s, \phi)$, $0 \leqslant s \leqslant 1$, which proceeds from the 'North pole', $\boldsymbol{r}_0(\theta = 0)$, a distance $a\theta$ down a meridian and let $\zeta^-(\theta, \phi)$ denote the phase factor associated with the path $\boldsymbol{r}(a, \pi - \theta s, \phi)$, $0 \leqslant s \leqslant 1$, which proceeds from the 'South pole', $\boldsymbol{r}_\pi(\theta = \pi)$, a distance $a\theta$ up a meridian. In each case it is assumed that $0 \leqslant \theta \leqslant \frac{1}{2}\pi$. These paths stay conveniently within one of the regions, $R_\pm$, and the gauge fields $\mathbf{W}^i$ in equation (5.89), are taken in the corresponding gauges to define $\zeta^\pm(\theta, \phi)$. Now consider a path, $\Lambda_\phi$, proceeding from $\boldsymbol{r}_0$, down the $\phi = 0$ (Greenwich) meridian to $\boldsymbol{r}_\pi$ and then back up the $\phi$ meridian to $\boldsymbol{r}_0$. In the $R_+$ gauge we would associate with $\Lambda_\phi$ the path-dependent phase factor:

$$h(s) = \zeta^+(\tfrac{1}{2}\pi, \phi)^{-1}\,\eta(s)\,\zeta^-(\tfrac{1}{2}\pi, \phi)\,\zeta^-(\tfrac{1}{2}\pi, 0)^{-1}\,\eta(0)^{-1}\,\zeta^+(\tfrac{1}{2}\pi, 0) \qquad \text{where} \qquad s = \phi/2\pi. \tag{5.90}$$

We have had to split up $\Lambda_\phi$ into segments in the northern and southern hemispheres and use the transition function, $\eta$, to relate them where $\Lambda_\phi$ crosses the 'equator'. The element $h(s)$ is perfectly well-defined and gauge-covariant in the sense that:

$$h(s) \rightarrow \xi(\boldsymbol{r}_0)\,h(s)\,\xi(\boldsymbol{r}_0)^{-1}$$

under a gauge transformation $\xi(\boldsymbol{r})$. Further $h(s)$, $0 \leqslant s \leqslant 1$, defines a closed loop in $H$ provided the single-valuedness assumption of Wu and Yang holds.

It is not difficult to see that the path $h$ of equation (5.90) agrees, at least up to homotopy, with that introduced in equation (5.59) if we are working in the framework of a Higgs field formulation. This demonstrates the truth of proposition (*a*), as it generalises $h$ to the Wu–Yang formulation. To see that $h$, so defined, is homotopic to $\eta$, consider:

$$h_\theta(s) = \zeta^+(\theta, \phi)^{-1}\,\eta(s)\,\zeta^-(\theta, \phi)\,\zeta^-(\tfrac{1}{2}\pi, 0)^{-1}\,\eta(0)^{-1}\,\zeta^+(\tfrac{1}{2}\pi, 0) \qquad \text{where} \qquad s = \phi/2\pi.$$

Using the Wu–Yang assumption again, $h$ describes a closed path in $H$ for each value of $\theta$, $0 \leqslant \theta \leqslant \frac{1}{2}\pi$. Further:

$$h_{\pi/2} = h \qquad \text{whereas} \qquad h_0 = \eta k$$

where

$$k = \zeta^-(\tfrac{1}{2}\pi, 0)^{-1}\,\eta(0)^{-1}\,\zeta^+(\tfrac{1}{2}\pi, 0)$$

and can be continuously changed to 1 since $H$ is connected. So we conclude that $\eta$ and $h$ are homotopic as claimed.

The formulation of Wu and Yang (1975) is closely related to the mathematical language of fibre bundles. Gauge theories have been discussed in this language by

a number of authors including Lubkin (1963), Trautman (1970), Ezawa and Tze (1976a, b, 1977) and Maison and Orfanidis (1977). In the Wu–Yang formulation the monopole is described by a *non-trivial* fibre bundle with structural group $H$ whilst in the Higgs field formulation it is described by a *trivial* fibre bundle with structural group $G$. It is this triviality which enables one to avoid a singularity inside the monopole. What we have described in this section is how the Higgs field enables one to *reduce* the structural group of the bundle from $G$ to $H$.

## 6. Microscopic properties of generalised monopoles

### *6.1. Elementary monopoles*

This section will necessarily be much shorter than the preceding one. It deals with a subject on which there is a much smaller, though growing, body of information. In the analysis of the macroscopic structure of monopoles we showed that the Higgs field could be dispensed with, but when the microscopic structure is examined, it seems to be essential for singularity-free, finite-energy monopole solutions. (Some authors, e.g. Troost and Vinciarelli (1976), would disagree with this statement.)

First we would like to formulate criteria which will ensure that we are considering a single *elementary monopole* rather than a bound state (or scattering state) of monopoles. Reasonable sufficient conditions appear to be (Corrigan *et al* 1976):

(i) There exists a Lorentz frame (the centre-of-mass frame) and a gauge in which all fields are time-independent.

(ii) Further, the origin of this frame may be so chosen that the solution is spherically symmetric with respect to the rotations generated by:

$$\boldsymbol{J}_0/\hbar = -\mathrm{i}\boldsymbol{r}\wedge\boldsymbol{\nabla}+\boldsymbol{t} \tag{6.1}$$

where $t^1$, $t^2$, $t^3$ are (constant) generators of the gauge group $G$, satisfying the angular momentum algebra:

$$[t^i, t^j]=\mathrm{i}\epsilon_{ijk}t^k. \tag{6.2}$$

Explicitly this means that the Higgs field $\phi$ and the gauge fields $\mathbf{W}^\mu$ satisfy:

$$D(\boldsymbol{J}_0)\,\phi=0 \tag{6.3}$$

$$[\boldsymbol{J}_0, \mathbf{W}^0]=0 \tag{6.4}$$

$$[J_0{}^i, \mathbf{W}^j]=\mathrm{i}\hbar\epsilon_{ijk}\mathbf{W}^k. \tag{6.5}$$

These conditions are all satisfied by the 't Hooft–Polyakov monopole discussed in §4, which also satisfies:

$$\mathbf{W}^0=0 \tag{6.6}$$

and this together with condition (i) above implies that, in the centre-of-mass frame of the monopole:

$$\mathbf{G}^{i0}=0. \tag{6.7}$$

The conditions (i) and (ii) are also satisfied by the dyon solutions of §4.8 which do not satisfy the further restrictions of equations (6.6) and (6.7).

Condition (ii) may ensure that the corresponding quantum-mechanical monopole has a definite spin angular momentum (Montonen and Olive 1977). On the classical level it will certainly enable the field equations to be reduced to purely radial ones.

In the next subsection we shall illustrate the procedure by considering monopoles in an SU(3) gauge theory. Then, in §6.3 we shall present some general deductions from the conditions introduced in this subsection, illustrating them from specific examples. Section 6.4 reviews some other work on spherical symmetry.

### *6.2. Monopoles in SU(3) gauge theories*

After SU(2), it is natural to consider SU(3) as the next most tractable gauge group. The simplest possibility for the Higgs field is to have it in the adjoint representation, and we shall consider this first. Then there are several possibilities for the exact symmetry group $H$, depending on the orbit of the Higgs field which constitutes $\mathcal{M}_0$, the set of minima of the self-interaction. The orbit of a given value $\phi$ of the Higgs field is determined by the eigenvalues of the corresponding matrix $\underline{\Phi}=\frac{1}{2}\phi_a\lambda^a$ in the Lie algebra of SU(3). (We use $\lambda^a$, $a=1, \ldots, 8$, to denote the conventional basis for traceless Hermitian $3\times 3$ matrices; see, for example, Gell-Mann and Ne'eman 1964.) The matrix $\underline{\Phi}$ may have either one, two or three distinct eigenvalues and then $H=\mathrm{SU}(3)$, $\mathrm{U}(2)$ or $\mathrm{U}(1)\times\mathrm{U}(1)$, respectively. In the first case, $\underline{\Phi}=0$, since it is traceless, and so $\mathcal{M}_0$ is topologically trivial; this case is uninteresting. The second case is the generic one for the general renormalisable self-interaction and is particularly interesting in view of the SU(2) subgroup. In this we can apply an SU(3) transformation to $\underline{\Phi}$ to take it parallel to $\lambda^8$ whilst for a general $\underline{\Phi}$ we can say that it is possible to take it into some linear combination of $\lambda^3$ and $\lambda^8$.

We shall discuss the case of $H=\mathrm{U}(2)$ in some detail. Here we have the situation described in §5.7 and we can identify $Q=e\hbar\underline{\Phi}/a$ as the electric charge operator in the Higgs vacuum as in equation (5.69). Here $a$ is the value of the length of $\underline{\Phi}$ in $\mathcal{M}_0$. The generators of $H$ orthogonal to $\underline{\Phi}$ as in equation (5.67) generate a colour group $K\simeq\mathrm{SU}(2)$. The quantisation condition of equation (5.71) takes the form:

$$\exp(\mathrm{i}gQ/\hbar)=\exp(2\pi\mathrm{i}K^3)\text{ or }1 \tag{6.8}$$

where $K^3$ is a generator of the colour group, $K$, normalised to have half-integral eigenvalues. Thus if $q_0$ is the smallest eigenvalue of $Q$, namely $e\hbar/2\sqrt{3}$:

$$\frac{q_0 g}{4\pi\hbar}=\tfrac{1}{4}M \tag{6.9}$$

where $M$ is an even or odd integer, respectively.

There are two gauge inequivalent choices for $\boldsymbol{t}$ in equation (6.1):

(*a*) $\boldsymbol{t}=(\frac{1}{2}\lambda^1, \frac{1}{2}\lambda^2, \frac{1}{2}\lambda^3)=\frac{1}{2}\boldsymbol{\lambda}$. The $t^a$ have eigenvalues $-\frac{1}{2}$, 0, $\frac{1}{2}$, and generate an SU(2) subgroup of SU(3). The general solution to equation (6.3) takes the form:

$$\underline{\Phi}(\boldsymbol{r})=\tfrac{1}{2}\boldsymbol{\lambda}.\hat{\boldsymbol{r}}\alpha(r)+\tfrac{1}{2}\lambda^8\beta(r). \tag{6.10}$$

Demanding that $\underline{\Phi}$ lie asymptotically in $\mathcal{M}_0$ yields $\underline{\Phi}(\boldsymbol{r})\to\underline{\Phi}_\infty(\hat{\boldsymbol{r}})$ as $r\to\infty$ with:

$$\underline{\Phi}_\infty(\hat{\boldsymbol{r}})=\tfrac{1}{2}a(\pm\sqrt{3}\boldsymbol{\lambda}.\hat{\boldsymbol{r}}-\lambda^8)$$

for $M=\pm 1$ or:

$$\underline{\Phi}_\infty(\hat{\boldsymbol{r}})=a\lambda^8$$

for $M=0$.

(*b*) $\boldsymbol{t}=(\lambda^7, -\lambda^5, \lambda^2)$. The $t^a$ have eigenvalues $-1$, 0, 1 and generate an SO(3)

subgroup of SU(3). Then the most general solution to equation (6.3) takes the form:

$$\begin{aligned}\underline{\Phi}(\boldsymbol{r})_{\alpha\beta} &= (\hat{r}_\alpha \hat{r}_\beta - \tfrac{1}{3}\delta_{\alpha\beta})\, A(r) + \mathrm{i}\epsilon_{\alpha\beta\gamma}\hat{r}_\gamma B(r) \\ &\equiv \underline{\Psi}_{\alpha\beta}{}^{(1)}(\hat{\boldsymbol{r}})\, A(r) + \underline{\Psi}_{\alpha\beta}{}^{(2)}(\hat{\boldsymbol{r}})\, B(r), \text{ say.}\end{aligned} \tag{6.11}$$

Demanding that $\underline{\Phi}$ lie asymptotically in $\mathcal{M}_0$ yields:

$$\underline{\Phi}_\infty(\hat{\boldsymbol{r}}) = \frac{\sqrt{3}}{4}\, a(\underline{\Psi}^{(1)} \pm \underline{\Psi}^{(2)})$$

for which $M = \pm 2$, or

$$\underline{\Phi}_\infty(\hat{\boldsymbol{r}}) = -\frac{\sqrt{3}}{2}\, a\underline{\Psi}^{(2)}$$

for which $M = 0$.

Notice that for case (*b*), $M$ is even in equation (6.9) so that the original and stricter Dirac quantisation condition is satisfied, whereas in case (*a*) it is not. Similar decompositions can be made for the gauge potentials satisfying the condition of equation (6.5) and further simplified using spherically symmetric gauge transformations as in §4.2. For further discussion and fuller details we refer the reader to the work of Corrigan *et al* (1976). Other work on this topic is contained in Wu and Wu (1974), Marciano and Pagels (1975), Chakrabarti (1975), Tyupkin *et al* (1975), Gursey (1976), Horvath and Palla (1976a), Sinha (1976) and Madore (1977). Analogous treatments of SU(4) monopoles have been given by Brihaye and Nuyts (1977), Kaku (1976) and Wilkinson (1977). There is related work on SU($N$) by Goldhaber and Wilkinson (1976) and Horvath and Palla (1976b, 1977).

If we relax the constraint that $\phi$ transform under the adjoint representation of SU(3) another interesting possibility is available to us, namely to arrange for $H = \mathrm{SO}(3)$. SO(3) is the smallest non-Abelian group which admits the possibility of a topologically stable monopole (since it has a non-trivial closed path, a rotation through $2\pi$) and SU(3) is the smallest simply connected group with an SO(3) subgroup. This, we argued in §5, is the criterion for a non-singular 'soliton-like' or monopole solution.

M Sato (private communication) has pointed out that we may arrange for $H = \mathrm{SO}(3)$ by taking $\phi$ to be in one of the two conjugate six-dimensional representations of SU(3). We may conveniently realise this representation by representing $\phi$ by a symmetric $3 \times 3$ matrix which transforms according to:

$$\phi \to u\phi u^{\mathrm{T}} \qquad \text{for} \qquad u \in \mathrm{SU}(3).$$

To obtain $H = \mathrm{SO}(3)$ one takes $\mathcal{M}_0$ to be the orbit of (some multiple of) the unit matrix, $1_3$. The little group of $\phi_0 = a1_3$ then consists of $3 \times 3$ matrices satisfying:

$$uu^{\mathrm{T}} = u\bar{u}^{\mathrm{T}} = 1_3$$

which imply $u \in \mathrm{SO}(3)$, the subgroup generated by $\lambda_7$, $-\lambda_5$ and $\lambda_2$. One may proceed to discuss this example in the way we have just discussed the case where $\phi$ is in the adjoint representation and, since the **6** representation is described by symmetric $3 \times 3$ matrices whilst the **8** representation is described by traceless $3 \times 3$ matrices, the formalism is rather similar. Again there are two cases to consider, depending on whether $\boldsymbol{t}$ defines an embedding of SU(2) or SO(3) in SU(3), corresponding to (*a*) and (*b*) above. We shall see in the next subsection that the solution corresponding to (*a*) $\boldsymbol{t} = \frac{1}{2}\boldsymbol{\lambda}$ is topologically stable whilst the other is not. As yet there is no complete discussion of this model in the literature.

*6.3. General properties of spherically symmetric monopoles*

In this subsection we shall present some results which correlate and generalise features of the specific cases we have discussed. In considering spherically symmetric monopole solutions for a general gauge group, $G$, it is not difficult to catalogue the gauge-inequivalent ways of embedding the rotation group generators $\boldsymbol{t}$ in the Lie algebra of $G$ using the theory of Dynkin. (For a recent application of this theory in the context of instantons see Bitar and Sorba (1977).) But this is not the whole story since $H_{\phi(r)}$ and, in particular, its orientation with respect to subgroup $R$ generated by $t^1$, $t^2$ and $t^3$ also plays a role. Specifically, $H_{\phi(r)}$ must have at least one generator in common with $R$, namely the radial component of $\boldsymbol{t}$, $\hat{\boldsymbol{r}}.\boldsymbol{t}=\hat{\boldsymbol{r}}.\boldsymbol{J}_0/\hbar$, using equation (6.1) since:

$$D(\hat{\boldsymbol{r}}.\boldsymbol{t})\,\phi(\boldsymbol{r})=0. \tag{6.12}$$

This is important because $\hat{\boldsymbol{r}}.\boldsymbol{t}$ completely determines the topological quantum number as an element of $\Pi_1(H)$ through the map (Olive 1976):

$$h(s)=\exp(\mathrm{i}\hat{\boldsymbol{r}}.\boldsymbol{t}4\pi s) \qquad 0\leqslant s\leqslant 1 \tag{6.13}$$

which is a closed path because $t^a$ has half-integral eigenvalues.

To prove that the path of equation (6.13) determines the topological quantum number note that we can always choose a gauge in which equation (5.76) holds without disturbing the spherical symmetry conditions, equations (6.3)–(6.5). Then equation (5.30) implies:

$$\boldsymbol{r}.\nabla\phi=0.$$

Taking the vector product of equation (6.3) with $\boldsymbol{r}$ thus yields:

$$r^2\nabla\phi-\mathrm{i}D(\boldsymbol{r}\wedge\boldsymbol{t})\,\phi=0$$

from which we deduce that a possible solution to equation (5.30) for $\mathbf{W}$ is, given $\phi$:

$$\mathbf{W}_{(0)}{}^i=\frac{1}{er^2}\,\epsilon_{ijk}r^j t^k. \tag{6.14}$$

The corresponding field strengths which would be calculated from $\mathbf{W}_{(0)}{}^i$ are:

$$\mathbf{G}_{(0)}{}^{ij}=\frac{1}{er^2}\,\hat{\boldsymbol{r}}.\boldsymbol{t}\epsilon_{ijk}\hat{r}^k. \tag{6.15}$$

Although the potentials and field strengths are not necessarily of the form of equations (6.14) and (6.15) the analysis of §§5.6 and 5.8 shows that we may use $\mathbf{W}_{(0)}{}^i$ and $\mathbf{G}_{(0)}{}^{ij}$ to calculate the topological quantum numbers as indicated in equations (5.78) and (5.83); this does indeed yield equation (6.13).

We may deduce a number of corollaries to the result contained in equation (6.13). Firstly, if the generators $\boldsymbol{t}$ have integral eigenvalues (so that they provide an embedding of SO(3) rather than SU(2) in $G$):

$$\exp(\mathrm{i}\hat{\boldsymbol{r}}.\boldsymbol{t}4\pi s) \qquad 0\leqslant s\leqslant \tfrac{1}{2}$$

defines a closed path in $H$ and we have the more stringent quantisation condition that only those elements of $\Pi_1(H)$ which are squares of other elements can be realised as topological quantum numbers by this sort of monopole solution.

Secondly, consider the case studied in §5.7 in which $H$ locally has the structure of $\mathrm{U}(1)\times K$. The topological quantum numbers are uniquely labelled by multiples,

$gQ/\hbar$, of the generator of the electromagnetic U(1) together with (the homotopy classes of) paths in $K$ from 1 to $k$, $\in Z(K)$, such that equation (5.71) is satisfied. Thus:

$$\hat{\boldsymbol{r}}.\boldsymbol{t}=\frac{gQ}{4\pi\hbar}+K_0 \tag{6.16}$$

where $K_0$ is a generator of the colour group, $K$. This can be checked directly (Olive 1976). It can be regarded as a generalisation of equation (2.18). Note that equation (6.16) immediately implies the quantisation condition of equation (5.71) since $\exp(4\pi i\hat{\boldsymbol{r}}.\boldsymbol{t})=1$. Further, if we are dealing with SO(3) embedding, i.e. $\hat{\boldsymbol{r}}.\boldsymbol{t}$ has integral eigenvalues, we have the stronger condition:

$$\exp(igQ/2\hbar)\in K \tag{6.17}$$

which is sufficient to explain why, when in §6.2 we considered SU(3) gauge theories, the monopoles in case (*b*) satisfied a stricter quantisation condition than those in case (*a*).

Now, consider the special case in which $H=\mathrm{U}(1)$ and there is no colour group. Equation (6.16) reduces to:

$$\hat{\boldsymbol{r}}.\boldsymbol{t}=gQ/4\pi\hbar. \tag{6.18}$$

Let $q_0$ denote the smallest non-zero eigenvalue of $Q$. The smallest non-zero eigenvalue of $\hat{\boldsymbol{r}}.\boldsymbol{t}$ is either $\frac{1}{2}$ or 1 depending on whether $\boldsymbol{t}$ provides in embedding of SU(2) or SO(3). Equating the smallest eigenvalues on the two sides of equation (6.18) we see that the magnetic charge of a spherically symmetric U(1) monopole must satisfy:

$$\frac{q_0 g}{4\pi\hbar}=\pm\tfrac{1}{2} \qquad \text{or} \qquad \pm 1 \tag{6.19}$$

respectively. Thus it has at most two Dirac units of magnetic charge whatever the nature of $G$ (Olive 1976). If $G=\mathrm{SU}(2)$, $q_0$ is one-half the gauge particle charge and $g$ can only be one Dirac unit. This result has been obtained in a number of other ways in the literature (Cremmer *et al* 1976, Weinberg and Guth 1976).

In the Salam–Weinberg model (Salam 1968, Weinberg 1967), where $G=\mathrm{SU}(2)\times\mathrm{U}(1)$ with $H=\mathrm{U}(1)$ generated by $Q$ not lying within either factor of $G$, spherically symmetric monopoles are impossible since the $\boldsymbol{t}$ must generate the SU(2) factor of $G$ and equation (6.18) then implies that $Q$ is an SU(2) generator, contrary to hypothesis.

In the model of the previous subsection, in which $G=\mathrm{SU}(3)$, $H=\mathrm{SO}(3)$, we see that equation (6.13) defines the trivial element of $\Pi_1(H)$ in case (*b*) but a non-trivial one in case (*a*) supporting the statements made about topological stability there.

Recently Wilkinson and Goldhaber (1977) have made progress towards finding all spherically symmetric monopoles for an arbitrary compact semi-simple gauge group. They follow the policy of first seeking 'point solutions' which are spherically symmetric and satisfy:

$$\mathscr{D}^i\phi=0 \tag{6.20}$$

everywhere but at the origin. These hopefully approximate exact finite-energy solutions asymptotically. (Earlier related work is contained in Goldhaber and Wilkinson (1976) and Bais and Primack (1977).) Alternatively we may present some of their conclusions within the framework of the sort of asymptotic assumptions made in §5.8, together with the assumptions of spherical symmetry set out in §6.1. In

particular, if we make the assumptions contained in equations (5.77) and (6.6) then, in the radial gauge of equation (5.76), equations (5.78) and (5.79) follow so that asymptotically $\mathbf{G}_{ij}$ has the inverse square law form:

$$\mathbf{G}_{ij}=\frac{1}{4\pi r^2}\,\epsilon_{ijk}\hat{r}^k\mathbf{G}(\hat{r}) \tag{6.21}$$

where $\mathbf{G}(\hat{r})$ is covariantly constant. (Wilkinson and Goldhaber remark that this holds for a point solution everywhere but the origin under suitable assumptions.) It is then straightforward to demonstrate that the generalised angular momentum:

$$J^i=\mathrm{i}\hbar\epsilon_{ijk}r^j\mathscr{D}^k+\frac{e\hbar}{4\pi}\hat{r}^i\mathbf{G}(\hat{r}) \tag{6.22}$$

satisfies the algebra:

$$[J^i, J^j]=\mathrm{i}\hbar\epsilon_{ijk}J^k \tag{6.23}$$

and the spherical symmetry assumption of equation (6.5) further implies:

$$[J_0{}^i, J^j]=\mathrm{i}\hbar\epsilon_{ijk}J^k. \tag{6.24}$$

From equations (6.23) and (6.24) and the fact that $\boldsymbol{J}_0$ satisfies an angular momentum algebra it follows that:

$$I^i=(J_0{}^i-J^i)/\hbar \tag{6.25}$$

satisfies an SU(2) algebra. If we assume equation (6.20) holds in an appropriate approximation:

$$D(\boldsymbol{J})\,\phi=0$$

and so

$$\boldsymbol{I}(\hat{r})=\boldsymbol{t}+\boldsymbol{r}\wedge\boldsymbol{W}-\frac{e}{4\pi}\,\mathbf{G}\hat{r} \tag{6.26}$$

generates an SU(2) (or SO(3)) subgroup of $H_\phi$. Thus we find that (Wilkinson and Goldhaber 1977) for a spherically symmetric monopole the generalised magnetic charge satisfies:

$$\frac{e}{4\pi}\,\mathbf{G}(\hat{r})=\hat{r}.(\boldsymbol{t}-\boldsymbol{I}(\hat{r})) \tag{6.27}$$

where $\boldsymbol{t}$ and $\boldsymbol{I}$ both satisfy SU(2) algebras, $\boldsymbol{t}$ being the internal generators of the spherical symmetry and $\boldsymbol{I}$ being generators of the little group $H_\phi$. This yields a necessary and sufficient condition for the existence of spherically symmetric point solutions. From equation (6.27) it is easy to verify the homotopic equivalence of the paths (5.83) and (6.13) specifying the topological quantum number of the solution, since the path:

$$\exp(-\mathrm{i}\hat{r}.\boldsymbol{I}4\pi s)\qquad 0\leqslant s\leqslant 1$$

is trivial in the SU(2) or SO(3) group generated by $\boldsymbol{I}$, which is a subgroup of $H$. Wilkinson and Goldhaber use their result to give a diagrammatic technique for finding point solutions to SU($N$) gauge theories.

### *6.4. Other solutions*

The spherical symmetry assumptions of equations (6.1)–(6.5) are unsatisfactory

as they stand since they lack a rigorous derivation from a reasonably general set of basic assumptions. It seems that so far alternative strategies have, in practice, led back to these assumptions. Weinberg and Guth (1976) formulated a more general, gauge-covariant expression of spherical symmetry which, they showed, retrieved only the known solutions in the 't Hooft–Polyakov case (see also O'Raifeartaigh 1979).

Michel *et al* (1977a, b) sought solutions which, like equation (4.17), were separable into products of radial and angular functions, but showed that there existed a gauge in which such solutions satisfied the spherical symmetry assumptions given here, equations (6.1)–(6.5). Indeed their assumptions are, in effect, more restrictive than the ones we have made since the solution of equation (6.10) does not satisfy the separability condition.

Another approach is to attempt to generalise the Bogomolny–Prasad–Sommerfield monopole of §4.7. The argument for the Bogomolny bound given in §4.6 generalises to any gauge group provided that the Higgs field lies in the adjoint representation. Then $H$ is necessarily locally of the form $\mathrm{U}(1)\times K$ as described in §5.7 and the argument shows that the mass of any monopole with U(1) magnetic charge $g$ satisfies:

$$M \geqslant a|g|.$$

One may seek to generalise the BPS monopole by attempting to saturate this bound when $V(\phi)\equiv 0$. It is natural to expect saturation to happen for some spherically symmetric solution, if at all. Czechowski (1977) has discussed this possibility for $G=\mathrm{SU}(3)$. An important open question is whether an analogous bound can be obtained if the Higgs field is not in the adjoint representation.

Another major task remaining at the level of classical solutions is that of constructing solutions describing two or more monopoles. Then their interactions can be studied. Some progress in this direction at the macroscopic level, made by Manton (1977), was briefly described in §4.7.

## 7. Epilogue

### *7.1. Solitons and scale transformations*

An interesting aspect of the preceding sections is the dependence of the discussion upon the dimensionality of space–time, which we alluded to in §3.4. Consider a field theory in $D$ space and one time dimensions of the sort discussed in §5; the topological quantum number associated with the boundary condition on the Higgs field corresponds to an element of $\Pi_{D-1}(G/H)$. The fact that we are dealing with $D=3$ enabled us to exploit the isomorphism of (5.47) which reflects the fundamental result of Cartan that $\Pi_2(G)=0$. On the other hand, we have not verified that there exist stable finite-energy classical solutions for values of $D$ other than 3. To attack this question we follow a line of argument originated by Derrick (1964) and elaborated by Coleman (1975b) and Faddeev (1976a) based on the exploitation of simple scale transformations. This enables one to dismiss many possible theories as being incapable of supporting stable finite-energy time-independent solutions.

Suppose that the energy can be written as the sum of three positive terms:

$$H = T_\phi + T_W + V \tag{7.1}$$

where

$$T_\phi[\phi, W] = \int \mathrm{d}^D x F(\phi)(\mathscr{D}^i\phi)^\dagger \mathscr{D}^i\phi \tag{7.2}$$

$$T_W[W] = \tfrac{1}{4}\int \mathrm{d}^D x G_a{}^{ij} G_{aij} \tag{7.3}$$

$$V[\phi] = \int \mathrm{d}^D x U(\phi). \tag{7.4}$$

It is assumed that $F$ and $U$ are positive functions of $\phi$ involving no derivatives. Note that the terms involving derivatives have the conventional quadratic form.

Since the integrands are non-negative, each of these terms must converge separately for a finite-energy solution. Under the scale transformation:

$$\phi(\boldsymbol{x}) \to \phi_\lambda(\boldsymbol{x}) = \phi(\lambda \boldsymbol{x}) \tag{7.5(a)}$$

$$W(\boldsymbol{x}) \to W_\lambda(\boldsymbol{x}) = \lambda W(\lambda \boldsymbol{x}) \tag{7.5(b)}$$

we find that $\mathscr{D}_\mu\phi(\boldsymbol{x}) \to \lambda\, \mathscr{D}_\mu\phi(\lambda \boldsymbol{x})$, $\mathbf{G}_{\mu\nu}(\boldsymbol{x}) \to \lambda^2 \mathbf{G}_{\mu\nu}(\lambda \boldsymbol{x})$. Consequently:

$$T_\phi[\phi_\lambda, W_\lambda] = \lambda^{2-D} T_\phi[\phi, W] \tag{7.6(a)}$$

$$T_W[W_\lambda] = \lambda^{4-D} T_W[W] \tag{7.6(b)}$$

$$V[\phi_\lambda] = \lambda^{-D} V[\phi]. \tag{7.6(c)}$$

For a static solution $H$ must be stationary with respect to arbitrary field variations and therefore, in particular, the scale transformation of equations (7.5). This will be impossible if all the terms in equation (7.1) increase or, equally, if they all decrease. The few possibilities remaining to us are illustrated graphically by the table where $\uparrow$ indicates that the term increases as $\lambda$ increases, $\downarrow$ that it decreases and 0 that it is independent of $\lambda$.

| $D$ | $T_\phi$ | $T_W$ | $V$ |
|---|---|---|---|
| 1 | $\uparrow$ | $\uparrow$ | $\downarrow$ |
| 2 | 0 | $\uparrow$ | $\downarrow$ |
| 3 | $\downarrow$ | $\uparrow$ | $\downarrow$ |
| 4 | $\downarrow$ | 0 | $\downarrow$ |
| $D \geqslant 5$ | $\downarrow$ | $\downarrow$ | $\downarrow$ |

Thus for $D=1$, $V$ must be present; the Sine-Gordon theory illustrates the sufficiency of this. For $D=2$ either all terms must be present or $T_\phi$ alone; the first of these possibilities is a Higgs model leading to vortex lines (Nielsen and Olesen 1973) and the second we shall discuss below. For $D=3$ all three terms must be present, yielding a Higgs model containing the monopole solutions discussed in §§4–6. For $D=4$ the only possibility is a pure gauge theory leading to the so-called instanton solutions. For $D \geqslant 5$ there are no possibilities within this class of theories.

The second possibility for $D=2$ is for the scalar field to realise the symmetry group $G$ in a non-linear way, its values lying on a manifold $\mathscr{M}$ which may be viewed as a coset space $G/H$ (under suitable assumptions) where $H$ is the subgroup of $G$ realised linearly (Coleman *et al* 1969, Callan *et al* 1969, Isham 1969).

What are the possible loopholes in this analysis? First one can seek time-dependent solutions. These are excluded for scalar fields when $D=3$ by the argument of §3.4 if we insist on non-trivial boundary conditions. But it is possible to accept trivial

boundary conditions using conventional Noether conservation laws to ensure stability (Lee 1976).

Secondly we could add positive terms with higher powers of derivatives (Skyrme 1961a, Faddeev 1976b), for example a term $T_\phi^{(4)}$ containing fourth powers of derivatives of $\phi$. Then:

$$T_\phi^{(4)}[\phi_\lambda] = \lambda^{4-D} T_\phi^{(4)}[\phi]$$

so that it behaves like $T_W$ in the table and can counterbalance $T_\phi$ or $V$ for $D=3$.

We shall now discuss briefly the structure of the topological quantum numbers associated with these possibilities. The Higgs model possibilities in $D=2, 3$ are classified by $\Pi_{D-1}(G/H)$. For the non-linear scalar field theory it follows from the finiteness of $T_\phi$ by the argument of §3.4 that:

$$|\nabla\phi| = O(r^{-D/2}).$$

So for $D \geqslant 2$, $\phi$ tends to a constant at spatial infinity which is independent of direction. So we may add one point to $D$-dimensional space, $\mathbb{R}^D$, namely the point at infinity, to compactify it, obtaining $S^D$. The scalar field then provides a map $S^D \to \mathcal{M}$, the manifold in which it takes values, and the topological quantum numbers are classified by $\Pi_D(\mathcal{M})$. Since:

$$\Pi_m(S^n) = \mathbb{Z}\delta_{mn} \qquad m \leqslant n$$

the simplest possibilities are to take $\mathcal{M} = S^2 = SO(3)/O(2)$, for $D=2$ and $\mathcal{M} = S^3 = SU(2) = SO(4)/SO(3)$ for $D=3$, which is indeed the non-linear model $\sigma$ considered by Skyrme (1961a).

For the pure gauge theories in $D=4$ the finiteness of $T_W$ requires $r^2 G_{ij} \to 0$ as $r \to \infty$. This implies that we may use the tangential component of:

$$W^i = \frac{i}{e}(\partial_i g)\, g^{-1}$$

for large $r$ to define $g(\hat{r}) \in G$ depending on the direction $\hat{r} \in S^3$. This provides a continuous map $S^3 \to G$ and so a topological quantum number in $\Pi_3(G)$. (Strictly speaking, this argument must be supplemented by an assumption of angular uniformity.)

Notice that this analysis has revealed two distinct types of topological conservation law (Faddeev 1976a). In the cases where gauge fields play a role the topological structure depends on non-trivial asymptotic behaviour leading to an element of a group $\Pi_{D-1}(X)$ defined by a map from the sphere at infinity, $S^{D-1}$. In the scalar field cases the topological structure depends on trivial asymptotic behaviour, enabling the space to be compactified from $\mathbb{R}^D$ to $S^D$ and an element of a group $\Pi_D(X)$ be defined by the values of the field throughout space. It should be remembered that even where a topological quantum number exists, the quantum-mechanical stability of the objects bearing it has not yet been established.

### *7.2. The quantum theory of monopoles*

Quantum field theories of Dirac monopoles were discussed some years ago (Schwinger 1966a, b, c, 1975, Zwanziger 1965) but the new developments reviewed in this article add extra ingredients to the discussion. We now have a classical picture of a monopole as having an internal structure and a finite mass (see §4). Could it be that renormalisation is finite or even unnecessary in such theories? This would

resolve the dilemma as to whether the Dirac condition should apply to the bare or the renormalised charge.

At present there is no answer to such questions. The work done so far is of a rather preliminary nature and we shall not try to review it in detail. The main techniques applied have been to consider quantum fluctuations about the 't Hooft–Polyakov classical solution (Christ *et al* 1976, Hasenfratz and Ross 1976) or have been semiclassical (Tomboulis and Woo 1976). These methods have been applied to the Sine-Gordon soliton (see Neveu (1977) for a review) but deny us the most important quantum-mechanical insight: the equivalence to the Thirring model, with the Thirring field creating the solitons (see §3).

This provokes the question as to whether there is an analogous result for monopoles. By this we mean an alternative but equivalent version of the field theory involving fields which create magnetic monopoles, with the electrically charged particles occurring as solitons. It seems likely that the local fields describing the magnetic monopoles would be non-local with respect to the local fields describing the electrically charged particles (that is, the two sorts of field would not commute at space-like separation) just as the Sine-Gordon and Thirring field operators are. The expectation of an analogy between the Sine-Gordon theory in two-dimensional space–time and the 't Hooft–Polyakov theory in four-dimensional space–time is enhanced by various points of similarity: the structure of topological quantum numbers in the two theories; the topological current associated with them as given by equations (3.8) and (4.58); the mass formulae of equations (3.4) and (4.42).

Montonen and Olive (1977) have speculated that in the Bogomolny–Prasad–Sommerfield case, discussed in §4.7, the dual field theory describing the monopoles is formally the same as the original one but with the roles of electricity and magnetism reversed. The Lagrangian would be a Georgi and Glashow (1972) one but with coupling constant being $g/\hbar$. The magnetic monopoles would be created by the charged gauge fields. The topological and Noether quantum numbers would have exchanged roles. These statements are as yet conjectural but there is some circumstantial evidence in favour. The symmetric mass formulae:

$$M_Q = |q|\, a \qquad M_g = |g|\, a$$

support the idea that the charged gauge particles may be solitons in a dual version of the theory. Further evidence is supplied by the forces between monopoles calculated by Manton (1977) which (asymptotically) cancel for like monopoles and behave like $g^2/4\pi r^2$ for unlike monopoles. This is precisely analogous to the forces between the charged particles because they can exchange both photons and massless Higgs particles to interact at large distance. The latter contribution is always attractive and either cancels or doubles the photon contribution.

For a more general exact symmetry group $H$ there may also be two dual formulations of the same quantum field theory but they will not in general be formally identical as in the $H = \mathrm{U}(1)$ case just described. Goddard *et al* (1976) have conjectured that the dual group $H^{\mathrm{v}}$ (see §5.8) plays a role in classifying monopole solutions when $H$ is the gauge group. The dual field theory would presumably have $H^{\mathrm{v}}$ as an exact gauge symmetry. Proof of such statements again involves deep questions in quantum field theory but there is some evidence in favour based on the generalised quantisation condition for magnetic charge.

These ideas have physical interest because the non-Abelian gauge theories play a role in both strong and weak interactions. A dual relationship between these respec-

tive theories would provide a newer and deeper concept of unification which would automatically explain the disparate strengths and the relative parity violation. It would put on a modern footing the old idea that nuclear matter is composed of monopoles (Schwinger 1968, 1969, Faddeev 1975).

To conclude, if these speculations are anywhere near correct we are only on the threshold of a development of a new and exciting theory.

## Acknowledgments

This review has grown out of lectures given by each of us in a number of different places (PG in Southampton, Cambridge and Salamanca, DIO in the Niels Bohr Institute, Copenhagen, and CERN, Geneva). We should like to thank our audiences for their help and interest. We have also gained much from discussions with many of our colleagues, particularly Edward Corrigan.

## Appendix 1. Aspects of homotopy theory

In this appendix we try to summarise some relevant results from homotopy theory (see, for example, Hilton 1953, Steenrod 1951). At various points of the discussion from §4.4 onwards we have seen the importance of the homotopy classes of maps, from the $n$-dimensional sphere $S^n$ to some topological space $Y$, which we have denoted by $\tilde{\Pi}_n(Y)$. As we remarked in §5.4 it is more convenient mathematically to restrict attention to maps, $S^n \to Y$, and homotopies, which leave fixed some particular point of $S^n$. For this purpose it is helpful (and permissible since we are concerned with equivalence under continuous deformation) to represent $S^n$ as the unit cube in $\mathbb{R}^n$:

$$I^n = \{\boldsymbol{x} = (x_1, \ldots, x_n) : 0 \leqslant x_r \leqslant 1 \text{ for } 1 \leqslant r \leqslant n\} \tag{A1.1}$$

with its boundary

$$\partial I^n = \{\boldsymbol{x} \in I^n : x_r = 0 \text{ or } 1 \text{ for some } r\} \tag{A1.2}$$

identified as a single point. We consider maps $\phi : I^n \to Y$ satisfying:

$$\phi(\boldsymbol{x}) = y_0 \qquad \text{for all} \qquad \boldsymbol{x} \in \partial I^n. \tag{A1.3}$$

Two such maps $\phi_1, \phi_2$ are said to be homotopic, for a given fixed point $y_0$, if there exists a continuous map $\Phi : I^n \times [0, 1] \to Y$ such that $\Phi(\boldsymbol{x}, t) = y_0$ for all $\boldsymbol{x} \in \partial I^n$ and $\Phi(\boldsymbol{x}, 0) = \phi_1(\boldsymbol{x})$, $\Phi(\boldsymbol{x}, 1) = \phi_2(\boldsymbol{x})$. This defines an equivalence relation on the set of maps $I^n \to Y$ satisfying (A1.3) and the equivalence classes so defined are denoted by $\Pi_n(Y, y_0)$. $\Pi_n(Y, y_0)$ can be given a natural group structure as follows: given two maps $\phi_1, \phi_2 : I^n \to Y$ satisfying (A1.3) we may define a third $\phi_{12}$, by:

$$\begin{aligned} \phi_{12}(x_1, x_2, \ldots, x_n) &= \phi_1(2x_1, x_2, \ldots, x_n) && 0 \leqslant x_1 \leqslant \tfrac{1}{2} \\ &= \phi_2(2x_1 - 1, x_2, \ldots, x_n) && \tfrac{1}{2} \leqslant x_1 \leqslant 1. \end{aligned}$$

It is straightforward to show that the homotopy class $[\phi_{12}]$ of $\phi_{12}$ depends only on the homotopy classes $[\phi_1]$ and $[\phi_2]$ of $\phi_1$ and $\phi_2$ respectively. Thus we may define a binary operation on $\Pi_n(Y, y_0)$ which we shall denote by

$$[\phi_{12}] = [\phi_1] + [\phi_2].$$

Again it is straightforward to verify that this operation satisfies the group axioms. Further if $n \geqslant 2$ we may demonstrate that this group is Abelian:

$$[\phi_1]+[\phi_2]=[\phi_2]+[\phi_1].$$

To do this use the fact that $I^n$ is topologically equivalent (homomorphic) to the unit ball $B^n=\{\boldsymbol{x}: |\boldsymbol{x}| \leqslant 1\}$. We may define a homotopy of $\phi_{12}$ by continuously rotating $B^n$ through $\pi$ about the second axis $n \geqslant 2$. After noting that we may similarly deform $\phi_1$ and $\phi_2$ this is sufficient to show $[\phi_{12}]=[\phi_{21}]$. In general $\Pi_1(Y, y_0)$ is not Abelian, though it necessarily is if $Y$ is a group.

The dependence of $\Pi_n(Y, y_0)$ on the fixed point $y_0$, is superficial if $Y$ is connected. Let $\rho:[0, 1] \to Y$ be a continuous path joining $y_0$ to $y_1$: $\rho(0)=y_0$, $\rho(1)=y_1$. Given $\phi_0: I^n \to Y$ satisfying (A1.3) we define $\phi_1: I^n \to Y$ by exploiting the equivalence of $I^n$ and $B^n$. If $\tilde{\phi}_0: B^n \to Y$ is the map corresponding to $\phi_0: I^n \to Y$ (under some particular homeomorphism of $I^n$ onto $B^n$) define $\tilde{\phi}_1: B^n \to Y$ by:

$$\begin{aligned} \tilde{\phi}_1(\boldsymbol{x}) &= \tilde{\phi}_0(2\boldsymbol{x}) && \text{if} && |\boldsymbol{x}| \leqslant \tfrac{1}{2} \\ &= \rho(2|\boldsymbol{x}|-1) && \text{if} && 1 \geqslant |\boldsymbol{x}| \geqslant \tfrac{1}{2}. \end{aligned}$$

The corresponding map $\phi_1: I^n \to Y$ satisfies $\phi_1(\boldsymbol{x})=y_1$ for all $\boldsymbol{x} \in \partial I^n$ and so defines an element $[\phi_1] \in \Pi_n(Y, y_1)$. One can show that $[\phi_1]$ depends only on $[\phi_0]$. In this way we obtain a map $\rho^*: \Pi_n(Y, y_0) \to \Pi_n(Y, y_1)$ and it is straightforward to demonstrate that this is a group isomorphism. The structure of $\Pi_n(Y, y_0)$ is thus independent of $y_0$ if $Y$ is connected (i.e. when such a path $\rho$ joining two points always exists). In general, we shall assume this henceforth and write $\Pi_n(Y, y_0) \equiv \Pi_n(Y)$ when the base point, $y_0$, plays no role. A corollary of this result is that a closed path, $\sigma$, beginning and ending at $y_0$ defines an automorphism of $\Pi_n(Y, y_0)$. Since the map $\rho^*$ depends only on the homotopy class of $\rho$ (with fixed end points) for a closed path $\sigma$, the automorphism $\sigma^*$ depends only on $[\sigma] \in \Pi_1(Y, y_0)$. Thus $\Pi_1(Y, y_0)$ acts as a group of automorphisms of $\Pi_n(Y, y_0)$.

Another situation in which $\Pi_n(Y, y_0)$ is independent of $y_0$ is when $Y$ is a group, not necessarily connected. Given a map $\phi_0$ satisfying (A1.3), $\phi_1(\boldsymbol{x})=y_1^{-1}y_0\phi(\boldsymbol{x})$ satisfies $\phi_1(\boldsymbol{x})=y_1$ for all $\boldsymbol{x} \in \partial I^n$. This procedure provides us with an isomorphism of $\Pi_n(Y, y_0)$ onto $\Pi_n(Y, y_1)$ since it may also be used to shift the fixed point of an homotopy from $y_0$ to $y_1$. In all our applications $Y$ will be either connected or a group.

As we have mentioned, *a priori* physically it is $\tilde{\Pi}_n(Y)$ that is of interest. Given any map $S^n \to Y$ we may regard it as a map $I^n \to Y$ satisfying:

$$\phi(\boldsymbol{x})=y \qquad \text{for all} \qquad \boldsymbol{x} \in \partial I^n$$

for some fixed $y \in Y$. Thus we may associate with $\phi$ an element of $\Pi_n(Y, y)$. By choosing arbitrarily a path, $\rho$, from $y$ to $y_0$ we may use $\rho^*$ to obtain an element of $\Pi_n(Y, y_0)$ for a definite $y_0$ fixed independently of $\phi$. This procedure depends only on the homotopy class of $\rho$. Thus the end result is ambiguous up to the action of $\Pi_1(Y, y_0)$ on $\Pi_n(Y, y_0)$. That is, if we divide $\Pi_n(Y)$ into equivalence classes (orbits) under the action of $\Pi_1(Y)$ we may unambiguously associate one of these classes with $\phi$. This argument may be developed to show that the classes so obtained are in one-one correspondence with the homotopy classes of $\phi$ without fixed points. The conclusion is that $\tilde{\Pi}_n(Y)$ can be identified with the collection of orbits of $\Pi_n(Y)$ under

$\Pi_1(Y)$. In particular, if $\Pi_1(Y)$ is trivial, i.e. $Y$ is simply connected, the distinction between $\Pi_n(Y)$ and $\tilde{\Pi}_n(Y)$ disappears:

$$\tilde{\Pi}_n(Y) = \Pi_n(Y) \qquad \text{if} \qquad \Pi_1(Y) = 0. \tag{A1.4}$$

It is clear that when (A1.4) holds the combination law on $\tilde{\Pi}_n(Y)$ induced by the group operation on $\Pi_n(Y)$ is the appropriate one for combining monopole topological quantum numbers in the sort of situation indicated in figure 4. In this case we are interested in the group structure of $\Pi_n(Y)$ as well as the nature of its elements. The cases in which (A1.4) fails are rather more obscure and are discussed in appendix 2.

Our analysis from §5.5 forward depends on the isomorphism theorems which exist for $\Pi_n(Y)$ when $Y$ is a coset space, $G/H$, and we shall now summarise these. In this situation $G$ acts transitively on $Y$ and $H$ is the little group of some (arbitrary) fixed $y_0 \in Y$. Denoting the action of $g \in G$ on $Y$ by $y \to gy$, $g \to \pi(g) = gy_0$ defines a projection $\pi: G \to Y$. The theorems are based on an attempt to *lift* a map $\phi: S^n \to Y$ to a map $\gamma: S^n \to G$ such that $\phi = \pi\gamma$. The usefulness of such an attempt is illustrated by considering $\phi: S^2 \to Y$. If this can be lifted to $\gamma: S^2 \to G$ such that $\phi = \pi\gamma$, $\phi$ must be homotopically trivial because, by a theorem of Cartan, $\gamma$ is homotopically trivial, so that there exists a homotopy $\Gamma: S^2 \times [0, 1] \to G$, of $\gamma$ onto a constant map $S^2 \to G$. $\Phi = \pi\Gamma$ then provides a similar homotopy for $\phi$. Thus if we could always lift any $\phi: S^2 \to Y$ we would deduce $\Pi_2(Y) = 0$ as well as $\Pi_2(G) = 0$. On the other hand, if we replace $S^n$ by $I^n$ we can find such a lift $\gamma: I^n \to G$ provided that we relax the equivalent of (A1.3) for $\gamma$. This statement is given in the following result which we merely quote.

*Lemma*: If $H$ is a closed subgroup of a Lie group $G$ and $\phi$ a continuous map $I^n \to Y = G/H$ satisfying (A1.3) then there exists a continuous map $\gamma: I^n \to G$ satisfying $\pi\gamma = \phi$ and

$$\gamma(\boldsymbol{x}) = 1 \qquad \text{if} \qquad \boldsymbol{x} \in \partial I^n \qquad \text{and} \qquad x_n \neq 1. \tag{A1.5}$$

(If the restriction $x_n \neq 1$ were omitted from (A1.5) we could have obtained the lift from $\phi: S^n \to G/H$ to $\gamma: S^n \to G$ which does not exist in general.) This lemma is a special case of a more general result reflecting the fact that $G/H$ is a 'fibre space' (Hilton 1953, p47). The restriction of $\gamma$ to $x_n = 1$ measures the extent to which we are unable to lift $\phi: S^2 \to Y$ to a map $S^2 \to G$. We will denote this restriction by $\hat{\gamma}$:

$$\hat{\gamma}(x_1, x_2, \ldots, x_{n-1}) = \gamma(x_1, x_2, \ldots, x_{n-1}, 1).$$

Since $\hat{\gamma}(\hat{\boldsymbol{x}})y_0 = y_0$ it defines a map $I^{n-1} \to H$ satisfying $\hat{\gamma}(\hat{\boldsymbol{x}}) = 1$ for $\hat{\boldsymbol{x}} \in \partial I^{n-1}$, where $\hat{\boldsymbol{x}} = (x_1, x_2, \ldots, x_{n-1})$, the homotopy class of which depends only on $\phi$. For if $\gamma_1$ and $\gamma_2$ satisfy (A1.5) and $\pi\gamma_1 = \pi\gamma_2 = \phi$:

$$\Gamma_{12}(\boldsymbol{x}, t) = \hat{\gamma}_1(\hat{\boldsymbol{x}})\, \gamma_1(\hat{\boldsymbol{x}}, t)^{-1}\, \gamma_2(\hat{\boldsymbol{x}}, t)$$

defines a homotopy between $\hat{\gamma}_1$ and $\hat{\gamma}_2$. So we map from functions $\phi: I^n \to Y$ satisfying (A1.3) to the element $[\hat{\gamma}]$ of $\Pi_{n-1}(G)$. Further $[\hat{\gamma}]$ depends only on the homotopy class $[\phi]$ of $\phi$. For given two homotopic functions $I^n \to Y$, $\phi_1$ and $\phi_2$ and satisfying (A1.3) with homotopy $\Phi: I^n \times [0, 1] \to Y$, by a result similar to the lemma we may lift $\Phi$ to a map $\Gamma: I^n \times [0, 1] \to G$ satisfying $\Gamma(\boldsymbol{x}, t) = 1$ if $\boldsymbol{x} \in \partial I^n$ and $x_n \neq 1$. In this way we obtain maps $\gamma_1(\boldsymbol{x}) = \Gamma(\boldsymbol{x}, 0)$ and $\gamma_2(\boldsymbol{x}) = \Gamma(\boldsymbol{x}, 1)$ satisfying (A1.5) and $\pi\gamma_1 = \phi_1$, $\pi\gamma_2 = \phi_2$, and a homotopy $\Gamma(\hat{\boldsymbol{x}}, 1, t) \in H$ between $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Thus the elements of $\Pi_{n-1}(H)$ defined by $\phi_1$ and $\phi_2$ are equal.

So we see that the lemma enables us to define a map from $\Pi_n(G/H)$ to $\Pi_{n-1}(H)$

for $n \geqslant 2$. It is straightforward to show that this map is a homomorphism, $f$ say. The image of $\Pi_n(G/H)$ under this homomorphism is not necessarily the whole of $\Pi_{n-1}(H)$. For any $[\hat{\gamma}] \in \Pi_{n-1}(H)$ we obtain in this way corresponds to a map $\hat{\gamma}: I^{n-1} \to H$ which is homotopic to a constant map in $G$ using the homotopy $\gamma(\hat{\boldsymbol{x}}, 1-t)$, noting that $\gamma(\hat{\boldsymbol{x}}, 0)=1$. Conversely, given a $\delta: I^{n-1} \to H$ satisfying $\delta(\hat{\boldsymbol{x}})=1$ for $\hat{\boldsymbol{x}} \in \partial I^{n-1}$ and homotopic to a constant in $G$, there exists a map $\Delta: I^{n-1} \times [0, 1] \to G$ such that $\Delta(\hat{\boldsymbol{x}}, t)=1$ for $\hat{\boldsymbol{x}} \in \partial I^{n-1}$ and $\Delta(\hat{\boldsymbol{x}}, 0)=1$. Then if $\phi(\boldsymbol{x})=\Delta(\hat{\boldsymbol{x}}, x_n)\, y_0$, $[\phi]$ is mapped onto $[\delta]$ under the homomorphism $f: \Pi_n(G/H) \to \Pi_{n-1}(H)$. We conclude that the image of $\Pi_n(G/H)$ under this homomorphism is precisely the subgroup of $\Pi_{n-1}(H)$ which corresponds to maps $S^{n-1} \to H$ which are trivial in $G$. Indeed we may define a homomorphism $i^*: \Pi_{n-1}(H) \to \Pi_{n-1}(G)$ by associating with each map $\phi: S^{n-1} \to H$, or rather its homotopy class in $\Pi_{n-1}(H)$, its homotopy class as a map $S^{n-1} \to G$. Then we have established that the image of $f$ is the kernel of $i^*$.

Further, $f$ is not necessarily one to one. For example, if $[\phi] \in \Pi_n(G/H)$ is mapped to zero in $\Pi_{n-1}(H)$, we have that $\hat{\gamma}: I^{n-1} \to H$, which satisfies $\hat{\gamma}(\hat{\boldsymbol{x}})=1$ if $\hat{\boldsymbol{x}} \in \partial I^{n-1}$, is homotopic in $H$ to a constant map. So there exists a map $\eta: I^{n-1} \times [0, 1] \to H$ satisfying $\eta(\hat{\boldsymbol{x}}, t)=1$ if $\hat{\boldsymbol{x}} \in \partial I^{n-1}$, $\eta(\hat{\boldsymbol{x}}, 0)=\hat{\gamma}(\hat{\boldsymbol{x}})$ and $\eta(\hat{\boldsymbol{x}}, 1)=1$. Then:

$$\gamma_1(\boldsymbol{x})=\gamma(\boldsymbol{x})\, \eta(\hat{\boldsymbol{x}}, x_n)^{-1}$$

defines a map $\gamma_1: I^n \to G$ satisfying $\gamma_1(\boldsymbol{x})=1$ if $\boldsymbol{x} \in \partial I^n$ and $\pi\gamma_1=\phi$. Thus the kernel of $f$ consists of precisely those maps which can be lifted to maps $S^n \to G$. Put another way the projection $\pi: G \to G/H$ associates with each map $S^n \to G$ a map $S^n \to G/H$ and this association induces a map $\pi^*: \Pi_n(G) \to \Pi_n(G/H)$ whose image is the kernel of $f$.

We have obtained a sequence of maps:

$$\Pi_n(H) \xrightarrow{i^*} \Pi_n(G) \xrightarrow{\pi^*} \Pi_n(G/H) \xrightarrow{f} \Pi_{n-1}(H) \xrightarrow{i^*} \Pi_{n-1}(G) \xrightarrow{\pi^*} \ldots$$

$$\ldots \xrightarrow{i^*} \Pi_2(G) \xrightarrow{\pi^*} \Pi_2(G/H) \xrightarrow{f} \Pi_1(H) \xrightarrow{i^*} \Pi_1(G) \xrightarrow{\pi^*} \Pi_1(G/H)$$

where each map in the sequence has as its kernel the image of the previous map. Such a sequence of homomorphisms is called an exact sequence. Since $\Pi_2(G)=0$ we obtain equation (5.47) if $\Pi_1(G)=0$ and equation (5.48) otherwise.

We may extend the sequence one stage further, if (A1.4) fails, by using the lemma to associate $\gamma(1)$ or rather its path component in $H$, with $\phi: I^1 \to G/H$. This path component only depends on $[\phi] \in \Pi_1(G/H)$. In this way we obtain a homomorphism $\Pi_1(G/H) \to \Pi_0(H)$ the group of path components of $H$, leading to equation (5.46) if equations (5.45) hold.

## Appendix 2. Disconnected exact symmetry groups

In §5.4 and appendix 1 we mentioned that the physically relevant $\tilde{\Pi}_2(\mathscr{M}_0)$ coincides with the second homotopy group $\Pi_2(\mathscr{M}_0)$ if $\mathscr{M}_0$ is simply connected but not, in general, otherwise. If $\Pi_1(\mathscr{M}_0) \neq 0$, it is necessary to consider the action of $\Pi_1(\mathscr{M}_0)$ on $\Pi_2(\mathscr{M}_0)$ and divide the latter into equivalent classes under this action, two elements of $\Pi_2(\mathscr{M}_0)$ being in the same equivalence class if they are related by an element of $\Pi_1(\mathscr{M}_0)$. Taking $\mathscr{M}_0=G/H$ as usual, with $G$ connected and simply connected we have the

isomorphisms (5.46) and (5.47), given by the exact sequence of appendix 1:

$$\Pi_2(\mathscr{M}_0) \simeq \Pi_1(H) \qquad \Pi_1(\mathscr{M}_0) \simeq \Pi_0(H).$$

The action of $\Pi_1(\mathscr{M}_0)$ on $\Pi_2(\mathscr{M}_0)$ can be translated into an action of $\Pi_0(H)$ on $\Pi_1(H)$. Given a closed path $h(s)$, $0 \leqslant s \leqslant 1$, in $H$ (with $h(0) = h(1) = 1$) and any $h_0 \in H$ we may construct a new closed path $h_0 h(s) h_0^{-1}$. This path will be homotopic to $h_1 h(s) h_1^{-1}$ if $h_0$ is connected to $h_1$ by a continuous path. Further the homotopy class of $h_0 h(s) h_0^{-1}$ depends only on that of $h(s)$. Indeed, $h_0$ defines an automorphism of $\Pi_1(H)$ which depends only on the path component of $h_0$ in $H$. It is not difficult to show that this is the automorphism corresponding to the action on $\Pi_2(\mathscr{M}_0)$ of the element of $\Pi_1(\mathscr{M}_0)$ corresponding to $h_0$.

We will illustrate this with an example (Goldstone 1976). Consider a theory in which the Higgs field is a three-dimensional complex vector $\phi = (\phi_1, \phi_2, \phi_3)$ and the potential function, $V$, has the symmetry group $G_0 = \mathrm{SO}(3) \times \mathrm{U}(1)$ whose action is:

$$\phi \rightarrow \exp(\mathrm{i}\alpha)\, R\phi \qquad (R, \exp(\mathrm{i}\alpha)) \in \mathrm{SO}(3) \times \mathrm{U}(1). \tag{A2.1}$$

In order to come within the framework of the above analysis we replace $G_0$ by its universal covering group $G = \tilde{G}_0 \simeq \mathrm{SU}(2) \times \mathbb{R}$. The action of $(u, \alpha) \in G$ is given by (A2.1) with $u \rightarrow R \equiv R(u)$ being the usual homomorphism of SU(2) onto SO(3). A possible form for $V$ is

$$V(\phi) = \tfrac{1}{2}\lambda\{(\phi^{\mathrm{T}}\phi)(\phi^{\dagger}\phi^{*}) - 2a^2\phi^{\dagger}\phi + a^4\}.$$

The set of minima, $\mathscr{M}_0$, of $V$ consist of those $\phi$ for which $\phi^{\dagger}\phi = a^2$ and $\phi^{*} = \exp(\mathrm{i}\beta)\,\phi$ for some $\beta \in \mathbb{R}$. It is easy to see that $G$ acts transitively on $\mathscr{M}_0$ and, consequently, $\mathscr{M}_0 = G/H$ where $H$ is the little group of $\phi_0 = a(1, 0, 0)$. The condition for $(u, \alpha) \in H$ is:

$$R(u)\,\phi_0 = \exp(-\mathrm{i}\alpha)\,\phi_0$$

which means that

$$H = \{(\exp\{\mathrm{i}\sigma_3\theta\}, 2\pi n), (\exp\{\mathrm{i}\sigma_3\theta\} \exp\{\mathrm{i}\sigma_2\pi/2\}, 2\pi n + \pi) : \theta \in \mathbb{R}, n \in \mathbb{Z}\}.$$

So we see that the connected part of $H$ is isomorphic to U(1), implying $\Pi_1(H) \simeq \mathbb{Z}$ and also $\Pi_0(H) \simeq \mathbb{Z}$. To calculate the action of $\Pi_0(H)$ on $\Pi_1(H)$ take representative points $h_n = (1, \pi n)$, $n$ even, and $(\exp\{\mathrm{i}\sigma_2\pi/2\}, \pi n)$, $n$ odd, from the components of $H$ and representative paths $\rho_m : \exp(2\pi\mathrm{i}\sigma_3 s m)$, $0 \leqslant s \leqslant 1$, $m \in \mathbb{Z}$, from the classes of $\Pi_1(H)$. Since:

$$\exp(\mathrm{i}\pi\sigma_2/2) \exp(2\pi\mathrm{i}m\sigma_3 s) \exp(-\mathrm{i}\pi\sigma_2/2) = \exp(-2\pi\mathrm{i}m\sigma_3 s)$$

the action of $h_n$ on $\rho_m$ is to take it to $\rho_{-m}$ if $n$ is odd but leave it unchanged if $n$ is even. This shows that whilst we can identify $\Pi_2(\mathscr{M}_0)$ with $\Pi_1(H) \equiv \{[\rho_m] : m \in \mathbb{Z}\}$ to obtain $\tilde{\Pi}_2(\mathscr{M}_0)$ we must also regard $[\rho_m]$ as equivalent to $[\rho_{-m}]$. Only the magnitude of $m$ is physically significant. (A somewhat similar example is given by Coleman (1975b).)

When $\Pi_1(\mathscr{M}_0) = 0$, the group operation on $\Pi_2(\mathscr{M}_0) = \tilde{\Pi}_2(\mathscr{M}_0)$ corresponds to the physical operation of combining the quantum numbers of distant monopoles. When this condition fails we no longer have a natural binary operation on $\tilde{\Pi}_2(\mathscr{M}_0)$ but we can regard its elements as subsets of $\Pi_2(\mathscr{M}_0) \simeq \Pi_1(H)$, which are disjoint and exhaustive, the orbits of $\Pi_1(H)$ under $\Pi_0(H)$. Given two such orbits, $C_1$ and $C_2$:

$$C_1 + C_2 = \{c_1 + c_2 : c_1 \in C_1, c_2 \in C_2\}$$

is a union of orbits. The topological quantum number of a monopole obtained by combining $C_1$ and $C_2$ will be that corresponding to one of the orbits $C \subset C_1 + C_2$. There is an ambiguity: $C_1$ and $C_2$ alone are insufficient to determine $C$, although they limit the possibilities. Specifically, in Goldstone's example discussed above, the elements of $\tilde{\Pi}_2(\mathcal{M}_0)$ may be labelled $C_m = \{[\rho_m], [\rho_{-m}]\}$, $m \geqslant 0$, and

$$C_m + C_n = C_{m+n} \cup C_{|m-n|}.$$

Either $C_{m+n}$ or $C_{|m-n|}$ could result from combining $C_m$ and $C_n$.

The physical interpretation that Coleman (1975a, b) gives to this sort of ambiguity is that whilst the monopoles are in the same homotopy class as the antimonopoles (and indeed are gauge-equivalent to them) in isolation, they reveal their distinct identities in combination. Another interpretation (Goldstone 1976), which seems more appropriate to us, is that the topological characteristics of the individual constituents are not sufficient to determine the topological characteristics of the combined system; one needs to know how they have been put together.

The question of combining or 'patching' solutions in a three-dimensional space–time has been discussed by Schonfield (1977).

## References


Abers E S and Lee B W 1973 *Phys. Rep.* **9C** 1–141

Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485–91

—— 1961 *Phys. Rev.* **123** 1511–24

Amaldi E and Cabibbo N 1972 *Aspects of Quantum Theory* ed Abdus Salam and E P Wigner (Cambridge: Cambridge University Press) pp185–212

Arafune J, Freund P G O and Goebel C J 1975 *J. Math. Phys.* **16** 433–7

Artru X 1977 *Nucl. Phys.* B **129** 415–28

Bais F A and Primack J R 1976 *Phys. Rev.* D **13** 819–29

—— 1977 *Nucl. Phys.* B **123** 253–73

Bitar K M and Sorba 1977 *Phys. Rev.* D **16** 431–7

Bogomolny E B 1976 *Sov. J. Nucl. Phys.* **24** 449–54

Bogomolny E B and Marinov M S 1976 *Sov. J. Nucl. Phys.* **23** 355–7

Boulware D G, Brown L S, Cahn R N, Ellis S D and Lee C 1976 *Phys. Rev.* D **14** 2708–27

Brandt R A and Primack J R 1977a *Phys. Rev.* D **15** 1175–7

—— 1977b *Phys. Rev.* D **15** 1798–802

Brihaye Y and Nuyts J 1977 *J. Math. Phys.* **18** 2177–90

Callan C G, Coleman S, Wess J and Zumino B 1969 *Phys. Rev.* **177** 2247–50

Callan C G, Dashen R F and Sharp D H 1968 *Phys. Rev.* **165** 1883–6

Chakrabarti A 1975 *Ann. Inst. Henri Poincaré* **23** 235–49

Christ N H 1975 *Phys. Rev. Lett.* **34** 355–8

Christ N H, Guth A H and Weinberg E J 1976 *Nucl. Phys.* B **114** 61–99

Coleman S 1973 *Laws of Hadronic Matter* (*Proc. 1973 Int. School of Physics 'Ettore Majorana'*) ed A Zichichi (New York: Academic) pp139–223

—— 1975a *Phys. Rev.* D **11** 2088–97

—— 1975b *New Phenomena in Subnuclear Physics* (*Proc. 1975 Int. School of Physics 'Ettore Majorana'*) ed A Zichichi (New York: Plenum) pp297–421

Coleman S, Parke S, Neveu A and Sommerfield C M 1977 *Phys. Rev.* D **15** 554

Coleman S, Wess J and Zumino B 1969 *Phys. Rev.* **177** 2239–47

Corrigan E and Olive D 1976 *Nucl. Phys.* B **110** 237–47

Corrigan E, Olive D, Fairlie D B and Nuyts J 1976 *Nucl. Phys.* B **106** 475–92

Cremmer E, Schaposnik F and Scherk J 1976 *Phys. Lett.* **65B** 78–80

Czechowski A 1977 *CERN Preprint* TH 2282

Dell'Antonio G F, Frishman Y and Zwanziger D 1972 *Phys. Rev.* D **6** 988–1007

Derrick G H 1964 *J. Math. Phys.* **5** 1252–4

Dirac P A M 1931 *Proc. R. Soc.* A **133** 60–72

—— 1948 *Phys. Rev.* **74** 817–30
Englert F and Brout R 1964 *Phys. Rev. Lett.* **13** 321–3
Englert F and Windey P 1976 *Phys. Rev.* D **14** 2728–31
Enz U 1963 *Phys. Rev.* **131** 1392–4
Ezawa Z F and Tze H C 1976a *J. Math. Phys.* **17** 2228–31
—— 1976b *Phys. Rev.* D **14** 1006–20
—— 1977 *Phys. Rev.* D **15** 1647–54
Faddeev L D 1975 *JETP Lett.* **21** 64–5
—— 1976a *Nonlocal, nonlinear and nonrenormalisable field theories* (*Proc. Int. Symp., Alushta*) (Dubna: Joint Institute for Nuclear Research) pp207–23
—— 1976b *Lett. Math. Phys.* **1** 289–93
Finkelstein D 1966 *J. Math. Phys.* **7** 1218–25
Finkelstein D and Misner C W 1959 *Ann. Phys., NY* **6** 240–3
Friedberg R, Lee T D and Sirlin A 1976a *Phys. Rev.* D **13** 2739–61
—— 1976b *Nucl. Phys.* B **115** 1–31
—— 1976c *Nucl. Phys.* B **115** 32–47
Gell-Mann M and Ne'eman Y 1964 *The Eightfold Way* (New York: Benjamin)
Georgi H and Glashow S L 1972 *Phys. Rev.* D **6** 2977–82
Gervais J L, Sakita B and Wadia S 1976 *Phys. Lett.* **63B** 55–8
Glashow S and Gell-Man M 1961 *Ann. Phys., NY* **15** 437–60
Goddard P, Nuyts J and Olive D 1977 *Nucl. Phys.* B **125** 1–28
Goldhaber A S 1965 *Phys. Rev.* **140** B1407–14
—— 1976 *Phys. Rev. Lett.* **36** 1122–5
Goldhaber A S and Smith J 1975 *Rep. Prog. Phys.* **38** 731–70
Goldhaber A S and Wilkinson D 1976 *Nucl. Phys.* B **114** 317–33
Goldstone J 1976 *Unpublished Lectures given at Cambridge University*
Guralnik G S, Hagen C R and Kibble T W K 1964 *Phys. Rev. Lett.* **13** 585–87
Gursey F 1976 *Gauge Theories and Modern Field Theories* ed R Arnowitt and P Nath (Cambridge, Mass.: MIT Press) pp369–76
Hasenfratz P and 't Hooft G 1976 *Phys. Rev. Lett.* **36** 1119–22
Hasenfratz P and Ross D A 1976 *Nucl. Phys.* B **108** 462–82
Higgs P W 1964a *Phys. Rev. Lett.* **12** 132–3
—— 1964b *Phys. Rev. Lett.* **13** 508–9
—— 1966 *Phys. Rev.* **145** 1156–63
Hilton P J 1953 *An Introduction to Homotopy Theory* (Cambridge: Cambridge University Press)
't Hooft G 1974 *Nucl. Phys.* B **79** 276–84
—— 1976 *Nucl. Phys.* B **105** 538
Horvath Z and Palla L 1976a *Phys. Rev.* D **14** 1711–4
—— 1976b *Nucl. Phys.* B **116** 500–24
—— 1977 *Phys. Lett.* **69B** 197–201
Hurst C A 1968 *Ann. Phys., NY* **50** 51–75
Isham C J 1969 *Nuovo Cim.* A **61** 188–202
Jackiw R and Rebbi C 1976 *Phys. Rev. Lett.* **36** 1116–9
Julia B and Zee A 1975 *Phys. Rev.* D **11** 2227–32
Kaku M 1976 *Phys. Rev.* D **13** 2881–3
Kibble T W K 1967 *Phys. Rev.* **155** 1557–61
Klimo P and Dowker J S 1973 *Int. J. Theor. Phys.* **8** 409–17
Lee T D 1976 *Phys. Rep.* **23C** 254–8
Lubkin E 1963 *Ann. Phys., NY* **23** 233–83
Madore J 1977 *Commun. Math. Phys.* **56** 115–23
Maison D and Orfanidis S J 1977 *Phys. Rev.* D **15** 3608–11
Mandelstam S 1962 *Ann. Phys., NY* **19** 25–66
—— 1968 *Phys. Rev.* **175** 1580–603
—— 1975 *Phys. Rev.* D **11** 3026–30
Manton N S 1977 *Nucl. Phys.* B **126** 525–41
Marciano W J and Pagels H 1975 *Phys. Rev.* D **12** 1093–5
Michel L, O'Raifeartaigh L and Wali K C 1977a *Phys. Lett.* **67B** 198–202
—— 1977b *Phys. Rev.* D **15** 3641–55
Mollenstedt G and Bayh W 1962 *Naturw.* **49** 81–2

Monastyrskii M I and Perelomov A M 1975 *JETP Lett.* **21** 43–4
Montonen C and Olive D 1977 *Phys. Lett.* **72B** 117–20
Neveu A 1977 *Rep. Prog. Phys.* **40** 709–30
Nielsen H B and Olesen P 1973 *Nucl. Phys.* B **61** 45–61
Olive D 1976 *Nucl. Phys.* B **113** 413–20
O'Raifeartaigh L 1976 *Nuovo Cim. Lett.* **18** 205–8
—— 1979 *Rep. Prog. Phys.* **42** to be published
Patrascioiu A 1975 *Phys. Rev.* D **12** 523–30
Poincaré H 1896 *C.R. Acad. Sci., Paris* **123** 530
Polyakov A M 1974 *JETP Lett.* **20** 194–5
Prasad M K and Sommerfield C M 1975 *Phys. Rev. Lett.* **35** 760–2
Saha M N 1936 *Ind. J. Phys.* **10** 145
—— 1949 *Phys. Rev.* **75** 1968
Salam A 1968 *Proc. 8th Nobel Symp.: Elementary Particle Theory* ed N Svartholm (New York: Wiley) pp337–67
Schiff L I 1966 *Phys. Rev. Lett.* **17** 714–6
—— 1967 *Phys. Rev.* **160** 1257–62
Schonfield J F 1977 *Nucl. Phys.* B **125** 381–403
Schrödinger E 1935 *Proc. R. Soc.* A **150** 465–77
Schwarz A S 1976 *Nucl. Phys.* B **112** 358–64
Schwinger J 1966a *Phys. Rev.* **144** 1087–93
—— 1966b *Phys. Rev.* **151** 1048–54
—— 1966c *Phys. Rev.* **151** 1055–7
—— 1968 *Phys. Rev.* **173** 1536–44
—— 1969 *Science* **165** 757–61
—— 1975 *Phys. Rev.* D **12** 3105–111
—— 1976 *Gauge Theories and Modern Field Theories* ed R Arnowitt and P Nath (Cambridge, Mass.: MIT Press) p363
Scott A C, Chu F Y F and McLaughlin D W 1973 *Proc. IEEE* **61** 1443–83
Shaw R 1955 *PhD Thesis* University of Cambridge
Sinha A 1976 *Phys. Rev.* D **14** 2016–22
Skyrme T H R 1958 *Proc. R. Soc.* A **247** 260–78
—— 1959 *Proc. R. Soc.* A **252** 236–45
—— 1961a *Proc. R. Soc.* A **260** 127–38
—— 1961b *Proc. R. Soc.* A **262** 237–45
Steenrod N E 1951 *The Topology of Fibre Bundles* (Princeton, NJ: Princeton University Press)
Sugawara H 1968 *Phys. Rev.* **170** 1659–62
Taylor J C 1976 *Gauge Theories of Weak Interactions* (Cambridge: Cambridge University Press)
Thirring W 1958 *Ann. Phys., NY* **3** 91–112
Tomboulis E and Woo G 1976 *Nucl. Phys.* B **107** 221–37
Trautman A 1970 *Rep. Math. Phys.* **1** 29–62
Troost W and Vinciarelli P 1976 *CERN Preprint* TH 2246
Tyupkin Yu S, Fateev V A and Shvarts A S 1975 *JETP Lett.* **21** 41–2
Utiyama R 1956 *Phys. Rev.* **101** 1597–602
Weinberg E J and Guth A H 1976 *Phys. Rev.* D **14** 1660–2
Weinberg S 1965 *Phys. Rev.* **138** B988–1002
—— 1967 *Phys. Rev. Lett.* **19** 1264–6
Wilkinson D 1977 *Nucl. Phys.* B **125** 423–44
Wilkinson D and Goldhaber A S 1977 *Phys. Rev.* D **16** 1221–31
Wu A C T and Wu T T 1974 *J. Math. Phys.* **15** 53–5
Wu T T and Yang C N 1969 *Properties of Matter under Unusual Conditions* ed H Mark and S Fernbach (New York: Interscience) pp344–54
—— 1975 *Phys. Rev.* D **12** 3845–57
—— 1976 *Nucl. Phys.* B **107** 365–80
Yang C N 1970 *Phys. Rev.* D **1** 2360
Yang C N and Mills R L 1954 *Phys. Rev.* **96** 191–5
Zumino B 1966 *Strong and Weak Interactions—Present Problems (Proc. 1966 Int. School of Physics 'Ettore Majorana')* ed A Zichichi (New York: Plenum) pp711–30
Zwanziger D 1965 *Phys. Rev.* **137** B647