Classical solutions of $SU(2)$ Yang–Mills theories

Alfred Actor

Abteilung Physik, Universität Dortmund, 4600 Dortmund, Federal Republic of Germany

A comprehensive review of the known classical solutions of $SU(2)$ gauge theories is presented. The author follows the historical development of this subject from its beginning (the first explicit solution found was an imbedded Abelian static Coulomb solution) up to the most recent work in the field (in particular the solutions which represent monopoles, instantons, and merons). As well as being a detailed survey, this article is intended to serve as a self-contained introduction to the subject.

CONTENTS

I. Introduction 461
 A. Quantum Yang–Mills theory 461
 B. Classical Yang–Mills theory 462
 C. Contents of this review 464
II. Some Results in Classical Minkowski Yang–Mills Theory 465
 A. Julia–Zee correspondence 465
 B. Fields at infinity and topology 466
 C. Self-Duality and the Bogomol'ny condition 468
 D. SU(2, C) gauge theory 469
III. Solutions with SU(2) Gauge Invariance 470
 A. Non-Abelian plane wave 470
 B. Ikeda–Miyachi solution 471
 C. Wu–Yang solution 471
 D. Rosen's *ansatz* 472
 E. ϕ^4 *ansatz* 473
 F. Witten's *ansatz* 473
IV. Solutions with U(1) Gauge Invariance 474
 A. The Higgs mechanism in classical and quantum field theory 474
 B. Solutions 476
  1. Treat's *ansatz* 476
  2. Prasad–Sommerfield–Bogomol'ny solution 476
  3. 't Hooft–Polyakov monopole 477
  4. Julia–Zee dyon 480
 C. Remarks 481
V. Solutions with No Residual Gauge Invariance 484
VI. Connection Between Yang–Mills Theory and ϕ^4 Theory 485
 A. *Ansatz* for the Yang–Mills potential 485
 B. de Alfaro–Fubini–Furlan solution 486
 C. Elliptic solutions 488
VII. Euclidean Solutions 489
 A. Introduction 489
  1. Euclidean SU(2) gauge theory 489
  2. Instanton solutions 491
  3. Meron solutions 495
  4. Elliptic solution 497
 B. *Ansätze* 497
  1. 't Hooft–Corrigan–Fairlie–Wilczek *ansatz* 497
  2. Witten's *ansatz* 498
 C. Belavin–Polyakov–Schwartz–Tyupkin instanton 499
 D. N-Instanton solutions 500
  1. 't Hooft's solution 500
  2. Witten's solution 502
  3. Atiyah–Hitchin–Drinfeld–Manin construction 503
 E. Two-Meron solution 504
  1. One Meron at infinity 504
  2. Two-Meron solution 504
 F. Multimeron configurations 506
 G. Elliptic solutions 507
  1. One-Meron solution 507
  2. Two-Meron solution 508
VIII. Conclusion 508
 A. Monopoles and dyons 508
 B. Instantons 509
 C. Merons 509
Acknowledgment 509
Appendix A: SU(2) Yang–Mills Theory 509
Appendix B: Imbedding SU(2) Solutions 511
Appendix C: Conformal Transformations and Yang–Mills Fields 511
Appendix D: The Dirac String 512
Appendix E: Electromagnetic Field Tensor in Yang–Mills Theory 513
Appendix F: Elliptic Functions 514
Appendix G: 't Hooft–Polyakov Differential Equations 515
  1. $g(r)=0$ 515
  2. $g(r)\neq 0$ 516
Appendix H: The Gribov Ambiguity 516
  1. Spherically symmetric fields 517
  2. Vacuum solutions 517
  3. Nonvacuum solutions 518
  4. One-Meron solution 518
  5. Instanton solution 519
  6. Vacuum tunneling in the Coulomb gauge 519
Appendix I. Topological Considerations 520
References 524

I. INTRODUCTION

A. Quantum Yang-Mills theory

Non-Abelian gauge theories were invented by Yang and Mills nearly twenty-five years ago (Yang and Mills, 1954). For most of this period it was not known whether any of the interactions observed in nature can be described by a non-Abelian gauge theory. Nevertheless, the elegance of these theories attracted interest. Quantization and renormalization were the central topics of research. For physically relevant theories with massive gauge bosons it turned out to be quite difficult to demonstrate renormalizability. First, the Higgs mechanism (Higgs, 1964) had to be discovered before one knew how to break the non-Abelian local gauge symmetry without introducing Goldstone bosons. Several years later the first proof ('t Hooft, 1971) was given that a Yang–Mills (hereafter YM) theory stays renormalizable when its local gauge symmetry is broken in this fashion. After renormalizability had been established, so that one had confidence in Feynman diagram calculational techniques, the door to quantitative YM phenomenology was finally open.

Many years earlier there had been attempts to describe physical phenomena in terms of YM theories. These were originally based on the assumption that the YM field could acquire a mass, dynamically, through its self-interaction. If this were true then YM fields would be short range, like the interactions between hadrons. Efforts were made to set up a YM theory of the strong interactions (see, for example, Sakurai, 1960; Schwinger, 1964). However, these theories were not successful: The basic assumption was wrong; YM fields do not acquire a mass through their self-interaction. The local gauge symmetry apparently has to be broken by hand, and this introduces Goldstone bosons unless the Higgs mechanism is used. Higgs-like models of the strong interactions could, perhaps, be devised. But with the widespread belief that hadrons are not "elementary," one nowadays has little interest in such models.

Following the discovery that the charged weak current has $V-A$ structure, so that the similarity between the weak and electromagnetic interactions became apparent, it was natural to try to combine these two theories within a larger gauge theory. Obviously the latter would have to be a non-Abelian gauge theory. Glashow (1961) constructed an $SU(2)\times U(1)$ model along these lines which had many attractive features, but lacked the vital Higgs fields which induce spontaneous gauge symmetry breakdown. These were later introduced by Weinberg (1967) and Salam (1968) into the $SU(2)\times U(1)$ model. The resulting field theory (the Weinberg–Salam model as it is now called) has turned out to be extraordinarily successful. This success has convinced most physicists that non-Abelian gauge theories of the weak and electromagnetic interactions are good physical theories. The coming generation of accelerators is expected to provide final experimental proof of this through the discovery (as particles) of the massive vector bosons that carry the charged and neutral weak interactions.

The many successes of the quark model have made it clear that hadrons are composite. With this realization it became necessary to find an adequate theory of the quark–quark interaction. One particular theory has emerged which seems to have a good chance of success, namely quantum chromodynamics (QCD) (see the review by Marciano and Pagels, 1978). This theory involves only quarks and an SU(3) gauge field. (The gauge symmetry is unbroken, so there are no Higgs fields present.) The quarks carry a charge called color, which by assumption cannot exist outside of a hadron since quarks have never been observed as free particles. The unknown mechanism which keeps the quarks (and also the gauge particles which are called gluons) inside hadrons is called *confinement*. The existence of this unproven property of YM theories is the major open question in QCD. Confinement is an infrared phenomenon, and YM theories are extremely singular in the infrared region. Therefore it is very difficult to do quantitative work on the confinement problem. In the ultraviolet region QCD is in much better shape because of "asymptotic freedom" (Gross and Wilczek, 1973; Politzer, 1973). In QCD, or any other YM theory with a semisimple gauge group, the effective coupling constant gets smaller with decreasing distance. Thus for deep inelastic lepton–hadron scattering one can do perturbation theory calculations. The QCD predictions are in agreement with a large body of experimental information. Asymptotic freedom also means that the effective YM coupling increases with increasing distance. Eventually the perturbation theory estimates on which all asymptotic freedom arguments are based break down, and one cannot follow this increase to large values of the effective coupling constant. However, it is believed by many that the effective coupling becomes strong enough to confine quarks.

The Weinberg–Salam model (or generalizations of it) and QCD are the two existing YM theories of real phenomenological importance. These theories can be formulated in terms of Feynman path integrals, i.e., functional integrals over all classical field configurations weighted by a factor exp(−action). If one knew everything about classical field configurations, then in principle all questions concerning the quantum theory could be answered. Partial information about classical fields might yield, at least, some insight into the quantum theory. This is the basic hope which motivates present research activity in classical YM theory.

B. Classical Yang-Mills theory

For many years there was little activity in the classical sector of YM theory. The first exact solution of the classical equations of motion of the pure SU(2) gauge theory was found by Ikeda and Miyachi (1962). This solution is the electromagnetic Coulomb solution *imbedded* in the larger theory, so it is not truly a YM solution. Loos (1965) observed that the imbedding works for any gauge group. Solutions of this type naturally did not arouse a great deal of interest.

A genuine non-Abelian YM solution was found by Wu and Yang (1968). Rosen (1972) rediscovered this same solution within the context of a general gauge theory. Like the Ikeda–Miyachi solution, the Wu–Yang solution is pointlike, i.e., the gauge potential behaves like $1/r$ everywhere. A very interesting property of the Wu–Yang solution is that it describes a (pointlike) non-Abelian magnetic monopole. The Wu–Yang monopole is not attached to a string (unlike a Dirac monopole). From this one might conclude that YM theories provide a natural setting for magnetic monopole solutions. (Several years later, when the nonsingular 't Hooft–Polyakov monopole was discovered, this became much clearer.)

YM theories admit static solutions more complicated than pointlike ones. The first explicit solution of this type was found by Treat (1967). In addition to an electric Coulomb-like component in the gauge potential, Treat's solution has short-range potentials which behave like e^{-Mr}. The constant M is essentially the value of a component of the gauge potential at spatial infinity. If this value were zero then the short-range potentials would become long range. One has a very similar situation in the nonsingular monopole and dyon solutions discussed below, where the nonvanishing field at infinity is a Higgs field. Treat's solution does not have the attractive physical properties of these later solutions, but it does contain some of their essential ingredients.

Important progress was made by Nielson and Oleson (1973), who introduced a "classical Higgs mechanism" into classical gauge theory. This mechanism is quite analogous to the quantum field-theoretic Higgs mechanism: It causes the classical gauge theory to become "massive" in the sense that certain components of the gauge potential must behave like e^{-Mr} at large r for the energy to be finite. In their SU(2) example, Nielson and Oleson used two Higgs triplets to make all gauge field components decrease exponentially away from an axis. The gauge field is essentially contained within a tube or "vortex." This very original calculation directly led to another discovery.

In 1974 the slow-growth period of classical YM theory came to an end. 't Hooft (1974) and Polyakov (1974) independently discovered a magnetic monopole solution of the SU(2) gauge theory with a Higgs triplet. This solution is nonsingular and has finite energy. It represents an extended, localized object with magnetic charge and topological stability. (The solution is characterized by a mapping $S^2 \to S^2$ of a sphere onto a sphere, covering the latter once. It is certainly the lowest-energy solution in this topological category, and therefore it must be stable.) These interesting properties, together with the fact that the YM theory involved has spontaneous symmetry breakdown (precisely the type of theory believed to be capable of unifying the weak and electromagnetic interactions), naturally attracted great interest. Much additional work was done on solutions of this type. An SU(2) dyon solution was found [this is the SU(2) monopole with an arbitrary electric charge]. Monopole and dyon solutions of theories with larger gauge groups were not difficult to find once the basic ingredients of the SU(2) solution were understood. [This involved the application of some basic results from homotopy group theory (unknown to nearly all physicists in 1974).] "No-go" theorems were established which clarify the extent to which the monopole solutions are unique. These and other developments are discussed in the non-Abelian monopole review by Goddard and Olive (1978). [See also the annotated monopole bibliography for the period 1973–1976 by Carrigan (1977).]

Unfortunately, one very important development did *not* occur. To this date, no one has found a multimonopole solution with acceptable physical properties (e.g., finite energy). Nor is it known whether such a solution exists. This makes one reluctant to accept the 't Hooft–Polyakov monopole as a particlelike object which might conceivably be realized in nature. One would at least like to know that two of them can exist with less than infinite separation. The lack of a multimonopole solution is mainly responsible for the greatly diminished interest in non-Abelian monopoles at present. Another reason has been the lack of a physical use for these solutions. Non-Abelian monopoles are very heavy (probably some thousands of GeV); one does not know what to do with them.

Let us now go over to Euclidean space-time (E^4) to mention a result which has stimulated even more interest than the SU(2) monopole. This is the extraordinary solution of the Euclidean SU(2) gauge theory found by Belavin, Polyakov, Schwartz, and Tyupkin (1975), called the "instanton" or "pseudoparticle." The main properties of the instanton solution are as follows:

(i) it is nonsingular and localized (symmetrically) in all directions in E^4 including the imaginary time axis (hence: instanton);

(ii) it is self-dual (which means that it carries zero energy);

(iii) it is characterized by a topological charge $q=1$ (i.e., characterized by a map $S^3 \to S^3$ which covers the latter sphere one time).

Moreover—and this is very important—exact solutions representing an arbitrary number of instantons have been found (Witten, 1977; 't Hooft, 1977c) with topological charge equal to the number of instantons. Thus one is compelled to accept instantons as "objects" which are present in YM theories and could lead to physical effects. As these objects carry zero energy, they are evidently some kind of localized vacuum fluctuation.

Instantons are solutions in Euclidean space-time, and this means they have something to do with tunneling in Minkowski space in the quantum YM theory. (Imaginary-time solutions of classical theories are usually interpreted as real-time tunneling in the corresponding quantized theory.) The N-instanton solution is a vacuum fluctuation with N units of a topological charge. What can one conclude from this information? Evidently there must exist an infinity of topologically distinct Minkowski vacua $|n\rangle$ with topological charge n. One instanton tunnels from $|n\rangle$ to $|n+1\rangle$, N instantons tunnel to $|n+N\rangle$, an anti-instanton tunnels to $|n-1\rangle$, and so on. This is actually true; the topological vacua $|n\rangle$ are easy to construct classically. Thus the vacuum in a YM theory is a great deal more complicated than one realized a few years ago. It used to be thought that the YM theory is based entirely on the $n=0$ perturbative vacuum (i.e., $W_\mu = 0$). Now we know that the YM vacuum has topologically distinct sectors, and that instantons tunnel between these, removing the degeneracy that would otherwise be present. The true YM vacuum is a superposition of the topological vacua, teeming with instantons. These large vacuum fluctuations lead to qualitatively new physics.

It has not been easy to unravel the physical implications of the new and quite intricate YM vacuum. Some aspects of the problem are partially understood. One important result is that massless fermions cause suppression of instanton tunneling, while massive fermions do not have this effect. Conversely, a massless fermion when placed in an instanton field becomes massive. Therefore, instantons seem to be connected with quark mass generation and chiral symmetry breakdown—one of the deep problems in hadron physics. Another deep problem is quark confinement. Polyakov (1975) was the first to suggest that instantons, through their disordering influence, might be responsible for quark confinement. [More precisely, by restoring the gauge symmetry of the YM vacuum they might be responsible for a phase transition from the massless (nonconfining) to the massive (confining) phase of the YM theory.] This idea has in the meanwhile been abandoned, as semiquantitative estimates indicate that instanton effects are not enough to confine quarks. Nevertheless, the situation with respect to confinement is still very unclear, and instantons may really play a key role. Many other possible physical effects of instantons have also been suggested. We do not want to list these here. The point we do wish to emphasize is that "instanton physics" is a very active and challenging field at present. This will surely continue to be the case for some time to come.

Another type of exact Euclidean YM solution has been much discussed lately. These are the "meron" solutions (de Alfaro, Fubini, and Furlan, 1976, 1977; Glimm and Jaffe, 1978a). A meron is a pointlike concentration of *one-half unit* of topological charge. Because of their singular nature meron solutions have infinite action, which makes their physical relevance somewhat obscure. However, multimeron solutions have been shown to exist (the only known explicit solutions describe two merons or a meron and an antimeron). Therefore, one is inclined to take them seriously. Unlike instanton solutions, meron solutions are not self-dual.

It is known that merons correspond to tunneling between two different vacua in real time. These vacua have topological charges $n=0$ and $n=\frac{1}{2}$, respectively. A short digression is necessary to explain how meron tunneling works.

Gribov (1977) has discovered an ambiguity in the Coulomb gauge formulation of YM theory. He shows that the gauge-fixing conditions $W_0=0, \partial_i W_i=0$ do *not* uniquely fix the potential W_μ. Even the Coulomb gauge vacuum is not unique. In addition to the perturbative vacuum $W_\mu=0$ there are (at least) two others $W_\mu \neq 0$ which are nonsingular pure-gauge potentials. Remarkably enough, these new Gribov vacua are localized; they have a size and position in three space (as if they were some sort of object). Moreover, they carry one-half unit of topological charge. Merons tunnel between the Gribov vacua and the $W_\mu=0$ vacuum, which has zero topological charge. Instantons on the other hand, tunnel from one Gribov vacuum to the other, but do not connect these with the $W_\mu=0$ vacuum. Therefore it seems that merons may be necessary to restore the gauge symmetry of the YM vacuum in the Coulomb gauge. This is a hint that they may play a vital role in the confinement problem.

Callen, Dashen, and Gross (1977, 1978a) have suggested that an instanton consists of two merons, and that instanton dissociation into meron pairs signals a phase transition of the YM theory into the confining phase. Instantons alone, they argue, cannot confine quarks (at least not the dilute gas approximation where semiclassical estimates can be made). But a plasma of logarithmically interacting merons might be able to do this. Such a plasma might generate a linear force between quarks. The arguments which lead to this result are at best semiquantitative, however. Moreover, it is not known whether instanton dissociation into merons is a physically meaningful concept. A good deal more has to be learned before we fully understand this confinement mechanism.

A rather more phenomenological statement of the effects of instantons and merons on quarks can be made (e.g., Callen, Dashen, and Gross, 1978b), namely, that a "bag" is formed which confines quarks by the following mechanism: Normally, the YM vacuum seeths with instantons and merons, for this is the state with lowest energy. (Recall that instantons and merons reduce, and perhaps eliminate, vacuum degeneracy by tunneling. Quantum mechanically this would be expected to lower the energy.) However, these topological objects cannot exist where strong YM fields are present. (This is the crucial observation in the present context. It has to be proven.) Inside hadrons there are certainly strong YM fields (in QCD) because of the quark sources. Therefore instantons and merons are expelled: A hadron is a bubble of three-space in which no instantons and merons "occur." This bubble cannot expand because this costs energy. Moreover, the quarks cannot come apart because such a bubble cannot form about a single quark.

By now the interest in classical YM theory has become so widespread that many workers are involved in the search for new solutions. At any time a completely new and unexpected type of YM solution might be found that could change the direction in which the subject is developing. Up to now (excepting the vortex solution) only three really important types of solution have been found: monopole, instanton, and meron. These have largely determined the subject as it now exists.

C. Contents of this review

In this paper the reader will find (a) a largely complete review of the known classical solutions of SU(2) gauge theories; (b) a discussion (where relevant) of the physical properties of these solutions; and (c) an extensive study of classical YM theory as a whole. To make these statements more specific we mention here some of the interesting work on classical YM theory which has *not* been included:

(i) stringlike or vortex solutions;
(ii) considerations which involve fermions moving in a classical gauge field, e.g., a monopole field;
(iii) the classification of YM fields and potentials;
(iv) the geometrical interpretation of gauge fields;
(v) solutions of theories with gauge groups larger than SU(2).
(vi) solutions with sources.

An adequate treatment of these topics would have doubled the length of this article.

In one respect this article is substantially complete. We shall discuss nearly every known solution of a classical SU(2) YM theory (excepting the vortex solutions) in some detail. The most important solutions (monopole, instanton, and meron) have been mentioned in the preceding subsection. Besides these there are many others, which do not have any obvious physical significance. We have included these for historical reasons or for the sake of completeness. It has been our goal to show clearly how each solution works—i.e., how it satisfies the nonlinear YM equations of motion. If there is more than one way to do this we usually choose the easier. Throughout this article we shall pay close attention to technical details. *Ansätze* for the YM potential will play a very important role. The properties of the most useful *ansätze* are studied in great detail.

Physical applications of classical YM theory begin with exact solutions. The physical properties of monopoles, instantons, and merons are particularly important, and these will be examined in some detail. Not all interesting calculations have been included, unfortunately, owing to natural limitations on the length of this manuscript. (For example, the problem of fermion motion in the field of a monopole has been studied, and also the corresponding Euclidean problem for instantons and merons. We do not report the results of these calculations.) As mentioned previously, the main activity at present is in the Euclidean domain. In Sec. VII we give an extensive introduction to this subject. Even here, however, a really complete review of all the work on all the various conjectured or established instanton effects was hardly feasible. Therefore we have been somewhat selective in Sec. VII, limiting our discussion largely to the aspects of instanton physics which seem to be fairly well understood. (Here one should perhaps read: "most familiar to the author.") In particular, with regard to instanton effects, we must apologize to authors whose work has not been cited.

Classical YM theory can be studied independently of exact solutions, of course. This is an interesting pursuit, because any results gained may lead to improvements in the path-integral formulation of quantum YM field theory. We do not report on the work in this direction.

The structure of our review is evident from the Table of Contents. Some technical points concerning Minkowski SU(2) YM theories are discussed in Sec. II. The following three sections are devoted to Minkowski solutions, which are, respectively, invariant under the full SU(2) gauge group; (Sec. III) invariant under a local U(1) subgroup; (Sec. IV) nongauge invariant (Sec. V). (In the last category no very interesting solution is known, but a number of nonexistence theorems are available.) In Sec. VI we discuss an important *ansatz* for the Minkowski SU(2) gauge potential in terms of a scalar field ϕ which must satisfy the ϕ^4 theory equation of motion. Section VII is an introduction to instanton physics. There we encounter Euclidean solutions for the first time. The known instanton and meron solutions are studied in great detail, together with more general aspects of Euclidean YM field theory. We emphasize the role of topological charge and the tunneling interpretation of Euclidean solutions. Finally, in the concluding section, we attempt to summarize the present situation in classical YM theory.

There are nine Appendices. In these a large number of mathematical details are discussed.

Several other reviews have appeared on the subject of classical YM theory. The reader should be aware of the following:

Jackiw (1977): review of soliton quantization including monopoles; discussion of topics in instanton physics;

Jackiw, Rebbi, and Nohl (1977): review of recent developments in classical YM theory;

Goddard and Olive (1978): review of non-Abelian monopole theory;

Callen, Dashen, and Gross (1978): physical application of instanton and meron solutions;

Crewther (1978): topological charge in YM theory and its consequences.

An extensive study of nonperturbative effects in QCD with emphasis on hadronic physics has recently been published (Shifman, Vainshtein, and Zakharov, 1979).

II. SOME RESULTS IN CLASSICAL MINKOWSKI YANG-MILLS THEORY

In this section we derive some technical results which are useful (some of them are essential) for understanding the *Minkowski-space* solutions in Sec. IV below. (Explicit solutions will not be discussed here. The reader who is not interested in full details can proceed directly to Sec. III.) The most important Minkowski-space solutions are the monopole and dyon solutions, which have nonvanishing Higgs fields at infinity. In a sense this behavior is the classical equivalent of spontaneous symmetry breakdown in the quantum theory—the Higgs fields explicitly break the local SU(2) gauge symmetry. Moreover, the Higgs fields provide the monopole and dyon solutions with their topological quantum number. The first two subsections are devoted to these aspects. We see how the gauge symmetry gets broken by the Higgs field. Then we show that for static solutions the time component W_0 of the gauge potential can do the same thing. In both cases a topological charge is involved. In subsection C the concept of a self-dual gauge field is introduced and shown to be equivalent to the Bogomol'ny condition. Also, we give Bogomol'ny's original derivation of this condition, which so nicely brings out its connection with energy minima and topological charge. Then we turn to complex SU(2) gauge fields. A number of interesting complex solutions are known; however, their interpretation is a problem. In subsection D some work of Wu and Yang on this open question is reviewed.

A. Julia-Zee correspondence

The SU(2) gauge theory with a Higgs triplet is defined by the Lagrangian

$$\mathcal{L} = -\tfrac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a + \tfrac{1}{2}D_\mu\phi_a D^\mu\phi_a - U(\phi)\,, \tag{2.1}$$

where

$$G^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + e\epsilon_{abc}W^b_\mu W^c_\nu\,, \tag{2.2}$$

$$D_\mu\phi_a = \partial_\mu\phi_a + e\epsilon_{abc}W^b_\mu\phi_c\,, \tag{2.3}$$

and the Higgs potential is

$$U(\phi) = \tfrac{1}{4}\lambda(m^2/\lambda - \phi^2)^2\,, \quad \phi^2 = \phi_a\phi_a\,. \tag{2.4}$$

The label $a = 1, 2, 3$ is an SU(2) label and $\mathcal{L}$ is invariant under local SU(2) transformations, with ϕ_a and W^a_μ both transforming like the adjoint representation.

In this classical theory there is, nevertheless, violation of local SU(2) gauge invariance. This is caused by the Higgs potential $U(\phi)$. The Higgs field must be nonvanishing at spatial infinity in order that the potential energy be zero there. Thus any physical solution must satisfy

$$\phi_a \to (m/\sqrt{\lambda})n_a(\hat{r})\,, \quad n_a n_a = 1\,, \quad r \to \infty\,. \tag{2.5}$$

This is like spontaneous symmetry breaking in the quantum theory, where one gives the Higgs field a nonzero vacuum expectation value $\langle\phi_a\rangle \neq 0$. (By assumption, there is vacuum at spatial infinity in the classical case.) If $\phi_a \neq 0$ at infinity then it necessarily selects a direction n_a in group space. This "breaks" local SU(2) gauge invariance (in the manifold of physical solutions) in the sense that any solution that satisfies Eq. (2.5) cannot be invariant under the full SU(2) gauge group. This solution will, however, be invariant under a U(1) subgroup of the SU(2) gauge group. The vector $n_a(r)$ determines this subgroup. In the quantum theory, the vacuum expectation value $\langle\phi_a\rangle \neq 0$ similarly determines the unbroken U(1) subgroup.

In the limit $m^2 \to 0, \lambda \to 0$ with $m^2/\lambda < \infty$ the Higgs potential $U(\phi)$ in Eq. (2.4) vanishes. In this limit the local SU(2) gauge symmetry of the classical solution may or may not be "restored." It is restored if the limiting value of m^2/λ is zero, but not otherwise.

Besides the Lagrangian (2.1) we want to consider another one:

$$\mathcal{L} = -\tfrac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a - \tfrac{1}{4}\lambda(m^2/\lambda + W^2)^2\,, \tag{2.6}$$

$$W^2 = W^a_\mu W^\mu_a\,. \tag{2.7}$$

This is a pure YM theory with a local gauge symmetry breaking potential. To minimize the potential energy at infinity we require

$$W^2 \to -m^2/\lambda\,, \quad r \to \infty\,. \tag{2.8}$$

This boundary condition plays essentially the same role as does Eq. (2.5) in forcing physical solutions to be noninvariant under SU(2) gauge transformations. When $m^2 \to 0, \lambda \to 0$ with $m^2/\lambda < \infty$ the gauge symmetry breaking term in Eq. (2.6) vanishes. However, the boundary condition (2.8) is only consistent with full gauge invariance if $m^2/\lambda = 0$.

The reason why we concern ourselves with the Lagrangian (2.6) and the boundary condition (2.8) is the Julia–Zee correspondence. Julia and Zee (1975), in their study of the dyon solution, observed that the gauge potential component W^a_0 enters the equations of motion very much as a Higgs field does. In fact, for $m^2 = 0$, $\lambda = 0, m^2/\lambda < \infty$ one can reinterpret W^a_0 as an imaginary Higgs field $i\phi_a$ or conversely ϕ_a as an imaginary gauge potential iW^a_0. This is only true for *static* fields, however. To see why, consider an arbitrary gauge transformation ω, under which the Higgs and gauge fields transform like

$$\phi \to \omega\phi\omega^{-1}\,,$$
$$W_\mu \to \omega W_\mu \omega^{-1} - (i/e)(\partial_\mu\omega)\omega^{-1}\,, \tag{2.9}$$

where 2×2 matrix notation is used (see Appendix A). If ω is time independent then $\partial_0\omega = 0$ and ϕ and W_0 transform in the same way. Therefore it is to be expected that W_0 and ϕ will contribute to gauge-invariant quantities (such as the Lagrangian) in much the same way. For time-dependent fields this will not be the case.

We now give a precise statement of the Julia–Zee correspondence. Let us consider a static solution of the theory (2.1) with the form

$$W'^a_0 = 0\,; \quad \phi'_a \text{ and } W'^a_i \neq 0\,; \tag{2.10}$$

and a static solution of the pure YM theory (2.6) with the form

$$W^a_0 \text{ and } W^a_i \neq 0\,. \tag{2.11}$$

Then these two solutions are *mathematically the same* if

$$W^a_0 = i\phi'_a\,, \quad W^a_i = W'^a_i\,. \tag{2.12}$$

This is true for any m^2 and λ; thus it is true in the limit $m^2 \to 0, \lambda \to 0$ with $m^2/\lambda < \infty$ when the theory (2.6) becomes the pure gauge theory. We now give two simple demonstrations of statement (2.12).

The first proof is to show that the two Lagrangians corresponding to the two solutions are identical. The potential terms are the same because $W^2 = -\phi'^2 + W'^2$ follows from Eq. (2.12). The kinetic terms are the same because $G^a_{ij} = G'^a_{ij}$ and

$$G^a_{0j} = -\partial_j W^a_0 + e\epsilon_{abc}W^b_0 W^c_j = -iD_j\phi'_a = -iD'_j\phi'_a\,, \tag{2.13}$$

so that

$$\begin{aligned}\tfrac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a &= \tfrac{1}{4}G^a_{ij}G^{ij}_a + \tfrac{1}{2}G^a_{0j}G^{0j}_a \\ &= \tfrac{1}{4}G'^a_{ij}G'^{ij}_a - \tfrac{1}{2}D'_j\phi'_a D'^j\phi'_a \\ &= \tfrac{1}{4}G'^a_{\mu\nu}G'^{\mu\nu}_a - \tfrac{1}{2}D'_\mu\phi'_a D'^\mu\phi'_a\,.\end{aligned}$$

Here we have used

$$D'_0\phi'_a = 0\,, \quad G'^a_{0j} = 0\,,$$

which only holds for static fields.

The second proof uses the equations of motion, which for the primed solution are

$$\begin{aligned}\partial^\nu G'^a_{\mu\nu} &= e\epsilon_{abc}[G'^b_{\mu\nu}W'^\nu_c - (D'_\mu\phi'_b)\phi'_c] \\ &\quad - W'^a_\mu\lambda(m^2/\lambda + W'^2 - \phi'^2)\,,\end{aligned} \tag{2.14}$$

$$\begin{aligned}\partial^\mu D'_\mu\phi'_a &= e\epsilon_{abc}(D'_\mu\phi'_b)W'^\mu_c \\ &\quad + \phi'_a\lambda(m^2/\lambda + W'^2 - \phi'^2)\,.\end{aligned} \tag{2.15}$$

For static fields satisfying Eq. (2.12) these equations can be rewritten

$$\begin{aligned}\partial^j G^a_{ij} &= e\epsilon_{abc}[G^b_{ij}W^j_c - (D_i\phi'_b)\phi'_c] \\ &\quad - W^a_i\lambda(m^2/\lambda - \phi'^2 - \mathbf{W}^2_a)\end{aligned} \tag{2.16}$$

$$\begin{aligned}\partial^j D_j\phi'_a &= e\epsilon_{abc}(D_i\phi'_b)W^i_c \\ &\quad + \phi'_a\lambda(m^2/\lambda - \phi'^2 - \mathbf{W}^2_a)\,.\end{aligned} \tag{2.17}$$

The static equations of motion for the unprimed solution are

$$\begin{aligned}\partial^j G^a_{ij} &= e\epsilon_{abc}[G^b_{ij}W^j_c + G^b_{i0}W^0_c] \\ &\quad - W^a_i\lambda(m^2/\lambda + W^2)\,,\end{aligned} \tag{2.18}$$

$$\begin{aligned}\partial^j G^a_{0j} &= e\epsilon_{abc}G^b_{0j}W^j_c \\ &\quad - W^a_0\lambda(m^2/\lambda + W^2)\,.\end{aligned} \tag{2.19}$$

Using Eq. (2.13) we see that Eqs. (2.16) and (2.17) are identical with Eqs. (2.18) and (2.19), which completes the proof.

B. Fields at infinity and topology

Static YM solutions may have nontrivial topological properties if a group-nonsinglet field is nonvanishing at infinity. Then the solution defines a map of the two-sphere at infinity into some well defined manifold; and in many cases this map is nontrivial. An elegant way to analyze this problem—in terms of homotopy groups—was first pointed out by Tyupkin, Fateev, and Schwartz (1975) and Monastyrskii and Perelomov (1975). We now review this analysis, trying to stress the generality of the considerations involved. The reader who is unfamiliar with homotopy groups will find an elementary treatment in Appendix I.

Let us begin with the SU(2) case. Recall the boundary condition (2.5) where the unit vector $n_a(\hat{r})$ depends on direction. This unit vector defines a mapping of the sphere at infinity S^2_∞ onto the unit sphere $S^2_{\rm I}$ in group space,

$$n_a(\hat{r}):\ S^2_\infty \to S^2_{\rm I}\ . \tag{2.20}$$

The mapping may cover $S^2_{\rm I}$ zero times, one time, or any integral number of times. (Other possibilities can be imagined, but they are not topologically different from the ones just mentioned which have an integral winding number. For example, if only the northern hemisphere of $S^2_{\rm I}$ gets covered, this is topologically the same as the map with zero winding number.) Thus any map (2.20) is characterized by an integer n which is sometimes called the "winding number." Maps with the same n are topologically equivalent, or homotopic, to one another. (Homotopic maps are continuously deformable into each other.) Maps with different n are inequivalent. The existence of this equivalence relation enables one to separate all maps (2.20) into equivalence classes. Remarkably enough, these equivalence classes themselves form the elements of a group, called the second homotopy group, which in this case is (see Appendix I)

$$\pi_2(S^2_{\rm I}) = Z\ . \tag{2.21}$$

The elements of Z are the integers, i.e., the winding number n. As every physical solution must satisfy the boundary condition (2.5), we can also separate the manifold of physical solutions into topologically distinct classes, i.e., solutions with the same winding number. Later on, in Sec. IV, we shall see that this winding number or "topological charge" is the magnetic charge of static solutions of the theory (2.1).

We have seen that physical solutions of the theory (2.1) belong to topological classes. Within each class, the solutions with lowest energy should be stable. If this were not so, then when one solution "decayed" into another with, say, lower n, the boundary condition would have to change. However, this would entail a discontinuous change in topology. One assumes that this is somehow forbidden by an infinite energy barrier. A better way to put it would be that a time-dependent solution that interpolates between two solutions with different n has infinite action. [To visualize this, think of changing the $n=1$ solution with $\hat{n}(\hat{r}) = \hat{r}$ into the $n=0$ solution with $\hat{n}(\hat{r}) = \hat{z}$.] Therefore, by introducing a nonzero Higgs field at infinity, and thereby equipping all physical solutions with maps $S^2 \to S^2$, we have arranged that certain of these solutions are "topologically" stable.

The preceding remarks concern the theory (2.1) with a Higgs field. However, we know from the Julia–Zee correspondence (2.12) that any static solution of the theory (2.1) implies the existence of a static solution of the theory (2.6). If the Higgs field satisfies the boundary condition (2.5) then in the pure YM theory the corresponding boundary condition would be

$$W^a_0 \to i(m/\sqrt{\lambda})n_a(\hat{r})\ ,\ r \to \infty\ . \tag{2.22}$$

It should be clear that static solutions which satisfy this boundary condition have exactly the same topological structure as the solutions of the theory (2.1) that we have been discussing up to now. Thus it is not really necessary to introduce Higgs fields to introduce topology. Nevertheless, Higgs fields are the most convincing way to do this, just as they provide the most convincing way known at present to break the local gauge symmetry in a YM theory.

The reader will, no doubt, have recognized that the homotopy group analysis of the simple example under discussion is rather an extravagance. One could just as well replace the key formula (2.21) with the primitive concept of the winding number. This is because a very simple mapping $(S^2 \to S^2)$ is involved. When one goes on to study larger gauge groups it is obvious that more complicated mappings $S^2 \to M$ will be encountered. Usually it is rather difficult to grasp the topological structure of these maps by merely inspecting them. This is a problem, because the topological structure of the map is closely related to the magnetic "charge" of the classical solution, and therefore we want to understand it. Fortunately, homotopy group theory is available to help us.

Let us consider a YM theory with gauge group G which is broken to a local subgroup H (in the quantum theory) by some set of Higgs fields ϕ_a. This theory will have a manifold M of constant vacuum solutions

$$M = \{\phi_a = V_a, W_\mu = 0 | \phi^2 \text{ fixed}\}\ . \tag{2.23}$$

Since H is unbroken, the constant fields $\phi_a = V_a$ are invariant under the action of H (by assumption). Another manifold whose elements are invariant under H (by construction, in this case) is the coset space G/H. These two manifolds are essentially the same

$$M \sim G/H\ . \tag{2.24}$$

Here topological equivalence is implied.

It may be taken for granted that all classical fields at spatial infinity take values which belong to M,

$$\phi_a \to \phi_a(\theta, \phi) \in M\ ,\quad r \to \infty\ , \tag{2.25}$$

for any solution with finite energy. This boundary condition of course defines a mapping of the sphere at infinity into the manifold M,

$$\phi_a(\theta, \phi):\ S^2_\infty \to M\ . \tag{2.26}$$

Two such maps are either homotopic (i.e., continuously deformable into one another) or not. Thus all maps (2.26) can be uniquely assigned to equivalence classes. These classes are the elements of the homotopy group $\pi_2(M)$ (see Appendix I). If we are able to calculate $\pi_2(M)$ then we will know quite a lot about the topology of the classical solutions of the theory. If $\pi_2(M) = 0$ is the trivial group then there can be no monopole solutions of the theory in question because the elements of $\pi_2(M)$ are labeled by the allowed monopole charges. If $\pi_2(M) \neq 0$ then monopolelike solutions may exist.

From Eq. (2.24) it follows that

$$\pi_2(M)=\pi_2(G/H)\,. \tag{2.27}$$

Thus if we can calculate $\pi_2(G/H)$ then we will know whether monopole solutions can exist. This is easily done for any simply connected group G, for there exists an identity

$$\pi_2(G/H)=\pi_1(H)\,. \tag{2.28}$$

Thus we only need $\pi_1(H)$, which is known for all Lie groups H (see Appendix I). If H is simply connected [like SU(n), for example] then $\pi_1(H)=0$ and there are no monopole solutions. If H is not simply connected then $\pi_1(H)$ is nontrivial, and there may be monopole solutions. For the frequently occurring case $H=\mathrm{U}(1)$ we have $\pi_1(H)=Z$, and there should be monopole solutions with a single integral charge. For $H=\mathrm{U}(1)\otimes\mathrm{U}(1)\otimes\cdots$ the monopoles would carry several different charges because $\pi_1(H)=Z\oplus Z\oplus\cdots$. Equation (2.28) has to be modified if G is not a simply connected group, but these cases can also be easily handled.

C. Self-duality and the Bogomol'ny condition

We shall call any Minkowski-space SU(2) gauge field *self-dual* if it satisfies the condition

$$\tilde G^a_{\mu\nu}\equiv\tfrac12\varepsilon_{\mu\nu\alpha\beta}G^{\alpha\beta}_a=\pm iG^a_{\mu\nu}\,, \tag{2.29}$$

or equivalently,

$$\pm iE^a_n=B^a_n\,, \tag{2.30}$$

where

$$E^a_n\equiv G^a_{0n}\,,\quad B^a_n\equiv-\tfrac12\varepsilon_{nij}G^a_{ij} \tag{2.31}$$

are the SU(2) "electric" and "magnetic" YM fields. The factor i in this definition of self-duality is unavoidable because we are working in Minkowski space where $\tilde{\tilde G}{}^a_{\mu\nu}=-G^a_{\mu\nu}$. Clearly, in Minkowski space any self-dual field configuration contains complex fields. (In Euclidean space E^4 the factor i is absent and self-dual fields can be real.)

Self-dual fields are interesting because they *automatically* satisfy the equations of motion of the pure gauge theory. In the SU(2) case the relation

$$\partial^\nu\tilde G^a_{\mu\nu}=e\varepsilon_{abc}\tilde G^b_{\mu\nu}W^\nu_c \tag{2.32}$$

is nothing more than an identity. For a self-dual field this identity becomes the equation of motion for the SU(2) gauge theory. Any potential W^a_μ which leads to a self-dual tensor $G^a_{\mu\nu}$ is therefore a solution of the equation of motion.

One can try to make use of this fact by searching for solutions of the self-dual equations, which are first order, rather than trying to solve the second-order equations of motion. Interesting solutions can be found in this way. But self-duality is a very special property which most solutions do not have. In the long run it might be more rewarding to try to solve the equations of motion directly.

Another comment: YM fields in any theory with an explicit local gauge symmetry breaking term in the Lagrangian cannot be self-dual. This statement is easy to verify.

Any self-dual solution in Minkowski space has a vanishing energy–momentum tensor

$$\theta_{\mu\nu}=-G^a_{\mu\lambda}G^{a\lambda}_\nu+\tfrac14g_{\mu\nu}G^a_{\alpha\beta}G^{\alpha\beta}_a$$
$$=-\tfrac14(G^a_{\mu\lambda}+i\tilde G^a_{\mu\lambda})(G^{a\lambda}_\nu-i\tilde G^{a\lambda}_\nu)\,. \tag{2.33}$$

The components of $\theta_{\mu\nu}$ are

$$\theta_{00}=\tfrac12(E^a_nE^a_n+B^a_nB^a_n)=\sum_i\theta_{ii}\,, \tag{2.34}$$

$$\theta_{0j}=-\varepsilon_{jmn}E^a_mB^a_n\,, \tag{2.35}$$

$$\theta_{ij}=-E^a_iE^a_j-B^a_iB^a_j+\delta_{ij}\tfrac12(E^a_nE^a_n+B^a_nB^a_n)\,. \tag{2.36}$$

Obviously $\theta_{\mu\nu}=0$ for any field configuration with $B^a_n=\pm iE^a_n$. Moreover, the Lagrangian

$$\mathcal{L}=-\tfrac14G^a_{\mu\nu}G^{\mu\nu}_a=\tfrac12(E^a_nE^a_n-B^a_nB^a_n) \tag{2.37}$$

is minimized (or maximized) by a self-dual solution.

Within the framework of the theory (2.1) Bogomol'ny (1976) introduced a condition which is closely analogous to the self-dual condition in a pure YM theory. The Bogomol'ny condition is

$$B^a_n\equiv-\tfrac12\varepsilon_{nij}G^a_{ij}=\pm D_n\phi_a\,. \tag{2.38}$$

Any static solution of this equation with $W^a_0=0$ is a solution of the equations of motion for the theory (2.1) in the limit $U(\phi)\to0$. The proof is simple.

All we need is the general correspondence (2.12) between a solution of the type considered by Bogomol'ny and a pure SU(2) gauge theory solution. From Eq. (2.12) we find

$$E^a_n=-iD_n\phi_a\,,$$

and then the Bogomol'ny condition (2.38) can be written

$$B^a_n=\pm iE^a_n\,,$$

which is precisely the condition (2.30) for a self-dual SU(2) gauge field. Q.E.D.

Now we give Bogomol'ny's derivation of condition (2.38). This derivation has two important features. It shows that (i) condition (2.38) minimizes the energy of a static solution with $W^a_0=0$, and (ii) that this minimum energy is proportional to the topological charge of the solution.

The total energy is

$$E=\int d^3x\,[\tfrac14G^a_{ij}G^a_{ij}+\tfrac12D_i\phi_aD_i\phi_a+U(\phi)$$
$$+\tfrac12G^a_{0j}G^a_{0j}+\tfrac12D_0\phi_aD_0\phi_a]\,. \tag{2.39}$$

The last two terms are absent for a static solution with $W^a_0=0$. Dropping these terms we rewrite E in the form

$$E=\int d^3x\,[\tfrac14(G^a_{ij}-\varepsilon_{ijn}D_n\phi_a)^2+\tfrac12\partial_nJ_n+U(\phi)]\,, \tag{2.40}$$

where

$$\partial_nJ_n=\varepsilon_{ijn}\partial_n[G^a_{ij}\phi_a]=\varepsilon_{ijn}G^a_{ij}D_n\phi_a\,. \tag{2.41}$$

Ignoring possible singularities Gauss's theorem then leads to

$$E=\int d\Omega\,\tfrac12\varepsilon_{ijn}[r^2\hat r_nG^a_{ij}\phi_a]_{r=\infty}$$
$$+\int d^3x\,[\tfrac14(G^a_{ij}-\varepsilon_{ijn}D_n\phi_a)^2+U(\phi)]\,. \tag{2.42}$$

The potential $U(\phi)$ vanishes in the limit $\lambda\to0$, $m^2\to0$, m^2/λ finite. In this limit the total energy is obviously minimized by functions which satisfy the Bogomol'ny condition (2.38). Then the energy is given by the first term in Eq. (2.41), which is proportional to the topological charge of the solution.

This is not difficult to show. All we need is a suitable definition of the physical magnetic field at large r, for example

$$B_n \equiv \frac{1}{2\phi} \varepsilon_{ijn} G^a_{ij} \phi_a . \tag{2.43}$$

(See Appendix E. For large r all definitions of the electromagnetic field tensor become the same.) The first term in Eq. (2.42) can therefore be written

$$\int d\Omega \, [\phi r^2 \hat{r}_n B_n]_{r=\infty} = (m/\sqrt{\lambda}) g ; \tag{2.44}$$

$$g \equiv \int d\Omega \, [r^2 \hat{r}_n B_n]_{r=\infty} , \tag{2.45}$$

where the boundary value $\phi = m/\sqrt{\lambda}$ of the Higgs field at infinity has been used. If the integral (2.45) is finite, then it is clearly equal to the magnetic charge of the solution. As we shall see in Sec. IV, this magnetic charge is proportional to a conserved topological charge.

A similar result can be established for static solutions with nonzero W^a_0 (Bogomol'ny, 1976; Coleman, Parke, Neveu, and Sommerfield, 1977). One has to assume that W^a_0 and ϕ_a are parallel in group space, so that $D_0\phi_a = 0$. Then Eq. (2.39) can be rewritten

$$E = \int d^3x \, [\tfrac{1}{4}(G^a_{ij} - \cos\theta \varepsilon_{ijn} D_n \phi_a)^2 + \tfrac{1}{2}(G^a_{0n} - \sin\theta D_n \phi_a)^2 + U(\phi) + \tfrac{1}{2}\cos\theta \partial_n J_n + \sin\theta \partial_n K_n] , \tag{2.46}$$

with J_n as in Eq. (2.41) and

$$\partial_n K_n = \partial_n [\phi_a G^a_{0n}] = G^a_{0n} D_n \phi_a + \phi_a [\partial_n G^a_{0n} - e\varepsilon_{abc} G^b_{0n} W^c_n] . \tag{2.47}$$

Here the square bracket is zero because of the equation of motion

$$\partial^\nu G^a_{\mu\nu} = e\varepsilon_{abc} [G^b_{\mu\nu} W^\nu_c - (D_\mu \phi_b)\phi_c]$$

and the assumption $D_0 \phi_b = 0$. Defining the physical electric field at large r by

$$E_n \equiv \hat{\phi}_a G^a_{0n} , \tag{2.48}$$

we find

$$\int d^3x \, \partial_n K_n = \int d\Omega \, [\phi r^2 \hat{r}_n E_n]_{r=\infty} = m/\sqrt{\lambda}\, q , \tag{2.49}$$

where

$$q \equiv \int d\Omega \, [r^2 \hat{r}_n E_n]_{r=\infty} \tag{2.50}$$

is the electric charge of the solution. Therefore the total energy of the solution is

$$E = (m/\sqrt{\lambda})(g\cos\theta + q\sin\theta) + \varepsilon , \tag{2.51}$$

$$\varepsilon = \int d^3x \, [\tfrac{1}{4}(G^a_{ij} - \cos\theta \varepsilon_{ijn} D_n \phi_a)^2 + \tfrac{1}{2}(G^a_{0n} - \sin\theta D_n \phi_a)^2 + U(\phi)] . \tag{2.52}$$

Choosing the arbitrary angle θ as follows:

$$\sin\theta = q/\sqrt{g^2+q^2} , \quad \cos\theta = g/\sqrt{g^2+q^2} , \tag{2.53}$$

we obtain an interesting formula for the total energy,

$$E = (m/\sqrt{\lambda})\sqrt{g^2+q^2} + \varepsilon . \tag{2.54}$$

Now suppose that a solution exists with $\varepsilon = 0$. Then

$$E = (m/\sqrt{\lambda})\sqrt{g^2+q^2} , \tag{2.55}$$

and the separate conservation of electric and magnetic charge implies that this static solution is *stable*.

A stable solution with $\varepsilon = 0$ can only exist in the limit $U(\phi) = 0$. The solution must also satisfy the first-order equations

$$G^a_{ij} = \cos\theta \varepsilon_{ijn} D_n \phi_a , \tag{2.56}$$

$$G^a_{0n} = \sin\theta D_n \phi_a , \tag{2.57}$$

where $\tan\theta = q/g$. These equations are a generalization of the Bogomol'ny condition (2.38). It is possible to trivially satisfy the second condition (2.57) by assuming that W^a_0 and ϕ_a are the same function up to a constant:

$$W^a_0 = -\sin\theta \phi_a . \tag{2.58}$$

This leaves Eq. (2.56), which can now be written

$$[gG^a_{0n}] = \tfrac{1}{2}\varepsilon_{ijn} [qG^a_{ij}] . \tag{2.59}$$

D. SU(2, C) gauge theory

Some general features of the problem of complex YM gauge potential and fields have been discussed by Wu and Yang (1975, 1976). These authors have shown that a gauge theory with *complex* potential and gauge group G is equivalent to another theory with *real* potentials whose gauge group is the complex extension of G. The complex extension of G is the group whose generators are L_a and iL_a, where L_a are the generators of G. The complex extension of SU(2) is the group of matrices

$$\omega = \exp[i\sigma_a(\phi_a + i\psi_a)/2] ,$$

where ϕ_a, ψ_a are real functions. This group is called SU(2, C). In general $\omega^+ \neq \omega^{-1}$ and the SU(2, C) matrices are nonunitary. The generators of SU(2, C) satisfy the commutation relations

$$[\tfrac{1}{2}\sigma_a, \tfrac{1}{2}\sigma_b] = -[\tfrac{1}{2}i\sigma_a, \tfrac{1}{2}i\sigma_b] = -i[\tfrac{1}{2}\sigma_a, \tfrac{1}{2}i\sigma_b] = i\varepsilon_{abc}\sigma_c/2 . \tag{2.60}$$

Consider a complex SU(2) gauge potential

$$W^a_\mu = u^a_\mu + iv^a_\mu , \tag{2.61}$$

where

$$u^a_\mu \equiv \mathrm{Re} W^a_\mu ; \; v^a_\mu \equiv \mathrm{Im} W^a_\mu \tag{2.62}$$

are real functions. The matrix form of the gauge potential is

$$W_\mu = \tfrac{1}{2}\sigma_a W^a_\mu = u_\mu + v_\mu \tag{2.63}$$

where

$$u_\mu \equiv \tfrac{1}{2}\sigma_a u^a_\mu , \; v_\mu \equiv (i/2)\sigma_a v^a_\mu . \tag{2.64}$$

Now suppose that we interpret the $i\sigma_a/2$ as generators of SU(2, C). Then W_μ in Eq. (2.63) becomes the matrix gauge potential for the gauge theory based on SU(2, C), and the components of this potential are the real functions u^a_μ and v^a_μ. Using this trick we have converted a complex gauge potential into a real one. The price we have paid is the introduction of the noncompact gauge group SU(2, C).

Returning to the SU(2) gauge theory we write down the field strength tensor following from the complex potential (2.61):

$$G^a_{\mu\nu} = U^a_{\mu\nu} + iV^a_{\mu\nu}. \tag{2.65}$$

Here the real and imaginary parts of $G^a_{\mu\nu}$ are

$$U^a_{\mu\nu} = \partial_\mu u^a_\nu - \partial_\nu u^a_\mu + e\varepsilon_{abc}(u^b_\mu u^c_\nu - v^b_\mu v^c_\nu),$$
$$V^a_{\mu\nu} = \partial_\mu v^a_\nu - \partial_\nu v^a_\mu + e\varepsilon_{abc}(u^b_\mu v^c_\nu + v^b_\mu u^c_\nu). \tag{2.66}$$

Now these real functions are the field strengths in the SU(2, C) theory. Moreover, the SU(2) theory equations of motion

$$\partial^\nu U^a_{\mu\nu} = e\varepsilon_{abc}(U^b_{\mu\nu}u^\nu_c - V^b_{\mu\nu}v^\nu_c),$$
$$\partial^\nu V^a_{\mu\nu} = e\varepsilon_{abc}(U^b_{\mu\nu}v^\nu_c + V^b_{\mu\nu}u^\nu_c), \tag{2.67}$$

are the equations of motion of the SU(2, C) gauge theory. These statements are trivially verified with the help of the SU(2, C) structure functions in Eq. (2.60).

There is a fundamental difference between the gauge theories based on SU(2) and SU(2, C). The latter group is noncompact, and this has important consequences. Perhaps the most important one is that the energy in the SU(2, C) theory is *not positive definite*. Wu and Yang (1976) have explained how this comes about. To construct $\mathcal{L}$, $\theta_{\alpha\beta}, \ldots$ in a gauge theory one has to define a scalar product in group space. This amounts to introducing a metric η_{ab}. For SU(2) the metric is simply $\eta_{ab} = \delta_{ab}$. But for SU(2, C) there is more freedom; the metric can be parametrized by a real angle θ. Let us introduce a two-component notation for the SU(2, C) gauge theory:

$$W^a_\mu = \begin{pmatrix} u^a_\mu \\ v^a_\nu \end{pmatrix}, \quad G^a_{\mu\nu} = \begin{pmatrix} U^a_{\mu\nu} \\ V^a_{\mu\nu} \end{pmatrix}. \tag{2.68}$$

Then the metric in this theory can be written

$$\eta_{ab} = \delta_{ab}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{2.69}$$

[The proof of this statement is to show that gauge invariance under SU(2, C) transformations is consistent with the metric (2.69) for arbitrary θ.] The Lagrangian constructed with this metric,

$$\mathcal{L} = -\tfrac{1}{4}\eta_{ab}G^a_{\mu\nu}G^{\mu\nu}_b$$
$$= \cos\theta[-\tfrac{1}{4}U^a_{\mu\nu}U^{\mu\nu}_a - \tfrac{1}{4}V^a_{\mu\nu}V^{\mu\nu}_a] + \sin\theta[-\tfrac{1}{2}U^a_{\mu\nu}V^{\mu\nu}_a], \tag{2.70}$$

is SU(2, C) invariant. Comparing this Lagrangian with the SU(2) Lagrangian calculated from the complex potential (2.65) one finds immediately that

$$\mathcal{L}[\mathrm{SU}(2,C)] = \mathrm{Re}\{e^{-i\theta}\mathcal{L}[\mathrm{SU}(2)]\}. \tag{2.71}$$

Similarly, for the SU(2, C) Hamiltonian one finds

$$\mathcal{H}[\mathrm{SU}(2,C)] = \mathrm{Re}\{e^{-i\theta}\mathcal{H}[\mathrm{SU}(2)]\}. \tag{2.72}$$

This Hamiltonian is real, but clearly it is not positive definite.

In Sec. V we show that static, real, finite-energy solutions of the SU(2) gauge theory do not exist. However, *complex* solutions with these properties do exist. An explicit example of such a complex solution due to Hsu and Mac (1977) is discussed in Sec. IV. The Hsu–Mac solution is self-dual and therefore the total energy is zero. Being complex, it can be reinterpreted as a real solution of the SU(2, C) gauge theory. This solution is an example of the general correspondence (2.12) between pure SU(2) solutions and solutions of the SU(2) gauge theory with a Higgs triplet. Equation (2.12) enables us to rephrase the nonexistence theorem in Sec. V as follows: There exists no static, finite-energy solution of the latter theory with *imaginary* Higgs field and real gauge field. Such a solution would be difficult to interpret physically.

III. SOLUTIONS WITH SU(2) GAUGE INVARIANCE

This section is devoted to Minkowski-space solutions of the equations of motion of the pure gauge theory

$$\partial^\nu G^a_{\mu\nu} = e\varepsilon_{abc}G^b_{\mu\nu}W^\nu_c, \tag{3.1}$$

which are invariant under the full SU(2) gauge group. Quite a variety of such solutions are known: non-Abelian plane waves, imbedded Abelian solutions, a sourceless non-Abelian magnetic monopole solution, a class of complex solutions (obtained from a certain *ansatz*), a real solution representing a Minkowski-space meron–antimeron pair, and an elliptic generalization of the meron–antimeron solution. These solutions are all interesting from the mathematical or historical viewpoint. Most of them have not found physical applications, however. An exception is the Wu–Yang monopole, which is the prototype of a non-Abelian string-free magnetic monopole solution. The physical meaning of the meron solutions is not yet understood. In this section we shall discuss the mathematical properties of these various solutions, with little mention of their possible physical applications.

A. Non-Abelian plane wave

A non-Abelian plane-wave solution of Eq. (3.1) has been given by Coleman (1977b). Consider the wave moving in the positive z direction with velocity $v = c$, whose potential has the form

$$W^a_1 = W^a_2 = 0,$$
$$W^a_0 = -W^a_3 = x_1F_a(x_0 - x_3) + x_2G_a(x_0 - x_3). \tag{3.2}$$

One can easily verify that the SU(2) field strengths are

$$G^a_{01} = G^a_{13} = -F_a,$$
$$G^a_{02} = G^a_{23} = -G_a,$$
$$G^a_{03} = G^a_{12} = 0.$$

Given $G^a_{\mu\nu}$ it is easy to see that Eq. (3.1) is satisfied, and therefore Eq. (3.2) is indeed a solution for arbitrary functions F_a and G_a. (The trivial generalization of this solution to an N-parameter gauge group involves $2N$ independent functions.) Note that the plane-wave solution (3.2) is essentially non-Abelian in nature.

Waves moving in different directions cannot be superimposed. This is to be contrasted with imbedded Abelian plane-wave solutions $W^a_\mu = \lambda_a \exp ik\cdot x(\lambda_a = \text{const})$ which can be superimposed for fixed λ_a.

B. Ikeda-Miyachi solution

The first explicit static solution of Eq. (3.1) was found by Ikeda and Miyachi (1962). This solution can be written in the form

$$W^a_0 = \delta_{a3}(A/r+B), \quad W^a_i = 0, \tag{3.3}$$

with A and B constant. Other, gauge-equivalent forms of this solution are given in the original paper. The potential (3.3) with $B=0$ is essentially a static, pointlike Coulomb potential.

We now verify the Ikeda–Miyachi solution (3.3). The simple calculation involved shows that a solution of this type can be found for any gauge group. From Eq. (3.3) one finds the field strengths

$$G^a_{0j} = -\delta_{a3}\partial_j(A/r+B); \quad G^a_{ij}=0. \tag{3.4}$$

The field equations are trivially satisfied except for one equation

$$\partial^j G^a_{0j} = \delta_{a3}\nabla^2(A/r+B) = 0,$$

which is satisfied everywhere except $r=0$. All we have needed up to this point is the antisymmetry of the SU(2) coupling constants ε_{abc}. But the coupling constants of any Lie group are antisymmetric in the relevant two indices, and therefore this result is generally valid.

For any gauge group one can make the *ansatz*

$$W^a_\mu = \lambda_a A_\mu, \quad \lambda_a = \text{const}. \tag{3.5}$$

This linearizes the field strengths and the equations of motion,

$$G^a_{\mu\nu} = \lambda_a(\partial_\mu A_\nu - \partial_\nu A_\mu), \quad \partial^\nu G^a_{\mu\nu} = 0. \tag{3.6}$$

Thus any Abelian solution can be imbedded in a YM potential.

The preceding discussion shows that any pure YM theory has a static solution in which only the long-range potential is excited. At this point it is natural to ask if this long-range component is *necessarily* present in all static solutions. We shall answer this question in some detail in Sec. V. Our conclusion is that, barring quite extreme boundary conditions at infinity, there will indeed be a long-range field component. This conclusion is not difficult to understand. Consider the pure SU(2) theory equation of motion (3.1), which we rewrite in the form

$$\partial^\nu(\partial_\mu W^a_\nu - \partial_\nu W^a_\mu) = e\varepsilon_{abc}[G^b_{\mu\nu}W^\nu_c - \partial^\nu(W^b_\mu W^c_\nu)]. \tag{3.7}$$

At large r a derivative ∂_i is roughly equivalent to a factor $1/r$. Therefore, barring possible cancellations, the equation of motion in the static case looks something like

$$\frac{1}{r^2}(\text{potential}) \sim \frac{1}{r}(\text{potential})^2 + (\text{potential})^3,$$

which implies

$$(\text{potential}) \sim 1/r.$$

Assuming things are this simple, it is clear that not all components of the potential can decrease exponentially. In terms of gauge symmetry the equivalent statement is that no solutions (with reasonable boundary conditions) exist which correspond to completely broken local SU(2) gauge invariance. There do exist solutions which correspond to local SU(2) broken to U(1), of course, and these solutions contain the expected long-range potential.

One might be tempted to interpret a pointlike solution like (3.3) as an electric charge at rest. This solution satisfies linear field equations, and it may look as though an Abelian formalism can be set up. But this is not possible. To see why not, let us imagine two of these charges separated by a great distance so that they do not interact. The potentials associated with these charges are $W^a_{1,2\mu} = \lambda^a_{1,2}A_{1,2\mu}$, where λ^a_1 and λ^a_2 are constant vectors in group space. The directions of these vectors are completely arbitrary because of gauge invariance. Now we bring one of the charges near to the other one. If λ^a_1 and λ^a_2 are parallel then we still have an Abelian problem with linear field equations. But there is no reason why λ^a_1 and λ^a_2 should be parallel since the individual charges are indifferent to the directions of these vectors. If λ^a_1 and λ^a_2 are not parallel, then the sum $W^a_\mu = \lambda^a_1 A_{1\mu} + \lambda^a_2 A_{2\mu}$ is not an exact solution of the equations of motion. As the charges move together, the nonlinear terms in Eq. (3.1) eventually become important and the solution worsens. For very small separation, W^a_μ is meaningless as a solution.

The point of this discussion is to convince the reader that even in this very simple case one cannot circumvent the essentially nonlinear nature of the problem. Two sources of YM fields may experience a mutual Coulomb interaction at large distances, but at small distances the interaction is bound to be more complicated. This is true even if the sources are pointlike.

C. Wu-Yang solution

Let us now go on to discuss some other pointlike solutions of the pure SU(2) theory. These solutions are obtained by introducing the following *ansatz* [first discovered by Wu and Yang, 1968, for $W^a_0=0$; subsequently extended by other authors (Julia and Zee, 1975; Hsu and Mac, 1977) to the case $W^a_0 \neq 0$],

$$\begin{aligned} eW^a_0 &= ir_a g(r)/r^2, \\ eW^a_i &= \varepsilon_{ain}r_n[1-h(r)]/r^2. \end{aligned} \tag{3.8}$$

This *ansatz* reduces the SU(2) equations of motion to the following coupled equations (see the discussion to follow)

$$\begin{aligned} r^2 g'' &= 2gh^2, \\ r^2 h'' &= h(h^2 - 1 + g^2). \end{aligned} \tag{3.9}$$

Constant g and h evidently imply unbroken local SU(2) gauge invariance because W^a_μ is a pointlike long-range potential in this case.

There are constant solutions of Eqs. (3.9). Two of them,

$$h=1,\ g=0 \text{ and } h=-1,\ g=0,$$

are vacuum solutions with $G^a_{\mu\nu}=0$ ($W^a_\mu = 0$ for the first one while for the second W_μ^a is pure gauge). The nontrivial constant solution is

$$h=0\,,\ g=C=\text{const}\,. \qquad (3.10)$$

When $h=0$ the remaining equation of motion $g''=0$ has the solution $g=C+Dr$ where C and D are constants. For nonzero D the potential W_0^a is nonvanishing at infinity, $W_0^a \to i\hat{r}_a D/e$. However, this contribution to W_0^a is trivial, as it can be removed by a gauge transformation. Therefore, up to a gauge equivalence, Eq. (3.10) is the only solution for $h=0$.

Let us now make a brief digression to bring the theory (2.1) with the Higgs triplet into our discussion. Equation (2.12) tells us that we can also interpret any static solution of the pure SU(2) theory as a solution of the larger theory (2.1), in a certain limit. Thus Eq. (3.10) also gives us a pointlike solution of the latter theory. The relevant *ansatz* corresponding to Eq. (3.8) is

$$e\phi_a = r_a g(r)/r^2\,,\ W_0^a \equiv 0\,,$$
$$eW_i^a = \varepsilon_{ain} r_n[1-h(r)]/r^2\,. \qquad (3.11)$$

The equations of motion obtained from the Lagrangian (2.1) are

$$\partial^\nu G_{\mu\nu}^a = e\varepsilon_{abc}[G_{\mu\nu}^b W_c^\nu - (D_\mu\phi_b)\phi_c]\,,$$
$$\partial^\mu D_\mu \phi_a = e\varepsilon_{abc}(D_\mu\phi_b)W_c^\mu + m^2\phi_a - \lambda\phi_a\phi^2\,. \qquad (3.12)$$

The *ansatz* (3.11) reduces these equations to

$$r^2 g'' = g(2h^2 - m^2r^2 + \lambda g^2/e^2)\,;$$
$$r^2 h'' = h(h^2 - 1 + g^2)\,, \qquad (3.13)$$

which coincide with Eqs. (3.9) when $m^2=0$, $\lambda=0$ as they should. For nonzero m^2, λ there are no constant solutions of Eqs. (3.13) (see Appendix H).

It is worthwhile going through the derivation of Eqs. (3.9) and (3.13); this is one of the important calculations in classical YM theory. From Eq. (3.11) one finds that $G_{0j}^a=0$, $D_0\phi_a=0$, and

$$eG_{ij}^a = -2\varepsilon_{aij}W + \varepsilon_{ijn}r_n r_a W^2 - [\varepsilon_{ain}r_j - \varepsilon_{ajn}r_i]r_n W'/r\,,$$
$$D_i\phi_a = \delta_{ia}(1-r^2W)f + r_a r_i(f' + rfW)/r\,,$$

where here we use the notation

$$eW_i^a = \varepsilon_{ain}r_n W(r)\,,\ e\phi_a = r_a f(r)\,.$$

The first equation in (3.12) (trivial for $\mu=0$) becomes, for $\mu=i$,

$$\varepsilon_{ain}r_n[W'' + (4/r)W' + 3W^2 - r^2W^3 + f^2(1-r^2W)] = 0\,,$$

while the second equation becomes

$$r_a[f'' + (4/r)f' + 4fW - 2r^2fW^2 + m^2f - \lambda r^2 f^3] = 0\,.$$

Changing variables,

$$W = (1-h)/r^2\,,\ f = g/r^2\,,$$

one easily obtains Eqs. (3.13).

Equations (3.8) and (3.11) define what is sometimes called the radial or "no string" gauge. This gauge is useful for the purpose of reducing the equations of motion to the forms (3.9) and (3.13) that can then be solved. However, it rather obscures certain important aspects of the solutions such as the number of components of the gauge potential which are excited in the solution. Another gauge, called the unitary gauge or "string gauge," can be introduced to clarify some of these things. It is related to the former gauge by the local SU(2) transformation

$$\omega = \begin{pmatrix} \cos(\theta/2) & e^{-i\phi}\sin(\theta/2) \\ -e^{i\phi}\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}\,, \qquad (3.14)$$

which rotates the $\hat{r}=(\theta,\phi)$ direction in group space (note that this direction is identified with the $\hat{r}$ direction in three-space) into the z-axis $\hat{z}=(0,0)$. This transformation is discontinuous along the negative z axis, and therefore the pure-gauge term in the transformed gauge potential is singular along this axis, the singularity being the Dirac string (see Appendix D). The string gauge potentials A_μ^a obtained from the *ansätze* (3.8) or (3.11) by the gauge transformation (3.14) are

$$A_0^a \text{ or } i\phi_a = \delta_{a3} ig(r)/er$$
$$A_i^1 = -(1/2er)h(r)[\hat{\phi}\cos\phi + \hat{\theta}\sin\phi]_i$$
$$A_i^2 = (i/2er)h(r)[-\hat{\phi}\sin\phi + \hat{\theta}\cos\phi]_i \qquad (3.15)$$
$$A_i^3 = -(1/2er)\tan\tfrac{1}{2}\theta[\hat{\phi}]_i\,,$$

where $\hat{\theta}$ and $\hat{\phi}$ are unit vectors associated with the polar angles θ and ϕ in three-space. A_i^3 is the familiar Dirac string potential (Appendix D), and $A_i^{1,2}$ are the remaining two components of the SU(2) potential. These components are essentially determined by the function $h(r)$. When spontaneous symmetry breaking occurs then $h(r)$ is an exponential function and $A_\mu^{1,2}$ are the massive gauge field components. But A_i^3 depends neither on $h(r)$ nor $g(r)$; it is determined entirely by the form of the *ansatz* (3.8) and is always pointlike and massless, regardless of whether the local SU(2) symmetry is broken or not.

We return now to the pointlike solution (3.10) above, which in the string gauge becomes

$$A_0^a \text{ or } i\phi_a = \delta_{a3} iC/er\,,$$
$$A_i^1 = A_i^2 = 0\,,$$
$$A_i^3 = -(1/2er)\tan\tfrac{1}{2}\theta[\hat{\phi}]_i\,. \qquad (3.16)$$

With $C=0$ this is the original solution obtained by Wu and Yang (1968), and in the form presented above it evidently corresponds to a magnetic monopole. For $C\neq 0$ the solution can be interpreted as a dyon with electric charge $Q=iC/e$ in the pure SU(2) theory.

D. Rosen's *ansatz*

Rosen (1972) introduced the *ansatz*

$$eW_i^a = (Cx_0, iCx_0, i)\partial_i\phi/\phi\,,\ W_0^a = 0\,, \qquad (3.17)$$

where C is a constant and $\phi=\phi(\mathbf{x})$ is a static function. The SU(2) field strengths are easily found to be

$$G_{ij}^a = 0\,,\ G_{0j}^3 = 0\,,\ G_{0j}^1 = -iG_{0j}^2 = C\partial_j\phi/\phi\,. \qquad (3.18)$$

The equations of motion (3.1) therefore reduce to

$$(1/\phi)\nabla^2\phi = 0\,. \qquad (3.19)$$

Any static solution of this equation leads to a complex solution of the YM theory.

A spherically symmetric solution of Eq. (3.19) is

$$\phi = a + b/r\,.$$

The nonzero gauge potentials are

$$eW_i^a = -(Cx_0, iCx_0, i)[br_i/r^2(ar+b)]\,,$$

and the corresponding field strengths are

$$G_{0j}^1 = -iG_{0j}^2 = -Cbr_j/er^2(ar+b)\,.$$

Rosen's *ansatz* is successful (in the sense that it drastically simplifies the YM equations of motion) because $G_{ij}^a = 0$. This implies that W_i^a in Eq. (3.17) is a pure-gauge potential. It is easy to verify that

$$eW_i^a = -i(\partial_i \omega)\omega^{-1}\,, \tag{3.20}$$

where the gauge transformation is

$$\omega = \tfrac{1}{2}(\phi + 1/\phi) + \hat{n}\cdot\sigma\tfrac{1}{2}(\phi - 1/\phi)\,,$$
$$\hat{n} = i(Cx_0, iCx_0, i)\,. \tag{3.21}$$

But $W_0^a = 0$ is not the corresponding time component of the pure-gauge potential, and therefore the *ansatz* (3.17) leads to nontrivial YM field configurations.

As an exercise, let us gauge-transform the potential (3.17) using the inverse of ω,

$$\omega^{-1} = \tfrac{1}{2}(\phi + 1/\phi) - \hat{n}\cdot\sigma\tfrac{1}{2}(\phi - 1/\phi)\,, \tag{3.22}$$

to obtain the equivalent *ansatz*

$$W_i^a = 0\,,\ \ W_0^3 = 0\,,\ \ W_0^1 = -iW_0^2 = -(C/2e)(\phi^2 - 1)\,. \tag{3.23}$$

The corresponding field strengths are

$$G_{ij}^a = 0,\ \ G_{0j}^3 = 0,\ \ G_{0j}^1 = -iG_{0j}^2 = (C/e)\partial_j\phi\,, \tag{3.24}$$

and one easily finds again the equation of motion $\nabla^2\phi = 0$.

E. ϕ^4 *ansatz*

In Sec. VI we shall discuss an *ansatz* due to 't Hooft (1976a, b, c), Corrigan and Fairlie (1977), and Wilczek (1977) which, like Rosen's *ansatz* above, is complex and involves one unknown scalar function. It is

$$eW_0^a = \pm i\partial_a\phi/\phi\,,\ \ eW_i^a = \varepsilon_{ian}\partial_n\phi/\phi \pm i\delta_{ai}\partial_0\phi/\phi\,. \tag{3.25}$$

This W_μ^a satisfies the YM equations of motion if ϕ satisfies

$$(1/\phi)\Box\phi + \lambda\phi^2 = 0\,, \tag{3.26}$$

where λ is an arbitrary constant. In particular, we can choose $\lambda = 0$, and then if ϕ is time independent it must satisfy the condition (3.19) in Rosen's *ansatz*. (In this case, the two *ansätze* must be gauge equivalent.) Clearly we can obtain a large class of complex Minkowski YM solutions from the *ansatz* (3.25). All harmonic functions (i.e., solutions of $\Box\phi = 0$) provide us with YM solutions, for example. There also exist elliptic solutions of Eq. (3.26) which lead to elliptic YM solutions. (See Sec. VI for more details.)

F. Witten's *ansatz*

In Sec. VII we discuss an *ansatz* for the Euclidean space-time gauge potential due to Witten (1977). There is also a useful Minkowski version of this *ansatz*, namely

$$eW_0^a = -\frac{x_a}{r}A_0\,,$$
$$eW_i^a = \varepsilon_{ian}\frac{x_n}{r^2}(1+\phi_2) + \frac{x_a x_i}{r^2}A_1 + \left(\delta_{ai} - \frac{x_a x_i}{r^2}\right)\frac{1}{r}\phi_1\,, \tag{3.27}$$

where the four *ansatz* functions depend on t and r. This *ansatz* is useful for finding real, time-dependent rotationally symmetric YM solutions. The equations of motion following from (3.27) are

$$\partial_0(r^2F_{01}) = 2(\phi_1 D_1\phi_2 - \phi_2 D_1\phi_1)\,,$$
$$\partial_1(r^2F_{01}) = 2(\phi_1 D_0\phi_2 - \phi_2 D_0\phi_1)\,,$$
$$r^2(D_0D_0 - D_1D_1)\phi_a = \phi_a(1 - \phi_1^2 - \phi_2^2)\,, \tag{3.28}$$

where

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu\,,$$
$$D_\mu\phi_a \equiv \partial_\mu\phi_a + \varepsilon_{ab}A_\mu\phi_b\,, \tag{3.29}$$

with indices μ, ν taking the values 0 and 1 and $\partial_0 = \partial/\partial x_0$, $\partial_1 = \partial/\partial r$. Notation (3.29) emphasizes the remarkable property of Witten's *ansatz*, that it reduces the SU(2) problem to an Abelian Higgs model problem in two space-time dimensions. To show what is meant by this statement let us consider local SU(2) gauge transformations of the form

$$\omega = \exp[\tfrac{1}{2}if(r,t)\hat{r}\cdot\sigma]\,. \tag{3.30}$$

Under this gauge transformation the *ansatz* functions transform as follows:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu f\,,$$
$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \to \begin{pmatrix}\phi'_2\\ \phi'_2\end{pmatrix} = \begin{pmatrix}\cos f & \sin f\\ -\sin f & \cos f\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\,. \tag{3.31}$$

Let us use this *ansatz* to find a real, nonsingular YM solution with finite energy and action (Actor, 1978c). Set $A_\mu = 0$ and $\phi_1 = 0$ so that the equation of motion remaining in (3.28) is

$$r^2(\partial_0\partial_0 - \partial_1\partial_1)\phi_2 = \phi_2(1 - \phi_2^2)\,. \tag{3.32}$$

A solution of this equation is

$$\phi_2 = \pm(1/2\sqrt{y^2})(1+x^2)\,,\ \ y^2 = \tfrac{1}{4}(1+x^2)^2 + r^2\,.$$

The corresponding potential is

$$eW_0^a = 0\,,\ \ eW_i^a = \varepsilon_{ian}(x_n/r^2)[1 \pm (1+x^2)/2\sqrt{y^2}\,]\,. \tag{3.33}$$

At $r = 0$ either W_i^a vanishes or it becomes pure gauge,

$$eW_i^a = \varepsilon_{ian}2x_n/r^2 = -i\,\mathrm{Tr}[\sigma_a(\partial_i\omega)\omega^{-1}]\,,$$
$$\omega = \omega^{-1} = -\sigma\cdot\hat{r}\,, \tag{3.34}$$

so there is no singularity at $r = 0$.

The solution (3.33) is a gauge-transformed version of a well known solution due to de Alfaro, Fubini, and Furlan (1976). Choosing $\pm = -$ in Eq. (3.33) we perform the gauge transformation (3.31) with

$$f = \mp\tan^{-1}[2r/(1+x^2)]\,,$$
$$\sin f = \mp r/\sqrt{y^2}\,,\ \ \cos f = -\phi_2 = (1+x^2)/2\sqrt{y^2}$$

and obtain new *ansatz* functions

$$A_0' = \mp r x_0/y^2\,,\quad A_1' = \pm (1/2y^2)(1 + x_0^2 + r^2)\,,$$
$$\phi_1' = \pm (r/2y^2)(1 + x^2)\,,\quad \phi_2' = -1 + r^2/y^2\,.$$

The new gauge potential is

$$eW_0^a = (1/y^2)[\pm x_0 x_a]\,,$$
$$eW_i^a = (1/y^2)[\varepsilon_{ian}x_n \pm \delta_{ai}\tfrac{1}{2}(1 + x^2) \pm x_a x_i]\,. \tag{3.35}$$

This solution has been found by several different methods, as we discuss in Sec. VI. It is the continuation to Minkowski space of the meron–antimeron solution in Euclidean space-time (de Alfaro, Fubini, and Furlan, 1977).

A more general (elliptic) solution can also be obtained from *ansatz* (3.27). Choose the *ansatz* functions to be (Actor, 1978c)

$$A_\mu = -\left(-\frac{b}{a}\right)^{1/2} E\left(u\sqrt{\frac{2}{a}}\right)\partial_\mu f\,,$$
$$\phi_1 = -\left(-\frac{b}{a}\right)^{1/2} E\left(u\sqrt{\frac{2}{a}}\right)\sin f\,,\quad \phi_2 = -\cos f\,, \tag{3.36}$$

with f as above. Here $E(u)$ is an elliptic function which satisfies $E'' + aE + bE^3 = 0$ where a and b are constants (see Appendix F). The argument u in Eq. (3.36) is $u = \tan^{-1}[2x_0/(1 - x^2)]$. It is not difficult to verify that the equations of motion (3.28) are satisfied by the *ansatz* (3.36). (Note that $\Box u = \Box f = 0$, $\partial u \cdot \partial f = 0$, $\partial u \cdot \partial u = -\partial f \cdot \partial f = 1/y^2$ where the notation is two dimensional.) The Abelian gauge field strength is

$$F_{01} = \pm \frac{1}{y^2}\frac{1}{a}\sqrt{-2b}\, E'\left(u\sqrt{\frac{2}{a}}\right). \tag{3.37}$$

Gauge-transforming with f we obtain new *ansatz* functions

$$A_\mu' = -\varepsilon\partial_\mu f\,,\quad \phi_1' = -\varepsilon \sin f \cos f\,,\quad 1 + \phi_2' = \varepsilon \sin^2 f\,, \tag{3.38}$$

where

$$\varepsilon = 1 + (-b/a)^{1/2} E(u\sqrt{2/a})\,. \tag{3.39}$$

This elliptic function satisfies the equation

$$\varepsilon'' + 2\varepsilon(1 - \varepsilon)(2 - \varepsilon) = 0\,.$$

The *Ansatz* functions (3.38) lead to the gauge potential (3.35) multiplied by ε. This real elliptic solution was first given by Lüscher (1977) and Schecter (1977).

We have not yet specified E. Two real, nonsingular solutions are obtained from $E = \mathrm{sn}, \mathrm{cn}$:

$$\varepsilon = 1 \pm \left[\frac{2k^2}{1 + k^2}\right]^{1/2} \mathrm{sn}\left(\frac{u\sqrt{2}}{\sqrt{1 + k^2}},\, k\right), \tag{3.40}$$

$$\varepsilon = 1 \pm \left[\frac{2k'^2}{1 + k'^2}\right]^{1/2} \mathrm{cd}\left(\frac{u\sqrt{2}}{\sqrt{1 + k'^2}},\, k'\right). \tag{3.41}$$

In (3.40), $\varepsilon \to 1$ in the limit $k \to 0$ and we recover the de Alfaro–Fubini–Furlan solution (3.35). In (3.41), $\varepsilon \to 0$ or 2 when $k \to 0$; both of these are vacuum solutions.

Ju (1978) has used *Ansatz* (3.27) to search for *static* YM solutions. He imposes the condition of self-duality, so that the (Minkowski) solutions he finds are necessarily complex. It turns out to be possible to find SU(2) gauge symmetry breaking solutions in this way. All the solutions found by Ju are singular at $r = 0$ except for the Prasad–Sommerfield-Bogomol'ny solution which we discuss in the following section. In fact, Ju's work provides a nice demonstration of the uniqueness of the latter solution.

IV. SOLUTIONS WITH U(1) GAUGE INVARIANCE

In this section we examine classical solutions which are invariant under a U(1) subgroup of the local SU(2) gauge group. Such solutions have one long-range component, corresponding to the unbroken U(1) gauge group, with the remaining YM field components being short range. The famous 't Hooft–Polyakov monopole solution belongs to this category. In subsection B we devote many pages to the study of the 't Hooft–Polyakov monopole and its generalization, the Julia–Zee dyon. These solutions have several interesting properties: They are topological solitons, carry string-free magnetic charge, etc. One aspect that we want to discuss separately is the analogy between the Higgs mechanism in quantized YM theory and the corresponding "classical Higgs mechanism" that is used to limit the gauge symmetry of the monopole and dyon solutions to U(1). Subsection A is devoted to this analogy. In subsection B we study the individual solutions.

A. The Higgs mechanism in classical and quantum field theory

Nielson and Oleson (1973) were the first to introduce a Higgs-like mechanism into classical YM theory. Everyone knows what this is in the quantum theory. A Higgs potential "spontaneously" breaks the local gauge invariance by inducing an asymmetry in the physical vacuum. Some neutral Higgs fields ϕ_a get nonzero vacuum expectation values $\langle\phi_a\rangle \neq 0$. These constants $\langle\phi_a\rangle$ are invariant under a definite subgroup H of the full gauge group G, but not under G itself. When $\langle\phi_a\rangle$ is subtracted from ϕ_a to make the Higgs fields ϕ_a physical ($\phi_a' = \phi_a - \langle\phi_a\rangle$), and the YM theory is then rewritten in terms of these physical fields ϕ_a', it turns out that certain components of the gauge potential W_μ become massive (Higgs, 1964). These components are determined uniquely by the Higgs field–gauge field coupling. All the other components of W_μ remain massless. (The number of massless components equals the number of generators of the invariance subgroup H of the vacuum solution.) Moreover, the charged Higgs fields ϕ_b with zero vacuum expectation value $\langle\phi_b\rangle = 0$ disappear from the theory. These fields get absorbed by the massive gauge fields, the latter thereby acquiring their extra degree of freedom. We now discuss the parallel situation in classical YM theory.

Within the context of classical field theory it is natural to identify the vacuum with the two-sphere at spatial infinity. Therefore the Higgs field should be nonzero on this sphere. One introduces a Higgs potential, e.g., $(m^2/\lambda - \phi^2)^2$, and requires that it vanish in the limit $r \to \infty$ like $O(r^{-4})$. Thus $\phi^2 - m^2/\lambda \to O(1/r^2)$

as $r\to\infty$. This amounts to the introduction of a boundary condition at infinity

$$\phi_a \to (m/\sqrt{\lambda})n_a(\theta,\phi)+O(1/r), \quad r\to\infty, \tag{4.1}$$

where $n_a(\theta,\phi)$ is a unit vector. As we have discussed in Sec. II.B, such a boundary condition provides the classical solutions of the theory with a definite topological character. (For example, the monopole solutions correspond to the mapping of a sphere onto a sphere.) There seems to be nothing in the quantum theory which is directly analogous to this topological aspect of the classical theory. Nevertheless we can draw an analogy between the expectation value $\langle\phi_a\rangle\neq 0$ in quantum field theory and the nonzero value (4.1) at infinity of ϕ_a in the classical theory. In both cases the "vacuum expectation value" is not invariant under the full gauge group G, but only under a subgroup $H\subset G$. In the classical case an element $h\in H$ is a local gauge transformation which does not change the unit vector $n_a(\theta,\phi)$.

Let us turn to gauge symmetry breakdown. In the quantum theory this is manifested by certain gauge field components becoming massive. The same thing happens in the classical theory: these components of W_μ^a also become massive, i.e., they fall off exponentially in r like $\exp(-Mr)$, where M is the "mass" of the component in question. The other components (one for each generator of H) remain massless, just as in the quantum theory. All finite-energy solutions will have this behavior. (This statement is the classical equivalent of the quantum field-theoretic statement that fields become massive.) In known solutions like the monopole and dyon one finds the short-range components explicitly. Their presence in the general case can be inferred from the equations of motion. At large r these equations can be simplified by substituting the boundary value (4.1). The resulting equations will only have solutions if particular components of W_μ^a are exactly zero (i.e., they vanish faster than any power r^{-n}). These components are the massive ones. Other components of W_μ^a need only vanish like $1/r$ and they are, of course, massless.

Quantum-mechanically, the "charged" Higgs fields (those whose initial vacuum expectation values are necessarily zero) do not appear in the final theory. Classically, the same thing happens: there are no charged Higgs fields in classical solutions. This is an automatic consequence of the way in which one defines the "neutral" direction in group space. In the SU(2) example to follow we show explicitly how this works.

The Higgs mechanism provides the surviving physical Higgs field ϕ_a' with a mass. Thus one would expect that ϕ_a' will also decrease exponentially at large r in a classical solution. This is indeed the case in the monopole and dyon solutions studied below.

Let us return to the SU(2) theory (2.1) and study in more detail the workings of the classical Higgs mechanism. We consider a hypothetical static solution with $W_0\equiv 0$ and nonzero W_i, ϕ_a, assuming the $r\to\infty$ behavior

$$e\phi_a \to \beta n_a(\hat{r}), \tag{4.2}$$

$$eW_i^a \to \varepsilon_{abc}[\partial_i n_b(\hat{r})]n_c(\hat{r}), \tag{4.3}$$

where $\beta=em/\sqrt{\lambda}$ and $n_a(\hat{r})$ is a unit vector. For this solution $D_0\phi_a\equiv 0$, $G_{0j}^a\equiv 0$, while for large r

$$eD_i\phi_a \to \beta[\partial_i n_a+\varepsilon_{abc}\varepsilon_{bde}n_c n_e(\partial_i n_d)] = \beta n_a n_d \partial_i n_d = 0,$$
$$eG_{ij}^a \to O(1/r^2). \tag{4.4}$$

Equation (4.2) should be regarded as a boundary condition which determines the behavior (4.3) of the gauge potential W_i^a. Indeed, if W_i^a had any other behavior for $r\to\infty$ [except for a possible term $A_i(r)n_a(\hat{r})$ that we ignore] then the covariant derivative $D_i\phi_a$ would be $O(1/r)$ and the energy would diverge.

As we have mentioned several times previously, the boundary condition (4.2) insures that any finite-energy solution is noninvariant under the SU(2) gauge group. However, there will be a local U(1) invariance group of this solution. To make this explicit we consider a local SU(2) transformation with the form

$$\omega=\exp[\tfrac{1}{2}if(\mathbf{x})\sigma_a\hat{\phi}_a(\mathbf{x})], \tag{4.5}$$

where

$$\hat{\phi}_a\equiv\phi_a/\phi, \quad \phi^2=\phi_a\phi_a \tag{4.6}$$

is a unit vector and $f(\mathbf{x})$ is any function. For a given ϕ_a the set of transformations (4.5) closes to form a U(1) subgroup of the SU(2) gauge. This group differs from point to point in space. At infinity the boundary condition (4.2) is unchanged by the U(1) subgroup. By definition, this is the invariance gauge group of the solution in question.

The Abelian gauge potential associated with the local U(1) gauge group (4.5) is

$$A_\mu\equiv\hat{\phi}_a W_\mu^a. \tag{4.7}$$

Under a transformation (4.5) one can easily verify (using formulas in Appendix A) that A_μ transforms as an Abelian gauge potential should:

$$A_\mu\to A_\mu+(1/2e)\partial_\mu f. \tag{4.8}$$

The two components of W_μ^a orthogonal to A_μ are

$$V_\mu^{1,2}\equiv e_{1,2}^a W_\mu^a \tag{4.9}$$

where the unit vectors $e_{1,2}^a$ and $e_3^a=\hat{\phi}_a(\hat{r})$ form an orthonormal basis in group space. Under a transformation (4.5) one can show, using results in Appendix A, that $V_\mu^{1,2}$ transform like a two-dimensional representation of the local U(1) gauge group,

$$\omega\begin{pmatrix}V_\mu^1\\ V_\mu^2\end{pmatrix}\omega^{-1}=\begin{pmatrix}\cos f & \sin f\\ -\sin f & \cos f\end{pmatrix}\begin{pmatrix}V_\mu^1\\ V_\mu^2\end{pmatrix}. \tag{4.10}$$

At large r it is easy to show that the potentials A_i and $V_i^{1,2}$ behave like

$$A_i\to 0, \tag{4.11}$$

and

$$V_i^1\to e_{2a}\partial_i n_a, \quad V_i^2\to -e_{1a}\partial_i n_a. \tag{4.12}$$

We see that A_i vanishes faster than any power r^{-n} while $V_i^{1,2}$ is $O(1/r)$. This seems to contradict our claim that A_μ is the gauge potential associated with the unbroken U(1) gauge group; the long-range component of the potential is evidently in $W_i^{1,2}$ and not in A_i. Curiously, this apparent contradiction is a gauge artifact. The boundary condition (4.2) determines a rather unusual gauge in which the topology of the solution is made explicit (i.e., the map $S^2 \to S^2$). [A good name for this gauge is the topological gauge. Sometimes it is called the no-string gauge (see Sec. III.C).] In this topological gauge the Abelian gauge potential is essentially zero. (Note that for small r there is no way to distinguish the massless and massive components of the potential, so it does not matter how A_μ behaves in this region.) Hence the long-range part of the solution is necessarily in the components $V_i^{1,2}$. How can one show this explicitly? It can be done by a gauge transformation to a more conventional gauge, where A_i and $V_i^{1,2}$ have the behavior at large r that one expects. A good example was given in Sec. III.C where $n_a(\hat{r}) = \hat{r}_a$ in the topological or no-string gauge. After the gauge transformation (3.14) to the string gauge, where $n_a(\hat{r}) = \delta_{a3}$, the gauge potential has the form (3.15) with the expected behavior as $r \to \infty$. Note that this gauge transformation changes the topology of the solution. Also it mixes massless and massive components of the potential. This explains how the $1/r$ term comes to be in $V_i^{1,2}$ in Eq. (4.12).

What about the "charged" Higgs fields, which are supposed to disappear into the massive components of the gauge potential? These components are identically zero, as they should be. By definition, the Abelian gauge potential $A_\mu = \hat{\phi}_a W_\mu^a$ determines the "neutral" direction in group space, namely $\hat{\phi}_a$. Thus, trivially, ϕ_a has no charged components.

A final comment concerns the boundary condition (4.2). Clearly it is necessary to specify the function $n_a(\hat{r})$ *before* trying to solve the equations of motion. This function is a distinct field variable which has to be distinguished from W_μ^a and ϕ_a. Therefore, by imposing nonzero boundary conditions at infinity, one introduces additional field variables into the problem. Several authors (Gervais, Sakita, and Wadia, 1976; Christ, Guth, and Weinberg, 1976; Wadia, 1977) have shown how these new variables can be included in the classical canonical formalism.

B. Solutions

Now we come to the interesting solutions of SU(2) gauge theories which incorporate SU(2) gauge symmetry breakdown, with an unbroken residual U(1) gauge group.

1. Treat's *ansatz*

The first explicit example of a YM gauge theory solution with short-range potentials was given by Treat (1967). This paper, which is not very often referred to, contains results which are really quite similar to the monopole-type solutions discovered several years later. In particular, using a specific *ansatz*, Treat was able to find an explicit solution of a pure YM theory with local symmetry breaking. In this solution W_0^a is nonvanishing at infinity, and the gauge potential has one long-range and two short-range components. [One should compare this solution with the Hsu–Mac solution of the pure SU(2) theory (see below).] It is more complicated than the known monopole solutions because its two basic functions depend on θ as well as on r. We shall not go into the details of this solution because it cannot be realized within a theory whose gauge group is smaller than SU(3), and therefore it falls outside the scope of this article.

Treat also introduced a different *ansatz* which can be realized within the SU(2) framework. He was not able to find an explicit solution using this *ansatz*, but he could demonstrate the existence of short-range potentials in any solution (of the type considered) which has a long-range component. For the sake of comparison with later work we now briefly review Treat's results.

A slightly generalized version of Treat's SU(2) *ansatz* is

$$\begin{aligned} eW_0^a &= \delta_{a1} G(\mathbf{r}), \\ eW_i^a &= \delta_{a2} n_i(\mathbf{r}) F(\mathbf{r}), \end{aligned} \tag{4.13}$$

where $n_i(\mathbf{r})$ is a unit vector that we do not yet specify. The field strengths are

$$\begin{aligned} eG_{0j}^a &= -\delta_{a1}\partial_j G + \delta_{a3} n_j G F, \\ eG_{ij}^a &= \delta_{a2}[\partial_i(n_j F) - \partial_j(n_i F)]. \end{aligned} \tag{4.14}$$

This *ansatz* is successful in simplifying the SU(2) equations of motion because G_{ij}^a and W_i^a are parallel in group space. The equations of motion are easily shown to be

$$\nabla^2 G = F^2 G, \tag{4.15}$$

$$-(\hat{n}\cdot\nabla) G F = (\hat{n}\cdot\nabla G) F, \tag{4.16}$$

$$\partial_j[\partial_i(n_j F) - \partial_j(n_i F)] = n_i G^2 F. \tag{4.17}$$

For $\hat{n}$, Treat chooses the unit vector $\hat{\phi}$ in spherical polar coordinates. The functions F and G depend on (r, θ) but not on ϕ. Then the second equation of motion (4.16) is trivially satisfied. Moreover, the $\hat{\phi}$ component of the third equation (4.17) is

$$\nabla^2 F = [1/r^2 \sin^2\theta - G^2] F, \tag{4.18}$$

while the $\hat{r}$ and $\hat{\theta}$ components are trivially satisfied. (Recall that $\nabla\cdot\hat{\phi} = 0$, $\nabla\phi = \hat{\phi}/r\sin\theta$, $\nabla^2\phi = 0$ and $\nabla^2\hat{\phi} = -\hat{\phi}/r^2\sin^2\theta$.) Any solution of this type is necessarily θ dependent.

Already from Eq. (4.15) we can see that F must be short range. Assume that $G \to A/r$ for $r \to \infty$. Then since $\nabla^2(1/r) = 0$ $(r \neq 0)$ is follows that F falls off faster than any power of r, that is to say, exponentially. Thus, for any solution obtained from the *ansatz* (4.13), the mere assumption that W_0^a is a Coulomb potential forces W_i^a to be a short-range potential.

2. Prasad-Sommerfield-Bogomol'ny solution

Next we discuss the Prasad–Sommerfield–Bogomol'ny solution, which we shall first write down as a solution of the pure SU(2) gauge theory. Hsu and Mac (1977) discovered the solution in this form. We return to the *ansatz* (3.8) leading to the equations of motion (3.9). An explicit solution of the latter is (Prasad and Sommerfield, 1975; Bogomol'ny, 1976)

$$h(r) = \beta r/\sinh\beta r\,,$$
$$g(r) = -1 + \beta r\cosh\beta r/\sinh\beta r\,, \tag{4.19}$$

as the reader can easily verify. β is an arbitrary constant here.

For $\beta \neq 0$ the solution (4.19) corresponds to broken local SU(2) symmetry because

$$h \to \beta r e^{-\beta r}\,,\ g \to \beta r\,,\ i\phi_a = W_0^a \to \hat{r}_a i\beta/e\,,\quad r\to\infty\,. \tag{4.20}$$

Thus β is the mass of the two YM field components which acquire a mass through the local gauge symmetry breaking. These components are given explicitly in the string gauge by Eq. (3.15). When $\beta \to 0$ we see that $h \to 1, g \to 0$, and the solution (4.19) becomes the vacuum solution $W_\mu^a = 0$. For $\beta \neq 0$ the solution (4.19) is regular at $r=0$ ($h=1$ and $g=0$ there). This solution has topological charge $n=1$, as we see from the boundary condition $W_0^a = \hat{r}_a(i\beta/e)$ at $r=\infty$. Furthermore, the solution in the Hsu–Mac form (4.19) is *self-dual* in the sense of Eqs. (2.29) and (2.30). This is discussed in Appendix E.

In the spirit of Eq. (2.12) we can reinterpret the solution above as a solution of the SU(2) theory with a Higgs triplet. The relevant *ansatz* is (3.12), and this is the form in which the solution was originally found by Prasad and Sommerfield (1975). The equations of motion (3.13) are satisfied when $m^2=0$, $\lambda=0$ but not otherwise. We recall that the Bogomol'ny condition (2.38) corresponds to self-duality in the pure SU(2) theory. As the solution (4.19) is self-dual, it also satisfies condition (2.38). Bogomol'ny independently found the solution by imposing this condition.

Bogomol'ny's derivation shows that the energy of the solution (4.19) is less than the energy of any other solution obtained from the *ansatz* (3.12). Moreover, this minimum energy is proportional to the topological charge of the solution. We have already discussed these points in Sec. II.C.

Protogenov (1977) also rederived the solution by solving the self-duality conditions (2.29). Moreover, he noticed the following generalizations of it,

$$h(r) = \pm\,\beta r/\sinh(\beta r + \varepsilon)\,,$$
$$g(r) = -1 + \beta r\cosh(\beta r + \varepsilon)/\sinh(\beta r + \varepsilon)\,, \tag{4.21}$$

where ε is an arbitrary constant. Equations (3.9) are satisfied, and this solution is self-dual. The large-r behavior is still given by Eq. (4.20) above. However, for $\varepsilon \neq 0$ the solution (4.21) is not regular at the origin: $h=0$ and $g=-1$ at $r=0$ and therefore the potentials (3.8) have $1/r$ singularities there. Note that in the limit $\varepsilon \to \infty$ the solution (4.21) becomes $h \to 0, g \to -1+\beta r$ which is a solution already encountered in Sec. III.

3. 't Hooft-Polyakov monopole

Now we come to the famous 't Hooft–Polyakov monopole solution. 't Hooft (1974) and Polyakov (1975) independently discovered this solution of the equations of motion (3.13) for nonzero m^2 and λ. These equations follow from the *ansatz* (3.11) which we give again here:

$$W_0^a = 0\,,\ \ \phi_a = r_a g(r)/er^2\,,$$
$$W_i^a = \varepsilon_{ain} r_n[1-h(r)]/er^2\,. \tag{4.22}$$

The important properties of the solution are

(1) W_μ^a and ϕ_a are nowhere singular,

(2) the long-range component in the solution corresponds to the electromagnetic field of a static magnetic monopole,

(3) the solution has finite energy and is believed to be stable,

(4) within a class of *ansätze* more general than the one in Eq. (4.22) above, the solution is unique.

Unfortunately, the solution cannot be given in closed form when m^2 and λ are nonzero (we discuss some of the reasons for this in Appendix G), but it can be obtained numerically. In the limit $m^2 \to 0, \lambda \to 0$ with m^2/λ fixed it becomes the Prasad–Sommerfield solution (4.19) above. The boundary conditions satisfied by the solution at large r are (see Appendix G)

$r \to \infty$:

$$h(r) \to A(m,\lambda,e) r e^{-\beta r}\,,\quad \beta \equiv [em/\sqrt{\lambda}\,]^{1/2} \tag{4.23}$$
$$g(r) \to (me/\sqrt{\lambda})r + D(m,\lambda,e)e^{-\sqrt{2}mr}\,, \tag{4.24}$$
$$W_i^a \to \varepsilon_{ain} r_n/er^2\,, \tag{4.25}$$
$$\phi_a \to \hat{r}_a[(m/\sqrt{\lambda}) + (D/er)e^{-\sqrt{2}mr}]\,. \tag{4.26}$$

A and D are constant. The main characteristics of the solution at large r are

$$U(\phi) = O(e^{-\sqrt{2}mr})\,, \tag{4.27}$$
$$G_{ij}^a = O(1/r^2)\,, \tag{4.28}$$
$$D_i\phi_a = O(1/r^2)\,. \tag{4.29}$$

From Eqs. (4.27)–(4.29) we see that the energy density θ_{00} is $O(1/r^4)$ at large r and therefore this contribution to the total energy is finite. The massive components of the gauge potential fall off like $e^{-\beta r}$. Roughly speaking, these massive components cannot penetrate into the region beyond $r \sim 1/\beta$ where the Higgs field takes its constant asymptotic value. The long-range component of course penetrates the Higgs field without difficulty.

The boundary conditions at small r are (see Appendix G)

$r \to 0$:

$$h(r) \to 1 + eB(m,\lambda,e)r^2\,, \tag{4.30}$$
$$g(r) \to eC(m,\lambda,e)r^2\,, \tag{4.31}$$
$$W_i^a \to -\varepsilon_{ain} r_n B(m,\lambda,e)\,, \tag{4.32}$$
$$\phi_a \to C(m,\lambda,e)r_a\,. \tag{4.33}$$

Here B and C are constants. The main characteristic at small r is that all functions are nonsingular: ϕ_a and W_i^a vanish, while

$$U(\phi) \to m^4/4\lambda\,, \tag{4.34}$$
$$G_{ij}^a \to \varepsilon_{aij} 2\beta\,. \tag{4.35}$$

[It is interesting to note that one can reinterpret the 't Hooft–Polyakov monopole solution as a solution of the pure YM theory (2.6) with nonzero m^2 and λ. Equation

(2.12) provides the necessary connection.]

Because of its importance we shall discuss this solution in considerable detail. Certain technical points are relegated to Appendices D, E, G, and I.

First of all we show why this solution has a natural interpretation as a monpole. 't Hooft and Polyakov demonstrated this in different ways. 't Hooft's approach was to search for a suitable definition of the electromagnetic field within the theory (2.1). This definition must be invariant under SU(2) gauge transformations for obvious reasons. 't Hooft proposed that the tensor

$$F_{\mu\nu}=\frac{1}{\phi}\phi_a G^a_{\mu\nu}-\frac{1}{e}\frac{1}{\phi^3}\varepsilon_{abc}\phi_a D_\mu\phi_b D_\nu\phi_c \tag{4.36}$$

be identified with the electromagnetic field tensor. This tensor can also be written in a more transparent form (Arafune, Freund, and Goebels, 1975):

$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu-(1/e)\varepsilon_{abc}\hat{\phi}_a\partial_\mu\hat{\phi}_b\partial_\nu\hat{\phi}_c\,, \tag{4.37}$$

where

$$A_\mu\equiv\hat{\phi}_a W^a_\mu\,,\ \ \hat{\phi}_a\equiv\phi_a/\phi\,. \tag{4.38}$$

Here A_μ is the massless component of the gauge potential W^a_μ. For the *ansatz* (3.12) one easily verifies that $A_\mu=0$ (in the no-string gauge the massless potential is *identically zero*). Also, $\hat{\phi}_a=\hat{r}_a$ and so

$$\begin{aligned}F_{0i}&=0\,,\\ F_{ij}&=-(1/e)\varepsilon_{abc}\hat{r}_a\partial_i\hat{r}_b\partial_j\hat{r}_c\\ &=-(1/e)\varepsilon_{ijk}r_k/r^3\end{aligned} \tag{4.39}$$

(see Appendix E). This is the electromagnetic field of a point magnetic monopole at rest with magnetic charge

$$g=1/e\,. \tag{4.40}$$

Now the minimum magnetic charge allowed by the Dirac quantization condition (see Appendix D) is $g=1/2e$ (e is clearly the basic unit of electric charge in the theory under discussion). The 't Hooft–Polyakov monopole has twice the minimum charge.

According to 't Hooft's definition the electromagnetic field tensor depends only on $\hat{\phi}_a$ in the no-string gauge. In the string gauge things are reversed. There $\hat{\phi}_a=\delta_{a3}$ and the Higgs term in Eq. (4.37) is zero. The massless component of the gauge potential is $A_\mu=W^3_\mu$ and

$$F_{\mu\nu}=\partial_\mu W^3_\nu-\partial_\nu W^3_\mu\,. \tag{4.41}$$

In quantum field theory one usually calls this the unitary gauge.

In Appendix E we discuss 't Hooft's electromagnetic tensor in more detail. Its definition is, perhaps, somewhat too singular, and alternative definitions are possible. However, in spite of its singular nature, it is a good definition which neatly separates the long-range part of the solution from the rest. Also, it emphasizes the role played by the Higgs field boundary condition at infinity in determining the long-range potential and the unbroken local U(1) gauge group associated with the solution.

Polyakov used the following, very direct, approach. He gauge-transformed his solution to the string gauge, obtaining the potentials (3.15) appropriate for this gauge. Since the function $h(r)$ is practically zero everywhere except near $r=0$, only the Dirac string potential,

$$A^3_i=-(1/2er)\tan\tfrac{1}{2}\theta[\hat{\phi}]_i\,, \tag{4.42}$$

remains important when one leaves the vicinity of the monopole. As we discuss in Appendix D, the object corresponding to this potential is a magnetic monopole at rest with magnetic charge $1/e$ and no electric charge.

Next we want to show why the monopole has finite energy. We follow Polyakov's discussion of this point because it nicely emphasizes the different roles played by the gauge and Higgs fields.

Let us consider a Higgs triplet theory with no gauge field:

$$\begin{aligned}\mathcal{L}&=\tfrac{1}{2}\partial_\mu\phi_a\partial^\mu\phi_a-U(\phi)\,,\\ U(\phi)&=\tfrac{1}{4}\lambda(m^2/\lambda-\phi^2)^2\,.\end{aligned} \tag{4.43}$$

Making the *ansatz*

$$\phi_a=r_a\,\phi(r)/r \tag{4.44}$$

for the Higgs field one obtains the equation of motion

$$\phi''+(2/r)\phi'-(2/r^2)\phi+m^2\phi-\lambda\phi^3=0\,. \tag{4.45}$$

This equation has a solution (of course, we cannot give a closed expression for it) with the behavior

$$\phi\to(m/\sqrt{\lambda})(1-1/m^2r^2+\cdots)\,,\quad r\to\infty\,, \tag{4.46}$$

$$\phi\to r\times\text{const}\,,\quad r\to 0\,. \tag{4.47}$$

Because the nonleading term in Eq. (4.46) is $O(1/r^2)$, the potential energy density $U(\phi)$ is $O(1/r^4)$ at large r and therefore the potential energy term is not divergent at infinity. However, the kinetic energy term is divergent there. From Eq. (4.44) we find

$$\nabla\phi_a=\phi\nabla(r_a/r)+r_a\mathbf{r}\phi'/r^2\,, \tag{4.48}$$

which leads to a term

$$\tfrac{1}{2}\phi^2\nabla(r_a/r)\cdot\nabla(r_a/r)\to(3m^2/2\lambda)(1/r^2) \tag{4.49}$$

in the energy integral. This term is linearly divergent. The difficulty here is clearly caused by the direction-dependent factor r_a/r in the *ansatz* (4.44): Different directions in SU(2) space are related to different direction in three-space.

A virtue of this infinite-energy solution [which Polyakov called the "hedgehog" because of the way it looks in SU(2) space] is that it is *stable*. There are two ways to see this. The potential energy is only important for small r, while the kinetic energy is negligible there, but not for large r. This peculiar situation is caused by the unique behavior of the Higgs field, which vanishes as $r\to 0$ but not as $r\to\infty$. Thus the solution is like a spherical cavity in an infinite medium (the Higgs field). Expansion of the cavity costs potential energy in the small-r region. Contraction of the cavity costs kinetic energy in the region outside the cavity.

Another argument is that the solution seems to be *topologically* stable. The boundary value $\phi_a(r=\infty)=\hat{r}_a m/\sqrt{\lambda}$ is nothing other than a mapping of the sphere at infinity onto the unit sphere in SU(2) space. This mapping covers the latter once; hence one says that it has a winding number (or topological charge) $n=1$. [A mapping which covers the unit sphere in SU(2) space n times has a winding number or topological charge equal to n. The mapping $\phi_a(r=\infty)=\delta_{a3}m/\sqrt{\lambda}$ maps the entire sphere at infinity onto the north pole of the SU(2) sphere, and this mapping has $n=0$.] Now there exists no continuous operation which changes n. The only way to change n, say from $n=1$ to $n=0$, is to perform a discontinuous operation. [A relevant example here is the discontinuous gauge transformation (3.14) between the no-string and string gauges, which changes $n=1$ into $n=0$.] Such an operation presumably involves infinite energy in some sense. Therefore, one argues, n cannot change in the normal time development of the system. That is, a topological charge like n is conserved. In particular, the charge $n=1$ of the hedgehog is conserved. This solution is presumably a minimum-energy field configuration because of its spherical symmetry. Therefore it should be stable.

Now let us switch on the SU(2) gauge field. This has a very crucial effect: The divergence in the kinetic energy of the hedgehog is removed by making all directions in SU(2) space equivalent [for we can now perform local as well as global SU(2) transformations]. The troublesome function $\partial_i\phi_a$ in Eq. (4.48) gets replaced by $D_i\phi_a$, and we can arrange that $D_i\phi_a=O(1/r^2)$ at large r so that the kinetic term becomes finite there. We make the *ansatz*

$$W_i^a=\varepsilon_{ain}r_n[1-h(r)]/er^2$$

and then

$$\begin{aligned}D_i\phi_a&=\partial_i\phi_a+e\varepsilon_{abc}W_i^b\phi_c\\&=\partial_i[(r_a/r)\phi]+\varepsilon_{abc}\varepsilon_{bin}r_cr_n(\phi/r^3)(1-h)\\&=(r_ar_i/r^2)\phi'+h\phi 1/r(\delta_{ai}-r_ar_i/r^2)\,.\end{aligned}\tag{4.50}$$

From the large-r behavior in Eqs. (4.23)–(4.26) it is clear that $D_i\phi_a$ is well behaved at infinity.

One easily verifies that all other terms in the energy integral are well behaved at infinity if the boundary conditions (4.25) and (4.26) are satisfied. Furthermore, there is no difficulty with $r=0$, because all functions are well behaved there. The total energy is thus finite.

In the limit $m^2\to 0,\lambda\to 0$ with m^2/λ finite we have the explicit Prasad–Sommerfield solution (4.19). The energy of this solution is

$$\int d^3x\,\theta_{00}=(4\pi/e^2)M_W\,,\quad M_W=\beta\,.\tag{4.51}$$

For nonzero m^2 and λ the energy is larger,

$$\int d^3x\,\theta_{00}=(4\pi/e^2)M_WC(\lambda/e^2)\,,\quad M_W=em/\sqrt{\lambda}\,,\tag{4.52}$$

where $C(\lambda/e^2)$ is a slowly increasing function of its argument, with $C(0)=1$. Some numerical estimates are $C(0.1)=1.1$ and $C(10)=1.44$ ('t Hooft, 1974), $C(0.5)=1.18$ (Julia and Zee, 1975), and $C(\infty)=1.787$ (Bogomol'ny and Marinov, 1976). The lower bound (4.51) on the monopole mass is $137M_W$. If M_W is the mass of the charged vector boson in the weak interactions, then the monopole mass is (at least) *several thousand GeV.*

The stability of the monopole solution is, of course, very important for its physical interpretation. If this solution is really stable then *any* small change in the SU(2) potential or the Higgs field (for fixed parameters m, λ, and e) must increase the total energy of the solution. To date, no one has been able to prove that this is the case (see, however, Yoneya, 1977; Coleman, Parke, Neveu, and Sommerfield, 1977). But there are other indications of stability, as we have seen. The monopole is obtained by the usual Higgs mechanism with a nonzero Higgs potential (unlike the Prasad–Sommerfield solution). Furthermore, the magnetic charge is essentially a topological quantum number that is conserved. If there is no field configuration with the same charge and lower energy, then the monopole ought to be stable. Most workers are willing to believe in its stability. One reason for this has been the demonstration that "nearby" solutions with a comparable degree of rotational symmetry do *not* exist.

One can look for "radial excitations" of the monopole, that is, different solutions of Eqs. (3.12). In the limit $m^2\to 0,\lambda\to 0$ with m^2/λ finite, at least, there are no such solutions (Frampton, 1976). Evidently, then, if other solutions are to be found, the *ansatz* (3.11) has got to be replaced by a more general one.

The following more general *ansatz* has been investigated (Weinberg and Guth, 1976; Cremmer, Schaposnik, and Scherk, 1976; O'Raifeartaigh, 1977; Michel, O'Raifeartaigh, and Wali, 1977):

$$e\phi_a=\phi(r)n_a(\theta,\phi)\,,\quad eW_i^a=W(r)A_{ai}(\theta,\phi)\,,\tag{4.53}$$

with $W_0^a=0$. Note that the r dependence is factorized from the angular dependence, and all indices are attached to the angular functions. For a solution with finite energy, one can show that only the magnetic charges $g=0$ and $1/e$ are possible. No solution with $g\geq 2/e$ and finite energy is possible within the *ansatz* (4.53). Furthermore, the solution with $g=1/e$ is identical with the 't Hooft–Polyakov solution.

The method of proof is, in principle, quite straightforward. From the assumptions of real fields and finite energy it follows that every term in the energy integral must be finite. These are examined one by one. The most crucial requirement is that $D_i\phi_a$ vanish sufficiently rapidly as $r\to\infty$ (as we have already seen). This couples the gauge and Higgs fields.

The most general proof was constructed by Michel, O'Raifeartaigh, and Wali (1977). These authors were able to show that, for any YM theory with a compact gauge group G and a Higgs field in any real representation of G, the gauge potential will vanish outside of a fixed SU(2) subgroup of G, while inside this subgroup the gauge potential coincides with the 't Hooft–Polyakov potential.

The result of Michel, O'Raifeartaigh, and Wali just mentioned admits the possibility that the Higgs field is not in the adjoint representation, i.e., that there are more than three Higgs fields. However, if there are more than three Higgs fields present, then the extra ones are necessarily copies of the fields in the basic triplet representation of SU(2). For example, when the gauge group is SU(2) then one can obtain all integral representations from the isotensors $\phi_a\phi_b\cdots\phi_n$ constructed from ϕ_a (see, for example, Shankar, 1976). If two different isovectors ϕ_a and ϕ_a' were used, with different behavior at infinity, then the local SU(2) gauge invariance would be broken completely: There would be no unbroken U(1) gauge group and no long-range component in the gauge potential hence no monopole.

4. Julia-Zee dyon

Julia and Zee (1975) showed how to give the 't Hooft–Polyakov monopole an electric charge. (A monopole with electric charge is called a dyon, following Schwinger.) The way to do this is to change the *ansatz* (3.11) by allowing W_0^a to be nonzero:

$$W_0^a = r_a f(r)/er^2\,,$$
$$\phi_a = r_a g(r)/er^2\,,$$
$$W_i^a = \varepsilon_{ain} r_n[1-h(r)]/er^2\,. \tag{4.54}$$

Then the equations of motion (3.12) become

$$r^2 f'' = f(2h^2)\,,$$
$$r^2 g'' = g(2h^2 - m^2 r^2 + \lambda g^2/e^2)\,,$$
$$r^2 h'' = h(h^2 - 1 + g^2 - f^2)\,. \tag{4.55}$$

Here we see very clearly that when $m^2=0, \lambda=0$ the functions f and g (that is, W_0^a and ϕ_a) play essentially the same role up to a factor i. Julia and Zee were the first to comment on this fact. Note that $f(r)$ and $g(r)$ are only indirectly related through $h(r)$.

Assuming that W_0^a is nonzero, we have nonzero G_{0j}^a, and therefore an electric field in addition to the magnetic field of the monopole. Easy to state in words, this extension of the 't Hooft–Polyakov solution only becomes meaningful when one has shown that a solution exists with $f(r)\neq 0$. Fortunately this can be done. Indeed, in the limit $m^2=0, \lambda=0$ with m^2/λ finite, an explicit solution of Eqs. (4.55) is known (Prasad and Sommerfield, 1975; Bogomol'ny, 1976),

$$h(r) = \beta r/\sinh\beta r\,,$$
$$f(r) = \sinh\gamma[-1 + \beta r\cosh\beta r/\sinh\beta r]\,,$$
$$g(r) = \cosh\gamma[-1 + \beta r\cosh\beta r/\sinh\beta r]\,, \tag{4.56}$$

where β and γ are arbitrary constants. Mathematically there is little difference between this solution and the one with $\gamma=0$ discussed earlier in connection with the pure monopole. W_0^a and ϕ_a are the same function up to a constant.

The solution (4.56) can be slightly generalized in the way suggested by Protogenov (1977) for the $\gamma=0$ case:

$$h(r) = \beta r/\sinh(\beta r + \varepsilon)\,,$$
$$f(r) = \sinh\gamma[-1 + \beta r\cosh(\beta r+\varepsilon)/\sinh(\beta r+\varepsilon)]\,,$$
$$g(r) = \cosh\gamma[-1 + \beta r\cosh(\beta r+\varepsilon)/\sinh(\beta r+\varepsilon)]\,.$$

For nonzero m^2 and λ, as in the case of the monopole, one cannot solve Eqs. (4.54) in closed form. But a solution can be found by numerical methods which satisfies the boundary conditions at infinity,

$r\to\infty$:

$$h(r)\to Ar\exp[-r\sqrt{\beta^2-M^2}]\,,$$
$$f(r)\to Mr + C_1 + O(1/r)\,,\quad C_1=\text{const}\,,$$
$$g(r)\to (em/\sqrt{\lambda})r+\cdots\,, \tag{4.57}$$

where $\beta=(em/\sqrt{\lambda})^{1/2}$, and M is a new parameter with the dimension of mass. If M is real then $M<\beta$. One can also choose M to be pure imaginary, in which case $|M|$ is arbitrary. This amounts to choosing W_0^a to be pure imaginary. The boundary conditions at $r\to 0$ are the same as for the monopole (f and g behaving in the same way, of course).

The dyon mass is finite. This is easy to verify since the only new contributions to the monopole energy integral come from $G_{0j}^a[\to O(1/r^2)$ for $r\to\infty]$ and $D_0\phi_a\equiv 0$ [as W_0^a and ϕ_a are parallel in SU(2) space].

To determine the electric charge of the dyon we must first find the electric field. At large r all definitions of the electromagnetic field tensor are the same (see Appendix E) and we can use the simple definition $F_{\mu\nu}=\hat{\phi}_a G_{\mu\nu}^a$. Then the dyon electric field at large r is

$$E_n = \hat{\phi}_a G_{0n}^a$$
$$= \hat{r}_a[-\partial_n W_0^a] = -\hat{r}_a\partial_n[r_a f(r)/er^2]$$
$$= -\partial_n[f(r)/er] = C_1 r_n/er^3\,, \tag{4.58}$$

where C_1 is the unknown constant in the boundary condition (4.57) for $f(r)$. This constant has to be found numerically. The dyon electric charge Q is

$$Q = (4\pi/e)C_1\,. \tag{4.59}$$

There is no indication that this charge is quantized at the classical level.

The dyon mass is a slowly increasing function of Q. Julia and Zee give the following numerical results for $\lambda/e^2=0.5$:

Q(dyon) =	0	$44e$	$169e$
M(dyon) =	$162M_W$	$171M_W$	$253M_W$

where M_W is the mass of the W boson in the theory. Because of the weak dependence of M(dyon) on Q, it seems that the decay of a dyon with electric charge Q into a dyon with charge $Q-e$ by W-boson emission is energetically forbidden. The dyon may be stable.

The exact solution (4.56) is easily shown to be stable, for its energy is given by Eq. (2.55),

$$E = (m/\sqrt{\lambda})(g^2+q^2)^{1/2}\,. \tag{4.60}$$

Here

$$q = Q = (4\pi/e)C_1 = -(4\pi/e)\sinh\gamma \tag{4.61}$$

is the electric charge in Eq. (4.59). Comparing Eq. (2.58) with Eqs. (4.56) we see that

$$\sin\theta = q/\sqrt{g^2+q^2} = -\tanh\gamma\,,$$
$$\sinh\gamma = -q/g\,,\quad \cosh\gamma = (1/g)\sqrt{g^2+q^2}\,.$$

The magnetic charge $g=4\pi/e$ defined by Eq. (2.45) differs from the magnetic charge $g=1/e$ used throughout this section. Thus

$$E = (4\pi/e)(m/\sqrt{\lambda})\cosh\gamma = (4\pi/e^2)M_W\cosh^2\gamma\,, \tag{4.62}$$

where $M_W = em/\sqrt{\lambda}\cosh\gamma$ is the vector-meson mass in the dyon solution.

C. Remarks

Before leaving the subject of SU(2) monopole and dyon solutions there are some additional things we wish to mention.

(1) No one has yet succeeded in finding a solution which represents a monopole with magnetic charge larger than $g = 1/e$. The no-go theorems mentioned earlier compel one to work with spherically nonsymmetric *ansätze* which lead to extremely complicated coupled differential equations. All efforts to circumvent these difficulties have been unsuccessful (see, for example, Manton, 1978b).

The no-go theorems say nothing about solutions with several monopoles at different locations, of course, because such field configurations are not rotationally symmetric. One can trivially imbed an arbitrary number of Dirac monopoles in the SU(2) theory by choosing the Abelian gauge $\phi_a = \delta_{a3}$ (Arafune, Freund, and Goebel, 1975). In this gauge, A^3_μ is the imbedded Abelian potential, which can be written as a sum of individual Dirac monopole potentials. The strings of these Dirac monopoles can, in fact, be gauge-transformed away. However, one does not obtain true non-Abelian multimonopole solutions by this procedure (see Bais, 1975).

In an interesting calculation Lohe (1978) has obtained a three-monopole solution of the SU(2) theory (2.1) in the limit $m^2 \to 0$, $\lambda \to 0$, m^2/λ finite. This limit is essential for the calculation, which involves Bäcklund transformations of a known explicit self-dual solution of the pure SU(2) gauge theory. The known solution is the Prasad–Sommerfield solution (4.19) in the form (3.8) discovered by Hsu and Mac. This complex static solution can be interpreted as a self-dual solution in a three-dimensional Euclidean space. Now some rather powerful results concerning self-dual solutions in four dimensions are known from work on the instanton problem (Corrigan *et al.*, 1978). It has been shown how to transform a given self-dual solution of the pure SU(2) gauge theory in E^4 into more complicated solutions with higher topological charge. Lohe noticed that parallel manipulations in the E^3 monopole problem can be performed. By means of Bäcklund transformations it is also possible to generate multimonopole solutions from the original one-monopole solution.

The details of the calculation are too complicated for us to reproduce here. Even the end formulas are quite lengthy. Therefore we shall be content with a description. First, the monopole solution in the Hsu–Mac form is gauge-transformed into the Yang R gauge (Yang, 1977) appropriate for this three-dimensional problem. Then the Bäcklund transformation of Corrigan, Fairlie, Yates, and Goddard (Corrigan *et al.*, 1978) is performed twice to obtain a real three-monopole solution. (The two-monopole solution obtained from one Bäcklund transformation turns out to be complex.) Solutions with more monopoles can be constructed by repeating the procedure. Unfortunately, the explicit three-monopole solution is so complicated that Lohe could not say where the monopoles are located; only that they are all positioned on the z axis. The same would be true of the higher monopole solutions.

Lohe's solution has been criticized by Bruce (1978), who points out that it has infinite energy. This is not altogether surprising, for it is known that in Euclidean space-time the solutions generated from the instanton solution by the Bäcklund transformation are singular. Nevertheless this is an interesting direction of research.

(2) An exact multimonopole solution would enable one to determine the interaction between 't Hooft–Polyakov monopoles to any degree of precision. Unfortunately this exact solution is not available, and one can only study the interaction by approximation methods. It is clear that two well separated monopoles, or a monopole and an antimonopole, will interact in first approximation via the Coulomb force for point magnetic charges because the magnetic field is the only long-range field in the monopole solution. More interesting is the problem of intermediate or small monopole separation. Magruder (1978) has investigated the two-monopole interaction using trial functions that reduce to the one-monopole solution at each monopole and that satisfy other necessary boundary conditions. He obtains upper and lower bounds on the interaction energy for arbitrary separation. It turns out that the interaction energy remains comparatively small even when the separation R goes to zero. Two monopoles experience a mutual repulsion which increases as R decreases; however, the interaction energy remains finite in the limit $R \to 0$. A monopole and an antimonopole simply annihilate.

In the limit $m^2, \lambda \to 0$ (m^2/λ finite) there is a remarkable change in the interaction of 't Hooft–Polyakov monopoles (Manton, 1977). The Higgs field becomes massless, and the attractive force associated with it becomes long range. This force exactly cancels the repulsive magnetic force between like magnetic charges, and doubles the attractive force between unlike charges. Thus two Prasad–Sommerfield–Bogomol'ny monopoles do not interact with a Coulomb force if they have the same magnetic charge. (This is rather obvious when one reinterprets these monopoles as Hsu–Mac dyons, which have an imaginary electric charge.) Magruder (1978) obtains an upper bound of order $O(1/R^2)$ on the interaction energy for like charges. A more restrictive bound would be useful. Bogomol'ny's result (2.42) and (2.45) that the total energy of a static solution with topological charge N is N times the monopole mass (in the Prasad–Sommerfield–Bogomol'ny limit) suggests without proving that there may in fact be no monopole–monopole interaction at all in this case. At the same time Eq. (2.42) also shows that 't Hooft–Polyakov monopoles definitely interact, and repulsively, for any separation.

Does *any* monopole–monopole interaction survive in the limit $m^2, \lambda \to 0$ (m^2/λ finite)? So far as we know, this question has not yet been answered. The best way to answer it would be to find an exact static two- (or N-) monopole solution in the limit considered. If this solution had the energy of two (or N) single monopoles then obviously there would be no interaction. Then the situation would be analogous to the N-instanton problem discussed in Sec. VII, where the action is N times the action of a single instanton (which means that instantons do not interact with instantons). On the other hand, there is an interaction between instantons and anti-instantons, and this would be analogous to the (doubled) Coulomb force between monopole and antimonopole.

In this connection it is interesting to mention Polyakov's demonstration of confinement in the $(2+1)$-dimensional Georgi–Glashow model (Polyakov, 1977). For this or any other theory, static solutions in $(3+1)$ dimensions are instanton solutions in $(2+1)$ dimensions. Polyakov was able to show that a dilute gas of these instantons can confine quarks in the $(2+1)$-dimensional model. He does this by reexpressing the functional integral in the quantum YM theory (approximated by only considering multimonopole configurations) as the partition function of a three-dimensional Coulomb gas. The Coulomb interaction between well separated monopoles seems to play an essential role in producing confinement. Because of the success of this calculation, it was originally hoped that instantons in $(3+1)$ dimensions might also confine quarks. This hope was based on an incomplete analogy, however. As mentioned above, instantons do not interact with instantons. Monopoles (perhaps) have this same property in the $m^2, \lambda \to 0$ (m^2/λ finite) limit, when they become self-dual (i.e., when they can be reinterpreted as Hsu–Mac dyons, which are self-dual). Only in this limit do monopoles become truly analogous to instantons in $(3+1)$ dimensions. But then they lose their Coulomb interaction, and the confinement argument seems to break down (we have not checked this in detail, however).

(3) In the gauge theory (2.1) we have found a soliton solution with magnetic charge $g=1/e$. By changing the sign of the Higgs field ϕ_a in this solution [note that this has no effect on the equations of motion (3.13)] we trivially obtain a solution with magnetic charge $g=-1/e$. [This is evident in Eqs. (4.36)–(4.40).] Therefore, there are two massive monopoles $M^\pm$ with opposite charge $g=\pm 1/e$ in the soliton sector of the theory (2.1). This magnetic charge is a topological charge and therefore it is conserved. The soliton sector of the theory is classical. At present little is known about how to quantize it.

The conventional interpretation of the theory (2.1) is that, after spontaneous symmetry breakdown, it contains two massive vector mesons $W^\pm$ with charge $Q=\pm e$ and a massless photon γ. Here Q is a Noether charge which is conserved because the electromagnetic current is conserved. We refer to this as the normal or perturbative sector of the theory (2.1). In this sector quantization proceeds according to standard rules.

There exists an interesting symmetry between the soliton and normal sectors of the theory (Montonen and Olive, 1977). Assume that the photon γ belongs also to the soliton sector. Then an SU(2) triplet (M^+, γ, M^-) can be constructed, and the theory in the soliton sector looks very much like the theory in the normal sector *if the monopole has spin one*. [The spin of the SU(2) monopole is unknown.] Note that the photon couples in exactly the same way to electric charge in the normal sector and to magnetic charge in the soliton sector. We summarize the situation as follows:

	Normal sector	Soliton sector
Particle	$W^\pm, \gamma, \sigma$	$M^\pm, \gamma, \sigma$
Charge	$\pm e, 0, 0$	$\pm \frac{1}{e}, 0, 0$
Mass	$M_W, 0, M_H$	$M_m, 0, M_H$

where σ is the Higgs field which remains after spontaneous symmetry breakdown. We have assumed that the Higgs field, like γ, plays the same role in both sectors of the theory. The masses are

$$M_W = em/\sqrt{\lambda}, \quad M_H = \sqrt{2}m,$$
$$M_m = (4\pi/e^2)M_W C(\lambda/e^2) = (4\pi/\sqrt{\lambda})gmC(\lambda g^2),$$

where $C(0)=1$ and $C(\lambda/e^2)$ is a slowly increasing function of its argument. Note that when the symmetry breaking is switched off ($m^2/\lambda \to 0$) both of the masses M_W and M_m vanish.

Montonen and Olive conjecture that if one knew how to quantize the soliton sector of the theory, the result would coincide with the normal quantized theory, but with $M^\pm$ playing the role of the massive vector boson $W^\pm$. Moreover, the SU(2) coupling would be $g=1/e$ and not e. The operator Lagrangian would be the one in Eq. (2.1) but with these changes. Essentially, the conjecture is that a kind of duality exists at the quantum level between the normal and soliton sectors of the theory. The duality operation is

$$W^\pm \leftrightarrow M^\pm, \quad e \leftrightarrow g = 1/e,$$

in which Noether and topological charges exchange roles. In the "dual" theory, where the monopole is elementary, the vector bosons $W^\pm$ are 't Hooft–Polyakov solitons.

This conjecture is supposed to hold in the limit $m^2 \to 0, \lambda \to 0$ with m^2/λ finite. Then the monopole mass is $M_m = 4\pi gm/\sqrt{\lambda}$ which (except for the 4π) is the mass obtained from $M_W = em/\sqrt{\lambda}$ by the replacement $e \to g = 1/e$. (The 4π is irrelevant; it can be absorbed by changing the units of electric charge.) For $\lambda \neq 0$ the constant $C(\lambda/e^2) > 1$ may spoil the argument. Montonen and Olive also cite Manton's result as evidence for their conjecture. Manton (1977) found that Prasad–Sommerfield–Bogomol'ny monopoles with like charge do not interact at long range because the effects of the massless Higgs field exactly cancel the repulsive magnetic force. This is, of course, just what one would have in quantum field theory.

We mention the Montonen–Olive conjecture because it nicely emphasizes certain global aspects of the SU(2) monopole problem. However, it is by no means clear that the monopole has spin one. On the contrary, a recent calculation of the zero modes of a Prasad–Sommerfield–Bogomol'ny monopole (Mottola, 1978) implies that the spin is, in fact, zero.

(4) If one insists on the electromagnetic interpretation of the SU(2) monopole and dyon solutions, then it seems unlikely that these particular solutions are realized in nature. The reason is that the SU(2) gauge group does not play a fundamental role in the weak and electromagnetic interactions of quarks and leptons. As

everyone knows, the smallest gauge group which is compatible with the existing data is the $SU(2)\times U(1)$ gauge group of the Weinberg–Salam model. Although extremely successful, this model with its two independent coupling constants is not expected to survive indefinitely as the best unified theory of weak and electromagnetic interactions. A larger compact group G should eventually emerge as the fundamental gauge group. Whichever group this is, it will have monopole and dyon solutions. [General topological arguments in support of this statement are given in Appendix I. Here we mean new solutions—and not merely imbedded SU(2) solutions—which belong to the group G.] The SU(2) solutions are interesting prototypes of these (possibly physical) more complicated solutions.

The reader may be curious to know if the Weinberg–Salam model has a monopole solution. The answer is no; there is no topological charge in this model. A Higgs doublet $\phi=(\phi_1,\phi_2)$ induces spontaneous symmetry breakdown in the Weinberg–Salam model; at infinity the boundary condition is

$$|\phi_1|^2+|\phi_2|^2=\text{const}.$$

Since ϕ_1 and ϕ_2 are complex, this is essentially the statement that a vector in a four-dimensional Euclidean space has fixed length. The tip of this vector traces out a sphere S^3. Because the map $S^2\to S^3$ has trivial topological structure [i.e., $\pi_2(S^3)=0$] there is no topological charge.

One can search for a nontopological monopole solution (Hsu, 1976). There exists such a solution in the Weinburg–Salam model which is algebraically very similar to the SU(2) monopole. The monopole charge turns out to be

$$g_m=(1/e)\sin^2\theta_W;\ \ \sin^2\theta_W=g'^2(g^2+g'^2)^{-1},$$

where g and g' are the couplings of the SU(2) and U(1) gauge groups, respectively, and e is the electric charge. It is not surprising that g_m depends on the mixing angle θ_W, because this angle determines the mixture of the initial SU(2) and U(1) gauge groups in the final unbroken U(1) gauge group. Note that the present experimental value $\sin^2\theta_W\approx\frac{1}{4}$ leads to a magnetic charge $g_m\approx\frac{1}{4}e$ which is one-half of the minimum Dirac unit $g_m=\frac{1}{2}e$. This indicates that the solution is unphysical.

(5) Certain (dyon) solutions of the SU(2) gauge theory with one Higgs triplet are mathematically equivalent to static solutions of a larger gauge theory with two Higgs triplets. Here, equivalence in the sense of Eq. (2.12) is meant. To demonstrate this we now derive a generalized version of the equivalence (2.12). Then we show how the dyon solutions can be reinterpreted as a monopole solution with two Higgs fields.

We base our discussion on the following Lagrangian, which is a generalization of (2.1) and (2.6):

$$\begin{aligned}\mathcal{L}=&-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a+\frac{1}{2}D_\mu\phi_a D^\mu\phi_a+\frac{1}{2}D_\mu\psi_a D^\mu\psi_a\\&-\frac{\lambda}{4}\left(\frac{m^2}{\lambda}-\psi^2\right)^2-\frac{\lambda'}{4}\left(\frac{m'^2}{\lambda'}-\phi^2+W^2\right)^2\\&-\frac{1}{2}e^2[\phi^2\psi^2-(\phi_a\psi_a)^2].\end{aligned}\tag{4.63}$$

Note the coupling of ϕ_a and ψ_a in the last term. For this term to vanish, in the vacuum solution for example, ϕ_a and ψ_a have to have the same direction in group space. Therefore in the theory (4.63) the vacuum is characterized by a single constant vector v_a in SU(2) space. This implies that local SU(2) symmetry can be broken to U(1), but not completely broken.

Again we consider two types of static solution, an unprimed one with fields

$$\phi_a\equiv 0,\ \ W^a_0, W^a_i,\ \text{and}\ \psi_a\neq 0\ ,\tag{4.64}$$

and a primed one with fields

$$W'^a_0\equiv 0,\ \ W'^a_i,\ \phi'_a,\ \text{and}\ \psi'_a\neq 0.\tag{4.65}$$

These two solutions are mathematically identical if we make the identification

$$W^a_0=i\phi'_a,\ \ W^a_i=W'^a_i,\ \ \psi_a=\psi'_a.\tag{4.66}$$

Proof: From Eq. (4.66) we find

$$G^a_{0j}=-iD'_j\phi'_a,\tag{4.67}$$

$$D_0\psi_a=ie\varepsilon_{abc}\phi'_b\psi'_c.\tag{4.68}$$

The Lagrangian for an unprimed solution of the type (4.64) can therefore be rewritten

$$\begin{aligned}\mathcal{L}=&-\frac{1}{4}G'^a_{ij}G'^{ij}_a+\frac{1}{2}D'_j\phi'_aD'^j\phi'_a+\frac{1}{2}D'_j\psi'_aD'^j\psi'_a\\&-\frac{\lambda}{4}\left(\frac{m^2}{\lambda}-\psi'^2\right)^2-\frac{\lambda'}{4}\left(\frac{m'^2}{\lambda'}-\phi'^2+W'^2\right)^2\\&-\frac{1}{2}e^2\varepsilon_{abc}\varepsilon_{amn}\phi'_b\psi'_c\phi'_m\psi'_n.\end{aligned}\tag{4.69}$$

It is easy to see that this Lagrangian is the same as the one obtained from Eq. (4.63) for a static solution of the primed type (4.65). This completes the proof.

In the dyon *ansatz* (4.54) we can replace $f(r)$ by $if(r)$. Then from Eq. (4.66) we find the corresponding *ansatz* for the theory with two Higgs triplets,

$$\begin{aligned}&\phi_a=r_af(r)/er^2,\ \ \psi_a=r_ag(r)/er^2,\\&W^a_0=0,\ \ W^a_i=\varepsilon_{ain}r_n[1-h(r)]/er^2.\end{aligned}\tag{4.70}$$

Choosing $\lambda'=0$ in the Lagrangian (4.63) it is easy to see that the equations of motion for this theory are reduced by the above *ansatz* to Eqs. (4.55), except for a sign change in the third equation,

$$r^2h''=h(h^2-1+g^2+f^2).$$

In the usual limit $m^2=0, \lambda=0, m^2/\lambda$ finite one finds from Eq. (4.56) the exact monopole solution

$$\begin{aligned}&h(r)=\beta r/\sinh\beta r\ ,\\&f(r)=\sin\gamma(-1+\beta r\coth\beta r),\\&g(r)=\cos\gamma(-1+\beta r\coth\beta r).\end{aligned}\tag{4.71}$$

Here γ is a mixing angle which determines the relative contributions of the two (practically identical) Higgs fields. One can switch off $\phi_a(\psi_a)$ by choosing $\gamma=0(\pi/2)$. Of course, when ψ_a and ϕ_a are required to satisfy specific boundary conditions at infinity, say

$$\psi\to m/\sqrt{\lambda},\ \ \phi\to m'/\sqrt{\lambda'}\ ,$$

then γ is determined,

$\tan\gamma = (m'/m)\sqrt{\lambda/\lambda'}$.

The essential features of the monopole solution (4.71) are clear. It has one unit of topological charge, coming from either Higgs field. Because $\phi_a = \tan\gamma\psi_a$ there is no ambiguity in the definition of $F_{\mu\nu}$. The total energy is finite.

V. SOLUTIONS WITH NO RESIDUAL GAUGE INVARIANCE

To gain a better perspective on the monopole and dyon solutions of YM theories we shall make a digression in this section. Our topic here is the complete breakdown of local SU(2) gauge invariance. Unfortunately, no interesting solution of a completely broken SU(2) gauge theory is known, and so our discussion is essentially a qualitative one. For the pure SU(2) gauge theory, however, some interesting mathematical results are known which limit the range of possible solutions. More generally, it is possible to make definite statements concerning the type of boundary conditions at infinity that are needed to achieve complete local gauge symmetry violation. In many cases it can be shown that these boundary conditions necessarily lead to infinite-energy solutions. The comments to follow summarize what is presently known about these aspects of SU(2) gauge theories.

(1) Suppose that there exists a static solution of the pure SU(2) YM theory with no residual local gauge invariance. This solution would have only short-range potentials; it would correspond to an object localized in space and constructed entirely from gauge fields—a classical "glueball." [One could, of course, also speculate on the existence of a classical glueball with a long-range component like the SU(2) monopole.] For a particle interpretation of the glueball to be possible the solution should have finite mass. But this property brings it into conflict with the following theorem (Coleman, 1975; Deser, 1975; Pagels, 1977):

Theorem:

There exists no static, real, finite-energy solution of the pure SU(2) gauge theory other than the vacuum solution. We give Pagels simple derivation of this result.

Recall that the energy–momentum tensor $\theta_{\mu\nu}$ for the SU(2) theory satisfies

$$\partial^\nu \theta_{\mu\nu} = 0, \quad \theta_{00} = \sum \theta_{ii} \geq 0 .$$

Assume that the energy

$$E = \int d^3x\, \theta_{00} \tag{5.1}$$

is finite and constant. This implies that

$$r^3 \theta_{\mu\nu} \to 0, \quad r \to \infty . \tag{5.2}$$

Now consider the quantity

$$D = \int d^3x\, x^\mu \theta_{\mu 0} = \int d^3x \left[x_0\theta_{00} - x_i\theta_{i0}\right] , \tag{5.3}$$

which is constant under the above assumptions;

$$\begin{aligned}\partial_0 D &= \int d^3x \left[\theta_{00} - x_i \partial_0 \theta_{i0}\right] \\ &= \int d^3x \left[\theta_{00} + x_i \partial_j \theta_{ij}\right] \\ &= \int d^3x \left[\theta_{00} + \partial_j (x_i \theta_{ij}) - \theta_{ii}\right] \\ &= \int d^3x\, \partial_j (x_i \theta_{ij}) = 0 ,\end{aligned} \tag{5.4}$$

by Gauss's theorem. If we make one more assumption, namely that θ_{i0} is independent of x_0 as would be the case for a static solution, then it follows immediately that the energy $E = 0$. The only real solution with $E = 0$ is the vacuum solution. Therefore, there are *no (static) classical glueballs*.

This proof says nothing about the existence of complex solutions, of course. Indeed, in Sec. IV we have discussed an explicit static, complex solution of the pure SU(2) gauge theory which has $E = 0$ because it is self-dual.

(2) The existence of time-dependent solutions with finite energy has been investigated by several authors (Coleman, 1977a; Weder, 1977; Magg, 1978). Coleman proved the following theorem.

Theorem:

The only finite-energy, real solution of the SU(2) gauge theory which satisfies

$$\lim_{r\to\infty} r^{3/2+\varepsilon} G^a_{\mu\nu} = 0 \quad (\varepsilon > 0) \tag{5.5}$$

uniformly in r and t (with $t > 0$) is the vacuum solution. The condition (5.5) is the requirement that no energy be radiated out to spatial infinity. Coleman has shown that the only solution of the pure SU(2) gauge theory with this property is the vacuum. Thus any solution with finite energy will, eventually, radiate its energy away to infinity.

Weder (1977) proved a more general version of Coleman's result.

Theorem:

There is no real, finite-energy solution of the pure SU(2) gauge theory for which there exists $\varepsilon, R, T > 0$ such that

$$E_R(x_0) \equiv \int_{r\leq R} d^3x\, \theta_{00} \geq \varepsilon, \quad \forall x_0 > T . \tag{5.6}$$

In words, this theorem states the following. For arbitrarily large x_0 the energy $E_R(x_0)$ inside the sphere $r \leq R$ *cannot* remain larger than the (arbitrarily small) number ε. Therefore, energy must be radiated outside the sphere, i.e., out to spatial infinity.

An explicit solution of the pure SU(2) theory which does radiate energy out to infinity, and which has finite energy, will be discussed in the next section. This is the de Alfaro–Fubini–Furlan solution in a real gauge. At $x_0 = -\infty$ the energy is distributed on the sphere at infinity. The fields move inward as x_0 becomes finite, until the energy is concentrated about the origin at time $x_0 \approx 0$. For $x_0 > 0$ everything proceeds in reverse. This is precisely the type of solution which is allowed by the preceding theorems. There is no localization; the boundary conditions at infinity are nonstatic; there is no indication of local gauge symmetry breakdown. Some long-range effect is evidently responsible for drawing the energy at infinity into a finite region about $r=0$ and then forcing it back out.

(3) In the pure SU(2) theory it is possible to completely break the local SU(2) gauge symmetry by introducing boundary conditions such as

$$W_0^a \to \beta n_a(\hat{r}), \quad r\to\infty, \qquad W_3^a \to \beta' n_a'(\hat{r}), \quad r\to\infty, \tag{5.7}$$

with $W_{1,2}^a \to 0$. Here $n_a(\hat{r})$ are two *different* unit vectors, each of which determines a local U(1) subgroup which would be unbroken if the other were not present. These two subgroups are incompatible, of course, and so no subgroup remains unbroken. Note that the boundary conditions (5.7) are quite unphysical within the context of a three-dimensional problem, because the field strength G_{03} does not vanish at infinity (unless $n_a = n_a'$),

$$G_{03}^a \to e\beta\beta'\varepsilon_{abc}n_b(\hat{r})n_c'(\hat{r}). \tag{5.8}$$

Therefore, boundary conditions such as (2.5) lead to badly divergent total energy. We see that it is not possible to arrange for complete symmetry breakdown in the pure SU(2) gauge theory.

(4) In the SU(2) theory with a Higgs triplet one can arrange for complete breakdown of local gauge invariance by imposing the following boundary conditions

$$W_0^a \to \beta n_a(\hat{r}), \quad r\to\infty, \qquad \phi_a \to \beta' n_a'(\hat{r}), \quad r\to\infty, \tag{5.9}$$

with $W_i^a \to 0$ as usual. When $n_a(\hat{r}) \neq n_a'(\hat{r})$ then no residual U(1) local subgroup is left unbroken. [If $n_a = n_a'$, as in the Julia–Zee dyon solution where $n_a = n_a' = \hat{r}_{a'}$ then a local U(1) subgroup is still unbroken.] The boundary conditions (5.9) lead to infinite energy, of course, because $D_0\phi_a$ does not vanish at infinity (for $n_a \neq n_a'$):

$$D_0\phi_a \to e\beta\beta'\varepsilon_{abc}n_b(\hat{r})n_c'(\hat{r}). \tag{5.10}$$

This is very much like the situation in the pure SU(2) theory [Eq. (5.8)]. We see that complete local gauge symmetry breakdown is not possible in the SU(2) theory with one Higgs triplet.

(5) Is it possible to spontaneously break the local SU(2) gauge symmetry completely in any SU(2) theory? The answer is no; at least, not in the type of theory we are considering here with an SU(2) gauge field and some number of Higgs triplets. Suppose we consider a theory with two uncoupled Higgs triplets ϕ_a and ψ_a, which have the boundary values at infinity

$$\phi_a \to \beta n_a(\hat{r}), \quad r\to\infty, \qquad \psi_a \to \beta' n_a'(\hat{r}), \quad r\to\infty. \tag{5.11}$$

For $n_a \neq n_a'$ there is no unbroken U(1) subgroup. But we again run into conflict with a nonvanishing energy density at $r=\infty$. The reason is that with a single gauge field W_μ^a one cannot arrange for both $D_\mu\phi_a$ and $D_\mu\psi_a$ to be $O(1/r^2)$ as $r\to\infty$. (Recall our discussion in Sec. IV.C.) This is only possible when $n_a = n_a'$.

VI. CONNECTION BETWEEN YANG-MILLS THEORY AND ϕ^4 THEORY

There exists a useful and interesting connection between the SU(2) YM theory and the scalar ϕ^4 theory. This connection is a specific *ansatz* for the YM potential W_μ^a in terms of a scalar field ϕ [see Eq. (6.1) below]. This *ansatz* reduces the complicated equations of motion of the YM theory to the single equation of motion for the ϕ^4 theory. Therefore one can find explicit solutions of the SU(2) theory by solving the much simpler scalar theory. In this section we shall review the known *Minkowski-space* solutions of this type. There is also a *Euclidean space-time* version of *ansatz* (6.1); the Euclidean solutions obtained from it are discussed in the following section. The *ansatz* was discovered by 't Hooft (1977b, c) in connection with the instanton problem [see also Corrigan and Fairlie (1977) and Wilczek (1977)]. Uy (1976) had previously given a solution of the pure SU(2) theory which corresponds to $\phi = xyz + C$. Rewriting Uy's solution in covariant form one is led immediately to the static version of the *ansatz*.

A. *Ansatz* for the Yang-Mills potential

In Minkowski space the *ansatz* for the SU(2) gauge potential W_μ^a is

$$eW_0^a = \pm i\partial_a\phi/\phi, \qquad eW_i^a = \varepsilon_{ian}\partial_n\phi/\phi \pm i\delta_{ai}\partial_0\phi/\phi, \tag{6.1}$$

where ϕ is a Lorentz scalar function which we assume has the dimension L^{-1} of a physical field. Equation (6.1) can also be written in the form

$$W_\mu^a = \eta_{a\mu\nu}\partial^\nu\phi/\phi, \tag{6.2}$$

where

$$\eta_{a\mu\nu} = \varepsilon_{0a\mu\nu} \mp ig_{a\mu}g_{\nu 0} \pm ig_{a\nu}g_{\mu 0} \tag{6.3}$$

is the Minkowski-space version of a tensor introduced by 't Hooft for the Euclidean problem. In Eq. (6.1) the potentials W_μ^a are complex for real ϕ. This is an unfortunate property of the *ansatz*, as one would like to find real solutions of the SU(2) Yang–Mills theory. But complex solutions are also interesting, and furthermore there exists the possibility that for a particular solution ϕ the SU(2) potentials (6.1) can be made real by a suitable *complex* SU(2) gauge transformation. For arbitrary ϕ this is not possible. But we shall discuss an explicit solution for which this is possible, and we will obtain the corresponding explicit, real solution of the pure SU(2) gauge theory.

The *ansatz* (6.1) is useful for the following reason. It reduces the equation of motion

$$\partial^\nu G_{\mu\nu}^a = e\varepsilon_{abc}G_{\mu\nu}^b W_c^\nu \tag{6.4}$$

for the pure SU(2) gauge theory to the much simpler equation (Corrigan and Fairlie, 1977; Wilczek, 1977)

$$(1/\phi)\partial_\mu\Box\phi = (3/\phi^2)\partial_\mu\phi\Box\phi. \tag{6.5}$$

Equation (6.5) can be integrated once to give

$$\Box\phi + \lambda\phi^3 = 0, \tag{6.6}$$

where λ is an arbitrary integration constant. Solutions of Eq. (6.6) are known, and in Eq. (6.1) these lead automatically to explicit solutions of the SU(2) gauge theory.

The *ansatz* (6.1) is often written with $\partial_\mu \phi/\phi$ replaced by $\partial_\mu \ln\phi$. We prefer $\partial_\mu\phi/\phi$ because we think that ϕ can be interpreted as a physical field, and not merely as an *ansatz* function. When one considers the extension to massive fields, it becomes apparent that ϕ has a natural interpretation as a Higgs-like field.

Let the SU(2) field have real mass m. Then the YM equations of motion are

$$\partial^\nu G^a_{\mu\nu} = e\varepsilon_{abc}G^b_{\mu\nu}W^\nu_c + m^2 W^a_\mu . \tag{6.7}$$

These are reduced by the *ansatz* (6.1) to (Actor, 1978a)

$$(1/\phi)(\Box + m^2)\partial_\mu\phi = (3/\phi^2)\partial_\mu\phi\Box\phi , \tag{6.8}$$

which in turn is satisfied if ϕ satisfies

$$\Box\phi - \tfrac{1}{2}m^2\phi + \lambda\phi^3 = 0 . \tag{6.9}$$

Now if the vector-meson mass m is real, then the mass term in the scalar equation (6.9) has the wrong sign. Equation (6.9) is the equation of motion for the ϕ^4 theory with spontaneous symmetry breakdown, whose Lagrangian is

$$\mathfrak{L} = \tfrac{1}{2}\partial^\mu\phi\partial_\mu\phi - \tfrac{1}{4}\lambda(\phi^2 - m^2/2\lambda)^2 . \tag{6.10}$$

The vacuum solution of this theory is $\phi = \pm m/\sqrt{2\lambda}$, and the corresponding YM field in Eq. (6.1) is $W^a_\mu = 0$.

We now give formulas for several quantities of interest, following from the *ansatz* (6.1). The YM field strengths are

$$\begin{aligned} eE^a_n \equiv eG^a_{0n} = \varepsilon_{nam}\left[\frac{1}{\phi}\partial_0\partial_m\phi - \frac{2}{\phi^2}\partial_0\phi\partial_m\phi\right] \\ \pm i\delta_{na}\left[\frac{1}{\phi}\partial_0^2\phi - \frac{1}{\phi^2}(\partial_0\phi\partial_0\phi + \partial_m\phi\partial_m\phi)\right] \\ \mp i\left[\frac{1}{\phi}\partial_n\partial_a\phi - \frac{2}{\phi^2}\partial_n\phi\partial_a\phi\right], \end{aligned} \tag{6.11}$$

$$eB^a_n = -\tfrac{1}{2}e\varepsilon_{nij}G^a_{ij} = \pm ieE^a_n + \delta_{an}(1/\phi)\Box\phi . \tag{6.12}$$

The self-duality condition $B^a_n = \pm iE^a_n$ evidently implies $\Box\phi = 0$, or $m^2 = 0$, $\lambda = 0$ in Eq. (6.9). The field strengths E^a_n and B^a_n are in general complex. However (remarkably enough) their squares are both real, and this means that the energy and Lagrangian densities obtained from the *ansatz* (6.1) are real, even though the potential W^a_μ is complex (see, for example, Bernreuther, 1977).

The Lagrangian density [see Eq. (2.37)] is given explicitly by

$$\begin{aligned} 2e^2\mathfrak{L} &= \frac{2}{\phi^2}[-\tfrac{1}{2}\Box\phi\Box\phi - \partial^\alpha\partial^\beta\phi\partial_\alpha\partial_\beta\phi] \\ &+ \frac{2}{\phi^3}[-\Box\phi\partial^\alpha\phi\partial_\alpha\phi + 4\partial^\alpha\partial^\beta\phi\partial_\alpha\phi\partial_\beta\phi] \\ &+ \frac{2}{\phi^4}[-3\partial^\alpha\phi\partial_\alpha\phi\partial^\beta\phi\partial_\beta\phi] \\ &= \Box\partial_\alpha(\partial^\alpha\phi/\phi) - \frac{1}{\phi}\Box\Box\phi \\ &+ \frac{4}{\phi^2}\partial^\alpha\phi\partial_\alpha\Box\phi - \frac{6}{\phi^3}\Box\phi\partial^\alpha\phi\partial_\alpha\phi . \end{aligned} \tag{6.13}$$

For $m^2 = 0$ we have

$$2e^2\mathfrak{L} = \Box\partial_\alpha(\partial^\alpha\phi/\phi) - 3\lambda^2\phi^4 . \tag{6.14}$$

Next, consider the pseudoscalar density $D(x)$ defined by

$$D(x) = -\tfrac{1}{4}iG^a_{\mu\nu}\tilde{G}^{\mu\nu}_a = -iE^a_n B^a_n . \tag{6.15}$$

$D(x)$ is simply related to $\mathfrak{L}(x)$ as we see from Eq. (6.12),

$$e^2(\mathfrak{L} \mp D) = \tfrac{1}{2}e^2(B^a_n - iE^a_n)^2 = \tfrac{3}{2}(\Box\phi/\phi)^2 . \tag{6.16}$$

In fact, for a self-dual solution $D = \pm\mathfrak{L}$.

Another quantity of primary interest is the energy–momentum tensor $\theta_{\mu\nu}(x)$ [see Eqs. (2.33)–(2.36)]. This is easily shown to be

$$e^2\theta_{\mu\nu} = \frac{\Box\phi}{\phi}\left\{\frac{4}{\phi^2}\partial_\mu\phi\partial_\nu\phi - \frac{2}{\phi}\partial_\mu\partial_\nu\phi + g_{\mu\nu}\left[\frac{1}{2\phi}\Box\phi - \frac{1}{\phi^2}\partial^\alpha\phi\partial_\alpha\phi\right]\right\}. \tag{6.17}$$

A self-dual solution has $\theta_{\mu\nu} = 0$ because $\Box\phi = 0$.

The total energy for the case $m^2 = 0$ is

$$\begin{aligned} e^2E &= \int d^3x\, e^2\theta_{00} \\ &= -6\lambda\int d^3x\,[\tfrac{1}{2}(\partial_0\phi)^2 + \tfrac{1}{2}(\nabla\phi)^2 + \tfrac{1}{4}\lambda\phi^4] , \end{aligned} \tag{6.18}$$

where we have neglected surface terms at infinity. Note that $E = -(6\lambda/e^2)E_s$, where

$$E_s = \int d^3x\,[\tfrac{1}{2}(\partial_0\phi)^2 + \tfrac{1}{2}(\nabla\phi)^2 + \tfrac{1}{4}\lambda\phi^4]$$

is the total energy for the scalar theory (6.10) with $m^2 = 0$. For real ϕ we see that $E > 0$ implies $\lambda < 0$.

One final comment: Eq. (2.12) tells us how to change any static solution of the pure SU(2) gauge theory into a static solution of the theory with a Higgs triplet. Therefore, if we make the *ansatz* (note that it is real)

$$eW^a_i = \varepsilon_{ian}\partial_n\psi/\psi, \quad eW^a_0 = 0, \quad e\phi_a = \partial_a\psi/\psi ,$$

for the latter theory then the equations of motion (3.12) (with $m^2 = 0$, $\lambda = 0$ of course) are satisfied by any ψ which is a solution of

$$-\nabla^2\psi + \gamma\psi^3 = 0 .$$

[To avoid a collision of notation we have renamed the *ansatz* function in (6.1) ψ, and the dimensionless constant in Eq. (6.6) γ.]

We now come to the known explicit solutions of Eq. (6.6) and the corresponding solutions of the unbroken YM theory. No explicit solutions of Eq. (6.9) are known to the author except for the rather trivial "plane-wave" type of solution.

B. de Alfaro-Fubini-Furlan solution

A solution of Eq. (6.6) which we shall discuss in the next section is

$$\phi = \left[\frac{(u-v)^2}{\lambda(x-u)^2(x-v)^2}\right]^{1/2} , \tag{6.19}$$

where u and v are constant four-vectors. This solution is singular at the points $x = u, v$. However, as remarked by de Alfaro, Fubini, and Furlan (1976), these singularities lie outside physical space if u and v are not real. Suppose that we choose $u=-v=(i,0)$. Then the solution (6.19) becomes

$$\phi=(-4/\lambda)^{1/2}[(1-x^2)^2+4x_0^2]^{-1/2}$$
$$=(-4/\lambda)^{1/2}\{[1+(x_0+r)^2][1+(x_0-r)^2]\}^{-1/2}. \quad (6.20)$$

This solution of Eq. (6.6), which is *nonsingular* everywhere, was first given by Castell (1972). de Alfaro, Fubini, and Furlan (hereafter DFF) recognized that a corresponding nonsingular solution of the YM theory exists. Now this is perfectly obvious when one is aware of *ansatz* (6.1). But DFF did not have this advantage, as the *ansatz* had not been discovered when they wrote their paper. Their construction of the solution involves conformal group manipulations which we shall not reproduce here. Instead, we shall discuss the nonsingular YM solution within the context of the *ansatz* (6.1).

For simplicity, we choose + in Eq. (6.1). Then the YM potential obtained from ϕ in Eq. (6.19) is

$$eW_0^a=-i\frac{2x_a(1-x^2)}{(1-x^2)^2+4x_0^2},$$
$$eW_i^a=-\varepsilon_{ian}\frac{2x_n(1-x^2)}{(1-x^2)^2+4x_0^2}-i\delta_{ai}\frac{2x_0(1+x^2)}{(1-x^2)^2+4x_0^2}. \quad (6.21)$$

For the field strengths we find

$$eE_n^a=4x_0\{\varepsilon_{nam}2x_m(1+x_0^2+r^2)+i4x_0(-\delta_{na}r^2+x_n x_a)\}[(1-x^2)^2+4x_0^2]^{-2}, \quad (6.22)$$

with B_n^a given by Eq. (6.12). The square of the electric field is real:

$$e^2E_n^aE_n^a=128x_0^2r^2/[(1-x^2)^2+4x_0^2]^2. \quad (6.23)$$

One easily verifies that $E_n^aB_n^a=0$, and therefore $D(x)=0$. From Eq. (6.16) we find that

$$e^2\mathcal{L}=24[(1-x^2)^2+4x_0^2]^{-2}. \quad (6.24)$$

The energy density is

$$e^2\theta_{00}=8[(1-x^2)^2+4x_0^2]^{-2}\left[\frac{4(1+x_0^2+r^2)^2}{(1-x^2)^2+4x_0^2}-1\right]. \quad (6.25)$$

At time $x_0=0$ this simplifies to

$$e^2\theta_{00}(x_0=0)=24/(1+r^2)^4. \quad (6.26)$$

The total energy is $E=3\pi^2/e^2$.

In the limit $x_0\to\pm\infty$ we note that ϕ behaves like

$$\phi\to(-4/\lambda)^{1/2}/x_0^2,\quad r \text{ finite},$$
$$\phi\to(-4/\lambda)^{1/2}/2r,\quad r=|x_0|.$$

The general space-time development of ϕ is as follows. At time $x_0=-\infty$, ϕ is distributed thinly and isotropically at spatial infinity. For finite x_0 the field moves inwards, until at time $x_0=0$ it is concentrated about the origin,

$$\phi(x_0=0)=(-4/\lambda)^{1/2}[1/(1+r^2)].$$

For $x_0>0$, everything proceeds in reverse.

The YM potential (6.21) exhibits similar space-time behavior. In the limit $x_0\to\pm\infty$ we see that the potential is distributed at spatial infinity with sufficient density for the total energy to be nonzero. At time $x_0=0$ the field is localized around the origin.

The complex DFF potential (6.21) can be transformed into a real solution of the SU(2) gauge theory. This real solution is (Lüscher, 1977; Schechter, 1977; Rebbi, 1978; Bernreuther, 1977)

$$eW_0^a=\pm(1/y^2)x_0x_a,$$
$$eW_i^a=(1/y^2)[\varepsilon_{ian}x_n\pm\delta_{ai}\tfrac{1}{2}(1+x^2)\pm x_a x_i],$$
$$y^2=\tfrac{1}{4}(1+x^2)^2+\mathbf{x}^2=\tfrac{1}{4}(1-x^2)^2+x_0^2. \quad (6.27)$$

There are various ways to obtain it.

(1) The SU(2) gauge theory is covariant under the $O(4,2)$ Minkowski conformal group, which is a transformation group in a six-dimensional Euclidean space. All of Minkowski space-time can be mapped onto a hypertorus in this six-dimensional space (Lüscher, 1977; Schechter, 1977):

$$\mathbf{r}=2\mathbf{x}/\sqrt{\lambda},\quad r_0=(1+x^2)/\sqrt{\lambda},\quad r_\mu r_\mu=1,$$
$$R_1=2x_0/\sqrt{\lambda},\quad R_0=(1-x^2)/\sqrt{\lambda},\quad R_iR_i=1,$$
$$\lambda\equiv(1+x^2)^2+4\mathbf{x}^2=(1-x^2)^2+4x_0^2.$$

Independent rotations of r_μ and R_i belonging to an $O(4)\times O(2)$ subgroup of the full conformal group leave this hypertorus invariant. The above mapping leads to an SU(2) gauge theory on the hypertorus, with equations of motion in the new variables which one can try to solve. If one requires that the (real) solution be invariant under the $O(4)\times O(2)$ invariance subgroup of the hypertorus, then the YM equations of motion reduce to a single elliptic differential equation in one variable whose real solutions are (i) a constant [this corresponds to solution (6.27) above] and (ii) an elliptic function (this corresponds to the elliptic generalization of the DFF solution discussed in the next subsection, in a real gauge). These solutions are not self-dual.

(2) The formalism just sketched can also be used to find a complex, self-dual solution of the SU(2) gauge theory (Rebbi, 1977). Requiring self-duality as well as $O(4)$ invariance leads to a first-order differential equation in the hypertoroidal variables. This equation has a simple complex solution. Projecting back to Minkowski space one obtains a self-dual (hence necessarily complex) solution, which has the remarkable property that its real part is also a solution, namely, Eq. (6.27).

(3) The solution (6.27) can be obtained from the one-meron solution of the Euclidean SU(2) gauge theory (see the following section) by a coordinate transformation (Bernreuther, 1977). We choose to give this derivation because it involves a minimum of formalism. Introduce Euclidean space-time coordinates y_μ and potentials $A_\mu^a(y)$ as follows:

$$y_0=\tfrac{1}{2}(1+x^2),\quad y_i=x_i;$$
$$W_0^a(x)=x_0A_0^a(y),\quad W_i^a(x)=-A_i^a(y)+x_iA_0^a(y).$$

Defining field strengths $A_{\mu\nu}^a(y)$ in the usual way, it is easy to show that

$$G_{0j}^a(x)=-x_0A_{0j}^a(y),$$
$$G_{ij}^a(x)=A_{ij}^a(y)-x_iA_{0j}^a(y)+x_jA_{0i}^a(y).$$

Let $A_\mu^a(y)$ be the one-meron potential (see Sec. VII):

$$eA_0^a(y)=\pm y_a/y^2,$$

$$eA_i^a(y) = -\varepsilon_{ian}y_n/y^2 \mp \delta_{ai}y_0/y^2 .$$

This Euclidean solution satisfies the conditions

$$y_\mu A_\mu^a = 0,\ \ y_\mu A_{\mu\nu}^a = 0\ ,\ \ y_\mu \partial'_\mu A_\alpha^a = -A_\alpha^a,\ \ y_\mu \partial'_\mu A_{\alpha\beta}^a = -2A_{\alpha\beta}^a,$$

where $\partial'_\mu = \partial/\partial y^\mu$. Given these conditions, it is easy to verify that the equations of motion for $W_\mu^a(x)$ are satisfied,

$$\partial^\nu G_{0\nu}^a - e\varepsilon_{abc}G_{0\nu}^b W_c^\nu = -x_0[\partial'_j A_{0j}^a - e\varepsilon_{abc}A_{0j}^b A_j^c] = 0\ ,$$

$$\partial^\nu G_{i\nu}^a - e\varepsilon_{abc}G_{i\nu}^b W_c^\nu = [\partial'_0 A_{i0}^a + \partial'_j A_{ij}^a] - e\varepsilon_{abc}[A_{i0}^b A_0^c + A_{ij}^b A_j^c] - x_i\{\partial'_j A_{0j}^a - e\varepsilon_{abc}A_{0j}^b A_j^c\} = 0$$

because the one-meron potential $A_\mu^a(y)$ is a solution of the Euclidean equations of motion. $W_\mu^a(x)$ is the real solution (6.27) above.

(4) de Alfaro, Fubini, and Furlan (1976) show that their original solution (6.21) is a continuation to Minkowski space of the *meron–meron* solution in Euclidean space-time. The same authors (de Alfaro *et al.*, 1977) show that the real solution (6.27) is a continuation to Minkowski space of the slightly different Euclidean solution which represents a *meron* and an *antimeron*. Moreover, they give the gauge transformation which connects these two Euclidean solutions.

(5) In Sec. III.F we have given a simple derivation of the real solution (6.27) in the $W_0^a = 0$ gauge using Witten's *ansatz*. There the elliptic generalizations of this solution are also derived.

C. Elliptic solutions

The nonsingular DFF solution described above can be generalized with the help of Jacobi elliptic functions. Cervero, Jacobs, and Nohl (1977) (hereafter CJN) were the first to show how to do this. Their approach was systematically developed by the author (Actor, 1978b). We now show how to construct elliptic solutions of the ϕ^4 and YM theories. The elliptic generalization of the DFF solution is then given as an example.

Suppose that we are searching for an elliptic solution of the equation

$$\Box\phi + \lambda\phi^3 = 0\ , \tag{6.28}$$

and that we have already found an explicit solution $f(x)$ of this same equation. We make the *ansatz*

$$\phi(x) = f(x)E(u(x), k)\ , \tag{6.29}$$

where $E(u, k)$ is an elliptic function which satisfies a differential equation of the form

$$E'' + a(k)E + b(k)E^3 = 0\ , \tag{6.30}$$

where $a(k)$ and $b(k)$ are constants (see Appendix F on elliptic functions). The argument $u = u(x)$ of the elliptic function has got to be determined. One easily verifies that Eq. (6.28) is satisfied if $f(x)$ and $u(x)$ satisfy

$$\Box f - (a\lambda/b)f^3 = 0\ , \tag{6.31}$$

$$2\partial^\alpha f\,\partial_\alpha u + f\Box u = 0\ , \tag{6.32}$$

$$\partial^\alpha u\,\partial_\alpha u - (\lambda/b)f^2 = 0\ . \tag{6.33}$$

We assume that $f(x)$ is a *known solution* of Eq. (6.31). This leaves Eqs. (6.32) and (6.33) to be satisfied by the unknown function $u(x)$. If this function exists then an elliptic generalization of $f(x)$ exists. Solving the conditions (6.32) and (6.33) for $u(x)$ is the essential step in the construction. Once $u(x)$ has been found, the elliptic generalization of the YM solution obtained from $f(x)$ follows automatically from *ansatz* (6.1).

There are 12 Jacobi elliptic functions (see Appendix F), and one might expect that 12 different elliptic generalizations of $f(x)$ exist. Actually, five of these are redundant. The remaining ones are (Actor, 1978b)

$$\phi = f\ \mathrm{sn}\left(\frac{\omega}{\sqrt{-1-k^2}}, k\right) = if\ \mathrm{sc}\left(\frac{\omega}{\sqrt{2-k'^2}}, k'\right), \tag{6.34}$$

$$\phi = f\ \mathrm{cn}\left(\frac{\omega}{\sqrt{-1+2k^2}}, k\right) = f\ \mathrm{nc}\left(\frac{\omega}{\sqrt{-1+2k'^2}}, k'\right), \tag{6.35}$$

$$\phi = f\ \mathrm{dn}\left(\frac{\omega}{\sqrt{2-k^2}}, k\right) = f\ \mathrm{dc}\left(\frac{\omega}{\sqrt{-1-k'^2}}, k'\right), \tag{6.36}$$

$$\phi = f\ \mathrm{ns}\left(\frac{\omega}{\sqrt{-1-k^2}}, k\right) = if\ \mathrm{cs}\left(\frac{\omega}{\sqrt{2-k'^2}}, k'\right), \tag{6.37}$$

$$\phi = f\ \mathrm{nd}\left(\frac{\omega}{\sqrt{2-k^2}}, k\right) = f\ \mathrm{cd}\left(\frac{\omega}{\sqrt{-1-k'^2}}, k'\right), \tag{6.38}$$

$$\phi = f\ \mathrm{sd}\left(\frac{\omega}{\sqrt{-1+2k^2}}, k\right) = if\ \mathrm{sd}\left(\frac{\omega}{\sqrt{-1+2k'^2}}, k'\right), \tag{6.39}$$

$$\phi = f\ \mathrm{ds}\left(\frac{\omega}{\sqrt{-1+2k^2}}, k\right) = if\ \mathrm{ds}\left(\frac{\omega}{\sqrt{-1+2k'^2}}, k'\right), \tag{6.40}$$

where $\omega = \sqrt{-a(k)}\,u(x)$, and Jacobi imaginary transformations have been used to obtain the second form of the solution with parameter k'. In each case, $f(x)$ satisfies Eq. (6.31) with the appropriate constants $a(k)$ and $b(k)$.

Proceeding now to the YM theory, we note that

$$\partial_\mu\phi/\phi = \partial_\mu f/f + (E'/E)\partial_\mu u \tag{6.41}$$

in Eq. (6.1). Therefore the YM potential obtained from ϕ has the form

$$eW_\mu^a(\phi) = eW_\mu^a(f) + (E'/E)G_\mu^a\ , \tag{6.42}$$

where $W(f)$ is the YM potential corresponding to f and

$$G_0^a = \pm i\partial_a u,\ \ G_i^a = \varepsilon_{ian}\partial_n u \pm i\delta_{ai}\partial_0 u\ . \tag{6.43}$$

In Eq. (6.42) we see that the elliptic YM potential will have singularities at points where the elliptic function $E(u)$ has zeros. This makes two of the elliptic solutions especially interesting, namely the ones with $E = \mathrm{dn}$ and $E = \mathrm{nd} = 1/\mathrm{dn}$. These elliptic functions have no zeros on the real-u axis, and they can therefore lead to YM solutions which are nonsingular.

Let us return to the DFF solution in subsection B above. The corresponding solution of Eq. (6.31) is

$$f(x) = (4b/a\lambda)^{1/2}h(x),\ \ h(x) \equiv [(1-x^2)^2 + 4x_0^2]^{-1/2}\ . \tag{6.44}$$

One can easily verify that the function $u(x)$ needed for the elliptic generalization of $f(x)$ is

$$u(x) = (1/\sqrt{a})\tan^{-1}[2x_0/(1-x^2)]\ . \tag{6.45}$$

Therefore, from Eqs. (6.34)–(6.40) we obtain seven elliptic solutions of the scalar theory.

To understand the properties of the elliptic YM solutions, it is important to notice that $f(x)$ and $u(x)$ in this example satisfy the two identities [which are extensions of (6.32) and (6.33), respectively]

$$f\partial_\alpha\partial_\beta u = \partial_\alpha u \partial_\beta f + \partial_\alpha f \partial_\beta u - g_{\alpha\beta}\partial^\mu f \partial_\mu u, \quad (6.46)$$

$$2af^2\partial_\alpha u \partial_\beta u = 4\partial_\alpha f \partial_\beta f - 2f\partial_\alpha\partial_\beta f - g_{\alpha\beta}\left(-\frac{\lambda a}{b} f^4 + \partial^\mu f \partial_\mu f\right). \quad (6.47)$$

The latter identity can also be written

$$(e^2 b/a\lambda)\theta_{\alpha\beta}(f) = 2af^2\partial_\alpha u \partial_\beta u - (a\lambda/2b) g_{\alpha\beta} f^4, \quad (6.48)$$

where $\theta_{\alpha\beta}(f)$ is the energy–momentum tensor calculated from the YM potential $W(f)$. This leads to a particularly simple formula for $\theta_{\alpha\beta}(f)$:

$$\theta_{\alpha\beta}(f) = (8/e^2)(4\hat{n}_\alpha \hat{n}_\beta - g_{\alpha\beta})h^4(x), \quad (6.49)$$

where $\hat{n}_\alpha$ is the unit vector

$$\hat{n}_\alpha = h(x)[(1-x^2)\delta_{\alpha 0} + 2x_0 x_\alpha]. \quad (6.50)$$

An important consequence of Eqs. (6.46) and (6.47) is that the energy-momentum tensor $\theta_{\alpha\beta}(\phi)$ for any of the elliptic YM solutions $W(\phi)$ is proportional to $\theta_{\alpha\beta}(f)$,

$$\theta_{\alpha\beta}(\phi) = -(bc/a^2)\theta_{\alpha\beta}(f), \quad (6.51)$$

where $c = c(k)$ is another constant associated with the elliptic functions (see Appendix F). For the elliptic solutions with $E = \text{dn}, \text{nd}$ that we are interested in, the proportionality constant in Eq. (6.51) is

$$-bc/a^2 = 2(1-k^2)/(2-k^2)^2. \quad (6.52)$$

After a somewhat lengthy calculation, one finds the pseudoscalar density $D(\phi)$ for any one of the elliptic YM solutions to be

$$\pm D(\phi) = (3\lambda^2/4e^2) f^4 [E^4 - 4c^2/b^2E^4]. \quad (6.53)$$

From Eq. (6.16) it follows that the action density is

$$\mathcal{L}(\phi) = -(3\lambda^2/4e^2) f^4 [E^4 + 4c^2/b^2E^4]. \quad (6.54)$$

For the solutions with $E = \text{dn}, \text{nd}$ we find

$$\pm D = \eta[48h^4(x)/e^2(1+k'^2)^2][\text{dc}^4(u,k') - k'^2\,\text{cd}^4(u,k')], \quad (6.55)$$

$$\mathcal{L} = -[48h^4(x)/e^2(1+k'^2)^2][\text{dc}^4(u,k') + k'^2\,\text{cd}^4(u,k')], \quad (6.56)$$

where

$$u = (1/\sqrt{1+k'^2})\tan^{-1}[2x_0/(1-x^2)], \quad (6.57)$$

and $\eta = +1(-1)$ for $E = \text{dn}(\text{nd})$. Note that we have chosen the second form of ϕ with parameter k' in Eqs. (6.36) and (6.38) because $a(k) = -(1+k'^2)$ is negative for these solutions.

The elliptic YM solution following from ϕ in Eq. (6.36) is the one found by CJN. The other solution with $E = \text{nd}$ is obviously gauge equivalent. CJN mention that their solution develops singularities when the parameter k exceeds a certain critical value k_c (numerically, $k_c^2 \approx 0.173$). To show that this happens, we note that $\text{cn}(u,k')$ in the denominator in Eqs. (6.55) and (6.56) vanishes at space-time points which satisfy the condition $u = K(k')$. The elliptic YM solution is therefore singular at these points. For $k = 0 (k' = 1)$ this condition cannot be satisfied because $K(1) = \infty$ while $\tan^{-1}2x_0/(1-x^2)$ cannot exceed π. But for $k \geq k_c$ the condition can be satisfied, so that there are singularities [k_c is defined by $K(k') = (1+k'^2)^{1/2}\pi$]. The CJN solution reduces to the nonsingular DFF solution for parameter $k = 0$ because $\text{dn}(u, 0) = 1$. As we see, this solution is the limit of a continuous sequence of nonsingular solutions.

VII. EUCLIDEAN SOLUTIONS

A. Introduction

Now we come to the instanton and meron solutions of the SU(2) gauge theory in Euclidean space-time. These exact solutions have attracted a great deal of interest since their discovery, and for good reason. In quantum-mechanical problems which have classical analogs, tunneling is bound to occur when there exist Euclidean (i.e., imaginary-time) solutions of the classical equations of motion. Usually an approximate description of the quantum tunneling effect can be obtained from these classical solutions. The classical instanton and meron solutions of the Euclidean YM theory therefore imply the existence of tunneling effects in the quantized theory. Traditional (perturbative) quantum field theory allows no room for such effects, and the evident conclusion is that perturbative field theory is only one sector of a larger mathematical structure in which tunneling occurs and is approximately described by instantons, merons, and perhaps by other Euclidean space objects which have yet to be found.

At first sight it is not obvious where one should look for the new tunneling effects in the quantized SU(2) gauge theory. One has to examine the details of the classical solutions to find this out. There are three basic types of Euclidean solution known (at present), which we give the names instanton, meron, and elliptic solutions. To appreciate the differences between these three types of solution it is necessary to understand certain topological properties of the Euclidean gauge theory. We first review these properties. Then we describe the various solutions and their physical interpretation. This discussion is a fairly self-contained introduction to instanton physics. Following this introduction we study the individual solutions in considerable detail. Two *ansätze* will be discussed separately from the solutions in subsection B.

1. Euclidean SU(2) gauge theory

In Euclidean space-time the gauge-theory Lagrangian is

$$\mathcal{L} = \tfrac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} = \tfrac{1}{2}[E^a_n E^a_n + B^a_n B^a_n],$$

$$E^a_n = G^a_{0n}, \quad B^a_n = -\tfrac{1}{2}\varepsilon_{nij}G^a_{ij}.$$

The energy–momentum tensor is

$$\theta_{\alpha\beta} = G^a_{\mu\alpha}G^a_{\mu\beta} - g_{\alpha\beta}\mathcal{L} = \tfrac{1}{4}(G^a_{\mu\alpha} + \tilde{G}^a_{\mu\alpha})(G^a_{\mu\beta} - \tilde{G}^a_{\mu\beta}).$$

Clearly $\theta_{\alpha\beta} = 0$ for any self-dual solution. A gauge-invariant quantity which plays an extremely important role is the pseudoscalar density

$$D \equiv \tfrac{1}{4}G^a_{\mu\nu}\tilde{G}^a_{\mu\nu} = -E^a_n B^a_n = (1/2e^2)\partial_\mu J_\mu. \quad (7.1)$$

D is the four-divergence of the current

$$(1/2e^2)J_\mu \equiv \tfrac{1}{2}\varepsilon_{\mu\alpha\beta\gamma}[W^a_\alpha\partial_\beta W^a_\gamma + \tfrac{1}{3}e\varepsilon_{abc}W^a_\alpha W^b_\beta W^c_\gamma]. \quad (7.2)$$

This result follows directly from the basic definition of D and the trivial identity

$$\varepsilon_{\mu\alpha\beta\gamma}\varepsilon_{abc}\varepsilon_{amn}W^b_\mu W^c_\nu W^m_\alpha W^n_\beta = 0.$$

In matrix notation the current (7.2) has the form

$$(1/2e^2)J_\mu = \mathrm{Tr}\,\varepsilon_{\mu\alpha\beta\gamma}[W_\alpha\partial_\beta W_\gamma + (2e/3i)W_\alpha W_\beta W_\gamma]. \quad (7.3)$$

Under a local gauge transformation

$$W_\mu \to W'_\mu = \omega W_\mu\omega^{-1} - (i/e)(\partial_\mu\omega)\omega^{-1},$$

J_μ transforms as follows

$$J_\mu \to J'_\mu = J_\mu + 2ie\,\mathrm{Tr}\,\varepsilon_{\mu\alpha\beta\gamma}\partial_\alpha(\partial_\beta\omega W_\gamma\omega^{-1}) - \tfrac{2}{3}\mathrm{Tr}\,\varepsilon_{\mu\alpha\beta\gamma}\partial_\alpha\omega\omega^{-1}\partial_\beta\omega\omega^{-1}\partial_\gamma\omega\omega^{-1}. \quad (7.4)$$

The second and third terms are divergence free, and $\partial_\mu J_\mu$ is gauge invariant, as it should be.

Integrating over D we obtain the *topological index or charge* of the Euclidean field configuration,

$$q \equiv \frac{e^2}{8\pi^2}\int d^4x\, D = \frac{1}{16\pi^2}\lim_{x^2\to\infty}\oint d\Omega_\mu\, J_\mu + q_{\rm sing}. \quad (7.5)$$

Here, in using Gauss's theorem, we have to admit the possibility that the field configuration is singular. Thus $q_{\rm sing}\equiv 0$ for any nonsingular solution. For singular solutions $q_{\rm sing}$ may or may not be zero. For example, $q_{\rm sing}=0$ for the singular one- and two-meron solutions. Note that q is different from a Noether charge, i.e., a charge of the form $\int d^3x\, J_0(\mathbf{x},t)$. It is impossible to make q time dependent. This is characteristic of a topological charge.

q can have any value from $-\infty$ to $+\infty$ depending on the YM field considered. It is useful to group all possible YM fields into classes, the fields within a class (and only these) having a given value of q. One can then ask: Are all values of q interesting? At the present time only classes with integer q or half-integer q are thought to be physically important. The fields in these classes describe collections of instantons and merons, respectively.

The total action is bounded from below by $|q|$ for any real Euclidean solution. This follows from the inequality

$$(G^a_{\mu\nu} \pm \tilde{G}^a_{\mu\nu})^2 \geq 0,$$

which implies that $\mathcal{L} \geq |D|$ and

$$W = \int d^4x\,\mathcal{L} \geq (8\pi^2/e^2)|q|.$$

In the $q=0$ topological sector the action W can be arbitrarily small; this is the perturbative sector. In the $|q|=1$ sector the action cannot be smaller than $8\pi^2/e^2$. This is the one-instanton sector; the one-instanton solution with action $W=8\pi^2/e^2$ saturates this lower bound. Similarly, the N-instanton solution saturates the lower bound in the $|q|=N$ sector. In general, it is clear that only self-dual solutions can saturate this bound (because $\mathcal{L}=|D|$ implies self-duality). It is conjectured, although not yet proven, that only self-dual YM solutions can have finite Euclidean action.

We now explain what q has to do with topology. Let us assume that the YM potential becomes gauge equivalent to the vacuum as $x^2\to\infty$,

$$eW_\mu \to -i(\partial_\mu\omega)\omega^{-1}, \quad x^2\to\infty. \quad (7.6)$$

This boundary condition associates an SU(2) group element ω with each point on the infinite sphere S^3_∞ in E^4. Therefore any solution with the behavior (7.6) will determine a map $S^3_\infty \to \mathrm{SU}(2)$ of this sphere into the SU(2) group manifold. Now the SU(2) group is topologically the same as a three-sphere S^3 in E^4. Therefore the solution determines a map $S^3_\infty \to S^3$. All maps of the three-sphere S^3 onto itself fall into homotopy classes labeled by a topological index or charge n (see Appendix I). The integer n is simply the number of times S^3 gets covered by the map of S^3_∞. Correspondingly, all solutions with the behavior (7.6) at infinity fall into homotopy classes labeled by n. For these solutions, the topological charge q defined in Eq. (7.5) is this integer n.

It is straightforward to prove that $q=n$ (Belavin *et al.*, 1975). For a pure-gauge potential the current J_μ is

$$J_\mu = -\tfrac{2}{3}\varepsilon_{\mu\alpha\beta\gamma}\,\mathrm{Tr}[\partial_\alpha\omega\omega^{-1}\partial_\beta\omega\omega^{-1}\partial_\gamma\omega\omega^{-1}].$$

Thus, for a solution which satisfies the boundary condition (7.6) at infinity, the surface integral in Eq. (7.5) becomes

$$q = -\oint_{x^2=\infty} d\Omega_\mu(1/24\pi^2)\varepsilon_{\mu\alpha\beta\gamma}\,\mathrm{Tr}[\partial_\alpha\omega\omega^{-1}\partial_\beta\omega\omega^{-1}\partial_\gamma\omega\omega^{-1}].$$

To simplify the integrand let us parameterize ω by three suitable angles $\theta_a (a=1,2,3)$. Then $\partial_\mu\omega = \partial_\mu\theta_a\partial_a\omega$ where from now on (until we finish the proof that $q=n$) $\partial_a\omega \equiv \partial\omega/\partial\theta_a$. The integral for q then becomes

$$q = -\oint_{x^2=\infty} d\Omega_\mu(1/4\pi^2)\varepsilon_{\mu\alpha\beta\gamma}\partial_\alpha\theta_1\partial_\beta\theta_2\partial_\gamma\theta_3 \times \mathrm{Tr}[\partial_1\omega\omega^{-1}\partial_2\omega\omega^{-1}\partial_3\omega\omega^{-1}].$$

Next, we need the following formula,

$$d\theta_1\,d\theta_2\,d\theta_3 = d\Omega_\mu\,\varepsilon_{\mu\alpha\beta\gamma}\partial_\alpha\theta_1\partial_\beta\theta_2\partial_\gamma\theta_3,$$

which may look complicated but is easy to understand. The surface element $d\Omega_\mu = x_\mu x^2\,d\Omega$ is parallel to x_μ and therefore $d\Omega_\mu\,\varepsilon_{\mu\alpha\beta\gamma}$ projects out the components of the derivatives $\partial_\alpha\theta_1\cdots$ perpendicular to x_μ. These perpendicular components are essentially the derivatives $\partial\theta_1/\partial\psi_i\cdots$, where ψ_i are three angles which characterize points on S^3_∞. The formula above is therefore essentially

$$d\theta_1\,d\theta_2\,d\theta_3 = (\text{Jacobian})\,d\psi_1\,d\psi_2\,d\psi_3.$$

[The reader who has difficulty with this formula should consider the corresponding U(1) problem, where a circle with angle ψ is mapped onto a circle with angle θ. Here the topological charge is given by the surface integral

$$q = -\frac{i}{2\pi}\oint_{x^2=\infty} d\Omega_\mu\,\varepsilon_{\mu\nu}(\partial_\nu\omega)\omega^{-1} = \frac{1}{2\pi}\oint_{x^2=\infty} d\Omega_\mu\,\varepsilon_{\mu\nu}\partial_\nu\theta,$$

where $\omega=\exp(i\theta)$ and $x_\mu=(x,y)=r(\cos\psi,\sin\psi)$ while $d\Omega_\mu = x_\mu d\psi$. The integrand is

$$d\Omega_\mu\,\varepsilon_{\mu\nu}\partial_\nu\theta = d\psi\,\varepsilon_{\mu\nu}x_\mu\partial_\nu\theta = d\psi\,(\partial\theta/\partial\psi) = d\theta,$$

and therefore

$$q=\frac{n}{2\pi}\int_0^{2\pi} d\theta = n$$

is the winding number n in this U(1) problem.] Returning to the SU(2) case we write q in the form

$$q=-\frac{n}{4\pi^2}\int d\theta_1\, d\theta_2\, d\theta_3 \,\mathrm{Tr}[\partial_1\omega\omega^{-1}\partial_2\omega\omega^{-1}\partial_3\omega\omega^{-1}].$$

If θ_a are the Euler angles,

$$\omega=\exp(\tfrac{1}{2}i\theta_1\sigma_3)\exp(\tfrac{1}{2}i\theta_2\sigma_2)\exp(\tfrac{1}{2}i\theta_3\sigma_3),$$

then one easily verifies that

$$\mathrm{Tr}[\partial_1\omega\omega^{-1}\partial_2\omega\omega^{-1}\partial_3\omega\omega^{-1}]=-\tfrac{1}{4}\sin\theta_2,$$

so that

$$q=\frac{2n}{16\pi^2}\int d\theta_1\, d\theta_2\, d\theta_3 \sin\theta_2 = n.$$

The extra factor 2 here comes from the fact that the Euler angle parametrization of ω corresponds to a proper rotation. The improper rotation $-\omega$ also has to be counted to obtain the full SU(2) group manifold.

For the purpose of illustration let us consider the simplest nontrivial case $n=1$. This is the one-instanton solution of Belavin *et al.* (1975), which satisfies the boundary condition (7.6) with

$$\omega=\sigma_\mu n_\mu,\quad \sigma_\mu=(1,i\sigma),\quad n_\mu=x_\mu/\sqrt{x^2}.$$

Here we see how ω maps the direction x_μ in E^4 onto the unit vector n_μ in another four-dimensional space R_4. The SU(2) group manifold in R_4 is the locus of points touched by n_μ, i.e., the unit sphere S^3. Obviously this sphere gets covered one time by the mapping.

We now explain how the topological charge q is related to tunneling in Minkowski space. For this purpose we introduce a second, *nonconserved* charge

$$Q_T(t)\equiv(1/16\pi^2)\int d^3x\, J_0. \tag{7.7}$$

This charge is only conserved if $D=0$ (assuming that surface integrals at infinity can be ignored). Therefore, nonconservation of Q_T is related to the presence of topological charge. Note that under a gauge transformation

$$Q_T\to Q_T'=Q_T-(1/24\pi^2)\int d^3x\,\varepsilon_{ijk}\,\mathrm{tr}[\partial_i\omega\omega^{-1}\partial_j\omega\omega^{-1}\partial_k\omega\omega^{-1}].$$

Thus for a pure-gauge potential $eW_\mu=-i(\partial_\mu\omega)\omega^{-1}$ the charge Q_T is

$$Q_T(t)=-(1/24\pi^2)\int d^3x\,\varepsilon_{ijk}\,\mathrm{tr}[\partial_i\omega\omega^{-1}\partial_j\omega\omega^{-1}\partial_k\omega\omega^{-1}]. \tag{7.8}$$

Now we want to show that Q_T is the homotopy class label associated with the gauge transformation ω. Let us consider a fixed time x_0, and assume that $\omega=\omega(x_0,\mathbf{x})$ satisfies the boundary condition $\omega\to 1$ for $r\to\infty$. Then $\omega(x_0,\mathbf{x})$ defines a map of three-space with all points at $r=\infty$ indentified into the SU(2) group manifold. Now three-space with all points at infinity identified is topologically equivalent to the unit sphere S^3 in a four-dimensional Euclidean space. Moreover, the group SU(2) has the topology of S^3. Therefore the local gauge transformation $\omega(x_0,\mathbf{x})$ defines a map $S^3\to S^3$. As we discuss in Appendix I, all maps of S^3 onto S^3 are characterized by an integral topological index n (n is the number of times the latter sphere gets covered). This means that (i) every gauge transformation ω carries an index n; (ii) all gauge transformations with different n are topologically inequivalent (i.e., they cannot be continuously deformed into one another); and (iii) gauge transformations with the same n are fully equivalent. The charge Q_T is the topological index of the gauge transformation. This can be proved using essentially the same argument just given for q. Introducing Euler angles θ_a we can rewrite Eq. (7.8) in the form (where $\partial_a\omega=\partial\omega/\partial\theta_a$)

$$\begin{aligned}Q_T&=-(1/2\pi^2)\int d^3x\,\varepsilon_{ijk}\partial_i\theta_1\partial_j\theta_2\partial_k\theta_3\\&\quad\times\mathrm{tr}[\partial_1\omega\omega^{-1}\partial_2\omega\omega^{-1}\partial_3\omega\omega^{-1}]\\&=-(n/2\pi^2)\int d\theta_1\,d\theta_2\,d\theta_3(-\tfrac{1}{4}\sin\theta_2)\\&=n.\end{aligned}$$

A local gauge transformation can change n. In contrast, q is absolutely gauge invariant. In general one can think of $Q_T(t)$ as a number which characterizes a YM field configuration in normal three-space at time t. For an arbitrary field, $Q_T(t)$ will not be an integer nor will it have any topological meaning. But for a pure gauge field $Q_T(t)$ is the homotopy class label of the gauge transformation.

We can now write down an expression which compactly summarizes the relationship between Euclidean solutions, Minkowski-space tunneling, and the two types of topological charge q and Q_T. This is

$$q=Q_T(t=+\infty)-Q_T(t=-\infty), \tag{7.9}$$

where we assume that $\mathbf{J}$ vanishes faster than $O(1/r^2)$ as $r\to\infty$. The evident meaning of this formula is that a Euclidean solution $W_\mu(x_0,\mathbf{x})$ interpolates between the real-time potentials $W_\mu(t=\pm\infty,\mathbf{x})$ in the distant past and future. Imaginary time x_0 plays the role of an interpolating parameter. In general, one cannot call this continuous interpolation a tunneling process because there is no barrier separating the initial and final field configurations. But when the fields $W_\mu(t=\pm\infty,\mathbf{x})$ are pure gauge, so that the charges $Q_T(t=\pm\infty)$ are homotopy class labels, then tunneling is the appropriate expression for $q\neq 0$. The initial and final states have zero energy and they are homotopically inequivalent for $q\neq 0$ so that a barrier must separate them. This is a tunneling situation.

2. Instanton solutions

An instanton (also called a pseudoparticle) is a localized, nonsingular solution of the Euclidean gauge theory with one unit of topological charge. The name instanton is derived from this localization in E^4 which corresponds to finite duration as well as spatial extension in Minkowski space, i.e., to an "event." All instanton solutions are self-dual, and therefore they have zero Euclidean energy. A self-dual solution in Minkowski space also has zero energy. This seems to suggest that the existence of instantons may be related to a tunneling effect in Minkowski space between different vacua. If only one vacuum exists, then of course there is no possibility for tunneling. More than one vacuum would be necessary. The topological nature of the instanton suggests that a denumerable infinity of vacua $|n\rangle$ might exist, the vacuum $|n\rangle$ having topological index n. Then the one-instanton solution would connect $|n\rangle$ with $|n+1\rangle$, the two-instanton solution (which is also self-dual) would connect $|n\rangle$ with $|n+2\rangle$, and so on.

We shall see immediately below that these speculations are correct; there are really infinitely many topologically inequivalent vacua. If the instanton solution did not exist then (other things being equal) these vacua would all be absolutely disconnected from one another, and all but one of them could be ignored. As it is, they all have to be taken into account.

It is not difficult to show the existence of an infinity of inequivalent vacua in the SU(2) gauge theory (Jackiw and Rebbi, 1976a; Callan, Dashen, and Gross, 1976). Choose a gauge with $W_0^a=0$. The (Minkowski-space) vacuum solution of the theory is then a pure-gauge potential,

$$eW_i(\mathbf{x}) = -i[\partial_i\omega(\mathbf{x})]\omega^{-1}(\mathbf{x})\,.$$

Here we must exclude time-dependent gauge transformations, as these would lead to nonzero W_0^a. It is technically advantageous (although not essential) to make the additional assumption that $\omega(\mathbf{x})\to 1$ as $r\to\infty$. Then, $\omega(\mathbf{x})$ defines a map of three-space with all points at infinity identified into the SU(2) group. As we have already discussed, these maps fall into homotopy classes labeled by an integer n. Correspondingly, the gauge transformations $\omega(\mathbf{x})$ fall into homotopy classes, and so do the pure-gauge potentials $eW_\mu=-i(\partial_\mu\omega)\omega^{-1}$. The totality of pure-gauge potentials determines the YM vacuum. This vacuum therefore consists of an infinity of sectors, which are distinguished from one another topologically in the way just described.

The traditional (i.e., perturbative) gauge-theory vacuum is the one with $n=0$, for this vacuum contains the potential $W_i^a=0$. All other vacuum potentials in the $n=0$ class are obtained from $W_i^a=0$ by gauge transformations that are continuously deformable to the identity $\omega_0(\mathbf{x})=1$.

A gauge transformation which is *not* continuously deformable to $\omega_0(\mathbf{x})=1$ is (Jackiw and Rebbi, 1976a)

$$\omega_1(\mathbf{x}) = \frac{r^2-\lambda^2}{r^2+\lambda^2} - i2\lambda\frac{\mathbf{x}\cdot\boldsymbol{\sigma}}{r^2+\lambda^2}\,.$$

This gauge transformation belongs to the class with $n=1$, as one can show by rewriting it in the form

$$\omega_1(\mathbf{x}) = \sigma_\mu n_\mu\,,\quad \sigma_\mu=(1, i\boldsymbol{\sigma})\,,$$

$$n_\mu = \left(\frac{r^2-\lambda^2}{r^2+\lambda^2}\,,\,-2\lambda\frac{\mathbf{x}}{r^2+\lambda^2}\right).$$

As $\mathbf{x}$ ranges over all of three-space, the unit vector n_μ traces out a sphere S^3 in a four-dimensional space. Thus ω_1 maps three-space, with all points at infinity identified ($\omega_1\to 1$ as $r\to\infty$), onto S^3, covering it once. The sphere S^3 here can also be identified with the SU(2) group manifold. Therefore we see that $\omega_1(x)$ is indeed an $n=1$ gauge transformation. A gauge transformation in the nth homotopy class is

$$\omega_n(\mathbf{x}) = [\omega_1(\mathbf{x})]^n\,.$$

Let us return to the instanton solution. This Euclidean solution (in the $W_0^a=0$ gauge) can be shown to have the following behavior for large negative and positive imaginary time:

$$-eW_i^a\to i(\partial_i\omega_1)\omega_1^{-1}\,,\quad x_0=it\to-\infty\,;$$

$$eW_i^a\to 0\,,\quad x_0=it\to+\infty\,.$$

In words, it interpolates between the $n=1$ vacuum field configuration in three-space and the $n=0$ configuration. There is no difference between three-space in E^4 and in Minkowski space-time, of course. Therefore the meaning of the instanton solution for Minkowski space-time is clear. It connects the $n=1$ vacuum with the $n=0$ vacuum via an exact solution of the imaginary-time equations of motion. This is the classical equivalent of tunneling. We have not said through what potential barrier the tunneling occurs. [See Bitar and Chang, 1978; Eylon and Rabinovici, 1977, for discussion of this point.] However, it is clear that an energy barrier must be present because a topological number gets changed. As we have seen, there does *not* exist a continuous sequence of zero-energy (i.e., pure-gauge) solutions which connects two vacuum solutions in different homotopy classes. Therefore it is necessary to go under an energy barrier, and this is precisely what the instanton does. [Bitar and Chang (1978) give an interesting discussion of the tunneling process in Minkowski space.]

Let us now try to construct the physical vacuum of the pure SU(2) gauge theory. This vacuum should be a superposition of all the different topological vacua. The latter we represent by vectors $|n\rangle$ in a Hilbert space. In this space the gauge transformation $\omega_1(\mathbf{x})$ which changes topological charge by one unit is represented by an operator G with the property

$$G\,|n\rangle = |n+1\rangle\,,\quad |n\rangle = G^n\,|0\rangle\,.$$

We also introduce a topological charge operator

$$Q_T \equiv (1/16\pi^2)\int d^3x\,J_0\,,$$

with the properties

$$Q_T|n\rangle = n|n\rangle\,,\quad [Q_T, G]=G\,.$$

The physical vacuum should be invariant under G, and this constrains it to have the following form, which is called a θ vacuum,

$$|\theta\rangle = \sum_n e^{in\theta}|n\rangle\,,$$

where $0\le\theta\le 2\pi$. Then

$$G|\theta\rangle = e^{i\theta}|\theta\rangle\,,$$

and the θ vacuum is invariant under G up to a phase. There is a continuum of these θ vacua and each one is a suitable ground state for the physical SU(2) gauge theory. Theories based on different θ vacua are distinct. Note that the θ vacua are orthogonal,

$$\langle \theta' | \theta \rangle = \sum_{n,m} e^{i(n\theta - m\theta')} \langle n | m \rangle$$

$$= \sum_n e^{in(\theta - \theta')} = 2\pi\delta(\theta - \theta') .$$

Here we have normalized the topological vacua according to $\langle n|m\rangle = \delta_{mn}$.

If there were no isolated instantons there would be no vacuum tunneling, and all of the $|n\rangle$ vacua would be degenerate in energy ($E_n = 0$). Instantons remove this degeneracy, in the same way that vacuum tunneling splits the ground state in the familiar symmetric double-well problem in ordinary quantum mechanics. The symmetric ground-state wave function has lower energy than the antisymmetric one. This implies that the θ vacua have energy $E(\theta) > 0$ and that $E(\theta)$ increases monotonically with θ (to avoid degeneracy). $|\theta = 0\rangle$ is the fully symmetric vacuum with lowest energy, and $|\theta = \pi\rangle$ is the maximally antisymmetric vacuum with highest energy. This can, of course, be verified explicitly (Callen, Dashen, and Gross, 1976, 1978).

The fact that the physical vacuum is $|\theta\rangle$ for some angle θ leads to an effective Lagrangian for the pure SU(2) gauge theory,

$$\mathcal{L}_{\text{eff}} = \tfrac{1}{4}[G^a_{\mu\nu} G^a_{\mu\nu} + i\theta (e^2/8\pi^2)\, G^a_{\mu\nu} \tilde{G}^a_{\mu\nu}] .$$

We give a qualitative proof below, but first let us point out an implication of the extra term in $\mathcal{L}_{\text{eff}}$. This gauge-invariant term violates P and T invariance. Therefore, the peculiar structure of the θ vacuum leads to spontaneous breakdown of these symmetries for $\theta \neq 0$. The strength of the breakdown is determined by the angle θ. Note that the extra term in $\mathcal{L}_{\text{eff}}$ is a four-divergence [see Eq. (7.1)]. Such a term is usually thought not to have any physical consequences, but this is clearly not the case here. The integral of the extra term is nontrivial even classically (being proportional to the topological charge).

Let us now show that $\mathcal{L}_{\text{eff}}$ has the form claimed above. First we write down the vacuum transition amplitude between topologically distinct vacua (Callen, Dashen, and Gross, 1976)

$$\lim_{x_0 \to \infty} \langle m | e^{-x_0 H} | n \rangle = \int d[W]_{m-n} \exp\left(- \int d^4x\, \mathcal{L} \right) ,$$

where $\mathcal{L} = \frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu}$ is the pure gauge theory Lagrangian, and for simplicity we forget about the gauge-fixing and ghost terms in the exponent. The functional integration here is restricted to Euclidean gauge fields which approach vacuum fields in the $(n - m)$ homotopy class at infinity. These are the only Euclidean gauge fields which can interpolate between the vacua $|n\rangle$ and $|m\rangle$. Next, we calculate the transition amplitude between θ vacua:

$$(1/2\pi) \lim_{x_0 \to \infty} \langle \theta' | e^{-x_0 H} | \theta \rangle$$

$$= (1/2\pi) \sum_{n,m} e^{i(n\theta - m\theta')} \lim_{x_0 \to \infty} \langle m | e^{-x_0 H} | n \rangle$$

$$= \delta(\theta' - \theta) \sum_m e^{-im\theta} \lim_{x_0 \to \infty} \langle m | e^{-x_0 H} | 0 \rangle$$

$$= \delta(\theta' - \theta) \sum_m e^{-im\theta} \int d[W]_m \exp\left(- \int d^4x\, \mathcal{L} \right)$$

$$= \delta(\theta' - \theta) \sum_m \int d[W]_m$$

$$\times \exp\left\{ - \int d^4x [\mathcal{L} + (e^2/8\pi^2)\, \theta \tfrac{1}{4} i G^a_{\mu\nu} \tilde{G}^a_{\mu\nu}] \right\}$$

$$= \delta(\theta' - \theta) \int d[W] \exp\left(- \int d^4x\, \mathcal{L}_{\text{eff}} \right) .$$

Here we have used Eq. (7.5) for the topological charge m. This calculation shows explicitly how the θ term comes to be in the effective Lagrangian in the theory built upon the θ vacuum.

Now let us introduce massless fermions into the gauge theory. This has a very important effect: namely, that *isolated instantons are strongly suppressed.* In other words, the presence of massless fermions destroys the tunneling between vacua $|n\rangle$, $|m\rangle$ with different topological charge, as first noticed by 't Hooft (1976a). This happens because, by introducing fermions, one has effectively introduced a conserved axial charge into the theory. Because of the triangle anomaly this conserved charge has an extra term of topological character in addition to the usual chiral charge Q_5. This extra term forces Q_5 to change when the topological index changes. But Q_5 cannot change in a vacuum tunneling situation, where no fermions are present, and therefore isolated instantons and anti-instantons are suppressed. (Bound pairs of these objects are not suppressed, however.) Note that the argument leading to this suppression of isolated instantons is only valid for *massless* fermions. When the fermions somehow acquire a non-negligible mass, then instantons and anti-instantons are liberated.

Consider N_f massless fermion doublets ψ_i. In the usual way one can define a *gauge-invariant* axial current

$$J^5_\mu = i \sum_i \bar{\psi}_i \gamma_\mu \gamma_5 \psi_i .$$

This current is not conserved. Instead, because of the well known triangle anomaly (Adler, 1969; Bell and Jackiw, 1972),

$$\partial_\mu J^5_\mu = -(e^2/16\pi^2) N_f G^a_{\mu\nu} \tilde{G}^a_{\mu\nu} = -(N_f/8\pi^2) \partial_\mu J_\mu ,$$

where J_μ is the current defined by Eq. (7.2) above. The usual axial charge,

$$Q_5 = \int d^3x\, J^5_0 ,$$

therefore does not commute with the Hamiltonian, although (by gauge invariance) it does commute with the topological charge operator G,

$$[Q_5, G] = 0 .$$

A *conserved* axial current can, however, be defined,

$$A_\mu \equiv J^5_\mu + (N_f/8\pi^2) J_\mu .$$

The corresponding conserved charge

$$Q_A = \int d^3x\, A_0 = Q_5 + 2N_f Q_T$$

commutes with the Hamiltonian, but it is not gauge invariant:

$$[Q_A, G] = 2N_f [Q_T, G] = 2N_f G\,.$$

Assuming that $Q_A|0\rangle = 0$ one easily verifies that $|n\rangle$ is an eigenvector of Q_A,

$$Q_A|n\rangle = 2nN_f|n\rangle\,.$$

Then, by considering the quantity $\langle n|\exp(-x_0 H)Q_A|m\rangle$, it is trivial to show that

$$\langle n|e^{-x_0 H}|m\rangle = \delta_{mn}\langle n|e^{-x_0 H}|n\rangle\,.$$

Thus tunneling between different vacua is suppressed when massless fermions are present.

Even though vacuum tunneling is suppressed in a theory with massless fermions, the physical vacuum in such a theory is still a θ vacuum. [This is certainly true if the limit (fermion mass) $\to 0$ is smooth. When the fermions are massive Q_A is not conserved, and vacuum tunneling exists, necessitating a θ vacuum in a theory with massive quarks.] However, all θ vacua are now physically equivalent. This is because Q_A is the operator which rotates θ,

$$e^{i\alpha Q_A}|\theta\rangle = \sum_n e^{in\theta} e^{i\alpha Q_A}|n\rangle$$
$$= \sum_n e^{in\theta} e^{i\alpha 2n N_f}|n\rangle = |\theta + 2\alpha N_f\rangle\,.$$

Since Q_A is conserved all θ vacua have the same energy. Indeed, they all define the same theory, even though $\langle\theta'|\theta\rangle = 0$ for $\theta' \neq \theta$. Consider the transition amplitude

$$(1/2\pi)\lim_{x_0\to\infty}\langle\theta'|e^{-x_0 H}|\theta\rangle$$
$$= \delta(\theta'-\theta)\lim_{x_0\to\infty}\sum_n \langle n|e^{-x_0 H}|n\rangle$$
$$= \delta(\theta'-\theta)\int d[W]_0 \exp\Big(-\int d^4x\, \mathcal{L}\Big)\,.$$

Here there is no extra term $\theta G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}$ in the effective Lagrangian. The theory is independent of θ. In particular, this means that P and T invariance have been restored by the introduction of massless fermions.

The nonconservation of Q_5 in the presence of instantons appears to solve an old problem. This is the $U_A(1)$ problem in Lagrangian theories with massless quarks (see Pagels, 1976, for a review). Such theories have a $U_A(1)$ symmetry whose generator is Q_5. There is no such symmetry in nature, nor is there a Goldstone boson which corresponds to broken $U_A(1)$. Instantons break this symmetry in a way which does not generate a Goldstone boson ('t Hooft, 1976a), and this seems to resolve the dilemma.

The rule $\Delta Q_5 + 2N_f \Delta Q_T = 0$ connects change of chirality with change in topological charge. Consider the one-instanton case $\Delta Q_T = 1$; the change in chirality is then $\Delta Q_5 = -2N_f$, where N_f is the number of fermion types. This means that N_f fermions have to change chirality when an isolated instanton occurs. We must therefore conclude that an *effective interaction exists between one instanton and $2N_f$ fermions* ('t Hooft, 1976a, b). This is a very important result, which (with due modifications) holds for any non-Abelian gauge theory. Such an interaction can lead to qualitatively new physics. 't Hooft (1976a) has given an amusing example. By applying his triangle anomaly argument to the Weinberg–Salam model he obtained a violation of baryon and lepton number, caused by instantons, which enables the reaction $p + n \to e^+ + \nu_\mu$ to take place. Of course, like all instanton transitions, the amplitude for this process contains a factor $\exp(-8\pi^2/e^2)$. This factor is practically zero if e is the electric charge, and so the reaction above will never occur. For the same reason, all instanton effects in gauge theories of the weak and electromagnetic interactions are negligible. (These are sometimes called "weak instanton" effects.) But in QCD the effective YM coupling can be much larger than the electric charge. One therefore looks for instanton effects in this theory.

Another comment concerning the effective instanton–fermion interaction: We mentioned earlier that the rule $\Delta Q_5 = -2N_f$ implies suppression of isolated instantons. Evidently this would not be the case if there were many massless fermions present and the instanton were not too small. But suppose that we are interested in small-scale effects deep within a hadron. Then, for a single instanton to be important it must (i) be of this same small scale, (ii) interact with several quarks at once. This is unlikely to happen, and the conclusion is again that isolated instantons are suppressed.

Now a word about quark confinement. Much of the attention instantons have received can be attributed to the suggestion (Polyakov, 1975, 1977) that they might provide a mechanism for confinement. A YM theory is thought of as having two phases. In phase (a) the gauge fields are the massless Goldstone bosons associated with a degenerate vacuum. There is no tunneling between different vacua to remove the degeneracy, i.e., no isolated instantons. As we have seen, this would correspond to massless quarks. In this phase the coupling is small (we are deep within the hadron) and there is no confinement. In the second phase (b) the vacuum degeneracy is removed by tunneling; isolated instantons can appear. This means that the quark masses are no longer negligible. The gauge fields are no longer Goldstone bosons; they are now massive. In phase (b) the coupling is large (the entire hadron is involved) and there is confinement.

At present, the concensus of opinion is that instantons do not confine quarks. Meuller (1978) has studied the confinement problem in a way which does not depend on the detailed mechanism which is responsible for confinement. His result is that the Euclidean YM field configurations which are needed for confinement cannot be solutions of the equations of motion. Specifically, he finds that the low-momentum Fourier components of these fields necessarily (and maximally) violate the equations of motion. Therefore these fields must deviate more and more from exact solutions as $x^2 \to \infty$. Taken literally, this rules out the many-instanton and many-anti-instanton configurations, because these are exact solutions for all x^2. It also rules out instanton–anti-instanton configurations, which are not exact solutions, because these configurations become increasingly good (as approximate solutions) when $x^2 \to \infty$.

Most quark confinement discussions (including Meuller's analysis above) are based on Wilson's quark-loop argument (Wilson, 1974). The idea is to calculate the average over all gauge fields of the quantity

$$I[W] = P \exp\left[-ie \oint_C dx_\mu W_\mu\right] \exp\left[-\tfrac{1}{4}\int d^4x G^a_{\mu\nu} G^a_{\mu\nu}\right],$$

where C is a large contour or quark loop in E^4 and P denotes path ordering of the matrices $W_\mu(x)$ along the contour. If C is rectangular with sides T and R ($T \gg R$) in, let us say, the $x_0 - x_3$ plane, then this average has the form

$$(1/Z)\int d[W] I[W] = e^{-T\varepsilon(R)}.$$

Here $\varepsilon(R)$ can be interpreted as the energy of a pair of massive quarks with spatial separation R. This is the function of interest for quark confinement. There is no confinement if $\varepsilon(R)$ does not grow with increasing R, for then it does not cost energy to pull the quarks apart. But if $\varepsilon(R)$ increases with R to arbitrarily large values then quarks may be confined. If instantons are assumed to dominate the functional integration over gauge fields one can calculate $\varepsilon(R)$. This is a rather technical calculation and we just state the results (Callen, Dashen, and Gross, 1977, 1978; Roth, 1977). An ensemble of well separated instantons without long-range interactions (i.e., a dilute gas of instantons) does not confine quarks but only leads to a renormalization of the quark mass. Ensembles with large, overlapping instantons are left out of the analysis because one does not know how to handle them. Presumably many large overlapping instantons and anti-instantons could transmit long-range forces, throughout the entire hadron for example. Trying to solve this probelm is like trying to solve the strong-coupling version of the confinement problem—one does not know how to do it. The hope of instanton physics is that confinement will already become a fact before the coupling gets strong. Instantons do not seem to accomplish this, but merons might (see below).

The generation of quark mass by instanton effects has been discussed in a number of papers. Caldi (1977) has estimated the quark mass with the help of the effective interaction Lagrangian for the instanton–$2N$-quark interaction worked out by 't Hooft (1976b). Carlitz (1978) obtained a formula for this mass,

$$m = C \int d\rho \, [\rho \bar{e}^4(\rho)]^{-2} e^{-8\pi^2/\bar{e}^2(\rho)},$$

(here $\bar{e}(\rho)$ is the running YM coupling, ρ is the instanton "size," and C is a known constant), by calculating the quark propagator in the presence of a dilute instanton gas. The same result was obtained in a calculation (Carlitz and Lee, 1978) of the electromagnetic (quark) current correlation function in the presence of a single instanton with all sizes. (This correlation function turns out to have the behavior for $x^2 \to \infty$ that is characteristic of massive quark fields.)

3. Meron solutions

A meron is a localized, singular solution of the Euclidean gauge theory with one-half unit of topological charge. The name meron derives from a Greek word meaning part (of a unit of topological charge). The topological charge of the meron is concentrated at the point where the solution is singular. This is to be contrasted with the instanton's nonsingular cloud of topological charge. Another difference between instanton and meron solutions is that the latter are not self-dual. Therefore we can expect that merons correspond to some new sort of tunneling in Minkowski space. Unfortunately, we shall not be able to say a great deal about this tunneling. The main difficulty is the absence of an explicit solution describing an arbitrary number of randomly located merons. Without this solution it is not easy to fully understand the role merons play in YM theory. Nevertheless, some parts of the picture are fairly clear.

Exact one- and two-meron solutions are known (de Alfaro, Fubini, and Furlan, 1976). The gauge potential of these solutions has the form

$$eW_\mu = \tfrac{1}{2}[-i(\partial_\mu g)g^{-1}],$$

which, as we see, is one-half of a pure-gauge potential. [It is conjectured (de Alfaro, Fubini, and Furlan, 1978) that all meron solutions have this form.] To investigate the tunneling property of merons let us calculate the current J_μ in Eq. (7.2) that determines the topological charge density D for the potential above. The result is

$$J_\mu = -\tfrac{1}{3}\varepsilon_{\mu\alpha\beta\gamma} \operatorname{tr}[\partial_\alpha g g^{-1} \partial_\beta g g^{-1} \partial_\gamma g g^{-1}].$$

This is one-half of the current obtained from a pure-gauge potential. Thus, if the gauge function $g(x) \in SU(2)$ defines a mapping at $|x| = \infty S^3 \to S^3$ with homotopy index n, then the meron potential above has topological charge $q = n/2$. This shows that the topological meaning of the charge q is basically the same for merons and instantons. However, there are very important differences. An instanton becomes pure-gauge only at infinity, while a meron solution is one-half of a pure gauge potential everywhere. Moreover, the tunneling behavior of merons is not simply related to the tunneling behavior of instantons. The latter tunnel between vacua constructed from the gauge functions ω_n that contain a size parameter (essentially the size of the instanton). Merons have no such parameter. (In Appendix H we show in detail how merons tunnel in the Coulomb gauge.)

A very important property of merons is their connection with magnetic monopoles. In the $W_0 = 0$ gauge the one-meron solution is

$$eW_i^a = -\varepsilon_{ian}(x_n/r^2)[1 - x_0/(x_0^2 + r^2)^{1/2}].$$

As $x_0 \to +\infty$ this potential vanishes, while for $x_0 \to -\infty$

$$eW_i^a \to -\varepsilon_{ian}(2x_n/r^2) = -i(\partial_i\omega)\omega^{-1}, \quad \omega = \omega^{-1} = \hat{r}\cdot\sigma.$$

Here we can see explicitly that the meron interpolates between two pure-gauge vacuum configurations. The one in the distant future is trivially zero. The one in the distant past is not quite trivial, but it is nonsingular at $r = 0$ and has topological charge $Q_T = 0$. For negative x_0 the interpolating three-space configurations are nonsingular at $r=0$. Note that the $x_0=0$ configuration is the Wu–Yang static point monopole [see Eq. (3.8) and below]. This is the most prominent three-space configuration through which the meron solution passes. At earlier (later) times the interpolating field is coming from (returning to) the vacuum through a sequence of configurations which somewhat resemble a monopole.

In the $W_0^a=0$ gauge the two-meron solution with both merons located on the imaginary time axis at points $x_0=a, b$ is

$$eW_i^a = -\varepsilon_{ian}\frac{x_n}{r^2}\left\{1+\frac{r^2+(x_0-a)(x_0-b)}{[(x-a)^2(x-b)^2]^{1/2}}\right\}.$$

For $x_0\to\pm\infty$ this solution becomes the same pure-gauge potential discussed above. If the separation $|a-b|$ is large then W_i^a becomes approximately a Wu-Yang point monopole at times $x_0=a, b$.

Clearly, the physical interpretation of these meron solutions is that colored magnetic monopoles with finite lifetimes can spontaneously appear in the YM vacuum. A gas of merons in E^4 corresponds to a continuously renewed gas of these short-lived monopoles in normal three-space. It has been argued (see Mandelstam, 1975) that a vacuum containing colored monopoles might lead to quark confinement. The mechanism would be that a superconducting state of colored monopoles comes into existence which expels colored electric fields. These fields are the gluon interactions between quarks, and if they are confined to tubes then these tubes are strings which bind the quarks together.

Meron solutions have infinite action because of their singularities. But one can smooth out the gauge potential in the neighborhood of each singularity to obtain an approximate solution with finite action, which only differs from the genuine solution in the immediate vicinity of the singular points. This is, of course, a valid field configuration for the path integral. An important fact is that the action of a smoothed meron pair increases with large separation R like $\ln R$ (Callen, Dashen, and Gross, 1977). This is most easily shown by using the solution $\phi=(\lambda x^2)^{-1/2}$, where ϕ is the usual scalar *ansatz* function. This solution represents a meron at $x=0$ and another at $x^2=\infty$ (see subsection E). From its action density $\mathcal{L}=3/2e^2|x|^4$ the total action of the meron pair is found to be

$$W=\int d^4x\,\mathcal{L}=\frac{3\pi^2}{e^2}\int_\varepsilon^R\frac{d|x|}{|x|}=\frac{3\pi^2}{e^2}\ln\left(\frac{R}{\varepsilon}\right),$$

where $\varepsilon>0, R<\infty$ represent the smearing of the point charges at the origin and infinity, respectively. The value of ε depends on the details of the smearing and is not important because R is large (R is the separation of the two merons). We see that the interaction between widely spaced merons is logarithmic.

A very interesting suggestion, due to Callen, Dashen, and Gross (1977, 1978a) is that an instanton may consist of two merons. (This might seem to imply that merons are, in some sense, more fundamental than instantons. Indeed, one is accustomed to think of pointlike charges as being more fundamental than extended ones.) The merons are tightly bound for small effective coupling e, but when this coupling increases to a certain critical value they will dissociate. This can be illustrated quite easily. The gauge potential for an instanton with center at the origin is

$$eW_\mu^I=[x^2/(x^2+v^2)][-i(\partial_\mu g)g^{-1}],$$

$$g=(\mathbf{x}\cdot\boldsymbol{\sigma}+ix_0)/\sqrt{x^2},$$

where v is the size of the instanton. The gauge potential for a meron at the origin is

$$eW_\mu^M=\tfrac{1}{2}[-i(\partial_\mu g)g^{-1}],$$

with the same matrix g. We see that

$$W_\mu^I=[x^2/(x^2+v^2)]2W_\mu^M,$$

and in the limit $v\to 0$ of vanishing instanton size the instanton becomes two merons. The reader is cautioned not to take this argument too literally because W_μ^I becomes pure gauge in the limit $v\to 0$. However, one can imagine that the merons become slightly separated, so that the limit of W_μ^I is not quite pure gauge.

The quark confinement picture proposed by Callen, Dashen, and Gross (1977, 1978a) has three phases. Here we are viewing the hadron from inside, as a quark would.

Phase I: Very small distances deep within the hadron are probed. The effective YM coupling $\bar{e}$ is very small according to the standard asymptotic freedom argument. Quark masses are practically zero. This means that instantons and anti-instantons are bound together [as explained in part (2) above], and tunneling between the topological vacua $|n\rangle$ is suppressed; the vacuum is degenerate. The perturbative ($n=0$) vacuum is the physical ground state for QCD with massless gluons and (practically) massless quarks.

Phase II: Somewhat larger distances (still deep within the hadron) are probed. The effective coupling $\bar{e}$ has increased beyond a small critical value and a phase transition has occurred: Quark masses are nonzero, instantons are now free, and there is tunneling between the topological vacua $|n\rangle$. The physical ground state $|\theta\rangle=\sum e^{in\theta}|n\rangle$ is a superposition of all of these vacua. Chiral symmetry is broken. Perturbative methods still work, although there are nonperturbative corrections.

Phase III: Distances approaching the hadronic scale are probed. $\bar{e}$ has increased beyond a second (still small) critical value and another phase transition has occurred. Instantons have dissociated into meron pairs: The resulting meron plasma confines quarks. We are near the surface of the hadron in this phase. (If the surface were passed the instanton and meron density would abruptly diverge.) Nonperturbative effects are overwhelmingly strong.

All three phases are analyzed with the help of semiclassical arguments. This means the effective YM coupling *must* be fairly small even in phase III. Otherwise the entire picture collapses.

It is possible to estimate the effective coupling at which a meron plasma might become important (Callen, Dashen, and Gross, 1977, 1978a). Merons are small and their interaction is weak (logarithmic). Therefore they can be treated like the particles in a gas with entropy roughly proportional to $\ln V$, where V is the volume occupied by the gas. This means that the probability of finding a meron within a large volume R^4, say with center at the origin, is roughly proportional to R^4. Suppose that this meron interacts with another which is outside the volume. The action of the pair is then $W = (3\pi^2/e^2)\ln R$. This leads to a total probability for finding the meron within R^4,

$$R^4 e^{-W} \sim R^{4-3\pi^2/e^2},$$

which vanishes (blows up) when $R\to\infty$ for $e^2/3\pi^2 < \frac{1}{4}$ ($>\frac{1}{4}$). The relatively small coupling $e^2/8\pi^2 = \frac{3}{32}$ therefore seems to be a critical value at which merons may become important. For smaller e the action of a meron pair with large separation R grows too rapidly for the merons to come apart. But at the critical coupling the action grows slowly enough that it is just compensated by the increasing entropy. This implies that the merons dissociate. Note that if the meron interaction were stronger than logarithmic, the entropy could not compensate for the diverging action, and there could be no dissociation.

Finally, we come to confinement. A meron plasma is supposed to confine quarks because the gauge potential of an individual meron behaves like $1/|x|$ for all $|x|\neq 0$. This means that the number of merons which make a significant contribution to a Wilson loop integral of area TR is proportional to TR^3. (Here merons which are further than R from the loop are not counted.) Ignoring interactions between merons this would naively seem to imply that the quark interaction energy is $\varepsilon(R)\sim R^3$ so that quarks are confined. Including the logarithmic interaction between merons reduces this to $\varepsilon(R)\sim R^\alpha$ with $\alpha\sim 1$ (see Callen, Dashen, and Gross, 1978a). Again, because of the logarithmic action, the confinement mechanism "turns on" at a critical value of the coupling constant which is still fairly small. This is all very qualitative, but nevertheless suggestive; particularly so when one observes that the argument above does *not* go through for a gas of small instantons. The instanton potential also falls off like $1/|x|$, but it is pure gauge and cannot contribute to the loop integral.

4. Elliptic solution

Instantons and merons are the two basic types of objects which have been found so far in E^4. But a more general exact solution of the Euclidean SU(2) theory is known which interpolates between these basic solutions. The more general solution involves Jacobi elliptic functions, and it is called an elliptic solution. The physical meaning of this solution is not yet understood. One interesting possibility is the conjecture (Callen, Dashen, and Gross, 1977, 1978) that instantons are two merons bound together, and a phase transition occurs in which these merons come unbound. If this is true then the elliptic solution may describe how this happens. The elliptic solution depends on a continuous parameter k, and for $k=1$ it reduces to the one-instanton solution while for $k=0$ it becomes the two-meron solution. These two quite different Euclidean YM field configurations are connected by a continuum of exact solutions which interpolate between them (Cervero, Jacobs, and Nohl, 1977). The topological charge distribution of the elliptic solution depends on k as follows. In the instanton limit $k=1$ there is, of course, a nonsingular cloud of topological charge in E^4 with total charge $q=1$. For $k\lesssim 1$ there appear two point singularities within the cloud. As $k\to 0$ these point singularities increase in strength, eventually becoming merons when $k=0$, while the cloud becomes weaker and totally disappears at $k=0$. This might be a description of the phase transition occurring when an instanton dissociates into two merons.

According to the elliptic two-meron solution an instanton with size λ would dissociate isotropically into merons with separation 2λ. In other words, the merons would come into existence out near the edge of the instanton. This is not *a priori* obvious; one could only have guessed that the meron separation would be $C\lambda$ for some constant C.

There may exist a much more general elliptic solution which interpolates between a $2N$-meron configuration and an N-instanton configuration. Clearly there are many different possibilities for pairs of merons to coalesce into instantons. The sizes of the instantons would depend on the separations of the various meron pairs. Until an explicit multimeron solution has been found, there is little chance of discovering this more general solution, unfortunately.

B. *Ansätze*

1. 't Hooft-Corrigan-Fairlie-Wilczek *ansatz*

We have already encountered the Minkowski-space version of this *ansatz* in Sec. VI. In Euclidean space it is

$$eW_0^a = \mp\partial_a\phi/\phi\,,$$
$$eW_i^a = \varepsilon_{ian}\partial_n\phi/\phi \pm \delta_{ai}\partial_0\phi/\phi\,. \tag{7.10}$$

The SU(2) equations of motion are reduced by this *ansatz* to Eq. (6.5), which we write here in the form

$$\partial_\mu(\Box\phi/\phi) = (2/\phi^2)\partial_\mu\phi\Box\phi\,.$$

If ϕ is a solution of

$$(1/\phi)\Box\phi + \lambda\phi^2 = 0\,, \tag{7.11}$$

then the gauge potential (7.10) is a solution of the YM equations of motion. λ is an arbitrary integration constant. The YM field strengths are

$$eE_n^a = \varepsilon_{nam}\left[\frac{1}{\phi}\partial_0\partial_m\phi - \frac{2}{\phi^2}\partial_0\phi\partial_m\phi\right] \pm \delta_{na}\left[\frac{1}{\phi}\partial_0^2\phi - \frac{1}{\phi^2}(\partial_0\phi\partial_0\phi - \partial_m\phi\partial_m\phi)\right] \pm\left[\frac{1}{\phi}\partial_a\partial_n\phi - \frac{2}{\phi^2}\partial_a\phi\partial_n\phi\right],$$

$$eB_n^a = \pm eE_n^a - \delta_{na}(1/\phi)\Box\phi\,. \tag{7.12}$$

The self-duality condition $E_n^a = \pm B_n^a$ implies $(\Box\phi)/\phi = 0$. [This statement is slightly too strong, as we shall see in the following subsection. $(\Box\phi)/\phi$ can be nonzero for a self-dual field if E_n^a and B_n^a are both proportional to δ_{na}.] From Eq. (7.12) we easily find

$$-\mathcal{L} = \pm D - (3/2e^2)(\Box\phi/\phi)^2\,. \tag{7.13}$$

Equations (6.13) and (6.14) for $\mathcal{L}$ are still valid here, and the energy–momentum tensor is given by Eq. (6.17). A particularly useful formula is

$$D=\pm(1/2e^2)\partial_\mu[\Box(\partial_\mu\phi/\phi)]. \tag{7.14}$$

Then from Eq. (7.5) we find a formula for the topological charge of a nonsingular solution:

$$q=\pm(1/16\pi^2)\int_{x^2=\infty} d\Omega\, x^2 x_\mu[\Box(\partial_\mu\phi/\phi)]. \tag{7.15}$$

The behavior of ϕ as $x^2\to\infty$ is closely related to the value of the topological charge q. Suppose that

$$\phi\to C/(x^2)^\alpha,\quad x^2\to\infty. \tag{7.16}$$

Then using Eq. (7.15) one easily verifies that

$$q=\pm\alpha. \tag{7.17}$$

If ϕ is a solution of $\Box\phi+\lambda\phi^3=0$ then only two values of α are allowed: (i) $\alpha=1$, which is the instanton solution with $\lambda=0$, and (ii) $\alpha=1/2$, which is the meron solution with λ arbitrary. Note that in deriving Eq. (7.17) we have ignored singularities. We now discuss the way in which singularities in ϕ can change q by an arbitrary integer.

Suppose that ϕ has a singularity at the point $x=v$ of the form

$$\phi\approx C/(x-v)^2,\quad \partial_\mu\phi/\phi\approx -2(x-v)_\mu/(x-v)^2. \tag{7.18}$$

Near this point the corresponding YM potential is pure gauge,

$$eW_\mu\approx -i(\partial_\mu\omega)\omega^{-1};$$
$$\omega=(1/\sqrt{(x-v)^2})[\boldsymbol{\sigma}\cdot(\mathbf{x}-\mathbf{v})\pm i(x_0-v_0)]. \tag{7.19}$$

There appears to be a singularity in W_μ at the point $x=v$, but it can be gauge transformed away. Near this point the YM field is gauge equivalent to the vacuum. However, the scalar function (7.18) is genuinely singular there. One consequence of this is that Eq. (7.14) for D breaks down at $x=v$. Let us calculate D near this point using ϕ in Eq. (7.18):

$$(e^2/8\pi^2)D\approx\pm(1/16\pi^2)\Box[-4/(x-v)^2]=\pm\delta(x-v). \tag{7.20}$$

According to this result there is a point topological charge with unit strength at $x=v$. But we have just seen that this is wrong; there is a vacuum at this point. To correct matters we must subtract the delta-function singularity from D. This is conveniently done as follows. In the expression (7.14) for D we can make the replacement $\phi\to(x-v)^2\phi$ and use the formula

$$\Box\Box\ln(x-v)^2=-4\pi^2\delta(x-v) \tag{7.21}$$

to obtain the correct result

$$D=\pm(1/2e^2)\Box\Box\ln\phi$$
$$\to\pm(1/2e^2)\Box\Box\ln[(x-v)^2\phi]=D\mp(2\pi^2/e^2)\delta(x-v). \tag{7.22}$$

This operation only changes D at the point $x=v$ where the original formula is wrong.

One can regard the operation just defined as a singular (but allowed) gauge transformation which eliminates the spurious singularity from D. Quite generally, this situation will arise whenever there is a *delta-function singularity with unit strength* in the topological charge density $(e^2/8\pi^2)D$. No matter what kind of YM solution is involved, near the singular point one can write the gauge potential in the form (7.19), which shows that it is nonsingular there. Thus the delta-function singularity in D must be spurious. For any point topological charge with strength $q\neq\pm1$ this is no longer true. Such a singularity is not spurious; the gauge field really does have a singularity.

2. Witten's *ansatz*

Witten (1977) introduced the following *ansatz*, for the Euclidean gauge potential,

$$-eW_0^a=\frac{x_a}{r}A_0,$$
$$-eW_i^a=\varepsilon_{ian}\frac{x_n}{r^2}(1+\phi_2)+\frac{x_a x_i}{r^2}A_1+\left(\delta_{ai}-\frac{x_a x_i}{r^2}\right)\frac{1}{r}\phi_1, \tag{7.23}$$

where $A_0(x_0r)$, $A_1(x_0,r)$, $\phi_1(x_0,r)$, and $\phi_2(x_0,r)$ are functions of $r=(x_1^2+x_2^2+x_3^2)^{1/2}$ and x_0. This *ansatz* is symmetric about the time axis in E^4. In three-dimensional space this corresponds to symmetry under spatial rotations.

The *ansatz* (7.23) is form invariant under the local U(1) subgroup of SU(2) with elements

$$\omega=\exp[\tfrac{1}{2}i\,f(x_0,r)\hat{r}\cdot\boldsymbol{\sigma}]. \tag{7.24}$$

Under this gauge transformation the *ansatz* functions transform as follows:

$$A_\mu\to A'_\mu=A_\mu-\partial_\mu f,$$
$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\to\begin{pmatrix}\phi'_1\\ \phi'_2\end{pmatrix}=\begin{pmatrix}\cos f & \sin f\\ -\sin f & \cos f\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}. \tag{7.25}$$

Here, and below, we use the notation

$$\partial_0=\partial/\partial x_0,\quad \partial_1=\partial/\partial r.$$

Equations (7.25) reveal the most interesting property of Witten's *ansatz*: It reduces the pure YM problem in E^4 to a simpler problem, namely the Abelian Higgs model in two space-time dimensions (with a curved metric $g_{\mu\nu}=r^2\delta_{\mu\nu}$). The Abelian gauge potential is $A_\mu=(A_0,A_1)$ and the complex (charged) Higgs field in the model is $\phi=\phi_1-i\phi_2$. The Abelian gauge group is, of course, the subgroup (7.24) of the SU(2) gauge group. To emphasize this property of his *ansatz* Witten introduces the notation

$$F_{\mu\nu}\equiv\partial_\mu A_\nu-\partial_\nu A_\mu,$$
$$D_\mu\phi_a\equiv\partial_\mu\phi_a+\varepsilon_{ab}A_\mu\phi_b, \tag{7.26}$$

where all indices take two values.

The YM field strengths obtained from Eq. (7.23) are

$$-eE_n^a=\varepsilon_{nak}x_k\frac{1}{r^2}D_0\phi_2+\frac{x_a x_n}{r^4}F_{01}+\left(\delta_{an}-\frac{x_a x_n}{r^2}\right)\frac{1}{r}D_0\phi_1, \tag{7.27}$$

$$eB_n^a=-\varepsilon_{nak}\frac{x_k}{r^2}D_1\phi_1+\frac{x_a x_n}{r^4}(1-\phi_1^2-\phi_2^2)+\left(\delta_{an}-\frac{x_a x_n}{r^2}\right)\frac{1}{r}D_1\phi_2.$$

Using these expressions one can easily calculate the action density of the Euclidean YM theory,

$$\mathcal{L} = 2\left[(1/2r^2)D_\mu\phi_a D_\mu\phi_a + \tfrac{1}{8}F_{\mu\nu}F_{\mu\nu} + (1/4r^2)(1-\phi_1^2-\phi_2^2)^2\right]. \quad (7.28)$$

The pseudoscalar density (7.1) is found to be

$$2e^2 D = (1/r^2)\partial_\mu J_\mu , \quad (7.29)$$

where

$$J_\mu = 2\varepsilon_{\mu\nu}[\varepsilon_{ab}\phi_a D_\nu\phi_b + A_\nu]. \quad (7.30)$$

The equations of motion following from Eq. (7.23) are

$$\partial_\mu(r^2F_{\mu\nu}) = 2\varepsilon_{ab}\phi_a D_\nu\phi_b ,$$
$$r^2 D_\mu D_\mu\phi_a = \phi_a(1-\phi_1^2-\phi_2^2) , \quad (7.31)$$

or more explicitly,

$$\partial_0(r^2F_{01}) = 2[\phi_1 D_1\phi_2 - \phi_2 D_1\phi_1] ,$$
$$\partial_1(r^2F_{01}) = -2[\phi_1 D_0\phi_2 - \phi_2 D_0\phi_1] ,$$
$$r^2[\partial_\mu(D_\mu\phi_a) + \varepsilon_{ab}A_\mu(D_\mu\phi_b)] = \phi_a(1-\phi_1^2-\phi_2^2) . \quad (7.32)$$

It is always possible to gauge-transform one of the fields ϕ_a to zero. For example, to make $\phi_1'=0$ we choose $f=\tan^{-1}(\phi_1/\phi_2)$ in Eq. (7.25). Setting $\phi_1=0$ in the equations of motion of course simplifies these equations considerably.

There exists a region of overlap between Witten's *ansatz* (7.23) and the ϕ^4 *ansatz* (7.10). If $\phi=\phi(x_0, r)$ depends only on r and x_0 then the two *ansätze* coincide if $A_1=\phi_1/r$ and

$$A_0 = \pm\partial_1\phi/\phi , \quad A_1 = \mp\partial_0\phi/\phi , \quad 1+\phi_2 = -r\partial_1\phi/\phi .$$

The gauge potential A_μ here automatically satisfies the Lorentz condition $\partial_\mu A_\mu=0$, which is essential for the construction of solutions in Witten's approach. Now an interesting question is: Under what conditions are the two *ansätze* equivalent? Manton (1978a) has answered this question. He shows that if $\partial_\mu A_\mu=0$ is satisfied, then *any* solution obtained from Witten's *ansatz* is gauge equivalent to a ϕ^4 solution. This result holds independently from assumptions such as self-duality.

Meron solutions correspond to $A_\mu=0$, $\phi_1=0$ (or gauge transformations thereof) in the *ansatz* (7.23). The equations of motion (7.32) then reduce to

$$r^2(\partial_0\partial_0 + \partial_1\partial_1)\phi_2 = \phi_2(1-\phi_2^2) ,$$

which is the equation of motion for a ϕ^4 theory in two space-time dimensions (with the curved metric $g_{\mu\nu} = r^2\delta_{\mu\nu}$). According to Manton's result above the gauge potentials obtained from solutions of this equation are gauge equivalent to potentials (7.10) for some $\phi = \phi(x_0, r)$.

An extension of Witten's *ansatz* has been introduced by Leznov and Saveliev (1978). The more general *ansatz* depends on a constant four-vector b_μ in such a way that it reduces to Witten's *ansatz* when $b_\mu=(1,0)$. The four *ansatz* functions depend on two variables $x_\mu b_\mu$ and $[x^2-(x_\mu b_\mu)^2]^{1/2}$ which become x_0 and r for $b_\mu=(1,0)$. Leznov and Saveliev discuss the symmetries of their *ansatz* and give equations of motion and other formulas of interest in terms of the *ansatz* functions. Moreover, they give an interesting expression for the N-instanton solution with all instantons on a line.

C. Belavin-Polyakov-Schwartz-Tyupkin instanton

The famous instanton solution of the Euclidean SU(2) gauge theory was found by Belavin, Polyakov, Schwartz, and Tyupkin (1975) (hereafter BPST). This solution is manifestly nonsingular,

$$eW_0^a = \mp\frac{2x_a}{x^2+v^2} ,$$
$$eW_i^a = -\varepsilon_{ian}[2x_n/(x^2+v^2)] \pm \delta_{ai}[2x_0/(x^2+v^2)]. \quad (7.33)$$

Moreover, it is self-dual:

$$eB_n^a = \pm eE_n^a = \delta_{an}4v^2/(x^2+v^2)^2 . \quad (7.34)$$

Therefore the energy–momentum tensor is identically zero. The potential (7.33) falls off like $O(1/\sqrt{x^2})$ as $x^2 \to \infty$ and normally this behavior would imply that the solution has infinite action. But in Eq. (3.34) we see that the field strengths decrease like $O(1/x^4)$ and, because the solution is nonsingular, this means that the action is finite. The field strengths decrease this rapidly for large x^2 because the instanton potential becomes a pure-gauge potential in the limit $x^2\to\infty$. To show this explicitly we can rewrite Eq. (7.33) in the form

$$eW_\mu = [x^2/(x^2+v^2)][-i(\partial_\mu g)g^{-1}] ; \quad (7.35)$$
$$g = (1/\sqrt{x^2})(\mathbf{x}\cdot\boldsymbol{\sigma} \mp ix_0) , \quad g^{-1} = (1/\sqrt{x^2})(\mathbf{x}\cdot\boldsymbol{\sigma} \pm ix_0) .$$

The instanton solution defines a particular mapping of the sphere S^3 at infinity onto the SU(2) group manifold. This mapping has topological index $q=\mp1$.

Actually, we are being somewhat careless with our terminology here because two topologically distinct field configurations are involved. The lower (upper) sign in Eqs. (7.33) and (7.35) gives the instanton (anti-instanton) solution with topological charge $q=+1$ ($q=-1$). In Eq. (7.34) we see that the instanton is self-dual and the anti-instanton is self-antidual. However, we find it convenient to discuss the two solutions together, and to loosely refer to an instanton that can have either charge, when this cannot lead to confusion.

To show that the potential (7.33) really is a solution of the equations of motion we observe that it can be written in the form

$$eW_0^a = \pm\partial_a\phi/\phi ,$$
$$eW_i^a = \varepsilon_{ian}\partial_n\phi/\phi \mp \delta_{ai}\partial_0\phi/\phi , \quad (7.36)$$

where

$$\phi = C/(x^2+v^2) , \quad \partial_\mu\phi/\phi = -[2x_\mu/(x^2+v^2)]. \quad (7.37)$$

Moreover, for $C=(8v^2/\lambda)^{1/2}$, ϕ is a solution of the equation $\Box\phi + \lambda\phi^3=0$. Since Eq. (7.36) coincides with the ϕ^4 *ansatz* (7.10) (with $\mp$ in place of $\pm$), it follows that the YM equations of motion are satisfied. Now the attentive reader will notice an apparent contradiction here, namely that ϕ satisfies $\Box\phi+\lambda\phi^3=0$ with nonzero λ and not $\Box\phi=0$. In subsection B it was mentioned that $\Box\phi=0$ is the condition for a self-dual solution based on the ϕ^4 *ansatz* (7.10). But there is no contradiction, as we now show. The reason is that the extra term in Eq. (7.12), which would seem to conflict with the self-duality of the solution, has the same form as the field strengths E_n^a and B_n^a in Eq. (7.34). In fact we can rewrite the self-duality equation (7.34) as

$$eB_n^a = \mp eE_n^a - \delta_{an}(\Box\phi/\phi)\,,$$

which is precisely Eq. (7.12) with the change $\pm \to \mp$ taken into account.

Now that we have the instanton solution in the form (7.36) we can easily calculate the topological charge. Indeed, this has already been done in Eq. (7.17). Since $\phi \to 1/x^2$ as $x^2 \to \infty$, it follows that $|q| = 1$. The topological charge density is easily found with the help of Eq. (7.14),

$$(e^2/8\pi^2)D = \mp[6v^2/\pi^2(x^2+v^2)^4]\,. \qquad (7.38)$$

Integration over all space gives $q = \mp 1$, as it should.

In Eq. (7.33) it is clear that the instanton is centered at $x_\mu = 0$. Translation in E^4, $x_\mu \to x_\mu + a_\mu$, simply moves the instanton to a new location. The constant v^2 in the denominator determines the size of space-time extent of the instanton. For $v^2 \to \infty$ it extends (thinly) over all of E^4; this is a large instanton. For $v^2 \approx 0$ the effects of the (small) instanton are concentrated near the point $x^2 = 0$. The fact that the SU(2) gauge theory has no built-in scale is responsible for this arbitrariness in the size of the instanton. Note that the instanton cannot become so small that it is pointlike, for when $v^2 = 0$ the instanton potential becomes pure gauge. Taking the limit $v^2 \to 0$ destroys the instanton.

There is a connection between the BPST solution above, and the "kink" solution of the ϕ^4 theory in two space-time dimensions (Marciano and Pagels, 1976; Calvo, 1977). This can be shown by a change of dependent and independent variables;

$$(2/\phi)(d\phi/dx^2) = (1/x^2)[f(y) - 1]\,,$$

$$y = \tfrac{1}{2}\ln(x^2/v^2)\,, f(y) = \tanh y = (v^2 - x^2)/(v^2 + x^2)\,.$$

The function $f(y)$ satisfies

$$f'' + 2f - 2f^3 = 0\,,$$

which is the static equation of motion for the two-dimensional theory with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\partial^\mu f \partial_\mu f - \tfrac{1}{2}(1-f^2)^2\,.$$

$f = \tanh y$ is the well known static kink solution of this theory. Translation of the kink, $y \to y + a$, corresponds to a change of scale in the instanton solution $v^2 \to v^2/\lambda$ with $a = \frac{1}{2}\ln\lambda$. The two vacua of the kink theory, $y = \pm\infty$, correspond, respectively, to the pure-gauge instanton potential at $x^2 = \infty$ and $W_\mu = 0$ at $x^2 = 0$ [see Eq. (7.35)].

It is generally conjectured that Euclidean YM solutions which are not self-dual (or self-antidual) cannot have finite action. (Nonsingular self-dual solutions, on the other hand, are guaranteed to have minimum action in any given topological sector.) Support for this conjecture can be found in the one-instanton sector. Let us restrict the discussion to solutions which are symmetric under rotations in E^4 (simultaneous gauge transformations are allowed). Many authors have investigated such field configurations (Hwa, 1977; Ball and Sen, 1977; Calvo, 1977b; Actor, 1978b; Mukherjee and Roy, 1978; Wada, 1978; Arik and Williams, 1978). It is found that the instanton solution is the *only one with finite action*. All other $O(4)$ symmetric solutions, self-dual or non-self-dual, have infinite action. The elliptic solutions which we discuss later show clearly how this comes about. For elliptic parameter $k = 1$ these solutions become the nonsingular instanton. But for all $k < 1$ the elliptic solutions are singular at $x^2 = 0$ and the action is (logarithmically) divergent. The value $k = 0$ is unique in this respect, and this illustrates the uniqueness of the instanton solution. (See subsection G below for more details.)

The conformal properties of the instanton solution have been investigated by Jackiw and Rebbi (1976b). Dilatations $x_\mu \to \lambda x_\mu$ change its size, $v^2 \to v^2/\lambda^2$. Thus a large instanton becomes a small one when E^4 is stretched enough. More interesting is the behavior under coordinate inversion, $x_\mu \to x_\mu/x^2$. This operation changes an instanton into an anti-instanton, hence $q = +1$ into $q = -1$. To prove this statement we note that $x_\mu/\sqrt{x^2}$, and hence the matrix g in Eq. (7.35), is unchanged by coordinate inversion. The coefficient in Eq. (7.35) becomes

$$\frac{x^2}{x^2+v^2} \to \frac{1/v^2}{x^2+1/v^2}\,.$$

Comparing the resulting expression with Eq. (7.39) below, which is a gauge-transformed version of Eq. (7.33), we see that coordinate inversion changes the sign of the instanton charge, and the size v into $1/v$. Jackiw and Rebbi go on to show that the instanton solution is invariant under the $O(5)$ subgroup of the $O(5,1)$ conformal group in E^4. This large symmetry group is another indication of the uniqueness of the instanton solution.

D. N-instanton solutions

1. 't Hooft's solution

The first multi-instanton solution we discuss is the one found by 't Hooft (1976c), and later generalized slightly by other authors (Jackiw and Rebbi, 1977a; Ansourian and Ore, 1977). The latter solution is still the most general explicit self-dual solution known. We emphasize that the solution to follow describes configurations of N instantons *or* N anti-instantons, but not a mixture of instantons and anti-instantons. It is generally believed that no exact solution exists which describes one instanton and one anti-instanton.

't Hooft was able to write down a multi-instanton solution after he had discovered an *ansatz* which linearizes the equations of motion, namely *ansatz* (7.10). Now this is not quite obvious when one looks at the BPST instanton solution in its original form (7.36), because ϕ in Eq. (7.37) satisfies $\Box\phi + \lambda\phi^3 = 0$ with $\lambda \neq 0$ and the problem is not linearized. However, 't Hooft noticed that the BPST solution can be written in a different form with a new scalar function ϕ which satisfies $(1/\phi)\Box\phi = 0$. This essentially linearizes the problem, as we now show.

Let us gauge-transform the BPST solution (7.35) using the inverse matrix g^{-1}. The new potential is

$$eW'_\mu = [x^2/(x^2+v^2)]g^{-1}[-i(\partial_\mu g)g^{-1}]g - i(\partial_\mu g^{-1})g$$
$$= [v^2/(x^2+v^2)][-i(\partial_\mu g^{-1})g]\,. \qquad (7.39)$$

Explicitly, the components of W'_μ coincide with the *ansatz* (7.10),

$$eW'^a_0 = \mp \partial_a\phi/\phi\,,$$
$$eW'^a_i = \varepsilon_{ian}\partial_n\phi/\phi \pm \delta_{ai}\partial_0\phi/\phi\,; \tag{7.40}$$

where

$$\phi = 1 + v^2/x^2\,,\quad \partial_\mu\phi/\phi = -2v^2x_\mu/x^2(x^2+v^2)\,. \tag{7.41}$$

This scalar function ϕ satisfies $(\Box\phi)/\phi = 0$ because

$$\Box[1/(x-a)^2] = -4\pi^2\delta(x-a)\,.$$

Therefore the instanton solution is self-dual, as we have already seen. Although ϕ is singular at $x^2=0$ this is not true of the YM potential W'_μ, as shown explicitly by Eq. (7.39). When $x^2\to 0$ the potential W'_μ becomes pure gauge, and there is no singularity.

The solution (7.40) and (7.41) represents an instanton with size $|v|$ centered at the origin. Given the BPST solution in this form, it is trivial to write down an exact solution of the equations of motion which represents N instantons with arbitrary sizes, centered at arbitrary points $x=a_n$ in E^4, namely

$$\phi = 1 + \sum_{n=1}^{N} \frac{b_n}{(x-a_n)^2}\,. \tag{7.42}$$

This scalar function satisfies

$$(1/\phi)\Box\phi = 0\,.$$

The corresponding YM field is self-dual and nonsingular (by the argument following Eq. (7.18)), and it has topological charge $q=N$ (as we show immediately). Therefore it is, as claimed, an N-instanton solution. Note that the parameters b_n are the (size)^2 of these instantons, and therefore they are positive.

Let us now prove that the scalar function (7.42) leads to topological charge $q=\pm N$ in the YM theory. For this we need Eqs. (7.13) and (7.14), which we write in the form

$$\pm 2e^2 D = -2e^2\mathfrak{L} = \Box\Box\ln\phi\,. \tag{7.43}$$

This is an explicit formula for the densities $\mathfrak{L}$ and D, both of which should be nonsingular because the YM solution is nonsingular. But there are delta-function singularities in Eq. (7.43), because Eq. (7.43) breaks down at the points where ϕ is singular as we have already mentioned in subsection B. To show this explicitly we write down the identity

$$\begin{aligned}\Box\Box\ln[(x-a_1)^2\cdots(x-a_N)^2\phi] &= \Box\Box\ln\phi + \Box\Box\ln(x-a_1)^2 + \cdots + \Box\Box\ln(x-a_N)^2 \\ &= \Box\Box\ln\phi - 4\pi^2\delta(x-a_1) - \cdots - 4\pi^2\delta(x-a_N)\,.\end{aligned} \tag{7.44}$$

The left-hand side here is nonsingular because $(x-a_1)^2\cdots(x-a_N)^2\phi$ is finite for $|x|<\infty$. Therefore the delta functions on the right must cancel singularities in $\Box\Box\ln\phi$. These singularities are not physical (because the YM field is nonsingular) and they have to be canceled. Away from these singular points, the factor $(x-a_1)^2\cdots(x-a_N)^2$ in the argument of the logarithm on the left in Eq. (7.44) has no effect whatever. Therefore the correct Lagrangian and pseudoscalar density for 't Hooft's solution are given by

$$\begin{aligned}\pm 2e^2D &= -2e^2\mathfrak{L} \\ &= \Box\Box\ln[(x-a_1)^2\cdots(x-a_N)^2\phi]\,.\end{aligned} \tag{7.45}$$

To complete the argument, we recall the result in Eqs. (7.16) and (7.17) above, which relates the behavior of the argument of ln as $x^2\to\infty$ to the topological charge. In Eq. (7.45) we see that the exponent of the argument is $\alpha=-N$, because of the N extra factors, and therefore the topological charge is $q=\pm\alpha=\mp N$, as it should be for an N-instanton solution.

Jackiw and Rebbi (1977a; see also Ansourian and Ore, 1977) pointed out that the scalar function (7.42) is not conformal invariant. In other words, this function, and the corresponding N-instanton solution of the YM theory, change their appearance under conformal transformations. A scalar function that is conformal invariant is

$$\phi = \sum_{n=0}^{N} b_n/(x-a_n)^2\,. \tag{7.46}$$

As Jackiw and Rebbi show, this function is form invariant under the full Euclidean conformal group. One easily verifies that it, too, gives rise to an N-instanton YM solution because the topological charge is $q=\mp N$. We can even change the solution (7.46) into the 't Hooft solution (7.42) by taking the limit $b_0\to\infty$, $a_0^2\to\infty$ with $b_0^2/a_0^2=1$. But this is a rather nontrivial limit, and it is not surprising that the parameters of the solution (7.46) no longer have an obvious physical interpretation. There are $5N+4$ parameters in Eq. (7.46) (an overall constant factor is irrelevant). This is four more than in the 't Hooft solution. The four extra parameters are necessary for conformal invariance.

Let us turn to the question of interactions between instantons. Because of self-duality, the N-instanton (or N-anti-instanton) solution saturates the lower bound on the total action,

$$W = (8\pi^2/e^2)\,|q| = (8\pi^2/e^2)N\,.$$

The action is therefore independent of the positions of the instantons. This means that instantons do not interact with instantons, nor anti-instantons with anti-instantons. [A correction to this statement should be made when the centers of two instantons are made to coincide. Then, as we see in Eq. (7.42), the two instantons merge to become one instanton with a new size parameter, and a unit of topological charge gets lost.]

There is a *logarithmic* interaction between instanton and anti-instanton. A rough argument demonstrates this. Consider an instanton and an anti-instanton with large separation R ($R\gg$ the instanton size). The gauge potential representing this situation is $A_\mu = W_\mu + \overline{W}_\mu$ where W_μ and $\overline{W}_\mu$ are the potentials of the instanton and anti-instanton. A_μ is an approximate solution of the equations of motion. The field strengths calculated from A_μ are, in an obvious notation,

$$A^a_{\mu\nu} = G^a_{\mu\nu} + \overline{G}^a_{\mu\nu} + e\varepsilon_{abc}[W^b_\mu\overline{W}^c_\nu - W^b_\nu\overline{W}^c_\mu]\,.$$

Now concentrate on the contribution of the last term to the total action. W_μ and $\overline{W}_\mu$ behave like $1/|x|$ and $1/|x-R|$, respectively (the instanton is located at the origin), and in calculating the action we find a contribution ~ln$R$ coming from the small-x region. Evidently this logarithmic interaction is attractive, for when the instanton and anti-instanton overlap they tend to cancel. When their centers coincide, and they have the same size, then the cancellation is complete and a pure vacuum is the result.

The 't Hooft N-instanton solution has $5N$ parameters, and its conformal-invariant generalization has $5N+4$ parameters. One may wish to know if further generalizations are possible, and if the maximum number of parameters is limited or unlimited. Answers to these questions have been given. The most general N-instanton solution [defined as a self-dual solution of the pure SU(2) gauge theory with topological charge $q=N$] depends on $8N-3$ parameters. These parameters have the following physical interpretation. $5N$ of them determine the position and size of the instantons. A further $3N$ parameters are needed to specify the orientations of the instantons in SU(2) space. [Instantons are SU(2) vectors because W_μ^a transforms like the adjoint representation under global SU(2) transformations.] But three of the orientation parameters are meaningless because a global SU(2) transformation cannot have any physical effect. This leaves $8N-3$ parameters (Jackiw and Rebbi, 1977b; Brown, Carlitz, and Lee, 1977). Note that *ansatz* (7.10) does not allow any freedom in the SU(2) orientation of the individual instantons. Essentially, the orientation is determined by the positions of all the instantons. Therefore $3N-3$ orientation parameters are missing, leaving the $5N$ parameters in 't Hooft's solution.

A more rigorous derivation of the number $8N-3$ has been given (Schwartz, 1977; Atiyah, Hitchin, and Singer, 1977). This work is based on a very fundamental theorem in mathematics known as the Atiyah–Singer index theorem. An interesting aspect of this approach is that normalizable zero-eigenvalue solutions of the Euclidean Dirac operator (with a self-dual gauge potential) determine the instanton number. This turns out to be equal to the number of such solutions with positive chirality or helicity minus the number with negative chirality. The explicit zero-mode solutions of the Dirac operator for the 't Hooft N-instanton solution have been found (Jackiw and Rebbi, 1977c; Grossman, 1977).

2. Witten's solution

The first multi-instanton solution was constructed by Witten (1977) with the help of *Ansatz* (7.23). Because this *ansatz* is $O(3)$ symmetric the instantons are necessarily arranged along the imaginary-time axis in E^4. 't Hooft's solution above is more general (and simpler), and for this reason Witten's solution has received less attention. Nevertheless, this solution has interesting mathematical properties and it deserves study.

Witten begins by imposing the self-duality condition $E_n^a = B_n^a$. From Eq. (7.27) we see that self-duality implies

$$\partial_0\phi_1 + A_0\phi_2 = \partial_1\phi_2 - A_1\phi_1\,, \tag{7.47}$$

$$\partial_0\phi_2 - A_0\phi_1 = -(\partial_1\phi_1 + A_1\phi_2)\,, \tag{7.48}$$

$$\partial_0 A_1 - \partial_1 A_0 = (1/r^2)(1-\phi_1^2-\phi_2^2)\,. \tag{7.49}$$

Then the gauge $\partial_\mu A_\mu = 0$ is chosen, which implies that A_μ has the form $A_0 = \partial_1\psi$, $A_1 = -\partial_0\psi$. Then Eqs. (7.47) and (7.48) become

$$[\partial_0 - (\partial_0\psi)]\phi_1 = [\partial_1 - (\partial_1\psi)]\phi_2\,,$$

$$[\partial_0 - (\partial_0\psi)]\phi_2 = -[\partial_1 - (\partial_1\psi)]\phi_1\,.$$

A change of dependent variable $\phi_1 = \chi_1 e^\psi$, $\phi_2 = \chi_2 e^\psi$ reduces the latter two equations to

$$\partial_0\chi_1 = \partial_1\chi_2\,,\quad \partial_0\chi_2 = -\partial_1\chi_1\,.$$

These are just Cauchy–Riemann equations that guarantee the differentiability or analyticity of $f(z) \equiv \chi_1 - i\chi_2$, $z \equiv r + ix_0$. Only the third self-duality equation (7.49) remains to be solved. In terms of the new variables it is

$$-r^2(\partial_0^2+\partial_1^2)\psi = 1 - e^{2\psi}(f^*f)\,. \tag{7.50}$$

A further change of variables,

$$\psi = \ln r - \tfrac{1}{2}\ln f^*f + \rho\,,$$

brings Eq. (7.50) into the form

$$(\partial_0^2+\partial_1^2)\rho = e^{2\rho}. \tag{7.51}$$

Here we have used the fact that

$$(\partial_0^2+\partial_1^2)\ln f^*f = 0$$

for any analytic function f.

The general solution of Eq. (7.51) is known,

$$\rho = -\ln[\tfrac{1}{2}(1-g^*g)] + \tfrac{1}{2}\ln|dg/dz|^2\,,$$

where $g=g(z)$ is any analytic function. Changing back to the variables f and ψ we see that

$$f = dg/dz\,,\quad \psi = -\ln[(1/2r)(1-g^*g)] \tag{7.52}$$

is a solution of Eq. (7.50). For ψ to be nonsingular, Witten chooses

$$g(z) = \prod_{n=0}^{N}\left(\frac{a_n - z}{a_n^* + z}\right),\quad \mathrm{Re}\,a_n > 0\,. \tag{7.53}$$

Because $\mathrm{Re}\,a_n > 0$, $g(z)$ has no poles in the (physical) half-plane $\mathrm{Re}\,z = r \geq 0$. Also, $gg^* = 1$ along the boundary $r=0$ so that ψ is not singular there.

Where are the instantons? It turns out that their positions and sizes are determined by the zeros of the function dg/dz. [This function has N zeros when g is given by Eq. (7.53).] Let $z=z_i$ be one of them. The rule is that $\mathrm{Im}\,z_i$ is the position of the instanton along the imaginary-time axis and $\mathrm{Re}\,z_i$ is its size parameter. This is not trivial to prove. Nevertheless, we can easily convince ourselves that the zeros of $f=dg/dz$ are quite unique points, for Eq. (7.50) is invariant under the transformation

$$f \to f' = hf\,,\quad \psi \to \psi' = -\tfrac{1}{2}\ln h^*h + \psi\,, \tag{7.54}$$

where h is any analytic function. The zeros of f are unchanged by this (gauge) transformation.

It is instructive to calculate the one-instanton potential. Take $N=1$ in Eq. (7.53) and let the parameters $a_0 = a_1 = \lambda$ be equal and real. The *ansatz* functions are most easily calculated from f' and ψ' defined by Eq. (7.54) with $h = (\lambda+z)^4$,

$$f' = h(dg/dz) = -4\lambda(\lambda - z)(\lambda + z)\,,$$

$$\psi' = -\ln[(1/2r)(1-g^*g)\sqrt{h^*h}] = -\ln[4\lambda(x^2+\lambda^2)]\,.$$

f' vanishes at $z=\lambda$, which means that the instanton is centered at $x_0=0$ and has size λ. The *ansatz* functions are

$$A_0 = -2r/(x^2+\lambda^2)\,,\ A_1 = 2x_0/(x^2+\lambda^2)\,,$$

$$\phi_1 = -(\lambda^2+x_0^2-r^2)/(x^2+\lambda^2)\,,\ \phi_2 = -2rx_0/(x^2+\lambda^2)\,. \quad (7.55)$$

To obtain the usual form of the instanton potential (7.33) (lower sign) we have only to perform the gauge transformation (7.25) with $f=-\pi/2$.

3. Atiyah-Hitchin-Drinfeld-Manin construction

Atiyah, Hitchin, Drinfeld, and Manin (1978) have shown how to construct the general self-dual solution of a Yang–Mills theory with arbitrary compact gauge group. This very general constructive procedure reduces the self-duality equations to purely algebraic conditions that are much easier to solve. An explicit, fully general solution of this algebraic problem has not been given. Nevertheless, the algebraic construction of Atiyah *et al.* is the next best thing to a complete solution of the self-duality problem. We now show how this construction works for the SU(2) gauge group. The reader is referred to recent papers by Corrigan, Fairlie, Goddard, and Templeton (1978) and Christ, Weinberg, and Stanton (1978) for a more detailed discussion.

For the gauge group SU(2) we have to work with quaternions. These are 2×2 complex matrices of the form

$$Q = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix},$$

which depend on four real numbers. It is convenient to introduce a basis

$$e_\mu = (1, -i\sigma)$$

and to write quaternions as follows,

$$Q = e_\mu Q_\mu = \begin{pmatrix} Q_0 - iQ_3 & -Q_2 - iQ_1 \\ Q_2 - iQ_1 & Q_0 + iQ_3 \end{pmatrix}.$$

Then $Q^+Q = Q_\mu Q_\mu I_2 = (\det Q)I_2$ and the inverse of Q is $Q^{-1} = Q^+/\det Q$. If $\det Q = 1$ then $Q \in \mathrm{SU}(2)$. A position vector x_μ in E^4 is represented by the quaternion

$$x = e_\mu x_\mu\,,\ e_\mu = \partial_\mu x\,.$$

The following properties of the basis quaternions e_μ are easy to verify:

$$e_\mu^+ e_\nu = i\sigma_a \eta_{a\mu\nu} + \delta_{\mu\nu} I_2\,,$$

$$e_\mu e_\nu^+ = i\sigma_a \bar\eta_{a\mu\nu} + \delta_{\mu\nu} I_2\,,$$

where η and $\bar\eta$ are the symbols introduced by 't Hooft (1976a).

The construction of Atiyah *et al.* begins with the *ansatz* for the SU(2) gauge potential

$$eW_\mu = iM^+\partial_\mu M = i[M_0^+\partial_\mu M_0 + \cdots + M_n^+\partial_\mu M_n]\,,$$

where $M=M(x)$ is a column vector of quaternions M_0, $M_1, \ldots, M_n$ with $n+1$ quaternion elements. This *ansatz* looks very much like a sum of pure-gauge terms (however quaternions are not SU(2) matrix elements unless they are unimodular). The quaternion vector M is required to satisfy the normalization condition

$$M^+M = M_0^+M_0 + \cdots + M_n^+M_n = I_2\,.$$

If $n=0$ then $M_0 \in \mathrm{SU}(2)$, and we simply have a pure-gauge potential. For $n>0$ this will not be the case. Note that the normalization condition implies

$$(\partial_\mu M^+)M + M^+(\partial_\mu M) = 0\,,$$

just as for an SU(2) matrix element. Therefore $W_\mu^+ = W_\mu$ is Hermitian and the gauge potential W_μ^a is real. An SU(2) gauge transformation ω of the gauge potential induces the following change in M,

$$M^+ \to \omega M^+ = (\omega M_0^+, \omega M_1^+, \ldots, \omega M_n^+)\,.$$

Therefore a gauge transformation changes the elements of M by a common unimodular factor.

So far we have only an *ansatz*. The quaternion elements of $M(x)$ have to be determined such that W_μ is a self-dual solution of the equations of motion. To this end let us introduce a matrix of quaternions with a very specific x dependence,

$$\Delta(x) \equiv A + Bx\,,$$

where A and B are *constant* quaternion matrices with dimension $(n+1)\times n$ and Bx means each element of B is multiplied by x. The constant parameters in A and B are the parameters of the solution: In other words, these matrices determine the solution. They cannot be chosen arbitrarily, however, for the algebraic construction to follow is only possible if $\Delta(x)$ satisfies the conditions

$$\Delta^+(x)\Delta(x) = R(x)\,,\quad \det R(x) \neq 0\,,$$

where $R(x)$ is an $n\times n$ matrix of real numbers (i.e., the elements of R are real numbers times the two-dimensional unit matrix I_2 so that they commute with σ). $M(x)$ is now required to satisfy the condition

$$M^+(x)\Delta(x) = 0\,.$$

This condition and the preceding one insure that $eW_\mu = iM^+\partial_\mu M$ is a self-dual solution of the field equations.

The latter statement is surprisingly easy to prove. First we calculate the field strengths,

$$-ieG_{\mu\nu} = \partial_\mu M^+(1-MM^+)\partial_\nu M - \partial_\nu M^+(1-MM^+)\partial_\mu M\,,$$

noting that $(1-MM^+)$ is a projection operator which annihilates M. Another projection operator with the same property is $\Delta(\Delta^+\Delta)^{-1}\Delta^+$. Moreover, the product of these operators in either order equals $\Delta(\Delta^+\Delta)^{-1}\Delta^+$. Therefore they are the same operator,

$$1 - MM^+ = \Delta(\Delta^+\Delta)^{-1}\Delta^+\,,$$

and the field strengths can be rewritten

$$\begin{aligned} -ieG_{\mu\nu} &= \partial_\mu M^+\Delta(\Delta^+\Delta)^{-1}\Delta^+\partial_\nu M - \partial_\nu M^+\Delta(\Delta^+\Delta)^{-1}\Delta^+\partial_\mu M \\ &= M^+\partial_\mu\Delta(\Delta^+\Delta)^{-1}\partial_\nu\Delta^+M - M^+\partial_\nu\Delta(\Delta^+\Delta)^{-1}\partial_\nu\Delta^+M \\ &= M^+Be_\mu(\Delta^+\Delta)^{-1}e_\nu^+B^+M - M^+Be_\nu(\Delta^+\Delta)^{-1}e_\mu^+B^+M \\ &= M^+B(\Delta^+\Delta)^{-1}[e_\mu e_\nu^+ - e_\nu e_\mu^+]B^+M \\ &= M^+B(\Delta^+\Delta)^{-1}[2i\sigma_a\bar\eta_{a\mu\nu}]B^+M\,. \end{aligned}$$

Here we have used

$$(\partial_\mu M^+)\Delta + M^+(\partial_\mu \Delta) = 0\ ,$$
$$\partial_\mu \Delta = B e_\mu\ ,\ (\Delta^+\Delta)^{-1} e_\mu = e_\mu (\Delta^+\Delta)^{-1}\ .$$

This last equation shows why $\Delta^+\Delta = R$ must have elements which are real numbers and not quaternions. When this is so then $G_{\mu\nu}$ is proportional to $\bar{\eta}_{a\mu\nu}$, i.e., it is self-antidual. To find a self-dual $G_{\mu\nu}$ we have only to replace x by x^+ (or e_μ by e_μ^+).

The construction of an explicit solution begins with a particular choice of constant matrices A and B such that $\Delta^+\Delta = R$, $|R| \neq 0$ are satisfied. These conditions determine the number of independent parameters in the solution, which for SU(2) turns out to be $8n-3$. Thus n here is the topological charge. To find the gauge potential one has to solve $M^+\Delta = 0$ for M. Unfortunately this is generally rather difficult. An exception is the 't Hooft solution below for which this step is easy.

To illustrate the construction of Atiyah *et al.* let us give the 't Hooft N-instanton solution in the new language:

$$A = \begin{pmatrix} \lambda_1 \cdots \lambda_n \\ a_1 \\ \quad a_2 \\ \qquad \ddots \\ \qquad\quad a_n \end{pmatrix},$$

$$B = \begin{pmatrix} 0 \cdots 0 \\ -1 \\ \quad -1 \\ \qquad \ddots \\ \qquad\quad -1 \end{pmatrix}.$$

Here $a_i = a_{i\mu} e_\mu$ specifies the ith instanton position and $\lambda_i = \lambda_i I_2$ is its size. One trivially verifies that $\Delta^+\Delta = R$ is satisfied. The condition $M^+\Delta = 0$ leads immediately to

$$\lambda_i M_0^+ + M_i^+(a_i - x) = 0\ ,$$

or

$$M_i^+ = [\lambda_i/(x-a_i)^2] M_0^+(x-a_i)^+\ .$$

Then normalizing $M^+M = 1$ we find

$$M_0 = 1/\sqrt{\phi}\ ,\ M_i = [\lambda_i/(x-a_i)^2](1/\sqrt{\phi})(x-a_i)\ ,$$

where

$$\phi = 1 + \lambda_1/(x-a_1)^2 + \cdots + \lambda_n/(x-a_n)^2$$

is the familiar scalar *ansatz* function for the 't Hooft solution. A short calculation then leads to the gauge potential

$$eW_\mu = -\tfrac{1}{2}\sigma_a \eta_{a\mu\nu}(\partial_\nu\phi/\phi)\ .$$

More complicated SU(2) solutions have been studied by Christ, Weinberg, and Stanton (1978). In particular, they address the problem of finding a (recognizable) N-instanton solution where the instantons have arbitrary SU(2) orientation. In the limit of large instanton separation an approximate solution can be given explicitly.

E. Two-meron solution

Exact one- and two-meron solutions are available. The former is a special case of the latter with one of the merons at infinity. We begin with this simpler solution.

1. One meron at infinity

The one-meron solution of the YM theory was first given by de Alfaro, Fubini, and Furlan (1976). This solution is easily written down with the help of the ϕ^4 *ansatz* (7.10). The function

$$\phi = 1/\sqrt{\lambda x^2} \tag{7.56}$$

satisfies $\Box\phi + \lambda\phi^3 = 0$ (Petiau, 1958) and therefore provides a solution of the YM theory, namely

$$eW_0^a = \pm x_a/x^2\ ,$$
$$eW_i^a = -\varepsilon_{ian} x_n/x^2 \mp \delta_{ai} x_0/x^2\ . \tag{7.57}$$

In matrix form,

$$eW_\mu = \tfrac{1}{2}[-i(\partial_\mu g)g^{-1}]\ ,$$
$$g = (1/\sqrt{x^2})(\mathbf{x}\cdot\sigma \pm i x_0)\ ,\quad g^{-1} = g^+\ . \tag{7.58}$$

As we see, the one-meron solution is a pure-gauge potential multiplied by $\frac{1}{2}$. (A pure-gauge potential multiplied by any numerical factor other than 1 is, of course, no longer gauge equivalent to the vacuum.) The name "meron" comes from the distribution of topological charge

$$(e^2/8\pi^2)D = \pm(1/16\pi^2)\Box\,\partial_\mu(\partial_\mu\phi/\phi)$$
$$= \pm(1/16\pi^2)\Box(-2/x^2) = \pm\tfrac{1}{2}\delta(x)\ . \tag{7.59}$$

This charge is concentrated at a single point in E^4 and has the strength $q = \pm\frac{1}{2}$.

The solution (7.57) represents a meron positioned at the origin. By translation invariance the meron can be moved to any other location. Note that by gauge-transforming the matrix potential (7.58) with $\omega = g^{-1}$,

$$eW'_\mu = g^{-1}\tfrac{1}{2}[-i(\partial_\mu g)g^{-1}]\,g - i(\partial_\mu g^{-1})g = \tfrac{1}{2}[-i(\partial_\mu g^{-1})g]\ , \tag{7.60}$$

we obtain the potential for a meron with the opposite topological charge.

The one-meron solution (7.57) can also be expressed in terms of Witten's *ansatz* (7.23) as follows,

$$(1/r)(1+\phi_2) = \mp A_0 = r/x^2 = -\partial_0 \tan^{-1}(r/x_0)\ ,$$
$$(1/r)\phi_1 = A_1 = \pm x_0/x^2 = \pm\partial_1 \tan^{-1}(r/x_0)\ . \tag{7.61}$$

Note that A_μ is pure gauge. Choosing $f = \pm\tan^{-1}(r/x_0)$ in Eq. (7.25) (i.e., $\sin f = \pm r/\sqrt{x^2}$, $\cos f = x_0/\sqrt{x^2}$) we obtain a simpler form of the one-meron potential with *ansatz* functions $A'_\mu = 0$, $\phi'_1 = 0$ and $\phi'_2 = -x_0/\sqrt{x^2}$. The new gauge potential is

$$eW_0'^a = 0\ ,\ eW_i'^a = -\varepsilon_{ian}(x_n/r^2)(1 - x_0/\sqrt{x^2})\ . \tag{7.62}$$

2. Two-meron solution

The scalar function for the two-meron solution is

$$\phi = [(a-b)^2/\lambda(x-a)^2(x-b)^2]^{1/2}\ . \tag{7.63}$$

This is a solution of $\Box\phi + \lambda\phi^3 = 0$. (To recover the one-meron solution we can take the limit $a \to 0, b \to \infty, \lambda \to \infty$ with b^2/λ finite.) It is easy to show that the YM potential calculated from ϕ represents two merons. We observe that ϕ satisfies

$$\partial_\alpha \phi/\phi = -[(x-a)_\alpha/(x-a)^2 + (x-b)_\alpha/(x-b)^2]\,. \tag{7.64}$$

Therefore

$$\partial_\alpha(\partial_\alpha\phi/\phi) = -2[1/(x-a)^2 + 1/(x-b)^2]\,,$$

and the topological charge distribution is

$$(e^2/8\pi^2)D = \pm(1/16\pi^2)\Box\partial_\alpha(\partial_\alpha\phi/\phi) = \pm\tfrac{1}{2}[\delta(x-a)+\delta(x-b)] \tag{7.65}$$

with the merons located at the arbitrary points a and b. The action density of this solution is, from Eq. (7.13),

$$\mathcal{L} = \pm D - (3/2e^2)[(a-b)^2/(x-a)^2(x-b)^2]^2\,. \tag{7.66}$$

The explicit two-meron potential is (de Alfaro, Fubini, and Furlan, 1976)

$$eW_0^a = \pm\left[\frac{(x-a)_a}{(x-a)^2} + \frac{(x-b)_a}{(x-b)^2}\right],$$

$$eW_i^a = -\varepsilon_{ian}\left[\frac{(x-a)_n}{(x-a)^2} + \frac{(x-b)_n}{(x-b)^2}\right] \mp \delta_{ai}\left[\frac{(x-a)_0}{(x-a)^2} + \frac{(x-b)_0}{(x-b)^2}\right]. \tag{7.67}$$

This is just the sum of two one-meron potentials.

Suppose that the two merons are positioned on the time axis in E^4 at points $a=(a,\mathbf{0})$, $b=(b,\mathbf{0})$. Then the two-meron solution (7.67) takes the form of Witten's *ansatz* (7.23) with

$$A_0 = \mp\frac{1}{r}(1+\phi_2) = \mp\left[\frac{r}{(x-a)^2} + \frac{r}{(x-b)^2}\right],$$

$$A_1 = \frac{1}{r}\phi_1 = \pm\left[\frac{x_0-a}{(x-a)^2} + \frac{x_0-b}{(x-b)^2}\right]. \tag{7.68}$$

Again A_μ is a pure-gauge potential,

$$A_\mu = \partial_\mu f\,,\; f = f_a + f_b\,,$$

$$f_a = \tan^{-1}\left(\frac{r}{x_0-a}\right),\quad f_b = \tan^{-1}\left(\frac{r}{x_0-b}\right). \tag{7.69}$$

Using the formulas

$$\sin f_a = \pm\frac{r}{\sqrt{(x-a)^2}},\quad \cos f_a = \frac{x_0-a}{\sqrt{(x-a)^2}},$$

it is easy to show that

$$\sin f = \mp\frac{r(x_0-a)+r(x_0-b)}{[(x-a)^2(x-b)^2]^{1/2}}, \tag{7.70}$$

$$\cos f = \frac{r^2-(x_0-a)(x_0-b)}{[(x-a)^2(x-b)^2]^{1/2}}, \tag{7.71}$$

$$\phi_1 = -\sin f\frac{(x_0-a)(x_0-b)+r^2}{[(x-a)^2(x-b)^2]^{1/2}}, \tag{7.72}$$

$$\phi_2 = \cos f\frac{(x_0-a)(x_0-b)+r^2}{[(x-a)^2(x-b)^2]^{1/2}}. \tag{7.73}$$

Then from the gauge transformation (7.25) we find new *ansatz* functions $A'_\mu = 0$, $\phi'_1 = 0$ and

$$\phi'_2 = \frac{r^2+(x_0-a)(x_0-b)}{[(x-a)^2(x-b)^2]^{1/2}}. \tag{7.74}$$

The new gauge potential is

$$eW_0'^a = 0\,,$$

$$eW_i'^a = -\varepsilon_{ian}\left(\frac{x_n}{r^2}\right)\left\{1+\frac{r^2+(x_0-a)(x_0-b)}{[(x-a)^2(x-b)^2]^{1/2}}\right\} \tag{7.75}$$

The solution we have been discussing represents two merons or two antimerons (see Eq. (7.65)). A similar solution is known which represents one meron and one antimeron (de Alfaro, Fubini, and Furlan, 1977). To derive this solution de Alfaro *et al.* begin with the one-meron potential (7.58) and make the conformal transformation

$$x_\mu \to y_\mu = a_\mu/2a^2 - (x+a)_\mu/(x+a)^2\,.$$

Then the potential (7.58) becomes

$$eW_\mu = \tfrac{1}{2}[-i(\partial_\mu G)G^{-1}]\,,$$

$$G = \frac{1}{\sqrt{y^2}}(\mathbf{y}\cdot\sigma \pm iy_0)\,. \tag{7.76}$$

Because

$$y_\mu/\sqrt{y^2} = [a^2(x+a)^2(x-a)^2]^{-1/2}[a_\mu(x+a)^2 - 2a^2(x+a)_\mu]\,,$$

the new solution is singular at the points $x=\pm a$. To show that the new solution represents a meron and an antimeron we choose the constant vector a to be $a=(a,\mathbf{0})$. Then after a straightforward calculation we find

$$eW_0^a = -4ax_0x_a/(x+a)^2(x-a)^2\,,$$

$$eW_i^a = [4a/(x+a)^2(x-a)^2]\,[-a\varepsilon_{ian}x_n - x_ax_i + \tfrac{1}{2}\delta_{ai}(x^2-a^2)]\,. \tag{7.77}$$

Near the singular points $x_0=\pm a$, $\mathbf{x}=0$ the new solution becomes

$$eW_0^a \approx \mp x_a/(x\mp a)^2\,,$$

$$eW_i^a \approx -\varepsilon_{ian}x_n/(x\mp a)^2 \pm \delta_{ai}(x_0\mp a)/(x\mp a)^2\,,$$

which is the potential for an antimeron at $x=a$ (a meron at $x=-a$). de Alfaro, Fubini, and Furlan also show that the meron–antimeron solution above can be obtained from the meron–meron solution by a gauge transformation.

Comparing Eq. (7.77) with Witten's *ansatz* (7.23) we see that the meron–antimeron solution corresponds to *ansatz* functions

$$A_0 = \frac{4ax_0r}{(x+a)^2(x-a)^2},$$

$$A_1 = -\frac{4a}{(x+a)^2(x-a)^2}\tfrac{1}{2}(x_0^2-r^2-a^2)\,,$$

$$\phi_1 = -\frac{4ar}{(x+a)^2(x-a)^2}\tfrac{1}{2}(x^2-a^2)\,,$$

$$\phi_2 = -\frac{[(x_0+a)(x_0-a)+r^2]^2}{(x+a)^2(x-a)^2}. \tag{7.78}$$

The Abelian potential here is pure-gauge:

$$A_\mu = \partial_\mu f\,,\; f = \tan^{-1}\left(\frac{r}{x_0+a}\right) - \tan^{-1}\left(\frac{r}{x_0-a}\right). \tag{7.79}$$

Moreover,

$$\phi_1 = \sin f \cos f\ , \quad \phi_2 = -\cos^2 f \tag{7.80}$$

where

$$\sin f = r\frac{(x_0 - a) - (x_0 + a)}{[(x+a)^2(x-a)^2]^{1/2}},$$

$$\cos f = \frac{(x_0 + a)(x_0 - a) + r^2}{[(x+a)^2(x-a)^2]^{1/2}}. \tag{7.81}$$

The gauge transformation (7.25) then leads to simpler *ansatz* functions $A'_\mu = 0$, $\phi'_1 = 0$, and $\phi'_2 = -\cos f$. The new form of the meron-antimeron potential is

$$eW_0'^a = 0\ ,$$

$$eW_i'^a = -\varepsilon_{ian}\frac{x_n}{r^2}\left\{1 - \frac{x^2 - a^2}{[(x+a)^2(x-a)^2]^{1/2}}\right\}. \tag{7.82}$$

Unlike the meron-meron potential (7.75), W_i^a here vanishes in the limit $x_0 \to \pm\infty$.

F. Multimeron configurations

The N-meron problem is still unsolved. The problem is, of course, to find an exact Euclidean solution which represents N merons located at arbitrary points in E^4. (A better statement of the problem would be to find an exact solution representing arbitrary numbers of merons and antimerons arranged in an arbitrary fashion. For, as we have seen above, exact meron-antimeron solutions exist.) The fact that this solution has not been found, although serious efforts have been made, suggests that it may be quite complicated. Hopefully, someone will discover a linearization of the problem which leads to a fairly general solution. We now review the results on multimeron configurations which are presently available.

To get some feeling for the difficulties involved in the construction of a multimeron solution, let us try to find one using the ϕ^4 *ansatz* (7.10). From Eq. (7.14) for the topological charge density it follows that if ϕ satisfies the condition

$$\partial_\mu \phi = -\phi V_\mu\ , \quad V_\mu = \sum_{i=1}^{N} \frac{(x - a_i)_\mu}{(x - a_i)^2}, \tag{7.83}$$

then the corresponding YM field describes N merons at the arbitrary points $x = a_i$.

Proof: From Eq. (7.83)

$$\partial_\mu(\partial_\mu\phi/\phi) = -\partial_\mu V_\mu = -2\sum_{i=1}^{N} 1/(x - a_i)^2\ ,$$

and therefore the density of topological charge is

$$\pm e^2/8\pi^2 D = (1/16\pi^2)\Box\left\{-2\sum_i 1/(x - a_i)^2\right\}$$

$$= \tfrac{1}{2}\sum_{i=1}^{N} \delta(x - a_i)\ , \tag{7.84}$$

which corresponds to N merons or antimerons. In general, however, condition (7.83) seems to be incompatible with the equation of motion $\Box\phi + \lambda\phi^3 = 0$. If one assumes that both equations are satisfied then they can be solved for ϕ,

$$\lambda\phi^2 = -V_\mu V_\mu + 2\sum_{i=1}^{N} 1/(x - a_i)^2. \tag{7.85}$$

For $N = 1, 2$ we recover the one- and two-meron solutions. But for $N > 2$ the scalar function ϕ in Eq. (7.85) satisfies neither condition (7.83) nor the equation of motion.

A scalar function which does satisfy Eq. (7.83) and therefore leads to an N-meron configuration is

$$\phi = \phi_1\phi_2 \cdots \phi_N$$

where ϕ_i is the one-meron scalar function $\phi_i = [\lambda(x - a_i)^2]^{-1/2}$. Because

$$\partial_\mu\phi_i = -\phi_i(x - a_i)_\mu/(x - a_i)^2\ ,$$

it follows trivially that Eq. (7.83) is satisfied. Moreover, ϕ is a solution of $\Box\phi + \lambda'\phi^3 = 0$ for a certain λ' (whose value is unimportant) near the singular points $x = a_i$, and therefore the YM potential obtained from ϕ above is a solution of the equations of motion in the vicinity of each meron. (In fact, this YM potential is just a sum of one-meron potentials.) But this is *not* a good approximation to a genuine N-meron solution. Obviously it is not a good approximation when the merons are bunched together. Neither is it a good approximation at large x^2. The scalar function behaves like $(x^2)^{-N/2}$ and for $N > 2$ this is inconsistent with the equation of motion $\Box\phi + \lambda\phi^3 = 0$. The approximate gauge potential is simply N times the one-meron potential at large x^2, and of course this is not a solution—the gauge potential is too large.

Having had no luck with the ϕ^4 *ansatz*, let us turn to Witten's *ansatz* (7.23) for the gauge potential. We have already seen that the one- and two-meron configurations (with the merons on the imaginary-time axis) are quite simple when expressed in terms of this *ansatz*. Glimm and Jaffe (1978a) have shown that an arbitrary configuration of merons and antimerons arranged along the imaginary-time axis can be just as easily represented. They give an explicit formula for the Abelian gauge potential A_μ. The two Higgs field components ϕ_a are expressed in terms of an unknown scalar function that must satisfy the equation

$$r^2(\partial_0^2 + \partial_1^2)\phi = \phi(1 - \phi^2)\ , \tag{7.86}$$

with appropriate boundary conditions. Note that Eq. (7.86) is the equation of motion remaining in Eq. (7.32) when $A_\mu = 0$, $\phi_1 = 0$, and $\phi_2 = \phi$. The reason why this is the basic equation to solve is that A_μ and ϕ_1 can always be gauge-transformed to zero, for any number of merons.

The construction of Glimm and Jaffe (1978a) is very simple. They choose the following pure-gauge form for the Abelian potential A_μ,

$$A_\mu = \partial_\mu f\ , \quad f = \sum_i \tan^{-1}\left(\frac{r}{x_0 - a_i}\right) - \sum_j \tan^{-1}\left(\frac{r}{x_0 - b_j}\right). \tag{7.87}$$

The corresponding Higgs field components are

$$\phi_1 = -\sin f\phi\ , \quad \phi_2 = \cos f\phi\ , \tag{7.88}$$

where ϕ is still undetermined. Performing the gauge transformation (7.25) one obtains new *ansatz* functions $A'_\mu = 0, \phi'_1 = 0$, and $\phi'_2 = \phi$, where ϕ satisfies the equation of motion (7.86) above. This is the unsolved part of the problem. Returning to the *ansatz* functions (7.87) and (7.88) we note that $F_{\mu\nu} = 0$ except at the singular points and therefore the first equation of motion in Eq. (7.31) implies that $\varepsilon_{ab}\phi_a D_\nu \phi_b = 0$. Then from Eqs. (7.29) and (7.30) we see that the topological charge density is

$$(e^2/8\pi^2)D = (1/8\pi^2 r^2)F_{01}.$$

It is easy to verify that

$$F_{01} = 2\pi\delta(r)\left[\sum_i \delta(x_0 - a_i) - \sum_j \delta(x_0 - b_j)\right]$$

for the pure-gauge potential (7.87) above; this is just Gauss's theorem in two dimensions (x_0, r). However, only the half-plane $r \geq 0$ is physical, and for this reason only one-half of each delta function counts. Thus

$$(e^2/8\pi^2)D = \tfrac{1}{2}\delta(r)\left[\sum_i \delta(x_0 - a_i) - \sum_j \delta(x_0 - b_j)\right], \qquad (7.89)$$

where $a_i = (a_i, 0), b_j = (b_j, 0)$. We emphasize that this result follows from two assumptions: (i) A_μ is given by Eq. (7.87). (ii) The relevant equation of motion is satisfied. To complete the construction one must solve the remaining equation of motion (7.86) with suitable boundary conditions (for example, $\phi(t, r=0) = \pm 1$). So far no one has been able to do this. The one- and two-meron solutions are known, of course. A numerical investigation of the four-meron solution has been done by Jacobs and Rebbi (1978). An existence theorem for the general solution has been established (Johnson *et al.*, 1978).

G. Elliptic solutions

In Sec. VI we have shown how to construct elliptic solutions of the ϕ^4 and YM theories using a known solution of the ϕ^4 theory as input. The discussion there applies also to Euclidean space-time. Using this method we now construct elliptic generalizations of the one- and two-meron solutions. The former solution is a special case of the latter (with one meron at infinity). Nevertheless, it is worth a separate discussion because of its $O(4)$ rotational invariance.

1. One-meron solution

In the one-meron case we begin with the solution of Eq. (6.31)

$$f(x) = (-b/a\lambda x^2)^{1/2}. \qquad (7.90)$$

The function $u(x)$ needed for the elliptic generalization is

$$u(x) = (1/\sqrt{-a})\ln f(x). \qquad (7.91)$$

Both conditions (6.32) and (6.33) are satisfied by $u(x)$. Therefore we can construct seven elliptic solutions of the ϕ^4 theory: the solutions (6.34)–(6.40). Most of these lead to singular YM solutions. The two interesting ones are (6.36) and (6.38), which we give again here:

$$\phi_1 = f_1 \,\mathrm{dn}\left(\frac{\ln f_1}{\sqrt{2-k^2}}, k\right),$$

$$f_1 = \left[\frac{2}{(2-k^2)\lambda x^2}\right]^{1/2}, \qquad (7.92)$$

$$\phi_2 = f_2 \,\mathrm{nd}\left(\frac{\ln f_2}{\sqrt{2-k^2}}, k\right),$$

$$f_2 = \left[\frac{2(1-k^2)}{(2-k^2)\lambda x^2}\right]^{1/2}. \qquad (7.93)$$

The solutions have no zeros for finite x^2.

Proceeding to the YM theory, we find from Eq. (6.42) that the gauge potential in this case is

$$W^a_\mu(\phi) = [1 + E'/E\sqrt{-a}\,]W^a_\mu(f). \qquad (7.94)$$

For the solutions $\phi_{1,2}$ the elliptic factor here is

$$E'/E = -\eta k^2 \,\mathrm{sn}u\,\mathrm{cn}u/\mathrm{dn}u, \qquad (7.95)$$

where $\eta = +1(-1)$ for $\phi_1(\phi_2)$ and $u = \ln f/\sqrt{2-k^2}$. This factor is nowhere singular. Let us now discuss the two elliptic solutions $W_\mu(\phi_{1,2})$. We need the following expressions for the topological charge density (Actor, 1978b), which can be derived from Eq. (7.14),

$$\frac{e^2}{8\pi^2}D(\phi_{1,2}) = \pm\tfrac{1}{2}\delta(x) \times\left\{1 - \eta\frac{k^2 sc}{4\sqrt{2-k^2}}\left[\frac{2d}{2-k^2} + \frac{2(1-k^2)}{(2-k^2)d^3} + \frac{4}{d}\right]\right\} \pm\eta\frac{3}{8\pi^2(2-k^2)^2x^4}\left[d^4 - \frac{(1-k^2)^2}{d^4}\right]. \qquad (7.96)$$

Note that the elliptic solutions correspond to a pointlike topological charge surrounded by a cloud of charge. The strengths of the point charge and the cloud both depend on the parameter k.

The solution $W_\mu(\phi_1)$ reduces to the one-meron solution when $k \to 0$. In the limit $k \to 1$ it becomes the instanton, because the first factor in Eq. (7.94) becomes

$$1 + d'/d = 1 - \tanh u = \frac{2x^2}{x^2 + 2/\lambda}$$

in this limit and thus

$$W_\mu(\phi_1, k=1) = \frac{x^2}{x^2 + 2/\lambda}[2W_\mu(f)],$$

which is one form of the instanton potential. We see that the elliptic solution $W_\mu(\phi_1)$ has topological charge $q = \frac{1}{2}$ for $k = 0$ and $q = 1$ for $k = 1$. For arbitrary k this solution corresponds to a point charge surrounded by a charge cloud, as already mentioned. When $k \to 0$ the cloud disappears and the point charge becomes the meron. When $k \to 1$ the point charge is switched off and the cloud becomes the instanton. This behavior can be verified directly from Eq. (7.96). When $k \to 0$ the second term vanishes and the first term becomes $\pm\frac{1}{2}\delta(x)$. For $k = 1$ the coefficient of the delta function vanishes (recall that $\eta = +1$ here) and the second term becomes the instanton charge density

$$\frac{e^2}{8\pi^2}D(\phi_1, k=1) = \pm\frac{3\lambda^2}{2\pi^2}\frac{1}{(1+\lambda x^2/2)^4}.$$

Note that $W_\mu(\phi_1)$ is singular at $x^2 = 0$ for all $k < 1$. The instanton solution is unique in this respect.

The other solution $W_\mu(\phi_2)$ also reduces to the one-meron solution when $k \to 0$, while for $k=1$ it becomes

$$W_\mu(\phi_2, k=1) = \frac{2}{2+\lambda x^2}[2W_\mu(f)] .$$

This is another form of the instanton potential. Thus both $W_\mu(\phi_1)$ and $W_\mu(\phi_2)$ interpolate between the one-meron and one-instanton solutions. For $k=1$ the topological charge density (7.96) is

$$\frac{e^2}{8\pi^2}D(\phi_2, k=1) = \pm\,\delta(x) \mp \frac{3\lambda^2}{2\pi^2}\frac{1}{(1+\lambda x^2/2)^4} .$$

We have learned that the delta-function singularity with unit strength can be ignored. This leaves the charge density corresponding to an instanton, as we expect.

2. Two-meron solution

To construct the generalized two-meron solution we begin with the following solution of Eq. (6.31):

$$f(x) = (-4bv^2/a\lambda)^{1/2}h(x) ,$$
$$h(x) \equiv [(x+v)^2(x-v)^2]^{-1/2} . \qquad (7.97)$$

In this case the function $u(x)$ is

$$u(x) = \omega(x)/\sqrt{-a} , \quad \omega(x) \equiv \tfrac{1}{2}\ln[(x-v)^2/(x+v)^2] , \qquad (7.98)$$

which satisfies conditions (6.32) and (6.33). The two elliptic solutions of the ϕ^4 theory which interest us are

$$\phi_1 = [8v^2/\lambda(2-k^2)]^{1/2}h(x)\,\mathrm{dn}(\omega(x)/\sqrt{2-k^2}, k) , \qquad (7.99)$$
$$\phi_2 = [8(1-k^2)v^2/\lambda(2-k^2)]^{1/2}h(x)\,\mathrm{nd}(\omega(x)/\sqrt{2-k^2}, k) . \qquad (7.100)$$

The YM potentials are given by Eq. (6.42). As these potentials are rather complicated we do not give them explicitly. (See Actor, 1978b, for full details.) The two identities (6.46) and (6.47) are satisfied by $f(x)$ and $u(x)$ and therefore Eqs. (6.48), (6.51), (6.53), and (6.54) are valid here as well, with modifications due to the singularities at $x=\pm v$. The quantity we are especially interested in is the topological charge density. After some computation this is found to be

$$\pm\frac{e^2}{8\pi^2}D(\phi_{1,2}) = \frac{1}{2}[\delta(x+v)+\delta(x-v)] - \eta\frac{k^2}{2\sqrt{2-k^2}}\frac{sc}{d}[\delta(x+v)-\delta(x-v)] + \eta\frac{6v^2}{\pi^2(2-k^2)^2}h^4(x)\left[d^4 - \frac{(1-k^2)^2}{d^4}\right] , \qquad (7.101)$$

where $\eta=+1(-1)$ for $\phi_1(\phi_2)$, and the elliptic functions all have argument $u(x)$ and parameter k. When $k=0$ this reduces to the two-meron charge density

$$\pm(e^2/8\pi^2)D(\phi_{1,2}, k=0) = \tfrac{1}{2}[\delta(x+v)+\delta(x-v)]$$

as it should. The charge distribution for $k=1$ is

$$\pm(e^2/8\pi^2)D(\phi_{1,2}, k=1) = \tfrac{1}{2}(1-\eta)[\delta(x+v)+\delta(x-v)] + \eta 6v^2/\pi^2(x^2+v^2)^4 .$$

$W_\mu(\phi_1)$ is the solution first discovered by Cervero, Jacobs, and Nohl (1977). For arbitrary k it represents two identical point charges surrounded by a cloud of charge with the same sign. In the limit $k\to 0$ the cloud dissipates, leaving the point charges that have become merons. In the limit $k\to 1$ the point charges disappear, leaving the cloud which has become an instanton.

$W_\mu(\phi_2)$ also interpolates between the instanton and two-meron solutions. It becomes the two-meron solution when $k\to 0$. In the limit $k\to 1$ the point charges increase to unit strength (so that they can be ignored) and the cloud becomes an instanton.

VIII. CONCLUSION

Three particularly interesting types of YM solution have been discussed in this review. We now briefly summarize the present theoretical situation with regard to each of these.

A. Monopoles and dyons

Developments have not matched the enthusiasm generated by the discovery of the 't Hooft–Polyakov monopole. It is known that solutions of this type generally exist for YM theories with arbitrary semisimple gauge group. All of these monopoles are topological solitons in $3+1$ dimensions, and theoretically they are extremely interesting. One can think of two obvious directions for future research.

(1) Phenomenological: Monopoles may be realized in nature as physical objects, as Dirac speculated many years ago. The YM monopoles would be attractive candidates because they are extended, nonsingular, string-free objects which seem to be stable. There would be the question of which gauge group is the correct one, of course just as for unified theories of the weak and electromagnetic interactions. Perhaps these would be related questions.

(2) Theoretical: The mathematical properties of YM monopoles could be studied independently of physical applications. One could try to set up a quantized theory of monopoles. The classical monopole–monopole interaction could be investigated, etc.

At the present time both of these directions are unfortunately blocked. SU(2) monopoles probably have masses of several thousand GeV, as we have seen in Sec. IV. While such a large mass may be commensurate with the large magnetic charge $g=1/e$, it makes monopoles impossible to produce experimentally. If there are none already on the earth, then no monopole will ever be found. Perhaps a huge monopole mass is the true reason why none have ever been seen. If this is the case, we shall never be able to verify it.

Theoretical work on monopoles is hindered by the absence of an exact multimonopole solution. All efforts to find such a solution with physically acceptable properties have failed. It is not known whether multimonopole solutions with finite energy exist. Without these solutions one can only superficially investigate the monopole–monopole interaction. Even solutions describing two or more pointlike Wu–Yang monopoles would be helpful; but no one has found these solutions either. The theoretical situation has rather stagnated, awaiting a clever *ansatz* or idea which may break things open.

B. Instantons

The most spectacular impact on YM theory has certainly been the discovery of the instanton solution. Exact multi-instanton solutions were found soon afterwards, and it is now clear that one must accept the presence of complicated ensembles of these objects in the YM vacuum. Instantons (which are mathematically unobjectionable solutions of the YM equations of motion carrying a unit of nontrivial topological charge, although they have zero energy) imply that the true YM vacuum consists of an infinity of topologically distinct sectors connected by instanton tunneling. The YM vacuum is therefore very complicated. Recognition of this fact has forced theoreticans to reexamine the existing YM formalism. A variety of possible physical instanton effects have been suggested.

The physical implications of instantons may be very important—this is still an open field for research. Weak instanton effects, i.e., those in unified models of the weak and electromagnetic interactions, seem to be negligible because of the small coupling. QCD instanton effects could be large, however. Massless quarks suppress instanton tunneling (for reasons which are not easy to understand). Conversely, if instantons are present then quarks have gotten masses through some dynamical mechanism. Instantons may therefore be related to chiral symmetry breakdown in QCD. Another fundamental problem is quark confinement. A dense ensemble of instantons could have such a disordering influence on the YM system that only short-range effects survive. In other words, when the density of instantons becomes large enough, there could be a phase transition from massless to massive gauge fields which signals the onset of confinement. This phase transition (if it exists) has not yet been formulated in a mathematically precise fashion. Many other possible effects of instantons have been proposed. These generally involve quarks in one way or another, e.g., quark–quark forces which affect hadron masses. None of these instanton effects has been firmly established.

C. Merons

To date, the physical meaning of meron solutions remains unclear. These solutions exist (the two-meron solution is known explicitly, and the multi-meron solution has been shown to exist) and one has to decide what to do with them. It could be argued that they are unphysical because they are singular; hence, the action is infinite, and in the path-integral formulation of quantum YM field theory such solutions ought to play a negligible role. But, as advocates of meron effects have stressed, this is not necessarily true, because the functional integration measure could possibly overcome the weight factor exp(−action). Under certain circumstances merons might dominate the path integral.

Elementary arguments in support of the physical relevance of merons have been presented in Appendix H. These are based on the Gribov ambiguity, i.e., the existence of additional vacua in the Coulomb gauge with topological charge $Q_T=\pm\frac{1}{2}$ besides the usual $Q_T=0$ vacuum. Instantons tunnel between the $Q_T=\frac{1}{2}$ and $Q_T=-\frac{1}{2}$ vacua, but this does not entirely remove the degeneracy of the vacuum. Merons are needed for this: they tunnel between the $Q_T=\pm\frac{1}{2}$ and $Q_T=0$ vacua, thereby completely restoring the vacuum symmetry. Thus, gluons are not necessarily Goldstone bosons when merons are present, and there may be a phase transition to a confining phase of the YM theory in which gluons are massive. Unfortunately, this simple picture is obscured by the singularity of the meron solutions. Merons do not have the impeccable mathematical properties of instantons, which make the latter so compelling. Before one can understand the physical role of merons, one has to understand why they are singular. So far, no one has satisfactorily explained this.

Further research on merons is hampered by the lack of an exact solution describing an ensemble of merons and antimerons. This is the outstanding problem which has to be solved before one can calculate multimeron effects. If no multimeron solution is found, then progress on merons will be difficult. It is clear that this problem is more difficult than the N-instanton problem. Meron solutions are not self-dual, and this makes them more truly *non-Abelian* than instanton solutions are. The non-pure gauge behavior of the one-meron potential at large $|x|$ makes it difficult to superimpose several merons and still satisfy the equations of motion.

There is one very interesting aspect of merons that should be kept in mind, namely their connection with YM magnetic monopoles. We have seen that a meron, when viewed at different times along the x_0-axis in E^4, starts off as a pure-gauge potential in three-space at $x_0=-\infty$, then becomes nontrivial for finite x_0, and at $x_0=0$ it is exactly a *Wu–Yang monopole*. For $x_0\to+\infty$ it decays into the vacuum again. This is probably a clue to the physical meaning of merons: These are tunneling solutions between vacuum configurations, whose least trivial intermediate configuration is the pointlike SU(2) monopole solution. If one forgets about tunneling, then merons can be thought of as short-lived YM monopoles. At this point one is reminded of the fact that the multimonopole problem, like the multimeron problem, is still unsolved. Quite possibly there is an important connection between the two.

ACKNOWLEDGMENT

I am indebted to D. Schröter for careful and patient typing of the manuscript.

APPENDIX A: SU(2) YANG-MILLS THEORY

Yang–Mills fields can be introduced in the following fashion (Yang and Mills, 1954; Utiyama, 1956). Consider a multiplet $\psi(x)$ which transforms locally under the action of some gauge group G according to the rule

$$\psi(x)\to\psi'(x)=\omega(x)\psi(x)\,, \tag{A1}$$

where $\omega(x)$ belongs to the relevant representation of G. Let us try to define a derivative $D_\mu\psi$ of ψ which has this same simple transformation property. We make the *ansatz*

$$D_\mu\psi = (\partial_\mu - ieW_\mu(x))\,\psi(x)\,, \tag{A2}$$

where $W_\mu(x)$ is a matrix function. By *assumption*, $D_\mu\psi$ transforms like

$$D_\mu\psi \to D'_\mu\psi' = \omega D_\mu\psi\,, \tag{A3}$$

where $D'_\mu = \partial_\mu - ieW'_\mu(x)$. This leads to

$$(\partial_\mu - ieW'_\mu)\omega = \omega(\partial_\mu - ieW_\mu)\,, \tag{A4}$$

which can be rewritten

$$W'_\mu = \omega W_\mu \omega^{-1} - (i/e)(\partial_\mu \omega)\omega^{-1}\,. \tag{A5}$$

$W_\mu(x)$ is the Yang-Mills potential in matrix form, and Eq. (A5) is the local transformation rule for this potential under the gauge group G. If the potential $W_\mu(x)$ is not introduced, then one cannot define the covariant derivative $D_\mu\psi$ with its simple tranformation property (A3). The existence of this covariant derivative is important, for it enables one to construct Lagrangian kinetic terms which are invariant under the gauge group G (e.g., $D^\mu\psi D_\mu\psi$ and $\bar\psi\gamma^\mu D_\mu\psi$ for scalar and spinor fields, respectively). One can then proceed to construct theories which are locally gauge invariant.

$W_\mu(x)$ is analogous to the four-potential in electromagnetism. The Yang-Mills forces, or field strengths, are *defined* as follows:

$$G_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu + (e/i)[W_\mu, W_\nu]\,. \tag{A6}$$

No experimental results lead one to this definition; but $G_{\mu\nu}$ has the simple transformation property

$$G_{\mu\nu} \to G'_{\mu\nu} = \omega G_{\mu\nu}\omega^{-1} \tag{A7}$$

under the gauge group G, as follows from Eq. (A5). $G_{\mu\nu}$ is therefore the "natural" generalization of the field strength tensor $F_{\mu\nu}$ in electromagnetism. Note that $G_{\mu\nu}$ is not invariant under gauge transformations, whereas $F_{\mu\nu}$ is gauge invariant. This is an important difference between Abelian and non-Abelian gauge theories.

Let us now particularize the discussion to the gauge group $G = SU(2)$. In the 2×2 representation

$$W_\mu = \tfrac{1}{2}\sigma_a W^a_{\ \mu}\,,\quad G_{\mu\nu} = \tfrac{1}{2}\sigma_a G^a_{\mu\nu}\,, \tag{A8}$$

where

$$G^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + e\varepsilon_{abc}W^b_{\ \mu}W^c_{\ \nu}\,. \tag{A9}$$

The local SU(2) gauge transformation $\omega(x)$ can be written

$$\omega = f_0 + i\sigma_b f_b\,,\quad \omega^{-1} = f_0 - i\sigma_b f_b\,,\quad f_0 f_0 + f_b f_b \equiv 1\,. \tag{A10}$$

For a real SU(2) transformation, $\omega^+ = \omega^{-1}$, f_0 and f_b are real. For a complex SU(2) transformation these functions are complex.

A Lagrangian which is invariant under any local real or complex SU(2) transformation is easily constructed,

$$\mathcal{L} = -\tfrac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu}\,. \tag{A11}$$

The equations of motion obtained from $\mathcal{L}$ are

$$\partial^\nu G^a_{\mu\nu} = e\varepsilon_{abc}G^b_{\mu\nu}W_c^{\ \nu}\,. \tag{A12}$$

Here we see two important new features of non-Abelian gauge theories which are absent in electromagnetism: (i) The equations of motion are nonlinear in the gauge potential. (ii) The gauge potential appears explicitly in the equations of motion. It seems that the YM potential plays a more basic role than the potential in an Abelian gauge theory. At least, this is true in the conventional formulation of the theory which directly involves the potential W^a_μ. In electromagnetism one can work exclusively with the field strengths **E** and **B**. Things are not so simple when the gauge group is non-Abelian. (For a formulation of YM theories directly in terms of field strengths see Halpern, 1977a, b.) In Abelian gauge theories the field strengths locally determine the gauge potential up to an arbitrary gauge transformation. The same is not true for non-Abelian gauge theories (Wu and Yang, 1975): two YM potentials that are gauge inequivalent can provide the same YM field strengths. This leads to an interesting problem, namely, the determination of all possible gauge potentials that yield a given field strength tensor. (For recent literature on this problem see Eguchi, 1976; Deser and Wilczek, 1976; Calvo, 1977a; Roskies, 1977.)

Local SU(2) gauge transformations are usually written in the 2×2 matrix form

$$\omega(x) = \exp[i\tfrac{1}{2}\sigma_a\theta_a(x)] = \cos\tfrac{1}{2}\theta(x) + i\hat n_a(x)\sigma_a \sin\tfrac{1}{2}\theta(x)\,, \tag{A13}$$

where $\hat n_a(x)$ is a unit vector defined by

$$\theta_a(x) \equiv \hat n_a(x)\theta(x)\,. \tag{A14}$$

It is useful to have gauge transformation formulas for the components of the gauge potential. The pure-gauge term is easily found to be

$$eW^a_\mu(\text{pure gauge}) \equiv -i\,\mathrm{Tr}\sigma_a(\partial_\mu\omega)\omega^{-1} = \tfrac{1}{2}\hat n_a\partial_\mu\theta + \tfrac{1}{2}\sin\theta(\partial_\mu\hat n_a) + \sin^2(\theta/2)\varepsilon_{abc}(\partial_\mu\hat n_b)\hat n_c\,. \tag{A15}$$

When $\hat n$=const, the pure-gauge term is simply

$$eW^a_\mu(\text{pure gauge}) = \hat n_a\tfrac{1}{2}\partial_\mu\theta\,. \tag{A16}$$

To calculate the term $\omega W_\mu\omega^{-1}$ in Eq. (A5) we need the formula

$$\omega\sigma_a\omega^{-1} = \sigma_a\cos\theta + \sin\theta\varepsilon_{abc}\hat n_b\sigma_c + 2\hat n_a(\hat n\cdot\sigma)\sin^2(\theta/2)\,. \tag{A17}$$

It follows that

$$\omega\sigma_a W^a_\mu\omega^{-1} = \cos\theta\sigma_a W^a_{\ \mu} + \sin\theta\varepsilon_{abc}W^a_\mu\hat n_b\sigma_c + 2\sin^2(\theta/2)\hat n_a W^a_\mu(\hat n\cdot\sigma)\,. \tag{A18}$$

Finally, Eq. (A5) can be written

$$W'^a_\mu = \cos\theta W^a_{\ \mu} + \sin\theta\varepsilon_{abc}W^b_\mu\hat n_c + 2\sin^2(\theta/2)\hat n_a(\hat n_b W^b_\mu) + (1/e)[\tfrac{1}{2}\hat n_a\partial_\mu\theta + \tfrac{1}{2}\sin\theta\partial_\mu\hat n_a + \sin^2(\theta/2)\varepsilon_{abc}(\partial_\mu\hat n_b)\hat n_c]\,. \tag{A19}$$

At any given space-time point x, the component of W^a_μ parallel to $\hat n_a(x)$ transforms like an Abelian gauge potential,

$$\hat n_a W'^a_\mu = \hat n_a W^a_{\ \mu} + (1/2e)\partial_\mu\theta\,. \tag{A20}$$

Projecting out the components of W^a_μ perpendicular to

$\hat{n}_a(x)$,

$$\hat{e}_1^a W^a_\mu \equiv [\varepsilon_{abc}\hat{e}_2^b\hat{n}_c] W^a_\mu .$$

$$\hat{e}_2^a W^a_\mu \equiv [\varepsilon_{abc}\hat{n}_b\hat{e}_1^c] W^a_\mu , \tag{A21}$$

we see that they transform like

$$\begin{pmatrix} \hat{e}_1^a\, W'^a_\mu \\ \hat{e}_2^a\, W'^a_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \hat{e}_1^a\, W^a_\mu \\ \hat{e}_2^a\, W^a_\mu \end{pmatrix} + \frac{1}{e}\sin\frac{\theta}{2} \begin{pmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} \hat{e}_1^a\, \partial_\mu \hat{n}_a \\ \hat{e}_2^a\, \partial_\mu \hat{n}_a \end{pmatrix}. \tag{A22}$$

Here $\hat{e}_1^a$, $\hat{e}_2^a$, and $\hat{n}_a = \varepsilon_{abc}\,\hat{e}_1^b\hat{e}_2^c$ are orthonormal basis vectors.

APPENDIX B: IMBEDDING SU(2) SOLUTIONS

Solutions of SU(2) gauge theories can be imbedded in any larger gauge theory. This is because all non-Abelian groups have at least one SU(2) subgroup, and this is the necessary requirement. The imbedding is rather trivial, as we discuss in this appendix. Our example—the YM theory with an arbitrary gauge group and a Higgs field in the adjoint representation—should adequately demonstrate the generality and triviality of the imbedding procedure. Some of the exact solutions of SU(3) and SU(4) gauge theories discussed in the recent literature are imbedded SU(2) solutions, while others are not. It is useful to be able to distinguish one type from the other.

Consider the gauge theory

$$\mathcal{L} = -\tfrac{1}{4}G^A_{\mu\nu}G^{\mu\nu}_A + \tfrac{1}{2}D^\mu\phi_A D_\mu\phi_A + U(\phi) , \tag{B1}$$

$$G^A_{\mu\nu} = \partial_\mu W^A_\nu - \partial_\nu W^A_\mu + e f_{ABC} W^B_\mu W^C_\nu , \tag{B2}$$

$$D_\mu\phi_A = \partial_\mu\phi_A + e f_{ABC} W^B_\mu \phi_C , \tag{B3}$$

where $A, B, C, \ldots = 1, \ldots, N$ and the gauge potential W^A_μ and Higgs field ϕ_A both transform according to the N-dimensional adjoint representation of the semisimple, compact gauge group. The structure constants f_{ABC} are real and completely antisymmetric. They determine the commutation relations of the Hermitian generators T_A of the gauge group,

$$[T_A, T_B] = i f_{ABC} T_C . \tag{B4}$$

The equations of motion obtained from the Lagrangian (B1) are

$$\partial^\nu G^A_{\mu\nu} = e f_{ABC}[G^B_{\mu\nu} W^\nu_C + (D_\mu\phi_B)\phi_C] , \tag{B5}$$

$$\partial^\mu D_\mu\phi_A = f_{ABC}(D_\mu\phi_B) W^\mu_C - (\phi_A/\phi) U'(\phi) . \tag{B6}$$

Suppose that the gauge group has an SU(2) subgroup, whose generators we name T_1, T_2, and T_3. Then the structure functions satisfy

$$f_{abc} = \varepsilon_{abc} , \tag{B7}$$

and

$$f_{Abc} = 0 , \quad A > 3 , \tag{B8}$$

where a, b, c take only the values 1, 2, or 3.

Now let us make the *ansatz*

$$W^A_\mu = W^a_\mu \quad A = a$$
$$W^A_\mu = 0 \quad A > 3$$
$$\phi_A = \phi_a \quad A = a$$
$$\phi_A = 0 \quad A > 3 , \tag{B9}$$

for the gauge and Higgs fields. Then it is easy to verify that

$$G^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + e\varepsilon_{abc} W^b_\mu W^c_\nu$$
$$G^A_{\mu\nu} = 0 , \quad A > 3 , \tag{B10}$$

$$D_\mu\phi_a = \partial_\mu\phi_a + e\varepsilon_{abc} W^b_\mu \phi_c$$
$$D_\mu\phi_A = 0 , \quad A > 3 , \tag{B11}$$

and furthermore that the equations of motion (B5) and (B6) become

$$\partial^\nu G^a_{\mu\nu} = e\varepsilon_{abc}[G^b_{\mu\nu} W^\nu_c + (D_\mu\phi_b)\phi_c] , \tag{B12}$$

$$\partial^\mu D_\mu\phi_a = e\varepsilon_{abc}(D_\mu\phi_b) W^\mu_c - (\phi_a/\phi) U'(\phi) , \tag{B13}$$

where $\phi^2 = \phi_a\phi_a$. These are precisely the equations of motion for the SU(2) theory. If these equations are satisfied then so are the equations of motion (B5) and (B6) of the larger gauge theory. This statement is obviously still true when the Higgs fields are identically zero and we are dealing with pure gauge theories.

The imbedding procedure just considered can easily be extended to subgroups H larger than SU(2). Any explicit solution of a gauge theory based on the group H can be imbedded in a similar larger gauge theory whose gauge group contains H as a subgroup.

The basic properties of the SU(2) solution are not affected by the imbedding. For example, a self-dual SU(2) solution or a local gauge symmetry breaking solution has this same property in the larger theory. Moreover, the correspondence (2.12) between static solutions of the SU(2) theory with and without a Higgs field is still valid after the imbedding. Thus we can change any imbedded static solution of the theory (B1) with $U(\phi) = 0$ into an imbedded static solution of the corresponding theory with no Higgs field. [Of course, one can also do this for more general static solutions of the theory (B1) than the simple imbedded type, because rule (2.12) can be generalized to an arbitrary gauge group.]

APPENDIX C: CONFORMAL TRANSFORMATIONS AND YANG-MILLS FIELDS

(1) Consider a general coordinate transformation $x_\mu \to y_\mu = y_\mu(x)$ where $y_\mu(x)$ is any function of x. Define a corresponding transformed YM field $\overline{W}^a_\mu$ by

$$W^a_\mu(x) \to \overline{W}^a_\mu(x) \equiv (\partial_\mu y^\alpha) W^a_\alpha(y) . \tag{C1}$$

Given this transformation rule it is easy to verify that the YM field strengths transform contravariantly,

$$G^a_{\mu\nu}(x) \to \overline{G}^a_{\mu\nu}(x) = (\partial_\mu y^\alpha)(\partial_\nu y^\beta) G^a_{\alpha\beta}(y) . \tag{C2}$$

This result follows trivially from the antisymmetry of $G_{\mu\nu}$. However, $\overline{W}^a_\mu(x)$ is not a solution of the YM equations of motion in general, even if $W^a_\mu(x)$ is a solution of these equations: the SU(2) gauge theory is not invariant under general coordinate transformations. [A generalized $O(4)$ gauge theory which is invariant under the general nonlinear group has been constructed by de Alfaro, Fubini, and Furlan, 1978.]

(2) The SU(2) gauge theory is Poincaré invariant. This means that if $x_\mu \to y_\mu$ is a Poincaré transformation, then $\overline{W}^a_\mu$ in Eq. (C1) is a solution if W^a_μ is a solution. However, the SU(2) gauge theory has a larger invariance group, namely the Minkowski conformal group, which is the group of transformations that leaves invariant the form $dx^\mu dx_\mu = 0$. This is a 15-parameter group, with the ten-parameter Poincaré group as a subgroup. Besides Poincaré transformations there are other kinds of transformations which leave $d^2x = 0$ invariant:

(i) Global scale changes, or dilatations,

$$x_\mu \to y_\mu = \lambda x_\mu \,, \tag{C3}$$

where λ is a dimensionless constant.

(ii) Inversions

$$x_\mu \to y_\mu = x_\mu / x^2 \,, \tag{C4}$$

for which

$$(\partial_\mu y_\alpha)(\partial^\mu y_\beta) = g_{\alpha\beta}/x^4 \,, \tag{C5}$$

$$d^4y = d^4x/x^8 \,. \tag{C6}$$

(iii) Special conformal transformations

$$x_\mu \to y_\mu = (1/\sigma(x))(x_\mu + c_\mu x^2) \,, \quad \sigma(x) = 1 + 2cx + c^2x^2 \,, \tag{C7}$$

(c_μ is a constant four-vector with dimension L^{-1}) for which

$$(\partial_\mu y_\alpha)(\partial^\mu y_\beta) = g_{\alpha\beta}/\sigma^2(x) \,, \tag{C8}$$

$$d^4y = d^4x/\sigma^4(x) \,. \tag{C9}$$

Special conformal transformations are equivalent to an inversion, followed by a translation, followed by another inversion. Thus inversions and special conformal transformations are not independent; and for the purpose of investigating group structure the latter are more convenient. The generators of the Minkowski conformal group consist of the ten Poincaré generations, the dilatation generator, and the four generators of special conformal transformations. This 15-parameter group is locally isomorphic to the group SO(4, 2). (See Mack and Salam, 1969 for an extensive list of references. One very readable paper is Wess, 1960.)

(3) Let us prove that the SU(2) gauge theory is conformal invariant. We already known that it is Poincaré invariant, and for the scale transformation (C3) the proof is trivial, so we restrict the proof to the special conformal transformation (C4). From Eqs. (C2) and (C5) we find the transformed SU(2) gauge theory Lagrangian

$$\overline{\mathcal{L}}(x) = -\tfrac{1}{4}\overline{G}^a_{\mu\nu}(x)\overline{G}^{\mu\nu}_a(x) = [1/\sigma^4(x)]\,\mathcal{L}(y) \,. \tag{C10}$$

Then using Eq. (C9) we find

$$d^4x\,\overline{\mathcal{L}}(x) = d^4y\,\mathcal{L}(y) \,. \tag{C11}$$

Therefore $\overline{W}^a_\mu(x)$ defined by Eq. (C1) is a solution of the equations of motion if $W^a_\mu(x)$ is a solution. Combined with Poincaré invariance and scale invariance, this proves the conformal invariance of the massless YM theory. It is obvious that the same proof holds for inversions. Thus the gauge theory is invariant under the inversion (C4).

(4) The massless ϕ^4 theory is also conformal invariant. (This is relevant here because of the importance of the ϕ^4 *ansatz* for the gauge potentials.) The transformed scalar field is

$$\overline{\phi}(x) \equiv [1/\sigma(x)]\,\phi(y) \,. \tag{C12}$$

Its gradient is

$$\partial_\mu\overline{\phi}(x) = (1/\sigma)\partial_\mu y^\alpha \partial'_\alpha\phi(y) - (1/\sigma^2)\partial_\mu\sigma\phi(y)$$

where $\partial'_\alpha \equiv \partial/\partial y^\alpha$. Using $\Box\sigma = 8c^2$, $\partial_\mu\sigma\partial^\mu\sigma = 4c^2\sigma$ one easily verifies that

$$\partial_\mu\overline{\phi}(x)\partial^\mu\overline{\phi}(x) = (1/\sigma^4)\partial'_\alpha\phi(y) \times \partial'^\alpha\phi(y) - \partial_\mu[(\partial^\mu\sigma/\sigma^3)\phi^2(y)] \,.$$

The transformed ϕ^4 theory Lagrangian is

$$\overline{\mathcal{L}}(x) = \tfrac{1}{2}\partial_\mu\overline{\phi}\partial^\mu\overline{\phi} - (\lambda/4)\overline{\phi}^4 = (1/\sigma^4)\mathcal{L}(y) \,, \tag{C13}$$

where we have dropped the divergence term. Conformal invariance then follows from

$$d^4x\,\overline{\mathcal{L}}(x) = d^4y\,\mathcal{L}(y) \,. \tag{C14}$$

APPENDIX D: THE DIRAC STRING

Many years ago Dirac (1931) showed how to incorporate magnetic monopoles into electromagnetic theory at the first-quantized level. He based his extended theory on a single four-vector potential A_μ. This would seem to minimize the departure from the conventional theory. However, a substantial departure turned out to be unavoidable. Dirac had to introduce the concept of a "string," i.e., a continuous line along which A_μ is singular. Any four-vector potential which provides the magnetic field associated with a magnetic monopole necessarily has such a singularity line. A simple example of such a potential is

$$A_0 = 0, \;\; \mathbf{A} = -\hat{\phi} g \sin\theta / r(1-\cos\theta) \,, \tag{D1}$$

which describes a monopole at rest at the origin because $\nabla \times \mathbf{A} = g\,\mathbf{r}/r^3$. The potential (D1) is singular along the positive z axis. Now this singularity line, or string, can have no physical meaning. One can tolerate its presence in the mathematics, and even derive an important result—the Dirac quantization condition—from the fact that it is unphysical. Nevertheless the string is a feature in the theory which many physicists have found disturbing. In the four and one-half decades since the publication of Dirac's paper many people have studied the problem of extending classical and quantum electrodynamics to include magnetic monopoles [Stevens (1970) has given an annotated bibliography for the period 1931–1970]. Elimination of the Dirac string has been one of the main goals. It is clear that new potentials have to be introduced to accomplish this.

The need for the Dirac string is immediately clear when one examines Maxwell's equations in the presence of electric and magnetic charge currents J_μ and g_μ, respectively,

$$\partial^\mu F_{\mu\nu} = J_\nu , \quad \text{(D2)}$$

$$\partial^\mu \tilde{F}_{\mu\nu} = -g_\nu , \quad \text{(D3)}$$

where

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{bmatrix} , \quad \text{(D4)}$$

and

$$\tilde{F}^{\mu\nu} = \tfrac{1}{2}\varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} = \begin{bmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z & E_y \\ -B_y & E_z & 0 & -E_x \\ -B_z & -E_y & E_x & 0 \end{bmatrix} . \quad \text{(D5)}$$

[The equations (D2) and (D3) are symmetric under the interchange of electric and magnetic quantities $\mathbf{E} \leftrightarrow \mathbf{B}$ $J_\mu \leftrightarrow g_\mu$, whereas the usual Maxwell's equations do not have this symmetry. The former therefore seem to enjoy some aesthetic advantage, it has sometimes been argued, despite the fact that the real world does not exhibit this symmetry in an obvious way.] Now let us introduce a four-vector potential,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu . \quad \text{(D6)}$$

We find at once from Eq. (D3) that

$$g_\nu = \varepsilon_{\nu\mu\alpha\beta}\partial^\mu \partial^\alpha A^\beta . \quad \text{(D7)}$$

In other words, $g_\nu \neq 0$ implies that

$$\partial_\mu \partial_\alpha A_\beta \neq \partial_\alpha \partial_\mu A_\beta ,$$

and therefore A_β must be singular. On any simply connected surface surrounding the monopole A_β need only be singular at one point on the surface. If one imagines an outward succession of such surfaces, then one is led to visualize a continuous line of points extending from the monopole to infinity, along which the four-potential is singular [recall Eq. (D1)].

The Dirac string can be visualized as an infinitely long, thin solenoid. Magnetic flux lines emanate in all directions from the monopole and return from infinity through the string. Clearly the position of the string and any motion it may experience are unphysical and undetectable. Therefore an electron or other particle should not exhibit unusual behavior in the vicinity of a string. In particular, the phase of the particles' wave function ψ should change by at most some integral multiple of 2π when a small closed loop is described about the string. Let us assume that the monopole is far away and that no other forces act on the particle. Then ψ satisfies the wave equation for a free particle. Write ψ in the form $\psi = \phi e^{i\beta}$ where ϕ is a function with a *definite phase* at every point. Then ϕ satisfies the wave equation for a particle in an electromagnetic potential $e\mathbf{A} = \nabla\beta$. This is entirely trivial if β is an integrable function, for there is no electromagnetic field. However, β will be nonintegrable near a string: for a small loop enclosing the string the change in β is

$$\oint d\beta = e\int d\mathbf{s}\cdot\nabla\times\mathbf{A} = e[\text{flux}] = 4\pi e g = 2\pi n .$$

The flux in this equation is the total magnetic flux within the string, which equals the total magnetic flux $\Phi = 4\pi g$ of the monopole. Dirac's famous quantization condition then follows immediately,

$$eg = n/2 . \quad \text{(D8)}$$

APPENDIX E: ELECTROMAGNETIC FIELD TENSOR IN YANG-MILLS THEORY

There is generally a long-range component in static solutions of YM theories which have an unbroken local U(1) gauge group. It seems natural (or attractive, in any case) to interpret this long-range component as an electromagnetic field. To do this one has to define the electromagnetic (EM) field tensor $F_{\mu\nu}$ in a gauge-invariant fashion (invariant under the full non-Abelian gauge group), using the ingredients available in the gauge theory in question. Here we discuss briefly how this can be done. The definition of $F_{\mu\nu}$, we emphasize, is not unique.

For the theory (2.1) with the Higgs triplet, 't Hooft (1974) has proposed the gauge-invariant definition

$$F_{\mu\nu} \equiv (1/\phi)\phi_a G^a_{\mu\nu} - (1/e)\varepsilon_{abc}(1/\phi^3)\phi_a(D_\mu\phi_b)(D_\nu\phi_c) . \quad \text{(E1)}$$

Arafune, Freund, and Goebel (1975) pointed out that this tensor can also be written in the more transparent form

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - (1/e)\varepsilon_{abc}\hat{\phi}_a\partial_\mu\hat{\phi}_b\partial_\nu\hat{\phi}_c , \quad \text{(E2)}$$

where

$$\hat{\phi}_a \equiv \phi_a/\phi, \;\; A_\mu \equiv \hat{\phi}_a W^a_\mu . \quad \text{(E3)}$$

Here A_μ is the massless component of the gauge potential. 't Hooft's definition (E1) has many good features. But it is singular at any point where $\phi = 0$.

We recall that the *ansatz* (3.12) leading to the 't Hooft–Polyakov solution is such that $A_\mu = 0$, $\hat{\phi}_a = \hat{r}_a$ for all $r > 0$. From Eq. (E2) we find immediately that

$$E_j = F_{0j} = 0 ,$$
$$B_i = (1/2e)\varepsilon_{ijk}\varepsilon_{abc}\hat{r}_a\partial_j\hat{r}_b\partial_k\hat{r}_c = r_i/er^3 , \quad \text{(E4)}$$

which is the static electromagnetic field of a monopole with magnetic charge $g = 1/e$. The "no-string gauge" form (3.12) of the monopole solution has the peculiar feature that the massless component A_μ of the gauge potential is identically zero, and therefore the EM component of the solution is not in the gauge potential W^a_μ. Instead it is in the Higgs field, as we see from Eq. (E4). This is a consequence of our gauge choice, and by means of the gauge transformation in Eq. (3.14) we can transfer the EM content of the solution back to the gauge field where one expects to find it. This results in the "string gauge" version of the solution given by Eq. (3.15). In this gauge $A_\mu = A^3_\mu$ is the usual Dirac string potential, and the Higgs field contributes nothing to $F_{\mu\nu}$.

The fact that the monopole field (E4) corresponds to a point magnetic charge, while the 't Hooft–Polyakov solution corresponds to an extended nonsingular object, might be regarded as a drawback of the definition (E2). This definition forces all of the magnetic charge to concentrate at the point $r=0$, where the Higgs field vanishes. A less singular definition may be preferable. One possibility is

$$F'_{\mu\nu}\equiv\hat{\phi}_a G^a_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu +[e\varepsilon_{abc}\hat{\phi}_a W^b_\mu W^c_\nu+W^a_\mu\partial_\nu\hat{\phi}_a-W^a_\nu\partial_\mu\hat{\phi}_a], \tag{E5}$$

with A_μ defined by Eq. (A3). In the unitary (or string) gauge where $\hat{\phi}_a=\delta_{a3}$ we have $A_\mu=W^3_\mu$ and

$$F'_{\mu\nu}=\partial_\mu W^3_\nu-\partial_\nu W^3_\mu+e(W^1_\mu W^2_\nu-W^1_\nu W^2_\mu). \tag{E6}$$

The nonlinear term here, which is absent in the EM tensor (E2), contains only short-range components. Therefore $F_{\mu\nu}$ and $F'_{\mu\nu}$ differ only in their short-range behavior.

Let us calculate $F'_{\mu\nu}$ for the monopole solution. First we work out the SU(2) field strengths. These are $E^{\prime a}_n=0$ and

$$eB^{\prime a}_n=-\delta_{an}(h'/r)+(r_a r_n/r^4)[1+rh'-h^2]. \tag{E7}$$

For the Prasad–Sommerfield solution (4.19) we find

$$B'_n=\hat{r}_a B^{\prime a}_n=r_n(1-h^2)/er^3 =(r_n/er^3)[1-(\beta r/\sinh\beta r)^2]. \tag{E8}$$

Therefore the definition (E5) above leads to a magnetic field B$'$ which corresponds to an extended distribution of magnetic charge concentrated about the origin rather than a pointlike distribution.

Next, let us consider the pure YM theory with no Higgs field. Both the definitions (E2) and (E5) above make use of the unit isovector $\hat{\phi}_a$ to define the isoscalar EM tensor $F_{\mu\nu}$. But now we have no Higgs field, and it may look as though one cannot construct a suitable EM field tensor. For unbroken local SU(2) gauge symmetry this is indeed the case. All components of the gauge field are long range, and no one of them can be distinguished from the others as being the EM field. However, when the local gauge symmetry is broken to U(1) then a unit isovector is always available for constructing $F_{\mu\nu}$. This is the isovector $n_a(\theta,\phi)$ in the boundary condition.

$$W^a_0\to\beta n_a(\theta,\phi),\quad r\to\infty, \tag{E9}$$

where $n_a n_a=1$. Replacing $\hat{\phi}_a$ by n_a in Eqs. (E2) and (E5) we obtain suitable EM field tensors for the pure SU(2) theory.

Consider the Hsu–Mac solution (4.19) of the pure SU(2) gauge theory. For this solution $n_a=\hat{r}_a$, and we find the same magnetic fields F_{ij} and F'_{ij} as before. But now there is an electric field in the game because

$$A_0=ig(r)/er \tag{E10}$$

is nonzero. From the definition (E5) we find

$$-ieE^{\prime a}_n=-ieG^a_{0n} =-\delta_{an}(1/r^2)gh+(r_a r_n/r^4)[g-rg'+gh]. \tag{E11}$$

From Eq. (4.19) it is easy to verify that

$$i\mathrm{E}'=\mathrm{B}'=(\mathrm{r}/er^3)[1-(\beta r/\sinh\beta r)^2]. \tag{E12}$$

The electric and magnetic fields are the same except for the factor i, and the Hsu–Mac solution is therefore self-dual.

The Hsu–Mac solution is, of course, just the Prasad–Sommerfield–Bogomol'ny solution with the Higgs field ϕ_a reinterpreted as iW^a_0. Because the Higgs field is massless in this solution it gives rise to a long-range electric field. When $m\neq 0$ the physical Higgs field is massive and this reinterpretation is not possible.

APPENDIX F: ELLIPTIC FUNCTIONS

There are twelve Jacobi elliptic functions (see, for example, Gradshteyn and Ryzhik, 1965, Abramovitz and Stegun, 1970). We introduce a generic name E $=E(u)=E(u,k)$ for them, where u is the variable argument and k is the parameter. These functions are solutions of the nonlinear differential equation

$$E''+aE+bE^3=0,$$

where the prime means d/du and $a=a(k)$, $b=b(k)$ are constants which depend on the parameter k. E also satisfies

$$(E')^2+aE^2+\tfrac{1}{2}bE^4=c.$$

The constants a, b, and c for the twelve Jacobi elliptic functions are as follows:

E	a	b	c
sn	$1+k^2$	$-2k^2$	1
cn	$1-2k^2$	$2k^2$	$1-k^2$
dn	$-(2-k^2)$	2	$-(1-k^2)$
ns	$1+k^2$	-2	k^2
nc	$1-2k^2$	$-2(1-k^2)$	$-k^2$
nd	$-(2-k^2)$	$2(1-k^2)$	-1
sc	$-(2-k^2)$	$-2(1-k^2)$	1
sd	$1-2k^2$	$2k^2(1-k^2)$	1
cs	$-(2-k^2)$	-2	$1-k^2$
cd	$1+k^2$	$-2k^2$	1
ds	$1-2k^2$	-2	$-k^2(1-k^2)$
dc	$1+k^2$	-2	k^2.

There are three basic Jacobi elliptic functions,

$$\mathrm{sn}(u,k)\equiv\sin\phi,\ \mathrm{cn}(u,k)\equiv\cos\phi,$$
$$\mathrm{dn}(u,k)\equiv d\phi/du=[1-k^2\sin^2\phi]^{1/2},$$

where ϕ is implicitly defined by the elliptic integral of the first kind,

$$u=\int_0^\phi dt[1-k^2\sin^2 t]^{-1/2}.$$

From the definitions above it follows immediately that

$$\mathrm{s}'=\mathrm{cd},\ \mathrm{c}'=-\mathrm{sd},\ \mathrm{d}'=-k^2\,\mathrm{sc}.$$

Here we have shortened the notation to s = sn, c = cn, d = dn with argument u and parameter k in each case. In the same notation,

$$\mathrm{s}^2=1-\mathrm{c}^2=(1-\mathrm{d}^2)/k^2,$$
$$\mathrm{c}^2=(1-1/k^2)+\mathrm{d}^2/k^2,$$

$$d^2 = 1 - k^2 s^2 = (1-k^2) + k^2 c^2 .$$

The constants a, b, and c above can be easily calculated from these results.

Jacobi elliptic functions are doubly periodic functions of the complex argument u. Specifically,

$$\mathrm{sn}(u + 4mK + i2nK') = \mathrm{sn}u ,$$

$$\mathrm{cn}(u + 4mK + 2n(K + iK')) = \mathrm{cn}u ,$$

$$\mathrm{dn}(u + 2mK + i4nK') = \mathrm{dn}u ,$$

where m and n are any integers and

$$K = K(k) = \int_0^{\pi/2} dt\,[1 - k^2 \sin^2 t]^{-1/2} ,$$

$$K' = K'(k) = K(k');\; k' = (1-k^2)^{1/2}$$

are complete elliptic integrals. Restricting u to be real we see that sn$u$, cn$u$, and dn$u$ have periods $4K$, $4K$, and $2K$, respectively. The shortest period corresponds to $k = 0$, when $K(0) = \pi/2$ and

$$\mathrm{sn}u = \sin u,\; \mathrm{cn}u = \cos u,\; \mathrm{dn}u = 1 .$$

$K(k) > \pi/2$ for $k > 0$; and in the limiting case with infinite period $K(\infty) = \infty$ we have

$$\mathrm{sn}u = \tanh u,\; \mathrm{cn}u = \mathrm{dn}u = 1/\cosh u .$$

The basic functions sn, cn, and dn are finite everywhere on the real-u axis and have the following zeros on this axis:

$$\mathrm{sn}(u,k) = 0 \quad \text{at } u = 2mK ,$$

$$\mathrm{cn}(u,k) = 0 \quad \text{at } u = (2m+1)K ,$$

$$\mathrm{dn}(u,k) \neq 0 \quad \text{for } k < 1 ,$$

where m is any integer. Only for parameter $k = 1$ does dn$u$ have zeros on the real axis, namely, at $u = \pm\infty$.

The Jacobi imaginary transformations

$$\mathrm{sn}(iu,k) = i\,\mathrm{sc}(u,k') ,$$

$$\mathrm{cn}(iu,k) = \mathrm{nc}(u,k') ,$$

$$\mathrm{dn}(iu,k) = \mathrm{dc}(u,k') ,$$

enable one to change from real to imaginary argument or conversely.

As an example of the use of elliptic functions, we consider the static equation of motion of the ϕ^4 theory with spontaneous symmetry breakdown,

$$\phi''(x) + m^2\phi(x) - \lambda\phi^3(x) = 0 .$$

A solution of this equation is

$$\phi = (m/\sqrt{\lambda})(-b/a)^{1/2} E(mx/\sqrt{a}, k) ,$$

where E is any of the 12 Jacobi elliptic functions. Choosing $E = $ sn and setting the parameter $k = 1$ we find the well known kink solution

$$\phi = (m/\sqrt{\lambda}) \tanh(mx/\sqrt{2}) .$$

APPENDIX G: 't HOOFT-POLYAKOV DIFFERENTIAL EQUATIONS

The equations of motion following from the 't Hooft-Polyakov *ansatz*

$$e\phi_a = (r_a/r^2)g(r);\; W_0^a = 0 ;$$
$$eW_i^a = \varepsilon_{ain}(r_n/r^2)[1 - h(r)] \tag{G1}$$

are

$$r^2 h'' = h(h^2 - 1 + g^2) , \tag{G2a}$$

$$r^2 g'' = 2gh^2 - m^2 r^2 g + (\lambda/e^2) g^3 . \tag{G2b}$$

Here we discuss some properties of these coupled differential equations. For $m^2 \neq 0$, $\lambda \neq 0$ they evidently cannot be solved analytically; at least, no one has succeeded up to now. The explicit Prasad–Sommerfield solution is known in the limit $m^2 = 0$, $\lambda = 0$ with m^2/λ finite, of course. One can verify that solutions of Eqs. (G2) are analytic in the coefficients m^2 and λ/e^2 near the values $m^2 = 0$, $\lambda/e^2 = 0$ because a power series expansion in these coefficients can be set up. Therefore it is possible to continue away from the Prasad–Sommerfield solution. To date this has only been done numerically.

We make two preliminary comments. For $m \neq 0$ it is trivial to verify that no *constant* solution of Eqs. (G2) exists which has $g \neq 0$. This means that for $m \neq 0$ no pointlike solution can be obtained from the *ansatz* (G1) with nonzero Higgs field. Of course one expects this: $m \neq 0$ means that the local gauge symmetry is broken, and so there must be short-range potentials present—i.e., the solution is nonpointlike. To obtain a pointlike solution one must either switch off the Higgs field entirely ($g = 0$), or switch off the gauge symmetry breaking ($m = 0$).

For $m = 0$ and arbitrary λ there is a constant solution of Eqs. (G2), namely

$$g^2 = 1/(1 - \lambda/2e^2),\; h^2 = 1/(1 - 2e^2/\lambda) .$$

The Higgs and gauge fields in this case correspond to unbroken local SU(2) symmetry ($\phi_a \to 0$ at infinity).

1. $g(r) = 0$

Let us first consider the case with no Higgs field, $g = 0$. Then we have the differential equation studied by Wu and Yang (1968),

$$r^2 h'' = h(h^2 - 1) . \tag{G3}$$

This equation has three constant solutions, $h = \pm 1$ and $h = 0$. The first two are vacuum solutions, while $h = 0$ in the *ansatz* (G1) corresponds to the gauge potential of a point magnetic monopole. Note that if h is a solution of Eq. (G3) then so is $-h$.

In a region where $h^2 \gg 1$, Eq. (G3) becomes approximately $r^2 h'' = h^3$, which has the solution

$$h = r_0\sqrt{2}/(r - r_0) . \tag{G4}$$

Therefore $h(r)$ has a singularity at some point $r = r_0$ in this region. When $h^2 \ll 1$, on the other hand, Eq. (G3) becomes approximately $r^2 h'' = -h$, which has the solution

$$h(r) = A\sqrt{r}\cos(\sqrt{3}\ln r) + B\sqrt{r}\sin(\sqrt{3}\ln r) \tag{G5}$$

with constant A and B. This function is poorly behaved in the limit $r \to \infty$ and therefore the boundary condition $h \to 0$ for $r \to \infty$ is unphysical.

Solutions of Eq. (G3) which are analytic about the points $r=0$ or $r=\infty$ will have the following power-series expansions (Rosen, 1972):

$$h=1+(ar)^2+\tfrac{3}{10}(ar)^4+\tfrac{1}{10}(ar)^6+\cdots, \tag{G6}$$

$$h=1+(b/r)+\tfrac{3}{4}(b/r)^2+\tfrac{11}{20}(b/r)^3+\cdots. \tag{G7}$$

However, no singular solution exists which connects these two power series except for the constant solution with $a=b=0$.

Wu and Yang found a numerical solution of Eq. (G3) with the behavior $h\to 1$ for $r\to\infty$ and $h\to 0$ for $r\to 0$. Equation (G5) gives the general form of h in the limit $r\to 0$. Although h is everywhere nonsingular in this solution, the gauge potential W_i^a in Eq. (G1) is singular at $r=0$. Besides this solution and the constant solutions $h=0, 1$ there are no other solutions of Eq. (G3) which are finite for all $0<r<\infty$.

2. $g(r)\neq 0$

Now let us see how the situation improves when the Higgs field is switched on. An essential change occurs in the equation for h'' when h is small: $r^2h''=h(g^2-1)$. The presence of g here enables h to vanish smoothly in the limit $r\to\infty$ [unlike Eq. (G5)]. Since $g\to(em/\sqrt{\lambda})r$ in this limit (for the energy to be finite) we see that h behaves like

$$h\to Ae^{-(me/\sqrt{\lambda})r}, \quad r\to\infty. \tag{G8}$$

This is, in fact, the behavior we wish to have because $M_W=em/\sqrt{\lambda}$ is the mass of the massive gauge-field components.

The "physical" Higgs field σ is defined as follows:

$$\phi_a\equiv\hat{r}_a(m/\sqrt{\lambda}+\sigma).$$

This field should acquire a mass $M_H=\sqrt{2}m$ through the Higgs mechanism. We can see that this happens from Eq. (G2b). Substituting $\sigma=\gamma/er$ in this equation we find

$$r^2\gamma''=2\gamma(h^2+m^2r^2)+(2me/\sqrt{\lambda})rh^2 + (3m\sqrt{\lambda}/e)r\gamma^2+(\lambda/e^2)\gamma^3.$$

At large r, assuming (G8), γ satisfies $\gamma''=2m^2\gamma$ and therefore

$$\gamma\to De^{-\sqrt{2}mr}, \quad r\to\infty. \tag{G9}$$

Next, consider the small-r region. It is easy to show that Eqs. (G2) are satisfied by the expansions

$$h(r)=1+Br^2+\cdots;\ g(r)=Cr^2+\cdots \tag{G10}$$

for some constants B and C. Note that these boundary conditions have the good property that ϕ_a and W_i^a in Eq. (G1) vanish with r, so there is no singularity at $r=0$.

The existence of the 't Hooft–Polyakov solution reduces to the following question: Does there exist a nonsingular solution of Eqs. (G2) that connects the boundary conditions at $r=\infty$ in Eqs. (G8) and (G9) with the boundary conditions at $r=0$ in Eq. (G10)? The answer is yes. A numerical solution has been obtained by computer. (For a good discussion of the numerical calculation see Bais and Primack, 1976.) The functions g/r and h in this solution interpolate smoothly and without nodes between their values

$$g/r=em/\sqrt{\lambda},\ h=0 \text{ at } r=\infty$$

and their values

$$g/r=0,\ h=1 \text{ at } r=0.$$

This solution changes smoothly into the Prasad–Sommerfield solution (4.19) when $m^2\to 0$, $\lambda\to 0$ with m^2/λ finite. Frampton (1976) has discussed the general form of the solution in this limit. The absence of nodes in $g(r)$ is obvious in this limit, for Eq. (G2b) becomes $g''=2gh^2/r^2\geq 0$. Thus g'' is everywhere non-negative, and $g(r)$ must become infinite as $r\to\infty$.

APPENDIX H: THE GRIBOV AMBIGUITY

Recently Gribov (1977) has discovered an ambiguity in the Coulomb-gauge formulation of non-Abelian gauge theories. In this gauge the potential W_μ^a is *not* uniquely determined by the usual gauge-fixing conditions

$$W_0^a=0;\ \partial_iW_i^a=0; \tag{H1}$$

there is some remaining gauge freedom. To investigate this we perform a time-independent gauge transformation $\omega^{-1}=\omega^{-1}(\mathbf{x})$ leading to a new potential with $W_0'=0$ and

$$W_i'=\omega^{-1}W_i\omega-(i/e)(\partial_i\omega^{-1})\omega. \tag{H2}$$

This potential will also satisfy the transversality condition $\partial_iW_i'=0$ if ω satisfies

$$\partial_i[(\partial_i\omega)\omega^{-1}]=(e/i)[(\partial_i\omega)\omega^{-1}, W_i], \tag{H3}$$

i.e., if the covariant divergence of the pure-gauge potential $-(i/e)(\partial_i\omega)\omega^{-1}$ is zero. Gribov shows that for rotationally symmetric W_i^a Eq. (H3) has nontrivial solutions—hence the ambiguity mentioned above. This lack of uniqueness in the gauge-fixing procedure leads to obvious technical problems in the quantum theory whose consequences have to be carefully investigated. The Gribov ambiguity is more than a technical problem, however—it is deeper than that. Very important physical questions in non-Abelian gauge theory seem to be connected with this ambiguity.

Gribov pointed out that there is a probable connection with the confinement problem. The conditions (H1) reduce the number of independent gauge-field components from 12 to six, which correspond to the six spin degrees of freedom of three massless vector bosons in the quantized theory. However, because of the ambiguity or remaining gauge freedom, there are in fact more than six independent functions present at the classical level. This seems to imply that the theory cannot be formulated in terms of massless fields alone.

We can make this argument more precise by noting that the SU(2) gauge theory has *three* rotationally symmetric vacua; the usual one $W_i=0$ and two other (Gribov) vacua with $W_i\neq 0$. (Beyond these there may be an infinity of nonsymmetric vacua.) The usual vacuum has topological charge $Q_T=0$ and the Gribov vacua have $Q_T=\pm\frac{1}{2}$. All three are gauge equivalent, and the Coulomb-gauge vacuum is initially degenerate. Instantons do not remove this degeneracy, for they only tunnel between the $Q_T=\frac{1}{2}$ and $Q_T=-\frac{1}{2}$ vacua.

However, merons tunnel between the $Q_T=\pm\frac{1}{2}$ vacua and $Q_T=0$ vacuum, thus restoring the symmetry of the vacuum and eliminating the need for Goldstone gauge quanta. Thus when merons are present the gauge theory may be in the confining phase. But this is not true when only instantons are present. We see that the discovery of the Gribov vacua provides additional support for the conjectured confinement mechanism of Callen, Dashen, and Gross (1977, 1978), where instantons dissociate into meron pairs at a critical value of the effective YM coupling when the theory goes into the confining phase.

1. Spherically symmetric fields

We first discuss the Gribov problem for spherically symmetric fields, using Witten's *ansatz* (3.27) for the Minkowski gauge potential which we give again here for convenience,

$$eW_0^a=-(x_a/r)A_0\;;$$
$$eW_i^a=\varepsilon_{ian}(x_n/r^2)(1+\phi_2)+(x_ax_i/r^2)A_1+(\delta_{ai}-x_ax_i/r^2)(1/r)\phi_1\,. \tag{H4}$$

We shall also need the transformation property

$$A_\mu\to A_\mu-\partial_\mu f,\qquad \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\to\begin{pmatrix}\cos f & \sin f\\ -\sin f & \cos f\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}, \tag{H5}$$

under the gauge transformation $\omega=\exp[\frac{1}{2}if(r,t)\hat{r}\cdot\sigma]$. $A_0=0$ in the Coulomb gauge, and the transversality condition is

$$\partial_1(r^2A_1)=2\phi_1,\;\;\partial_1=\partial/\partial r\,. \tag{H6}$$

After the gauge transformation (H5) the transversality condition becomes

$$\partial_1[r^2(A_1-\partial_1 f)]=2(\cos f\phi_1+\sin f\phi_2)\,, \tag{H7}$$

or using (H6)

$$\partial_1(r^2\partial_1 f)=4\sin\tfrac{1}{2}f[\sin\tfrac{1}{2}f\phi_1-\cos\tfrac{1}{2}f\phi_2]\,. \tag{H8}$$

Here $f=f(r)$ if we wish to remain in the Coulomb gauge.

A comparatively trivial example of the Gribov ambiguity can already be given. Consider any configuration with $A_1=\phi_1=0$ so that Eq. (H6) is identically satisfied. A different configuration is obtained by the gauge transformation $f=\pm\pi$, where $A'_\mu=A_\mu$, $\phi'_a=-\phi_a$. The transversality condition (H6) is still satisfied. Note that $f=\pm\pi$ is the gauge transformation

$$\omega=e^{\pm i(\pi/2)\hat{r}\cdot\sigma}=\pm i\hat{r}\cdot\sigma\,. \tag{H9}$$

In particular, the pure-gauge potential constructed from ω is transverse. One does not have this freedom in an Abelian gauge theory.

Nontrivial gauge transformations f which vanish at $r=0$ are more interesting. Let us suppose that

$$f\approx(Mr)^\alpha,\quad \alpha>0,\;\; r\approx 0\,.$$

If ϕ_a is nonsingular at $r=0$ then from Eq. (H8) we obtain a formula for α

$$\alpha(\alpha+1)+2\phi_2(r=0)=0\,. \tag{H10}$$

The potential W_i^a in Eq. (H4) will be singular at $r=0$ if $\phi_2(r=0)\neq\pm1$, i.e., if there is not vacuum at $r=0$. $\phi_2=+1$ does not provide a real value of α, but $\phi_2=-1$ leads to $\alpha=1$ in Eq. (H10) and therefore

$$f\approx Mr,\;\; r\approx 0\,. \tag{H11}$$

Quite generally, if the potential is well behaved at the origin, the Gribov gauge transformation f will have the behavior (H11) at small r. Note that the mere existence of such a gauge transformation implies the existence of a scale or "mass" M in the small $-r$ region. M is, of course, arbitrary because of scale invariance.

2. Vacuum solutions

The usual Coulomb-gauge vacuum is $W_i=0$. Gribov discovered two additional vacua with *nonzero* gauge potential W_i, which are gauge equivalent to the usual vacuum. The new vacua, which we shall refer to as Gribov vacua, have topological charge $Q_T=\pm\frac{1}{2}$, while the usual vacuum has $Q_T=0$. This reminds one of the situation in the temporal gauge $W_0=0$ (discussed in Sec. VII.A.) where there are infinitely many topologically distinct vacua. The difference is that the temporal gauge was never supposed to be a complete gauge specification, while the Coulomb gauge was previously thought to uniquely determine the gauge potential.

The new vacua are obtained from the usual one ($A_\mu=\phi_1=0$ and $\phi_2=-1$ in Witten's *ansatz*) by a gauge transformation f. From Eq. (H8) we see that f must satisfy

$$\frac{d}{dr}\left(r^2\frac{df}{dr}\right)-2\sin f=0\,. \tag{H12}$$

To visualize the solutions of this equation it helps to introduce the variable $t=\ln r$. Then Eq. (H12) becomes

$$\ddot{f}+\dot{f}-2\sin f=0\,, \tag{H13}$$

where the dot means d/dt. This is the equation of motion of a damped pendulum, as discussed by Wadia and Yoneya (1976), Gribov (1977), and others. The pendulum points upwards when $f=0$ (mod$2n\pi$) and downwards when $f=\pm\pi$ (mod$2n\pi$). One can also think in terms of a particle moving with friction in a potential $V=-4\sin^2\frac{1}{2}f$ where f is the position of the particle. The usual vacuum corresponds to the particle remaining at the top of the potential, i.e., at $f=0$. The Gribov vacua correspond to the particle starting at $f=0$ in the distant past ($t=-\infty$ or $r=0$) with zero velocity and sliding downhill to the left or right, coming to rest at the bottom of the potential $f=\pm\pi$ in the distant future ($t=\infty$ or $r=\infty$). Moreover, the solutions with positive and negative f are identical except for the sign. The positive f solution satisfies the boundary conditions

$$f\to Me^t,\quad t\to-\infty\,, \tag{H14}$$
$$f\to\pi+Ae^{-t/2}\cos(\pm\tfrac{1}{2}\sqrt{7}\,t+B),\;\; t\to+\infty\,. \tag{H15}$$

[Jackiw, Muzinich, and Rebbi (1978) give the numerical solution $f(t)$ over the full range $-\infty\leq t\leq\infty$. The particle slides down, slightly overshoots the bottom of the potential because its velocity is nonzero, but the friction soon eliminates the residual motion. Therefore, the commonly made statement that $|f|<\pi$ for finite t is slightly incorrect.]

For large r (where $f=\pi$) the Gribov vacuum potential has the form

$$eW_i^a=\varepsilon_{ian}(2x_n/r^2)=-i\,\mathrm{Tr}\sigma_a(\partial_i\omega)\omega^{-1}\,, \tag{H16}$$

with $\omega=\hat{r}\cdot\sigma$. This is gauge equivalent to the usual vacuum $W_i=0$. For small r [where $f=Mr-(Mr)^3/30+\cdots$] we find

$$eW_i^a=-\delta_{ai}M+\varepsilon_{ian}\tfrac{1}{2}x_nM^2+O(r^2)\,. \tag{H17}$$

This potential correctly yields $G_{\mu\nu}=0$. Curiously, if we kept only the first (constant) term then W_i^a would be a vacuum solution of the massive YM theory with mass $=i\sqrt{2}M$.

Now we calculate the topological charge of the Gribov vacua. The pure-gauge *ansatz* (H4) with

$$A_0=0,\ A_1=-df/dr,\ \phi_1=-\sin f,\ \phi_2=-\cos f, \tag{H18}$$

can also be written in the form

$$eW_i=-i(\partial_i\omega)\omega^{-1},\ \ \omega=e^{-i(f/2)\hat{r}\cdot\sigma}\,. \tag{A19}$$

From the definition (7.8) of the topological charge Q_T one finds after some algebra (Abbott and Eguchi, 1977)

$$J_0=\frac{2}{r^2}\frac{d}{dr}(f-\sin f)\,,$$

and therefore

$$Q_T=\frac{1}{16\pi^2}\int d^3x\,J_0=\frac{1}{2\pi}[f-\sin f]_{r=0}^{r=\infty}\,. \tag{H20}$$

For the Gribov vacua with $f\to\pm\pi$ as $r\to\infty$ we see that $Q_T=\pm\frac{1}{2}$.

If we drop the transversality condition then it is easy to show that the three symmetric Coulomb-gauge vacua are gauge equivalent to the infinity of symmetric temporal gauge vacua with $Q_T=n$ (Abbott and Eguchi, 1977). Moreover, one can construct an additional set of vacuum potentials by replacing f in Eqs. (H18)–(H20) by nf. These new potentials have topological charge $Q_T=n/2$ (Abbott and Eguchi, 1977).

3. Nonvacuum solutions

We now consider the general field configuration $\phi_1=A_\mu=0$, $\phi_2=\phi_2(r,x_0)$. The transversality condition is

$$\partial_1(r^2\partial_1 f)+2\sin f=0\,, \tag{H21}$$

and the equation of motion for ϕ_2 is

$$r^2(\partial_0^2-\partial_1^2)\phi_2=\phi_2(1-\phi_2^2)\,. \tag{H22}$$

Here we are interested in the r dependence of f, and the time x_0 is regarded as a parameter.

Gribov (1977) has discussed the solutions of Eq. (H21) without restricting ϕ_2 to be a solution of the equation of motion (H22). The mechanical analog problem here is

$$\ddot{f}+\dot{f}+2\sin f\phi_2(t)=0\,, \tag{H23}$$

which is a particle moving in a more complicated potential $V(f)$ than in the vacuum problem. Evidently we can obtain many different solutions by choosing the function $\phi_2(t)$ suitably. Suppose that $\phi_2\to-1$ for $t\to\pm\infty$ so that in the distant past and future we have the same potential $V(f)$ as before. For finite time t we can change the shape of the new potential $V(f)$ as we wish. The particle begins at $f=0$ as before, and it may return to $f=0$ at the end of its motion, or slide down to $f=\pi$ as before, or end up at the new position $f=2\pi$, and so on. Turning things around we may say that there exist many functions ϕ_2 for which the Gribov ambiguity exists. These functions may or may not be solutions of the equations of motion—that is a problem for future study. The important point is that the Gribov ambiguity is present at a general level for nonvacuum fields.

Bender, Eguchi, and Pagels (1978) have examined the problem of finding infinitesimal Gribov gauge transformations for the static potential $\phi_2=\phi_2(r)$, $\phi_1=A_\mu=0$. The change of variable $f(r)=(1/r)y(r)$ in Eq. (H21) leads to the Schrödinger-like equation for infinitesimal $y(r)$,

$$y''+[E+(2/r^2)\phi_2]y=0\,,$$

where $E=0$ is the energy of interest. Bender, Eguchi, and Pagels discuss the solutions of this Schrödinger equation assuming that the function $\phi_2(r)$ which determines the "potential" $V=-2\phi_2/r^2$ is monotonic with $\phi_2(0)=-1$. For $\phi_2(\infty)>0$ (<0) there are, of course, bound states (no bound states). Moreover, if $\phi_2(\infty)>\frac{1}{8}$ there is an accumulation of bound states at $E=0$. This amounts to an accumulation of infinitesimal gauge transformations which satisfy the transversality condition. The authors argue that this accumulation phenomenen implies an interaction energy growing like r^3 between nonsinglet sources, and therefore confinement.

4. One-meron solution

The meron solution has topological charge $q=\frac{1}{2}$, and one would expect that it describes tunneling between the Gribov vacuum and the normal vacuum. One known form of the meron solution in the Coulomb gauge does *not* have this property:

$$eW_i^a=\varepsilon_{ian}\frac{x_n}{r^2}(1+x_0/\sqrt{x^2})\,. \tag{H24}$$

This solution interpolates between the vacuum $W_i=0$ at $x_0=-\infty$ and the vacuum $eW_i^a=\varepsilon_{ian}2x_n/r^2$ at $x_0=+\infty$, both of which are $Q_T=0$ vacua. However, there exists a gauge-equivalent form of this solution which does tunnel between the two vacua (Chiu, Kaul, and Takasugi, 1978).

The one-meron solution corresponds to $\phi_2=x_0/\sqrt{x^2}$ in Eq. (H4) with the other *ansatz* functions zero. The transversality constraint (H8) on any gauge transformation of this solution is

$$\partial_1(r^2\partial_1 f)+2\sin f\,x_0/\sqrt{x^2}=0\,. \tag{H25}$$

The nontrivial solutions of this equation are time dependent, $f=f(r,x_0)$, and therefore the gauge-transformed one-meron solutions are not in the Coulomb gauge because $W_0\neq0$. However, in the distant past $x_0\to-\infty$, Eq. (H25) reduces to the vacuum transversality condition (H12) whose solutions are the Gribov vacua and the usual vacuum. Moreover, in the distant future $x_0\to+\infty$ we see that $f\to\bar{f}+\pi$ where $\bar{f}$ satisfies Eq. (H12). Therefore, a nontrivial solution $f(r, x_0)$ of Eq. (H25) may very well interpolate between the Gribov vacuum and the normal vacuum. But this interpolating potential is not in the Coulomb gauge.

How do the solutions of Eq. (H25) behave? First of all we note that at time $x_0=0$ f satisfies $\partial_1(r^2\partial_1 f)=0$ (for $r\neq 0$), which implies that f is constant everywhere in space. (The solution $f=c/r$ is unacceptable.) Next, consider the small $-r$ region where Eq. (H25) becomes

$$\partial_1(r^2\partial_1 f)+(x_0/|x_0|)2\sin f=0\,.$$

For $x_0<0$ ($x_0>0$) the nontrivial solution is $Ar(Ar+\pi)$. In the limit $r\to\infty$ the transversality condition becomes $\partial_1(r^2\partial_1 f)=0$ (for finite x_0) and f is either constant or $O(1/r)$. Evidently, a nontrivial solution which is continuous at $x_0=0$ must have one of the forms (Chiu, Kaul, and Takasugi, 1978)

$$f=\begin{cases}\pi & x_0<0\\ g(r,|x_0|) & x_0>0\end{cases},\qquad f=\begin{cases}\pi-g(r,|x_0|) & x_0<0\\ 0 & x_0>0\end{cases},\tag{H26}$$

where

$$g(r,0)=\pi \text{ for } r\neq 0,\quad g(0,|x_0|)=0\,,$$
$$g(r,|x_0|)\to \text{Gribov function},\quad |x_0|\to\infty\,.$$

Both solutions tunnel between the normal vacuum and the Gribov vacuum. These solutions are discussed in more detail by Chiu, Kaul, and Takasugi (1978). The topological charge density is still $\frac{1}{2}\delta^4(x)$, as for the original meron solution. In contrast, the solution (H24) above has zero topological charge. [This has been absorbed by the singular gauge transformation used to obtain the form (H24) from the original solution.]

5. Instanton solution

In Sec. VII.D we have shown that in Witten's (Euclidean) notation the instanton solution has the form

$$A_0=-2r/(x^2+\lambda^2),\quad A_1=2x_0/(x^2+\lambda^2)\,,$$
$$\phi_1=2rx_0/(x^2+\lambda^2),\quad \phi_2=-(\lambda^2+x_0^2-r^2)/(x^2+\lambda^2)\,.\tag{H27}$$

A gauge transformation to the $A_0'=0$ gauge is

$$\bar f=-(2r/\sqrt{r^2+\lambda^2})\tan^{-1}(x_0/\sqrt{r^2+\lambda^2})\,.\tag{H28}$$

The new *ansatz* functions A_1' and ϕ_a' are somewhat cumbersome, and we do not give them explicitly. However, it is interesting to note that in the limits $x_0\to\pm\infty$ these functions become $A_1'=-\partial_1\bar f$, $\phi_1'=-\sin\bar f$, $\phi_2'=-\cos\bar f$ where $\bar f=\mp\pi r/\sqrt{r^2+\lambda^2}$, which corresponds to a pure gauge potential with gauge function

$$\omega=\exp[\pm\tfrac{1}{2}i(\pi r/\sqrt{r^2+\lambda^2})\hat r\cdot\sigma]\to\pm i\hat r\cdot\sigma,\quad r\to\infty\,.\tag{H29}$$

Here we see that, in the $W_0=0$ gauge, the instanton interpolates between two Gribov-like vacua at large r. As we demonstrate shortly, there exists a Coulomb-gauge form of the instanton solution which exactly interpolates between the two Gribov vacua with topological charge $=\pm\frac{1}{2}$. These two vacua are therefore connected by instanton tunneling (Sciuto, 1977; Abbott and Eguchi, 1977).

First we make a parenthetical remark. Any function of r can be added to $\bar f$ in Eq. (H28) without changing $A_0'=0$. Let us consider the gauge transformation

$$\bar f=-(2r/\sqrt{r^2+\lambda^2})\,[\tan^{-1}(x_0/\sqrt{r^2+\lambda^2})+\tfrac{1}{2}\pi]$$

with the limiting behavior

$$\bar f\to 0,\quad x_0\to-\infty\,;$$
$$\bar f\to -2\pi r/\sqrt{r^2+\lambda^2},\quad x_0\to+\infty\,.$$

In the limit $x_0\to-\infty$ the gauge-transformed potential obtained from (H27) becomes the vacuum $W_\mu=0$. In the limit $x_0\to+\infty$ the same potential becomes pure gauge with gauge function

$$\omega=\exp[-\tfrac{1}{2}i(2\pi r/\sqrt{r^2+\lambda^2})\hat r\cdot\sigma]$$
$$=\cos(\pi r/\sqrt{r^2+\lambda^2})-i\hat r\cdot\sigma\sin(\pi r/\sqrt{r^2+\lambda^2})\,.$$

This is essentially the gauge transformation ω_1 used in Sec. VII to construct the topological vacua $|n\rangle$. Here we have explicit verification of a claim made in Sec. VII that the instanton tunnels between the vacua $|0\rangle$ and $|1\rangle$ in the $W_0=0$ gauge.

Now on to the Gribov problem. We begin in the gauge (H27), noting that the transversality condition $\partial_1(r^2\partial_1 A_1)=2\phi_1$ is *not* satisfied in this gauge. We have to find a gauge function $f=f(r,x_0)$ which satisfies the transversality condition (H8) (Wadia and Yoneya, 1976)

$$\partial_1(r^2\partial_1 f)-2\sin f=\frac{4rx_0}{x^2+\lambda^2}(1-\cos f)-\frac{4r^2}{x^2+\lambda^2}\sin f\,.\tag{H30}$$

In the limit $|x_0|\to\infty$ this condition becomes the condition (H12) for the Gribov vacuum. Moreover, among the *ansatz* functions (H27) only $\phi_2\to-1$ remains nonzero in this limit. Therefore the transformed instanton solution must go smoothly into one of the Gribov vacua when $x_0\to\pm\infty$. It is clear that different Gribov vacua are connected in this way becuase f in Eq. (H30) must change sign at $x_0=0$. Therefore the solution of Eq. (H30) has the form (Sciuto, 1977; Abbott and Eguchi, 1977)

$$f=\begin{cases}-g(r,|x_0|,\lambda) & x_0<0\\ g(r,|x_0|,\lambda) & x_0>0\end{cases},\tag{H31}$$

where

$$g\to 0,\quad \partial_0 g\to 0;\quad x_0\to 0\,;$$
$$g\to\text{Gribov function};\quad |x_0|\to\infty\,.$$

6. Vacuum tunneling in the Coulomb gauge

We have seen that merons tunnel between the Gribov vacua and $W_\mu=0$, while instantons only tunnel from one Gribov vacuum to the other. This seems to imply that merons play a physical role in YM theory, as they are necessarily present in the ground state of lowest energy. On the other hand, all meron solutions are singular and thus have infinite action. This obscures their physical interpretation. The situation with respect to merons would be clarified if one understood why they are singular. Ideally, meron advocates would like to be able to show that merons have to be singular on physical grounds. No proof of this is known to the author. However, we believe that the singular nature of merons is dictated by their connection with the Gribov vacua (Actor, 1979).

A Gribov vacuum has size and position. Its size is essentially $1/M$, where M is the parameter in Eq. (H14). In the limit $M\to\infty$ the Gribov vacuum becomes pointlike and singular:

$$f=\begin{matrix}0 & r=0\\ \pi & r>0\end{matrix}. \tag{H32}$$

If it were not for the point $r=0$ this singular Gribov vacuum would coincide with the perturbative vacuum $W_\mu=0$ up to a trivial gauge transformation. Thus its topological charge $Q_T=\frac{1}{2}$ is concentrated at the singular point $r=0$. Now the limit $M\to\infty$ is only a scale change, which one is allowed to make because of conformal invariance. (This scale change of course does not affect the pointlike meron solution.) Obviously a nonsingular Euclidean solution could not interpolate between the pointlike Gribov vacuum and $W_\mu=0$. However, the singular meron solution has just the right form for this interpolation. Indeed, the pointlike nature of merons enables them to tunnel from Gribov vacua of any size to $W_\mu=0$.

The meron's topological charge is concentrated at a point in E^4. From this we conclude that the meron tunneling in Eq. (H26) proceeds in the following way. Consider the tunneling W_μ (Gribov) $\to W_\mu=0$ to be explicit. At $x_0=-\infty$ there is a Gribov vacuum with arbitrary size $1/M$. For finite $x_0<0$ this configuration *shrinks* until it is finally pointlike. At $x_0=0$ the pointlike meron cancels the singularity in the pointlike configuration and what remains is a nonsingular configuration (with $Q_T=0$) which dissolves into the $W_\mu=0$ vacuum as $x_0\to\infty$. (For more details see Actor, 1979.)

Merons are the only known YM solutions with half-integral topological charge. There probably does not exist a nonsingular solution with $q=\frac{1}{2}$ which could interpolate between a finite Gribov vacuum and $W_\mu=0$. Like merons, this hypothetical solution would have equally strong self-dual and self-antidual parts at large x^2. However, it is generally believed that finite-action, non-self-dual solutions do not exist.

Next consider instanton tunneling, as in Eq. (H31). Here the instanton interpolates smoothly from one finite Gribov vacuum to the other. If we make a scale change such that the Gribov vacua become pointlike, then the interpolating configuration must also become singular, i.e., a two-meron configuration. This is an indication that instantons are in some sense equivalent to meron pairs.

What about multi-instanton solutions? Ademollo, Napolitano, and Sciuto (1978) pointed out that there is no room for these solutions in the Coulomb gauge, for the only symmetric vacua are the ones with $Q_T=0$, $\pm\frac{1}{2}$. If we try to write down a two-instanton solution in this gauge, then at some time $x_0=T$ it must have a discontinuity at which Q_T changes by one unit. Suppose the two instantons are very far apart, and the tunneling begins in the $Q_T=-\frac{1}{2}$ vacuum and proceeds to the $Q_T=+\frac{1}{2}$ vacuum at $x_0=T$. At this time the potential must change discontinuously back to the $Q_T=-\frac{1}{2}$ Gribov vacuum. Then the second instanton tunnels continuously to the $Q_T=+\frac{1}{2}$ vacuum again. For multi-instantons the discontinuity is repeated. Multimeron solutions would also have to have discontinuities.

APPENDIX I: TOPOLOGICAL CONSIDERATIONS

(1) Sometimes it is useful to think of a local SU(2) gauge transformation as a mapping of space-time onto the SU(2) group. This helps one to visualize its topological structure. Although most gauge transformations are trivial in a topological sense, there are certain ones with very interesting topological properties. These are gauge transformations which define a mapping of some topologically nontrivial object in space-time onto a similar object in the SU(2) group manifold. The topologically nontrivial objects we have in mind here are loops, spheres, (tori?), etc. The mapping of a loop onto a loop, or a sphere onto a sphere, can cover the latter n times where n is any integer. Mappings with different n are topologically inequivalent, and this is an important distinction.

(2) The manifold of all space-time points has no interesting topological structure. However, the SU(2) group manifold is topologically nontrivial: It is topologically equivalent to S^3. To see this we recall that an arbitrary SU(2) transformation can be written

$$\omega=\sigma_\mu f^\mu;\quad \sigma_\mu=(i,\sigma),\quad f_\mu^*=f_\mu\,. \tag{I1}$$

Because $\omega^+=\omega^{-1}$ it is necessary that the four real functions f_μ satisfy

$$f_0^2+f_i^2=1\,. \tag{I2}$$

Thus one can interpret f_μ as a unit vector in a four-dimensional Euclidean space. The manifold of points touched by the tip of this vector is the unit sphere S^3. This sphere is the SU(2) group manifold.

(3) The topological structure of the SU(2) group manifold is of immediate relevance for Euclidean solutions of the SU(2) gauge theory. In Euclidean space-time any physical solution W_μ of the Yang–Mills equations of motion should become pure gauge as $x^2\to\infty$,

$$eW_\mu\to -i(\partial_\mu\omega)\omega^{-1}\,. \tag{I3}$$

The gauge transformation ω defines a mapping of the sphere at infinity, S_∞^3, into the SU(2) group manifold S^3. This mapping can cover the latter sphere zero times: All points on S_∞^3 are mapped onto one group element, say the identity element, so that $W_\mu\to 0$ as $x^2\to\infty$. It can cover the latter sphere once: each point on S_∞^3 is mapped onto a different point on S^3. It can cover the latter sphere n times: then n points on S_∞^3 are mapped onto one point on S^3. The general situation is clear. Euclidean solutions of the SU(2) gauge theory which satisfy the boundary condition (I3) automatically define a mapping $S_\infty^3\to S^3$ that has a topological index n equal to an integer. A mathematical expression of this fact is

$$\pi_3(\mathrm{SU}_2)=\pi_3(S^3)=Z\,. \tag{I4}$$

Here Z is the group of additive integers, and π_3 is the third homotopy group. Before saying what this group is, we mention another interesting case.

(4) In the SU(2) monopole solutions discussed in Sec. IV the Higgs field satisfies the boundary condition at infinity $\phi_a \to \hat{r}_a m/\sqrt{\lambda}$. Such a boundary condition is needed to make the potential energy vanish as $r \to \infty$. (Specifically, $\phi_a \phi_a \to m^2/\lambda$ is necessary.) This particular boundary condition defines a map of the sphere at spatial infinity, S^2_∞, onto a sphere S^2 in the SU(2) manifold. [Here we regard the SU(2) manifold as a solid ball in a three-dimensional Euclidean space. The sphere S^2 is determined by $\phi_a \phi_a = \text{const}$ where ϕ_a is a vector in this space.] The latter sphere is covered one time. Other boundary conditions can easily be imagined in which the sphere S^2 is covered n times. The mathematical expression of this is

$$\pi_2(S^2) = Z\,, \tag{I5}$$

where π_2 is the second homotopy group. We now proceed to a discussion of the homotopy groups.

(5) Let us begin with the simplest homotopy group, π_1, making no pretense of giving a mathematically rigorous treatment. We are only trying to explain what homotopy groups are. For an excellent review of results on homotopy groups, with many references to the mathematical literature, see Boya, Cariñena, and Mateos (1978). A standard (although slightly outdated) reference is Steenrod (1951).

Definition of $\pi_1(X)$: The first homotopy group is a group whose elements are equivalence classes. The elements which are collected into these equivalence classes belong to the set of all maps of the unit circle S^1 into a topological space X. This unit circle, or loop, is given a direction, say counterclockwise, and some point P on the circle is chosen to play a special role. One can think of P as the start (and end) point of a circular path traversed in the chosen counterclockwise direction. Moreover, a base point N in the space X is chosen to play a special role. The map $S^1 \to X$ is such that P always gets mapped onto N; otherwise the mappings are arbitrary (but continuous). Two maps are said to be *equivalent*, or *homotopic*, if they can be continuously deformed into one another. This is a meaningful equivalence, which for mappings of loops can be easily visualized. Given this equivalence, one can readily group all maps $S^1 \to X$ into equivalence classes. *Each equivalence class is an element of $\pi_1(X)$, and these elements form a group.* This is the essential idea to grasp in connection with all homotopy groups; the higher ones are defined analogously.

The identity element of $\pi_1(X)$ is the map $S^1 \to N$ of the entire loop S^1 onto the point N in X, together with all maps which are homotopic to this one (i.e., all maps whose image in X can be continuously contracted to the point N). If X is simply connected, then, by definition, $\pi_1(X)$ consists only of the identity element. If X is not simply connected, then $\pi_1(X)$ is nontrivial. For the sake of illustration we now consider a particular topological space X.

As an example (see Roman, 1975) let us take X to be Euclidean two-space with a hole cut out, so that it is not simply connected. Then we have the situation shown in Fig. 1. In Fig. 1(a) we see the type of map which belongs to the identify element of $\pi_1(X)$. In Fig. 1(b) we see another type of map which is clearly in a different homotopy class, and Fig. 1(c) shows yet another map which belongs to a third group element of $\pi_1(X)$. It should be clear that the various homotopy classes, or group elements, are distinguished by the number of times the image of S^1 winds around the hole. This number can be any positive or negative integer n, because the loops are directed. There are an infinite number of group elements in $\pi_1(X)$, labeled by the integer n.

Group multiplication is defined by taking a representative map from each of the two group elements in question (clearly it is irrelevant which maps are chosen) and joining their images in X into a single loop [Figs. 1(d), (e), and (f)]. The latter generally determines a new group element with a new winding number n. But if one of the original group elements is the identity then the winding number does not change [Fig. 1(d)]. Moreover, because the loops are directed, two loops with opposite direction can annihilate [Fig. 1(f)]. This shows how inverse group elements are defined.

Group manifolds are topological spaces, and one can calculate the first homotopy group $\pi_1(G)$ for any group G. Some examples are

$$\pi_1(U_1) = \pi_1(S^1) = Z\,, \quad U_1 \sim S^1 \tag{I6}$$

$$\pi_1(U_1 \times U_1) = \pi_1(U_1) \times \pi_1(U_1) = Z \times Z\,; \tag{I7}$$

$$\pi_1(SU_2) = \pi_1(S^3) = 0; \quad SU_2 \sim S^3 \tag{I8}$$

$$\pi_1(U_2) = \pi_1(S^3 \times S^1) = \pi_1(S^3) \times \pi_1(S^1) = Z; \quad U_2 \sim S^3 \times S^1\,. \tag{I9}$$

These examples are easy to understand. The first one is essentially the example in Fig. 1. Note that $\pi_1(S^3) = 0$ (the trivial group) because S^3 is simply connected. In examples (I7) and (I9) we see that when S^1 is mapped into a product space the result is a product group π_1. This is not difficult to visualize.

(6) The nth homotopy group is defined as follows.

Definition of $\pi_n(X)$: The nth homotopy group is the set of equivalence classes of maps of the unit sphere S^n in n dimensions into a topological space X. The north pole N of S^n gets mapped onto a chosen base point P in X, just as for π_1. Two maps are equivalent, or homotopic, if they are continuously deformable into one another. This determines the equivalence classes which are the group elements of $\pi_n(X)$. Group multiplication is defined by the joining of images in X, just as for π_1. Inverse group elements are defined by maps with the opposite "direction." Unfortunately, all of this is difficult to visualize for the higher homotopy groups. This is the reason why we went to some pains to explain the group π_1, where things are easily visualized. The higher homotopy groups can then be understood, to some extent, by analogy with π_1. Note the following results concerning π_n:

$$\pi_n(A \times B) = \pi_n(A) \times \pi_n(B)$$

$$\pi_n(S^m) = 0, \quad n < m$$

$$\pi_n(S^n) = Z$$

$$\pi_n(S^1) = 0, \quad n > 1$$

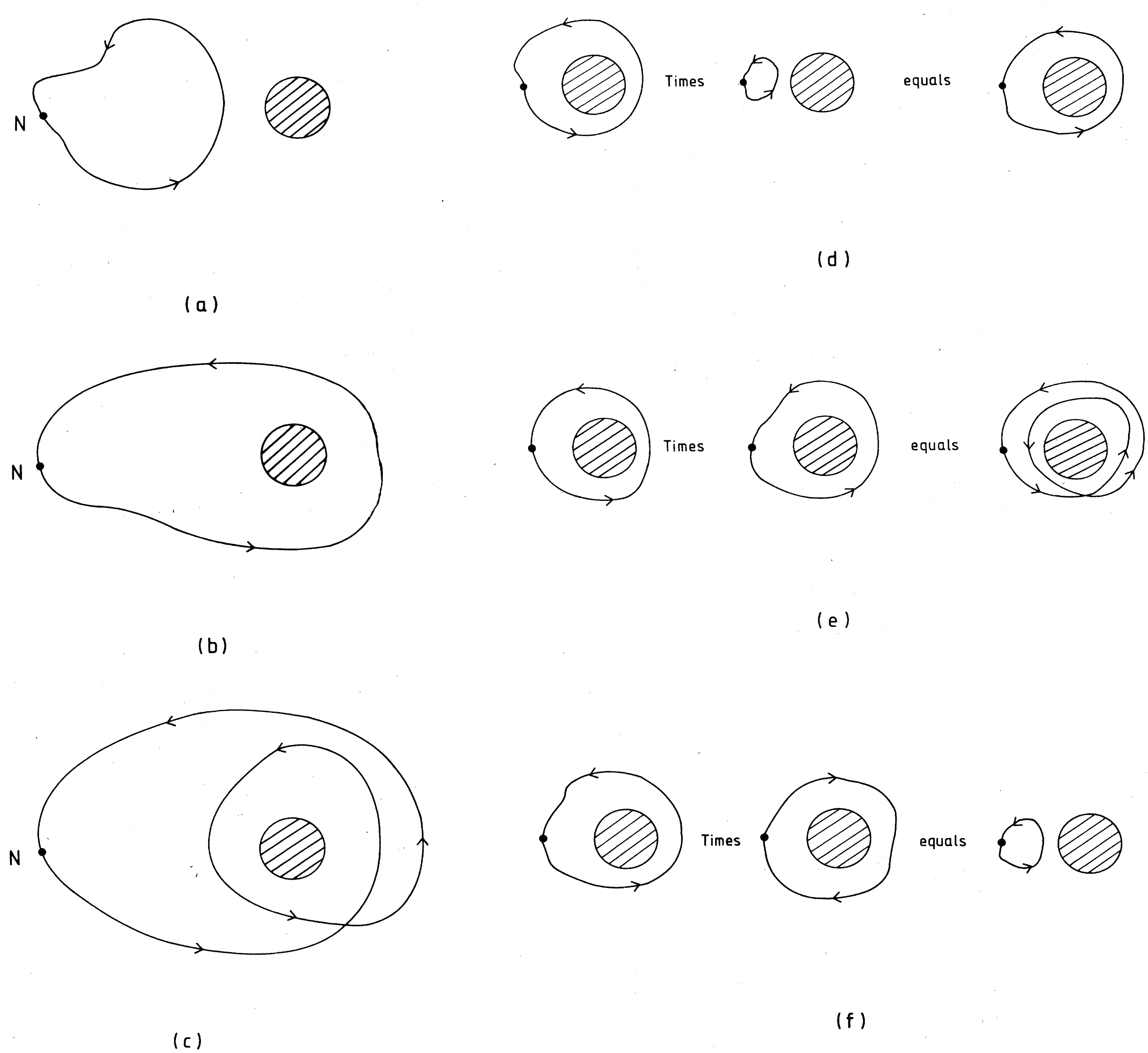

FIG. 1. (a)–(c) Representative mappings belonging to the elements of the first homotopy group $\pi_1(X)$ with winding number 0, 1, and 2, respectively. (d)–(f) Examples of group multiplication. Here the image space X is a plane with a hole cut out.

$$\pi_2(S^n) = 0, \quad n \neq 2$$

$$\pi_3(S^n) = 0, \quad n \neq 2, 3$$

$$\pi_3(S^2) = \pi_3(S^3) = Z\,.$$

The topological spaces X of interest in YM theory are, of course, non-Abelian group manifolds G. The base point N in G is always chosen to be the identity element of the group. It is generally a very nontrivial exercise to calculate the homotopy groups $\pi_n(G)$. We quote without proof some interesting results:

$$\pi_1(SO_n) = Z_2, \quad n > 2$$

$$\pi_1(U_n) = Z, \quad n \geq 1$$

$$\pi_1(SU_n) = 0, \quad n \geq 2$$

$$\pi_1(SU_n/Z_n) = Z_n, \quad n \geq 2$$

$$\pi_2(G) = 0, \quad G = \textit{any} \text{ Lie group}$$

$$\pi_3(G) = Z, \quad G = \textit{any simple} \text{ Lie group}$$

$$\pi_3(SO_4) = Z \times Z$$

$$\pi_3(SO_n) = Z, \quad n > 4$$

$$\pi_n(U_1) = 0, \quad n > 1$$

$$\pi_n(U_n) = \begin{matrix} 0 & n \text{ even} \\ Z & n \text{ odd} \end{matrix}$$

$$\pi_{2n}(U_n) = Z_{n!}$$

Here Z_n is the cyclic group of order n.

(7) The power of homotopy group methods in classical YM theory is beautifully illustrated by the following theorem (Tyupkin, Fateev, and Schwartz, 1975; Monastryskii and Perelomov, 1975):

Theorem: A YM theory with gauge group G and a Higgs multiplet which breaks the local gauge invariance, leaving an unbroken local subgroup H, can have monopole solutions only if the second homotopy group of the factor space G/H is nontrivial, i.e., if

$$\pi_2(G/H) \neq 0 . \tag{I10}$$

Moreover, if G is a simply connected group [like SU(n)] then this condition takes the very simple form

$$\pi_1(H) \neq 0 \tag{I11}$$

because for simply connected G

$$\pi_2(G/H) = \pi_1(H) . \tag{I12}$$

As the calculation of π_1 is generally quite easy, Eq. (I11) provides an extremely practical, as well as general, criterion for the *existence* of monopole solutions. If the subgroup H is simply connected, then $\pi_1(H) \equiv 0$ and no monopole solution is possible. If H is multiply connected, then monopole solutions *may* exist. This general topological argument provides no guarantee of existence, of course, nor does it provide insight into the functional nature of solutions. Essentially one is told where to look for solutions, and where not to look.

Eq. (I11) does provide information on the possible values of the monopole charge. The group elements of $\pi_1(H)$ are essentially labeled by these values. For example, in the 't Hooft–Polyakov monopole problem the gauge group $G = \mathrm{SU}(2)$ is simply connected, and $H = \mathrm{U}(1)$. From Eq. (I6) we see that $\pi_1(H) = Z$ in this case, and the integer n which labels the group elements of Z represents the allowed monopole charges. (Of course, only the $n = 1$ solution has been found to date. We know from the no-go theorems described in Sec. IV that solutions with $n > 1$ are necessarily very complicated, and it may be practically impossible to find them. But the topological analysis suggests that they exist.) Another example is provided by the SU(3) gauge theory with an octet of Higgs bosons. In this theory one can break the local SU(3) gauge invariance down to an unbroken local subgroup $H = U_2$ or $H = U_1 \times U_1$ (see Corrigan, Olive, Fairlie, and Nuyts, 1976; Sinha, 1976). Equations (I7) and (I9) show us what happens in these two cases. For $H = U_2$ there is a single monopole charge, while for $H = U_1 \times U_1$ there are two independent charges (the elements of $Z \times Z$ are labeled by two independent integers).

One can easily understand why condition (I10) above involves the second homotopy group. In general, the Higgs field ϕ_a is nonzero on the sphere at infinity, S^2_∞, to minimize the potential energy there. This implies a boundary condition

$$\phi_a(r = \infty) = f_a(\hat{r}), \quad f_a f_a = \text{const} , \tag{I13}$$

which clearly defines a mapping of S^2_∞ onto the group manifold of G. More precisely, this map is such that the sphere S^2_∞ gets mapped onto the manifold of cosets G/H, and this is why $\pi_2(G/H)$ and not $\pi_2(G)$ is the relevant homotopy group. This is the only difficult aspect of condition (I10) to grasp. (Besides the two references above, the reader might consult Coleman, 1975; Shankar, 1976; Sinha, 1976). The basic idea is as follows. By definition, the boundary condition (I13), written in matrix form, commutes with the elements of the unbroken subgroup H (recall our discussion in Sec. IV). Thus the group elements in H have no effect on the map $S^2_\infty \to G$. They can therefore be factored out from the group manifold G—this is precisely the meaning of the coset space G/H—without affecting the topological nature of the map. What remains is the map $S^2 \to G/H$.

We do not know of any really simple derivation of the identity (I12). One can convince oneself that this formula is plausible, however. It must hold for any subgroup H of G. For the trivial subgroups $H = e$ (the identity element) and $H = G$ the formula tells us that

$$\pi_2(G/e) = \pi_2(G) = \pi_1(e) = 0 ,$$

$$\pi_2(G/G) = \pi_2(e) = \pi_1(G) = 0 .$$

The second is obviously a correct statement; $\pi_1(G) \equiv 0$ because G is simply connected. The first statement is also correct because $\pi_2(G) = 0$ for any Lie group G.

(8) Homotopy group methods can help one to clarify and understand other problems in classical YM theory besides monopoles. We mention two more examples. (a) Equation (I4) implies the existence of instanton solutions of the Euclidean SU(2) theory which are labeled by an integral topological charge. A very modest input, namely that Euclidean solutions become pure gauge at infinity, is sufficient to make this statement. Now static solutions in Minkowski space can become pure gauge as $r \to \infty$, and one might ask: Can instantonlike solutions be found in this case? The answer is no. Maps of S^2_∞ into the SU(2) group are characterized by the homotopy group

$$\pi_2(SU_2) = \pi_2(S^3) = 0 .$$

Therefore, no instantonlike solution exists in three-space. (b) Do there exist instantonlike solutions in E^4 of theories with gauge groups G larger than SU(2)? Imbedded SU(2) solutions do exist, as we have explained in Appendix B. But one is interested in more general solutions, which are characteristic of the group G and not the group SU(2). Now we have seen that $\pi_3(G) = Z$ for any simple Lie group G. This extremely general result suggests that instanton solutions beyond the SU(2) ones exist for all such groups. [See Bitar and Sorba (1977) for a study of imbedded SU(2) instanton solutions in larger gauge theories. Bernard, Christ, Guth, and Weinberg (1977) have given a quite general discussion of instanton solutions of arbitrary gauge theories based on the Atiyah–Singer index theorem. An explicit instanton solution of the SU(3) gauge theory which is not an imbedded SU(2) solution has been found by Bais and Weldon (1978).]

REFERENCES

Abbott, L. F., and T. Eguchi, 1977, Phys. Lett. B **72**, 215.
Abramovitz, M., and I. Stegun, 1970, *Handbook of Mathematical Functions* (Dover, New York).
Actor, A., 1978a, Lett. Math. Phys. **2**, 275.
Actor, A., 1978b, Ann. Phys. (N.Y.) (to be published).
Actor, A., 1978c, 1979, Dortmund preprints.
Ademollo, M., E. Napolitano, and S. Sciuto, Nucl. Phys. B 134, 477.
Adler, S. L., 1969, Phys. Rev. **177**, 2426.
Ansourian, M. M., and F. R. Ore, Jr., 1977, Phys. Rev. D 16, 2662.
Arafune, J., P. G. O. Freund, and C. J. Goebel, 1975, J. Math. Phys. **16**, 433.
Arik, M., and P. Williams, 1978, Westfield College preprint.
Atiyah, M. F., N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, 1978, Phys. Lett. A **65**, 185.
Atiyah, M., N. J. Hitchin, and I. M. Singer, 1977, Proc. Natl. Acad. Sci. USA **74**, 2662.
Bais, F. A., 1976, Phys. Lett. B **64**, 465
Bais, F. A., and J. Primack, 1976, Phys. Rev. D **13**, 819.
Bais, F. A., and H. A. Weldon, 1978, Phys. Rev. D **18**, 561.
Ball, J., and A. Sen, 1977, Phys. Rev. D **15**, 2405.
Belavin, A. A., A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, 1975, Phys. Lett. B **59**, 85.
Bell, J. S., and R. Jackiw, 1972, Nuovo Cimento A **60**, 47.
Bernard, C. W., N. H. Christ, A. H. Guth, and E. Weinberg, 1977, Phys. Rev. D **16**, 2967.
Bender, C. M., T. Eguchi, and H. Pagels, 1978, Phys. Rev. D **17**, 1086.
Bernreuther, W., 1977, Phys. Rev. D **16**, 3609.
Bitar, K. M., and P. Sorba, 1977, Phys. Rev. D **16**, 431.
Bitar, K. M., and S.-J. Chang, 1978, Phys. Rev. D **17**, 486.
Bogomol'ny, E. B., 1976, Sov. J. Nucl. Phys. **24**, 449.
Bogomol'ny, E. B., and M. S. Marinov, 1976, Sov. J. Nucl. Phys. **23**, 355.
Boya, L. J., J. F. Cariñena, and J. Mateos, 1978, Fortschr. Phys. **26**, 175.
Brown, L. S., R. D. Carlitz, and C. Lee, 1977, Phys. Rev. D **16**, 417.
Bruce, D. J., 1978, Nucl. Phys. B **142**, 253.
Caldi, D. G., 1977, Phys. Rev. Lett. **39**, 121.
Callen, C., R. Dashen, and J. Gross, 1976, Phys. Lett. B 63, 334.
Callen, C., R. Dashen, and J. Gross, 1977, Phys. Lett. B 66, 375.
Callen, C., R. Dashen, and J. Gross, 1978a, Phys. Rev. D 17, 2717.
Callen, C., R. Dashen, and J. Gross, 1978b, Phys. Lett. B 78, 307.
Calvo, M., 1977a, Phys. Rev. D **15**, 1733.
Calvo, M., 1977b, Phys. Rev. D **16**, 1061.
Carlitz, R. D., 1978, Phys. Rev. D **17**, 3225.
Carlitz, R. D., and C. Lee, 1978, Phys. Rev. D **17**, 3238.
Carrigan, R. A., 1977, Fermilab preprint.
Castell, L., 1972, Phys. Rev. D **6**, 536.
Cervero, J., L. Jacobs, and C. R. Nohl, 1977, Phys. Lett. B **69**, 351.
Chiu, C., R. Kaul, and E. Takasugi, 1978, Phys. Lett. B **76**, 615.
Christ, N. H., A. H. Guth, and E. J. Weinberg, 1976, Nucl. Phys. B **114**, 61.
Christ, N. H., E. J. Weinberg, and N. K. Stanton, 1978, Phys. Rev. D **18**, 2013.
Coleman, S., 1975, Erice Lectures.
Coleman, S., 1977a, Commun. Math. Phys. **55**, 113.
Coleman, S., 1977b, Phys. Lett. B **70**, 59.
Coleman, S., S. Parke, A. Neveu, and C. M. Sommerfield, 1977, Phys. Rev. D **15**, 544.
Corrigan, E., and D. Fairlie, 1977, Phys. Lett. B **67**, 69.
Corrigan, E., D. Fairlie, S. Templeton, and P. Goddard, 1978 1978, Nucl. Phys. B **140**, 31
Corrigan, E., D. B. Fairlie, R. G. Yates and P. Goddard, 1978, Phys. Lett. B **72**, 354.
Corrigan, E., D. Olive, D. Fairlie, and J. Nuyts, 1976, Nucl. Phys. B **106**, 475.
Cremmer, E., F. Schaposnik, and J. Scherk, 1976, Phys. Lett. B **65**, 78.
Crewther, R. J., 1978, Schladming Lectures, CERN-TH 2522.
de Alfaro, V., S. Fubini, and G. Furlan, 1976, Phys. Lett. B **65**, 163.
de Alfaro, V., S. Fubini, and G. Furlan, 1977, Phys. Lett. B **72**, 203.
de Alfaro, V., S. Fubini, and G. Furlan, 1978, Phys. Lett. B **73**, 463.
Deser, S., 1975, Phys. Lett. B **64**, 463.
Deser, S., and F. Wilczek, 1976, Phys. Lett. B **65**, 391.
Dirac, P. A. M., 1931, Proc. R. Soc. Lond. A **133**, 60.
Eguchi, T., 1976, Phys. Rev. D **13**, 1561.
Eylon, Y., and E. Rabinovici, 1977, Phys. Rev. D **16**, 2660.
Frampton, P. H., 1976, Phys. Rev. D **14**, 528.
Gervais, J. L., B. Sakita, and S. Wadia, 1975, Phys. Lett. B **63**, 55.
Glashow, S., 1961, Nucl. Phys. **22**, 579.
Glimm, J., and A. Jaffe, 1978a, Phys. Lett. B **73**, 167.
Glimm, J., and A. Jaffe, 1978b, Phys. Rev. Lett. **40**, 277.
Goddard, P., and D. Olive, 1978, Rep. Prog. Phys. **41**, 1357.
Gradshteyn, I. S., and I. W. Ryzhik, 1965, *Table of Integrals, Series and Products* (Academic, New York).
Gross, D. J., and F. Wilczek, 1973, Phys. Rev. D **8**, 3633.
Grossman, B., 1977, Phys. Lett. A **61**, 86.
Gribov, V. N., 1977, Lecture at the 12th Winter School of the Leningrad Nuclear Physics Institute (SLAC-TRANS-176) [See also Nucl. Phys. B **139** (1978)].
Halpern, M. B., 1977a, Phys. Rev. D **16**, 1978.
Halpern, M. B., 1977b, Phys. Rev. D **16**, 3515.
Higgs, P. W., 1964, Phys. Lett. **12**, 132.
Hsu, J. P., 1976, Phys. Rev. Lett. **36**, 646.
Hsu, J. P., and E. Mac, 1977, J. Math. Phys. **18**, 100.
Hwa, R., 1977, Phys. Rev. D **15**, 2341.
Ikeda, M., and Y. Miyachi, 1962, Prog. Theor. Phys. **27**, 474.
Jackiw, R., 1977, Rev. Mod. Phys. **49**, 681.
Jackiw R., C. Nohl, and C. Rebbi, 1977, in *Particles and Fields*, edited by D. H. Boal and A. N. Kamal (Plenum, New York).
Jackiw, R., and C. Rebbi, 1976a, Phys. Rev. Lett. **37**, 172.
Jackiw, R., and C. Rebbi, 1976b, Phys. Rev. D **14**, 517.
Jackiw, R., and C. Rebbi, 1977a, Phys. Rev. D **15**, 1642.
Jackiw, R., and C. Rebbi, 1977b, Phys. Lett. B **67**, 189.
Jackiw, R., and C. Rebbi, 1977c, Phys. Rev. D **16**, 1052.
Jackiw, R., I. Muzinich, and C. Rebbi, 1978, Phys. Rev. D **17**, 1576.
Jacobs, L., and C. Rebbi, 1978, Phys. Rev. D **18**, 1137.
Johnson, T., O. McBryan, F. Zirilli, and J. Hubbard, 1978, Commun. Math. Phys. (to be published).
Ju, I., 1978, Phys. Rev. D **17**, 1637.
Julia, B., and A. Zee, 1975, Phys. Rev. D **11**, 2227.
Leznov, A. N., and M. V. Saveliev, 1978, Phys. Lett. B **76**, 108.
Lohe, M. A., 1978, Nucl. Phys. B **142**, 236.
Loos, H. G., 1965, Nucl. Phys. **72**, 677.
Lüscher, M., 1977, Phys. Lett. B **70**, 321.
Mack, G., and A. Salam, 1969, Ann. Phys. (NY) **53**, 174.
Magg, M., 1978, J. Math. Phys. **19**, 991.
Magruder, S. F., 1978, Phys. Rev. D **17**, 3257.
Mandelstam, S., 1975, Phys. Rep. **23C**, 245.
Manton, N. S., 1977, Nucl. Phys. B **126**, 525.
Manton, N. S., 1978a, Phys. Lett. B **76**, 111.
Manton, N. S., 1978b, Nucl. Phys. B **135**, 319.
Marciano, W., and H. Pagels, 1976, Phys. Rev. D **14**, 531.
Marciano, W., and H. Pagels, 1978, Phys. Rep. **36C**, 137.
Meuller, A. H., 1978, Phys. Rev. D **17**, 1605.

Michel, L., L. O'Raifeartaigh, and K. C. Wali, 1977, Phys. Rev. D **15**, 3641.
Monastryrskii, M. I., and A. M. Perelomov, 1975, JETP Lett. **21**, 43.
Montonen, C., and D. Olive, 1977, Phys. Lett. B **72**, 117.
Mottola, E., 1978, Phys. Lett. B **79**, 242.
Mukherjee, A., and P. Roy, 1978, Tata preprint.
Nielson, H. B., and P. Oleson, 1973, Nucl. Phys. B **61**, 45.
O'Raifeartaigh, L., 1977, Lett. al Nuovo Cimento **18**, 205.
Pagels, H., 1976, Phys. Rev. D **13**, 343.
Pagels, H., 1977, Phys. Lett. B **68**, 466.
Petiau, G., 1958, Suppl. Nuovo Cimento **9**, 542.
Politzer, H., 1973, Phys. Rev. Lett. **30**, 1346.
Polyakov, A. M., 1975a, Sov. Phys.-JETP **41**, 988.
Polyakov, A. M., 1975b, Phys. Lett. B **59**, 82.
Polyakov, A. M., 1977, Nucl. Phys. B **120**, 429.
Prasad, M. K., and C. M. Sommerfield, 1975, Phys. Rev. Lett. **35**, 760.
Protogenov, A. P., 1977, Phys. Lett. B **67**, 62.
Rebbi, C., 1978, Phys. Rev. D **17**, 483.
Roman, P., 1975, *Some Modern Mathematics for Physicists and Other Outsiders*, Vol. 1 (Pergamon, New York).
Rosen, G., 1972, J. Math. Phys. **13**, 595.
Roskies, R., 1977, Phys. Rev. D **15**, 1731.
Roth, M. W., 1977, Fermilab preprint.
Sakurai, J. J., 1960, Ann. Phys. (NY) **11**, 1.
Salem, A., 1968, *Proceedings of the 8th Nobel Symposium*, edited by N. Svartholm (Almqvist and Wiksells, Stockholm).
Schechter, B. M., 1977, Phys. Rev. D **16**, 3015.
Schwartz, A. S., 1977, Phys. Lett. B **67**, 172.
Schwinger, J., 1964, Rev. Mod. Phys. **36**, 609.
Sciuto, S., 1977, Phys. Lett. B **71**, 129.
Shankar, R., 1976, Phys. Rev. D **14**, 1107.
Shifman, M. A., A. I. Vainshtein, and V. I. Zakharov, 1979, Nucl. Phys. B **147**, 385, 448, 519.
Sinha, A., 1976, Phys. Rev. D **14**, 2016.
Steenrod, N., 1951, *The Topology of Fiber Bundles* (Princeton University, Princeton, New Jersey).
Stevens, A. S., 1977, Phys. Lett. B **67**, 172.
't Hooft, G., 1971, Nucl. Phys. B **35**, 167.
't Hooft, G., 1974, Nucl. Phys. B **79**, 276.
't Hooft, G., 1976a, Phys. Rev. Lett. **37**, 8.
't Hooft, G., 1976b, Phys. Rev. D **14**, 3432.
't Hooft, G., 1976c (unpublished).
Treat, R. P., 1967, Nuovo Cimento A **50**, 871.
Tyupkin, Yu. S., V. A. Fateev, and A. S. Schwartz, 1975, JETP-Lett. **21**, 42.
Utiyama, R., 1956, Phys. Rev. **101**, 1597.
Uy, Z. E. S., 1976, Nucl. Phys. B **110**. 389
Wada, H., 1978, University of Tokyo-Komaba, preprint.
Wadia, S., 1977, Phys. Rev. D **15**, 3615.
Wadia, S., and T. Yoneya, 1976, Phys. Lett. B **66**, 341.
Weder, R., 1977, Commun. Math. Phys. **57**, 113.
Weinberg, E. J., and A. H. Guth, 1976, Phys. Rev. D **14**, 1660.
Weinberg, S., 1967, Phys. Rev. Lett. **19**, 1264.
Wess, J., 1960, Nuovo Cimento **18**, 1086.
Wilczek, F., 1977, in *Quark Confinement and Field Theory*, edited by D. Stump and D. Weingarten (Wiley, New York).
Wilson, K. G., 1974, Phys. Rev. D **10**, 2445.
Witten, E., 1977, Phys. Rev. Lett. **38**, 121.
Wu, T. T., and C. N. Yang, 1968, in *Properties of Matter Under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York).
Wu, T. T., and C. N. Yang, 1975, Phys. Rev. D **12**, 3843.
Wu, T. T., and C. N. Yang, 1976, Phys. Rev. D **13**, 3233.
Yang, C. N., 1977, Phys. Rev. Lett. **38**, 1377.
Yang, C. N., and R. L. Mills, 1954, Phys. Rev. **96**, 191.
Yoneya, T., 1977, Phys. Rev. D **16**, 2567.