The lattice gauge theory approach to quantum chromodynamics*

John B. Kogut

Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801

The author reviews in a pedagogical fashion some of the recent developments in lattice quantum chromodynamics. This review emphasizes explicit examples and illustrations rather than general proofs and analyses. It begins with a discussion of the heavy-quark potential in continuum quantum chromodynamics. Asymptotic freedom and renormalization-group improved perturbation theory are discussed. A simple dielectric model of confinement is considered as an intuitive guide to the vacuum of non-Abelian gauge theories. Next, the Euclidean form of lattice gauge theory is introduced, and an assortment of calculational methods are reviewed. These include high-temperature expansions, duality, Monte Carlo computer simulations, and weak coupling expansions. A Λ-parameter calculation for asymptotically free-spin models is presented. The Hamiltonian formulation of lattice gauge theory is presented and is illustrated in the context of flux tube dynamics. Roughening transitions, Casimir forces, and the restoration of rotational symmetry are discussed. Mechanisms of confinement in lattice theories are illustrated in the two-dimensional electrodynamics of the planar model and the U(1) gauge theory in four dimensions. Generalized actions for SU(2) gauge theories and the relevance of monopoles and strings to crossover phenomena are considered. A brief discussion of the continuity of fields and topological charge in asymptotically free lattice models is presented. The final major topic of this review concerns lattice fermions. The species doubling problem and its relation to chiral symmetry are illustrated. Staggered Euclidean fermion methods are discussed in detail, with an emphasis on species counting, remnants of chiral symmetry, Block spin variables, and the axial anomaly. Numerical methods for including fermions in computer simulations are considered. Jacobi and Gauss-Siedel inversion methods to obtain the fermion propagator in a background gauge field are reviewed. Langevin, pseudofermion Monte Carlo methods, and random walk methods are discussed, as are proposals for including the fermion determinant in computer simulations. Finally, spontaneous chiral symmetry breaking in lattice models is discussed. The review ends with brief remarks about present research directions.

CONTENTS

I. Introduction and Overview 775
II. Non-Abelian Dynamics in Continuum Quantum Chromodynamics 777
  A. Asymptotic freedom and the heavy-quark potential 777
  B. A dielectric model of confinement 782
  C. Renormalization-group improved perturbation theory and the heavy-quark potential 783
III. Lattice Gauge Theory Basics 785
IV. Calculational Methods for the Lattice Gauge Theory 788
  A. Euclidean high-temperature (strong coupling) expansions 788
  B. Euclidean lattice duality 791
  C. Monte Carlo techniques for pure gauge theories. The string tension 792
    1. Metropolis algorithm 793
      a. Correlations between successive configurations 793
      b. $1/\sqrt{N}$ 794
      c. Finite-size effects 794
    2. Heat bath algorithm 794
  D. Λ parameters and asymptotic freedom 795
V. Hamiltonian Lattice Gauge Theory, Flux Tube Dynamics, and Continuum String Models 800
  A. The transfer matrix and the Hamiltonian limit 800
  B. Relativistic thin strings, delocalization, and Casimir forces 802
  C. Roughening and the restoration of spatial symmetries in lattice gauge theory 803
VI. Topological Excitations and Confinement in Lattice Theories 806
  A. Vortices and confinement in the electrodynamics of the planar model 806
  B. U(1) gauge theory in four dimensions 810
  C. Monopoles, crossover, and confinement in non-Abelian models in four dimensions 811
  D. Topology and continuity in asymptotically free lattice theories 814
VII. Lattice Fermions 816
  A. Free fields on a spatial lattice. Species doubling 817
  B. Fermions and bosons on Euclidean lattices 820
  C. Staggered Euclidean fermions 821
  D. Block derivatives and axial symmetries 823
  E. Staggered fermions and remnants of chiral symmetry 824
VIII. Numerical Methods for Lattice Fermions 825
  A. Jacobi and related methods of calculating the fermion propagator 825
  B. The Langevin equation and pseudofermion Monte Carlo 826
  C. The fermion propagator and random walks 827
  D. Calculating the fermion determinant 828
IX. Chiral Symmetry Breaking on the Lattice 829
X. Concluding Remarks 834
Acknowledgments 835
References 835

I. INTRODUCTION AND OVERVIEW

The last two years have seen considerable progress in lattice gauge theory (Wilson, 1974; Polyakov, 1975; Wegner, 1971). The special features of this formulation of gauge theory have been better understood technically and it has been used as a calculational device to understand some nonperturbative phenomena of quantum chromodynamics. We now have numerical evidence for confinement in SU(2) and SU(3) gauge theories (Creutz, 1980), chiral symmetry breaking in quantum chromodynamics (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Kogut *et al.*, 1982) and some rough, but encouraging, meson and baryon mass calculations (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten and Petch-

*This is an extended version of lecture series given at Les Houches and St. Andrews Summer Schools, 1982.

Copyright © 1983 The American Physical Society

er, 1981; Weingarten, 1982). There are good reasons to hope that in the near future precise calculations of the properties of quantum chromodynamics and related gauge theories will be accomplished, and that the strong interactions will be "understood."

The lectures on which this article is based were intended as a pedagogical review of the developments leading to these grandiose expectations. Not all the important topics or contributions will be covered—there are simply too many. Rather, a few will be singled out and will be presented in a rather detailed form intended for those readers relatively new to the subject. It is hoped that one can learn from this review. However, in order that we may concentrate on recent, interesting developments in quantum chromodynamics, I have omitted background material on lattices and statistical mechanics. There are reviews in the literature which might be consulted as introductory material to the present lecture series (Kogut, 1979; Kadanoff, 1977). An acquaintance with elementary facts, such as the correspondence between four-dimensional statistical mechanics language and field theory, the relevance of second-order phase transitions to the existence and properties of the continuum limits of lattice models, will be assumed here.

This review has the following structure: We begin with a review of the asymptotic freedom of continuum gauge theory. Asymptotic freedom is the special feature of gauge theories which underlies most of the modern advances in field theory and the relation of field theory to nature. The heavy-quark potential is calculated and the Callan-Symanzik function derived. The idea of "renormalization-group improved perturbation theory" is introduced in this context. A simple dielectric model of confinement is presented and special features of flux tube dynamics are discussed. Next, lattice gauge theory is introduced and its weak and strong coupling properties are reviewed. These include the emergence of ordinary Yang-Mills theory at weak coupling and confinement at strong coupling. Then we turn to a more detailed look at lattice theories and their calculated methods. Strong coupling character expansions are introduced and illustrated for the heavy-quark potential. Duality is illustrated for the three-dimensional Ising model, and the relation between interfaces and string models is discussed. Computer simulation methods and Monte Carlo algorithms for pure gauge (no fermions) theories are reviewed. The Metropolis and heat bath algorithms are introduced and discussed, and some critical remarks (and warnings!) about finite lattice calculations are made. Finally, the Λ parameters of asymptotically free models are introduced and the mass scale determination in lattice gauge theory calculations is discussed. Weak coupling expansions in lattice models are illustrated in the $SU(N)\times SU(N)$ spin models in two dimensions. Next we take a more detailed look at lattice string and flux tube dynamics. The transfer matrix is introduced and the Hamiltonian form of lattice gauge theory obtained. Flux tubes are discussed in this context, and the notions of roughening and the universal $1/R$ potential of the fluctuating string are obtained in the continuum and on the lattice. The restoration of rotational symmetry is illustrated in Hamiltonian lattice gauge theory as the weak coupling, continuum limit is approached. The next major topic concerns the role of topological excitations on confinement. A two-dimensional model with vortices is discussed, and the compact version of lattice electrodynamics is analyzed in terms of its topological excitations. Then SU(2) and related non-Abelian lattice models are analyzed from this point of view. The significance of monopole loops and strings in crossover and confinement mechanisms is discussed and illustrated in analysis and computer simulations. Topological charge is discussed in the context of asymptotically free spin models in order to see when the continuum idea of continuity is a good guide to lattice physics and vice versa. The last third of this set of lectures deals with lattice fermions. Free fields in $1+1$ dimensions are discussed first, the species doubling problem is illustrated, and its connection to continuous chiral symmetries is drawn. Different lattice fermion techniques are introduced and their species doubling problems and chiral symmetries discussed. The Euclidean "staggered" fermion technique is developed in detail, with its remnants of continuous and discrete chiral symmetries exposed explicitly. The idea of "block" fermion variables and actions is reviewed and the explicit lattice breaking of the axial flavor-singlet symmetries is noted. Numerical methods for studying lattice fermions and gauge fields are reviewed. The emphasis is on the calculation of the fermion propagator in a background gauge field configuration. Exact Jacobi, Gauss-Siedel, and conjugate-gradient iterative methods are reviewed. Then Monte Carlo, Langevin, and random walk methods of calculating the propagator approximately are given. Finally, some current ideas on including dynamical fermion loops (fermion polarization effects) in computer simulations, the calculation of the fermion determinant, are reviewed. The final topic concerns chiral symmetry breaking in quantum chromodynamics and related theories. Strong coupling calculations and Monte Carlo computer experiments are discussed. Some calculations of chiral symmetry breaking in a wider context—in quantum chromodynamics at finite temperature, in pure electrodynamics—are discussed as examples where lattice methods can be applied to a wide range of rich, nonperturbative phenomena. The review ends with a brief look at present research directions.

Although it is impossible to predict the new topics that lattice methods will grapple with in the near future, many workers in the field share the opinion that both conceptual and calculational improvements are needed to deal with fermions. As discussed in the text, all of the lattice fermion methods involve some compromise in basic principles. Left-right asymmetric theories cannot be treated by present techniques, so symmetry breaking in weak interaction theories cannot be studied by nonperturbative lattice methods. This is a pity, since symmetry breaking is so mysterious in these theories and is handled in such a contrived, artificial fashion in most models of grand unified gauge theories. In addition, even in cases where the

lattice fermion formulations are adequate, the computational methods are not very good. Computer simulations with fermions present new problems which have been met before in statistical mechanics and nuclear physics, but have never really been mastered. If the present enthusiasm for lattice quantum chromodynamics yields new computational methods, there will be considerable impact on other fields.

There is growing interest in using lattice methods to study supersymmetric theories, gravity, etc. These topics will not be studied explicitly in the text, because only preliminary impressions are at hand. Although there are many obstacles to such investigations—the lattice itself breaks the dearest symmetries in these models—there is so little known about these theories that any exploratory method should be applied to them.

Perhaps it is safe to say that lattice methods have taught us a great deal about the dynamics of pure gauge fields and that there is good reason to hope that they will yield an approximate solution of quantum chromodynamics. In the long run, however, one hopes that lattice methods will develop into a reliable tool for exploratory field theory investigations and will shed light on the next generation of problems in grand unification, cosmology, etc. This hope underlies my interest in the subject.

II. NON-ABELIAN DYNAMICS IN CONTINUUM QUANTUM CHROMODYNAMICS

A. Asymptotic freedom and the heavy-quark potential

To begin we consider the heavy-quark potential in perturbation theory (Susskind, 1976; Fischler, 1977). This potential has played a singularly important role in lattice gauge theory. Here we shall find that SU_3 gauge theory is asymptotically free (Politzer, 1973; Gross and Wilczek, 1973; 't Hooft, 1972) in a particularly transparent fashion.

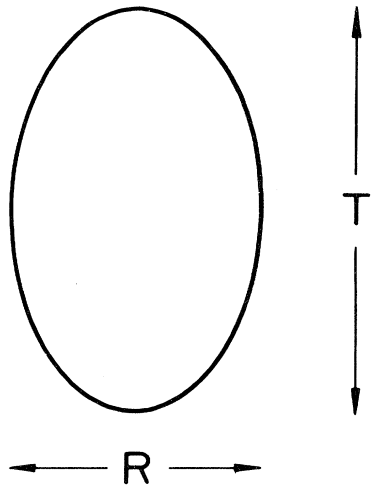

FIG 1. Closed world line of a heavy quark.

To obtain the heavy-quark potential consider a thought experiment. First, adiabatically separate a $Q\bar{Q}$ pair to a relative distance R. Then hold this configuration for a time $T\to\infty$. Finally, bring the quarks back together and let them annihilate. The world line of the quarks is shown in Fig. 1. The Euclidean amplitude for the process is the matrix element of the evolution operator $\exp(-HT)$ between the initial and final states,

$$\langle i\,|\,e^{-HT}\,|\,f\rangle \ . \tag{2.1}$$

Here $|\,i\rangle$ and $|\,f\rangle$ represent a $Q\bar{Q}$ pair a distance R apart and H is the Hamiltonian of the SU_3 gauge theory. The evolution operator damps out in time and does not oscillate, because we are formulating the problem in Euclidean space. Equation (2.1) can be expressed as a path integral,

$$\langle i\,|\,e^{-HT}\,|\,f\rangle=\int[DA_\mu^a][Dc_a][Dc_a^*]\exp\left[-S+ig\int A_\mu^a J_\mu^a d^4x\right]\Big/\int[DA_\mu^a][Dc_a][Dc_a^*]e^{-S}\ , \tag{2.2}$$

where S is the action of the SU_3 gauge theory, J_μ^a $(a=1,2,\ldots,8)$ is an external current describing the world lines of the heavy quarks, and the functional integrals run over the gauge field A_μ^a and Fadeev-Popov ghost fields c_a. For the path shown in Fig. 1, Eq. (2.2) simplifies to

$$\langle i\,|\,e^{-HT}\,|\,f\rangle=\int[DA_\mu^a](Dc_a][Dc_a^*]\exp\left[-S+ig\oint_C A_\mu^a\tfrac{1}{2}\lambda^a dx_\mu\right]\Big/\int[DA_\mu](Dc_a][Dc_a^*]e^{-S}\ , \tag{2.3}$$

where the contour C traces out the closed path of the figure. Since we shall meet such expressions repeatedly, we give the numerator and denominator of Eq. (2.3) separate names,

$$\langle i\,|\,e^{-HT}\,|\,f\rangle=Z(J)/Z(0)\ . \tag{2.4}$$

All this formalism will prove convenient when we evaluate the amplitude for the thought experiment in perturbation theory. Since $|\,i\rangle$ and $|\,f\rangle$ are identical and since the process is static, the left-hand side of Eq. (2.4) is simply

$$e^{-V(R)T}\langle i\,|\,f\rangle\ , \tag{2.5}$$

and $V(R)$ is the heavy-quark potential! So, taking the logarithm of Eq. (2.4) we have

$$V(R)=-\lim_{T\to\infty}\frac{1}{T}\ln\left\langle \mathrm{tr}P\exp\left[ig\oint A_\mu^a\tfrac{1}{2}\lambda^a dx_\mu\right]\right\rangle \times\langle\mathrm{tr}1\rangle^{-1}\ , \tag{2.6}$$

where P ("path ordered") reminds us that the order of the operators is important and the $\langle\mathrm{tr}1\rangle$ accounts for the

normalization of the initial and final states in Eq. (2.5)—there are three colors and factors of $1/\sqrt{3}$ are needed to normalize the initial and final states.

The quantity $Z(J)$, or

$$\left\langle \mathrm{tr}P \exp\left[ig \oint A_\mu^a \tfrac{1}{2}\lambda^a dx_\mu \right] \right\rangle \ , \tag{2.7}$$

is the "gauge-invariant loop correlation function"—the Wilson correlation function (Wilson, 1974; Polyakov, 1975; Wegner, 1971)—which serves as an order parameter for lattice gauge theories.

Now we turn to the perturbative calculation of Eq. (2.6). We will use the Feynman gauge in these calculations. The perturbation theory rules are familiar, with, perhaps, a few exceptions concerning the static quark lines (Susskind, 1976; Fischler, 1977). Free sources are described by the rules in Fig. 2. There i and j label the color indices of the quarks. The interaction between the static quarks and gluons are represented by the graphs in Fig. 3.

Now let's enumerate the graphs which contribute to Eq. (2.6). To second order we have the single-gluon exchange term (Fig. 4). The fourth-order graphs are shown in Fig. 5. Because of the noncommutativity of the vertices, we must keep the first and second contributions, Figs. 5(a) and 5(b), separate.

The second-order graph is

$$\frac{1}{2!} g^2 \mathrm{tr} T^\alpha T^\alpha \int dx_\mu dy_\mu D(x-y)^2 \ , \tag{2.8}$$

where $D(x-y)$ is a massless propagator and $T^\alpha = \lambda^\alpha/2$. The 1/2! takes care of the overcounting when we integrate over the entire source loop. Then we can call the integral in Eq. (2.8) I_a and write the result as

$$g^2 \mathrm{tr} T^\alpha T^\alpha I_a \ . \tag{2.9}$$

Similarly, the expressions for the other graphs in Fig. 5 are conveniently written by pulling out the group-theory factors and coupling constant and leaving an integral over the propagators behind,

$$\begin{aligned} &\frac{1}{2!} g^4 \mathrm{tr} T^\alpha T^\alpha T^\beta T^\beta I_b \ , \\ &\frac{1}{2!} g^4 \mathrm{tr} T^\alpha T^\beta T^\alpha T^\beta I_c \ , \\ &g^4 \mathrm{tr} T^\alpha T^\beta T^\gamma f^{\alpha\beta\gamma} I_d \ , \\ &g^4 \mathrm{tr} T^\alpha T^\alpha I_e \ , \\ &g^4 \mathrm{tr} T^\alpha T^\alpha I_f \ , \end{aligned} \tag{2.10}$$

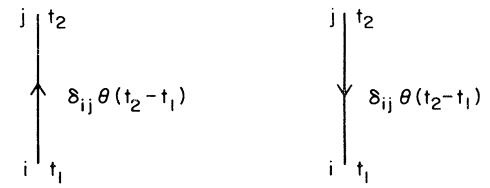

FIG. 2. Perturbation theory rules for a static quark and antiquark.

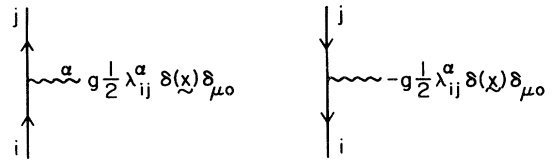

FIG. 3. Interaction vertices for a static quark (antiquark) and a gluon in SU(3) gauge theory.

where each term corresponds to a graph in Figs. 5(a)—5(e). The 1/2! are counting factors—when one was integrating over the sources the identical bosons would be overcounted.

First let's do an Abelian model calculation of such graphs, neglecting internal quark loops. The result will illustrate the connection between the loop correlation function and the heavy-quark potential. Now the only fourth-order effects are Figs. 5(a) and 5(b). There are no noncommuting T^α's, so these two graphs add trivially,

$$\frac{1}{2!} g^4 (I_b + I_c) \ . \tag{2.11}$$

The integrals $I_b + I_c$ differ only in the Θ functions. But obviously the Θ functions which organize the path ordering sum to unity, as illustrated in Fig. 6. So in the Abelian model,

$$Z = 1 + g^2 I_a + \frac{1}{2!} g^4 I_a^2 + \cdots \ , \tag{2.12}$$

and one can argue without difficulty that to all orders

$$Z = e^{g^2 I_a} \ . \tag{2.13}$$

Below we shall compute the integral over single photon exchange,

$$I_a \simeq T \frac{1}{4\pi R} \ . \tag{2.14}$$

Therefore, the heavy quark potential is

$$-\lim_{T\to\infty} \frac{1}{T} \ln Z = -\frac{g^2/4\pi}{|\mathbf{R}|} \ , \tag{2.15}$$

as expected.

Consider the vacuum polarization correction to the Abelian model shown in Fig. 5(e). This graph is a correction to the photon propagator. The momentum space propagators are shown in Fig. 7. The one-loop integral is

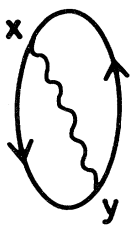

FIG. 4. Single-gluon exchange contribution to the heavy-quark world line.

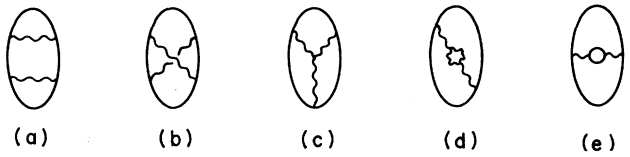

FIG. 5. Fourth-order contributions to the heavy-quark amplitude.

(Bjorken and Drell, 1964)

$$I_{\mu\nu}(q)=(-1)e^2\int\frac{d^4k}{(2\pi)^4}\times\mathrm{tr}\frac{\gamma^\mu(\not p+m)\gamma^\nu(\not p-\not q+m)}{(p^2+m^2)[[p-q)^2+m^2]}\ . \tag{2.16}$$

Set $m=0$. The loop does two things—it renormalizes the charge and it changes Coulomb's law,

$$-\frac{g_{\mu\nu}}{q^2}\to-\frac{g_{\mu\nu}}{q^2}\left[1+\frac{\alpha}{3\pi}\ln(q^2/M^2)\right],\quad \alpha=\frac{e^2}{4\pi}\ . \tag{2.17}$$

Fourier transforming to configuration space, we get

$$\frac{e^2}{4\pi R}\to\frac{e^2}{4\pi R}\left[1-\frac{2\alpha}{3\pi}\ln(R/a)\right]\ , \tag{2.18}$$

where $M=1/a$ is the ultraviolet cutoff parameter. To make sense of this result—the dependence of the coupling on a cutoff—we must appeal to the renormalizability of quantum electrodynamics and recall that order by order in perturbation theory the cutoff dependence in the potential can be absorbed into a redefinition of the coupling constant. The coupling constant then becomes an uncalculable parameter which in principle must be fitted to experiment. To do this we make a renormalization prescription: at $R=a$, $V(R)=\alpha/R$ should have a specified value. One still measures the potential relative to Coulomb's law,

$$\frac{\alpha}{R}\left[1-\frac{2\alpha}{3\pi}\ln(R/a)\right]\equiv\frac{\alpha(R)}{R}\ , \tag{2.19}$$

and the effective, or "running," coupling constant depends on the *physical* length scale R,

$$\alpha(R)=\alpha(a)-\frac{2}{3\pi}\alpha^2(a)\ln(R/a)\ , \tag{2.20}$$

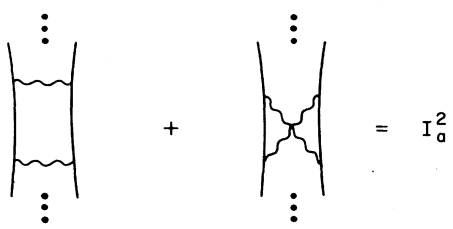

FIG. 6. Two-gluon exchange graphs.

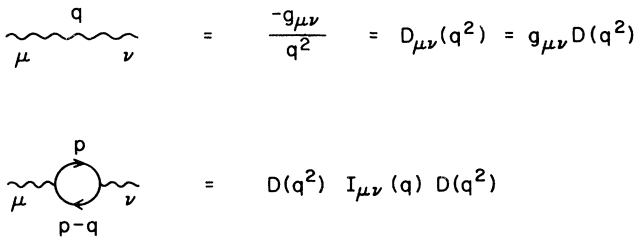

FIG. 7. Amplitudes for single-gluon exchange and vacuum polarization.

to this order in perturbation theory. Note that

$$R\frac{\partial}{\partial R}\alpha(R)=-\frac{2}{3\pi}\alpha^2(R) \tag{2.21}$$

to this order. The minus sign in Eq. (2.21) means that as R increases the effective coupling decreases. The physical reason for the sign is screening—the effective charges of the static heavy quarks are decreased by vacuum polarization (Fig. 8).

One writes with greater generality,

$$-R\frac{\partial}{\partial R}\alpha(R)=\beta[\alpha(R)]\ ,\qquad \beta(\alpha)=\frac{2}{3\pi}\alpha+\cdots\ , \tag{2.22}$$

as depicted in Fig. 9, where the arrow indicates the movement of $\alpha(R)$ as R increases. So, with α set on one-length scale, the theory has smaller effective couplings on larger scales—in this sense the theory is "infrared stable." Quantum fluctuations are relatively unimportant at larger length scales and classical physics works well. Of course, the same cannot be said of the ultraviolet!

Now let us return to the non-Abelian calculation. Consider the non-Abelian character of graphs in Figs. 5(a) and 5(b). We use the notation tr→tr/tr1 to accommodate the denominator in the expression for the loop correlation function. Since

$$T^\alpha T^\alpha=\tfrac{1}{4}\lambda^\alpha\lambda^\alpha=\tfrac{4}{3} \tag{2.23}$$

for SU(3), graph a can be written

$$\frac{1}{2!}g^4\mathrm{tr}T^\alpha T^\alpha T^\beta T^\beta I_b=\frac{1}{2!}(g^2\mathrm{tr}T^\alpha T^\alpha)^2 I_b\ . \tag{2.24}$$

This cannot be done for the graph in Fig. 5(b)—the non-Abelian character plays a role here. But the sum is as fol-

FIG. 8. Screening of charged impurities in ordinary matter.

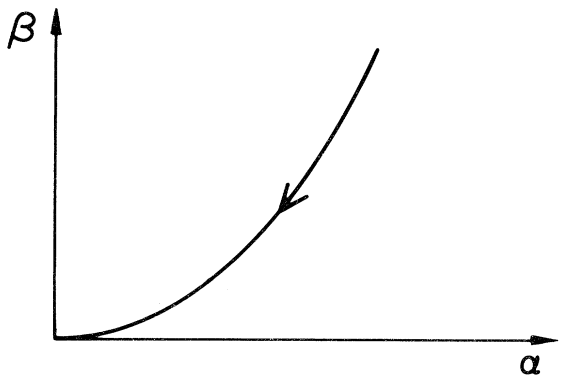

FIG. 9. The Callan-Symanzik function for electrodynamics.

lows:

$$(a)+(b)\equiv\frac{1}{2!}g^4\,\mathrm{tr}(T^\alpha T^\alpha T^\beta T^\beta)(I_b+I_c)$$
$$+\frac{1}{2!}g^4\,\mathrm{tr}T^\alpha[T^\beta,T^\alpha]T^\beta I_c \ . \tag{2.25}$$

If we use the fact that $I_b+I_c=I_a^2$, we can isolate the "Abelian-like" part of the calculation,

$$(a)+(b)=\frac{1}{2!}(g^2\,\mathrm{tr}T^\alpha T^\alpha)^2(I_a)^2$$
$$+\frac{1}{2!}g^4\,\mathrm{tr}T^\alpha[T^\beta,T^\alpha]T^\beta I_c \ . \tag{2.26}$$

But

$$\mathrm{tr}T^\alpha[T^\beta,T^\alpha]T^\beta=\tfrac{1}{2}\,\mathrm{tr}[T^\beta,T^\alpha][T^\beta,T^\alpha]$$
$$=\tfrac{1}{2}f^{\beta\alpha\gamma}f^{\beta\alpha\gamma'}\,\mathrm{tr}T^\gamma T^{\gamma'} \ . \tag{2.27}$$

Since $f^{\beta\alpha\gamma}f^{\beta\alpha\gamma'}=3\delta^{\gamma\gamma'}$ we have

$$(a)+(b)=\frac{1}{2!}(g^2\,\mathrm{tr}T^\alpha T^\alpha)^2I_a^2+\tfrac{3}{4}g^4\,\mathrm{tr}T^\alpha T^\alpha I_c \ . \tag{2.28}$$

Collecting everything, we have

$$Z=1+g^2\,\mathrm{tr}T^\alpha T^\alpha I_a+\frac{1}{2!}(g^2\,\mathrm{tr}T^\alpha T^\alpha I_a)^2$$
$$+\tfrac{3}{4}g^4\,\mathrm{tr}T^\alpha T^\alpha I_c+\tfrac{3}{2}g^4\,\mathrm{tr}T^\alpha T^\alpha I_d$$
$$+g^4\,\mathrm{tr}T^\alpha T^\alpha(I_e+I_f) \ , \tag{2.29}$$

where we have simplified the graph of Fig. 5(c) with the algebra

$$(c)=g^4\,\mathrm{tr}T^\alpha T^\beta T^\gamma f^{\alpha\beta\gamma}I_d$$
$$=\tfrac{1}{2}g^4\,\mathrm{tr}[T^\alpha,T^\beta]T^\gamma f^{\alpha\beta\gamma}I_d$$
$$=\tfrac{1}{2}g^4f^{\alpha\beta\zeta}f^{\alpha\beta\gamma}\,\mathrm{tr}T^\zeta T^\gamma I_d$$
$$=\tfrac{3}{2}g^4\,\mathrm{tr}T^\alpha T^\alpha I_d \ . \tag{2.30}$$

To obtain the potential to $O(g^4)$ we need $\ln Z$,

$$\ln Z=g^2\,\mathrm{tr}T^\alpha T^\alpha[I_a+\tfrac{3}{4}g^2I_c+\tfrac{3}{2}g^2I_d+g^2(I_e+I_f)] \ . \tag{2.31}$$

Now we will do the integrals. We need do them well enough to extract the term linear in T. Terms which do not grow this rapidly with T—end effects—will be discarded immediately.

Consider the integral in Fig. 4. The Euclidean integral is easy to estimate:

$$I_a=\tfrac{1}{2}\oint dx_\mu dy_\mu D(x-y)$$
$$\to\frac{1}{4\pi^2}\int_{-T/2}^{T/2}\frac{dt_1dt_2}{(t_1-t_2)^2+\mathbf{R}^2} \ , \tag{2.32}$$

where the $\frac{1}{2}$ was absorbed by the two terms shown in Fig. 10. The last two sketches in that figure give integrals which have no $\mathbf{R}$ dependence and are dropped. The remaining integral is simply

$$I_a\sim T\frac{1}{4\pi|\mathbf{R}|} \ . \tag{2.33}$$

Next we need I_c. There are three distinct types of graphs shown in Fig. 11, labeled (1), (2), and (3). Note that (1) has no $\mathbf{R}$ dependence so we can ignore it, and

$$(2)=\frac{4}{(4\pi^2)^2}\int dt_1dt_2dt_3dt_4$$
$$\times\frac{\Theta(t_2-t_3)\Theta(t_3-t_1)}{|t_1-t_2|^2[(t_3-t_4)^2+\mathbf{R}^2]} \ . \tag{2.34}$$

Integrate first over t_4,

$$\int_{-T/2}^{T/2}\frac{dt_4}{(t_4-t_3)^2+\mathbf{R}^2}=\frac{\pi}{R} \ . \tag{2.35}$$

Now change variables, $\tau=t_2-t_3$, $\sigma=t_1-t_3$, and $t_1-t_2=\sigma-\tau$. Therefore,

$$(2)=\frac{4}{(4\pi^2)^2}\frac{\pi}{R}\int dt_3d\tau\,d\sigma\frac{\Theta(\tau)\Theta(-\sigma)}{|\tau-\sigma|^2}$$
$$=\frac{4}{4\pi^2}\frac{T}{4\pi R}\int d\tau\,d\sigma\frac{\Theta(\tau)\Theta(-\sigma)}{|\tau-\sigma|^2} \ , \tag{2.36}$$

where $\int dt_3=T$. This integral has a logarithmic divergence when $|\tau-\sigma|\to0$. Cut it off by requiring $|\tau-\sigma|\geq a$, a temporal lattice spacing. It also has an infrared divergence when $|\tau-\sigma|\to\infty$, which is cut off by T, the total time of the process. To leading logarithmic accuracy,

$$(2)=\frac{1}{\pi^2}\frac{T}{4\pi R}\ln(T/a) \ . \tag{2.37}$$

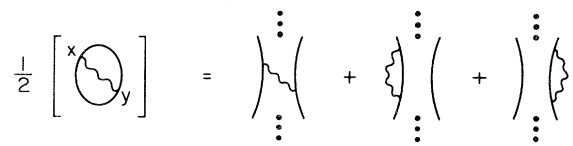

FIG. 10. Contributions to single-gluon exchange in a loop correlation function.

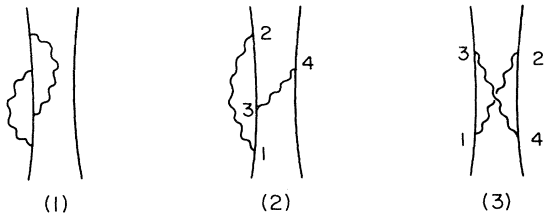

FIG. 11. Contributions to double-gluon exchange.

This looks like a catastrophe! We expected $V(R)T$, not $T\ln T$. The next graph, however, cures the problem:

$$(3)=-\frac{4}{4\pi^2}\int dt_1 dt_2 dt_3 dt_4 \times\frac{\Theta(t_3-t_1)\Theta(t_2-t_4)}{[(t_1-t_2)^2+\mathbf{R}^2][(t_3-t_4)^2+\mathbf{R}^2]} , \tag{2.38}$$

where the relative minus of (2) and (3) occurs because there is one coupling between the source and antisource in (2) and two such couplings in (3). The calculation is similar to (2), but there are no ultraviolet divergences—R plays the role of the cutoff, as is clear from the expression above. $\ln T$ occurs, however, for the same reason as in (2). So

$$(3)=-\frac{1}{\pi^2}\frac{T}{4\pi R}\ln(T/R) \tag{2.39}$$

and the sum is

$$(2)+(3)=\frac{1}{\pi^2}\frac{T}{4\pi R}\ln(R/a) . \tag{2.40}$$

The $\ln T$'s have cancelled and the answer has the correct structure. The ultraviolet divergence signals coupling constant renormalization, as we shall see.

The cancellation between (2) and (3) can be understood physically—the real source of radiation in Fig. 11 is a color singlet *dipole,* while each graph (2) and (3) records the radiation from an *unscreened* source or antisource. If the color source were not a singlet, the $\ln T$ would persist and mean that the system could radiate long-wavelength gluons copiously.

The infrared finiteness of the heavy-quark potential illustrates a more general result. Detailed analyses of spin models which have a global continuous symmetry group have shown that correlation functions which are *group singlets* are infrared finite to all orders in perturbation theory (David, 1980). The local gauge invariance of the loop correlation function is believed to render it infrared finite in a similar way.

Next consider I_d. Recall the triple-gluon vertex in the Feynman gauge. The sources are purely timelike ($\mu=\sigma=\nu=0$), which imples that the vertex vanishes identically. This graph, shown in Fig. 12, does not contribute to the heavy-quark potential!

Finally, consider I_e. This is very special to the non-Abelian theory. It is another modification of the gluon

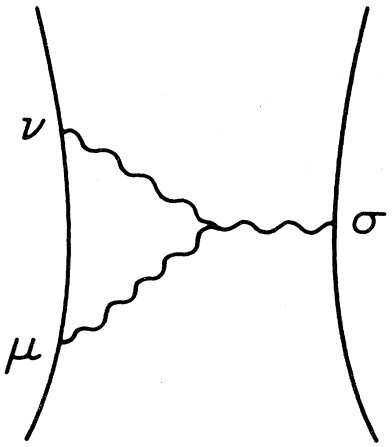

FIG. 12. The triple-gluon vertex.

propagator. Calculating the graph shown in Fig. 13, one finds (Hughes, 1979)

$$\frac{g^4}{16\pi^2}\frac{10}{3}\frac{3\,\mathrm{tr}T^\alpha T^\alpha}{4\pi R}\ln(R/a) . \tag{2.41}$$

The most important feature of this result is that it has the "wrong" sign—opposite to QED!

Collecting everything, we get

$$V(R)=-\frac{\mathrm{tr}T^\alpha T^\alpha}{4\pi R}\left[g^2+\frac{22g^4}{16\pi^2}\ln(R/a)\right] . \tag{2.42}$$

And we identify a running coupling constant,

$$g^2(R)=g^2(a)+\frac{22}{16\pi^2}g^4(a)\ln(R/a) , \tag{2.43}$$

and a Callan-Symanzik function (Callan, 1970; Symanzik, 1970),

$$\beta(g)=-R\frac{\partial}{\partial R}g^2(R)=-\frac{11}{16\pi^2}g^3(R)+\cdots . \tag{2.44}$$

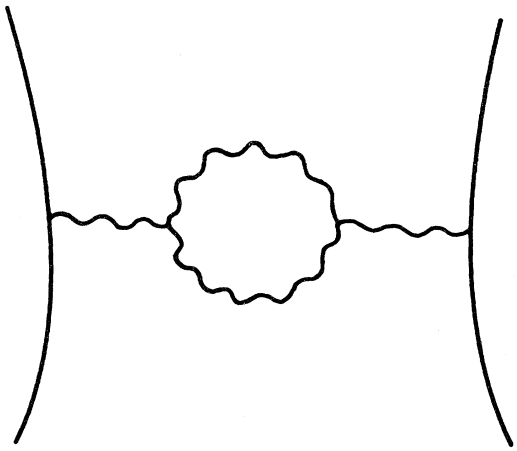

FIG. 13. Gluon polarization contribution to the quark loop amplitude.

The minus sign in Eq. (2.44), depicted in Fig. 14, implies that the theory becomes more strongly coupled at larger distances than classical estimates would have suggested. This suggests that confinement is a possibility in such theories.

Why, physically, is Yang-Mills asymptotically free? The special graph was gluon vacuum polarization (and its Fadeev-Popov ghost partner) (Fischler, 1977). It has the opposite sign to fermion (or spin-0 boson) vacuum polarization. Only spin-1 gluons have this property (Coleman and Gross, 1973). The only renormalizable, asymptotic freedom field theories in four dimensions are non-Abelian gauge theories. They probably play a very special role in nature.

B. A dielectric model of confinement

We have understood the sign of fermion vacuum polarization as screening. So, the sign of gluon vacuum polarization is due to "antiscreening." It is amusing and instructive to make an antiscreening model of confinement (Kogut and Susskind, 1974; Lee, 1981, Chap. 17).

Recall static electricity and magnetism in a *polarizable* material (Landau and Lifschitz, 1975). There is a dielectric constant ε and in ordinary electricity and magnetism, there is a strict inequality $\varepsilon \geq 1$ ($\varepsilon = 1$ for the vacuum). Static electric and magnetic effects are described by $\mathbf{D}$, the displacement vector, $\mathbf{E}$, the electric field, and $\mathbf{P}$, the polarization vector. Put static impurities of charge q into the system. They are the sources of $\mathbf{D}$, $\nabla \cdot \mathbf{D} = 4\pi q \delta(x)$. The medium responds to the impurity and screens it in part. Dipoles form in the material; their density is $-\mathbf{P}$. The net $\mathbf{E}$ field is $\mathbf{E} = \mathbf{D} - 4\pi\mathbf{P}$. The response of the medium may be *linear,* so $\mathbf{P} \propto \mathbf{E}$, and $\mathbf{E}$ is less than it otherwise would be. This is written $\mathbf{E} = \mathbf{D}/\varepsilon$, $\varepsilon \geq 1$, and ε is the "dielectric constant" characterizing the material.

To model quantum chromodynamics we might imagine $\varepsilon < 1$ and perhaps almost zero, say. If we place a *small* test charge $\delta q > 0$ into this medium, antiscreening charges come in as shown in Fig. 15. But the antiscreening charges repel one another, so a hole of vacuum ($\varepsilon = 1$) appears around the δq (Lee, 1981)! The size of the hole will depend on δq and the detailed properties of the vacuum.

Note that this is completely different from quantum

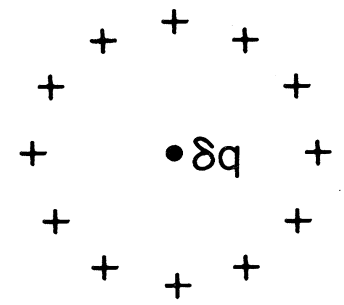

FIG. 15. Antiscreening in a dielectric model of confinement.

electrodynamics with $\varepsilon \geq 1$. Then the screening charges can move onto the positive impurity and the net cancellation of charge allows the hole to shrink to zero (Fig. 16). Thus in the electrodynamics of materials,

$$\mathbf{D} = \frac{\delta q}{4\pi r^2}\hat{r} \quad \text{and} \quad \mathbf{E} = \frac{\delta q/\varepsilon}{4\pi r^2}\hat{r} \ , \tag{2.45}$$

and there is no hole—just a Coulomb field of attenuated field strength.

Let's make the $\varepsilon > 1$ case more quantitative (Lee, 1981). Let us choose to consider $\delta q > 0$ again and suppose a spherical region of "vacuum" occurs around it, as shown in Fig. 17. Then, in the two regions,

inside	outside
$D = E = \frac{\delta q}{4\pi r^2}$	$D = \frac{\delta q}{4\pi r^2}$, $E = \frac{\delta q/\varepsilon}{4\pi r^2}$

Recall that the energy density in the material is $\frac{1}{2}\mathbf{D}\cdot\mathbf{E}$. Inside the "vacuum" region the energy density reduces to the vacuum answer; outside it is scaled down by $1/\varepsilon$. The electrical energy of the configuration relative to the vacuum is

$$W_{\text{el}} = \tfrac{1}{2}[(\delta q)^2/\varepsilon - (\delta q)^2] \int_R^\infty \frac{d^3r}{(4\pi)^2 r^4}$$
$$= \tfrac{1}{2}(\delta q)^2 \left[\frac{1}{\varepsilon} - 1\right] \frac{1}{4\pi R} \ , \tag{2.46}$$

where R is the radius of the "hole." There is also the energy change of the material—the appearance of a hole

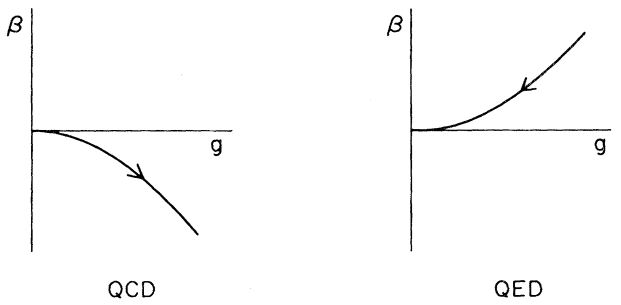

FIG. 14. The Callan-Symanzik functions for QCD (ultraviolet stable, infrared unstable) and QED.

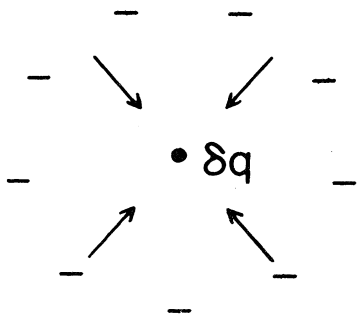

FIG. 16. Screening in an ordinary material.

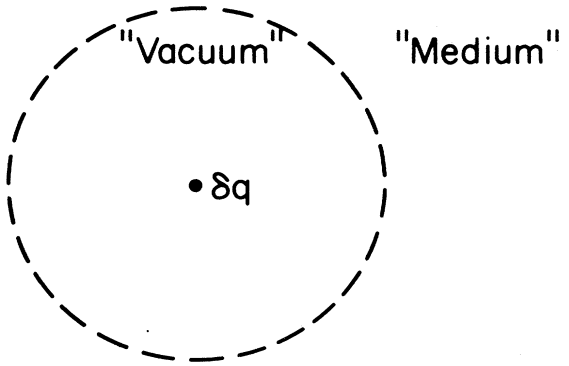

FIG. 17. An idealized model of antiscreening.

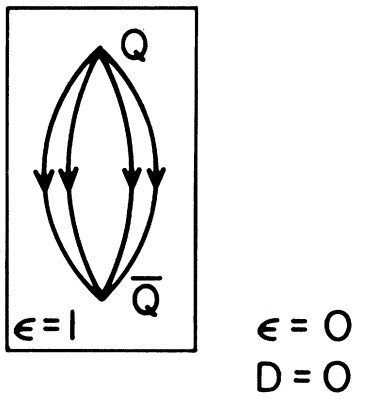

FIG. 19. A heavy meson in a confined quark model.

costs a constitutive energy proportional to the volume of the hole. The surface of the hole can also cost additional energy,

$$W_{\text{hole}}=\left[\frac{4\pi}{3}R^3\right]\rho+(4\pi R^2)\sigma+\cdots\ , \tag{2.47}$$

where ρ, σ, etc., are positive constants. The equilibrium configuration of electric and material energies is determined by minimizing the total energy $W_{\text{el}}+W_{\text{hole}}=W$. Look at W graphically in Fig. 18. It is easy to minimize W and find

$$W(R_{\text{eq}})=\frac{4}{3}\left[\frac{\delta q^2}{2\varepsilon}\right]^{3/4}(4\pi\rho)^{1/4}$$

if ε is very small, and the surface tension σ vanishes.

So, in this model the energy of the isolated quark approaches ∞ as $\varepsilon\to 0$. The ε dependence here is trivial to understand. $W_{\text{el}}\sim 1/\varepsilon R$ and $W_{\text{hole}}\sim R^3$, so $1/\varepsilon R_{\text{eq}}\sim R_{\text{eq}}^3$, where the two curves cross. Therefore, $R_{\text{eq}}\sim(1/\varepsilon)^{1/4}$. Since $W_{\text{hole}}\sim R^3$ and $W_{\text{el}}\sim 1/\varepsilon R$, we have $W(R_{\text{eq}})\sim(1/\varepsilon)^{3/4}$.

As $\varepsilon\to 0$, the physical sector of the theory will consist of only zero-charge states. Consider a $Q\bar{Q}$ pair. Since **D** lines flow from source to sink, we can make a *finite* energy field plus vacuum configuration by confining the flux in a tube (Fig. 19). Since the system is neutral, there is no divergent energy from $\int \mathbf{D}\cdot\mathbf{E}\,d^3r$. In the lowest energy configuration a hole forms again. As the $Q\bar{Q}$ are separated, the flux must flow between them. Because of the one-dimensional geometry of the tube (Fig. 20), the $Q\bar{Q}$ feel a linear confining potential.

Note that in the bound state the Q and $\bar{Q}$ experience finite forces of a conventional strength—they move quasi-freely, but they are confined. This is called "soft confinement"—the concept is compatible with the ideas of the naive quark model, the successes of the parton model, etc. When we turn to theories of quarks, we pass from Abelian charges to SU_3 color. The idea of "neutral" goes over to group "color" singlet. For SU_3 we can have color singlet states of finite energy, while color nonsinglets are confined. $\bar{Q}^a Q^a$ is a singlet, $3\times\bar{3}=1+8$, and $\varepsilon_{abc}Q^aQ^bQ^c$ is a singlet, $3\times3\times3=1+8_A+8_S+10$. These operators can generate meson and baryon states, respectively, in a naive quark model (Fig. 21).

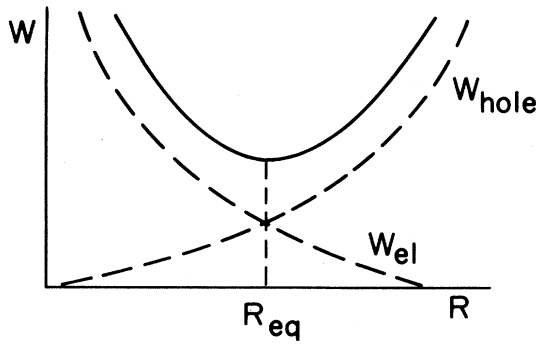

FIG. 18. Total energy for a static impurity in the dielectric model of confinement.

C. Renormalization-group improved perturbation theory and the heavy-quark potential

Now let's return to our perturbation theory calculation of the heavy-quark potential and improve it with the renormalization group (Gell-Mann and Low, 1954). The analysis that follows illustrates a very general methodology (Wilson and Kogut, 1974).

The heavy-quark potential as we calculated it has the functional dependence

$$V(R,a,g)\ . \tag{2.48}$$

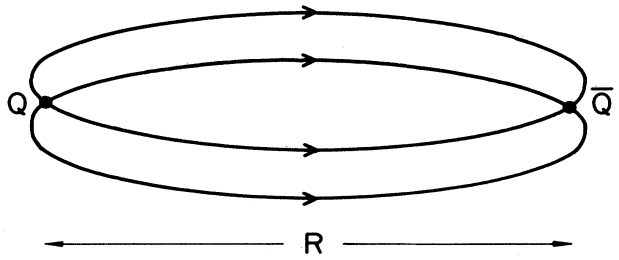

FIG. 20. A flux tube.

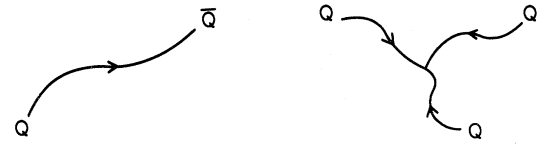

FIG. 21. A meson and baryon in an SU(3) string model.

But we can imagine fixing V at a certain physical length scale. Then, if we change the cutoff a, V must remain fixed, since it is, after all, a real, physical energy! This means that g must depend on a in a precise way. We require that

$$a\frac{d}{da}V(R,a,g)=0 \ . \tag{2.49}$$

Writing out the total derivative gives us

$$\left[a\frac{\partial}{\partial a}+a\frac{\partial g}{\partial a}\frac{\partial}{\partial g}\right]V(R,a,g)=0 \ . \tag{2.50}$$

But recall that $-a\,\partial g/\partial a=\beta(g)$ (Callan, 1970; Symanzik, 1970), so we have

$$\left[a\frac{\partial}{\partial a}-\beta(g)\frac{\partial}{\partial g}\right]V(R,a,g)=0 \ . \tag{2.51}$$

Also,

$$V=\frac{\mathrm{tr}T^{\alpha}T^{\alpha}}{4\pi R}\left[g^2+\frac{22g^4}{16\pi^2}\ln(R/a)\right] \ . \tag{2.52}$$

So substituting this perturbative result into the renormalization group equation (2.51), we get

$$-\frac{22g^4}{16\pi^2}-\beta(g)2g=0+\cdots \ , \tag{2.53}$$

where the ellipsis in Eqs. (2.53)—(2.55) represents terms of higher order. Therefore,

$$\beta(g)=-\frac{11}{16\pi^2}g^3+\cdots \ , \tag{2.54}$$

which determines the rate of change of g with the cutoff a,

$$\frac{\partial g}{\partial \ln a}=\frac{11}{16\pi^2}g^3+\cdots \ . \tag{2.55}$$

This equation can be integrated:

$$-\frac{1}{2}\left[\frac{1}{g^2}-\frac{1}{g_0^2}\right]=\frac{11}{16\pi^2}\ln(a/a_0) \ ,$$
$$g^2(a)=\frac{g_0^2}{1-\frac{22}{16\pi^2}g_0^2\ln(a/a_0)} \ . \tag{2.56}$$

We recognize the Taylor-series expansion of the numerator as our explicit perturbation theory calculation improved to arbitrary orders in g_0^2. The renormalization-group argument has summed up the "leading logs"—if we calculated to higher order, the highest powers of $\log(a/a_0)$ in each order of perturbation theory would generate a geometric series. The renormalization-group argument shows how such a tedious calculation must organize itself in a sensible, renormalizable theory.

$g^2(a)$ is the relevant measure of interactions at length-time scales of approximately a (measured in units of a_0). Let's illustrate this point by using the renormalization-group idea to calculate $V(R,a,g)$ to leading logarithmic accuracy in R. We had the renormalization-group equation,

$$\left[a\frac{\partial}{\partial a}-\beta(g)\frac{\partial}{\partial g}\right]V(r,a,g)=0 \ . \tag{2.51}$$

But, by dimensional analysis, $RV(R)$ depends on a only through R/a, so

$$a\frac{\partial}{\partial a}(RV)=-R\frac{\partial}{\partial R}(RV)$$
$$=-RV-R^2\frac{\partial}{\partial R}V \ . \tag{2.57}$$

Substituting into Eq. (2.51) gives

$$-RV-R^2\frac{\partial}{\partial R}V-\beta(g)\frac{\partial}{\partial g}RV=0 \ . \tag{2.58}$$

This is the result we wanted—it states that changing R and changing g are related (for given a).

It is convenient to use dimensionless parameters—scale R with a parameter λ and consider $V(\lambda R,g,a)$. Equation (2.58) then becomes (Fischler, 1977)

$$\left[\lambda\frac{\partial}{\partial\lambda}+\beta(g)\frac{\partial}{\partial g}\right]V(\lambda R,a,g)=-V(\lambda R,a,g) \ . \tag{2.59}$$

The solution for V is nicely expressed in terms of the running coupling constant. let

$$\lambda\frac{\partial}{\partial\lambda}g(\lambda)=-\beta[g(\lambda)] \ . \tag{2.60}$$

Then

$$V(\lambda R,g,a)=\frac{1}{\lambda}V[R,g(\lambda),a],\quad g(\lambda)\,|_{\lambda=1}=g \tag{2.61}$$

solves Eq. (2.59), as is easily checked. This equation shows how a change in the physical length scale can be absorbed in a change in the coupling constant.

An explicit one-loop expression for $g^2(\lambda)$ is

$$\lambda\frac{\partial}{\partial\lambda}g=\frac{11}{16\pi^2}g^3\rightarrow g^2(\lambda)=\frac{g^2}{1-\frac{22}{16\pi^2}g^2\ln\lambda} \ . \tag{2.62}$$

Now we can obtain the renormalization-group improved $V(R,g,a)$. If we take our renormalization condition to be

$$V(a,g,a)\equiv\frac{1}{4\pi a}g^2 \ , \tag{2.63}$$

and if we choose $\lambda=R/a$ and replace R by a in the renormalization-group equation (2.61), we have

$$V(R,g,a)=\frac{1}{R/a}V[a,g(R/a),a]$$
$$=\frac{1}{R/a}\frac{1}{4\pi a}g^2(R/a)=\frac{1}{4\pi R}g^2(R/a)$$
$$=\frac{1}{4\pi R}\frac{g^2}{1-\frac{22}{16\pi^2}g^2\ln(R/a)} \quad . \tag{2.64}$$

So, the running coupling constant enters the potential as the strength of Coulomb's law to *all* orders in perturbation theory. We have also recorded the one-loop result in the last line of Eq. (2.64).

We now see

(1) The physical significance of the running coupling constant.

(2) Asymptotic freedom as a physical effect, i.e., as $R/a\to 0$,

$$V(R,g,a)\to\frac{1}{4\pi R}\frac{16\pi^2/22}{\ln(R/a)} \quad ,$$

vanishing slightly faster than Coulomb's law.

III. LATTICE GAUGE THEORY BASICS

Now we will review SU(2) lattice gauge theory. I assume that everyone has some familiarity with this subject. Past reviews (Kogut, 1979; Kadanoff, 1977) contain introductory material.

The discussion will stress several basic properties of the theory. SU(N) lattice gauge theory's basic building blocks are SU(N) group elements which are assigned to the links of a hypercubic lattice—a discrete form of space-time. In the path integral formulation of the model these variables are freely integrated over. Since the group volumes are finite, such integrals are well defined and have simple invariance properties. This feature of the theory is crucial to its conceptual simplicity. For example, it allows gauge invariance to be stated precisely in the full fluctuating theory. In this sense it is simpler and conceptually clearer than the weak coupling perturbative formulation based on the Lie algebra of SU(N). The theory's formulation in terms of group elements underlies its well-defined strong coupling, confining features, as well.

The idea of local gauge invariance can be stated elegantly on the lattice. It serves several purposes. It dictates the form of the interactions in the theory much as local gauge invariance does in the classical continuum theory. Using local gauge invariance and requiring locality of the interactions, one can invent lattice actions whose classical continuum limits reproduce the Yang-Mills theory (Yang and Mills, 1954) and whose strong coupling limits confine quarks. The demonstration of these two points constitutes the core of this section.

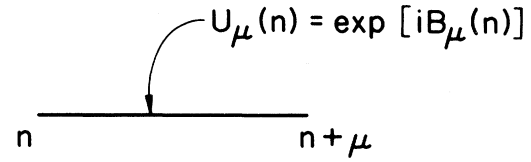

FIG. 22. The gauge degree of freedom on a lattice link.

Let us consider a four-dimensional hypercubic Euclidean lattice with spacing a. On each link we place an SU(2) matrix, as shown in Fig. 22,

$$U_\mu(n)=\exp[iB_\mu(n)] \quad , \tag{3.1}$$

where $B_\mu(n)=\frac{1}{2}ag\tau_i A^i_\mu(n)=\frac{1}{2}ag\boldsymbol{\tau}\cdot\mathbf{A}_\mu(n)$. Each link carries a direction, (n,μ) with $\mu=1$, 2, 3, or 4. If we denote a link in the backward direction, we associate with it $U^{-1}_\mu(n)$,

$$U_{-\mu}(n+\mu)\equiv U^{-1}_\mu(n) \quad . \tag{3.2}$$

The $[U_\mu(n)]_{ij}$ are SU(2) rotation matrices. One can imagine a color frame of reference at each site. Following Yang and Mills (Yang and Mills, 1954), we will require that the orientation of the frames in color SU(2) space be locally arbitrary. A rotation in color space can be done at each site with an SU(2) matrix,

$$G[\chi(n)]=\exp[-i\tfrac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{\chi}(n)] \quad . \tag{3.3}$$

The theory will have a local gauge invariance, and under a gauge transformation,

$$U_\mu(n)\to G(n)U_\mu(n)G^{-1}(n+\mu) \quad , \tag{3.4}$$

which we recognize as the simplest local generalization of a global SU(2) invariance—a rotation is applied to the SU(2) axes, and SU(2) matrices undergo a similarity transformation. Note the n and $n+\mu$ in Eq. (3.4) and the local character of the invariance group. Writing out the transformation law with indices, we get

$$[U_\mu(n)]_{ij}\to\sum_{kl}\{\exp[-i\tfrac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{\chi}(n)]\}_{ik}\{\exp[i\tfrac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{\chi}(n+\mu)]\}_{jl}[U_\mu(n)]_{kl} \quad . \tag{3.5}$$

Now we need an action which incorporates this local symmetry. A little thought indicates that the action should be built out of the products of U matrices taken around *closed* paths. These quantities are locally gauge invariant because the SU(2) color indices are then all contracted into local group singlets. We know that plaquettes, elementary squares, are the most local closed contours, so a candidate lattice action is (Wilson, 1974; Polyakov, 1975; Wegner, 1971)

$$S=-\frac{1}{2g^2}\sum_{n,\mu,\nu}\mathrm{tr}U_\mu(n)U_\nu(n+\mu)$$
$$\times U_{-\mu}(n+\mu+\nu)U_{-\nu}(n+\nu)+\mathrm{H.c.} \tag{3.6}$$

What is the physics of this model? Let's check that

(1) Its classical continuum limit ($a \to 0$) is ordinary Yang-Mills theory.
(2) Its strong coupling limit confines quarks.

With these results we will have a cutoff, finite, gauge-invariant formulation of gauge fields which can be used to study the theory for all a and all g.

To take the classical limit, we Taylor expand the slowly varying fields $B_\mu(n)$ appropriate to a long-wavelength approximation,

$$\begin{aligned} B_\nu(n+\mu) &\cong B_\nu(n)+a\partial_\mu B_\nu(n)+O(a^2) \\ B_{-\mu}(n+\mu+\nu) &= -B_\mu(n+\nu) \\ &\cong -[B_\mu(n)+a\partial_\nu B_\mu(n)]+O(a^2) \\ B_{-\nu}(n+\nu) &= -B_\nu(n) \ . \end{aligned} \tag{3.7}$$

Then

$$UUUU \cong e^{iB_\mu} e^{i(B_\nu + a\partial_\mu B_\nu)} e^{-i(B_\mu + a\partial_\nu B_\mu)} e^{iB_\nu} \ . \tag{3.8}$$

Now we use the operator identity,

$$e^x e^y = \exp(x+y+\tfrac{1}{2}[x,y]+\cdots) \ . \tag{3.9}$$

So,

$$\begin{aligned} UUUU &\cong \exp(i(B_\mu+B_\nu+a\partial_\mu B_\nu)-\tfrac{1}{2}[B_\mu,B_\nu]) \\ &\quad \times \exp(-i(B_\mu+B_\nu+a\partial_\nu B_\mu)-\tfrac{1}{2}[B_\mu,B_\nu]) \\ &\cong \exp(ia(\partial_\mu B_\nu-\partial_\nu B_\mu)-[B_\mu,B_\nu]) \ . \end{aligned} \tag{3.10}$$

But

$$\begin{aligned} B_\mu(n) &= \tfrac{1}{2}ag\tau_i A_\mu^i(n) \equiv agA_\mu(n) \ , \\ A_\mu(n) &\equiv \tfrac{1}{2}\tau\cdot\mathbf{A}(n) \ . \end{aligned} \tag{3.11}$$

So,

$$\begin{aligned} ia(\partial_\mu B_\nu-\partial_\nu B_\mu)-[B_\mu,B_\nu] &= ia^2 g\{\partial_\mu A_\nu-\partial_\nu A_\mu \\ &\quad +ig[A_\mu,A_\nu]\} \ . \end{aligned} \tag{3.12}$$

We identify the conventional Yang-Mills field strength $F_{\mu\nu}$ here. So

$$UUUU \cong e^{ia^2 gF_{\mu\nu}} \ , \tag{3.13}$$

with corrections in the exponent higher order in a^2. These corrections will not contribute in the classical continuum limit. For smooth, classical fields we look at excitations having long wavelengths compared to a, so

$$a^2 gF_{\mu\nu} \ll 1 \ , \tag{3.14}$$

and we can simplify the tr$UUUU$ further,

$$\begin{aligned} \mathrm{tr}UUUU &\cong \mathrm{tr}e^{ia^2 gF_{\mu\nu}} \\ &= \mathrm{tr}(1+ia^2 gF_{\mu\nu}-\tfrac{1}{2}a^4 g^2 F_{\mu\nu}^2+\cdots) \\ &= \mathrm{tr}1-\tfrac{1}{2}a^4 g^2 \mathrm{tr}F_{\mu\nu}^2+\cdots \ , \end{aligned} \tag{3.15}$$

where $\mathrm{tr}F_{\mu\nu}=0$, because tr(generator)$=0$. Finally,

$$\begin{aligned} \mathrm{tr}F_{\mu\nu}^2 &= \mathrm{tr}[\tfrac{1}{2}\tau_i(\partial_\mu A_\nu^i-\partial_\nu A_\mu^i-g\varepsilon_{kli}A_\mu^k A_\nu^l)] \\ &\quad \times [\tfrac{1}{2}\tau_{i'}(\partial_\mu A_\nu^{i'}-\partial_\nu A_\mu^{i'}-g\varepsilon_{k'l'i'}A_\mu^{k'}A_\nu^{l'})] \ , \end{aligned} \tag{3.16}$$

where we used $[\tau_i,\tau_j]=2i\varepsilon_{ijk}\tau_k$. Next, we recall that $\mathrm{tr}\tau_i\tau_j=2\delta_{ij}$, so Eq. (3.16) becomes

$$\mathrm{tr}F_{\mu\nu}^2 = \tfrac{1}{2}(\partial_\mu A_\nu^k-\partial_\nu A_\mu^k-g\varepsilon_{kij}A_\mu^i A_\nu^j)^2 \ , \tag{3.17}$$

which is the square of the gauge-covariant field strength tensor. Now the action becomes

$$\begin{aligned} S &\cong \frac{1}{2g^2}\int \frac{d^4x}{a^4} a^4 g^2 \tfrac{1}{2}(\partial_\mu A_\nu^k-\partial_\nu A_\mu^k-g\varepsilon_{ijk}A_\mu^i A_\nu^j)^2 \\ &\quad +O(a^2) \ , \end{aligned} \tag{3.18}$$

where we replaced

$$\sum_{n,\mu\nu} \to \int \frac{d^4x}{a^4}\sum_{\mu\nu} \ , \tag{3.19}$$

and where a 2 has appeared from the Hermitian conjugate. Collecting everything, we find

$$S = \tfrac{1}{4}\int d^4x (F_{\mu\nu}^i)^2 \ , \tag{3.20}$$

where

$$F_{\mu\nu}^i = \partial_\mu A_\nu^i-\partial_\nu A_\mu^i-g\varepsilon^{ijk}A_\mu^j A_\nu^k \ , \tag{3.21}$$

which is the usual Euclidean action of classical Yang-Mills. This result evokes several remarks:

(1) The final result is Euclidean O(4) invariant! Where has the difference between the cubic invariance of the original action and the O(4) invariance of the classical continuum limit gone? Into the higher-order a terms! Those operations are classified as "irrelevant"—they do not affect the continuum limit (Wilson and Kogut, 1974). This underscores the fact that we have much arbitrariness in the construction of lattice actions—one must simply engineer them so that the correct continuum limit occurs.

(2) The final result involves $F_{\mu\nu}^i$, the standard gauge-covariant field strength tensor of Yang-Mills. The *local* invariance built into the lattice action guarantees this.

Now let us consider the opposite extreme of the theory—the strong coupling limit. Is it characterized by quark confinement? As we reviewed in the first section, the loop correlation function

$$\left\langle \mathrm{tr}P\exp\left[ig\oint A_\mu^a \tfrac{1}{2}\lambda^a dx^\mu\right]\right\rangle \ , \tag{3.22}$$

is directly related to the heavy-quark potential. On the lattice this operator becomes

$$\prod_c U_\mu(n) \ , \tag{3.23}$$

where c denotes a closed contour of links. We should calculate

$$\left\langle \prod_c U_\mu(n) \right\rangle = \int \prod_{n,\mu} [dU_\mu(n)] \prod_c U_\mu(n) e^{-S} \left[\int \prod_{n,\mu} [dU_\mu(n)] e^{-S} \right]^{-1} , \tag{3.24}$$

where we are denoting the integral over the SU(2) group on each link generically. In fact, it is the invariant Haar measure which has the properties (Wilson, 1974; Polyakov, 1975; Wegner, 1971),

$$\begin{aligned} &\int [dU] = 1 , \\ &\int [dU] f(U) = \int [dU] f(U_0 U) , \end{aligned} \tag{3.25}$$

where U_0 is an arbitrary element of the group and f is an arbitrary but sensible function.

To evaluate $\langle \prod_c U_\mu(n) \rangle$ at strong coupling, $\beta = g^{-2} \ll 1$, we can write

$$e^{-S} = \prod_b e^{-\beta \,\mathrm{tr} UUUU} , \tag{3.26}$$

where $\prod_p$ denotes a product over plaquettes and we can expand the exponential. This leads us to integrals of products of U matrices. To obtain the leading strong coupling behavior of $\langle \prod_c U_\mu(n) \rangle$ we need only two integrals,

$$\begin{aligned} &\int [dU] U_{ij} = 0 \\ &\text{and} \\ &\int [dU] U_{ij} U^\dagger_{kl} = \tfrac{1}{2} \delta_{il} \delta_{jk} . \end{aligned} \tag{3.27}$$

The first integral is zero, since the integrand has no group-invariant piece—U_{ij} transforms as $2 \times \bar{2}$. The second integral is also easy to understand. The product of Kronecker symbols is determined by group invariance, and the normalization factor $\frac{1}{2}$ follows from the unitary character of each U matrix.

Now the leading behavior of $\langle \prod_c U_\mu(n) \rangle$ is easy to find. We must expand the factor $\prod_p e^{-\beta \,\mathrm{tr} UUUU}$ on each plaquette until the minimal area enclosed in the contour C is "tiled" once! Consider a 2×2 contour, Fig. 23, for illustration. The inner plaquettes in the figure indicate the factors $\beta \,\mathrm{tr} UUUU$ obtained from the expansion of $\prod_p e^{-\beta \,\mathrm{tr} UUUU}$. Consider link 1. We have the integral

$$\int U_{ij} U^\dagger_{\beta\alpha} [dU] = \tfrac{1}{2} \delta_{i\alpha} \delta_{j\beta} .$$

Calculating it, we are left with integrals over the remaining links, as shown in Fig. 24. Continuing to the other integrals, we find, finally,

$$\left\langle \prod_c U_\mu(n) \right\rangle = \left[\frac{1}{g^2} \right]^{N_c} = \exp(-\ln g^2 \times \mathrm{Area}) , \tag{3.28}$$

where N_c is the number of plaquettes—its "area" measured in dimensionless units—in the minimal surface determined by the contour c. If we take a rectangular contour of width R and length T and use the relation between the potential and the correlation function discussed in Sec. II, we find

$$V(R) = \sigma |R| \tag{3.29}$$

and identify the string tension at strong coupling

$$\sigma = \ln g^2 + \cdots . \tag{3.30}$$

In a later section on calculational methods I will indicate how this calculation can be improved and pushed to higher orders.

In Sec. II we obtained Coulomb's law modified by logarithms in perturbation theory. In the language of this section that result reads

$$\left\langle \prod_c U_\mu(n) \right\rangle \sim e^{-\alpha T/4\pi R \ln R} , \tag{3.31}$$

where $\alpha = g^2/4\pi$. Another possible behavior for the loop integral correlation function is the "perimeter law,"

$$\left\langle \prod_c U_\mu(n) \right\rangle \sim e^{-2mT} . \tag{3.32}$$

Then

$$V(R) \sim 2m \tag{3.33}$$

for large R. Clearly the underlying physics of this possibility is short-range finite forces between the quarks. The quarks can propagate freely in this case.

The loop correlation function is the order parameter for pure gauge theories. The "area" law labels a disordered phase—clearly the U matrices are strongly disordered in the strong coupling calculation illustrated above. By comparison, the "perimeter" law labels an ordered phase. The loop correlation function will reappear regularly throughout this review.

What about other matrix elements? For example, it is natural to ask whether $\langle U_\mu(n) \rangle$ is an interesting quantity. However, if $\langle U_\mu(n) \rangle \neq 0$, the theory would be spontaneously breaking gauge invariance. Is this possible? Can local symmetries break down spontaneously? Only global symmetries which act on an infinite number of degrees of freedom in a thermodynamic limit can break down spontaneously in systems with local coupling among ordinary degrees of freedom. In fact, there is a theorem by S. Elitzur which states this result quite rigorously (Elitzur, 1975). Thus gauge-*variant* operators like $U_\mu(n)$ or

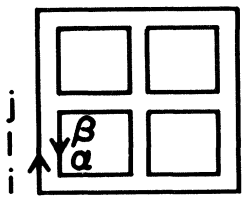

FIG. 23. Strong coupling graph for a 2×2 heavy-quark loop.

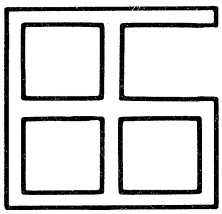

FIG. 24. Same as Fig. 23 after one-link variable integration.

$U_\mu(n)U_\nu(n)$ have vanishing matrix elements and are not interesting. We will not consider them further.

IV. CALCULATIONAL METHODS FOR THE LATTICE GAUGE THEORY

A. Euclidean high-temperature (strong coupling) expansions

The simplest high-temperature (β small) expansion for SU(2) gauge theory illustrated in the preceding section can be generalized and made into a systematic, useful tool for gauge theories. I will develop these expansions here and illustrate them for the planar loop correlation function of Ising lattice gauge theory in three dimensions.

The method begins by exploiting the group invariance of a theory's partition function to reorganize its traditional high-temperature series using character expansions. The ideas of connected and disconnected graphs can be stated precisely, and diagrammatic rules for connected matrix elements can be given. These results resemble the cluster expansions of classical statistical mechanics and Feynman diagram expansions of field theory (Wortis, 1974). However, since the products of U matrices do not satisfy a simple generalization of Wick's theorem—i.e., the expectation value of the product of several U matrices is not just the sum of products of expectations values of products of U matrices taken two at a time—the proofs of the general properties of these expansions are rather elaborate. For example, the proof of the extensive character of the expansion for the free energy requires more technology than the equivalent proof in weak coupling continuum perturbation theory. For that reason, we will not venture into proofs, but will illustrate their ingredients by example.

One should be aware of the vast literature on this subject. Rigorous proofs giving lower bounds on the radii of convergence of the high-temperature expansions of various matrix elements in various models exist (Osterwalder and Seiler, 1978). Such proofs usually rely on two crucial properties of the lattice theories: (1) upper bounds on the rate of growth of the number of graphs with the order of perturbation theory, and (2) the boundedness of the matrix elements of products of U matrices. Once domains of analyticity have been determined, the expansions can be used in practical calculations which search for phase transitions, crossover phenomena, etc. These less rigorous applications of the strong coupling calculations will be considered from time to time in the remainder of this article.

We consider a generic lattice gauge theory

$$S=\sum_p \beta d_\chi^{-1}\chi(U_p)=\sum_p S_p \ , \tag{4.1}$$

where χ is the real part of the character of U_p in a faithful representation of the gauge group G, $d_\chi=\chi(1)$ is the dimension of that representation. Since e^S is an invariant function of U_p, we can write it in a character (generalized Fourier) expansion (Balian *et al.*, 1975),

$$\begin{aligned} e^S&=\prod_p e^{S_p} \ , \\ e^{S_p}&=\exp[\beta d_\chi^{-1}\chi(U_p)]=\sum_\nu c_\nu(\beta)\chi_\nu(U_p) \ , \end{aligned} \tag{4.2}$$

where the sum ν extends over the set of nonequivalent irreducible unitary representations of the compact group G. Using the orthogonality relations of group characters, we discover the expansion coefficients to be

$$c_\nu(\beta)=\int [dU]\chi_\nu(U^{-1})e^{S_p(U)} \ . \tag{4.3}$$

This formalism is useful, because the simple properties of the characters $\chi_\nu(U_p)$ help us when computing matrix elements.

Some simple character expansions are

(1) Z_2. $S_p=\beta\sigma(p)$. Since $\sigma(p)=\pm 1$, the character expansion is obvious,

$$\exp(S_p)=\cosh\beta+\sinh\beta\sigma(p) \ . \tag{4.4}$$

(2) SU(2). Now $S_p=\frac{1}{2}\beta\,\mathrm{tr}U_p$. We can work out everything explicitly without using sophisticated group theory analysis. In the fundamental representation a link variable U can be parametrized as

$$U=u_0+i\mathbf{u}\cdot\boldsymbol{\sigma} \ , \quad |u_0|^2+|\mathbf{u}|^2=1 \tag{4.5}$$

or

$$U=\cos(\phi/2)+i\sin(\phi/2)\hat{n}\cdot\boldsymbol{\sigma} \ . \tag{4.6}$$

So,

$$S_p=\beta\cos(\phi/2) \ . \tag{4.7}$$

The character in the jth representation, $U_J=e^{i\mathbf{J}\cdot\hat{n}\phi}$,

$$\chi_j=\sum_{m=-j}^{+j} e^{im\phi}=\frac{\sin(2j+1)\phi/2}{\sin\phi/2} \ , \tag{4.8}$$

and the group measure depends on $\phi/2$, as

$$[dU]=d\phi(\sin\tfrac{1}{2}\phi)^2/\pi \ , \quad 0<\phi<2\pi \ . \tag{4.9}$$

So, the coefficients of the character expansion are

$$\begin{aligned} c_j(\beta)&=\frac{1}{\pi}\int_0^{2\pi} d\phi(\sin\tfrac{1}{2}\phi)^2\frac{\sin(2j+1)\phi/2}{\sin\phi/2} \\ &\quad\times e^{\beta\cos\phi/2} \\ &=2(2j+1)\beta^{-1}I_{2j+1}(\beta) \ , \end{aligned} \tag{4.10}$$

where we have identified the integral representation of a

modified Bessel function.

Since we usually want only ratios of Z's to compute matrix elements, it is convenient to write for Z_2

$$e^{S_p} \to 1 + \tanh\beta\sigma(p)$$

and for SU(2) (4.11)

$$e^{S_p} \to 1 + \sum_{j\neq 0} (2j+1)\frac{I_{2j+1}(\beta)}{I_1(\beta)}\chi_j[U(p)] .$$

Notice that for small β, $\tanh\beta \sim \beta$ and $I_{2j+1}(\beta)/I_1(\beta) \sim \beta^{2j}$, so these expressions give high-temperature expansions. The expansion coefficients are $\tanh\beta$ and $I_{2j+1}(\beta)$, respectively. Further, expanding $\tanh\beta$ and $I_{2j+1}(\beta)/I_1(\beta)$ in powers of β gives the old-fashioned high-temperature expansions. Sometimes using $\tanh\beta$ or $I_{2j+1}(\beta)$ as an expansion parameter gives a more useful, better behaved series (Samuel, 1980—character expansions are studied systematically here).

Let's look at the string tension in Z_2 gauge theory in three dimensions as an illustration (Münster, 1981). Consider the loop correlation function $\prod_c \sigma$ in Fig. 25. We compute several terms in its small-β expansion from the expression

$$\left\langle \prod_c \sigma \right\rangle = \sum e^{\beta S} \prod_c \sigma \Big/ \sum e^{\beta S}$$
$$= \sum \prod_p [1 + \tanh\beta\sigma(p)] \prod_c \sigma(l)$$
$$\times \left[\sum \prod_p [1 + \tanh\beta\sigma(p)] \right]^{-1} . \quad (4.12)$$

The leading term in the small-β expansion of Eq. (4.12) occurs when the plane is "tiled." It contributes

$$(\tanh\beta)^{RT} . \quad (4.13)$$

In general, we expect

$$\left\langle \prod_c \sigma \right\rangle \cong e^{-\alpha RT} , \quad (4.14)$$

where α is the string tension and where α is the excess free energy per area due to the presence of the quarks. So, to leading order,

$$\alpha = -\ln\tanh\beta . \quad (4.15)$$

Next we meet graphs in Eq. (4.12) which involve more than the minimal number of plaquettes. The simplest such graph is shown in Fig. 26. It involves four more plaquettes than the minimal surface. The "house" in the figure can occur on either side of the sheet (factor of 2) and it can occur in RT positions on the sheet. So this class of graphs contributes

$$2RT\tanh^4\beta(\tanh\beta)^{RT} . \quad (4.16)$$

With two houses on the sheet,

$$\frac{1}{2!}(2RT)(2RT-6)\tanh^8\beta(\tanh\beta)^{RT} , \quad (4.17)$$

where the -6 counts the "excluded volume"—given one "house" on the sheet, the second "house" is excluded from six positions. Summing the *leading* terms of this class of graphs of n nonoverlapping "houses" on the sheet, we get

$$(\tanh\beta)^{RT}\left[1 + 2RT\tanh^4\beta + \frac{1}{2!}(2RT\tanh^4\beta)^2 + \cdots\right] = e^{\ln\tanh\beta RT} e^{2\tanh^4\beta RT} . \quad (4.18)$$

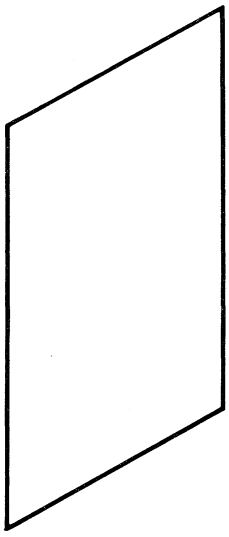

FIG. 25. A rectangular heavy-quark loop.

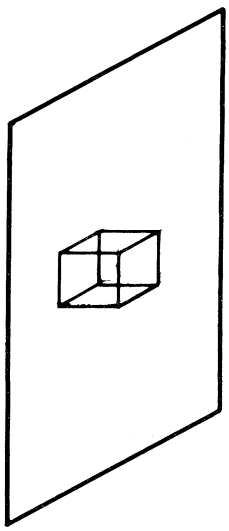

FIG. 26. Fourth-order contribution to the heavy-quark loop. The "house" lies on the sheet.

So, through this order,

$$\alpha = -\ln\tanh\beta - 2\tanh^4\beta \ . \tag{4.19}$$

We shall have to account for the excluded volumes [the factor -6 in Eq. (4.17)] in this set of graphs when we compute higher orders! In connected graph expansions these excluded volume effects are incorporated directly into the calculational rules (Münster, 1981).

Now consider $\tanh^6\beta$ graphs. We cannot neglect the denominator, which behaves as

$$1 + V\tanh^6\beta \tag{4.20}$$

from the graphs shown in Fig. 27. V is the volume of the four-dimensional lattice. In the numerator,

$$(\tanh^6\beta)(V - 2RT)(\tanh\beta)^{RT} , \tag{4.21}$$

where $-2RT$ is an excluded volume effect—the free cube avoids the minimal surface. But there are also rectangular houses on the sheet (Fig. 28) which contribute to the same order:

$$2[2RT\tanh^6\beta(\tanh\beta)^{RT}] \ . \tag{4.22}$$

So, in the numerator,

$$(\tanh^6\beta)(V + 2RT)(\tanh\beta)^{RT} \ . \tag{4.23}$$

But expanding up the denominator cancels the V dependence, and we are left with

$$2RT\tanh^6\beta(\tanh\beta)^{RT} \ . \tag{4.24}$$

This exponentiates as before, giving

$$\alpha = -\ln\tanh\beta - 2\tanh^4\beta - 2\tanh^6\beta \ . \tag{4.25}$$

Now for eight order. The graphs are shown schematically in Fig. 29. The first graph is the excluded volume effect which we ignored when we exponentiated Eq. (4.17)! The graphs of Fig. 29 contribute

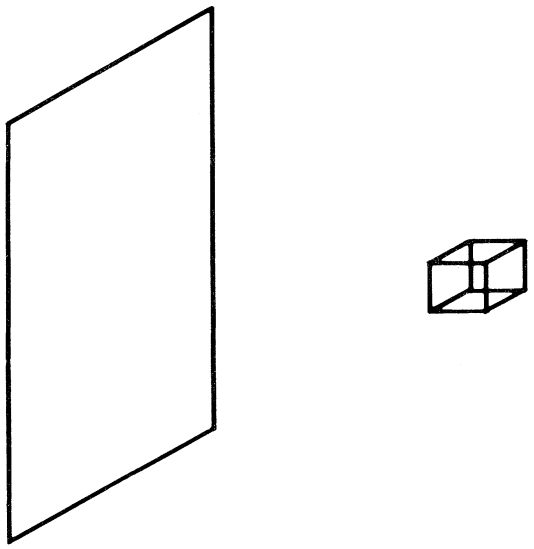

FIG. 27. Sixth-order disconnected graph.

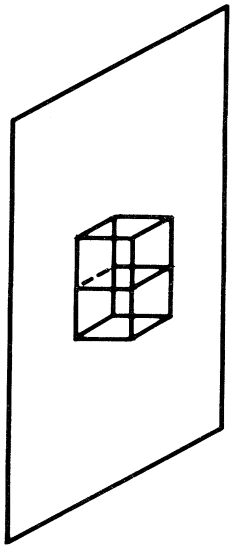

FIG. 28. Sixth-order connected graph.

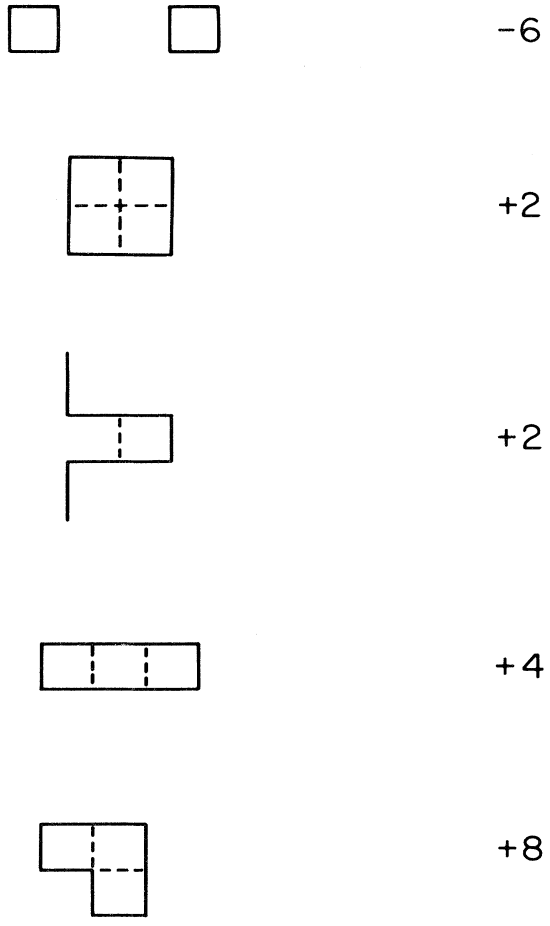

FIG. 29. Eighth-order graphs and counting factors. The first contribution is an excluded volume effect.

$10RT\tanh^8\beta(\tanh\beta)^{RT}$. So,

$$\alpha = -\ln\tanh\beta - 2\tanh^4\beta - 2\tanh^6\beta - 10\tanh^8\beta - \cdots . \tag{4.26}$$

Note that the fluctuations *decrease* the string tension. The physics behind this calculation will be discussed more thoroughly in later sections.

We compare the answer with SU_2 gauge theory in three dimensions,

$$\alpha = -\ln u - 2u^4 + 4u^6 - 6u^4 v - 10u^8 + \cdots , \tag{4.27}$$

where

$$u = c_{1/2} , \quad v = c_1 , \quad w = c_{3/2} \tag{4.28}$$

in the notation of Eq. (4.10).

B. Euclidean lattice duality

A standard but powerful approach to model field theories consists in developing mappings between strongly coupled and weakly coupled systems. A great deal of progress has been made in two-dimensional physics in this way. The mapping between the sine-Gordon and the massive Thirring model is a particularly interesting case (Coleman, 1975). One constructs fermion fields from bosons through a nonlocal map and finds that the two local field theories, one at weak coupling and the other at strong coupling, are copies of one another.

Similar constructions can be made in Abelian lattice systems in all dimensions and they are sometimes quite useful (Savit, 1980—a thorough review of lattice duality). The construction depends on the detailed geometry of the lattice, the number of space-time dimensions and the lattice action. The most famous example of such mappings is the self-duality of the two-dimensional Ising model—a mapping between its high- and low-temperature phases which allows one to locate the model's critical point (assuming its uniqueness) (Kramers and Wannier, 1941).

Here we shall review the fact, first discovered by Wegner (Wegner, 1971), that Ising lattice gauge theory in three dimensions is dual to the ordinary three-dimensional Ising model. This is a useful result, because much is known about the Ising model. We shall see that the loop correlation function maps into the interface of the Ising model. This result will give us additional insight into the string tension of the Ising gauge model and string dynamics in general.

Given a three-dimensional cubic lattice, the duality map will generate a "dual lattice" with the associations

$$\begin{aligned} &\text{site} \leftrightarrow \text{dual cube} \\ &\text{link} \leftrightarrow \text{dual plaquette} \\ &\text{plaquette} \leftrightarrow \text{dual link} \\ &\text{cube} \leftrightarrow \text{dual site} . \end{aligned} \tag{4.29}$$

We begin with the partition function for the gauge theory,

$$\begin{aligned} Z &= \sum_{\{s_l\}} \prod_p e^{\beta S_p} , \\ &= \sum_{\{s_l\}} \prod_p \cosh\beta(1 + S_p\tanh\beta) \\ &= \sum_{\{s_l\}} \prod_p \cosh\beta \sum_{K_p=\pm 1} (S_p\tanh\beta)^{(K_p+1)/2} , \end{aligned} \tag{4.30}$$

where $S_p = \sigma\sigma\sigma\sigma$ and where we introduced an Ising plaquette variable K_p in the last line. Now we calculate the $\sum_{\{s_l\}}$ sum. A given term in the K_p sum contributes only if S_p is raised to an even power. So, the sum of the $\frac{1}{2}K_p$ for the four plaquettes containing the link l must sum to zero mod2:

$$Z = \sum_{K_p} \prod_l \delta_2 \left[\frac{1}{2} \sum_{p : l\in\partial p} K_p \right] \prod_{p'} (\tanh\beta)^{(K_{p'}+1)/2} . \tag{4.31}$$

Make the dual association $K_p = K_{l^*}$ shown in Fig. 30. The constraint reads,

$$\frac{1}{2} \sum_{l^*} K_{l^*} = 0(\text{mod}2) . \tag{4.32}$$

So the "curl" of K_{l^*} should be zero mod2. This constraint is solved by introducing Ising spins on the dual sites:

$$K_{l^*} = -s_{i^*}s_{i^*+\mu^*} . \tag{4.33}$$

So,

$$\begin{aligned} Z &= \sum_{\{s_{i^*}\}} \prod_l \tanh\beta^{(-s_{i^*}s_{i^*+\mu^*}+1)/2} , \\ Z &= \sum_{\{s_i\}} \exp\left[-\frac{1}{2}\ln\tanh\beta \sum_{\langle ij\rangle} s_i s_j \right] , \end{aligned} \tag{4.34}$$

which is the Ising model with $\beta^* = -\frac{1}{2}\ln\tanh\beta$. Here $\langle ij\rangle$ labels bonds on a cubic lattice.

This is a duality map between the high-T phase of one model and the low-T phase of the other. It could easily

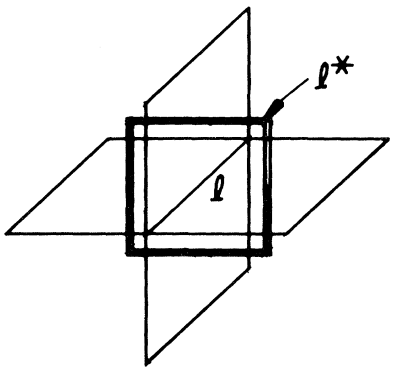

FIG. 30. Dual-link construction for the three-dimensional Ising model. The link l^* pierces the plaquette p of the original lattice.

have been discovered in their expansions (Kogut, 1979). In a low-T expansion of the Ising model one writes Z in an expansion in the number of flipped spins. A typical graph defines a surface of broken bonds as shown in Fig. 31. But this figure can be associated with a strong coupling graph of the Ising gauge model which is a closed surface of plaquettes, as indicated in the figure.

We can also see how matrix elements are mapped one to another under duality. Consider the Wilson loop (Wilson, 1974; Polyakov, 1975; Wegner, 1971),

$$W(J)=Z(J)/Z(0)=\left\langle \prod_c \sigma \right\rangle , \tag{4.35}$$

which we can write as

$$W(J)=\left\langle \prod_S s_p \right\rangle , \tag{4.36a}$$

where $s_p=\sigma\sigma\sigma\sigma$ and S is a surface bounded by the contour c. The logic and structure of the duality map are as before, leading to

$$W(J)=\left\langle \exp\left[-2\beta^* \sum_S s_i s_j\right]\right\rangle_* , \tag{4.36b}$$

which is the dual Ising model with a surface S of *anti*ferromagnetic bonds. The sites i and j in Eq. (4.36b) are nearest neighbors which label bonds piercing the surface S. So, the matrix element describes an *interface,* and

$$W(J)=e^{\Delta F} , \tag{4.37}$$

where ΔF is the excess free energy due to the interface (Balian *et al.,* 1975). The string tension of lattice gauge theory is then the surface tension of the interface. At low T the dominant spin configuration is shown in Fig. 32. As T increases, the boundary between the antiparallel spins wanders and thickens due to fluctuations. We will examine the physics of interfaces and sheets in more detail later.

Understanding ΔF is the key to understanding the phase transition in the three-dimensional Ising model. At low T the dominant spin configuration in the partition function sum has most spins up, say, and small, improbable blocks of spins down. The finite surface tension α suppresses these regions. As $T\to T_c$, the regions of false vacuum grow; and, finally, at T_c, α vanishes, the magnetization vanishes, and the system becomes disordered. In fact, as T approaches T_c from below, α is believed to vanish in a simple way. Recall the bulk correlation length ξ which controls the spin-spin correlations in the system,

$$\langle s_i s_j\rangle \sim \langle s_i\rangle^2[1-A\exp(-|i-j|/\xi)] . \tag{4.38}$$

For $T\lesssim T_c$, correlation length scaling (Kogut, 1979) maintains that the only relevant length in the system is ξ itself. Then dimensional analysis suggests that α vanishes as ξ^{-2}. But ξ diverges with a critical index ν in the scaling region, $\xi\sim|T-T_c|^{-\nu}$, so α should vanish as $\alpha\sim|T-T_c|^{2\nu}$. The four-dimensional, non-Abelian analog of this scaling law plays a central role in lattice quantum chromodynamics and will receive considerable attention in this review.

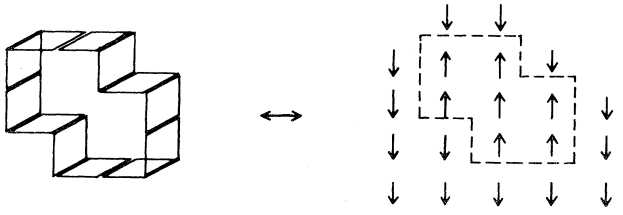

FIG. 31. A two-dimensional slice of a block of flipped spins in the three-dimensional Ising model and its corresponding graph in the Ising gauge model.

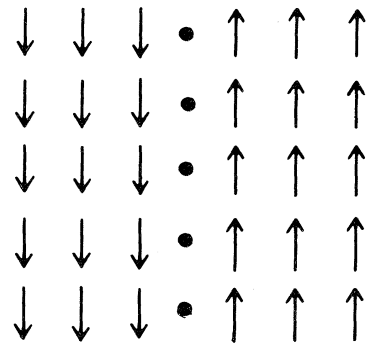

FIG. 32. A low-temperature planar interface in an Ising model.

C. Monte Carlo techniques for pure gauge theories. The string tension

Computer simulations and Monte Carlo evaluation of expectation values have been used in statistical mechanics and many-body problems since the availability of large computers (Binder, 1979). The idea of the method is to replace the computation of the expectation value of a quantity Θ,

$$\langle\Theta\rangle=\int[\pi\, dU]e^{-S}\Theta\Big/\int[\pi\, dU]e^{-S} , \tag{4.39}$$

which involves a very high dimensional integral, by an average over gauge configurations $\{U\}_i$,

$$\langle\Theta\rangle=\sum_i \Theta(\{U\}_i)\Big/\sum_i 1 , \tag{4.40}$$

where the configurations $\{U\}_i$ are distributed by the Boltzmann law e^{-S}. Note that each configuration $\{U\}_i$ means a set of U matrices on *all* the links of the lattice. The replacement of Eq. (4.39) by Eq. (4.40) is an example of the technique of "importance sampling"—if one attacked the integral $[\pi\, dU]$ directly, much effort would be wasted on configurations with large action which do not contribute significantly to the average $\langle\Theta\rangle$. The difficult aspect of the problem is now shifted to the generation of

an ensemble of configurations $\{U\}_i$ which are distributed according to the Boltzmann weight. Several computer algorithms exist for doing this. We will look at two of them: (1) Metropolis algorithm and (2) heat bath algorithm.

In both algorithms one begins with a configuration $\{U\}$. Then a single-link variable is varied and a new configuration $\{U'\}$ is generated. The new configuration replaces the old one if the rules of the algorithm are satisfied. After all the link variables have been sampled, one has "swept through" the lattice once. The aim of both procedures is to generate, after many sweeps, a gauge configuration $\{U\}$, at thermal equilibrium—it is a member of a Boltzmann distribution. $\Theta(\{U\}_1)$ can be calculated for this configuration and a contribution to Eq. (4.40) recorded. Next the algorithm can be applied many more times to $\{U\}_1$ until a new, statistically independent member of a Boltzmann distribution, $\{U\}_2$, is generated. $\Theta(\{U\}_2)$ is computed and recorded. Finally, this procedure is repeated many times, and an estimate of $\langle\Theta\rangle$ is obtained.

Now let us consider each procedure in detail.

1. Metropolis algorithm

Given a configuration $\{U\}$, we make a new one $\{U\}'$. For example, in an Ising system one would flip a spin at a given site. The procedure is then the following:

Compute the *change* ΔS in the action S. If it is less than zero, the new configuration is accepted. If it is larger than zero, the new configuration is accepted with the conditional probability $\exp(-\Delta S)$. [In practice, one picks a random number x, $0<x<1$. If $\exp(-\Delta S)>x$, the change is accepted, otherwise not.]

All this occurs at a given site. Then another site, or link, is chosen, and the procedure is repeated.

Let's review the argument that this procedure eventually brings the system into thermal equilibrium (Binder, 1979), i.e., the probability to find a configuration $\{U\}$ after N sweeps, is

$$P_N(U)\underset{N\to\infty}{\to} e^{-S(U)}\,. \tag{4.41}$$

To do this, consider the relationship between the Nth and the $(N+1)$th sweeps,

$$\begin{aligned}P_{N+1}(U)&=\sum_{\{U'\}}W(U'\to U)P_N(U')+\left[1-\sum_{\{U'\}}W(U\to U')\right]P_N(U)\\&=P_N(U)+\sum_{\{U'\}}[P_N(U')W(U'\to U)-P_N(U)W(U\to U')]\,,\end{aligned} \tag{4.42}$$

where $W(U\to U')$ is the probability for the transition $U\to U'$. A stationary probability distribution will then satisfy

$$P_N^S(U)W(U\to U')=P_N^S(U')W(U'\to U)\,, \tag{4.43}$$

which implies that

$$P_{N+1}^S(U)=P_N^S[(U)]\,. \tag{4.44}$$

Note that a stationary distribution is also stable—i.e., if

$$\frac{P_N(U_a)}{P_N(U_b)}>\frac{W(U_b\to U_a)}{W(U_a\to U_b)}\,, \tag{4.45}$$

then $P_{N+1}(U_a)<P_N(U_a)$ and $P_{N+1}(U_b)>P_N(U_b)$. Therefore, successive applications of the method bring us closer to satisfying detailed balance Eq. (4.43). Now Metropolis chooses

$$W(U\to U')=\begin{cases}1\,, & \text{if } S(U)>S(U')\\ e^{-[S(U')-S(U)]}\,, & \text{if } S(U)<S(U')\,.\end{cases} \tag{4.46}$$

So then

$$\frac{W(U_b\to U_a)}{W(U_a\to U_b)}=e^{-[S(U_a)-S(U_b)]}\,. \tag{4.47}$$

So, from Eq. (4.43),

$$P_N^S(U)=P^{eq}(U)=ce^{-S(U)} \tag{4.48}$$

as desired!

The final ingredient in the Metropolis scheme is the specification of the transition $U\to U'$. In the case of models like gauge theories with variables which belong to a continuous group, we clearly must invent a procedure which can cover the group space uniformly. In this way the Haar measure in the partition function will be respected. There are many ways to accomplish this (Wilson, 1980). For example, in the case of SU(3) gauge theory one might begin by generating a table of random SU(3) matrices and their Hermitian adjoints. (Or the table might be biased toward the identity—this is convenient if one is simulating the theory at large β.) Then, repeated applications of matrices chosen from this table to any arbitrary SU(3) matrix come arbitrarily close to any other member of the group (Wilson, 1980). Using this approach, we find the Metropolis step easy to execute—we specify a link, change its link variable by multiplication by a randomly chosen matrix from the table, and execute the standard algorithm, as discussed above. In practice, one also updates the table of random matrices regularly to avoid any unforeseen correlations or unlikely coincidences.

There are, however, many practical problems in applying this algorithm.

a. Correlations between successive configurations

It may take many sweeps before a statistically independent configuration is made. One can repeat the Metropolis algorithm at each site many times before moving on to the next site to proceed more efficiently. This problem, however, is particularly difficult near a second-order phase transition where ξ, the correlation length, is large. This is called *"critical slowing down"* (Binder, 1979). (It

is also an experimental problem!) The natural spatial and temporal correlation lengths diverge at T_c, so one needs large lattices, of course, and more troubling, many sweeps before a *new* member of a statistical ensemble is generated. Data analysis, connected correlation functions plotted against Monte Carlo sweep number, is necessary to monitor potential difficulties (Binder, 1979). Similar problems with the method occur in models with *metastable states*—the procedure may find a metastable rather than the equilibrium state and remain there for many sweeps if the tunneling probability between states is small.

b. $1/\sqrt{N}$

The procedure gives $\langle\Theta\rangle$ with statistical errors. The uncertainty in the mean values behaves as $c/\sqrt{N}$ by the central limit theorem. N is the number of *statistically independent* samples in the average, and c is a number which is essentially the standard deviation in the set of measurements, i.e., the susceptibility χ of the quantity Θ. This causes at least two problems:

(i) We need many statistically independent configurations, since $1/\sqrt{N}$ is a slowly falling function of N. The method may not be practical.

(ii) Near the continuum limit critical point, susceptibilities such as χ grow and eventually diverge.

c. Finite-size effects

This is not related specifically to the Monte Carlo procedure but affects all the numerical work on finite systems. Near T_c, $\xi\to\infty$ and the size of the system begins to play a role—thermodynamic functions are no longer characteristic of the bulk system. Typically, in the most straightforward Monte Carlo simulations, one works with parameters chosen so that $a\ll\xi\ll L$. The inequality $\xi\ll L$ can frequently be relaxed by using finite-size scaling theory (Hamer and Barber, 1980; Nightingale, 1976). Then it is necessary to know how to parametrize the finite-size effects, and extrapolations to $L\to\infty$ are needed to get physical results. The inequality $a\ll\xi$ can also be relaxed by modifying the lattice action to fit the continuum action more precisely (Symanzik, 1982).

2. Heat bath algorithm

Let us consider a given configuration $\{U\}$ and one link l in particular. All of the links which interact with l have fixed U matrices and provide a background. On l we choose U' from the gauge group G with a probability density proportional to the Boltzmann factor,

$$dP(U')=e^{-\beta S(U')}dU' . \tag{4.49}$$

Clearly, this algorithm satisfies detailed balance and so leads to thermal equilibrium. It is more effective than the Metropolis procedure, because if the Metropolis procedure were repeated many times on a given link, it would eventually distribute U' according to the Boltzmann weight, Eq. (4.49). The heat bath is harder to implement, however, because one needs the group measure explicitly. It has fewer problems with temporal correlations, because the new U matrix is not related to the old U matrix, as it was in the Metropolis method.

Let me illustrate the procedure for the Ising model. Consider site i and its interactions with its neighbors,

$$S(i)=\sigma(i)[\sigma(i+\hat{1})+\sigma(i-\hat{1})+\sigma(i+\hat{2})+\sigma(i-\hat{2})] . \tag{4.50}$$

The nearest neighbors are fixed. The probability for $\sigma(i)$ is

$$dP[\sigma(i)]=\frac{e^{-\beta S(i)}}{\sum_{\sigma(i)=\pm1}e^{-\beta S(i)}} . \tag{4.51}$$

Given a configuration, we easily compute this and tabulate $dP(+1)$ and $dP(-1)$. Now pick a random number between zero and unity. Call it x. If $x<dP(+1)$, assign spin $+1$ to the site. Otherwise, assign spin -1. This procedure, by construction, gives a Boltzmann distribution for $\sigma(i)$ in the heat bath provided by its nearest neighbors.

SU(2) (Creutz, 1980) and SU(3) (Pietarinen, 1981) heat baths exist. The SU(2) procedure takes advantage of the simplicity of Pauli algebra and is quite slick. An SU(3) algorithm, which uses SU(2) subgroups of SU(3) (Cabbibo and Marinari, 1982), is far superior to the SU(3) Metropolis program (Wilson, 1980).

Now consider Monte Carlo estimates of the string tension. First, the procedure: Measure Wilson loops $W(I,J)=\langle\prod_c U_\mu(n)\rangle$ for a rectangular, planar contour C, as shown in Fig. 33. In the limit $J\to\infty$, I fixed but large, the heavy quark potential is obtained:

$$V(I)=-\frac{1}{J}\ln\langle W(I,J)\rangle . \tag{4.52}$$

In principle this may *not* be a practical calculation, be-

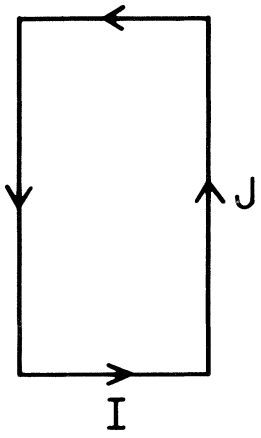

FIG. 33. A heavy-quark loop with rectangular dimensions $I\times J$.

cause $W(I,J,)$ has a factor $e^{-2m(I+J)}$ due to the self-energy of the heavy quark so W is infinitesimal for large loops! We are interested in a possible area term $e^{-\alpha IJ}$. Assume that $W(I,J)$ is dominated by just these terms for relatively *small* loops and consider the ratio of measurements (Creutz, 1980)

$$\frac{W(I,J)W(I-1,J-1)}{W(I,J-1)W(I-1,J)} . \tag{4.53}$$

Then if

$$W(I,J)=e^{-\alpha IJ-m(I+J)2} , \tag{4.54}$$

the ratio is

$$\rho^{-\alpha} \tag{4.55}$$

So, consider (Creutz, 1980)

$$\chi(I,J)=-\ln\left[\frac{W(I,J)W(I-1,J-1)}{W(I,J-1)W(I-1,J)}\right] . \tag{4.56}$$

Then, assuming Eq. (4.55),

$$\chi(I,J)=\alpha . \tag{4.57}$$

If $W(I,J)$ has more complicated behavior, such as $e^{-J/I}$ expected from the fluctuations of the surface, then

$$\chi(I,J)=\alpha+\mathscr{P} , \tag{4.58}$$

where $\mathscr{P}$ stands for power-law corrections. In practice, $\chi(I,J)$ has been measured for $I,J=1,2,3$ on lattices 6^4 for SU(3) (Pietarinen, 1981) and 10^4–16^4 for SU(2) (Stack, 1983). α has been extracted for a range of g. Since $\alpha\sim[\text{mass}]^2$, it should behave as

$$\alpha=C\Lambda_L^2 \tag{4.59}$$

in the weak coupling scaling region of the theory. Λ_L is a *physical mass* scale of the theory. As will be discussed in the next section, the renormalizability of the theory determines its dependence on the lattice spacing a and coupling constant g for g near zero,

$$\Lambda_L=\frac{1}{a}(\beta_0 g^2)^{-\beta_1/2\beta_0^2}e^{-1/2\beta_0 g^2} , \tag{4.60}$$

where

$$\beta_0=\frac{11}{3}\left[\frac{N}{16\pi^2}\right] , \quad \beta_1=\frac{34}{3}\left[\frac{N}{16\pi^2}\right]^2 \tag{4.61}$$

in SU(N) gauge theory.

So, Monte Carlo measurements of α should satisfy this *scaling* law. Equation (4.60) of the continuum limit and the constants C of Eq. (4.59) can be obtained for SU(2) and SU(3). There is a narrow window in g^2 where g is small enough that the weak coupling scaling law holds and where the correlations in the system are small enough that the computer measurements are not obscured by finite-size effects. Some Monte Carlo data [34] are shown in Fig. 34. Fits of the measurements give (Creutz, 1980)

$$\begin{aligned}\sqrt{\alpha}&=(79\pm12)\Lambda_L ,\\ \sqrt{\alpha}&=(220\pm66)\Lambda_L ,\end{aligned} \tag{4.62}$$

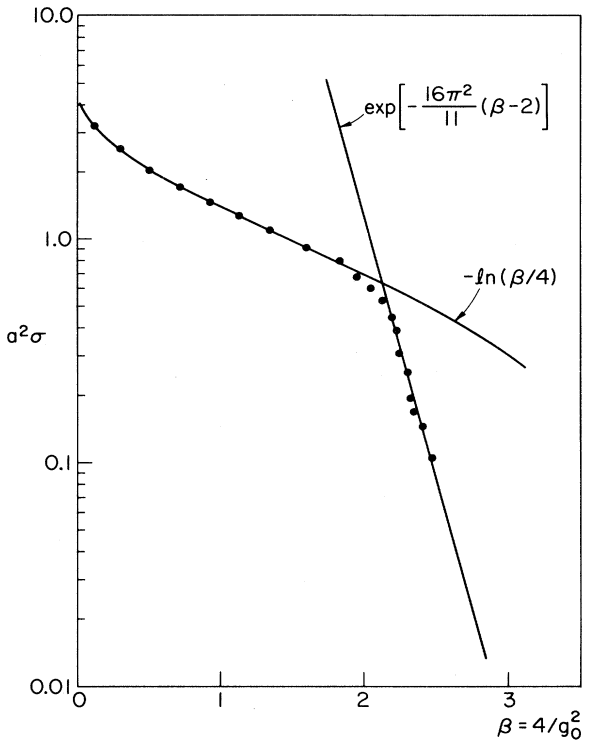

FIG. 34. SU(2) string tension Monte Carlo data vs coupling, $\beta=4/g^2$.

for SU(2) and SU(3), respectively. These large numbers of proportionality are artifacts of lattice regularization. Comparing these to momentum space regularization of the continuum model, one can calculate (Hasenfratz and Hasenfratz, 1980; Dashen and Gross, 1981), as will be illustrated in the next section,

$$\begin{aligned}\Lambda^{\text{mom}}/\Lambda_L&=57.5 ,\\ \Lambda^{\text{mom}}/\Lambda_L&=83.5 ,\end{aligned} \tag{4.63}$$

for SU(2) and SU(3), respectively. So, the calculation predicts the ratio of two nonperturbative mass scales of the continuum model,

$$\begin{aligned}\sqrt{\alpha}&=(1.3\pm.3)\Lambda^{\text{mom}} ,\\ \sqrt{\alpha}&=(2.5\pm1.0)\Lambda^{\text{mom}} ,\end{aligned} \tag{4.64}$$

also for SU(2) and SU(3), respectively. Λ^{mom} is measured (roughly!) in fits to deep inelastic scaling through a measurement of corrections to free field behavior at short distance. $\sqrt{\alpha}$ can be identified with the coefficient of the linear potential needed to explain heavy-quark spectroscopy. If $\sqrt{\alpha}\simeq450$ MeV, then Eq. (4.64) gives $\Lambda^{\text{mom}}\approx180$ MeV. These are not unreasonable numbers, even though the lattice calculation has ignored dynamical quarks.

D. Λ parameters and asymptotic freedom

The lattice provides a gauge-invariant regularization scheme for field theories. In asymptotically free theories

the continuum limit of the lattice model is found in the $g \to 0$ limit (Kogut, 1979; Kadanoff, 1977). $g^*=0$ is the infrared unstable fixed point of the theory. The neighborhood of $g^*=0$ can be studied using perturbation theory, suitably improved by the renormalization group if necessary, and the approach of the lattice theory to its continuum version can be understood in great detail. The vicinity of $g^*=0$ where perturbative analysis is adequate is called the "scaling region" of the theory. Asymptotically free scaling laws were discussed in Sec. II. The simplest such result predicts how the mass gap of the theory depends upon the cutoff and the bare coupling constant. The analysis can be done for any regularization scheme. We require that the mass m be a renormalization group invariant,

$$\frac{d}{da}m=0 , \tag{4.65}$$

where a might be a lattice or continuum cutoff. By dimensional analysis,

$$m=\frac{1}{a}f(g) \tag{4.66}$$

in a theory without an intrinsic scale. Substituting into Eq. (4.65), we find

$$f'(g)=-f(g)/\beta(g) , \tag{4.67}$$

where $\beta(g)$ in the Callan-Symanzik function,

$$\beta(g)=-a\frac{\partial g}{\partial a} . \tag{4.68}$$

In the vicinity of the fixed point $g^*=0$, $\beta(g)$ can be computed perturbatively,

$$a\frac{\partial g}{\partial a}=\beta_0 g^3+\beta_1 g^5+\cdots , \tag{4.69}$$

with

$$\beta_0=\frac{11}{3}\left[\frac{N}{16\pi^2}\right] , \quad \beta_1=\frac{34}{3}\left[\frac{N}{16\pi^2}\right]^2 , \tag{4.70}$$

in SU(N) gauge theory. If we use Eq. (4.69), the formal integral for $f(g)$ following from Eq. (4.67),

$$f(g)=\exp\left[-\int^g \frac{dg'}{\beta(g')}\right] , \tag{4.71}$$

can be evaluated:

$$f(g)\propto(\beta_0 g^2)^{-\beta_1/2\beta_0^2}\exp(-\tfrac{1}{2}\beta_0 g^2) . \tag{4.72}$$

Therefore the mass gap's dependence on the bare coupling is determined—this is a scaling law. It is analogous to the relationship for the correlation length in a statistical mechanics setting,

$$\xi\alpha \mid T-T_c \mid^{-\nu} , \quad T\approx T_c . \tag{4.73}$$

The critical index ν is not computable accurately in perturbation theory if the scale-invariant theory at the critical point T_c is interacting. It is a wonderful feature of gauge theories that the scaling laws can be obtained using ordinary perturbation theory!

Note that the mass gap's dependence on g is nonanalytic—mass generation in the theory is a nonperturbative effect. It was necessary to use the renormalization group to obtain this result. The Callan-Symanzik function itself is expandable in powers of g.

One of the aims of lattice gauge theory is the calculation of the mass spectrum of quantum chromodynamics. Equation (4.72) predicts how those masses must depend on g such that finite masses result in the continuum limit. The importance of verifying the scaling laws in nonperturbative lattice calculations is clear.

It is also necessary to relate the lattice regularization procedures, couplings, and cutoffs, to those of continuum methods. Then mass scales can be compared and physical predictions made. The Λ-parameter calculations done here are an important part of that program. I shall illustrate such calculations for asymptotically free spin models in two dimensions. These models are simple enough that all the calculations can be done analytically both on the lattice and in the continuum.

Consider Pauli-Villars regulation of a continuum model and lattice regularization of the same model. Both should give the same physics. We require that the renormalized couplings g_R^2 be identical, for the Pauli-Villars and lattice cases, respectively,

$$\begin{aligned} &g_R^2=Z_1^{-1}g^2 , \quad Z_1=1+B(M)g^2+\cdots , \\ &g_R^2=Z_{1L}^{-1}g_L^2 , \quad Z_{1L}=1+B_L(a)g_L^2+\cdots , \end{aligned} \tag{4.74}$$

for the two models. The functions $B(M)$ and $B_L(a)$ can be computed in perturbation theory. They depend logarithmically on the cutoff,

$$\begin{aligned} &B(M)=-\beta_0 \ln M+C , \\ &B_L(a)=-\beta_0 \ln(1/a)+C_L , \end{aligned} \tag{4.75}$$

where $\beta_0>0$ in an asymptotically free theory. Thus $B(M)$ expresses the asymptotic freedom of the model—keeping g_R constant and increasing M requires that g be made smaller. In the language of Callan and Symanzik, we hold g_R^2 fixed,

$$M\frac{\partial}{\partial M}g_R^{-2}=0=M\frac{\partial}{\partial M}(Z_1 g^{-2})=M\frac{\partial}{\partial M}[g^{-2}+B(M)] . \tag{4.76}$$

So,

$$0=-\frac{1}{g^4}M\frac{\partial}{\partial M}g^2-\beta_0\to M\frac{\partial}{\partial M}g^2=-\beta_0 g^4 , \tag{4.77}$$

which is the one-loop Callan-Symanzik equation. It can be integrated to obtain the renormalization-group trajectory,

$$\begin{aligned} &g^2(M)\cong\frac{1}{\beta_0 \ln(M/\Lambda)} , \quad \ln(M/\Lambda)\gg 1 \\ &g_L^2(a)\cong\frac{1}{\beta_0 \ln(1/a\Lambda_L)} , \quad \ln(1/a\Lambda_L)\gg 1 \end{aligned} \tag{4.78}$$

in each case. The Λ parameters set the scale in each logarithm. They are simply related, because

$$\frac{1}{g_R^2}=\frac{Z_1}{g^2}=\frac{Z_{1L}}{g_L^2}\ , \tag{4.79}$$

or

$$\frac{1}{g^2}+B(M)=\frac{1}{g_L^2}+B_L(a)\ ,$$
$$\beta_0\ln\left[\frac{M}{\Lambda}\right]+B(M)=\beta_0\ln\left[\frac{1}{a\Lambda_L}\right]+B_L(a)\ , \tag{4.80}$$

or

$$\frac{\Lambda}{\Lambda_L}=aM\exp\left[\frac{1}{\beta_0}[B(M)-B_L(a)]\right]$$
$$=\exp\left[\frac{1}{\beta_0}(C-C_L)\right]\ , \tag{4.81}$$

using the notation of Eq. (4.75). So, if we calculate the divergent and the *finite* parts of one-loop coupling constant renormalization, we can relate the Λ parameters of the two cutoff procedures.

Since Λ parameters set the scale of scale breaking in deep inelastic scattering, this calculation is worthwhile for quantum chromodynamics, as discussed in the preceding section.

I shall take this opportunity to illustrate *weak* coupling lattice calculations (Kogut and Shigemitsu, 1981). Consider the SU(N)×SU(N) spin model in two dimensions,

$$S=-\frac{1}{g^2}\sum_{x,\mu}\mathrm{tr}\{[U(x)U^\dagger(x+\mu)-1]+\mathrm{H.c.}\}\ ,$$
$$Z=\int\left[\prod_x dU\right]e^{-S}\ , \tag{4.82}$$

where $[dU]$ is the group-invariant integration measure. The SU(N) matrices $U(x)$, which reside on the lattice sites, fluctuate on all length scales from a to ∞. We wish to integrate out the high-frequency modes and obtain the effective action for the low-frequency modes. To make this division we use the *background field method* (DeWitt, 1967). Write

$$U(x)=e^{ig\phi(x)}U^{\mathrm{cl}}(x)\ , \tag{4.83}$$

where U^{cl} solves the classical equations of motion and $\phi(x)$ parametrizes the quantum fluctuations of $U(x)$. Define

$$\phi(x)=\lambda^\alpha\phi^\alpha(x)\ ,\quad \alpha=1,2,\ldots,N^2-1 \tag{4.84}$$

$$[\lambda^\alpha,\lambda^\beta]=if^{\alpha\beta\rho}\lambda_\rho\ ,$$
$$\mathrm{tr}\lambda_\alpha\lambda_\beta=\tfrac{1}{2}\delta_{\alpha\beta}\ . \tag{4.85}$$

Substituting into the action, we get

$$S=-\frac{1}{g^2}\sum_{x,\mu}\mathrm{tr}\{[U^{\mathrm{cl}}(x)U^{\mathrm{cl}\dagger}(x+\mu)-1]+\mathrm{H.c.}\}$$
$$-\frac{1}{g^2}\sum_{x,\mu}\mathrm{tr}[(e^{-ig\phi(x+\mu)}e^{ig\phi(x)}-1)U^{\mathrm{cl}}(x)U^{\mathrm{cl}\dagger}(x+\mu)+\mathrm{H.c.}]\ . \tag{4.86}$$

Since we are going to do perturbation theory in g^2, S and $[dU]$ should be expanded in powers of g^2,

$$e^{-ig\phi(x+\mu)}g^{ig\phi(x)}=\exp\{-ig\nabla_\mu\phi(x)+\tfrac{1}{2}g^2[\phi(x+\mu),\phi(x)]+O(g^3)\}$$
$$=1-ig\nabla_\mu\phi(x)+\tfrac{1}{2}g^2[\phi(x+\mu),\phi(x)]-\tfrac{1}{2}g^2[\nabla_\mu\phi(x)]^2+O(g^3)\ , \tag{4.86a}$$
$$dU=\prod_{\alpha=1}^{N^2-1}d\phi^\alpha[1+O(g^2\phi^2)]\ . \tag{4.86b}$$

Equation (4.86b) states that the curvature of the group manifold does not contribute to lowest order.

A convenient parametrization of U^{cl} is

$$U^{\mathrm{cl}}(x)U^{\mathrm{cl}}(x+\mu)=\exp[iF_\mu(x)]\ ,\quad F_\mu(x)=\lambda^\alpha F_\mu^\alpha(x)\ , \tag{4.87}$$

where $F_\mu(x)$ is Hermitian. It will prove sufficient to expand this result in powers of $F_\mu(x)$. Collecting everything, we find

$$S=\frac{1}{g^2}S^{\mathrm{cl}}+S_0+S_{\mathrm{int}}+O(g^2,F^4)\ , \tag{4.88}$$

where

$$S^{\mathrm{cl}}=-\tfrac{1}{2}\sum_{x,\mu}(F_\mu^\alpha)^2\ ,$$
$$S_0=\sum_{x,\mu}\tfrac{1}{2}(\nabla_\mu\phi^\alpha)^2\ , \tag{4.89}$$
$$S_{\mathrm{int}}=\sum_{x,\mu}[-\tfrac{1}{2}\mathrm{tr}(\lambda^\alpha\lambda^\beta\lambda^\gamma\lambda^\delta)\nabla_\mu\phi^\alpha\nabla_\mu\phi^\beta F_\mu^\gamma F_\mu^\delta$$
$$+\tfrac{1}{2}f^{\alpha\beta\gamma}\phi^\alpha(x+\mu)\phi^\beta(x)F_\mu^\gamma]\ .$$

Thus we have a classical free field piece, a quantum free field piece, and interactions. It is convenient to modify S_0,

$$S_0\to\sum_{x,\mu}[\tfrac{1}{2}(\nabla_\mu\phi^\alpha)^2+\tfrac{1}{2}m^2\phi^{\alpha 2}]\ , \tag{4.90}$$

to avoid spurious infrared problems in the midst of a calculation. The limit $m \to 0$ will be smooth for the gauge-invariant quantities we calculate.

So far our manipulations hold for the continuum model or a lattice model. Consider a Pauli-Villars regulated continuum calculation first. The ∇_μ becomes an ordinary differential $a\partial_\mu$ and the $S_{\rm int}$ simplifies. Its first term behaves as a^2, since $F_\mu \sim a$, $\nabla_\mu \phi \sim a\partial_\mu \phi$, $\sum_{x,\mu} \sim \int d^2x/a^2$, and it vanishes in the continuum limit. The second term survives. We define

$$F_\mu^\gamma(x) \to F_\mu^\gamma(x)/a \tag{4.91}$$

and find

$$S_{\rm int}^{\rm cont} = \tfrac{1}{2} f^{\alpha\beta\gamma} \int \partial_\mu \phi^\alpha \phi^\beta F_\mu^\gamma d^2x \ .$$

Now we calculate the continuum Z through one-loop,

$$\begin{aligned} Z &= \int \pi\, d\phi^\alpha \exp\left[-\frac{1}{g^2}S^{\rm cl} - S_0 - S_{\rm int}\right][1+O(g^2,F^4)] \\ &= \exp\left[-\frac{1}{g^2}S^{\rm cl}\right] \int \pi\, d\phi^a e^{-S_0}(1 - S_{\rm int} + \tfrac{1}{2}S_{\rm int}^2 + \cdots) \\ &= \exp\left[-\frac{1}{g^2}S^{\rm cl}\right](1 + \langle S_{\rm int}\rangle + \tfrac{1}{2}\langle S_{\rm int}^2\rangle + \cdots) \ . \end{aligned} \tag{4.92}$$

But $\langle S_{\rm int}\rangle = 0$, because it transforms as a gradient. The second-order term is

$$\begin{aligned} \tfrac{1}{2}\langle S_{\rm int}^2\rangle &= \tfrac{1}{8}\int d^2y\, d^2y' f^{\alpha\beta\gamma} f^{\alpha'\beta'\gamma'} F_\mu^\gamma(y) F_\nu^{\gamma'}(y') \langle \partial_\mu \phi^\alpha(y) \phi^\beta(y) \partial_\nu \phi^{\alpha'}(y') \phi^{\beta'}(y')\rangle \\ &= \tfrac{1}{8} N \int d^2y\, d^2y' F_\mu^\gamma(y) F_\nu^\gamma(y') [G(y-y')\partial_\mu \partial_\nu' G(y-y') - \partial_\mu G(y-y')\partial_\nu' G(y-y')] \ , \end{aligned} \tag{4.93}$$

where we used $f^{\alpha\beta\gamma} f^{\alpha\beta\gamma'} = N\delta_{\gamma\gamma'}$, and we recall

$$G(y) = \int \frac{d^2k}{(2\pi)^2} \frac{e^{iky}}{k^2+m^2} \ . \tag{4.94}$$

The background field can be chosen slowly varying,

$$\partial_\mu F_\nu \ll F_\nu \ , \tag{4.95}$$

so

$$\begin{aligned} \tfrac{1}{2}\langle S_{\rm int}^2\rangle &= \tfrac{1}{4} N \int d^2y\, d^2y' F_\mu^\gamma(y) F_\nu^\gamma(y') \\ &\quad \times G(y-y')(-\partial_\mu \partial_\nu) G(y-y') \ . \end{aligned} \tag{4.96}$$

Since we are letting F vary much more slowly than G, $F_\nu^\gamma(y')$ can be replaced inside the integral by $F_\nu^\gamma(y)$ and the d^2y' integral can be done explicitly. Then the replacement $-\partial_\mu \partial_\nu \to -\frac{1}{2}\delta_{\mu\nu}$ is valid, $\partial_\mu \partial_\mu G(y-y') = -\delta(y-y')$ and we have

$$\begin{aligned} \tfrac{1}{2}\langle S_{\rm int}^2\rangle &= \tfrac{1}{4} N G(0) \tfrac{1}{2} \int F_\mu^\gamma(y) F_\mu^\gamma(y) d^2y \\ &= \tfrac{1}{4} N G(0) S^{\rm cl} \ . \end{aligned} \tag{4.97}$$

Then the partition function is

$$\begin{aligned} Z &= \exp\left[-\frac{1}{g^2}S^{\rm cl}\right][1 + \tfrac{1}{4}NG(0)S^{\rm cl}] \ , \\ Z &= \exp\left[-\left(\frac{1}{g^2} - \tfrac{1}{4}NG(0)\right)\tfrac{1}{2}\int d^2y\, F_\mu^\gamma G_\mu^\gamma\right] \end{aligned} \tag{4.98}$$

to this order. We recognize coupling constant renormalization—the low-energy modes fluctuate with a renormalized coupling constant,

$$\frac{1}{g_R^2} = \frac{Z_1}{g^2} \ , \quad Z_1 = 1 - \tfrac{1}{4}g^2 NG(0) \ . \tag{4.99}$$

But this is still a formal expression, since $G(0)$ is ultraviolet divergent. We choose the Pauli-Villars (PV) regulator scheme—introducing a negative metric $\hat{\phi}^\alpha$ field with mass M coupled to F just as ϕ^α is. The calculation above goes through with the obvious final replacement.

$$Z_1 = 1 - \tfrac{1}{4}g^2 N G^{\rm PV}(0) \ , \tag{4.100}$$

where

$$G^{\rm PV}(0) = \int \frac{d^2k}{(2\pi)^2}\left[\frac{1}{k^2+m^2} - \frac{1}{k^2+M^2}\right] \ . \tag{4.101}$$

So,

$$Z_1 = 1 - \frac{Ng^2}{8\pi}\ln(M/m) \ . \tag{4.102}$$

From this we learn that the theory is asymptotically free,

$$g_R^2 = \frac{1}{Z_1}g^2 = g^2 + \frac{Ng^4}{8\pi}\ln(M/m) \ , \tag{4.103}$$

with the crucial + sign traced to the positive curvature of the compact group manifold (Polyakov, 1975). And the Callan-Symanzik function is identified,

$$\begin{aligned} M\frac{\partial}{\partial M}g_R^2 = 0 &= M\frac{\partial}{\partial M}g^2 + \frac{Ng^4}{8\pi} \\ &\to M\frac{\partial}{\partial M}g^2 = -\frac{N}{8\pi}g^4 \ . \end{aligned} \tag{4.104}$$

Now we repeat this calculation using the lattice regulator. We face again

$$Z=\exp\left[-\frac{1}{g^2}S^{\rm cl}\right](1+\langle S_{\rm int}\rangle+\tfrac{1}{2}\langle S_{\rm int}^2\rangle+\cdots)\,. \tag{4.105}$$

The first term is

$$\langle S_{\rm int}\rangle=\sum_{x,\mu}-\tfrac{1}{2}{\rm tr}(\lambda^\alpha\lambda^\beta\lambda^\gamma\lambda^\delta)F_\mu^\gamma F_\mu^\delta\langle\nabla_\mu\phi^\alpha\nabla_\mu\phi^\beta\rangle$$
$$+\tfrac{1}{2}f^{\alpha\beta\gamma}F_\mu^\gamma\langle\phi^\alpha(x+\mu)\phi^\beta(x)\rangle\,. \tag{4.106}$$

Its second piece is zero, because the average of a vector quantity vanishes identically. The first piece does not vanish. Only the terms with $\alpha=\beta$ and $\gamma=\delta$ contribute. For these indices we have the replacement

$${\rm tr}\lambda^\alpha\lambda^\beta\lambda^\gamma\lambda^\delta=\frac{1}{8}\frac{2}{N}\delta^{\alpha\beta}\delta^{\gamma\delta}+\cdots\,. \tag{4.107}$$

So

$$\tfrac{1}{2}\langle S_{\rm int}^2\rangle=\tfrac{1}{8}f^{\alpha\beta\gamma}f^{\alpha'\beta'\gamma'}\sum_{yy'\mu\nu}F_\mu^\gamma(y)F_\nu^{\gamma'}(y')\langle\phi^\alpha(y+\mu)\phi^\beta(y)\phi^{\alpha'}(y'+\nu)\phi^{\beta'}(y')\rangle+O(F^4)$$
$$=\tfrac{1}{8}N\sum_{yy'\mu\nu}F_\mu^\gamma(y)F_\nu^\gamma(y')[G(y-y'+\mu-\nu)G(y-y')-G(y-y'+\mu)G(y-y'-\nu)]\,, \tag{4.112}$$

where we used $f^{\alpha\beta\gamma}f^{\alpha\beta\gamma'}=N\delta_{\gamma\gamma'}$. Now we must simplify this collection of propagators. Let $\Delta=4-2\cos k_1-2\cos k_2+m^2a^2$. In momentum space the propagators in Eq. (4.112) become

$$\int\frac{d^2k}{(2\pi)^2}\int\frac{d^2k'}{(2\pi)^2}\frac{e^{ik_\mu}(e^{-ik_\nu}-e^{-ik'_\nu})(e^{ik(y-y')}e^{ik'(y-y')})}{\Delta(k)\Delta(k')}\,. \tag{4.113}$$

The background field is slowly varying, so we replace $F_\nu^\gamma(y')\to F_\nu^\gamma(y)$, and the y' integral is trivial,

$$\tfrac{1}{2}\langle S_{\rm int}^2\rangle=\tfrac{1}{8}N\sum_{y\mu\nu}F_\mu^\gamma(y)F_\nu^\gamma(y)\int\frac{d^2k}{(2\pi)^2}\frac{e^{ik_\mu}(e^{-ik_\nu}-e^{ik_\nu})}{\Delta^2(k)}\,. \tag{4.114}$$

The momentum integral vanishes unless $\mu=\nu$, so

$$\tfrac{1}{2}\langle S_{\rm int}^2\rangle=\tfrac{1}{4}NS^{\rm cl}G(1)\,, \tag{4.115}$$

where

$$G(1)=\int\frac{d^2k}{(2\pi)^2}\frac{e^{ik_1}(e^{-ik_1}-e^{ik_1})}{\Delta^2(k)} \tag{4.116}$$

can be shown (Kogut and Shigemitsu, 1981). Collecting all this, we have the partition function with the high frequencies integrated out,

$$Z=\exp\left[-\frac{Z_{1L}}{g^2}S^{\rm cl}\right]\,, \tag{4.117}$$

with

$$Z_{1L}=1-g^2\left[\tfrac{1}{4}NG_L(1)+\frac{N^2-1}{8N}\right]$$
$$=1-g^2\left[\tfrac{1}{4}NG_L(0)+\frac{N^2-1}{8N}-\frac{N}{16}\right]\,. \tag{4.118}$$

$$\langle S_{\rm int}\rangle=-\frac{1}{8N}\sum_x\tfrac{1}{2}(F_\nu^\gamma)^2\langle(\nabla_\mu\phi^\alpha)^2\rangle$$
$$=-\frac{1}{8N}\sum_x\tfrac{1}{2}(F_\nu^\gamma)^2 4(N^2-1)[G(0)-G(1)]\,, \tag{4.108}$$

where

$$G(x)=\langle\phi^1(x)\phi^1(0)\rangle$$
$$=\int_{-\pi}^{\pi}\frac{d^2k}{(2\pi)^2}\frac{e^{ikx}}{4-2\cos k_1-2\cos k_2+m^2a^2} \tag{4.109}$$

is the free scalar propagator on the lattice. From this expression or the discrete form of the differential equation $(\partial_\mu\partial_\mu+m^2a^2)G(x)=-\delta(x)$, one easily checks that

$$G(1)-G(0)=-\tfrac{1}{4}+O(m^2a^2)\,. \tag{4.110}$$

So

$$\langle S_{\rm int}\rangle=\frac{N^2-1}{8N}\sum_x\tfrac{1}{2}(F_\nu^\gamma)^2=\frac{N^2-1}{8N}S^{\rm cl}\,. \tag{4.111}$$

And finally we have the second-order term,

Since (Kogut and Shigemitsu, 1981)

$$G_L(0)=\frac{1}{4\pi}\ln\left[\frac{32}{m^2a^2}\right]+O(m^2a^2)\,, \tag{4.119}$$

we have,

$$Z_{1L}=1-g^2\left[\frac{1}{16\pi}N\ln\left[\frac{32}{m^2a^2}\right]+\frac{N^2-1}{8N}-\frac{N}{16}\right]\,. \tag{4.120}$$

Comparing this with the Pauli-Villars calculation, we have the same one-loop, Callan-Symanzik equation, as expected. The *finite* pieces of the renormalization are different. In the language of our earlier definitions,

$$C^{\rm PV}=0\,,$$
$$C^L=-\frac{1}{16\pi}\ln 32-\frac{1}{16N^2}(N^2-2)\,. \tag{4.121}$$

So

$$\Lambda^{PV}/\Lambda_L = \exp\left[8\pi\left[\frac{1}{16\pi}\ln 32+\frac{1}{16N^2}(N^2-2)\right]\right]$$
$$=\exp\left[\ln\sqrt{32}+\frac{\pi}{2N^2}(N^2-2)\right]$$
$$=\sqrt{32}e^{\pi(N^2-2)/2N^2}. \quad (4.122)$$

And there is a huge change in scale between the methods, $\Lambda^{PV} \gg \Lambda^L$. Since

$$g^2(M)=\frac{1}{\beta_0 \ln(M/\Lambda)}, \quad g_L^2(a)=\frac{1}{\beta_0 \ln(1/a\Lambda_L)}, \quad (4.123)$$

the bare lattice coupling is much less than the bare Pauli-Villars coupling for the same renormalized physics. The region in the complex coupling constant plane where weak coupling and semiclassical physics apply is much smaller in the lattice theory than in a continuum theory. The onset of nonperturbative effects begins at a much smaller coupling in the lattice approach than in the usual continuum parametrizations of the same theory.

A final word about the absolute scale chosen in the Λ-parameter definition. With the choice

$$\Lambda_L=\frac{1}{a}(\beta_0 g^2)^{-\beta_1/2\beta_0^2} e^{-1/2\beta_0 g^2} \quad (4.124)$$

one can check that

$$g_L^2(a)=\frac{1}{\beta_0 \ln(1/a\Lambda_L)+\beta_1/\beta_0 \ln(\ln 1/a\Lambda_L)}. \quad (4.125)$$

So, with the choice of scale in Λ_L's definition, the two-loop correction does not rescale the length scale conventions chosen in the one-loop calculations—i.e., there are no corrections of the form $\beta_1 \times$const in the denominator of $g_L^2(a)$. Therefore, this convention gives the widest range of validity to the one-loop calculation.

The huge change in scales between the lattice model and the continuum model is a common phenomenon. It depends on the detailed lattice action used. The SU(2) and SU(3) gauge theory results were quoted in Sec. IV.C and were useful in setting the scale in the string tension Monte Carlo calculations.

V. HAMILTONIAN LATTICE GAUGE THEORY, FLUX TUBE DYNAMICS, AND CONTINUUM STRING MODELS

A. The transfer matrix and the Hamiltonian limit

In addition to the partition function approaches to field theory discussed to this point, one might also investigate the space of states and the spectrum of lattice gauge theory. To do this systematically we should consider the transfer matrix and the Hamiltonian, time continuum limit, of the theory (Kogut, 1979).

The partition function approach of lattice gauge theories gave a lattice regularization of

$$L=-\tfrac{1}{4}\int F^a_{\mu\nu}F^a_{\mu\nu}d^3x. \quad (5.1a)$$

The formal replacement was

$$L=-\tfrac{1}{4}\int F^a_{\mu\nu}F^a_{\mu\nu} \to \frac{1}{4g^2a^4}\mathrm{tr}P\exp\left[igC\oint_C A^\alpha_\mu \frac{\tau^\alpha}{2}dx^\mu\right]$$
$$\to \frac{1}{4g^2a^4}\mathrm{tr}UUUU. \quad (5.1b)$$

We chose simple plaquettes for the contours C in Eq. (5.1b).

To make a Hamiltonian theory, we distinguish between plaquettes with a t link and those without. Those without are spatial plaquettes and are left alone in this argument. The temporal links are special—we want to organize things so that the partition function can be written in the form

$$Z=\int \langle\{U_F\}|\hat{T}|\{U_{N-1}\}\rangle\left[\prod dU_{N-1}\right]\langle\{U_{N-1}\}|\hat{T}|\{U_{N-2}\}\rangle\cdots\langle\{U_1\}|\hat{T}|\{U_i\}\rangle, \quad (5.2)$$

where $\hat{T}$ is the theory's transfer matrix. This equation is visualized in Fig. 35. Clearly, $\hat{T}$ evolves the system one link in the "time" direction. $\hat{T}$ will be simple if we choose temporal links to be trivial, $e^{iB_0}=1$. Therefore, we work in the temporal gauge $A_0=0$. We recall from continuum quantization that in this case gauge invariance is imposed as a constraint on the Hilbert space of states—i.e., Gauss's law is implemented

$$G_\alpha(\mathbf{n})|\mathrm{phys}\rangle=0, \quad (5.3)$$

where $G_\alpha(\mathbf{n})$ is the generator of local rotations in color space at the site $\mathbf{n}=(x,y,z)$. An explicit construction of G_α will be obtained below.

For temporal plaquettes Eq. (5.1b) simplifies,

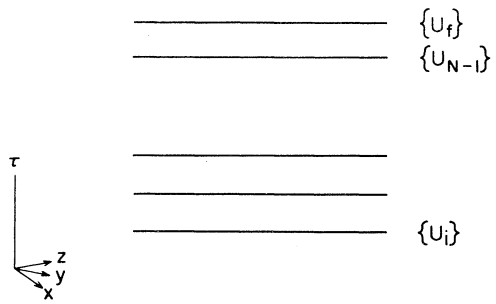

FIG. 35. A transfer matrix decomposition of the partition function. $\{U\}$ indicates a three-dimensional gauge field configuration at a given τ.

$$\mathrm{tr}U^\dagger(t_{i+1})U(t_i)+\mathrm{H.c.}=-\mathrm{tr}[U^\dagger(t_{i+1})-U^\dagger(t_i)]\times[U(t_{i+1})-U(t_i)]+\mathrm{const} \tag{5.4}$$

and we can construct the "velocity" term of the Lagrangian. The temporal loops shown in Fig. 36 contribute to the Lagrangian density,

$$-\frac{1}{4g^2a^2}\mathrm{tr}\left[\frac{U^\dagger(t_{i+1})-U^\dagger(t_i)}{a_t}\right]\left[\frac{U(t_{i+1})-U(t_i)}{a_t}\right], \tag{5.5}$$

where a_t is the temporal lattice spacing. Taking $a_t\to 0$, we obtain a conventional quantum-mechanical picture of the lattice gauge theory,

$$L=\sum_{\text{links}}\frac{a}{4g^2}\mathrm{tr}\dot{U}^\dagger\dot{U}+\sum_{\text{plaq}}\frac{1}{4ag^2}\mathrm{tr}UUUU+\mathrm{H.c.}\ , \tag{5.6}$$

where the replacement $\int d^3x\to\sum a^3$ has been made. Now we can form the Hamiltonian, passing from "velocity" to "momentum" variables by canonical procedures.

$$H=\sum\left[\dot{U}_{ij}^\dagger\frac{\partial L}{\partial\dot{U}_{ij}^\dagger}+U_{ij}\frac{\partial L}{\partial\dot{U}_{ij}}\right]-L$$
$$=\sum_{\text{links}}\frac{a}{4g^2}\mathrm{tr}\dot{U}^\dagger\dot{U}-\sum_{\text{plaq}}\frac{1}{4ag^2}(\mathrm{tr}UUUU+\mathrm{H.c.})\ , \tag{5.7}$$

which describes a system of coupled "tops." To quantize this system we must identify independent degrees of freedom carefully. One might use Euler angles, but this is awkward. Instead we shall eliminate $\dot{U}$ in favor of the generators of local gauge rotations (Coleman, 1976; Creutz, 1977). If an infinitesimal gauge rotation is made at site $\mathbf{n}$,

$$U(\mathbf{n},i)\to\left[1+i\varepsilon\frac{\tau^\alpha}{2}\right]U(\mathbf{n},i)\ , \tag{5.8}$$

where i labels the direction of the link emanating from site $\mathbf{n}$. We call the quantum generators of these transformations $E^\alpha(\mathbf{n},i)$. So

$$[E^\alpha(\mathbf{n},i),U(\mathbf{m},j)]=\tfrac{1}{2}\tau^\alpha U(\mathbf{m},j)\delta_{ij}\delta_{\mathbf{nm}} \tag{5.9}$$

in a slightly abbreviated notation. Since the E^α's generate SU(2) rotations,

$$[E^\alpha(\mathbf{n},i),E^\beta(\mathbf{m},j)]=i\varepsilon^{\alpha\beta\gamma}E^\gamma(\mathbf{n},i)\delta_{ij}\delta_{\mathbf{nm}}\ . \tag{5.10}$$

Now we want to eliminate $\dot{U}^\dagger$ and $\dot{U}$ in favor of the E^α variables. Since E^α generates the local gauge rotation which is a symmetry of L,

$$E^\alpha=\frac{\partial L}{\partial\dot{U}_{ij}}\left[i\frac{\tau^\alpha}{2}U\right]_{ij}+\frac{\partial L}{\partial\dot{U}_{ij}^\dagger}\left[-i\frac{\tau^2}{2}U\right]_{ij}^\dagger$$
$$=i\frac{a}{4g^2}(\mathrm{tr}\dot{U}^\dagger\tfrac{1}{2}\tau^\alpha U-\mathrm{H.c.})\ . \tag{5.11}$$

Next consider $E^\alpha E^\alpha$. To compute this, the quadratic Casimir operator, we need two identities,

$$U^\dagger U=1\ ,\quad \dot{U}^\dagger U+U^\dagger\dot{U}=0\ , \tag{5.12a}$$

and

$$\tau^\alpha_{ij}\tau^\alpha_{kl}=2\delta_{il}\delta_{jk}-1\delta_{ij}\delta_{kl}\ , \tag{5.12b}$$

and we compute

$$E^\alpha E^\alpha=\frac{a^2}{2g^4}\mathrm{tr}\dot{U}^\dagger\dot{U}\ . \tag{5.13}$$

Collecting everything (Kogut and Susskind, 1975), we have

$$H=\sum_{\text{links}}\frac{g^2}{2a}E^\alpha E^\alpha-\frac{1}{4ag^2}\sum_{\text{plaq}}(\mathrm{tr}UUUU+\mathrm{H.c.})\ . \tag{5.14}$$

The basic commutation relations

$$[E^\alpha(\mathbf{n},i),U(\mathbf{m},j)]=\tfrac{1}{2}\tau^\alpha U(\mathbf{m},j)\delta_{ij}\delta_{\mathbf{nm}}\ ,$$
$$[E^\alpha(\mathbf{n},i),E^\beta(\mathbf{m},j)]=i\varepsilon^{\alpha\beta\gamma}E^\gamma(\mathbf{n},i)\delta_{ij}\delta_{\mathbf{nm}}\ , \tag{5.15}$$

specify the model with the definition of physical states

$$\sum_{j=-3}^{3}E^\alpha(\mathbf{n},j)\,|\,\mathrm{phys}\rangle=0\ . \tag{5.16}$$

Note that in the partition function language time-independent gauge transformations are

$$\prod_{n_0=-\infty}^{\infty}G(n_0,\mathbf{n})\ . \tag{5.17}$$

Since $U=1$ on temporal links and

$$U=1\to e^{i(\tau/2)\chi(\mathbf{n})}Ue^{-i(\tau/2)\chi(\mathbf{n})}=1\ , \tag{5.18}$$

the axial gauge is preserved appropriately for arbitrary time-independent gauge functions $\chi(\mathbf{n})$. Clearly also

$$[E^\alpha(\mathbf{n},i),H]=0 \tag{5.19}$$

and the H is gauge invariant.

Now we are ready for applications. We want to understand the distribution of electric flux responsible for the linear confining potential. Consider a $Q\bar{Q}$ state. Let the

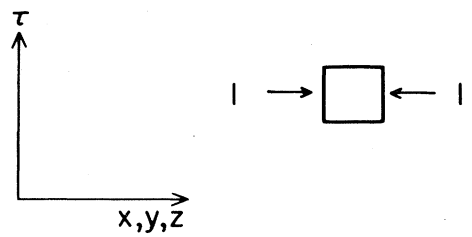

FIG. 36. A temporal plaquette in the gauge $A_0=0$.

Q and $\bar{Q}$ be heavy and let them be created by the operators $\psi^\dagger(n)$ and $\psi(n)$—they are sources and sinks of color flux of one unit. The generator of a color rotation is now generalized to

$$\sum E^\alpha(\mathbf{n},i) \rightarrow \sum E^\alpha(\mathbf{n},i) + \psi^\dagger(\mathbf{n}) \tfrac{1}{2}\tau^\alpha \psi(\mathbf{n}) \tag{5.20}$$

at the site $\mathbf{n}$. We need a gauge-invariant operator to describe the heavy $Q\bar{Q}$ state. The operator

$$\psi^\dagger(\mathbf{n}) \left\{ \prod_{\text{path}} U \right\} \psi(\mathbf{n}+\mathbf{R}) , \tag{5.21}$$

where the "path" extends from $\mathbf{n}$ to $\mathbf{n}+\mathbf{R}$ but is otherwise arbitrary, satisfies this requirement. The local contraction of color indices guarantees the local gauge invariance of this operator. Now consider the theory at strong coupling $g^2 \gg 1$. Then the leading term in H is

$$H_0 = \frac{g^2}{2a} \sum_l E_l^2 . \tag{5.22}$$

The vacuum has each link in a color singlet state,

$$E_l^\alpha \mid 0\rangle = 0 . \tag{5.23}$$

So the $Q\bar{Q}$ state of minimal energy will have $\prod_{\text{path}}$ over the fewest number of links, since each link with a U matrix costs energy,

$$E^2 U \mid 0\rangle = \tfrac{1}{2}\tau^\alpha \tfrac{1}{2}\tau^\alpha U \mid 0\rangle = \tfrac{3}{4} U \mid 0\rangle , \tag{5.24}$$

as follows from the commutation rule for SU(2),

$$[E^\alpha, U] = \tfrac{1}{2}\tau^\alpha U . \tag{5.25}$$

Thus the energy of the $Q\bar{Q}$ state is

$$V(R) = \frac{g^2}{2a} \frac{3}{4} \frac{R}{a} = \alpha R , \tag{5.26}$$

which gives us the strong coupling limit of the string tension in this theory,

$$\alpha = \frac{3}{8} \frac{g^2}{a^2} . \tag{5.27}$$

Note that requiring α to be independent of a implies that $g \sim a$—stronger coupling on a coarser lattice. This is the same qualitative behavior as found at weak coupling, where asymptotic freedom implied stronger coupling on coarser lattices. This strong coupling calculation of α can be carried to high orders in the natural expansion parameter, $1/g^2$, of the Hamiltonian theory (Kogut *et al.*, 1979, 1980). These results complement the Monte Carlo simulations and also suggest that the continuum limit of the pure gauge theory confines quarks with a finite, fixed, physical string tension. But what about the spatial distribution of the flux responsible for confinement? At strong coupling the flux exists in a thin tube between the two on-axis quarks. Is this a possible state of flux in the continuum limit? To gain some intuition into this question we turn to a simple string model.

B. Relativistic thin strings, delocalization, and Casimir forces

Consider a structureless thin string with its ends pinned down at $\mathbf{x} = -\frac{1}{2}\mathbf{R}$ and $\mathbf{x} = \frac{1}{2}R$. It can be described by a two-component vector field $\xi(t,z)$, $-\frac{1}{2}R \le z \le \frac{1}{2}R$, $\xi(t,-\frac{1}{2}R) = \xi(t,\frac{1}{2}R) = 0$, as in Fig. 37. We want an effective action describing the low-frequency modes of this string. Let's make the following assumptions about it (Lüscher, 1981):

(1) S_{eff} must be local, $S_{\text{eff}} = \int dz\, dt\, L$, where L depends on ξ and its derivatives.

(2) L should be invariant to the following space-time symmetries:

(a) Poincaré transformations in the (t,z) plane.

(b) O(2) rotations and translations of ξ.

Now let's write the possible terms in L. Property (2b) precludes mass terms $\sim \xi^2$ for the string. So L must be made up of derivatives $\partial_\mu \xi, \partial_\mu \partial_\nu \xi, \ldots$ $(\mu,\nu = 0.1)$. Then

$$S_{\text{eff}} = \int_{-R/2 \le z \le R/2} dz\, dt \{ \partial_\mu \xi \cdot \partial^\mu \xi + b \partial_\mu \partial^\mu \xi \cdot \partial_\nu \partial^\mu \xi + c (\partial_\mu \xi \cdot \partial^\mu \xi)^2 + \cdots \} . \tag{5.28}$$

The only relevant or marginal operator is $\partial_\mu \xi \cdot \partial^\mu \xi$—the other operators are irrelevant (Wilson and Kogut, 1974), i.e., the parameters b and c carry dimensions of the underlying spatial cutoff to various powers. Order by order in perturbation theory these irrelevant operators do not contribute to the long-wavelength physics of the string. One can check that when these operators are inserted into Feynman diagrams the infrared behavior of the graph improves. So, for studying wavelengths $\lambda \gg a$, the intrinsic width of the string, it suffices to take only

$$S_{\text{eff}} = \int dz\, dt\, \partial_\mu \xi \cdot \partial^\mu \xi . \tag{5.29}$$

Now we can ask whether the state with $\langle \xi(z,t) \rangle = 0$ is stable to quantum fluctuations. We calculate the variance in $\xi(z,t)$ for any z between $-\frac{1}{2}R$ and $\frac{1}{2}R$,

$$\langle \xi^2(z,t) \rangle \sim \int \frac{d^2k}{k^2} \sim \ln(R/a) . \tag{5.30}$$

So $\langle \xi^2(z,t) \rangle \rightarrow \infty$ as $R \rightarrow \infty$ and the string is delocalized by fluctuations as shown in Fig. 38! This explicit calculation is an example of the more general Mermin-Wagner theorem (Mermin and Wagner, 1966): spontaneous breakdown of a continuous global symmetry in a theory with local couplings in two dimensions is not possible.

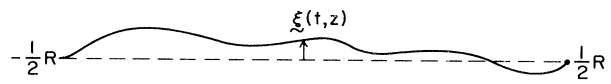

FIG. 37. A thin string labeled with a vector field $\xi(t,z)$.

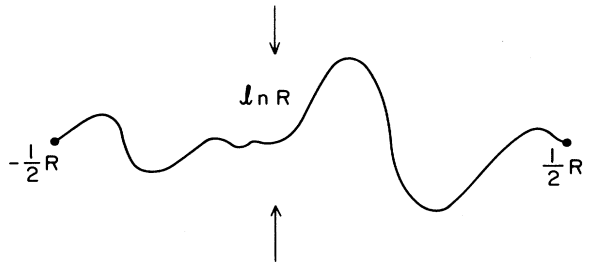

FIG. 38. A thin string with a divergent mean width $\sim \ln R$.

A second physical effect of these massless modes is the existence of a universal $1/R$ potential in the string channel of the theory (Lüscher, 1981),

$$V(R)=\alpha \mid R \mid -\frac{(d-2)\pi/24}{R}+\cdots . \tag{5.31}$$

The $1/R$ correction term is a one-dimensional Casimir effect which is easily calculated. For thin, structureless strings it is truly universal—it depends only on the space-time dimension d in which the string lives! We begin with a free, massless field in a two-dimensional box of length R,

$$S=\tfrac{1}{2}\int_{0\le z\le R} dz\,dt\,\partial_\mu\xi\cdot\partial^\mu\xi . \tag{5.32}$$

Think of it as a box of massless bosons, one flavor for each transverse direction. Each transverse mode contributes $\frac{1}{2}\hbar\omega$ to the ground-state energy,

$$\Delta V(R)=\tfrac{1}{2}\sum_n \varepsilon_n , \quad \varepsilon_n=\pi n/R , \quad n=0,1,2,\ldots , \tag{5.33}$$

when we have counted the massless normal modes for a field with fixed boundary conditions $\xi(0,t)=\xi(R,t)=0$. The sum over normal modes diverges—the fluctuations shift the ground-state energy density. To separate off this term we first do the sum with a convergence factor $e^{-\lambda n}$ and let $\lambda\to 0$ at the end. (Actually, for a string of thickness a the largest physically sensible value for n is $\sim R/a$, so that $\varepsilon\lesssim\pi/a$. Therefore the minimal value for λ is $\lambda_{\min}\sim a/R$.) Now

$$\begin{aligned}\Delta V(\lambda,R)&=\frac{\pi}{2R}\sum_n n e^{-\lambda n}=-\frac{\pi}{2R}\frac{d}{d\lambda}\sum_n e^{-\lambda n}\\&=-\frac{\pi}{2R}\frac{d}{d\lambda}\left[\frac{1}{e^{\lambda}-1}\right]\\&=-\frac{\pi}{2R}\sum_n B_n\frac{(n-1)\lambda^{n-2}}{n!} ,\end{aligned} \tag{5.34}$$

where B_n are Bernoulli numbers. The $n=2$ term gives the only finite, nonzero term which survives in the $\lambda\to 0$ limit,

$$\Delta V(R)=-\frac{\pi}{2R}\frac{B_2}{2}=-\frac{\pi/24}{R} , \tag{5.35}$$

for each transverse degree of freedom. For $(d-2)$ transverse directions we have the advertised result,

$$\Delta V(R)=-\frac{(d-2)\pi/24}{R} . \tag{5.36}$$

Note that the $n=0$ term in Eq. (5.34) simply renormalizes the string tension additively.

C. Roughening and the restoration of spatial symmetries in lattice gauge theory

How do these features—the delocalization of the string and the universal $1/R$ potential—occur in the lattice-regulated theory? There is only cubic symmetry on the spatial lattice, so straight strings are possible at strong coupling. As $g\to 0$, however, the magnetic effects in the Hamiltonian become important; and at some critical coupling, the roughening point g_R the string will delocalize (Hasenfratz *et al.*, 1981; Itzykson *et al.*, 1980; Lüscher *et al.*, 1981). For the relativistic string the lowest-energy transverse excitation costs zero energy in the $R\to\infty$ limit. On the lattice the lowest-energy transverse excitation is shown in Fig. 39, and it has an energy $(g^2/2a)\frac{3}{4}(N+1)$ in the strong coupling limit. This is an energy of $\frac{3}{4}g^2/2a$ above that of a straight string of N links. Call the state a "kink" and $m_k=(g^2/2a)(\frac{3}{4}+\cdots)$ where the $+\cdots$ means that the mass can be calculated in perturbation theory (Kogut *et al.*, 1981). The kink should be thought of as a particle—its energy is localized and it can carry momentum. The transverse distribution of electric flux is bounded, because the transverse wandering of the string is inhibited by a finite energy barrier, m_k. As g decreases from strong coupling, however, m_k decreases and eventually vanishes at $g=g_R$. Then there is no energy barrier inhibiting the transverse wandering of the string. For all $g\le g_R$ the lattice string is delocalized and resembles more closely the continuum string. Roughening can therefore also be seen in a calculation of the mean transverse width of the string,

$$\langle r_\perp^2\rangle=\frac{\langle r_\perp^2\mathbf{E}_{\parallel}^2\rangle}{\langle\mathbf{E}_{\parallel}^2\rangle} , \tag{5.37}$$

where $\mathbf{E}_{\parallel}^2$ is the electric flux squared on a link parallel to the straight string but displaced in the transverse direction by $r_\perp$ links. One can develop strong coupling series for

$$\langle r_\perp^2\rangle\alpha , \tag{5.38}$$

which measures the transverse width of the string in units of the physical length $1/\sqrt{\alpha}$.

It is useful to gain a broader perspective on roughening transitions before continuing with the discussion of non-Abelian gauge theories. Since roughening concerns the long-wavelength vibrations of strings, the internal compo-

FIG. 39. Lowest-energy transverse string excitation on a lattice.

sition of the string should be irrelevant to this phenomenon and simple models should expose all the essential physics. With this in mind, let us consider the three-dimensional Ising model. Calculations indicate that the roughening temperature T_R is considerably less than the bulk critical temperature T_c (Weeks *et al.*, 1973). If we use the duality relations developed in the fourth section, this means that the string roughens deep inside the confining region of the three-dimensional Ising gauge theory, Fig. 40. (Recall that the strong coupling $g \gg 1$ region of the gauge model maps onto the low-temperature region of the spin model.) Since the string tension $\alpha(g)$ maps onto the excess free energy in the interface sector of the Ising spin model, any structural phase transitions in the interface, such as roughening, should appear as nonanalyticities in $\alpha(g)$ for g well inside the confining phase where $\alpha(g)$ itself is large! These nonanalyticities are easy to understand—for $T < T_R$ there are no massless modes in the theory, but for $T_R \leq T < T_c$, massless, transverse excitations occur.

The physics of the roughening transition is made particularly transparent in the "solid-on-solid" approximation (Gilmer and Bennema, 1972; van Beijeren, 1975). Consider the interface at finite T and the fluctuations which contribute to the partition function. Now, neglect all bulk fluctuations which are not connected to the interface (the boundary between up and down spins). In addition, neglect fluctuations of the interface with overhangs. The acceptable fluctuations are shown in Fig. 41 for a continuum membrane. A cross-sectional slice of the fluctuating interface in the Ising model is shown in Fig. 42. Since no overhangs are allowed, the height of each column of "down" spins above or below the position of the interface at zero temperature $T=0$ is a single-valued, well-defined function. Call it n_{i^*} where i^* labels a "dual site" of the two-dimensional $T=0$ interface. Since only broken bonds in Fig. 42 contribute to the action, the partition function of the interface is simply

$$Z = \sum_{\{n_{i^*}\}} \exp\left[-\frac{4}{T}\sum_{\langle i^* j^*\rangle} |n_{i^*} - n_{j^*}|\right]. \tag{5.39}$$

Clearly, this is the natural lattice version of the thin string discussed in the preceding subsection!

How well does Eq. (5.39) approximate the real interface of the three-dimensional Ising model? Monte Carlo studies indicate that for $T \approx T_R$, the most likely values of $n_{i^*} - n_{j^*}$ are 0, $+1$, or -1 for nearest neighbors. Therefore, the neglect of overhangs appears justified. In addition, since T_R is much less than T_c ($T_R \approx \frac{1}{2} T_c$), bulk fluctuations are unlikely, as well. Since $n_{i^*} - n_{j^*} = 0, \pm 1$, it is tempting to make another approximation (Jose *et al.*, 1977),

$$Z = \sum_{\{n_{i^*}\}} \exp\left[-\frac{4}{T}\sum_{\langle i^* j^*\rangle} |n_{i^*} - n_{j^*}|^2\right]. \tag{5.40}$$

This is the "discrete Gaussian model." As we shall see in a different context in the next section, it is dual to the famous planar model. The classic work of Kosterlitz and Thouless (1973) therefore applies and indicates

(1) There are massless modes on the interface for $T > T_R$. They are delocalized for $T > T_R$.
(2) The excess free energy has an essential singularity, $\exp(-B/\sqrt{T - T_R})$, at the roughening temperature.

These clear results strengthen our belief that the physics behind the roughening transition is quite simple and general. Evidence for these results in more straightforward, but opaque, computer simulations and calculational programs is growing.

Now let's return to gauge theories and strings. The nonanalyticity of $\alpha(g)$ poses problems for Monte Carlo simulations and strong coupling expansions (Hasenfratz *et al.*, 1981; Itzykson *et al.*, 1980; Lüscher, 1981). The Monte Carlo simulations are subject to finite-size effects

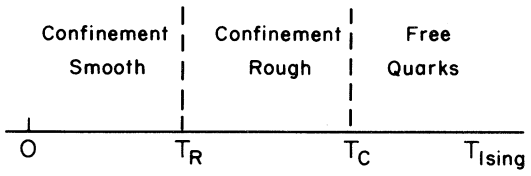

FIG. 40. The phases in the string sector of the three-dimensional Ising model.

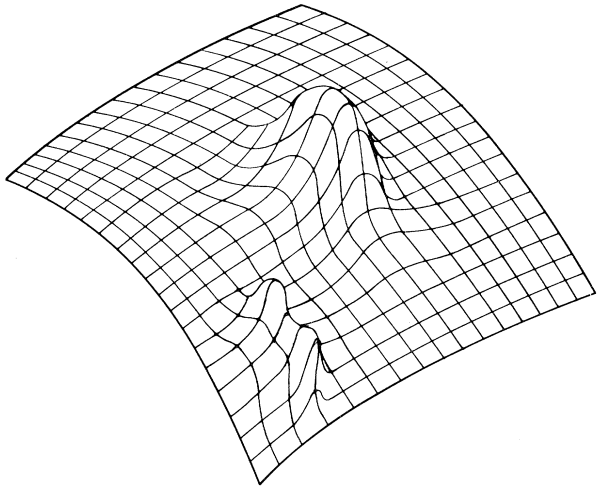

FIG. 41. Smooth fluctuations of an interface.

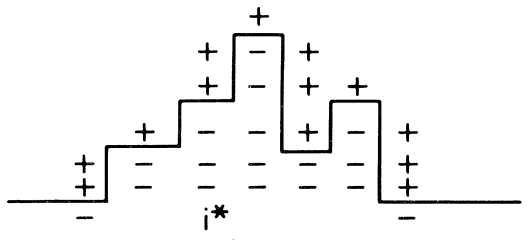

FIG. 42. A slice of a fluctuating interface in the three-dimensional Ising model. The dual-site variable i^* labels the column, height, of the interface.

because of the massless modes, and the strong coupling expansions have a radius of convergence bounded by g_R. Note that these structural phase transitions occur because of the nonlocal character of the matrix element, the loop correlation function, and that they occur *only* in the string sector of the theory.

One can discuss the roughening transition from the perspective of symmetry restoration (Lüscher, 1981). For $g \gtrsim g_R$ continuous translations perpendicular to the string become good symmetries of the system. This is a corollary to the fact that the transverse excitations of the interface become massless as g passes g_R from below—the *long*-wavelength surface vibrations have a relativistic massless energy-momentum relation for $g \gtrsim g_R$. This fact leads us to consider the restoration of full rotational symmetry in the string sector of the theory (Kogut *et al.*, 1981). For purposes of illustration consider the $(2+1)$-dimensional SU_3 gauge theory. Let the $Q\bar{Q}$ be off-axis, as shown in Fig. 43. At strong coupling the flux travels by a route of minimal distance. The leading order heavy-quark potential is

$$V(x,y)=\frac{g^2}{2a}\frac{4}{3}(|x|+|y|)\,, \tag{5.41}$$

which has equipotentials shown in Fig. 44. The equipotentials are not analytic near the axes. It is easy to see that this is because $m_k \neq 0$—the sharp edges in the equipotentials will disappear at g_R (Kogut *et al.*, 1981). But in general there is bad breaking of rotational symmetry. But rotational symmetry should be restored as $g \to 0$, in the scaling region of the lattice theory. Let us improve the large $g^2 \gg 1$ calculation of V and see this happen in perturbation theory.

We face a problem in degenerate perturbation theory—there are many, $(|x|+|y|)!/|x|!|y|!$, paths of equal length which contribute to the leading order calculation of V. To calculate the first-order correction to V we must diagonalize the perturbation in this subspace,

$$H = H_0 - x_s H'\,, \tag{5.42a}$$

where

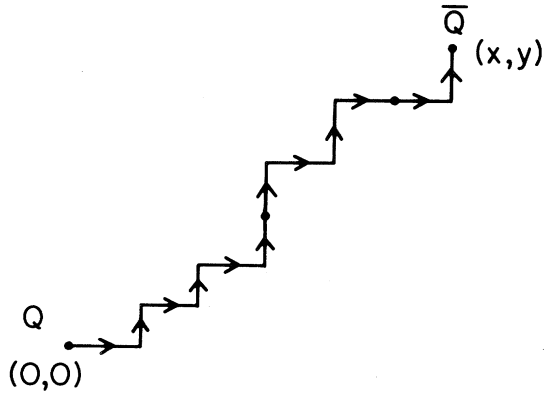

FIG. 43. A path of a flux tube between off-axis heavy quarks on a lattice.

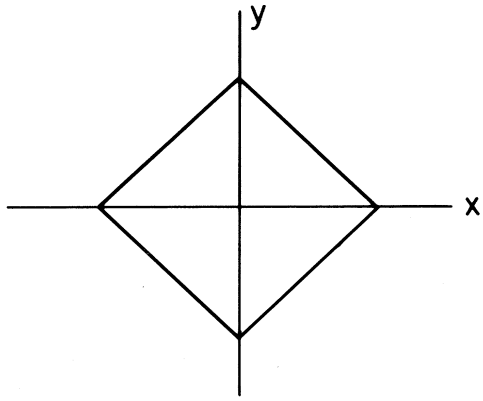

FIG. 44. An equipotential at strong coupling.

$$H_0=\frac{g^2}{2a}\sum_l E_l^2\,,\quad H'=\sum_{\text{plaq}}(\text{tr}UUUU+\text{H.c.})\,,$$
$$x_s=\frac{2}{g^4}\,. \tag{5.42b}$$

What is the effect of H' on paths of flux with corners as shown in Fig. 45? On the two links in Fig. 45 where two U matrices act, one must decompose the product into irreducible representations. Since $3\times\bar{3}=1+8$, there is a singlet piece and the flux path shown in Fig. 45 results with a weight of $\frac{1}{3}$. These are the processes of importance—they mix different members of the degenerate subspace. We recognize all this as a familiar problem—particles (fermions) hopping along a chain. To make this observation quantitative, label a path as shown in Fig. 46 and make the definitions

y link with flux$\to$fermion at site i ,

x link with flux$\to$absence of fermion at site i .

Thus we imagine a box of length $L=|x|+|y|$ containing $|y|$ fermions. It is clear from the perturbation theory exercises done above that the perturbation allows a "fermion" to hop one site to a nearest neighbor if that site is initially unoccupied. Therefore, the restriction of H' to the degenerate subspace is

$$H'=\tfrac{1}{3}\sum_i a_i^\dagger a_{i+1}+\text{H.c.}\,,\quad |y|=\sum a_i^\dagger a_i\,. \tag{5.43}$$

This standard one-dimension hopping Hamiltonian is diagonalized by passing to momentum space,

$$a_i=\sum_n a_n\phi_n(i),\phi_n(i)=\frac{1}{\sqrt{2(L+1)}}\sin\left[\frac{\pi n i}{L+1}\right]\,,$$
$$n=1,2,\ldots,L\,. \tag{5.44}$$

Then

$$H'=\tfrac{2}{3}\sum_n \cos\left[\frac{\pi n}{L+1}\right]a_n^\dagger a_n\,. \tag{5.45}$$

For our application, the first $|y|$ fermion levels are filled, so the heavy quark potential is

$$V(x,u)\simeq\frac{g^2}{2a}\left[\frac{4}{3}(|x|+|y|)-x_s\frac{2}{3}\sum_{n=1}^{|y|}\cos\left[\frac{\pi n}{L+1}\right]\right]$$
$$\simeq\frac{g^2}{2a}\left[\frac{4}{3}(|x|+|y|)-\frac{4}{3g^4}\left[\frac{\sin\left[\frac{\pi(|y|+1)}{2(|x|+|y|+1)}\right]\sin\left[\frac{\pi(|x|+1)}{2(|x|+|y|+1)}\right]}{\sin\left[\frac{\pi}{2(|x|+|y|+1)}\right]}-1\right]\right]. \quad (5.46)$$

It is interesting to plot the equipotentials of V. This is shown in Fig. 47. The restoration of rotational symmetry is clear. It occurs at a coupling close to the values where the string tension begins to scale according to asymptotic freedom. This is as expected and is very satisfactory. Higher-order corrections allow one to discuss smaller values of g than can sensibly be handled by our first-order formula (Kogut *et al.*, 1981). Computer simulations also show restoration of rotational symmetry (Stack, 1982).

A few final comments: One can expand the formula Eq. (5.46) for V in powers of $1/L$ and find power-law corrections. The strength of the $1/L$ term is close to that predicted by the thin string, continuum model discussed in the preceding section. And the off-axis matrix element, therefore, has all the qualitative properties of the continuum string! For all $g\neq\infty$, it is rough. Therefore, the tension of the off-axis string should be an analytic function of g^2, $0<g^2<\infty$. This point can be proved in models of fluctuating thin strings (Kogut and Sinclair, 1981). This makes it a more desirable matrix element—but more difficult to work with—than the on-axis string.

VI. TOPOLOGICAL EXCITATIONS AND CONFINEMENT IN LATTICE THEORIES

The purpose of this and the next several sections of this review is to identify the mechanisms in lattice gauge theories which lead to quark confinement. Topological excitations—vortices and monopoles—will play an important role here. These objects all have spin model analogs and are helpful in labeling and understanding phase diagrams in a host of systems. We will discuss systems of increasing complexity—the electrodynamics of the planar model in two dimensions, U(1) gauge theory in four dimensions, and finally SU(2) and SU(3) gauge theories in four dimensions. Our final topic in this line of thinking will be a discussion of the role of the topological charge in lattice systems which are asymptotically free.

A. Vortices and confinement in the electrodynamics of the planar model

Let us look at a very simple example in which topological excitations lead to confinement in a lattice theory. It will also provide an example in which confinement occurs in the continuum limit only for a definite range of the bare parameters of the cutoff theory.

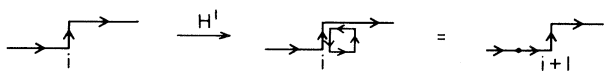

FIG. 45. The action of the perturbation H' on a kink in the string. The net result is the movement of the kink by one link.

Consider the Abelian Higgs model in 1+1 dimension and freeze the length of the Higgs field. The continuum Lagrangian is

$$L=|(\partial_\mu-igqA_\mu)\phi|^2-\tfrac{1}{4}F_{\mu\nu}^2, \quad (6.1)$$

where $\phi=\kappa e^{i\Theta}$ and q is the charge of the Higgs field. For fixed κ,

$$L=\kappa^2(\partial_\mu\Theta-gqA_\mu)^2-\tfrac{1}{4}F_{\mu\nu}^2. \quad (6.2)$$

When we make a gauge transformation, $D_\mu=A_\mu+(1/qg)\partial_\mu\Theta$, the Lagrangian becomes

$$L=-\tfrac{1}{4}F_{\mu\nu}^2+g^2q^2\kappa^2D_\mu^2, \quad (6.3)$$

which describes a free massive vector meson. If we couple external test charges into such a model, there will be just short-range, screened forces—obviously no confinement.

Now let us make a lattice action visualized in Fig. 48 based on Eq. (6.1) with periodicity in both Θ and $B_\mu=agA_\mu$ (Jones *et al.*, 1979),

$$S=2\kappa^2\sum_{r,\mu}\cos[\Delta_\mu\Theta(r)+qB_\mu(r)]+\frac{1}{2g^2a^2}\sum_{r,\mu,\nu}\cos(\Delta_\mu B_\nu-\Delta_\nu B_\mu). \quad (6.4)$$

The important properties of S are

(1) local gauge invariance, $B_\mu\to B_\mu+1/q\Delta_\mu\Lambda(r)$, $\Theta\to\Theta-\Lambda(r)$, and

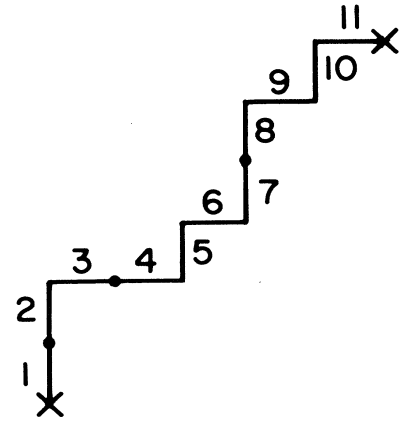

FIG. 46. An off-axis string with links labeled successively.

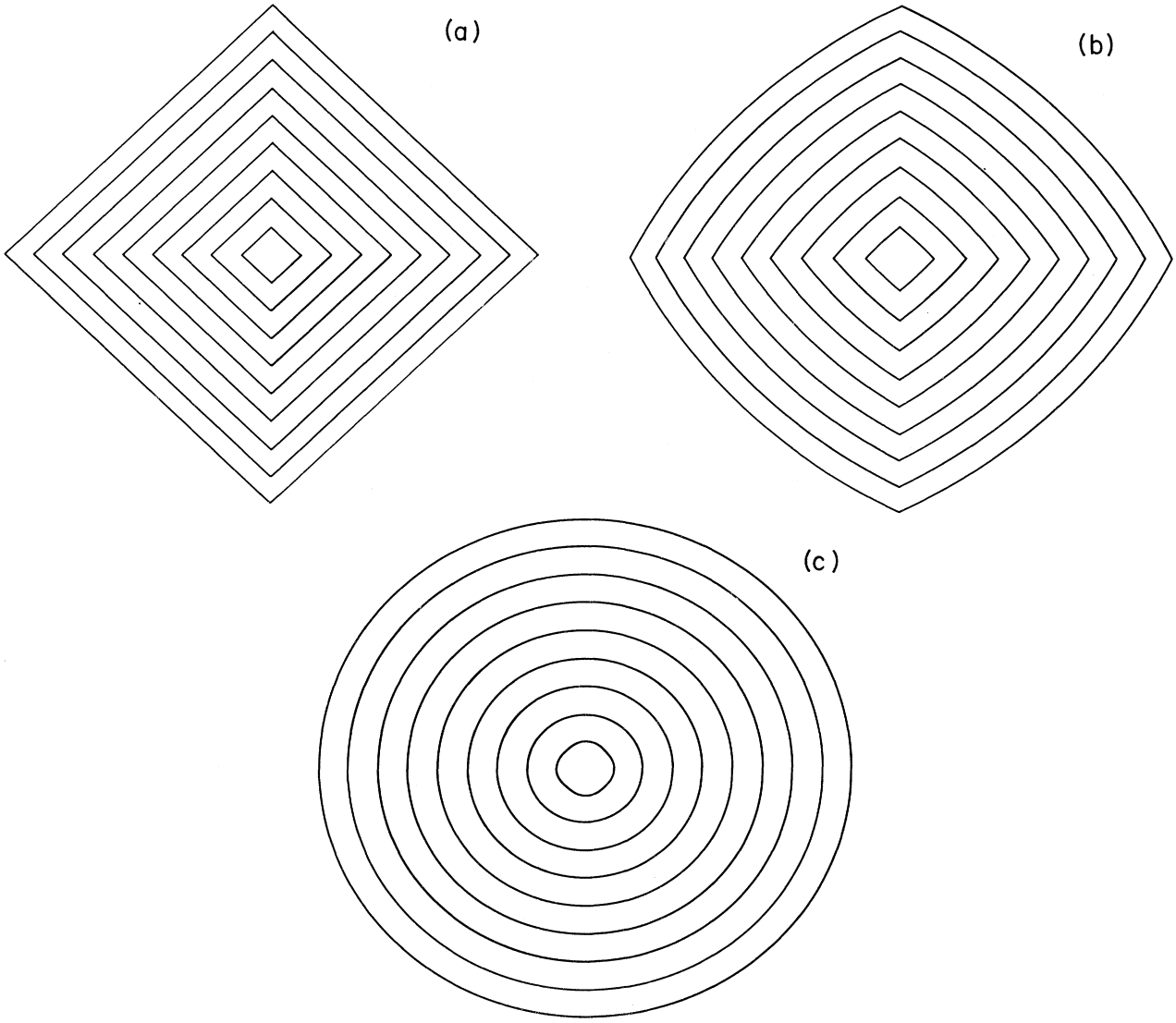

FIG. 47. The equipotentials at (a) $g=\infty$, (b) $g=1.19$, and (c) $g=1.00$.

(2) periodicity.

Because of property (2), we shall see that the theory has vortices which tend to disorder the theory leading to confinement for $a\neq 0$. However, in the continuum limit we shall also see that the effects of the vortices are felt only for $\kappa^2<\kappa_c^2$. For $\kappa^2>\kappa_c^2$ the vortices are suppressed in the continuum limit and the classical analysis sketched above is qualitatively reliable.

Our detailed analysis of this lattice model will use tricks borrowed from the two-dimensional planar model of statistical mechanics (Kosterlitz and Thouless, 1973). Before we begin a technical discussion, let's summarize the simple topological ideas behind it and the physics in the results. To test for confinement we place static charges of magnitudes p and $-p$ a distance r apart. For $p=1,2,\ldots,q-1$ we shall find a long-range confining potential, $V(r)\sim|r|$, between the charges. For $p=q$, however, the coefficient of the $|r|$ potential will vanish. The reason for this is screening—when the charge of the static external probe equals that of the dynamical charged fields of the theory, a dynamical charge binds to the static charge, making a chargeless state and the flux tube disappears. The fact that vortices can lead to this kind of

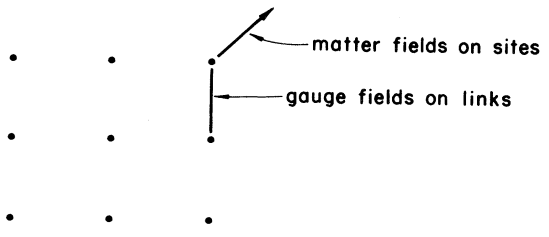

FIG. 48. The lattice degrees of freedom of the two-dimensional electrodynamics of the planar model. The variables $\Theta(r)$ reside on sites, and the gauge fields are on links.

physics is easy to understand. A vortex is a singularity in the field Θ with the property (Kosterlitz and Thouless, 1973)

$$\oint_C \nabla_\mu \Theta \, dx^\mu = 2\pi \tag{6.5}$$

if the closed contour C surrounds the singularity. Inspecting Eq. (6.4) we see that this field configuration will have finite action if

$$qg \oint_C A_\mu dx^\mu = 2\pi \ . \tag{6.6}$$

Now, consider the effect of such a vortex on the static impurities of charges p and $-p$. We must calculate the Wilson loop,

$$\left\langle \exp\left[ipg \oint A_\mu dx^\mu\right]\right\rangle = \frac{1}{Z}\int \left[\prod_r d\Theta(r)\right]\left[\prod_{r,\mu} dA_\mu(r)\right] e^{-S+ipg\oint A_\mu dx^\mu} \ . \tag{6.8}$$

Let the contour C be a rectangle of width R and length T. Then the heavy-quark potential is

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \ln\langle e^{ipg\oint A_\mu dx^\mu}\rangle \tag{6.9}$$

for an impurity of charge $p = 0,1,2,\ldots,q$.

It is convenient to label the contour C with a vector field $J_\mu(r)$:

$$J_\mu(r) = \begin{cases} +1 & \text{if link } r \to r+\mu \text{ is on } C \\ -1 & \text{if link } r+\mu \to r \text{ is on } C \\ 0 & \text{otherwise} \ . \end{cases} \tag{6.10}$$

Then we can write

$$pg \oint A_\mu dx^\mu \to p \sum_{r,\mu} B_\mu J_\mu \ . \tag{6.11}$$

We define

$$Z(J) = \int \left[\prod_r d\Theta(r)\right]\left[\prod_{r,\mu} dA_\mu(r)\right] e^{-S-ip\sum B_\mu J_\mu} \ . \tag{6.12a}$$

Then, as in previous sections, we can write

$$\left\langle \exp\left[ipg \oint A_\mu dx^\mu\right]\right\rangle = Z(J)/Z(0) \ . \tag{6.12b}$$

Now we want to do the functional integrals. We replace them with *periodic* Gaussian functions, rendering the integrals tractable and preserving the periodicity and local gauge invariance:

$$e^{-2\kappa^2[1-\cos(\Delta_\mu\Theta - qB_\mu)]} \to \sum_{n_\mu=-\infty}^{\infty} e^{in_\mu(\Delta_\mu\Theta - qB_\mu)} e^{-n_\mu^2/4\kappa^2} \ , \tag{6.13}$$

$$e^{-1/2g^2a^2[1-\cos(\Delta_\mu B_\nu - \Delta_\nu B_\mu)]} \to \sum_{l_{\mu\nu}=-\infty}^{\infty} e^{il_{\mu\nu}(\Delta_\mu B_\nu - \Delta_\nu B_\mu)} \times e^{-g^2a^2 l_{\mu\nu}^2/4} \ .$$

This is just a Fourier series analysis of the periodic functions on the left-hand side and the replacement of the Fourier coefficients with Gaussians in the conjugate variables. This nice trick is due to J. Villain (1975).

Now the integrals $\int \pi \, d\Theta, \pi \, dA_\mu$ are simple—they generate constraints among the conjugate variables,

$$Z(J) = \prod_{r,\mu\nu} \sum_{l_{\mu\nu}(r)} \prod_{r'\mu'} \delta[\Delta^\mu l_{\mu\nu}(r') + pJ_\mu(r') + qn_\mu(r')] \prod_{r''} \sum_{n_\mu(r'')} \prod_{r'''} \delta[\Delta^\mu n_\mu(r''')] \exp[-n_\mu^2(r'')/4\kappa^2 - g^2a^2 l_{\mu\nu}^2(r)/4] \ . \tag{6.14}$$

One can solve the Kronecker constraints. $\Delta^\mu n_\mu = \Delta^\mu J_\mu = 0$ implies that n_μ, J_μ are curls of integer-valued scalar fields,

$$n_\mu = \varepsilon_{\mu\nu}\Delta_\nu n \ , \quad J_\mu = \varepsilon_{\mu\nu}\Delta_\nu J \ . \tag{6.15}$$

A particular solution of the constraint $\Delta^\nu l_{\mu\nu} + pJ_\mu + qn_\mu = 0$ is

$$l_{\mu\nu} = m^\mu (m\cdot\Delta)^{-1}(pJ^\nu + qn^\nu) - m^\nu (m\cdot\Delta)^{-1}(pJ^\mu + qn^\mu) \ , \tag{6.16}$$

where $(m\cdot\Delta)^{-1}$ is a line integral in the m^μ direction. If $m^\mu = (0,1)$, then

$$(m\cdot\Delta)^{-1} f(r_1,r_2) = \sum_{m'=-\infty}^{r_2} f(r_1,m') \ . \tag{6.17}$$

The choice of m will not affect the physics. The Gaussian factors in Eq. (6.13) are nicely written in terms of n and J:

$$n_\mu n_\mu=(\Delta_\mu n)^2\ ,\qquad l_{\mu\nu}l_{\mu\nu}=2(pJ+qn)^2\ . \tag{6.18}$$

So, finally,

$$Z(J)=\sum_{n(r)=-\infty}^{\infty}\exp\left[-\frac{1}{4\kappa^2}\sum_r[\Delta_\mu n(r)]^2-\frac{q^2g^2a^2}{2}\sum_r\left[n(r)+\frac{p}{q}J(r)\right]^2\right]. \tag{6.19}$$

This is a discrete Gaussian model. For $g=0$, it reduces to the interface roughening model which is dual to the planar model, as mentioned in the preceding section (Jose *et al.*, 1977).

If $g^2a^2\gg 1$, the interface roughening model, Eq. (6.19), is a good expression of the physics of the model, because the sum over $n(r)$ can be truncated. For $g^2a^2\to 0$, the continuum limit, this representation is not convenient. We do another Fourier transformation using the Poisson summation formula,

$$\sum_{l=-\infty}^{\infty}h(l)=\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}d\phi\,h(\phi)e^{2\pi im\phi}\ . \tag{6.20}$$

Applying this, we obtain

$$Z(J)=\int_{-\infty}^{\infty}\left[\prod_r d\phi(r)\right]\sum_{m(r)=-\infty}^{\infty}\exp\left[-\frac{1}{4\kappa^2}\sum_r(\Delta_\mu\phi)^2-\frac{g^2a^2q^2}{2}\sum_r\left[\phi+\frac{p}{q}J\right]^2+2\pi i\sum_r m(r)\phi(r)\right], \tag{6.21}$$

and this form of the partition function will prove most useful. Set $J=0$ here, and we notice that it describes a free *massive* scalar field coupled to an integer-valued field. We recognize the mass $q^2g^2a^2$ as due to the Higgs mechanism as in the continuum analysis. The $m(r)$ will be identified with vortices which tend to disorder the system. They destroy the Higgs mechanism and replace it with confinement.

Let us evaluate $Z(J=0)$. ϕ appears only in a quadratic form, so the integral is easy,

$$Z(J=0)=\sum_{m(r)=-\infty}^{\infty}\exp\left[-4\pi^2\kappa^2\sum_{r,r'}m(r)G(r-r';m_va)m(r')\right], \tag{6.22}$$

where $m_v^2=2q^2g^2\kappa^2$ and

$$G(r;m_va)=\int_{-\pi}^{\pi}\frac{dk_x}{(2\pi)}\int_{-\pi}^{\pi}\frac{dk_y}{(2\pi)}\frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{(4-2\cos k_x-2\cos k_y+m_v^2a^2)}\ , \tag{6.23}$$

which is the lattice version of the massive propagator, $(\nabla^2+m_v^2a^2)^{-1}$. So the vortices interact through a *short-range,* screened potential. This will lead to an important simplification. For a range of parameters the vortex concentrations will be small. Then they will not often lie within a distance $1/m_va$ of one another. In those cases, Z is well approximated by just keeping the $r=r'$ term in the double sum. The self-energy of a vortex is

$$G(0,m_va)=-\frac{1}{2\pi}\ln(m_va)+\text{const} \tag{6.24}$$

and

$$Z(J=0)\approx\sum_{m(r)}\exp\left[2\pi\kappa^2\ln(m_va)\sum_r m(r)^2\right]. \tag{6.25}$$

This equation means that there is an activation energy $2\pi\kappa^2\ln(m_va)$ for a vortex and a probability of finding one at a specified site of $(m_va)^{2\pi\kappa^2}$. Two vortices interact only if they lie within $1/m_va$ of one another. Given one at site 1, the probability that a second one lies close enough to interact with it is

$$(m_va)^{2\pi\kappa^2}(m_va)^{-2}=(m_va)^{2\pi\kappa^2-2}\ . \tag{6.26}$$

If this is a small number, a dilute gas approximation is trustworthy. Note that as $a\to 0$, this condition is violated for $\kappa^2<1/\pi$. So, we could naively guess that for $\kappa^2>1/\pi$, the continuum limit would be free of the effects of vortices, and for $\kappa^2<1/\pi$ it would not.

Now let's verify that vortices disorder the system. We return to the expression $Z(J)$. Let $\phi\to\phi+(p/q)J$. Recalling that J is a step function, changing variables and dropping terms which could not contribute $\sim RT$ to the exponent, we have

$$Z(J)=\sum_{m(r)}\exp\left[2\kappa^2\pi\ln(m_va)\sum_r m^2(r)-2\pi i\frac{p}{q}\sum_r^{\text{inside}}m(r)\right]. \tag{6.27}$$

If a is small, the vortex activation energy is large and the sum over $m(r)$ can be truncated at $0,\pm1$. This is why the Poisson summation is useful. Now,

$$\sum_{m(r)} \exp\left[2\pi\kappa^2 \ln(m_v a) m^2(r) - 2\pi i \frac{p}{q} m(r)\right] = 1 + 2e^{2\pi\kappa^2 \ln(m_v a)} \cos\left[2\pi \frac{p}{q}\right]$$
$$\approx \exp\left[2e^{2\pi\kappa^2 \ln(m_v a)} \cos\left[2\pi \frac{p}{q}\right]\right] . \tag{6.28}$$

So,

$$\frac{Z(J)}{Z(0)} \approx \exp\left\{2(m_v a)^{2\pi\kappa^2} \sum_{r}^{\text{inside}} \left[\cos\left[2\pi \frac{p}{q}\right] - 1\right]\right\} \tag{6.29}$$

or,

$$\frac{Z(J)}{Z(0)} \approx \exp\left[-2(m_v a)^{2\pi\kappa^2}\left[1 - \cos 2\pi \frac{p}{q}\right] RT\right] , \tag{6.30}$$

and the heavy-quark potential is

$$V(R) = 2m_v^2 (m_v a)^{2\pi\kappa^2 - 2}\left[1 - \cos 2\pi \frac{p}{q}\right] R , \tag{6.31}$$

where we have restored physical units.

So, we have confinement.

(1) If $p = 1, 2, \ldots, q-1$, the sources experience a linear confining potential.

(2) If $p = q$, the string tension vanishes—a dynamical charged particle binds to the impurity and there are no long-range effects.

Note that if $\kappa^2 > 1/\pi$, the tension vanishes in the continuum limit. This is the same condition exposed before—the vortices become infinitely dilute and disappear—the periodicity in the lattice variables become irrelevant, and the naive classical continuum ideas are qualitatively reliable.

For $\kappa^2 < 1/\pi$, this analysis fails—the vortices interact and must be treated dynamically. Numerically, work on the model indicates that the vortices form a plasma for $\kappa^2 < 1/\pi$ and that the string tension saturates rather than diverges in the $a \to 0$ limit (Jones *et al.*, 1979). Then confinement is a property of the continuum limit. A self-consistent screening theory *à la* Kosterlitz-Thouless or a Kosterlitz renormalization group must be made.

B. U(1) gauge theory in four dimensions

This theory, a lattice version of quantum electrodynamics, confines for strong coupling as can be checked using the methods of Sec. IV. For weak coupling we expect it to reduce to ordinary electrodynamics, which does not confine. So, it should be a two-phase system with a critical coupling g_c. We can ask

(1) What is g_c? Is the phase transition second order? Is there a continuum confining theory there?

(2) What is the nature of the confining mechanism? How does it drive the transition at g_c?

Using the tricks developed in the preceding section, we can answer these questions. Consider Abelian U(1) lattice gauge theory in four dimensions (Banks *et al.*, 1977),

$$S = \beta \sum_{r,\mu,\nu} [1 - \cos\Theta_{\mu\nu}(r)] , \tag{6.32}$$

where $\Theta_\mu(r)$ is an angular variable on the link $r \to r + \mu$ and $\Theta_{\mu\nu}(r) = \Delta_\mu \Theta_\mu(r) - \Delta_\nu \Theta_\mu(r)$, as shown in Fig. 49. We want to compute the excess free energy when a pair of quarks is placed in the system, so we consider the loop correlation function,

$$\left\langle \exp\left[ie \oint_C A_\mu dx^\mu\right]\right\rangle = \frac{Z(J)}{Z(0)} , \tag{6.33}$$

and the heavy-quark potential

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \ln\left\langle \exp\left[ie \oint_C A_\mu dx^\mu\right]\right\rangle . \tag{6.34}$$

We need $Z(J)$ as usual. As in the $(1+1)$-dimension problem, we write

$$Z(J) = \int_0^{2\pi} \left[\prod_{r,\mu} d\Theta_\mu(r)\right] \exp\left[-\beta \sum_{r,\mu,\nu} [1 - \cos\Theta_{\mu\nu}(r)] + i \sum_{r,\mu} \Theta_\mu(r) J_\mu(r)\right] \tag{6.35}$$

and make the Villain replacement,

$$e^{-\beta(1-\cos\Theta_{\mu\nu})} \to \sum_{l_{\mu\nu}=-\infty}^{\infty} e^{il_{\mu\nu}\Theta_{\mu\nu}} e^{-l_{\mu\nu}^2/2\beta} . \tag{6.36}$$

We calculate the $\int d\Theta_\mu(r)$ integrals generating Kronecker symbol constraints,

$$Z(J) = \sum_{l_{\mu\nu}(r)=-\infty}^{\infty} \left[\prod_{r,\mu} \delta[\Delta_\nu l_{\nu\mu}(r) + J_\mu(r)]\right] \exp\left[-\sum_{r,\mu,\nu} l_{\mu\nu}^2(r)/2\beta\right] , \tag{6.37}$$

and solve the constraint,

$$l_{\mu\nu} = n_\mu (n\cdot\Delta)^{-1} J_\nu - n_\nu (n\cdot\Delta)^{-1} J_\mu + \varepsilon_{\mu\nu\lambda\kappa} \Delta_\lambda l_\kappa , \tag{6.38}$$

where n_μ is an arbitrary unit vector which will disappear from the formalism soon. We substitute into $Z(J)$. There is a

sum over the integral-valued field l_κ. We use the Poisson formula and change variables from l_κ to ϕ_κ and an integer-valued vector field m_κ. The $[d\phi_\kappa]$ integral is Gaussian, so we do it in the usual way. Finally (Banks *et al.*, 1977),

$$Z(J)=\exp\left[-\frac{1}{2\beta}\sum_{rr'}J_\mu(r)v(r-r')J_\mu(r')\right]$$
$$\times\sum_{m_\mu(r)=-\infty}^{\infty}\left[\prod_r\delta[\Delta_\mu m_\mu(r)]\right]\exp\left[-2\pi^2\beta\sum_{r,r'}m_\mu(r)v(r-r')m_\mu(r')\right.$$
$$\left.+2\pi i\sum_{rr'}m_\mu(r)\varepsilon_{\mu\nu\lambda\kappa}n_\nu v(r-r')\Delta_\lambda(n\cdot\Delta)^{-1}J_\kappa(r')\right], \tag{6.39}$$

where $v(r)$ is the four-dimensional massless propagator. We can interpret the conserved vector field $m_\mu(r)$ as closed loops as in Fig. 50 because $\Delta_\mu m_\mu=0$. These are the world lines of *magnetic monopoles* (Banks *et al.*, 1977). Note that the electric loops in Eq. (6.39) interact with strength $g_E^2=1/2\beta$ and that the magnetic loops interact with strength $g_m^2=2\pi^2\beta$. Therefore, $g_E g_m=\pi$, the familiar reciprocal relation of Dirac. Also, m_μ interacts with $\Delta_\lambda\tilde{F}_{\mu\lambda}$, the divergence of the dual of the field strength tensor. So, m_μ is a source of $\Delta_\lambda\tilde{F}_{\mu\lambda}$ which we recognize as *magnetic current* ($\Delta_\lambda F_{\mu\lambda}$ is electric current).

We see a natural disordering mechanism here. At large β, weak coupling, the loops are suppressed by a large activation energy. Presumably, they are then irrelevant for the large scale behavior of the theory which reduces to free, nonperiodic QED, without confinement. However, for small β, large coupling, the magnetic loops are not suppressed. They are macroscopic in length, and there is a finite probability of finding a monopole at any three-cube. These monopoles disorder the system, suppressing $Z(J)$ leading to the area law and confinement—like in the $(1+1)$-dimension model, presumably. They provide an electric Meissner effect (Mandelstam, 1976; 't Hooft, 1975).

One can estimate the critical coupling in the U(1) gauge theory following Kosterlitz and Thouless (1973). We consider a loop of L steps. Its activation energy is $2\pi^2\beta v(0)L$. The number of loops of length L through a given point behaves as 7^L, roughly. The free energy of such a loop is

$$F=-ST+E=2\pi^2v(0)L-T\ln 7^L$$
$$=L[2\pi^2v(0)-T\ln 7], \tag{6.40}$$

which becomes negative at

$$T_c=\frac{2\pi^2v(0)}{\ln 7}. \tag{6.41}$$

Since F becomes negative for $T\geq T_c$, the system becomes a monopole loop condensate for these couplings. Since $v(0)\approx 0.155$, $T_c=g_c^2\approx 1.57$.

In this picture we have a monopole condensate and confinement for $g^2>1.57$. This crude argument has neglected $1/R$ potentials in Eq. (6.39) and has overcounted loops, etc. These approximations appear to be subdominant effects, however, if the transition occurs where the loops are sufficiently dilute (Banks *et al.*, 1977).

This physical picture has good support from Monte Carlo simulations (DeGrand and Toussaint, 1980). Even the estimate of T_c is quite acceptable.

C. Monopoles, crossover, and confinement in non-Abelian models in four dimensions

What are the essential ingredients in the pure SU(2) gauge theory that lead to confinement?

Some insight into this question can be gained by considering variant actions on the lattice. For SU(2),

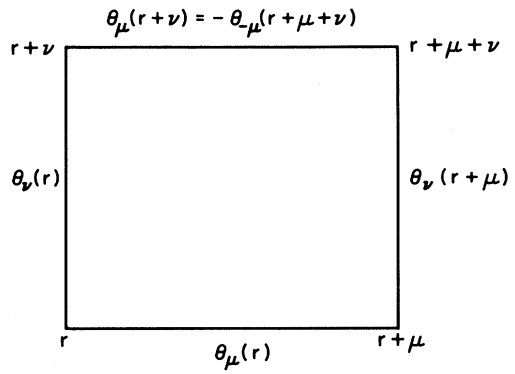

FIG. 49. Gauge field degrees of freedom in compact QED.

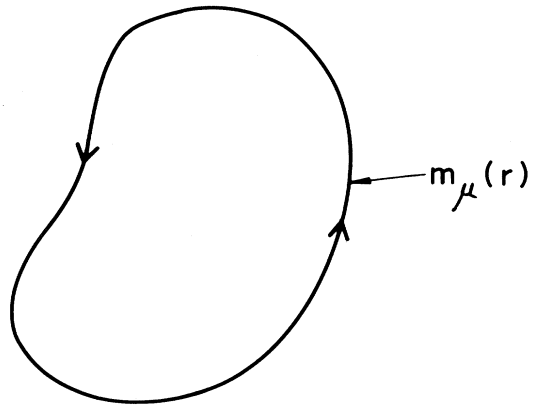

FIG. 50. A closed monopole loop.

$$S=\sum_p \mathrm{tr}_F U_p \ , \tag{6.42}$$

where tr_F is the trace (or character) in the fundamental representation of the group. Of course, in continuum theories only the Lie group is significant, not the representation. However, strong cutoff theories are sensitive to the global structure of the group, so they depend on the precise form of the action. For example, choose

$$S=\sum_p \mathrm{tr}_A U_p \ , \tag{6.43}$$

where tr_A is the trace (or character) in the adjoint representation. The smooth field, continuum limits of both theories Eqs. (6.42) and (6.43) are identical. The first is SU(2) lattice gauge theory, the second is an $SU(2)/Z(2)=SO(3)$ theory. Locally these groups are the same, but they have different global characters. The SO(3) theory has *monopoles* with finite action, while the SU(2) theory does not. A particularly convenient SO(3) action is (Halliday and Schwimmer, 1981)

$$Z=\sum_{\sigma(p)=\pm 1}\int [dU]\exp\left[\beta\sum_p \mathrm{tr}U_p\sigma(p)\right] , \tag{6.44}$$

where the tr is in the fundamental and we have added a Z_2 variable on plaquettes. Now the sign of $\mathrm{Tr}U_p$ is irrelevant, so the local gauge group is SO(3) as claimed.

Let us compare configurations and energetics between SU(2) and SO(3) (Brower *et al.*, 1981). We let $\eta(p)=\mathrm{sgn}\,\mathrm{tr}U_p$ and consider a set of plaquettes having $\eta(p)=-1$ in a configuration with all other plaquettes having $\eta(p)=+1$. In an Abelian model

$$\prod_p U=\exp\left[ie\oint_C \mathbf{A}\cdot d\mathbf{x}\right]=\exp\left[ie\int_s d\mathbf{S}\cdot\mathbf{B}\right] ,$$

so it is natural to identify one unit of Z_2 flux going along along a "string" piercing the $\eta(p)=-1$ plaquettes in Fig. 51. The string ends, and around either end we can construct a three-cube and consider the quantity

$$\eta(\partial c)=\prod_{p\in\partial c}\eta(p) \ . \tag{6.45}$$

Since all plaquettes other than those shown in Fig. 51 have $\eta(p)=+1$, $\eta(\partial c)=-1$ only on *two* cubes at the ends of the string. It is natural to make the identification

$$\eta(\partial c)=e^{ie\Phi} \ , \tag{6.46}$$

where Φ is the magnetic flux, so the ends of the strings can be thought of as sources of magnetic flux—we have a monopole pair with a string between them in Fig. 51!

Note that $\eta(\partial c)=-1$ is not possible in a purely Abelian model—links occur twice in the product Eq. (6.45) and $\eta(\partial c)=+1$ for *all* configurations—so these monopoles occur because of the *non-Abelian* character of the lattice action.

We have exposed two topologically significant objects:

(1) *Vortices*—closed strings. They cost an action which grows as the string length in SU(2), but they are invisible (do not occur) in SO(3) gauge theories.

(2) *Monopoles*—They occur singly in SO(3), but only in bound pairs with an action which grows as the length between them in SU(2).

We could also make *thick vortices* and *monopoles*. They are expected to be relevant at weak coupling, as will be discussed briefly below.

Let's do all this more formally (Halliday and Schwimmer, 1981). For the SO(3) model,

$$Z=\sum_{\sigma(p)=\pm 1}\int \pi\, dU\exp\left[\beta\sum_p \mathrm{tr}U_p\sigma(p)\right] . \tag{6.47}$$

We write

$$\sigma(p)=e^{i\pi n_{\mu\nu}} \ , \tag{6.48}$$

where $n_{\mu\nu}=0$ or ± 1, $n_{\mu\nu}=-n_{\nu\mu}$ for the plaquette shown in Fig. 52. We can write the field $n_{\mu\nu}$ as the sum of two pieces—a curl and a monopole current density,

$$n_{\mu\nu}=\Delta_\mu m_\nu-\Delta_\nu m_\mu+\varepsilon_{\mu\nu\rho\sigma}\frac{a^\rho M^\sigma}{a\cdot\Delta} \ , \tag{6.49}$$

with

$$M^\sigma=\varepsilon^{\sigma\rho\mu\nu}\Delta_\rho n_{\mu\nu} \ ; \tag{6.50}$$

and a^ρ is an arbitrary vector on the dual lattice along the string. M^ρ is divergenceless,

$$\Delta_\rho M^\rho=0 \ , \tag{6.51}$$

and m_μ is perpendicular to the string, $a_\mu m^\mu=0$. This implies that

$$a^\mu n_{\mu\nu}=a\cdot\Delta m_\mu \ , \tag{6.52}$$

so

$$m_\nu=a^\mu n_{\mu\nu}/a\cdot\Delta \tag{6.53}$$

which insures the integrability and single valuedness of m_ν.

Now we substitute $\sigma(p)=\exp(i\pi n_{\mu\nu})$ into Z, absorb

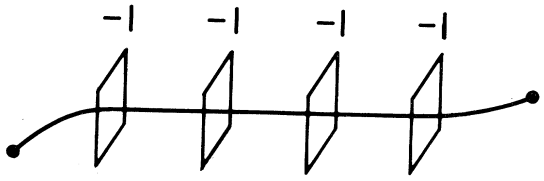

FIG. 51. A thin Dirac string and frustrated plaquettes.

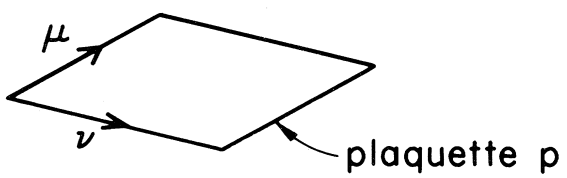

FIG. 52. A plaquette p with orientation vectors μ and ν.

$\Delta_\mu m_\nu - \Delta_\nu m_\mu$ into $\mathrm{tr}U_p$, and find

$$Z = \sum_{M^\sigma} \int [dU] \exp\left[\beta \sum_p \mathrm{tr}U_p \exp(\pi i \varepsilon_{\mu\nu\rho\sigma} a^\rho M^\sigma / a \cdot \Delta)\right] . \tag{6.54}$$

We can identify M^ρ as a conserved (mod2) monopole current. We return to the monopole counter for this model,

$$\eta(\partial c) = \prod_{p \in \partial c} \sigma(p) . \tag{6.55}$$

At low $1/\beta$, $\mathrm{tr}U_p$ and σ_p will have identical signs, so this monopole counter definition is natural for this action. Choosing the cube to be at equal time, we let a^μ point in the z direction. It is then trivial to compute

$$\eta(\partial c) = \exp(\pi i M_4) . \tag{6.56}$$

So, wherever $M_4 = \pm 1$, there is a non-Abelian monopole!

The SO(3) model therefore has monopole loops which are coupled to gauge fields, as shown in Fig. 53.

So it is plausible, based on our experience with compact QED, that this model will have a phase transition which the SU(2) model will not experience!

By suppressing the monopoles we can make a model which interpolates between SU(2) and SO(3). Consider (Halliday and Schwimmer, 1981)

$$S = \beta \sum_p \mathrm{tr}U(\partial p)\sigma(p) + \mu \sum_p \sigma(\partial c) . \tag{6.57}$$

The added term is a "monopole activation energy." We can examine this theory's phase diagram in the (β,μ) plane.

(1) $\beta = 0$. It reduces to a hypergauge theory which is dual to the four-dimension Ising model (Wegner, 1971; Balian *et al.*, 1975b). It has a second-order phase transition. In the language of this section, the phase transition is a result of the competition between the chemical potential of the loops and their entropy.

(2) $\beta \to \infty$. The theory becomes frozen (pure gauge transformation) and is trivial.

(3) $\mu = 0$. Pure SO(3) again.

(4) $\mu \to \infty$. Then $\sigma(\partial c) \to 1$, monopoles are suppressed ($M^\sigma \to 0$, $n_{\mu\nu} \to \Delta_\mu m_\nu - \Delta_\nu m_\mu$), and the SU(2) theory results.

The expected phase diagram is shown in Fig. 54. This phase structure was confirmed by crude Monte Carlo studies (Halliday and Schwimmer, 1981). The transition is first order for the pure SO(3) axis and also along the transition line for small μ.

For the $\mu = 0$ model one can calculate the density of monopole current,

$$M \equiv 1 - \langle \sigma(\partial c) \rangle . \tag{6.58}$$

For small β,

$$M = 1 - 4\left[\frac{I_2(\beta)}{I_1(\beta)}\right]^6 + O(\beta^{18}) . \tag{6.59}$$

For large β, $\sigma(\partial c)$ is $+1$ on the dominant saddle point and $M = 0$, to all orders in β^{-1}. These large- and small-β results again suggest the condensation mechanism and a phase transition.

How important are monopoles in the pure SU(2) model? We define another density of monopole current,

$$\bar{M} \equiv 1 - \left\langle \prod_{p \in \partial c} \eta(p) \right\rangle ,$$
$$\eta(p) = \mathrm{sgn}[\mathrm{tr}U(\partial p)] , \tag{6.60}$$

as discussed earlier, and take the mixed model at $\mu = 1$. Monte Carlo data suggest that its monopole content is almost identical to the usual SU(2) model (Halliday and Schwimmer, 1981). $\bar{M}$ changes dramatically in the crossover region, as shown in Fig. 55. It can be fit with an exponential for $\beta \gtrsim 2.0$, $e^{-2.1\beta}$.

One can also make a vortex counter. A thin vortex frustrates, $\mathrm{sgn}\,\mathrm{tr}U_p < 1$, plaquettes on a closed path. The smallest closed path of this variety circulates through six plaquettes in four dimensions. We can look for a vortex

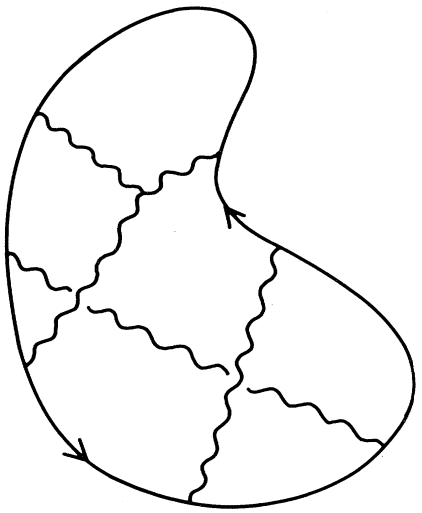

FIG. 53. A closed monopole current loop coupled to a gauge field.

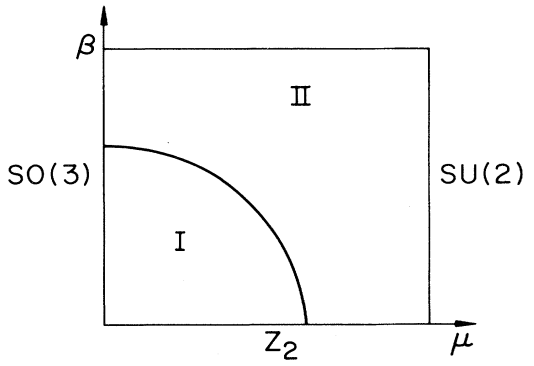

FIG. 54. The phase diagram of the generalized SO(3) model.

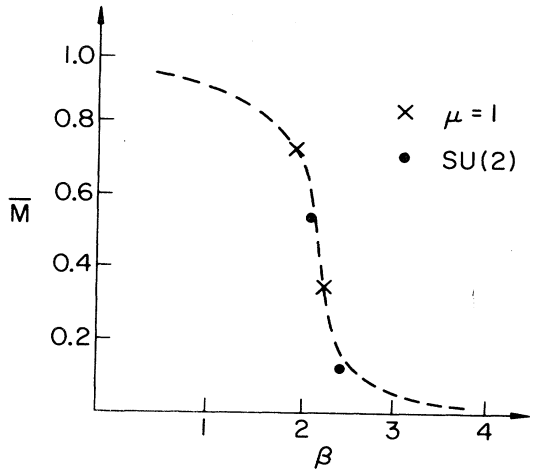

FIG. 55. The monopole density in the generalized model at $\mu=1$ and in the pure SU(2) model.

at this length scale with the operator

$$E_p=\text{sgn}\,\text{tr}U(p)\ . \tag{6.61}$$

The density of these small, thin vortices is large at strong coupling and falls rapidly in the crossover region of the pure SU(2) model (Halliday and Schwimmer, 1981).

All this still leaves us with the questions of how the SU(2) theory avoids a phase transition and why the SU(2) model confines at weak coupling? The idea which emerges from these studies is that as β increases there are vortices of greater size and thickness (Mack and Petkova, 1980). They are condensed and disorder the loop correlation function on all length scales. To see these larger topological objects on the lattice we consider the rectangles shown in Fig. 56. If $\text{tr}U(P_{\text{rec}})$ has $\text{sgn}\,\text{tr}U(P_{\text{rec}})<0$, there is a "thick" string passing through it. Since in the notation of Fig. 56

$$\text{tr}U(P_{\text{rec}})\neq\text{tr}U(P_{\text{sq.no.1}})\,\text{tr}U(P_{\text{sq.no.2}})\ , \tag{6.62}$$

the presence of a thick vortex is independent of the thin ones. This is a specifically non-Abelian effect!

It appears that small monopoles disappear from the pure SU(2) theory at $\beta\simeq 2.0$ and then asymptotic freedom sets in and growing topological objects dominate. This is an intriguing conjecture deserving more sudy.

Another two-parameter model which addresses the same issues is

$$S=\beta\sum_p\left[1-\tfrac{1}{2}\,\text{tr}_F U(p)\right]+\beta_A\sum_p\left[1-\tfrac{1}{3}\,\text{tr}_A U(p)\right]\ , \tag{6.63}$$

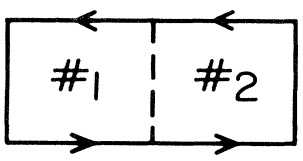

FIG. 56. A six-link rectangular loop as a probe for "thick" strings.

where $\text{tr}_A U(p)=|\,\text{tr}_F U(p)\,|^2-1$. Consider the $\beta-\beta_A$ phase plane. The limit $\beta_A\to\infty$ reduces the finite energy fluctuations in the model to those having $\text{tr}_F U(p)=\pm 1$. The theory reduces to Z_2 gauge theory. The phase diagram is shown in Fig. 57. The transition lines are first order in nature. The Z_2 transition occurs at $\beta=\frac{1}{2}\ln(1+\sqrt{2})=0.44\ldots$ by a familiar duality argument (Wegner, 1971; Balian *et al.*, 1975b).

Crossover from weak to strong coupling in the pure SU(2) model occurs for $\beta\approx 2.2$. There is a smooth but large specific-heat peak in that region (Nauenberg and Lautrup, 1980). We see from Fig. 57 that this behavior is a remnant of the end of the line of first-order transitions in the mixed model, Eq. (6.63) (Bhanot and Creutz, 1981)!

Increasing β_A suppresses the monopoles—they are not condensed in the high-β_A region of Fig. 57. Increasing β suppresses the vortices. Across the first-order line both density operators experience a discontinuous drop (Bhanot and Creutz, 1981).

The specific heat peak of the pure SU(2) model, Eq. (6.42), is associated with the sharp drop of the monopole and vortex concentrations. These small topological objects decouple abruptly at $\beta\simeq 2.0$, causing the specific-heat peak. It is believed that once they have disappeared only larger topological excitations remain, whose structure is less distorted by the lattice. Their population is thought to coexist with asymptotic freedom.

D. Topology and continuity in asymptotically free lattice theories

Let us recall the O(3) sigma model in 1 + 1 dimensions. The classical action is

$$S_{\text{cl}}=\tfrac{1}{2}\int d^2x\,\partial_\mu\mathbf{s}\cdot\partial_\mu\mathbf{s}\ ,\quad |\,\mathbf{s}(x)\,|^2=1\ , \tag{6.64}$$

where $\mathbf{s}(x)$ is a unit vector in three-space. Clearly, $\mathbf{s}(x)$ would be a free field were it not for the constraint $|\,\mathbf{s}(x)\,|^2=1$. This theory has several surprising proper-

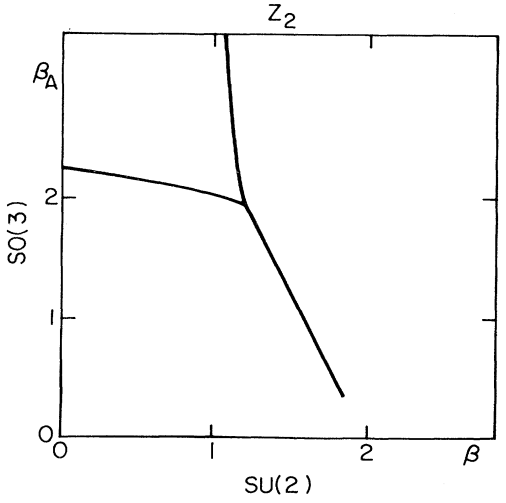

FIG. 57. Phase diagram of a mixed SO(3)-SU(2) model.

ties. It is asymptotically free, and it has a nontrivial topological charge Q, which is our main concern here (Polyakov, 1975). It is given by the expression

$$Q=\frac{1}{4\pi}\int d^2x\ \mathbf{s}\cdot(\partial_1\mathbf{s}\times\partial_2\mathbf{s}) \tag{6.65}$$

and can assume only integer values, $Q=0,\pm1,\pm2,\ldots$, which label distinct sectors of the theory. The different topological sectors cannot be connected by continuous deformations of the field $\mathbf{s}(x)$. This result can be understood by exposing the geometric meaning of Q. We take a boundary condition on $\mathbf{s}$ so that $\mathbf{s}\to(1,0,0)$ when $x\to\infty$ in any direction. The base space of the model is then essentially a two-sphere, S_2. The field $\mathbf{s}$ becomes a mapping of the base space S_2 to the sphere of $|\mathbf{s}|=1$. Such mappings can cover the unit sphere an integral number of times, and we recognize Q as computing this number.

A field configuration with $Q=\pm1$ can be drawn easily, as in Fig. 58. It is given by (Susskind, 1977)

$$\begin{aligned} s_1&=1-2e^{-x^2}\ ,\\ s_2&=x_1 f(x^2)\ ,\\ s_3&=x_2 f(x^2)\ ,\quad f(x^2)=\frac{4}{x^2}e^{-x^2}(1-e^{-x^2})\ . \end{aligned} \tag{6.66}$$

Note that $s(0)=(-1,0,0)$ and that as x covers the base space, $\mathbf{s}$ covers S_2 just once.

Field configurations having $Q=+1$ which are local minima of the action are instantons (Polyakov, 1975). Their classical action is for the O(3) model

$$S_{\text{inst}}=4\pi \tag{6.67}$$

and for the generalized CP^{n-1} models (Eichenherr, 1978; d'Adda *et al.*, 1978),

$$S_{\text{inst}}=2\pi n\ . \tag{6.68}$$

Finally, we recall the asymptotic freedom scaling law for the O(3) model. We consider a Pauli-Villars cutoff M and a coupling $g^2(M)$. Then the correlation length ξ, the reciprocal of the mass gap m, depends on M and $g^2(M)$, as (Polyakov, 1975)

$$\xi=\frac{1}{m}\propto\frac{1}{M}g^2e^{2\pi/g^2}\ . \tag{6.69}$$

Now let us consider the lattice model. Do any of these topological ideas survive? Since continuity concepts are not so clear on a lattice, one naively would think not. However, if we specialize to asymptotically free models where the critical point is at zero coupling, some remnants of the ideas do survive. In the vicinity of the critical-point matrix elements of the fields are very smooth in many cases. In particular, let us consider the lattice action

$$S=\sum_{\langle ij\rangle}(1-\mathbf{s}_i\cdot\mathbf{s}_j)\ ,\quad |\mathbf{s}|^2=1\ . \tag{6.70}$$

In a theory with the Boltzmann weight $e^{-\beta S}$, one can compute the internal energy at weak coupling in powers of $g^2=1/\beta$,

$$\langle 1-\mathbf{s}_i\cdot\mathbf{s}_j\rangle=\frac{1}{2\beta}+O\left[\frac{1}{\beta^2}\right]\ . \tag{6.71}$$

This equation implies that as $g^2\to0$ the "most likely" fields become arbitrarily smooth and ordinary continuum notions of continuity are useful!

To make the topological properties of the lattice model precise we need a lattice construction of the topological charge Q_L in which (Berg and Lüscher, 1981; Lüscher, 1982)

(1) Q_L has only integer values and is composed of a sum of *local* charge densities.

(2) Q_L has the correct continuum limit.

(3) The lattice definition of Q_L should be unambiguous except, perhaps, on a set of spin configurations of measure zero, and these "exceptional" spin configurations should have an action which is bounded from below by a finite constant.

For O(3) the construction is particularly transparent. Let us first consider a triangular lattice. The naive construction which approximates the continuum charge density is the following: At the center of each triangle we associate the spherical area enclosed by the three nearest spins as shown in Fig. 59. Then we define $A=\alpha_1+\alpha_2+\alpha_3-\pi$. The sign of the area should be $\sigma=\text{sgn}[\mathbf{s}_1\cdot(\mathbf{s}_2\times\mathbf{s}_3)]$, and the local charge density is

$$q(x^*)=\sigma A/4\pi\ . \tag{6.72}$$

Then

$$Q_L=\sum_{x^*}q(x^*) \tag{6.73}$$

is a good lattice definition of the topological charge.

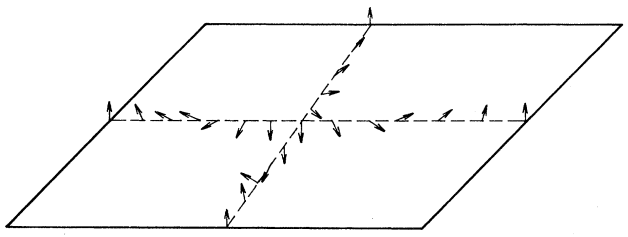

FIG. 58. A spin configuration approximating an instanton in the two-dimensional O(3) model.

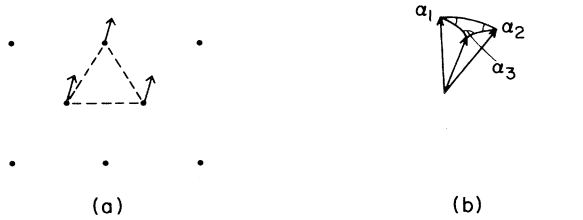

FIG. 59. (a) A spin configuration in the O(3) model on a triangular lattice, and (b) the construction of the local topological charge density.

Only if $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ lie in a given plane does the definition become ambiguous if (Berg and Lüscher, 1981; Lüscher, 1982)

$$1+\mathbf{s}_1\cdot\mathbf{s}_2+\mathbf{s}_2\cdot\mathbf{s}_3+\mathbf{s}_1\cdot\mathbf{s}_3\leq 0\ . \tag{6.74}$$

Such an exceptional configuration has at least one bond which is very unfavorable—i.e., the configuration has a large action and is very unlikely at weak coupling.

For a square lattice, one makes triangles and copies this construction. The triangles can be chosen in two distinct ways, so we can construct two local topological charge densities,

$$\begin{aligned} q_1(x^*)&=\frac{1}{4\pi}[\sigma A(s_1s_2s_4)+\sigma A(s_2s_3s_4)]\ , \\ q_2(x^*)&=\frac{1}{4\pi}[\sigma A(s_1s_2s_3)+\sigma A(s_1s_3s_4)]\ . \end{aligned} \tag{6.75}$$

Similar constructions have been made for CP^{n-1} models in two dimensions (Berg and Lüscher, 1981; Lüscher, 1982) and SU(N) gauge theories in four dimensions (Lüscher, 1982). In each case the exceptional configurations have at least one very high energy bond for which a lower bound can be computed.

Now the question of interest is whether in the continuum limit only smooth lattice field configurations survive which have classical partners. The following difficulty can arise. There may be lattice spin configurations having $Q\neq 0$ which have topological charge densities isolated to the neighborhood of only one site, and these configurations may survive in the continuum limit. This in fact happens in the O(3) model! The configuration shown in Fig. 60 has $Q=1$ and action $S_d=6.69\ldots$. Its probability of occurrence behaves as $\sim e^{-\beta S_d}$. In physical units its probability per area $\sim e^{-\beta S_d}/e^{4\pi\beta}\sim e^{-\beta(S_d-4\pi)}$ [recall from Eq. (6.69) that $\xi\sim e^{2\pi\beta}$]. Since $S_d<4\pi$, these "dislocations" do *not* become unlikely in the continuum limit, and the continuum limit of the topological charge is *unphysical* (Berg and Lüscher, 1981; Lüscher, 1982). Note that this is a violation of universality in the sense that the short distance lattice artifacts are surviving the continuum limit. This is a serious problem, because they still destroy the asymptotic freedom scaling laws for matrix elements involving the topological charge.

For CP^{n-1} models with $n>4$, $S_d>4\pi$, and the naive continuum limit is obtained. Similarly, for gauge theories in four dimensions these "dislocations" do not contribute in the continuum limit.

Although the dislocations remain in the O(3) model, they do not contribute significantly to the ordinary magnetic properties, such as the magnetic susceptibility. These quantities scale as expected from asymptotic freedom (Shenker and Tobochnik, 1980). Why is there such a dramatic difference between ordinary magnetic properties of the model and the model's topological features? Note that the magnetic susceptibility gets contributions from smooth spin configurations when $\beta\to\infty$. However, the topological susceptibility, $\chi_t=\langle Q^2\rangle/\text{volume}$, does not—it gets contributions from only an *exponentially* small fraction $O[\exp(-4\pi\beta)]$, of the slowly varying fields. Therefore, field configurations which vary rapidly on the scale of the lattice spacing can contribute significantly to χ_t, while the magnetic susceptibility is, to good approximation, unaffected.

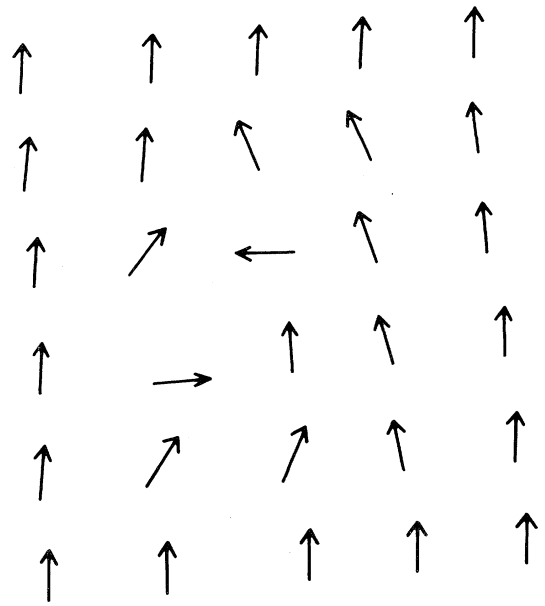

FIG. 60. An "exceptional" spin configuration.

VII. LATTICE FERMIONS

Placing the Dirac equation on a space-time lattice presents some surprisingly difficult problems. In this and the next few sections these issues will be addressed primarily through examples. The reason for an explicit, illustrative presentation reflects the tricky elements in this topic—generic or formal discussions of lattice fermions are notoriously untrustworthy! We will first consider free boson and fermion fields on a spatial lattice (continuum time) in $1+1$ dimensions to see the "species doubling" problem in its simplest setting. The doubling problem is characteristic of a more general pattern of phenomena which have been studied extensively by H. B. Nielsen and N. Ninomiya (1981). Then we shall work through the Euclidean form of the "naive" or "staggered" fermion technique in two dimensions. The emphasis here is on the remnants, both continuous and discrete, of flavor chiral symmetry and the role of the anomaly in the fermion technique. Then numerical methods for lattice fermions will be reviewed. This review will include brief discussions of sparse matrix inversion methods, the Langevin equation, and pseudofermion Monte Carlo and Von Neumann—Ulam random walk methods. Finally, the fermion determinant that appears in the path integral formulation of lattice gauge theory will be discussed, and numerical methods for including it in computer simulations will be reviewed.

Some of these topics are under active research at this time. The "best" fermion method and the "best" numerical approach for dealing with fermion loops are certainly not settled at this time. The discussion here is meant to bring together some of the main ideas and problems in a rapidly developing subject.

A. Free fields on a spatial lattice. Species doubling

To begin, consider $1+1$ dimensions and a free boson field $\phi(x,t)$. The Klein-Gordon equation is

$$\ddot{\phi}=\nabla^2\phi-m^2\phi\ , \tag{7.1}$$

which implies the energy-momentum relation

$$E^2=\mathbf{p}^2+m^2 \tag{7.2}$$

for plane waves. To place Eq. (7.1) on a spatial lattice (time continuum), we make the replacement,

$$\begin{aligned}a^2\nabla^2\phi&\rightarrow\phi(n+1)+\phi(n-1)-2\phi(n)\\&=(d^+ +d^- -2)\phi(n)\ ,\end{aligned} \tag{7.3}$$

where $d^{\pm}$ are shift operators, $d^{\pm}\phi(n)=\phi(n\pm1)$. The lattice Klein-Gordon equation is now

$$a^2\ddot{\phi}=(d^+ +d^- -2)\phi-a^2m^2\phi\ . \tag{7.4}$$

We consider a plane-wave solution to Eq. (7.4),

$$\phi=e^{ikna+iEt} \tag{7.5}$$

on the lattice shown in Fig. 61, and we construct the Brillouin zone symmetrically about $k=0$:

$$-\pi<ka<\pi\ . \tag{7.6}$$

Substituting Eq. (7.5) into Eq. (7.4) gives

$$-E^2\phi=\frac{(e^{ika}+e^{-ika}-2)}{a^2}\phi-m^2\phi\ . \tag{7.7}$$

Therefore, the lattice model has an energy-momentum relation

$$E^2=m^2-2(\cos ka-1)/a^2\ . \tag{7.8}$$

For small $ka\ll1$,

$$E^2=m^2+k^2+O(k^4a^2)\ , \tag{7.9}$$

and the proper energy-momentum relation results in the continuum limit. For finite a, the energy-momentum relation leads to the curve in Fig. 62 and its negative. It is clear from the figure that as $a\rightarrow0$, only the vicinity of $ka\simeq0$ is relevant and that the ordinary energy-momentum relation is obtained and relativity restored. All is well.

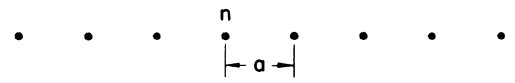

FIG. 61. A space discrete-time continuum lattice in $1+1$ dimensions.

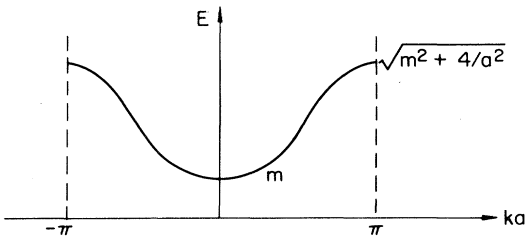

FIG. 62. The energy-momentum relation for free bosons.

Now let us consider the Dirac equation ($m_0=0$) with the conventions

$$\psi=\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix},\quad \gamma_0=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix},\quad \alpha=\gamma_5=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}. \tag{7.10}$$

In the continuum

$$i\dot{\psi}=-i\alpha\partial_z\psi=-i\gamma_5\partial_z\psi\ . \tag{7.11}$$

It is convenient to consider the chiral eigenstates

$$\gamma_5\chi_\pm=\pm\chi_\pm\ . \tag{7.12}$$

For plane waves we choose,

$$\psi_\pm=e^{-ikz+iEt}\chi_\pm\ , \tag{7.13}$$

so

$$E=\pm k\ ,\quad -\infty<k<\infty\ , \tag{7.14}$$

as shown in Fig. 63. We have left- and right-moving fermions and antifermions.

Next we consider the simplest lattice form of the Dirac equation. We place

$$\psi=\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}$$

on each site of a spatial lattice and replace ∂_z by a discrete difference,

$$i\dot{\psi}(n)=\frac{-i}{2a}\gamma_5[\psi(n+1)-\psi(n-1)]\ . \tag{7.15}$$

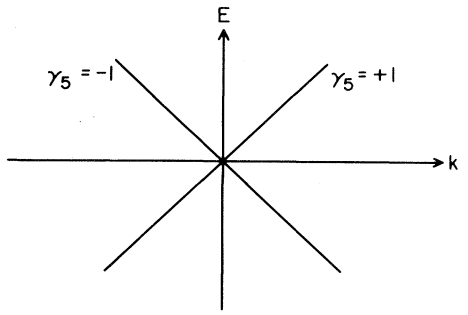

FIG. 63. The spectrum of the free, continuum Dirac equation in $1+1$ dimensions.

Again let

$$\psi_{\pm}=e^{i(-kna+Et)}\chi_{\pm}\ ,\quad \gamma_5\chi_{\pm}=\pm\chi_{\pm}\ , \tag{7.16}$$

so Eq. (7.15) implies

$$-E=\pm\frac{i}{2a}(e^{ika}-e^{-ika})\ , \tag{7.17}$$

which gives

$$E=\pm\sin(ka)/a\ . \tag{7.18}$$

For small $ka\ll 1$, Eq. (7.18) becomes

$$E\simeq\pm k+O(k^3a^2)\ . \tag{7.19}$$

However, for $ka=\pi-k'a$, $k'a\ll 1$, we also have low-energy excitations,

$$E\simeq\mp k'+O(k'^3a^2)\ . \tag{7.20}$$

The full energy-momentum dispersion relation is shown in Fig. 64. We can identify 2 two-component Dirac particles in the continuum limit! Note that the total chiral charge of the fermions is zero.

This catastrophe is best seen by considering the field ψ_+ on the lattice. It describes the excitations shown in Fig. 65. There are a right-mover ($k\simeq 0$) and a left-mover ($k\simeq\pi$). But since ordinary chirality (helicity for particles) is just velocity in 1 + 1 dimensions, the finite energy content of the lattice field χ_+ is a *pair* of fermions with net chirality zero ("species doubling")!

It is important to ask whether the unexpected properties of the lattice Dirac equation are general in character or particular to the detailed construction of this example. Experience with lattice methods convinced workers in the field long ago that the problem is generic and that it is closely related to continuous chiral symmetry. These impressions were made precise by H. B. Nielsen and M. Ninomiya (1981), who have presented a number of "no go" results. An interesting example is the following.

Attempt to couple a left-handed fermion ψ_L to the gauge field A_μ,

$$S=\int d^4x\{-i\bar{\psi}_L\dot{\psi}_L+\bar{\psi}_l\sigma[-i\partial-\mathbf{A}(x)]\psi_L\}\ . \tag{7.21}$$

Put this theory on a lattice with

(1) *Exact* conservation of the chiral charges associated with a compact, *continuous* group [chiral $SU_3\times SU_3$, chiral U(1) of quantum chromodynamics (QCD), for example].

(2) *Locality* in the interactions and hopping terms.

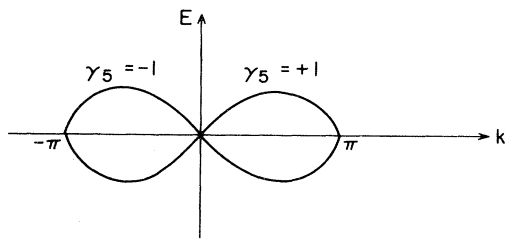

FIG. 64. The spectrum of the native lattice Dirac equation.

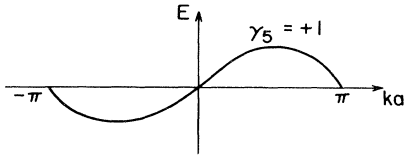

FIG. 65. The spectrum of the field ψ_+ on the lattice.

Then it follows, using the continuity of the energy momentum relation in the Brillouin zone, that there is an equal number of left- and right-handed Weyl fermions in this lattice theory. The proof generalizes those features observed in the (1 + 1)-dimension example presented above. Note that the locality assumption is important here. As in the (1 + 1)-dimension example, it means that the energy-momentum relation is smooth and that the species counting is simple. Changing the precise form of Eq. (7.15), for example, using a slightly different lattice kinetic energy, changes the shape of the energy-momentum relation in Fig. 64 but does not change those global, general features which lead to species doubling. This simple observation underlies the exacting arguments in the literature. One can "evade" the Nielsen-Ninomiya result by using a nonlocal lattice derivative (Drell *et al.*, 1976). But such theories have not been studied extensively by statistical mechanics methods and may have problems with relativity in the continuum limit. They will not be discussed here.

Consider two approaches to beat the species doubling problem:

(1) Thinning the degrees of freedom.

(2) Lifting the energy at the edges of the Brillouin zone from zero.

Approach 1: "Staggered Fermions."

Place a single-component Fermi field $\phi(n)$ on each site (Banks *et al.*, 1976),

$$\{\phi(n),\phi(m)\}=0\ ,\quad \{\phi^\dagger(n),\phi(m)\}=\delta_{nm}\ , \tag{7.22}$$

and describe its dynamics with the Hamiltonian,

$$H=-\frac{i}{2a}\sum[\phi^\dagger(n)\phi(n+1)-\phi^\dagger(n+1)\phi(n)]\ . \tag{7.23}$$

This Hamiltonian gives the equation of motion, $i(\partial/\partial t)\phi=[H,\phi]$,

$$\dot{\phi}(n)=-\frac{1}{2a}[\phi(n+1)-\phi(n-1)]\ . \tag{7.24}$$

To identify a single two-component Dirac field decompose the lattice into an even sublattice (n is an even integer) and an odd sublattice (n is an odd integer),

$$\begin{aligned}&\psi_1(n)=\phi(n),\quad n\text{ even}\\&\psi_2(n)=\phi(n),\quad n\text{ odd}\end{aligned} \tag{7.25}$$

where

$$\psi=\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}.$$

Now the equation of motion becomes

$$\dot{\psi}_1(n)=-\frac{1}{2a}[\psi_2(n+1)-\psi_2(n-1)]\,,$$
$$\dot{\psi}_2(n)=-\frac{1}{2a}[\psi_1(n+1)-\psi_1(n-1)]\,, \tag{7.26}$$

which should be compared to the continuum equation,

$$\dot{\psi}=-\gamma_5\partial_z\psi=-\begin{bmatrix}0&1\\1&0\end{bmatrix}\partial_z\psi$$

or

$$\dot{\psi}_1=-\partial_z\psi_z \quad\text{and}\quad \dot{\psi}_2=-\partial_z\psi_1\,. \tag{7.27}$$

The correspondence between Eqs. (7.26) and (7.27) is clear.

The staggered fermion method avoids the species doubling problem by doubling the size of the unit cube in real space and effectively halving the size of the Brillouin zone.

Let's write fermion bilinears in the new language of $\phi(n)$ and identify the symmetries of the lattice theory:

$$\int\psi^\dagger\psi\,dz=\int(\psi_1^\dagger\psi_1+\psi_2^\dagger\psi_2)dz\to\sum_n\phi^\dagger(n)\phi(n)\,,$$
$$\int\bar{\psi}\psi\,dz=\int(\psi_1^\dagger\psi_1-\psi_2^\dagger\psi_2)dz\to\sum_n(-1)^n\phi^\dagger(n)\phi(n)\,,$$
$$\int\bar{\psi}\gamma_5\psi\,dz=\int(\psi_1^\dagger\psi_2-\psi_2^\dagger\psi_1)dz\to\sum_n(-1)^n[\phi^\dagger(n)\phi(n+1)-\phi^\dagger-(n+1)\phi(n)]\,, \tag{7.28}$$
$$\int\psi^\dagger\gamma_5\psi\,dz=\int(\psi_1^\dagger\psi_2+\psi_2^\dagger\psi_2)dz\to\sum_n\phi^\dagger(n)\phi(n+1)+\phi^\dagger(n+1)\phi(n)\,.$$

Now we consider the symmetries of

$$H=-\frac{i}{2a}\sum[\phi^\dagger(n)\phi(n+1)-\phi^\dagger(n+1)\phi(n)]\,.$$

1. Translation of the spatial lattice by an even number of sites. This symmetry should be related to ordinary translations. The generator of ordinary translations in the continuum theory is

$$p_z=-i\int\psi^\dagger\partial_z\psi\,dz=-i\int(\psi_1^\dagger\partial_z\psi_1+\psi_2^\dagger\partial_z\psi_2)dz\,, \tag{7.29}$$

which does not mix ψ_1 and ψ_2. So its lattice version should not mix the two sublattices, and

$$p_z\to\sum_n\phi^\dagger(n+2)\phi(n)+\phi^\dagger(n)\phi(n+2) \tag{7.30}$$

is the corresponding lattice generator. Translations by *two* lattice spacings correspond to ordinary continuum translations.

2. Translation by an odd number of sites. Again $H\to H$ by inspection. Note from Eq. (7.28) that the generator of translations by one site is $\int\psi^\dagger\gamma_5\psi\,dz$. So this lattice fermion method has discrete γ_5 symmetry. Note that

$$\gamma_5\begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}=\begin{bmatrix}\psi_2\\ \psi_1\end{bmatrix}$$

interchanges ψ_1 and ψ_2. These are chiral rotations through $\pi/2$ radians, $\exp(i\gamma_5\pi/2)=i\gamma_5$.

So, just as the lattice leaves over only discrete translations as symmetries, it admits only discrete chiral rotations. A mass term,

$$m_0\sum(-1)^n\phi^\dagger(n)\phi(n)\,, \tag{7.31}$$

is invariant to shifts by an even number of lattice spacings (translations) but not by an odd number of spacings (discrete chiral transformations). Note that the theory with $m_0=0$ has discrete γ_5 symmetry, so a bare mass term cannot be generated by interactions, if they also possess the γ_5 symmetry.

Note that we have "solved" the species doubling problem at the expense of continuous chiral symmetry. This is consistent with the general Nielsen-Ninomiya constraints.

Approach 2: "Brute Force."

Now we will leave two-component fermions on the spatial lattice but will add terms to H so that E vs k does not have secondary minima at $ka=\pm\pi$ (Wilson, 1977). Let

$$H=-\frac{i}{2a}\sum_n\psi^\dagger(n)\alpha[\psi(n+1)-\psi(n-1)]$$
$$+\frac{B}{2a}\sum_n\bar{\psi}(n)[2\psi(n)-\psi(n+1)-\psi(n-1)]$$
$$+\frac{m}{2a}g^2\sum_n\bar{\psi}(n)\psi(n)\,. \tag{7.32}$$

Note that the second term is a "boson" kinetic energy term multiplied by a. In a classical continuum limit

$$\frac{1}{a}\sum_n \bar{\psi}(n)[2\psi(n)-\psi(n+1)-\psi(n-1)]\to\frac{1}{a}\int \frac{dz}{a} a\bar{\psi}a^2\nabla^2\psi$$
$$=a\int \bar{\psi}\nabla^2\psi\, dz$$
$$\to 0 \quad \text{as } a\to 0 . \tag{7.33}$$

If we omit the last term in H for the moment, the equation of motion for ψ is

$$i\dot{\psi}(n)=-\frac{i}{2a}\gamma_5[\psi(n+1)-\psi(n-1)] + \frac{B}{2a}\gamma_0[2\psi(n)-\psi(n-1)-\psi(n+1)] . \tag{7.34}$$

For a plane wave

$$\psi=e^{i(-kna+Et)}\chi \tag{7.35}$$

the equation of motion becomes

$$-E\chi=-\frac{i}{2a}\gamma_5(e^{-ika}-e^{ika})\chi + \frac{B}{2a}\gamma_0(2-e^{ika}-e^{-ika})\chi ,$$
$$-E\chi=-\gamma_5\frac{\sin ka}{a}\chi+2B\gamma_0\frac{\sin^2(ka/2)}{a}\chi . \tag{7.36}$$

So

$$E^2=\frac{\sin^2 ka}{a^2}+4B^2\frac{\sin^4(ka/2)}{a^2} . \tag{7.37}$$

Letting $ka\to 0$, we have $E^2\simeq k^2+B^2k^4a^2+\cdots$, which is acceptable. Near the edge of the Brillouin zone, $ka\approx\pm\pi$,

$$E^2\simeq 4B^2/a^2 , \tag{7.38}$$

and the boson kinetic energy has opened a gap in the spectrum. In this case, the species doubling problem has been solved with an *irrelevant* operator (Wilson and Kogut, 1974). What has the cost been?

(1) H explicitly breaks continuous *and* discrete γ_5 invariance! There is no remnant of one of the most important symmetries of hadron physics.

(2) Interactions will induce a bare mass for the fermions—hence the third term is added in Eq. (7.32), and a fine tuning must be made so that the continuum limit has finite mass particles in it!

B. Fermions and bosons on Euclidean lattices

Consider the boson propagator in four-dimensional Euclidean space,

$$(m^2-\partial_\mu\partial_\mu)\Delta(x-x')=\delta^4(x-x') . \tag{7.39}$$

We label sites on a four-dimensional lattice,

$$x_\mu=n_\mu a,\quad n_\mu=0,\pm1,\pm2,\ldots,\quad \mu=1,2,3,4 \tag{7.40}$$

and replace derivatives with finite differences,

$$x=\sum_\mu\frac{\partial^2}{\partial x_\mu^2}\to\frac{1}{a^2}\sum_\mu(d_\mu^+ +d_\mu^- -2) , \tag{7.41}$$

where $d_\mu^\pm$ is a shift operator by one lattice spacing in the direction μ.

Let $\Delta_n=a^2\Delta(n)$. Then the Klein-Gordon equation for the propagator becomes

$$a^{-2}m^2\Delta_n=\sum_\mu(\Delta_{n+\mu}+\Delta_{n-\mu}-2\Delta_n)/a^4 =\delta_{n0}/a^4 , \tag{7.42}$$

or, in dimensionless form,

$$(8+m^2a^2)\Delta_n-\sum_\mu(\Delta_{n+\mu}+\Delta_{n-\mu})=\delta_{n0} . \tag{7.43}$$

Passing to momentum space in the usual way, we find

$$\Delta(p)=\frac{1}{1-K\sum_\mu(e^{ip_\mu a}+e^{-ip_\mu a})},\quad K\equiv\frac{1}{8+m^2a^2} , \tag{7.44}$$

or

$$\Delta(p)=\frac{1}{1-2K\sum_\mu\cos p_\mu a} . \tag{7.45}$$

The Brillouin zone covers the range $-\pi/a\le p_\mu a\le\pi/a$ for each $\mu=1$, 2, 3, or 4. If we continue back to Minkowski space $p_0\to iE$, we can identify the particle spectrum with the poles of $\Delta(E,\mathbf{p})$,

$$\Delta(E,\mathbf{p})=\frac{1}{1-2K\cosh Ea-2K\sum_i\cos p_i a} . \tag{7.46}$$

The only poles of Eq. (7.46) in the Brillouin zone occur near the origin, Ea, and $p_ia\ll 1$. In this vicinity we have

$$\Delta(E,\mathbf{p})\simeq\frac{1}{1-8K-K(E^2a^2-\mathbf{p}^2a^2)} , \tag{7.47}$$

which has a pole at

$$E^2=\mathbf{p}^2+(1-8K)/Ka^2 . \tag{7.48}$$

This is the relativistic $E-\mathbf{p}$ relation if $m^2=(1-8K)/Ka^2$. So a proper nontrivial continuum limit is obtained by letting $K\to\frac{1}{8}$ *from below*, $1-8K>0$, and taking $a\to 0$ so that

$$m^2=\left[\frac{1-8K}{K}\right]\Big/a^2 \tag{7.49}$$

is held fixed. Note that for small $K\ll\frac{1}{8}$, fixed a, the lattice theory is essentially static. Expansions in K about

$K=0$ can be done.

Now let us consider free fermion propagation. The continuum's propagator equation reads,

$$(\gamma^\mu \partial'_\mu - m)G(x'-x)=\delta^4(x'-x)\ ,$$
$$\{\gamma^\mu,\gamma^\nu\}=2\delta^{\mu\nu}\ . \tag{7.50}$$

The continuum action is

$$A=\int L\,d^4x=\tfrac{1}{2}\int \bar\psi\gamma^\mu \overleftrightarrow{\partial}_\mu \psi\, d^4x+\int \bar\psi m\psi\, d^4x$$
$$=\int \bar\psi(\gamma^\mu\partial_\mu - m)\psi\, d^4x\ . \tag{7.51}$$

A "naive" lattice action is

$$A=\sum_{n,\mu}\frac{-1}{2a}[\bar\psi(n)\gamma_\mu\psi(n+a_\mu)-\bar\psi(n+a_\mu)\gamma_\mu\psi(n)]a^4$$
$$-m\sum_n \bar\psi(n)\psi(n)a^4\ . \tag{7.52}$$

Since we can anticipate a species doubling problem, add a second-order hopping term,

$$\frac{r}{2a}\sum_{n,\mu}[\bar\psi(n)\psi(n+a_\mu)+\bar\psi(n+a_\mu)\psi(n)-2\bar\psi(n)\psi(n)]a^4\ . \tag{7.53}$$

Now A becomes

$$A=\sum_{n,\mu}\frac{1}{2a}[\bar\psi(n)(r-\gamma_\mu)\psi(n+a_\mu)$$
$$+\bar\psi(n+a_\mu)(r+\gamma_\mu)\psi(n)]a^4$$
$$+\sum_n\left[m-\frac{8r}{a}\right]\bar\psi(n)\psi(n)a^4\ . \tag{7.54}$$

We introduce dimensionless fields $\psi(n)\to a^{-3/2}\psi(n)$,

$$A=\sum_{n,\mu}\bar\psi(n)(r-\gamma_\mu)\psi(n+\mu)+\bar\psi(n+\mu)(r+\gamma_\mu)\psi(n)$$
$$+(ma-8r)\bar\psi(n)\psi(n)\ . \tag{7.55}$$

The propagator equation is then

$$[(r-\gamma_\mu)d_\mu^- +(r+\gamma_\mu)d_\mu^+ +(ma-8r)]G(n)=\delta_{n0}\ . \tag{7.56}$$

Let $K\equiv -1/(ma-8r)$. The propagator in momentum space is

$$G(p)=\frac{K}{1-K\sum_\mu[(r+\gamma_\mu)e^{ip_\mu}+(r-\gamma_\mu)e^{-ip_\mu}]}\ . \tag{7.57}$$

Let us consider the propagator for some special choices of parameters. First, $m=0$, $r=0$. Then

$$G(p)=\frac{1}{\sum_\mu \gamma_\mu \sin p_\mu}=\frac{\sum_\mu \gamma_\mu \sin p_\mu}{\sum_\mu \sin^2 p_\mu}\ . \tag{7.58}$$

It is clear that $G(p)$ has periodic structure in the Brillouin zone: there are particles at the edges of the Brillouin zones, as well as at the origin. In two dimensions the low-energy species occur at the edges of the Brillouin zone shown in Fig. 66. A similar figure exposes 2^d species when d dimensions are made discrete.

Second, $m=0$, $r\neq 0$. Then

$$G(p)=\frac{1}{2\left[\sum_\mu \gamma_\mu \sin p_\mu\right]-2r\sum_\mu \cos p_\mu+8r}\ . \tag{7.59}$$

Now the additional zeros are removed but the action Eq. (7.55) has neither continuous or discrete γ_5 symmetry.

Note that the choice $r=1$ is quite special. Then

$$[(1-\gamma_\mu)d_\mu^+ +(1+\gamma_\mu)d_\mu^- -8]G(n)=\delta_{n0}\ , \tag{7.60}$$

and $1+\gamma_\mu$ and $1-\gamma_\mu$ are *projection* operators. We let

$$P_\mu^\pm=\tfrac{1}{2}(1\pm\gamma_\mu)\ . \tag{7.61}$$

Then

$$(P_\mu^\pm)^2=P_\mu^\pm,\quad P_\mu^+P_\mu^-=P_\mu^-P_\mu^+=0\ . \tag{7.62}$$

With $r=1$, $K\to\frac{1}{8}$ gives the correct relativistic propagator,

$$G\sim\frac{1}{p}\ . \tag{7.63}$$

C. Staggered Euclidean fermions

Now I shall devote considerable space to the Euclidean version of "staggered" fermions. The discussion will be made in two dimensions for free fields but the features of the methods we concentrate on extend to four dimensions in the theory with gauge fields. I am most interested in counting species and exposing the remnants of chiral symmetry in the method. Unlike the Hamiltonian version of the method, there is a *continuous* remnant of chiral symmetry, in addition to discrete γ_5 operations! This is important, since it allows lattice studies of the Goldstone mechanism to make sense!

We choose a γ matrix representation in two dimensions,

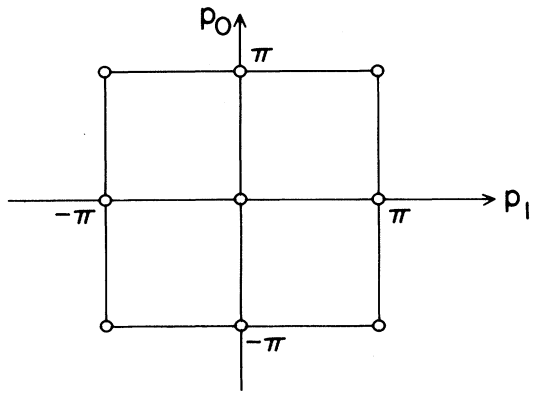

FIG. 66. The Brillouin zone of a two-dimensional Euclidean square lattice.

$$\gamma_0=\begin{bmatrix}0&1\\1&0\end{bmatrix}=\sigma_1,\quad \gamma_1=\begin{bmatrix}0&-i\\i&0\end{bmatrix}=\sigma_2\,,$$
$$\gamma_5=\begin{bmatrix}1&0\\0&-1\end{bmatrix}=\sigma_3\,. \tag{7.64}$$

The naive action is

$$S=\sum_{i,\mu}\bar{\psi}(i)\gamma_\mu[\psi(i+\mu)-\psi(i-\mu)],\quad i=(n_0,n_1)\,. \tag{7.65}$$

To thin the degrees of freedom, we "spin diagonalize" this expression (Kawamoto and Smit, 1981). We define $\chi(n_0,n_1)$:

$$\psi=\gamma_0^{|n_0|}\gamma_1^{|n_1|}\chi\,, \tag{7.66}$$

and independently,

$$\bar{\psi}=\bar{\chi}\gamma_1^{|n_1|}\gamma_0^{|n_0|}\,. \tag{7.67}$$

So,

$$S=\sum_{i,\mu}\bar{\chi}(i)(-1)^{\phi(i,\mu)}[\chi(i+\mu)-\chi(i-\mu)]\,, \tag{7.68}$$

where

$$(-1)^{\phi(i,0)}=\gamma_1^{|n_1|}\gamma_0^{|n_0|}\gamma_0\gamma_0^{|n_0+1|}\gamma_1^{|n_1|}=+1\,,$$
$$(-1)^{\phi(i,1)}=\gamma_1^{|n_1|}\gamma_0^{|n_0|}\gamma_1\gamma_0^{|n_0|}\gamma_1^{|n_1+1|}=(-1)^{|n_0|}\,. \tag{7.69}$$

Since S is "spin-diagonal," one can thin the degrees of freedom, and one need keep only a *single* complex field $\chi(i)$ at each site! This trick of spin diagonalization shows the intimate relation between the two fermion methods.

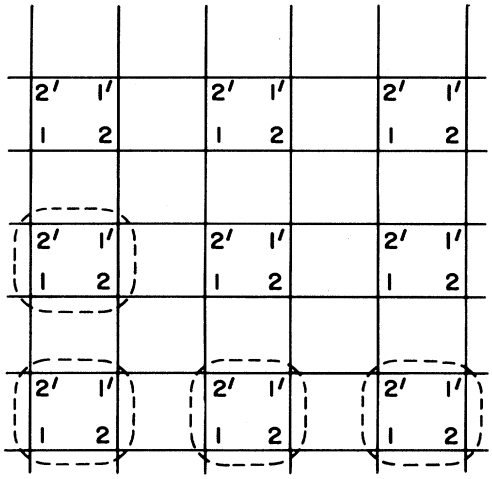

FIG. 67. Sites, blocks, and degrees of freedom for staggered Euclidean fermions.

We label the lattice as two interleaving lattices—one primed and one unprimed, as in Fig. 67, as an aid to identifying the multiple species hiding in Eq. (7.68). The dotted block in the figure repeats itself throughout the lattice by translation. We consider the four sites of the block and label them with γ matrices, as suggested by Eq. (7.66),

site		
origin	$\psi_1^a=\chi(1)\begin{bmatrix}1\\0\end{bmatrix}$,	$\bar{\psi}_1^a=(1,0)\bar{\chi}(1)$
γ_0	$\psi_2^a=\chi(2)\gamma_0\begin{bmatrix}1\\0\end{bmatrix}=\chi(2)\begin{bmatrix}0\\1\end{bmatrix}$,	$\bar{\psi}_2^a=(0,1)\bar{\chi}(2)$
γ_1	$\psi_2^b=\chi(2')\gamma_1\begin{bmatrix}1\\0\end{bmatrix}=i\chi(2')\begin{bmatrix}0\\1\end{bmatrix}$,	$\bar{\psi}_2^b=(0,1)[-i\bar{\chi}(2')]$
$\gamma_0\gamma_1$	$\psi_1^b=\chi(1')\gamma_0\gamma_1\begin{bmatrix}1\\0\end{bmatrix}=i\chi(1')\begin{bmatrix}1\\0\end{bmatrix}$,	$\bar{\psi}_1^b=(1,0)[-i\bar{\chi}(1')]$,

(7.70)

where the labels a and b will help us identify physical degrees of freedom.

The action can be simplified:

$$S=\bar{\chi}(1)[\nabla_0\chi(2)+\nabla_1\chi(2')]+\bar{\chi}(2)[\nabla_0\chi(1)-\nabla_1\chi(1')]+\bar{\chi}(1')[\nabla_0\chi(2')-\nabla_1\chi(2)]+\bar{\chi}(2')[\nabla_0\chi(1')-\nabla_1\chi(1)]+\mathscr{B}$$
$$=\bar{\psi}_1^a(\nabla_0\psi_2^a-i\nabla_1\psi_2^b)+\bar{\psi}_2^a(\nabla_0\psi_1^a+i\nabla_1\psi_1^b)+\bar{\psi}_1^b(\nabla_0\psi_2^b-i\nabla_1\psi_2^a)+\bar{\psi}_2^b(\nabla_0\psi_1^b+i\nabla_1\psi_1^a)+\mathscr{B}\,, \tag{7.71}$$

where $\mathscr{B}$ represents other blocks and where $\bar{\chi}(1)\nabla_0\chi(2)$ means $\bar{\chi}(1)(d_\mu^+-d_\mu^-)\chi(1)$ with $\mu=0$, etc. The site labeling in Eq. (7.71) follows Fig. 67 closely. We can diagonalize the action. Let

$$u_i=(\psi_i^a+\psi_i^b)/\sqrt{2},\quad \tilde{d}_i=(\psi_i^a-\psi_i^b)/\sqrt{2}\,. \tag{7.72}$$

So

$$u_i+\tilde{d}_i=\sqrt{2}\psi_i^a,\quad u_2-\tilde{d}_i=\sqrt{2}\psi_i^b\,. \tag{7.73}$$

Substituting into Eq. (7.71), we have, after some algebra,

$$\begin{aligned} S &= \bar{u}_1(\nabla_0 u_2 - i\nabla_1 u_2) + \bar{u}_2(\nabla_0 u_1 + i\nabla_1 u_1) \\ &\quad + \bar{\tilde{d}}_1(\nabla_0 \tilde{d}_2 + i\nabla_1 \tilde{d}_2) + \bar{\tilde{d}}_2(\nabla_0 \tilde{d}_1 - i\nabla_1 \tilde{d}_2) \\ &= \bar{u}Du + \bar{\tilde{d}}\tilde{D}\tilde{d} , \end{aligned} \tag{7.74}$$

where

$$D = \gamma_0\nabla_0 + \gamma_1\nabla_1, \quad \tilde{D} = \gamma_0\nabla_0 - \gamma_1\nabla_1 . \tag{7.75}$$

Finally, we can make a linear transformation on $\tilde{d}$ to obtain a standard Dirac operator,

$$\tilde{d} = i\gamma_5\gamma_1 d = i\sigma_3\sigma_2 d = \sigma_1 d = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} d . \tag{7.76}$$

Then

$$\bar{\tilde{d}}\tilde{D}\tilde{d} = \bar{d}\sigma_1\tilde{D}\sigma_1 d = \bar{d}(\gamma_0\nabla_0 - \gamma_0\gamma_1\nabla_1\gamma_0)d = \bar{d}Dd . \tag{7.77}$$

Now

$$S = \bar{u}Du + \bar{d}Dd , \tag{7.78}$$

which suggests that two fermions, an isodoublet, emerge in the continuum limit.

D. Block derivatives and axial symmetries

The degrees of freedom relevant to the continuum limit occupy a singie square, as shown in Fig. 67. If we introduce a "block derivative," the continuum limit physics can be made even clearer than in Eq. (7.78) (Duncan *et al.*, 1982). The final result of this exercise will give a flavor-diagonal expression from which we will read off the continuum species and symmetries. We recall

$$\begin{aligned} S &= \bar{\psi}_1^a(\nabla_0\psi_2^a - i\nabla_1\psi_2^b) + \bar{\psi}_2^a(\nabla_0\psi_1^a + i\nabla_1\psi_1^b) \\ &\quad + \bar{\psi}_1^b(\nabla_0\psi_2^b - i\nabla_1\psi_2^a) + \bar{\psi}_2^b(\nabla_0\psi_1^b + i\nabla_1\psi_1^a) . \end{aligned} \tag{7.79}$$

Let's begin to simplify this expression by examining the discrete differences here in greater detail. We consider three sites—$\overset{1}{x}$, $\overset{2}{x}$, and $\overset{3}{x}$—and definitions of a derivative,

$$\begin{aligned} \nabla f(2) &= \tfrac{1}{2}[f(3) - f(1)] , \\ \nabla^2 f(2) &= f(3) + f(1) - 2f(2) . \end{aligned} \tag{7.80}$$

Then

$$\begin{aligned} f(3) - f(2) &= \tfrac{1}{2}[f(3) - f(1)] \\ &\quad + \tfrac{1}{2}[f(3) + f(1) - 2f(2)] \\ &= \nabla f + \tfrac{1}{2}\nabla^2 f , \end{aligned} \tag{7.81}$$

$$\begin{aligned} f(2) - f(1) &= \tfrac{1}{2}[f(3) - f(1)] \\ &\quad - \tfrac{1}{2}[f(3) + f(1) - 2f(2)] \\ &= \nabla f - \tfrac{1}{2}\nabla^2 f . \end{aligned} \tag{7.82}$$

Now we introduce a block derivative. In the action, $\bar{\psi}_1^a\nabla_0\psi_2^a$ occurs. It consists of two terms—one coupling degrees of freedom entirely inside a block and another connecting blocks. We consider the horizontal axis shown in Fig. 68. The difference $\psi_2^a(x_2) - \psi_2^a(x_1)$ occurs in Eq. (7.79). But using the algebra above, we arrive at

$$\begin{aligned} \psi_2^a(x_2) - \psi_2^a(x_1) &= \tfrac{1}{2}[\psi_2^a(x_3) - \psi_2^a(x_1)] \\ &\quad - \tfrac{1}{2}[\psi_2^a(x_3) + \psi_2^a(x_1) - 2\psi_2^a(x_2)] \\ &= \hat{\nabla}\psi_2^a - \tfrac{1}{2}\hat{\nabla}^2\psi_2^a , \end{aligned} \tag{7.83}$$

where "hats" indicate intrablock derivatives.

If we use block derivatives, the variables ψ_i^a, ψ_i^b $(i = 1,2)$ can be thought to live at the *centers* of each block. The variables are now naturally organized to account for the multiple flavors which appear in the continuum limit. This formalism also lends itself most naturally to real-space renormalization-group methods.

Introducing the block derivatives into S, we get

$$\begin{aligned} S &= \bar{\psi}_1^a[(\hat{\nabla}_0\psi_2^a - \tfrac{1}{2}\hat{\nabla}_0^2\psi_2^a) - i(\hat{\nabla}_1\psi_2^b - \tfrac{1}{2}\hat{\nabla}_1^2\psi_2^b)] + \bar{\psi}_2^a[(\hat{\nabla}_0\psi_1^a + \tfrac{1}{2}\hat{\nabla}_0^2\psi_1^a) + i(\hat{\nabla}_1\psi_1^b - \tfrac{1}{2}\hat{\nabla}_1^2\psi_1^b)] \\ &\quad + \bar{\psi}_1^b[(\hat{\nabla}_0\psi_2^b + \tfrac{1}{2}\hat{\nabla}_0^2\psi_2^b) - i(\hat{\nabla}_1\psi_2^a + \tfrac{1}{2}\hat{\nabla}_1^2\psi_2^a)] + \bar{\psi}_2^b[(\hat{\nabla}_0\psi_1^b - \tfrac{1}{2}\hat{\nabla}_0^2\psi_1^b) + i(\hat{\nabla}_1\psi_1^a + \tfrac{1}{2}\hat{\nabla}_1^2\psi_1^a)] . \end{aligned} \tag{7.84}$$

So

$$\begin{aligned} S &= \bar{u}\hat{D}u + \bar{\tilde{d}}\hat{\tilde{D}}\tilde{d} + \bar{\psi}_1^a\left[-\tfrac{1}{2}\hat{\nabla}_0^2\psi_2^a + \frac{i}{2}\hat{\nabla}_1^2\psi_2^b\right] + \bar{\psi}_2^a\left[\tfrac{1}{2}\hat{\nabla}_0^2\psi_1^a - \frac{i}{2}\hat{\nabla}_1^2\psi_1^b\right] \\ &\quad + \bar{\psi}_1^b\left[\tfrac{1}{2}\hat{\nabla}_0^2\psi_2^b - \frac{i}{2}\hat{\nabla}_1^2\psi_2^a\right] + \bar{\psi}_2^b\left[-\tfrac{1}{2}\hat{\nabla}_0^2\psi_1^b + \frac{i}{2}\hat{\nabla}_1^2\psi_1^a\right] . \end{aligned} \tag{7.85}$$

The first two terms in Eq. (7.85) can be written as $\bar{u}\hat{D}u + \bar{d}\hat{D}d$ using Eqs. (7.76) and (7.77). The terms in large parentheses in Eq. (7.85) can also be simplified by transforming them to u's and d's. It is convenient to use flavor isospin notation,

$$f = \begin{bmatrix} u \\ d \end{bmatrix} , \tag{7.86}$$

and flavor matrices,

$$T_0 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad T_1 = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} . \tag{7.87}$$

Then some algebra yields the final result,

$$S = \bar{f}\hat{D}f + \bar{f}\gamma_5 T_\mu \tfrac{1}{2}\hat{\nabla}_\mu^2 f . \tag{7.88}$$

This form of the action has the virtues advertised

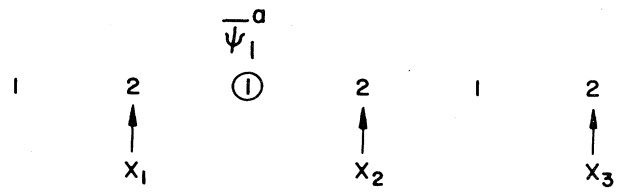

FIG. 68. Labels and coupling of degrees of freedom along one axis for staggered fermions.

above. The first term is free of flavor mixing effects and displays the two species in the clearest possible way. The second term is an irrelevant boson kinetic energy term with an important twist—the matrices $\gamma_5 T_\mu$ occur! The matrix character of the second term has several important consequences. It insures the existence of a *continuous* remnant of chiral symmetry, and it breaks explicitly the axial flavor-neutral symmetries. These points will be made explicit below in greater detail.

E. Staggered fermions and remnants of chiral symmetry

To expose the vector and axial symmetries of the staggered fermions we reconsider the action written in its most primitive form, Eq. (7.68).

We consider even and odd states separately and a $U(1)\times U(1)$ global invariance group, where the first U(1) acts on even sites, the second U(1) on odd sites,

$$\chi(n)\rightarrow\begin{cases}U_e\chi(n), & n \text{ even}\\ U_0\chi(n), & n \text{ odd}\end{cases} \tag{7.89}$$

$$\bar{\chi}(n)\rightarrow\begin{cases}\bar{\chi}(n)U_0^+, & n \text{ even}\\ \bar{\chi}(n)U_e^+, & n \text{ odd}\,.\end{cases} \tag{7.90}$$

This is obviously a symmetry of

$$S=\bar{\chi}(1)[\nabla_0\chi(2)+\nabla_1\chi(2')]+\bar{\chi}(2)[\nabla_0\chi(1)-\nabla_1\chi(1')]+\bar{\chi}(1')[\nabla_0\chi(2')-\nabla_1\chi(2)]$$
$$+\bar{\chi}(2')[\nabla_0\chi(1')+\nabla_1\chi(1)]+\mathscr{B}\,, \tag{7.91}$$

since only even (odd) sites are coupled to odd (even) sites. We write this symmetry in terms of conventional, continuum fields,

$$\psi_1\rightarrow U_0\psi_1,\quad \psi_2\rightarrow U_e\psi_2\,,$$
$$\begin{bmatrix}u_1\\u_2\end{bmatrix}=\begin{bmatrix}\psi_1^a+\psi_1^b\\ \psi_2^a+\psi_2^b\end{bmatrix}\rightarrow\begin{bmatrix}U_0 & u_1\\ U_e & u_2\end{bmatrix} \tag{7.92}$$
$$=\tfrac{1}{2}(U_0+U_e)\begin{bmatrix}u_1\\u_2\end{bmatrix}+\tfrac{1}{2}(U_0-U_e)\begin{bmatrix}u_1\\-u_2\end{bmatrix}.$$

Similarly,

$$\begin{bmatrix}\tilde{d}_1\\ \tilde{d}_2\end{bmatrix}\rightarrow\tfrac{1}{2}(U_0+U_e)\begin{bmatrix}\tilde{d}_1\\ \tilde{d}_2\end{bmatrix}$$
$$+\tfrac{1}{2}(U_0-U_e)\gamma_5\begin{bmatrix}\tilde{d}_1\\ \tilde{d}_2\end{bmatrix}. \tag{7.93}$$

But $\tilde{d}=\sigma_1 d$, and $\gamma_5=\sigma_3$ in this basis, so

$$\begin{bmatrix}\tilde{d}'_1\\ \tilde{d}'_2\end{bmatrix}=\sigma_1\begin{bmatrix}d'_1\\ d'_2\end{bmatrix}=U_+\sigma_1\begin{bmatrix}d_1\\ d_2\end{bmatrix}+U_-\gamma_5\sigma_1\begin{bmatrix}d_1\\ d_2\end{bmatrix}, \tag{7.94}$$

where $U_\pm=(U_0\pm U_e)/2$. So

$$d'=\sigma_1U_+\sigma_1 d+\sigma_1U_-\gamma_5\sigma_1 d\,. \tag{7.95}$$

But $\sigma_1\gamma_5\sigma_1=-\gamma_5$, so

$$d'=U_+d-U_-\gamma_5 d\,,$$
$$u'=U_+u+U_-\gamma_5 u\,. \tag{7.96}$$

We let

$$T_3=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix},$$

and use the isospin notation $f=\binom{u}{d}$. Then

$$f'=U_+f+U_-\gamma_5T_3f\,. \tag{7.97}$$

In summary, the action,

$$S=\bar{f}\hat{D}f+\bar{f}\gamma_5T_\mu\tfrac{1}{2}\hat{\nabla}_\mu^2 f\,, \tag{7.98}$$

has the continuous symmetry

$$f'=U_+f+U_-\gamma_5T_3f\,. \tag{7.99}$$

So, the original $U_e\times U_0$ symmetry contains a vector piece $1_D\times 1_F$, fermion number conservation, and an axial piece $\gamma_5\times T_3$, one component of axial isospin. The second piece forbids a mass term in S. When it breaks spontaneously in more than two dimensions, a pion will appear in the usual Goldstone fashion!

We note that the irrelevant operator exposed in the Block derivative expression breaks the full axial isospin symmetries to a smaller set of continuous symmetries. It breaks the axial flavor-neutral symmetries explicitly.

In addition to the continuous $U_0\times U_e$ symmetries there are discrete symmetries of interest in S. If we translate the system by an odd number of sites in either direction and transform phases appropriately, the action will not change. Therefore these symmetries are square roots of translations and will be identified with λ_5, T_i, and their products. These discrete symmetries forbid fermion mass terms, isospin breaking mass terms, etc.

We consider a translation along a body diagonal

$$\chi(t,z)\to i(-1)^t\chi(t+1,z+1)\ ,$$
$$\bar{\chi}(t,z)\to i(-1)^t\bar{\chi}(t+1,z+1)\ . \tag{7.100}$$

One can check explicitly that S, Eq. (7.91) is unchanged by this transformation. In the notation of Fig. 67 the symmetry operation is

$$\chi(1)\to i\chi(1'),\ \chi(2)\to -i\chi(2')\ ,$$
$$\chi(1')\to -i\chi(1),\ \chi(2')\to i\chi(2)\ . \tag{7.101}$$

So

$$\psi_1^a\to\psi_1^b,\ \psi_2^a\to-\psi_2^b,\ \psi_1^b\to\psi_1^a,\ \psi_2^b\to-\psi_2^a\ , \tag{7.102}$$

or

$$u_1\to u_1,\ u_2\to -u_2,\ d_2\to -d_2,\ d_1\to d_1\ . \tag{7.103}$$

So

$$\begin{bmatrix}u_1\\u_2\end{bmatrix}\to\gamma_5\begin{bmatrix}u_1\\u_2\end{bmatrix},\quad \begin{bmatrix}d_1\\d_2\end{bmatrix}\to\gamma_5\begin{bmatrix}d_1\\d_2\end{bmatrix}\ , \tag{7.104}$$

or

$$f\to\gamma_5 f\ . \tag{7.105}$$

This translation is therefore a flavor singlet γ_5 transformation.

This is the only remnant of the continuous U_A symmetry in the lattice action.

This lattice fermion method has the species we want, an isodoublet degenerate in mass, and has the correct symmetries in the continuum limit. On the lattice, however, it has only *remnants* of the full chiral symmetries. There is one continuous axial flavor symmetry, and if the gauge field dynamics drives spontaneous symmetry breaking, a Goldstone pion should occur even in the cutoff model. In the continuum limit, full chiral symmetry should appear and should be spontaneously broken, as well. Then additional Goldstone pions should evolve as $a\to 0$. The lattice fermion method has enough symmetry to guarantee this naturally.

VIII. NUMERICAL METHODS FOR LATTICE FERMIONS

Now we are ready to put interactions back into the fermion action. Our aim is to calculate composite propagators describing the mesons and baryons of quantum chromodynamics. We shall see that a basic ingredient in such a calculation is the fermion Green's function G in a given background gauge field configuration. So first we will discuss methods to calculate G numerically and finally review methods to incorporate fermion dynamics into Monte Carlo calculations.

A. Jacobi and related methods of calculating the fermion propagator

In an abbreviated notation (but one I hope to be clear), the fermion propagator G in a background U field satisfies

$$[\not{D}(U)+m]G=\delta^2(x-y)\ . \tag{8.1}$$

So we face a generic problem: Find the vector x which satisfies $Mx=b$. Here x represents a fermion field which lives on sites. It can be thought of as a complex vector of dimension equal to the number of sites of the lattice.

This problem can be solved with relaxation methods. A most elementary method is due to Jacobi. We let t equal the number of sweeps through the lattice and consider the differential equation

$$\frac{dx}{dt}=-[Mx(t)-b]\ . \tag{8.2}$$

If we invent an iterative procedure which approaches $\dot{x}=0$, then we will have $x=M^{-1}b$, as desired.

The matrix M can be written in the form $M=1-K$. This is natural in our problem—the 1 labels the "no hopping" term in the kinetic energy. K is near diagonal and includes the hopping pieces of the lattice action. The precise form of all this depends on the lattice fermion method used.

For the computer $dt=\varepsilon$ and Eq. (8.2) becomes

$$\frac{x_{n+1}-x_n}{\varepsilon}=-[(1-K)x_n-b]\ . \tag{8.3}$$

So

$$x_{n+1}=\varepsilon b+(1-\varepsilon+\varepsilon K)x_n\ . \tag{8.4}$$

Iterating this equation generates successive terms in the Neumann series,

$$\frac{1}{1-(1-\varepsilon+\varepsilon K)}\varepsilon b\equiv\frac{1}{1-K}b\ , \tag{8.5}$$

as desired. So the method will work if $||(1-\varepsilon+\varepsilon K)||<1$, where $||\ ||$ denotes the maximum eigenvalue of the enclosed operator.

Typically, $||K||\approx 2d/2m$, where d is the dimension of space-time and m is the bare fermion mass. The operator K is imaginary, so its eigenvalues lie in the region shown in Fig. 69. If we choose ε small enough, the eigenvalues

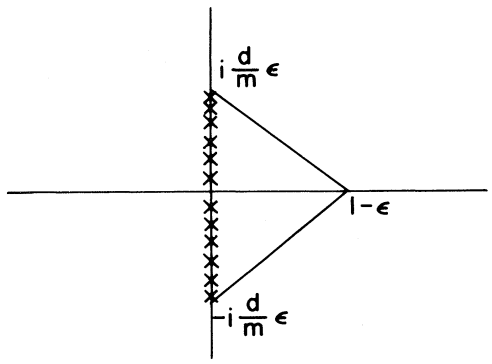

FIG. 69. Region of convergence for the Jacobi matrix inversion method.

of $1-\varepsilon+\varepsilon K$ will lie within the unit circle, and the method will converge (Stone, 1982). This approach has several problems, however.

(1) As $m\to 0$ (real physics), ε must be taken small, and many iterations are necessary for convergence.

(2) Convergence, even when it occurs, is not monotonic.

(3) There is an art to choosing the "best" ε, the "over-relaxation" parameter. Optimization procedures exist (Hageman and Young, 1979).

An improvement over the Jacobi method is the *Gauss-Siedel* iteration procedure (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten, 1982). Here the variables $x_n(i)$ are ordered. For example, one labels the lattice sites sequentially with sensible boundary conditions. In the Gauss-Siedel procedure when $x_n(i+1)$ is updated the most recently calculated $x_n(1), x_n(2), \ldots, x_n(i)$ are used on the right-hand side of Eq. (8.4). This frequently speeds up convergence considerably, since more information is used in each step as compared to the Jacobi method, in which the values $x_n(i)$ are updated only after each sweep through the lattice. The following terminology is used and should be clear, Gauss-Siedel is the cyclic relaxation and Jacobi is the simultaneous relaxation.

There are other relaxation methods which are more sophisticated than these. Experience has shown that conjugate-gradient methods (Kogut *et al.*, 1982)

(1) Converge more regularly in a wider class of problems than either Jacobi or Gauss-Siedel.

(2) Proceed much faster, especially in the weak coupling region for Euclidean staggered fermions.

The conjugate-gradient method is a generalization of the method of steepest descents. One must dig deeply into the literature to find useful, explicit discussions of the method (Hestenes and Stiefel, 1952).

B. The Langevin equation and pseudofermion Monte Carlo

We recall Langevin's classic work on Brownian motion and the approach to equilibrium of classical statistical ensembles (Langevin, 1908). In the simplest application one considers a set of particles q_i which has an equilibrium distribution $\exp[-H(q_i)/\sigma]$. A stochastic equation, the Langevin equation, which gives this distribution as $t\to\infty$ is (Langevin, 1908)

$$\dot{q}_i=-\frac{\partial H}{\partial q_i}+\eta_i(t)\ ,$$
$$\langle\langle \eta_i(t)\eta_j(t')\rangle\rangle=2\sigma\delta_{ij}\delta(t-t')\ , \tag{8.6}$$

where the white noise η brings in the fluctuations leading to a probability distribution for q_i with the desired width σ. In Eq. (8.6) $\langle\langle\ \rangle\rangle$ indicates an average over Gaussian stochastic white noise.

We can use this formalism if we invent an equivalent boson problem whose correlation function gives the fermion propagator in the background U-field configuration (Fucito *et al.*, 1981).

Consider a two-component $\phi_i(x,t)$, where x labels sites on a four-dimensional Euclidean lattice and t is the number of sweeps through the lattice,

$$\dot{\phi}_1=(\not{D}+m)\phi_1+\eta\ ,$$
$$\dot{\phi}_2=(-\not{D}+m)\phi_2+\eta \tag{8.7}$$

(same η) where

$$\langle\langle \eta(x,t)\eta^*(x',t')\rangle\rangle=2\delta(t-t')\delta(x-x')\ . \tag{8.8}$$

We will show that an average over the noise produces the fermion Green's function,

$$\lim_{t\to\infty}\langle\langle \phi_1(x,t)\phi_2^*(y,t)\rangle\rangle=G(x,y;U)\ . \tag{8.9}$$

This method of calculation has some advantages over iterative procedures. $G(x,y;U)$ is obtained for all x and y at the same time. The relaxation methods give G only for a source point fixed. It also has disadvantages. In particular, more statistical errors occur here because of its inherent noise.

Let's prove the claim Eq. (8.9) (Stone, 1982). Solve both equations in Eq. (8.7) formally,

$$\phi_1(x,t)=\int dx'dt'\sum_\lambda \phi_\lambda(x)\phi_\lambda^*(x')e^{-|t-t'|\lambda}\Theta(t-t')\eta(x',t')\ ,$$
$$\phi_2(y,t)=\int dy'dt''\sum_\lambda \phi_\lambda(y)\phi_\lambda^*(y')e^{-|t-t''|\lambda^*}\Theta(t-t'')\eta(y't'')\ . \tag{8.10}$$

We observed here that the spectrum of $(\not{D}+m)$ is $\{\lambda,\phi_\lambda\}$ and of $(-\not{D}+m)$ is $\{\lambda^*,\phi_\lambda\}$.

Now

$$\langle\langle \phi_1(x,t)\phi_2^*(y,t)\rangle\rangle=\int dx'dy'dt'dt''\sum_\lambda \phi_\lambda(x)\phi_\lambda^*(x')\sum_\rho \phi_\rho^*(y)\phi_\rho(y')e^{-|t-t'|\lambda}e^{-|t-t''|\lambda}$$
$$\times\langle\langle \eta(x',t')\eta^*(y',t'')\rangle\rangle\Theta(t-t')\Theta(t-t'')$$
$$=\int dx'dt'\sum_{\lambda\rho}\phi_\lambda(x)\phi_\lambda^*(x')\phi_\rho^*(y)\phi_\rho(x')2\Theta(t-t')e^{-2|t-t'|\lambda}\ . \tag{8.11}$$

But

$$\int dx' \phi_\lambda^*(x')\phi_\rho(x') = \delta_{\lambda\rho} , \tag{8.12}$$

so

$$\langle\langle \phi_1(x,t)\phi_2^*(y,t)\rangle\rangle = \sum_\lambda \int dt' e^{-2|t-t'|\lambda} 2\Theta(t-t')\phi_\lambda(x)\phi_\lambda^*(y)$$

$$= \sum_\lambda \frac{\phi_\lambda(x)\phi_\lambda^*(y)}{\lambda} \equiv G(x,y;U) , \tag{8.13}$$

which is the desired result. So the differential equations, Eq. (8.7), can be solved numerically, an average over noise taken, and $G(x,y;U)$ obtained.

An ordinary Monte Carlo method with bosons can also be used to obtain the fermion propagator. The fermion piece of the action reads generically

$$S_{\text{fermion}} = \sum_{i,j} \bar{\psi}_i (\not{D} + m)_{ij} \psi_j . \tag{8.14}$$

Suppose that $\not{D}+m$ is positive. Then

$$G(x,y;U) = \langle \phi(x)\phi^*(y)\rangle$$

$$= \frac{1}{Z} \int [d\bar{\phi}][d\phi]\phi(x)\phi^*(y) \times \exp\left[- \sum_{ij} \bar{\phi}_i (\not{D}+m)_{ij}\phi_j \right] . \tag{8.15}$$

If $\not{D}+m$ is not positive, we should consider $(-\not{D}+m)(\not{D}+m) = -\not{D}^2+m^2$. Then the "pseudofermion" calculation, Eq. (8.15), gives $(-\not{D}^2+m^2)^{-1}$, and we can apply $(-\not{D}+m)$ to the result to obtain $(\not{D}+m)^{-1}$. The problem with the method can be computer time—a Monte Carlo program for $-\not{D}^2+m^2$ involves second-nearest-neighbor coupling and is somewhat complicated.

C. The fermion propagator and random walks

The connection between random walks, path integrals, and propagators is a profitable one in many field-theory, statistical mechanics settings.

Here we want to invert $[\not{D}(U)+m]_{ij}$ and obtain $G(U)_{ij}$. Write $D = [\not{D}(U)+m]$, $D = 1-H$. Suppose that maximum eigenvalue of H, $\lambda_{\max}(H)$, has magnitude which is less than unity. Then

$$D^{-1} = (1-H)^{-1} = \sum_{k=0}^{\infty} H^k \tag{8.16}$$

is a meaningful Neumann series. In indices,

$$(D^{-1})_{ij} = 1 + H_{ij} + H_{ik}H_{kj} + H_{ik}H_{kl}H_{lj} + \cdots . \tag{8.17}$$

Let's devise a stochastic method to calculate Eq. (8.17) (Forsythe and Liebler, 1950; Wasow, 1951; Kuti, 1982—this reference advocates the von Neumann—Ulam method for fermion problems). We visualize H_{kl} as a particle hopping from site l to site k. Typically, H_{kl} will be nonzero only if k and l are nearest neighbors or next-nearest neighbors. Then each term in Eq. (8.17) can be visualized as in Fig. 70. Next we consider H_{kl} and factor it,

$$H_{kl} = p_{kl} h_{kl} , \tag{8.18}$$

with no summation, where $p_{kl} \geq 0$ and is the probability of the moving particle to hop from k to l. Let there also be a finite probability that the particle stop at site l,

$$p_l = 1 - \sum_{k=1}^{n} p_{lk} .$$

Now let us consider a walk $i \to k \to l \to p \to \cdots \to m \to j$. We lay down the rule that it contributes

$$h_{ik}h_{kl}\cdots h_{mj}p_j^{-1} \tag{8.19}$$

to $(D^{-1})_{ij}$. Then the average over all walks reproduces $(D^{-1})_{ij}$, because the probability of the route $i \to k \to l \to \cdots \to m \to j$ is

$$p_{ik}p_{kl}\cdots p_{mj}p_j , \tag{8.20}$$

so the total contribution to $(D^{-1})_{ij}$ is for n terms

$$\sum_n \sum_{k,l,\ldots,m} (p_{ik}p_{kl}\cdots p_{mj}p_j)(h_{ik}h_{kl}\cdots h_{mj}p_j^{-1}) = \sum (p_{ik}h_{ik})(p_{kl}h_{kl})\cdots(p_{mj}h_{mj})$$

$$= \sum H_{ik}H_{kl}\cdots H_{mj} = \sum_n (H^n)_{ij} , \tag{8.21}$$

which reproduces the Neumann series!

The variance in calculation of $(D^{-1})_{ij}$ can also be calculated. We let

$$R_{ij} = p_{ij}h_{ij}^2 , \tag{8.22}$$

with no sum and assume $\max_i |\lambda_i(k)| < 1$. We define

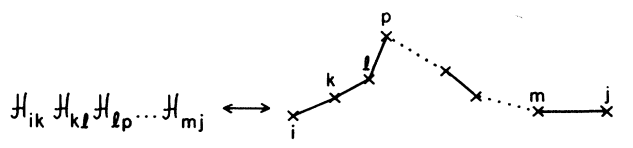

FIG. 70. A pictorial representation of the matrix inversion method of Von Neumann and Ulam.

$T=(1-R)^{-1}$. Then (Forsythe and Liebler, 1950; Wasow, 1951)

$$\sigma_{ij}^2=T_{ij}p_{ij}^{-1}-(D^{-1})_{ij}^2 . \tag{8.23}$$

If $\max_i |\lambda_i(k)| \geq 1$, then $\sigma_{ij}^2=\infty$.

The advantage of this random walk method is that the required number of walks to produce a result with a given statistical accuracy does not depend strongly on the size of the matrix. The method appears to be well suited to Euclidean fermion methods discussed in Sec. VII.B, but not to the staggered fermion method.

D. Calculating the fermion determinant

The original problem, quantum chromodynamics, had an action of the generic form

$$S=\sum_{i,j}\bar{\psi}_i[\not{D}(U)+m]_{ij}\psi_j+S_0(U) . \tag{8.24}$$

It is not practical to use computer simulation techniques on S directly. The anticommuting properties of the Euclidean fermion fields $\{\psi_j,\psi_i\}=0$ would require the storage of phases between fermions at all the sites of the lattice. This is hopeless even for small lattices. However, since ψ_i appears quadratically in Eq. (8.24) it can be integrated out,

$$\begin{aligned}\int\prod_i[d\psi][d\bar{\psi}]e^{-S}&=e^{-S_{\text{eff}}}\\&=\det[\not{D}(U)+m]e^{-S_0(U)}\\&=\exp-\{S_0-\operatorname{tr}\ln[\not{D}(U)+m]\} ,\end{aligned} \tag{8.25}$$

and we are left with just a U-field problem. Unfortunately, the effective action in Eq. (8.25) is nonlocal and is therefore very difficult to deal with. Direct evaluation of the determinant is not practical, since the matrix $[\not{D}(U)+m]$ is $N\times N$, and N is the number of links in the lattice. In addition, the determinant consists of $\sim N^3$ terms with intricate cancellations between terms. Because of these problems, approaches dealing with Eq. (8.25) are approximate in character and must exploit the sparse nature of the matrix $\not{D}(U)+m$.

We consider a local Metropolis algorithm for Eq. (8.25) (Scalapino and Sugar, 1981). Change the configuration $\{U\}$ to $\{\tilde{U}\}$ so that they differ *at only one link*. We must compare $S_{\text{eff}}[U]$ and $S_{\text{eff}}[\tilde{U}]$,

$$e^{-S_0(\tilde{U})-S_0(U)}\frac{\det[\not{D}(\tilde{U})+m]}{\det[\not{D}(U)+m]} , \tag{8.26}$$

to update link variables. Let us look at the ratio of determinants. Let us call $\not{D}(U)+m=G^{-1}(U)$ and $D(\tilde{U})+m=\tilde{G}^{-1}\equiv G^{-1}(U)+\delta$. Then

$$\frac{\det[\not{D}(\tilde{U})+m]}{\det[\not{D}(U)+m]}=\frac{\det(G^{-1}+\delta)}{\det G^{-1}}=\det(1+G\delta) . \tag{8.27}$$

Here G is the $N\times N$ matrix, the full lattice fermion propagator in the gauge field U. The size of the matrix δ depends on the procedure; but for local fermion derivatives it is small, because $\{U\}$ and $\{\tilde{U}\}$ differ on only one link (Scalapino and Sugar, 1981). Generically,

$$\delta\simeq\begin{bmatrix}0&0&\cdots&0\\&0&1&0\\\vdots&-1&0&\vdots\\0&\cdots&&0\end{bmatrix}=\begin{bmatrix}0&0&0\\0&\delta_2&0\\0&0&0\end{bmatrix} . \tag{8.28}$$

In this idealized case we would have

$$\det(1+G\delta)=1+\operatorname{tr}G\delta_2+\det G\delta_2 \tag{8.29}$$

exactly. Therefore, we need only $G_{ij}(U)$ for $i=j$ or nearest-neighbor sites to carry out a simulation!

Unfortunately, the procedure is still slow compared to pure gauge theory simulations, because each time a link is updated $G_{ii}(U)$ and $G_{i,i+\mu}(U)$ are needed. It is hoped that the random walk methods are fast enough to be practical here, but certainly the more straightforward iterative procedures are hopelessly too slow. In addition, δ_2 is not such a small matrix in practice. If a second-order formalism must be used in place of Eq. (8.26) to guarantee convergence, then next-to-nearest couplings occur in $\not{D}^2(U)+m^2$ and increase the number of nonzero elements in δ_2 considerably in four dimensions.

Luckily, a nice simplification can be made here if only *small* gauge field changes $\{U\}\to\{\tilde{U}\}$ are made (Fucito *et al.*, 1981). We return to the expression

$$S_{\text{eff}}=S_0(U)-\operatorname{tr}\ln[\not{D}(U)+m] . \tag{8.30}$$

If $\{\tilde{U}\}$ and $\{U\}$ are almost identical, then we can expand Eq. (8.26),

$$\begin{aligned}S_{\text{eff}}(\tilde{U})-S_{\text{eff}}(U)&=S_0(\tilde{U})-S_0(U)\\&\quad-\sum_{i,j}G_{ij}(U)\frac{\delta\not{D}(U)_{ij}}{\delta U}(\tilde{U}-U)\\&\quad+O[(\tilde{U}-U)^2] .\end{aligned} \tag{8.31}$$

This result suggests a modified Metropolis algorithm (Fucito *et al.*, 1981). We consider a gauge field configuration $\{U\}$ and calculate $G_{ij}(U)$ by the pseudofermion method. In other words, we update the pseudofermion field ϕ_i with the action,

$$S(\phi)=\sum_{i,j}\phi_i^*[\not{D}(U)+m]_{ij}\phi_j , \tag{8.32}$$

n times and collect $\overline{\phi_i^*\phi_j}$ in the configuration $\{U\}$. Next we update the gauge field with the action,

$$S(U)=S_0(U)-\sum_{i,j}\overline{\phi_i^*\phi_j}[\not{D}(U)+m]_{ij} , \tag{8.33}$$

and obtain a new field configuration $\{\tilde{U}\}$. We repeat the procedure many times. Neglecting errors $O[(\tilde{U}-U)^2]$, we obtain the correct results when $n\to\infty$.

In practice one finds that relatively small values

of n give good estimates for $\overline{\phi_i^* \phi_j}$. This is not surprising, since $\phi_i^* \phi_j$ is a local matrix element which is ultraviolet divergent. However, as $m \to 0$, the convergence deteriorates rapidly!

Methods of including the dynamical effects of light fermions in computer simulations are actively being pursued by several research groups. There may be many surprises in these investigations. The role of topological excitations in the fermion determinant will certainly be investigated in the near future.

To date, most results on the spectrum of quantum chromodynamics neglect internal fermion loops. Note that the fermion determinant actually occurs in the effective action raised to a power equal to the number of fermion flavors N_f. So, in the theoretical limit $N_f \to 0$, the fermion determinant is unity and calculations can be done efficiently. One should recognize this as a fancy way of suppressing internal fermion loops—each loop has a coefficient proportional to N_f, a counting factor. There are good phenomenological reasons to expect the $N_f \to 0$ limit to be a good approximation for quantum chromodynamics—dynamical fermion loops appear to be unimportant to hadron spectroscopy (for example, the naive quark model is successful, and widths of hadronic masses are typically small compared to the masses themselves). The $N_f \to 0$ limit does, however, suffer from an excess of symmetry—the axial anomaly is lost in this limit, and the η' should be degenerate with the pion. Anyway, in the $N_f \to 0$ limit the hadron mass spectrum does not involve any calculation more difficult than $G(U)$, since meson and baryon propagators are just products of quark propagators. The class of graphs contributing to a meson propagator is illustrated in Fig. 71. Preliminary results on such calculations are very encouraging (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten, 1982).

IX. CHIRAL SYMMETRY BREAKING ON THE LATTICE

Now we will apply the lattice fermion methods of the last section to problems of chiral symmetry breaking in quantum chromodynamics and related field theories. We shall begin by reviewing arguments for chiral symmetry breaking in continuum gauge theories. Then we consider the strong coupling large-N (color) limit of quantum chromodynamics on the lattice and see that chiral symmetry is spontaneously broken. Naive lattice fermions are used in this discussion. Then finally we consider Monte Carlo data which address problems such as the following: (1) What is the interplay of chiral symmetry breaking and confinement? (2) Does the chiral symmetry breaking "most attactive channel" hypothesis work? (3) Does chiral symmetry breaking occur in quantum electrodynamics at strong coupling? Because of the Nielsen-Ninomiya theorem (Nielsen and Ninomiya, 1981) we will not be able to face left-right asymmetric models. This is a shame. However, we can still pose questions relevant to the hierarchy problem and patterns of chiral symmetry breaking and discuss the essential physics of the phenomenon in a few cases.

First let's review some of the ideas concerning the physics of chiral symmetry breaking. There is an argument due to A. Casher (1979) which is abstracted from $(1+1)$-dimension models (Casher *et al.*, 1974), which claims that confinement implies chiral symmetry breaking. To begin, we consider a bound $q\bar{q}$ pair in an s wave. The quarks have zero bare mass and the theory's Lagrangian is chirally symmetric. Suppose that the confining force is spin independent, as in lattice gauge theory, and that bound $q\bar{q}$ pairs in s waves exist. Suppose that we can think about this bound state semiclassically as in a quark model. Then the claim is that this scenario has spontaneously broken chiral symmetry. Let us consider a turning point in the world line of the bound quark (Fig. 72). The quark turns around, but its spin does not—therefore its chirality (chirality is identically equal to helicity for quarks) changes sign. So chirality is not a good quantum number, and we must have a vacuum with chiral symmetry breaking!

What is the physical mechanism behind this result? It

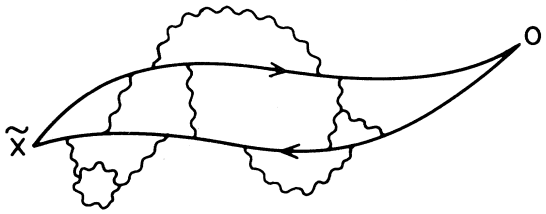

FIG. 71. A meson propagator in the $N_f \to 0$ approximation.

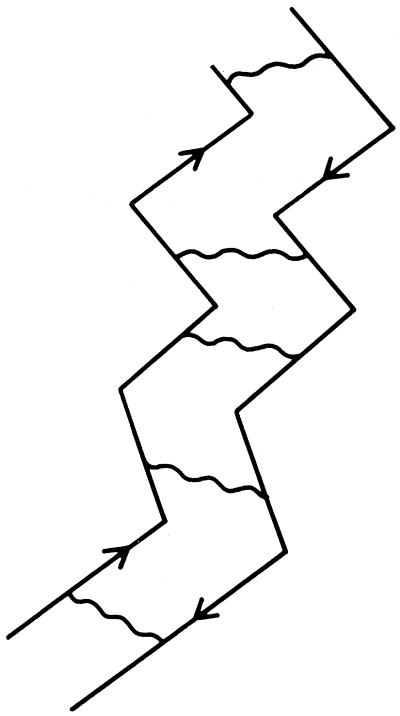

FIG. 72. World lines of a s-wave bound $q\bar{q}$ pair.

must be that the vacuum is a $q\bar{q}$ condensate with indefinite chirality and that when the quark in the s-wave bound state turns around it has actually absorbed chirality from the vacuum (Casher, 1979).

This intuitive picture strongly suggests but does not prove that confinement implies chiral symmetry breaking. More formal arguments which arrive at similar conclusions in more restricted cases (e.g., $N\to\infty$) have also been given (Coleman and Witten, 1980).

Other authors have presented models, physical pictures, and approximate calculations of a different sort in which chiral symmetry breaking occurs due to *short*-range interactions. The original Nambu–Jona-Lasinio four-Fermi coupling models were of this type (Nambu and Jona-Lasinio, 1961). The interaction is attractive and has zero range. It was argued that if the strength of the interaction is above a critical value, $g^2>g_c^2$, a condensate of a chiral symmetry-breaking operator forms, the fermions develop a dynamical mass nonperturbatively, and a triplet of massless pions results as the Goldstone bosons (spin waves) of the spontaneously broken symmetry operators. All of this has been studied more formally recently (Cornwell *et al.*, 1974). These recent references consider self-consistent equations for the dynamical generation of a fermion mass. They compare the vacuum energies for massive fermions to that with massless ones in perturbation theory in the gauge coupling. The general analysis is hopelessly complicated to handle analytically, so they are forced to truncate the perturbation theory series to the lowest-order effect, $O(g^2)$. They find that if g^2 is large enough, then chiral symmetry breaking occurs. Unfortunately, $g_c^2/4\pi\sim O(1)$, so their analysis is not consistent. Anyway, their philosophy is that short-distance attraction is sufficient for chiral symmetry breaking—they sum only one-gluon exchange graphs explicitly.

This mechanism has been suggested as an ingredient in understanding the hierarchy problem of unified gauge theories (Raby *et al.*, 1980). We consider single-gluon exchange between quarks (Fig. 73) in a higher representation with quadratic Casimir C_f. Perhaps these quarks condense when $C_f g_c^2/4\pi\sim 1$. In an asymptotically free theory this means that quarks with "slightly" different color charges can condense and decouple from the remaining light degrees of freedom at very different length scales. This is easily seen. Let the momentum transfer of the gluon be $q^2<0$. Then

$$C_f g^2(q^2)/4\pi\sim C_f\frac{1}{4\pi\beta_0\ln(-q^2/\Lambda^2)}\sim 1\ , \tag{9.1}$$

where

$$\beta_0=\frac{11}{3}\left[\frac{N}{16\pi^2}\right] \tag{9.2}$$

in a pure SU(N) gauge theory. So the quarks condense with characteristic relative momentum $-q^2=\Lambda^2 e^{12\pi C_f/11N}$. The exponential dependence on C_f is the important point here—the relevant length scale of the condensate is *exponentially sensitive* to the color charges of the quarks. So in a given gauge theory with quarks in several representations we can in principle generate a hierarchy of length scales in a natural fashion.

However, it is not clear that these estimates are reliable. Is g^2 small enough that one-gluon exchange is reliable and asymptotic freedom can be used? The exponent in the formula above that gave the hierarchal formula relied on these facts very sensitively. If C_f is large, do screening and internal quark loops in general change these estimates? Certainly if C_f is too large they do—even asymptotic freedom is spoiled by too many quarks of large color charges!

Luckily, the dynamical, nonperturbative effects of interest here are estimable in lattice gauge theory. Let's begin by considering strong coupling in the Euclidean four dimensions formulation. In particular, let us consider SU(N) gauge theory in d space-time dimensions. We can make a simple calculation when both N and d are large. Although this limit is far from real physics, it will show us that chiral symmetry breaking is possible and that disorder, and not confinement, is the essential ingredient (Blairon *et al.*, 1980).

The action for "naive" fermions is

$$S=\tfrac{1}{2}\sum_{r,n_\mu}\bar{\psi}(r)\gamma_\mu n_\mu U(r,n_\mu)\psi(r+n_\mu) +m\sum_r\bar{\psi}(r)\psi(r)+\frac{1}{2g^2}\sum_{plq}\mathrm{tr}UUUU+\mathrm{H.c.}\ , \tag{9.3}$$

where the sum over the unit vector n_μ includes all the sites $r+n_\mu$ neighboring r. Suppose that $g^2\gg 1$, so that the tr$UUUU$ term can be dropped. We have added a small fermion mass term as a symmetry-breaking effect in Eq. (9.3) to help search for spontaneous symmetry breaking in the usual fashion. At the end of the calculation we let $m\to 0$ and see if the limit is smooth and if dynamical mass generation has occurred. We will do perturbation theory with

$$S=S_0+S_{\mathrm{int}}\ , \tag{9.4}$$

where

$$S_0=m\sum_r\bar{\psi}(r)\psi(r)\ ,\qquad S_{\mathrm{int}}=\tfrac{1}{2}\sum_{r,n_\mu}\bar{\psi}\gamma_\mu n_\mu U\psi\ , \tag{9.5}$$

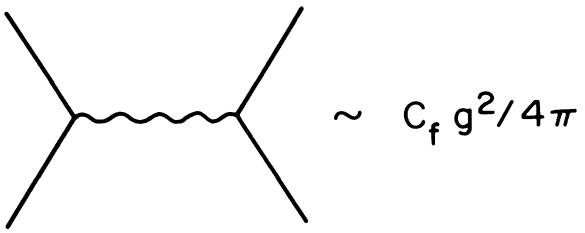

FIG. 73. Single-gluon exchange between quarks in a color representation with quadratic Casimir C_f.

and treat the hopping term, $S_{\rm int}$, to all orders. Readers will recognize the calculation as a graphical approach to a self-consistent fermion mean-field theory.

The ingredients in perturbation theory are

$$\langle \psi_\beta^b(r)\bar{\psi}_\alpha^a(r)\rangle = \frac{1}{m}\delta_{\alpha\beta}\delta^{ab} , \tag{9.6}$$

where α and β are color indices running from 1 to N, and a and b are Dirac indices running from 1 to $2^{d/2}$. The expectation value in Eq. (9.6) is taken with respect to S_0. When $S_{\rm int}$ is treated as a perturbation—i.e., when $\exp(S_0+S_{\rm int})$ is expanded in powers of $S_{\rm int}$—it is clear that local gauge invariance implies that only closed fermion loops contribute. Each link in such a loop contributes a factor,

$$-\tfrac{1}{2}n_\mu\gamma_\mu U(r,n_\mu) , \tag{9.7}$$

to the amplitude. In addition, the integrals over the gauge fields project the product of U matrices on each link onto the identity matrix. To obtain the leading N dependence in this calculation the only U-matrix integral we need is

$$\int [dU]U_{ij}U_{kl}^\dagger = \frac{1}{N}\delta_{il}\delta_{jk} . \tag{9.8}$$

In these graphical rules the only consequence of Fermi statistics is the association of -1 to each closed fermion loop.

Now we can begin doing graphs. The zeroth-order contribution for $\langle\bar{\psi}(0)\psi(0)\rangle$ is

$$\frac{NC}{m} , \tag{9.9}$$

where C is the number of space-time components of ψ, $2^{d/2}$. To second order in $S_{\rm int}$ the fermion can hop to a nearest-neighbor site and back, so it contributes

$$\frac{NC}{m}\left[\frac{2d}{4m^2}\right] . \tag{9.10}$$

The factor $2d$ counts the number of nearest neighbors. Note that the fermion path encloses zero area and that the integral

$$\int [dU]U_{ij}U_{kl}^\dagger = \frac{1}{N}\delta_{il}\delta_{jk}$$

has been done.

At higher orders complicated graphs occur. However, for large N we shall see that graphs with many zero-area "petals," as shown in Fig. 74(a), dominate. A graph with n petals contributes

$$A_n = \frac{NC}{m}\left[\frac{d}{2m^2}\right]^n . \tag{9.11}$$

We could select just this set of graphs, let $n\to\infty$, and compute $\langle\bar{\psi}\psi\rangle$. However, it is better to make a self-consistent calculation. Then the end of each petal of each graph must be treated as the source for an additional flower! This more ambitious calculation gives a fermion mean-field theory.

(a) (b)

FIG. 74. (a) Dominant graphs contributing to $\langle\bar{\psi}\psi\rangle$ for $N\to\infty, d\to\infty$, and (b) a subdominant graph.

So, with this improvement the graphs are decorated to (Blairon *et al.*, 1980)

$$\begin{aligned} A_0 &= \frac{NC}{m} , \\ A_1 &= \frac{NC}{m}\left[\frac{d}{2m}\frac{\langle\bar{\psi}\psi\rangle}{NC}\right] , \\ &\vdots \\ A_n &= \frac{NC}{m}\left[\frac{d}{2m}\frac{\langle\bar{\psi}\psi\rangle}{NC}\right]^n . \end{aligned} \tag{9.12}$$

Summing all these graphs for $\langle\bar{\psi}\psi\rangle$ gives

$$\langle\bar{\psi}\psi\rangle = \frac{CN}{m-\frac{d}{2NC}\langle\bar{\psi}\psi\rangle} , \tag{9.13}$$

which has a smooth $m\to 0$ limit, yielding $\langle\bar{\psi}\psi\rangle = -iCN\sqrt{2/d}$. Mapping ψ back to Minkowski space ψ_m, we obtain

$$\langle\bar{\psi}_m\psi_m\rangle = -i\langle\bar{\psi}\psi\rangle = -CN\sqrt{2/d} , \tag{9.14}$$

which gives spontaneous symmetry breaking of the usual sort.

A few comments about this calculation are in order. Note the mean-field character of the calculation: $\langle\bar{\psi}\psi\rangle$ at one site is determined in terms of $\langle\bar{\psi}\psi\rangle$ at nearest-neighbor sites self-consistently. Note that the "petals" in A_n do not overlap and that our counting factor is good only for large dimensions. This calculation is the first nontrivial term in a $1/d$ expansion. Vacuum fluctuations—disconnected graphs—did not appear in this calculation. One can check that such graphs enclosing zero area cancel between the numerator and the denominator in the formal expression for $\langle\bar{\psi}\psi\rangle$ at large d.

Another illuminating calculation one can do in the fermion mean-field approximation gives the pion propagator,

$$D(r-r') = \langle(\bar{\psi}\gamma_5\psi)_r(\bar{\psi}\gamma_5\psi)_{r'}\rangle . \tag{9.15}$$

The graphs which dominate are shown in Fig. 75 and they can be summed by methods similar to those for

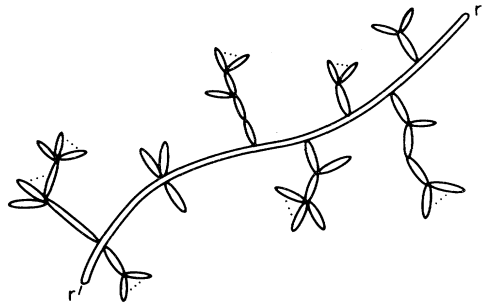

FIG. 75. Strong coupling graphs contributing to the pion propagator.

$\langle \bar{\psi}\psi \rangle$, yielding (Blairon *et al.*, 1980)

$$D(k) \sim \frac{4NC}{k^2 + \frac{4d}{NC} m \langle \bar{\psi}\psi \rangle} \tag{9.16}$$

as $k^2 \to 0$. So, letting the explicit symmetry breaking vanish, $m \to 0$, we have a Goldstone pion. For $m \neq 0$ we have a strong coupling version of the current algebra relation

$$f_\pi^2 m_\pi^2 = m \langle \bar{\psi}_m \psi_m \rangle \tag{9.17}$$

in the form

$$\left[\frac{NC}{4d} \right] m_\pi^2 = m \langle \bar{\psi}_m \psi_m \rangle \ . \tag{9.18}$$

Our last detail is to check that these graphs which include zero area give the dominant contribution at $N \to \infty$. The argument for this is similar to the analogous result in the continuum $N \to \infty$ limit. For example, compare Fig. 74(a) with Fig. 74(b), which has a hole. The dominant graph, Fig. 74(a), behaves as N, which simply counts the number of quark colors circulating in the closed, infinitely narrow fermion loop. The graph with the hole behaves as O(1). This can be seen explicitly, using the U matrix integral formula provided above. Alternatively, we can note that *both* branches of the internal loop must be local color singlets separately, so there is an additional constraint which eliminates one power of N. A systematic argument can be made which generalizes this observation (Blairon *et al.*, 1980).

Another perspective on the mean-field graphs is the following: We consider the $N_f \to 0$ limit discussed in the preceding section. It is clear that the "flower" graphs of the mean-field approximation can be written as the continuous world line of a single quark. It is also clear that graphs such as Fig. 74(b) are subdominant in the $N_f \to 0$ limit. So this strong coupling calculation suggests that the $N_f \to 0$ limit is adequate for studying chiral symmetry breaking, at least qualitatively. This observation will become more important when we discuss numerical work below.

These are all amusing and instructive results. However, they apply only for $g \gg 1$ and depend, therefore, on the nonuniversal features of this lattice action. To proceed we turn to computer simulations which can reach the scaling limit $g^2 \to 0$. We consider also the $N_f \to 0$ limit of the theory, so internal fermion loops are neglected. The $N_f \to 0$ limit is also implicitly taken in most past work on the subject of chiral symmetry breaking—certainly the single-gluon exchange calculations. Of course the $N_f \to 0$ limit has some serious limitations. For example, it ignores color dynamical fermions (screening) and probably overestimates the tendency for condensates to form.

In the $N_f \to 0$ limit the Monte Carlo calculation of $\langle \bar{\psi}\psi \rangle$ is relatively simple (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten, 1982; Kogut *et al.*, to appear). For each field configuration $\{U\}$ one computes $G(x,0;\{U\})$. Averaging over many gauge field configuration then gives the propagator at a chosen value of g^2. $\langle \bar{\psi}\psi \rangle$ is proportional to $\overline{G(0,0;\{U\})}$, where the bar indicates the average over a statistical ensemble of gauge field configurations computed by ordinary Monte Carlo methods at g^2.

First we consider the evidence for chiral symmetry breaking in QCD-SU_3 gauge fields with quarks in the fundamental representation of the gauge group (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten, 1982). $\langle \bar{\psi}\psi \rangle$ is calculated for various g^2, and we expect it to satisfy the weak coupling scaling laws for $g^2 \lesssim 1$. By dimensional analysis, $\langle \bar{\psi}\psi \rangle$ should scale as Λ_L^3, if the operator was renormalization-group invariant. It is not (in fact $m\bar{\psi}\psi$ is), and the dimensional analysis result is modified by a logarithm,

$$\langle \bar{\psi}\psi \rangle = C(g^2/6\pi)^{-4/11} \Lambda_L^3 \ . \tag{9.19}$$

The data and expected scaling law are shown in Fig. 76. As for the string tension, the data strongly suggest that chiral symmetry breaking occurs in the continuum limit. Furthermore, the magnitude of the matrix element is in fairly good agreement with phenomenology (Hamber and Parisi, 1981; Marinari *et al.*, 1981; Weingarten, 1982).

Next we can ask whether confinement is necessary for chiral symmetry breaking, and we can identify the energy scale responsible for chiral symmetry breaking (Kogut *et al.*, 1982). A nice way to do this is to find the tempera-

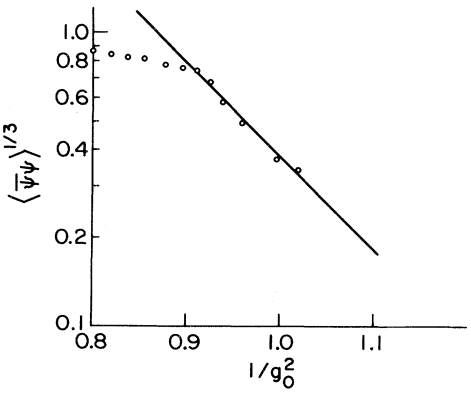

FIG. 76. $\langle \bar{\psi}\psi \rangle^{1/3}$ vs $1/g_0^2$ Monte Carlo data in the $N_f \to 0$ approximation for SU_3 gauge theory.

ture T_χ where chiral symmetry is restored and compare it to the deconfining temperature T_D. Finite temperature simulations are done on asymmetric lattices with the number of temporal links N_t related to the physical temperature,

$$N_t=(aT)^{-1}\,, \tag{9.20}$$

as readers will recall from continuum field theory (Fetter and Walecka, 1971). By varying g^2 one changes the physical length scale in the system and effectively changes the temperature. Alternatively, one can change N_t for fixed g^2 and produce different physical temperatures. The temperature T_χ above which $\langle\bar\psi\psi\rangle$ vanishes can be obtained from these measurements. The deconfinement transition is obtained by measuring the excess free energy of an isolated quark in the system. With periodic boundary conditions for the U matrices in the t direction (and antiperiodic for the Fermi field) one calculates the expectation value of the Wilson line, $\langle \mathrm{tr}\prod_{n_0=1}^{N_t} U_0(n_0,\mathbf{n})\rangle$ (Yaffe and Svetitsky, 1982). For $T<T_D$ the flux tube originating from the static source of color at $\mathbf{n}$ costs a finite amount of energy per unit length. Therefore, the matrix element vanishes as the linear dimension of the spatial lattice. If $T>T_D$ the gluons are presumed to form a plasma which can screen the color of the heavy quark—a flux tube does not form, and the quark simply has a finite self-mass dependent on g^2. It can be argued that the deconfining phase transition in SU(N) gauge theory is associated with the spontaneous symmetry breaking of global Z_N symmetry (Yaffe and Svetitsky, 1982). In Fig. 77 we show $\langle\bar\psi\psi\rangle$ and the Wilson line $\langle W\rangle$ for SU(2). $\langle\bar\psi\psi\rangle$ for massless quarks is inferred from calculations of G with $m\to 0$ and the limit $m\to 0$ is taken by standard (essentially linear) extrapolation procedures. Since $\langle W\rangle$ can be positive or negative, the absolute value of W measured in each gauge field configuration is averaged over the gauge field ensemble to obtain $\langle W\rangle$. (This procedure is analogous to inferring the magnetization of the Ising model from an ensemble of spin configurations where the global magnetization could be either "up" or "down.") We see from Fig. 77 that for SU(2) chiral symmetry breaking may persist slightly beyond the deconfining point $1\lesssim T_\chi/T_D\lesssim 1.3$. The SU(3) curves shown in Fig. 78 are quite different. It appears that both transitions are first order and that they are coincident. The first-order character of the SU(3) deconfining transition is not unexpected. The relevant symmetry for an SU(N) deconfinement is transition in Z_N. Z_2, Ising systems in three dimensions have second-order transitions, while Z_3, three-state Potts models have first-order transitions. Less is known about the chiral transitions. However, since the SU(3) deconfining transition is first order, it is not suprising to see it carry the chiral transition with it.

Additional computer simulations on lattices of different sizes show that the physical temperatures scale with asymptotic freedom,

$$T_\chi=C_\chi\Lambda_L,\quad T_D=C_D\Lambda_L\,. \tag{9.21}$$

For SU(2), $1\lesssim C_\chi/C_D,=1.3$; and for SU(3), $1\lesssim C_\chi/C_D\lesssim 1.05$.

Computer simulations measuring $\langle\bar\psi\psi\rangle$ for quarks in higher representations have also been done (Kogut *et al.*, 1982). For SU(2) the color representations $j=1$, $\frac{3}{2}$, and 2 have been studied in addition to $j=\frac{1}{2}$. The higher representations all show chiral symmetry breaking! Zero-temperature simulations give the couplings g_j^2, where $\langle\bar\psi\psi\rangle$ begins to turn on rapidly,

$$j(j+1)g_j^2\approx 3.5\,. \tag{9.22}$$

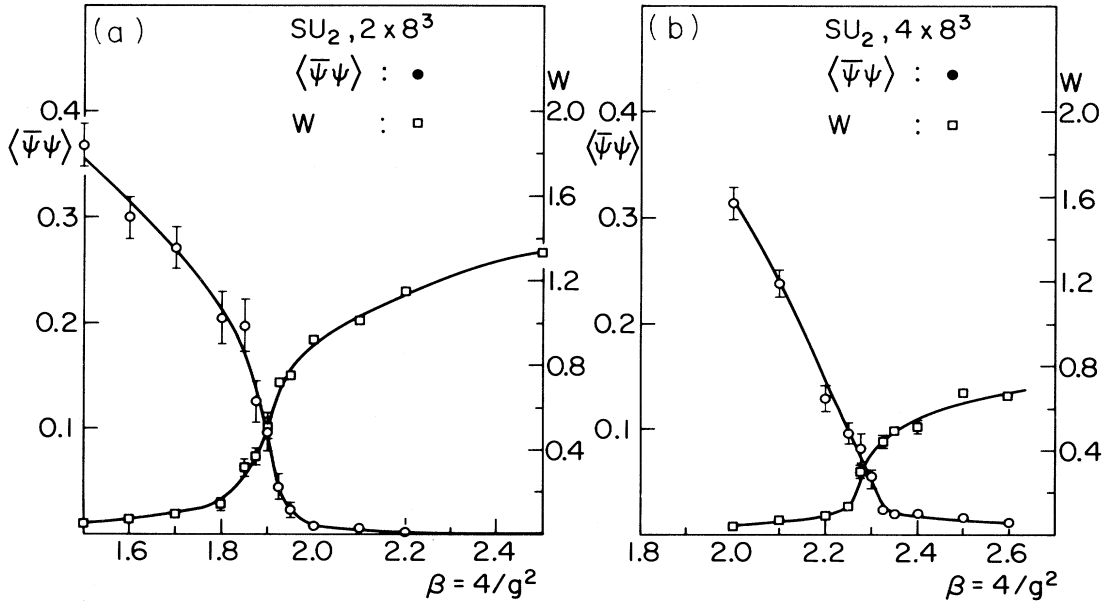

FIG. 77. $\langle\bar\psi\psi\rangle$ and the Wilson line vs $\beta=4/g^2$ on a 2×8^3 and a 4×8^3 lattice for SU(2) gauge theory.

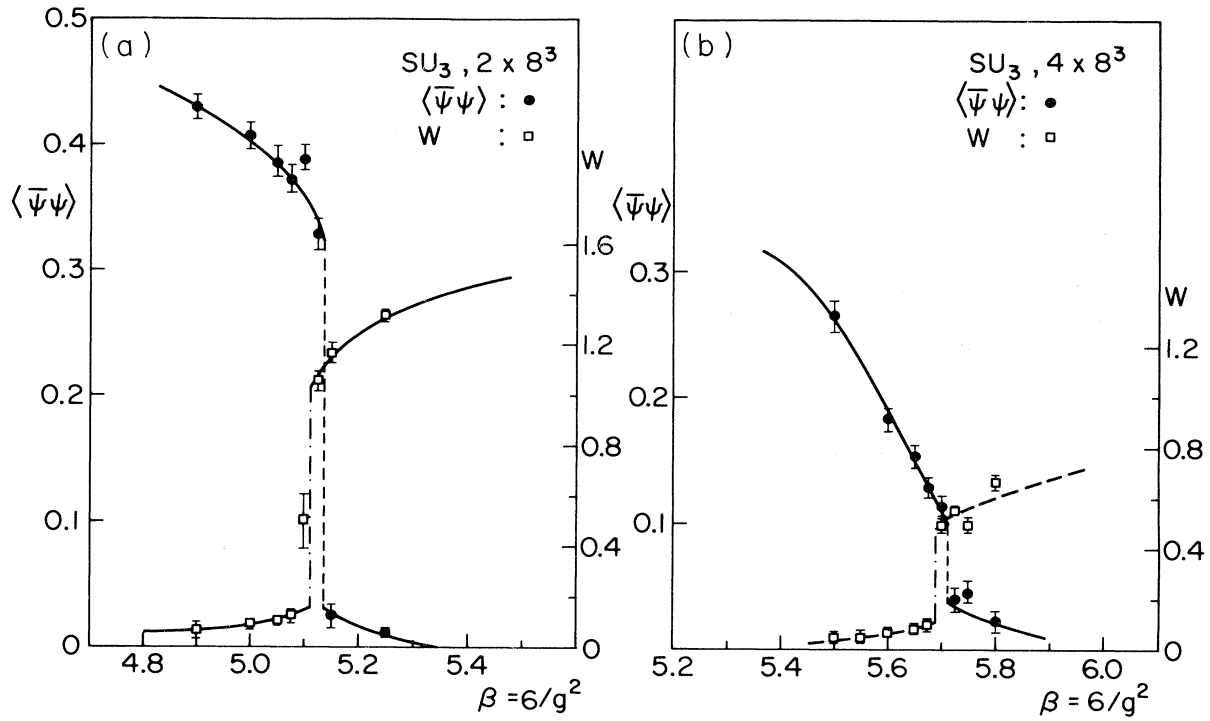

FIG. 78. The same as Fig. 77 except for SU(3) and $\beta=6/g^2$.

The right-hand side of this formula is (roughly) independent of j. It therefore supports the simple picture of chiral symmetry breaking discussed earlier in this section. And exponentially disparate mass scales are expected of these condensates of different color representations.

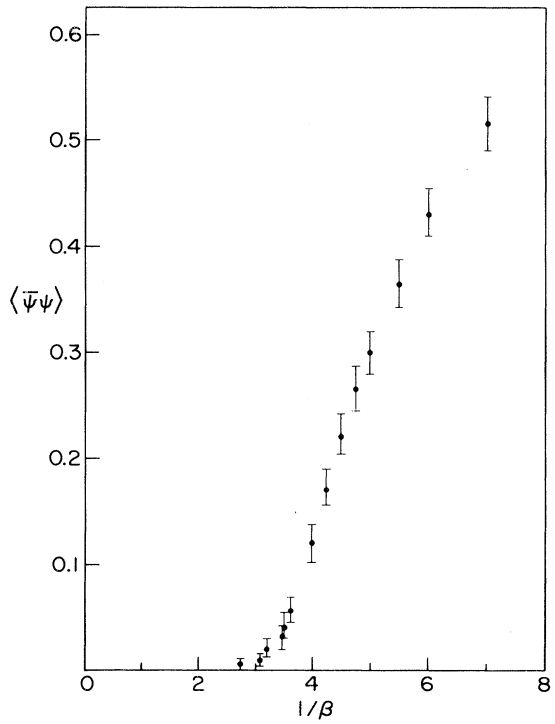

FIG. 79. $\langle\bar{\psi}\psi\rangle$ vs $e_0^2=1/\beta$ for electrodynamics formulated on a four-dimensional lattice.

Finally, chiral symmetry breaking has been explored in pure lattice electrodynamics without vortices, i.e., the Villain form of the action, as discussed in Sec. VI.B, with the magnetic monopole loops *discarded* (Shenker *et al.*, 1983). This model does not confine—it is a free field and generates a $1/R$ potential exactly—and it has no topological excitations. However, as shown in Fig. 79, it breaks chiral symmetry at $e_0^2\approx 3.7$, and $\langle\bar{\psi}\psi\rangle$ turns on with a critical index $\beta=0.64\pm0.10$, $\langle\bar{\psi}\psi\rangle\sim(e^2-e_0^2)^\beta$, which is close to the canonical dimension 0.5 expected of an order parameter in a four-dimensional free field. It remains to compare this result in detail with one-loop continuum field theory estimates described earlier in this lecture. But this calculation shows quite forcefully that strong, short-distance forces are sufficient to break chiral symmetry in four dimensions.

And perhaps these lattice calculations also suggest that symmetry-breaking physics beyond QCD can be attacked numerically.

X. CONCLUDING REMARKS

There is a great deal of activity in lattice gauge theory at the moment. Several groups are attempting to improve the rough spectrum calculations of Parisi and collaborators and to include fermion loops into those calculations. Larger lattices are needed to reduce finite-size effects, and greater statistical accuracy is required. A first-principles calculation of the spectrum and static matrix elements of quantum chromodynamics is not far off. Other groups

are studying symmetry breaking in quantum chromodynamics and related theories. First-principles nuclear matter calculations and matter in extreme environments seem possible. The Bielefeld group (Satz, 1982) is particularly active in this field.

Analytic work aims at the $N \to \infty$ properties of the theory. Recent work by T. Eguchi and K. Kawai (1982) has stirred considerable activity, and a firmer grasp of the $N \to \infty$ pure SU(N) gauge theory seems possible.

Alternate approaches through string models are also beautiful and may yield breakthroughs (Polyakov, 1981).

Certainly there are many things to look forward to in this field!

ACKNOWLEDGMENTS

The author thanks Ralph Simmons, Chairman of the Physics Department of the University of Illinois, for his support and for the opportunity to present an earlier version of this work as an advanced graduate course. He thanks J. B. Zuber and the Les Houches Summer School and D. J. Wallace and the Scottish Workshop in Particle Physics for invitations to present parts of these lecture notes to graduate and postdoctoral students.

REFERENCES

Balian, R., J. M. Drouffe, and C. Itzykson, 1975a, Phys. Rev. D **11**, 2098.
Balian, R., J. M. Drouffe, and C. Itzykson, 1975b, Phys. Rev. D **11**, 2104.
Banks, T., J. Kogut, and R. Myerson, 1977, Nucl. Phys. B **129**, 493.
Banks, T., J. Kogut, and L. Susskind, 1976, Phys. Rev. D **13**, 1043.
Berg, B., and M. Lüscher, 1981, Nucl. Phys. B **190** [FS3], 412.
Bhanot, G., and M. Creutz, 1981, Phys. Rev. D **24**, 3212.
Binder, K., editor, 1979, *Monte Carlo Methods in Statistical Physics* (Springer, New York).
Bjorken, J., and S. Drell, 1964, *Relativistic Quantum Mechanics* (McGraw-Hill, New York).
Blairon, J.-M., R. Brout, F. Englert, and J. Greensite, 1980, Universite Libre de Buxelles preprint, September.
Brower, R., D. Kessler, and H. Levine, 1981, Phys. Rev. Lett. **47**, 621.
Cabibbo, N., and E. Marinari, 1982, University of Rome preprint No. 301, June 21.
Callan, C., 1970, Phys. Rev. D **2**, 1541.
Casher, A., 1979, Phys. Lett. B **83**, 395.
Casher, A., J. Kogut, and L. Susskind, 1974, Phys. Rev. D **10**, 732.
Coleman, S., 1975, Phys. Rev. D **11**, 2088.
Coleman, S., 1976, private communication.
Coleman, S., and D. J. Gross, 1973, Phys. Rev. Lett. **31**, 851.
Coleman, S., and E. Witten, 1980, Phys. Rev. Lett. **45**, 100.
Cornwell, J., R. Jackiw, and E. Tomboulis, 1974, Phys. Rev. D **10**, 2428.
Creutz, M., 1977, Phys. Rev. D **15**, 1128.
Creutz, M., 1980a, Phys. Rev. D **21**, 2308.
Creutz, M., 1980b, Phys. Rev. Lett. **45**, 313.
d'Adda, A., M. Lüscher, and P. Di Vecchia, 1978, Nucl. Phys. B **146**, 63.
Dashen, R., and D. J. Gross, 1981, Phys. Rev. D **23**, 2340.
David, F., 1980, Phys. Lett. B **96**, 371.
DeGrand, T. A., and D. Toussaint, 1980, Phys. Rev. D **22**, 2478.
DeWitt, B. S., 1967, Phys. Rev. **162**, 1195.
Drell, S. D., M. Weinstein, and S. Yankielowicz, 1976, Phys. Rev. D **14**, 487.
Duncan, A., R. Roskies, and H. Vaidya, 1982, University of Pittsburgh preprint PITT-82-6.
Eguchi, T., and H. Kawai, 1982, Phys. Rev. Lett. **48**, 1063.
Eichenherr, H., 1978, Nucl. Phys. B **146**, 214.
Elitzur, S., 1975, Phys. Rev. D **12**, 3978.
Fetter, A. L., and J. D. Walecka, 1971, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York).
Fischler, W., 1977, Nucl. Phys. B **129**, 157.
Forsythe, G. E., and R. A. Leibler, 1950, M.T.A.C. (Vol. 4), 127.
Fucito, R., E. Marinari, G. Parisi, and C. Rebbi, 1981, Nucl. Phys. B **180** [FS2], 369.
Gell-Mann, M., and F. Low, 1954, Phys. Rev. **95**, 1300.
Gilmer, G. H., and P. Bennema, 1972, J. Appl. Phys. **43**, 1347.
Gross, D. J., and F. Wilczek, 1973, Phys. Rev. Lett. **30**, 1343.
Hageman, R., and A. Young, 1979, *Applied Iterative Methods* (Academic, New York).
Halliday, G., and A. Schwimmer, 1981a, Phys. Lett. B **101**, 327.
Halliday, G., and A. Schwimmer, 1981b, Phys. Lett. B **102**, 337.
Hamber, H., and G. Parisi, 1981, Phys. Rev. Lett. **47**, 1792.
Hamer, C. J., and M. N. Barber, 1980, J. Phys. A **13**, L169.
Hamer, C. J., and M. N. Barber, 1981, J. Phys. A **14**, 241.
Hasenfratz, A., E. Hasenfratz, and P. Hasenfratz, 1981, Nucl. Phys. B **180**, 353.
Hasenfratz, A., and P. Hasenfratz, 1980, Phys. Lett. B **93**, 165.
Hasenfratz, A., P. Hasenfratz, Z. Kunszt, and C. B. Lang, 1982, Phys. Lett. B **110**, 289.
Hestenes, M. R., and E. Stiefel, 1952, J. Res. Natl. Bur. Stand. **49**, 409.
Hughes, R., 1979, Phys. Lett. B **97**, 246.
Itzykson, C., M. Peskin, and J.-B. Zuber, 1980, Phys. Lett. B **95**, 259.
Jones, D. R. T., J. Kogut, and D. K. Sinclair, 1979, Phys. Rev. D **19**, 1882.
Jose, J., L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, 1977, Phys. Rev. B **16**, 1217.
Kadanoff, L. P., 1977, Rev. Mod. Phys. **49**, 267.
Kawamoto, N., and J. Smit, 1981, Nucl. Phys. B **192**, 100.
Kogut, J., 1979, Rev. Mod. Phys. **51**, 659.
Kogut, J., R. B. Pearson, J. L. Richardson, J. Shigemitsu, and D. K. Sinclair, 1981, Phys. Rev. D **23**, 2945.
Kogut, J., R. B. Pearson, and J. Shigemitsu, 1979, Phys. Rev. Lett. **43**, 484.
Kogut, J., R. B. Pearson, and J. Shigemitsu, 1980, Phys. Lett. B **98**, 63.
Kogut, J., and J. Shigemitsu, 1981, Nucl. Phys. B **190**, 365.
Kogut, J., and D. K. Sinclair, 1981, Phys. Rev. D **24**, 1610.
Kogut, J., M. Stone, H. W. Wyld, W. R. Gibbs, J. Shigemitsu, S. H. Shenker, and D. K. Sinclair, 1983, Phys. Rev. Lett. **50**, 393.
Kogut, J., M. Stone, H. W. Wyld, J. Shigemitsu, S. H. Shenker, and D. K. Sinclair, 1982, Phys. Rev. Lett. **48**, 1140.
Kogut, J., M. Stone, H. W. Wyld, J. Shigemitsu, S. H. Shenker, and D. K. Sinclair, The Scales of Chiral Symmetry Breaking in SU(2) Gauge Theory, Ohio State University preprint, to ap-

pear.
Kogut, J., and L. Susskind, 1974, Phys. Rev. D **9**, 3501.
Kogut, J., and L. Susskind, 1975, Phys. Rev. D **11**, 395.
Kosterlitz, J. M., and D. J. Thouless, J. Phys. C **6**, 118.
Kramers, H. A., and G. H. Wannier, 1941, Phys. Rev. **60**, 252.
Kuti, J., 1982, Phys. Rev. Lett. **49**, 183.
Landau, L. D., and E. M. Lifschitz, 1975, *Electrodynamics of Continuous Media* (Pergamon, Oxford).
Langevin, P. 1908, C. R. Acad. Sci. Paris **146**, 530.
Lee, T. D., 1981, *Particle Physics and Introduction to Field Theory* (Harwood, New York).
Lüscher, M., 1981, Nucl. Phys. B **180**, 317.
Lüscher, M., 1981, Bern preprint BUTP-10, to appear in Commun. Math. Phys.
Lüscher, M., 1982, Nucl. Phys. B **200** [FS4], 61.
Lüscher, M., G. Münster, and P. Weisz, 1981, Nucl. Phys. B **180** [FS2], 1.
Mack, G., and V. B. Petkova, 1980, Ann. Phys. (N.Y.) **125**, 117.
Mandelstam, S., 1976, Phys. Rep. C **23**, 245.
Marinari, E., G. Parisi, and C. Rebbi, 1981, Phys. Rev. Lett. **47**, 1795.
Mermin, N. D., and H. Wagner, 1966, Phys. Rev. Lett. **17**, 1133.
Metropolis, N., A. W. Rosenbluth, A. H. Teller, and E. Teller, 1953, J. Chem. Phys. **21**, 1087.
Münster, G., 1981, Nucl. Phys. B **128** [FS2], 23.
Nambu, Y., and G. Jona-Lasinio, 1961, Phys. Rev. **122**, 345.
Nauenberg, M., and B. Lautrup, 1980, Phys. Rev. Lett. **45**, 1755.
Nielsen, H. B., and M. Ninomiya, 1981, Nucl. Phys. **185**, 20.
Nightingale, M. P., 1976, Physica A **83**, 561.
Osterwalder, K., and E. Seiler, 1978, Ann. Phys. (N.Y.) **110**, 440.
Pietarinen, E., 1981, Nucl. Phys. B **190**, 349.
Politzer, H. D., 1973, Phys. Rev. Lett. **30**, 1346.
Polyakov, A. M., 1975a, Phys. Lett. B **59**, 79.
Polyakov, A. M., 1975b, Phys. Lett. B **59**, 82.
Polyakov, A. M., 1981, Phys. Lett. B **103**, 207.
Raby, S., S. Dimopoulos, and L. Susskind, 1980, Nucl. Phys. B **169**, 373.
Samuel, S., 1980, J. Math. Phys. **21**, 1842.
Satz, H., 1982, Phys. Rep. **88**, 321.
Savit, R., 1980, Rev. Mod. Phys. **52** 453.
Scalapino, D. J., and R. L. Sugar, 1981, Phys. Rev. Lett. **46**, 519.
Shenker, S. H., M. Stone, J. Kogut, J. Bartholomew, and J. Sloan, 1983, Chiral Symmetry Breaking in Lattice Electrodynamics, University of Chicago preprint, to appear.
Shenker, S., and J. Tobochnik, 1980, Phys. Rev. B **22**, 4462.
Stack, J. D., 1983, Phys. Rev. D **27**, 412.
Stone, M., 1982, private communication.
Susskind, L., 1977, "Coarse grained quantum chromodynamics," in *Weak and Electromagnetic Interactions at High Energy,* edited by R. Balian and C. H. Llewellyn Smith, Les Houches 29, 1976 (North-Holland, Amsterdam), p. 207.
Swendsen, R. H., 1978, Phys. Rev. B **18**, 492.
Symanzik, K., 1970, Commun. Math. Phys. **49**, 424.
Symanzik, K., 1982, in *Mathematical Problems in Theoretical Physics,* edited by R. Schroder, Lecture Notes in Physics, 153 (Springer, Berlin).
't Hooft, G., 1972, unpublished remarks at the Marseilles Conference on Gauge Theories.
't Hooft, G., 1976, in *High Energy Physics,* edited by A. Zichichi, Palermo, 1976 (Editrice Compositori, Bologna).
van Beijeren, H., 1975, Commun. Math. Phys. **40**, 1.
Villain, J., 1975, J. Phys. C **36**, 581.
Wasow, W. R., 1951, M.T.A.C. (Vol. 5), 78.
Weeks, J. D., G. H. Gilmer, and H. J. Leamy, 1973, Phys. Rev. Lett. **31**, 549.
Wegner, F., 1971, J. Math. Phys. **12**, 2259.
Weingarten, D., 1982, Phys. Lett. B **109**, 57.
Weingarten, D., and D. Petcher, 1981, Phys. Lett. B **99**, 33.
Wilson, K. G., 1974, Phys. Rev. D **14**, 2455.
Wilson, K. G., 1977, in "Quarks: from paradox to myth," in *New Phenomena in Subnuclear Physics,* edited by A. Zichichi, Erice, 1975 (Plenum, New York), p. 13.
Wilson, K. G., 1980, in "Monte-Carlo calculations for the lattice gauge theory," in *Recent Developments in Gauge Theories,* edited by G. 't Hooft *et al.,* Cargèse, 1979 (Plenum, New York), p. 363.
Wilson, K. G., and J. Kogut, 1974, Phys. Rev. C **12**, 75.
Wortis, M., 1974, in *Phase Transitions and Critical Phenomena,* edited by C. Domb and M. S. Green (Academic, London), Vol. 3, p. 113.
Yaffe, L. G., and B. Svetitsky, 1982, Phys. Rev. D **26**, 963.
Yang, C. N., and R. L. Mills, 1954, Phys. Rev. **96**, 1605.