(1973); S. Weinberg, *ibid.* 7, 2887 (1973); R. Jackiw, Ref. 4.

[19]The analysis of the physical interpretation of $\Gamma(\phi, G)|_{\text{static}}$ is an adaptation to the present context of the corresponding argument for $\Gamma(\phi)|_{\text{static}}$. That discussion is due to K. Symanzik, Commun. Math. Phys. 16, 48 (1970). We

learned it from S. Coleman, in proceedings of the Lectures given at the International Summer School of Physics "Ettore Majorana," 1973 (unpublished).

[20]R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D (to be published).

Confinement of quarks*

Kenneth G. Wilson

Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14850

(Received 12 June 1974)

A mechanism for total confinement of quarks, similar to that of Schwinger, is defined which requires the existence of Abelian or non-Abelian gauge fields. It is shown how to quantize a gauge field theory on a discrete lattice in Euclidean space-time, preserving exact gauge invariance and treating the gauge fields as angular variables (which makes a gauge-fixing term unnecessary). The lattice gauge theory has a computable strong-coupling limit; in this limit the binding mechanism applies and there are no free quarks. There is unfortunately no Lorentz (or Euclidean) invariance in the strong-coupling limit. The strong-coupling expansion involves sums over all quark paths and sums over all surfaces (on the lattice) joining quark paths. This structure is reminiscent of relativistic string models of hadrons.

I. INTRODUCTION

The success of the quark-constituent picture both for resonances and for deep-inelastic electron and neutrino processes makes it difficult to believe quarks do not exist. The problem is that quarks have not been seen. This suggests that quarks, for some reason, cannot appear as separate particles in a final state. A number of speculations have been offered as to how this might happen.[1]

Independently of the quark problem, Schwinger observed many years ago[2] that the vector mesons of a gauge theory can have a nonzero mass if vacuum polarization totally screens the charges in a gauge theory. Schwinger illustrated this result with the exact solution of quantum electrodynamics in one space and one time dimension, where the photon acquires a mass $\sim e^2$ for any nonzero charge e [e has dimensions of $(\text{mass})^{1/2}$ in this theory]. Schwinger suggested that the same effect could occur in four dimensions for sufficiently large couplings.

Further study of the Schwinger model by Lowenstein and Swieca[3] and Casher, Kogut, and Susskind[4] has shown that the asymptotic states of the model contain only massive photons, not electrons. Nevertheless, as Casher *et al.* have shown in detail, the electrons are present in deep-inelastic processes and behave like free pointlike

particles over short times and short distances. The polarization effects which prevent the appearance of electrons in the final state take place on a longer time scale (longer than $1/m_\gamma$, where m_γ is the photon mass).

A new mechanism which keeps quarks bound will be proposed in this paper. The mechanism applies to gauge theories only. The mechanism will be illustrated using the strong-coupling limit of a gauge theory in four-dimensional space-time. However, the model discussed here has a built-in ultraviolet cutoff, and in the strong-coupling limit all particle masses (including the gauge field masses) are much larger than the cutoff; in consequence the theory is far from covariant.

The confinement mechanism proposed here is soft (long-time scale). However, in the model discussed here the cutoff spoils the possibility of free pointlike behavior for the quarks.

The model discussed in this paper is a gauge theory set up on a four-dimensional Euclidean lattice. The inverse of the lattice spacing a serves as an ultraviolet cutoff. The use of a Euclidean space (i.e., imaginary instead of real times) instead of a Lorentz space is not a serious restriction; the energy eigenstates (including scattering states) of the lattice theory can be determined from the "transfer-matrix" formalism as has been discussed by suri[5] and reviewed by Wilson and Kogut.[6] A brief discussion of the

transfer-matrix method is given in Sec. III.

In Schwinger's speculations about four dimensions, the photon mass would be zero for any charge e less than a critical coupling e_c; for $e > e_c$ the photon mass would be nonzero and vary with e. Figure 1 shows how a plot of m_γ vs e might look. The point e_c is a point of nonanalyticity. Similar nonanalytic points, called critical points, occur in solid-state physics at certain types of phase transitions. Consider, for instance, a ferromagnet in the absence of an external field. For any temperature above the Curie temperature T_C, the spontaneous magnetization M is 0. Below T_C, M is nonzero and a function of temperature. At T_C there may be either a first-order phase transition (in which case M is discontinuous at T_C) or a second-order phase transition (critical point) for which $M \rightarrow 0$ as $T \rightarrow T_C$ from either side.

By analogy with the solid-state situation one can think of the transition from zero to nonzero photon mass as a change of phase: this analogy is best understood by imagining the particles of quantum electrodynamics to be the excitations of a medium (the ether). In this case it is the ether which undergoes a change of phase at e_c. There is again a question whether this change of phase is first-order (cf. Fig. 2) or second-order (Fig. 1). (Coleman and Weinberg[7] have found a nontrivial example of a first-order transition in another context.)

The model discussed in this paper is a single Abelian gauge field coupled (with strength g) to massive quarks. In weak coupling the gauge field behaves like a normal free zero-mass field (despite modifications introduced in the lattice quantization) and the quarks are unbound. In strong coupling the gauge field is massive and the quarks are bound, showing the existence of the second phase. Thus there should be a phase transition at some intermediate value of g. Nothing is known about this transition at the present time.

The quantization procedure and strong-coupling approximation described in this paper can be applied to non-Abelian gauge theories also. This will be explained briefly in Sec. III.

An extraordinary feature of the strong-coupling expansion of the lattice theory (see Sec. IV) is that it has the same general structure as the relativistic string models of hadrons.[8] The vacuum expectation values of the gauge theory involve (in the strong-coupling expansion) sums over all quark paths and sums over all surfaces connecting these paths; the surfaces are generated by the gauge field treated in strong coupling. The paths and surfaces are defined on a discrete lattice. There are geometrical difficulties in relat-
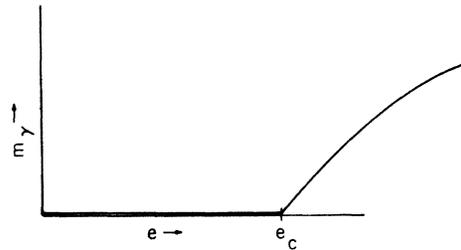

FIG. 1. Speculative plot of photon mass vs renormalized charge e, in unknown units. The transition at e_c is second-order (see text).

ing the surfaces on the lattice to the continuum surfaces of the string models; it is not known at present whether these difficulties can be overcome.

In Sec. II the nature of the quark binding mechanism will be discussed, qualitatively. In Sec. III the gauge theory will be formulated on a discrete lattice, both classically and quantum mechanically. In Sec. IV the strong-coupling expansion for the lattice gauge theory is explained. In Sec. V a cursory discussion of weak coupling is given. In Sec. VI there is a brief discussion of the problem of Lorentz invariance and the relation to string models.

II. QUARK BINDING MECHANISM

The binding mechanism will be explained in this section using the Feynman path-integral picture. The path-integral framework will be used in an intuitive rather than a formal way. Consider the current-current propagator

$$D_{\mu\nu}(x) = \langle \Omega | TJ_\mu(x) J_\nu(0) | \Omega \rangle , \qquad (2.1)$$

whose Fourier transform determines the e^+-e^- annihilation cross section into hadrons. Assume that the currents $J_\mu(x)$ are built from quark fields as in the quark-parton model. Assume that the quarks interact through a single gauge field. (The restriction to one field is only for simplicity.) In the Feynman path-integral picture the propagator
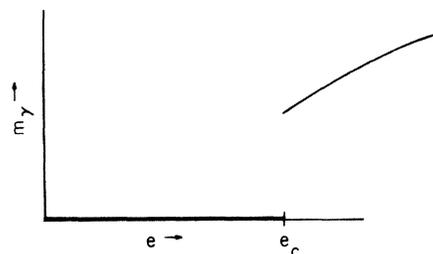

FIG. 2. Speculative plot of photon mass vs renormalized charge e if there is a first-order transition at e_c.

$D_{\mu\nu}(x)$ is given by a weighted average over all possible classical quark paths and all possible classical values of the gauge field. The currents $J_\mu(x)$ and $J_\nu(0)$ are thought of as producing a quark pair at the origin which later annihilate at the point x: One has to sum over all paths joining the points 0 and x for each of the pair of quarks.

An example of paths for the quark and antiquark are shown in Fig. 3. The vacuum can also emit and absorb quark pairs; this leads to further closed loops as illustrated in Fig. 4. All possible loops must be summed over too. It is also possible to have independent loops for the points 0 and x (Fig. 5), but this possibility will not be important here.

The weight associated with a given quark path or set of paths includes a factor of $\exp[ig\oint A_\mu(x) \times ds^\mu]$, where $A_\mu(x)$ is the gauge field. Here $\oint \cdots ds^\mu$ is a line integral or a sum over line integrals for each of the quark-antiquark loops, including the loop joining the points 0 and x. The constant g is the coupling constant of the gauge theory. There are further weight factors independent of A. Finally, independently of the quark paths there is another weight factor, namely, the exponential of the free action for the gauge field. The combined weight factor is then averaged over all quark paths and all gauge fields $A_\mu(x)$ to give the current-current propagator.

In order that quarks exist as separate final-state particles it must be possible to have quark-antiquark loops with well-separated quark and antiquark lines, at least when x and 0 are far apart. This is illustrated in Fig. 6(a). If the quark and antiquark paths are unlikely to separate beyond a fixed size, say 10^{-13} cm [see Fig. 6(b)], then clearly no detector will see a quark or antiquark in isolation.

It is assumed in this discussion that vacuum loops are not important. If vacuum loops are important enough then space will be filled with a high *density* of vacuum-produced quark-antiquark pairs. In particular, there will be many quark-antiquark pairs inside a detector of macroscopic size. The question then is whether there can be an excess of quarks over antiquarks, or vice versa, in a region of macroscopic size. This is
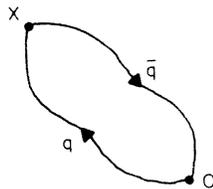

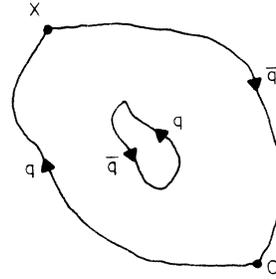FIG. 3. An example of quark (q) and antiquark ($\bar{q}$) paths connecting the points 0 and x.

FIG. 4. Example of current loop (as in Fig. 3) with an extra vacuum loop.

a more difficult question to answer and will not be discussed in this paper.

Note that x must be large: If x is small there is little likelihood of finding a large size loop. This may seem a bit peculiar: One expects quarks to appear in the final state of e^+-e^- annihilation only at large virtual-photon momentum q if they appear at all, and large q means small x, not large x. The answer to this paradox has two parts. First, the important paths in the Feynman path integral bear no detailed relation to possible physical final states (the paths are paths of bare particles, not physical particles). Secondly, large x does not necessarily mean small q. In fact the study of whether well-separated quark-antiquark paths exist for large x is really a search for a quark-antiquark threshold in e^+-e^- annihilation, which would contribute a term $\sim \exp(2mi\sqrt{x^2})$ to the current-current propagator for *large* x, where m is the quark mass. (Here $\sim$ means up to a power of x^2.) Such a term corresponds in momentum space to the singularity at the threshold $q^2 = (2m)^2$.

Suppose the gauge-field averaging is performed before the quark-paths averaging. Then one determine the average over all gauge fields of the weight factor $\exp[i\oint gA_\mu(y)ds^\mu]$ weighted further with the exponential of the free gauge-field action. For an Abelian gauge theory this average can be computed explicitly: It is
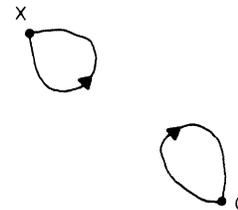

FIG. 5. Example of separate quark loops for the points 0 and x. (Integration over the gauge field produces gauge propagators which connect these loops.)

$$\exp\left[-g^2 \oint ds^\mu \oint ds'^\nu D_{\mu\nu}(y - y')\right],$$

where $D_{\mu\nu}(y - y')$ is the free gauge-field propagator.

The quark binding mechanism can be seen by comparing the above expression for one space dimension and three space dimensions. In three space dimensions this calculation gives no binding (the binding occurs only with a modified gauge-field action: see Sec. IV), while there is binding in one space dimension. In three space dimensions $D_{\mu\nu}(y - y')$ behaves as $(y - y')^{-2}$. In consequence large values of $(y' - y)$ are negligible in the double line integral. Hence, the double integral is proportional to P, where P is the length of the loop. Unfortunately the integral is divergent at $y' = y$; a cutoff is needed for the integral to make sense. Since a cutoff will be introduced anyway in this paper, this is not a major concern. For simple loops, the perimeter P is roughly of order $(x^2)^{1/2}$ (ignoring the case that x is close to the light cone). Thus one has an exponential of the type one expects when free quarks are present.

In one space dimension, $D_{\mu\nu}(y - y')$ behaves as $\ln[(y - y')^2]$ for $y' - y$ large, and y' and y can freely range separately over the loop. In this case the double integral is proportional to P^2. Now the gauge-field average behaves as e^{-icP^2}, where c is a constant. In this case the contribution of large loops is heavily suppressed and there are no free quarks. [The case of nearby quark-antiquark pairs as in Fig. 6(b) is special—in this case large $y - y'$ is unimportant due to cancellation between the quark path and the nearby and oppositely directed antiquark path. In this case the double integral behaves as P, not P^2, but in this case there are no isolated quarks.]

In the strongly coupled lattice gauge theory described in later sections, the gauge-field average of $\exp[ig \oint A_\mu(x) ds^\mu]$ behaves as $\exp(ic'A)$, where A is the enclosed area of the loop. This heavily suppresses large loops, such as in Fig. 6(a), where A is of order P^2. One can think of one factor P as being roughly $(x^2)^{1/2}$, the other P as being analogous to a mass multiplying $(x^2)^{1/2}$. Since $P \to \infty$ as $x \to \infty$, the quark-antiquark threshold is at infinite mass.

In all these calculations one can have a large loop if there is a nearby vacuum loop (Fig. 7). In this case one always gets $e^{ic''P}$ behavior. For example, in the strong-coupling case the relevant enclosed area is the area between the two loops which is proportional to the perimeter P provided the separation of the two loops is fixed independently of P. This is in accord with Schwinger's picture. While an isolated well-separated loop
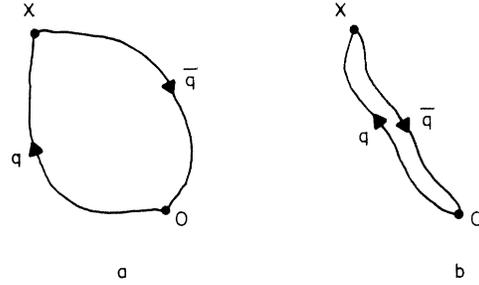

FIG. 6. (a) Loop with well-separated quark and antiquark. (b) Loop with small separation between quark and antiquark.

may be suppressed (due to P^2 or A dependence in the exponential) a loop closely shielded by a vacuum-polarization loop is always unsuppressed.

The binding mechanism proposed here is soft: The exponential damping is associated with large size loops having large areas. The behavior of small loops is irrelevant to the binding mechanism. Also for small loops both their area and perimeter are small and neither is of great importance in an exponential.

The mechanism discussed here works equally well for Dirac quarks or scalar quarks. This is in contrast to the Higgs mechanism which can wipe out the charge of scalar particles only.

III. LATTICE QUANTIZATION OF GAUGE FIELDS

A. Classical action on a lattice

In this section the gauge-field action (space-time integral of the Lagrangian) will be defined on a discrete lattice with spacing a in Euclidean space-time. The simplest way to proceed is to consider a continuum action, substitute finite-difference approximations for derivatives, and replace the space-time integral by a sum over the lattice sites. However, the result of this is an action which is not gauge-invariant for nonzero a. Because of the vagaries of renormalization this is likely to mean that the quantized theory still lacks
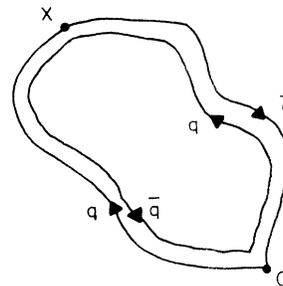

FIG. 7. Quark-antiquark loop with nearby vacuum loop.

gauge invariance in the limit $a \to 0$ (if such a limit exists). The alternative is to formulate gauge invariance for a lattice theory, and tinker with the action so that it is gauge invariant for any a. This alternative will be pursued here.

For convenience a charged Dirac field ψ coupled to a single gauge field A_μ will be discussed in detail. Generalizations to non-Abelian gauge groups will be noted later.

On the lattice the fields are ψ_n, $\bar{\psi}_n$, and $A_{n\mu}$, where n is a four-vector with integer components referring to points on a simple hypercubic lattice. A simple action on the lattice for the Dirac field (ignoring the gauge field for now) has the form

$$A_\psi = -a^4 \sum_n m_0 \bar{\psi}_n \psi_n$$
$$+ \tfrac{1}{2} a \sum_n \sum_\mu \bar{\psi}_n \gamma_\mu (\psi_{n+\hat{\mu}} - \psi_{n-\hat{\mu}}) , \qquad (3.1)$$

where m_0 is the bare mass; $\hat{\mu}$ is a unit vector along the axis μ; $a^4 \sum_n$ replaces the space-time integration of the continuum theory, and $(\psi_{n+\hat{\mu}} - \psi_{n-\hat{\mu}})/2a$ replaces $\nabla \psi$. There is no over-all factor of i due to the Euclidean metric. A gauge transformation on the lattice can be defined as follows:

$$\psi_n \to e^{i y_n g} \psi_n , \qquad (3.2)$$

$$\bar{\psi}_n \to e^{-i y_n g} \bar{\psi}_n , \qquad (3.3)$$

$$A_{n\mu} \to A_{n\mu} - (y_{n+\hat{\mu}} - y_n)/a , \qquad (3.4)$$

where g is the coupling constant and y_n is arbitrary. The terms $\bar{\psi}_n \psi_{n+\hat{\mu}}$ and $\bar{\psi}_n \psi_{n-\hat{\mu}}$ are not invariant to this transformation; the corresponding gauge-invariant expressions are $\bar{\psi}_n \psi_{n+\hat{\mu}}$ $\times \exp(iagA_{n\mu})$ and $\bar{\psi}_n \psi_{n-\hat{\mu}} \exp(-iagA_{n-\hat{\mu},\mu})$. Thus, a gauge-invariant form for A_ψ is

$$A_\psi = \tfrac{1}{2} a^3 \sum_n \sum_\mu (\bar{\psi}_n \gamma_\mu \psi_{n+\hat{\mu}} e^{iag A_{n\mu}} - \bar{\psi}_{n+\hat{\mu}} \gamma_\mu \psi_n e^{-iag A_{n\mu}})$$
$$- a^4 \sum_n m_0 \bar{\psi}_n \psi_n . \qquad (3.5)$$

It is convenient to define

$$B_{n\mu} = ag A_{n\mu} . \qquad (3.6)$$

In the action A_ψ, the field $B_{n\mu}$ acts like an angular variable: A_ψ is periodic in $B_{n\mu}$ with period 2π. The free gauge-field action will be defined to preserve this property. This does not mean that $A_{n\mu}$ is an angular variable in the continuum limit. Owing to the relation (3.6), $A_{n\mu}$ becomes an infinitesimal angular variable $A_\mu(x)$ for $a \to 0$; such a variable has the range $-\infty < A_\mu(x) < \infty$ without any periodicity.

A gauge-invariant lattice approximation for $\nabla_\mu A_\nu - \nabla_\nu A_\mu$ is

$$F_{n\mu\nu} = (A_{n+\hat{\mu},\nu} - A_{n\nu} - A_{n+\hat{\nu},\mu} + A_{n\mu})/a . \qquad (3.7)$$

It is convenient to define a rescaled form of $F_{n\mu\nu}$:

$$f_{n\mu\nu} = a^2 g F_{n\mu\nu}$$
$$= B_{n\mu} + B_{n+\hat{\mu},\nu} - B_{n+\hat{\nu},\mu} - B_{n\nu} . \qquad (3.8)$$

A simple lattice action for the gauge field which preserves periodicity is

$$A_B = \frac{1}{2g^2} \sum_{n\mu\nu} e^{if_{n\mu\nu}} . \qquad (3.9)$$

In the continuum limit, $f_{n\mu\nu} \to 0$ due to the factor a^2 in the definition of $f_{n\mu\nu}$. Thus for small a, one can write

$$A_B \simeq \frac{1}{2g^2} \sum_{n\mu\nu} (1 + if_{n\mu\nu} - \tfrac{1}{2} f_{n\mu\nu}^2 - \cdots) . \qquad (3.10)$$

The constant term is irrelevant. The linear term in $f_{n\mu\nu}$ is 0 because $f_{n\mu\nu}$ is odd in the indices μ and ν. The quadratic term gives

$$A_B \simeq -\tfrac{1}{4} a^4 \sum_{n\mu\nu} F_{n\mu\nu}^2 , \qquad (3.11)$$

which is the conventional gauge-field action in a lattice approximation. The terms involving $f_{n\mu\nu}^3$, $f_{n\mu\nu}^4$, etc. all vanish for $a \to 0$ even after removing a factor a^4 to convert $\sum_n$ into an integral.

The full action may now be written

$$A = -c \sum_n \bar{\psi}_n \psi_n + K \sum_n \sum_\mu (\bar{\psi}_n \gamma_\mu \psi_{n+\hat{\mu}} e^{i B_{n\mu}} - \bar{\psi}_{n+\hat{\mu}} \gamma_\mu \psi_n e^{-i B_{n\mu}}) + \frac{1}{2g^2} \sum_n \sum_{\mu\nu} e^{if_{\mu\nu}} , \qquad (3.12)$$

with $c = m_0 a^4$, $K = a^3/2$. This action reduces to the usual continuum action for $a \to 0$; for finite a it is gauge invariant and periodic in the gauge field. Note, however, that the continuum limit is a *classical* limit in which the lattice variables ψ_n, $\bar{\psi}_n$, and $A_{n\mu}$ approach continuum functions $\psi(x)$, $\bar{\psi}(x)$, and $A_\mu(x)$ with $x = na$. The continuum limit of the quantized theory is much harder to discuss owing to renormalization problems.

B. Quantization

The problem of principal interest here is the quantization of the gauge field. Therefore, the gauge field will be quantized by itself to start with. Later the quantization of ψ will be discussed. At the end of this section the generalizations to non-Abelian gauge theories will also be described.

The quantization of the lattice gauge theory will

be carried out in two steps. The first step will
be to define a lattice version of Euclidean vacuum
expectation values, starting from a lattice ver-
sion of the Feynman path integral. The second
step will be to define a quantum theory on the lat-
tice, which will allow the introduction of a real
time variable and the definition of particle states
and scattering amplitudes. In both cases the lat-
tice provides an ultraviolet cutoff and there is no
Lorentz invariance. Lorentz invariance can only
be achieved in the limit a (lattice spacing) $\to 0$, if
at all, and in practice this is a difficult limit to
evaluate.

As discussed in Sec. I, one would like to calcu-
late the gauge-field average of $\exp[ig\oint A_\mu(x)ds^\mu]$
weighted with the gauge-field action. On a lattice
the line integral becomes a sum over a closed
path P on the lattice (see, e.g., Fig. 8). The sum
has the form: $i\sum_P(\pm)B_{n\,\mu}$, where a particular $B_{n\,\mu}$
is present in the sum if the path connects the sites
n and $n+\hat\mu$ ($-B_{n\,\mu}$ appears if the path goes from
$n+\hat\mu$ to n).

On the lattice, an average over all gauge fields
involves integrating over all values of the $B_{n\mu}$ for
all n and μ. Normally one would have integrals
over an infinite range: $-\infty < B_{n\mu} < \infty$, but because
of the periodicity in $B_{n\mu}$ there is no point to in-
tegrating over more than a single period. Thus
the lattice version of the gauge-field average is

$$I(P) = Z^{-1}\left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu}\right)$$

$$\times \exp\left(i\sum_P(\pm)B_{n\mu} + \frac{1}{2g^2}\sum_{n\mu\nu}e^{if_{n\mu\nu}}\right),$$

$$(3.13)$$

with

$$Z = \left(\prod_m\prod_\nu\int_{-\pi}^{\pi} dB_{m\nu}\right)\exp\left(\frac{1}{2g^2}\sum_{n\mu\nu}e^{if_{n\mu\nu}}\right).$$

$$(3.14)$$

Note that no gauge-fixing term has been added
to the action. The finite range of $B_{n\mu}$ makes a
gauge-fixing term unnecessary. In continuum
gauge theories where $A_\mu(x)$ has an infinite range
$[-\infty < A_\mu(x) < \infty]$ a gauge-fixing term is essential
to have a convergent functional integral. The rea-
son for this is that the volume in path-integral
space generated by all possible gauge transforma-
tions is infinite; the gauge-fixing term provides
a convergence factor in this volume.[10] In the lat-
tice theory the total volume of integration is finite
if the lattice itself is of finite extent; no conver-
gence factor is required. For a lattice of infinite
extent there are divergences due to the infinite
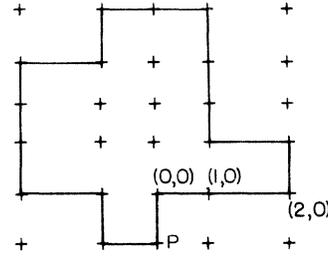number of integrations (in other words, the in-

FIG. 8. Example of a lattice path P.

finite lattice volume), but these divergences are
normally removed by the division by Z (this divi-
sion is equivalent to removing all vacuum loops
in perturbation theory).

One can define more conventional vacuum ex-
pectation values in a similar manner. For in-
stance, one can define a propagator. In the ab-
sence of a gauge-fixing term it is awkward to de-
fine a propagator for the gauge field $B_{n\mu}$ itself;
instead one can define a gauge-invariant propa-
gator as

$$D_{n\mu\pi,\sigma\tau} = Z^{-1}\left(\prod_m\prod_\nu\int_{-\pi}^{\pi}dB_{m\nu}\right)$$

$$\times\exp\left(if_{n\mu\pi} - if_{\sigma\sigma\tau} + \frac{1}{2g^2}\sum_{m\nu\rho}e^{if_{m\nu\rho}}\right).$$

$$(3.15)$$

This is a propagator for the operator $e^{if_{n\mu\nu}}$; it is
defined only for the lattice points n of a Euclidean
space-time lattice. If the lattice spacing is a, this
means the propagator is defined only for imag-
inary times of the form in_0a, where n_0 is an in-
teger.

A theory defined only for discrete imaginary
values of the time leaves much to be desired.
Fortunately, one can generalize the theory to de-
fine a Hamiltonian for a quantized theory. The
particle eigenstates and scattering amplitudes of
the theory can then be obtained, in principle, by
diagonalizing the Hamiltonian. The Hamiltonian
will be defined using the transfer-matrix formal-
ism. Only a brief discussion of the transfer-ma-
trix approach will be given here. For a review of
the ideas see Wilson and Kogut.[6] A detailed dis-
cussion including approximate calculation of sin-
gle-particle energies and scattering amplitudes
in a simple scalar field theory is given by suri.[5]

Consider the expression for Z. Introduce finite
bounds on the lattice coordinate n_0, say

$$-N \le n_0 \le N .$$

$$(3.16)$$

Introduce periodic boundary conditions (see be-
low). Then one can write Z as the trace of a ma-
trix V, more precisely

$$Z = \operatorname{Tr} V^{2N+1} . \tag{3.17}$$

This formula is made possible by the fact that each term in the action A involves no more than two adjacent values of n_0.

To set up the matrix V, one must first understand the space on which it acts. The space used here is the space of all functions $\psi(B^{\rightarrow}_{ni})$ (periodic in each $B^{\rightarrow}_{ni}$ with period 2π), where the index i runs from 1 to 3 only, and the lattice variable $\vec{n}$ has only three components (n_1, n_2, n_3). The matrix V will be defined as a function of two sets of arguments, say $B^{\rightarrow}_{ni}$ and $B^{\prime}_{ni}$, these two sets of arguments referring to the space-time fields B_{ni} for two adjacent values of n_0. Matrix multiplication of two V's involves integrations over a set of variables $\{B^{\rightarrow}_{ni}\}$. Define V as

$$V = \prod_{\vec{m}} \int_{-\pi}^{\pi} dB^{\rightarrow}_{\vec{m}0} e^{U} . \tag{3.18}$$

The quantity U as written out below looks complicated, but all it is is that part of the action A referring to a given nearest-neighbor pair of values for n_0; terms in A referring to a single value of n_0 are divided equally between the matrices connecting n_0 to $n_0 + 1$ and n_0 to $n_0 - 1$. The result is

$$U = \frac{1}{4g^2} \sum_{\vec{n}} \sum_i \sum_j (e^{if^{\prime}_{nij}} + e^{if^{\rightarrow}_{nij}})$$
$$+ \frac{1}{2g^2} \sum_{\vec{n}} \sum_i (e^{if^{\rightarrow}_{nio}} + e^{if^{\rightarrow}_{noi}}) , \tag{3.19}$$

where

$$f^{\rightarrow}_{nij} = B^{\rightarrow}_{ni} + B^{\rightarrow}_{n+\hat{i},j} - B^{\rightarrow}_{n+\hat{j},i} - B^{\rightarrow}_{nj} \tag{3.20}$$

and

$$f^{\rightarrow}_{nio} = B^{\rightarrow}_{ni} + B^{\rightarrow}_{n+i,0} - B^{\prime}_{ni} - B^{\rightarrow}_{no} . \tag{3.21}$$

With the definition of V given here, the trace $\operatorname{Tr} V^{2N+1}$ is easily seen to reproduce all the integrations involved in the equation for Z, and the sum of the $2N+1$ exponents U reproduces the action A except for some additional terms coupling the boundary $n_0 = N$ to the boundary $n_0 = -N$ to achieve a periodic structure.

Note that the 0 components of $B_{n\mu}$ have received special treatment. This is because there are no terms in A involving B_{n0} for more than one value of n_0; this makes it possible to include integrations over B_{n0} in the definition of V rather than in the definition of matrix multiplication.

The matrix V is used to define the quantized theory. Briefly, this is accomplished as follows. V is a Hermitian matrix, i.e.,

$$V(B, B')^* = V(B', B) . \tag{3.22}$$

[This result can be verified by close examination of Eqs. (3.18)–(3.21). One must remember that the variables $B^{\rightarrow}_{n0}$ are integrated out in the definition of V. This means that in forming the complex conjugate of V one can also make the change of variable $B^{\rightarrow}_{n0} \rightarrow -B^{\rightarrow}_{n0}$.] Hence V has a complete orthogonal set of eigenstates Ψ and eigenvalues λ. The Hamiltonian H is now *defined* as follows: The eigenstates of H are the eigenstates of V; the eigenvalues of H are given by

$$E = -a^{-1} \ln\lambda , \tag{3.23}$$

where λ is the corresponding eigenvalue of V. The reason for the factor a^{-1} will be evident shortly. The reason for using the logarithm is so that the energies of multiparticle scattering states will be the sum of single-particle energies (see suri[5]).

A problem arises with this definition if V has any negative eigenvalues λ. If this were to happen, H would have complex eigenvalues. This did not happen in the case studied by suri[5]; whether it happens here the author does not know; this question must be studied further. Even if V has negative eigenvalues, they may be irrelevant in the limit $a \rightarrow 0$ if such a limit exists.

The definition (3.23) means that

$$V = e^{-aH} . \tag{3.24}$$

This means V is the operator which propagates a state through an imaginary time ia. It is a consequence of this that the propagator $D_{n\mu\pi, \sigma\tau}$ is a vacuum expectation value for imaginary time $in_0 a$ in the theory with Hamiltonian H. For proof of this statement see Refs. 5 and 6.

The lattice quantization procedure can be extended to non-Abelian gauge theories. This is done as follows. In place of the single variable $B_{n\mu}$, one has a set of variables $B^\alpha_{n\mu}$ where α is an internal index. For each n and μ, $B^\alpha_{n\mu}$ is to parametrize an element $b_{n\mu}$ of the gauge group. In place of $\exp(iB_{n\mu})$ one substitutes $U(b_{n\mu})$, where U is the unitary matrix representing $b_{n\mu}$ in the quark representation. The product $\bar\psi_n \gamma_\mu \psi_{n+\hat\mu} \exp(iB_{n\mu})$ is replaced by $\bar\psi_n \gamma_\mu U(b_{n\mu})\psi_{n+\hat\mu}$. A gauge transformation is defined by a set of group transformations y_n; under these transformations

$$\psi_n \rightarrow U(y_n)\psi_n , \tag{3.25}$$

$$\bar\psi_n \rightarrow \bar\psi_n U^\dagger(y_n) , \tag{3.26}$$

$$b_{n\mu} \rightarrow y_n b_{n\mu} y_{n+\hat\mu}^{-1} , \tag{3.27}$$

$$U(b_{n\mu}) \rightarrow U(y_n) U(b_{n\mu}) U^\dagger(y_{n+\hat\mu}) , \tag{3.28}$$

where $y_n b_{n\mu} y_{n+\hat\mu}^{-1}$ is computed according to the multiplication law of the group. A simple gauge-invariant action for the gauge field is

$$A_B = \frac{1}{2g^2} \sum_{n\mu\nu} \text{Tr} \, U_A(b_{n\mu} b_{n+\hat{\mu},\nu} b_{n+\hat{\nu},\mu}^{-1} b_{n\nu}^{-1}) \, ,$$

$$(3.29)$$

where the unitary transformation U_A is taken from the adjoint representation of the group. (Any representation will do as well.) In the classical continuum limit, $b_{n\mu}$ becomes an infinitesimal group transformation $[B_{n\mu}^\alpha = gaA_\mu^\alpha(na)$ with A_μ^α fixed as $a \to 0]$; in this limit one can show, with some effort, that A_B reduces to the standard continuum Yang-Mills action. When the non-Abelian theory is quantized, the integrations over $b_{m\nu}$ are group integrations over all elements of the group. These are compact integrations (for compact groups) and no gauge-fixing term is required.

Since quark binding can be illustrated using Abelian theory, the non-Abelian theory will not be studied further.

Finally, the quantization of the Dirac field will be discussed. It is convenient initially to quantize the Dirac field in an analogous manner to the gauge field. The only problem that occurs is to define "integration" for a Fermi field. This can be done.[11]

The property of the integral that is crucial for quantization is translational invariance in the integration variable. For example, when quantizing a scalar field ϕ on a lattice the field averaging involves the integral $\int_{-\infty}^{\infty} d\phi_n$ which has the translational invariance

$$\int_{-\infty}^{\infty} d\phi_n f(\phi_n + J_n) = \int_{-\infty}^{\infty} f(\phi_n) d\phi_n \qquad (3.30)$$

for any integrable function f and any constant J_n. It is this translational invariance that makes the Feynman path integral provide a realization of Schwinger's action principle (see, e.g., Ref. 12 for further discussion). Analogously one needs to define an integration over Fermi fields with the same translational invariance. Stated abstractly, one wants to define a bracket operation $\langle \cdots \rangle$ defined on functions of purely anticommuting Fermi fields ψ_n with the property

$$\langle f(\psi_n + \eta_n, \bar{\psi}_n + \bar{\eta}_n) \rangle = \langle f(\psi_n, \bar{\psi}_n) \rangle \, , \qquad (3.31)$$

where η_n and $\bar{\eta}_n$ are anticommuting c-numbers (these have been introduced by Schwinger). The bracket operation should produce a number for every function f; it should also be a linear operation. Thus for a finite lattice it is sufficient to specify the bracket $\langle \cdots \rangle$ for all monomials in the ψ_n and $\bar{\psi}_n$. Because of the anticommutation rules, ψ_n^2 and $\bar{\psi}_n^2$ vanish (more correctly, $\psi_{n\alpha}^2$ and $\bar{\psi}_{n\alpha}^2$ vanish where $\psi_{n\alpha}$ and $\bar{\psi}_{n\alpha}$ are any component of ψ_n and $\bar{\psi}_n$), therefore, there are only a finite number of possible monomials. It is now easy to see that

the bracket $\langle \cdots \rangle$ must vanish for all products of ψ's and $\bar{\psi}$'s, except the product containing all possible ψ's and $\bar{\psi}$'s. For example, suppose there are two lattice sites 0 and 1, and ψ_n and $\bar{\psi}_n$ are single component fields. Then the brackets must be

$$\langle 1 \rangle = 0 \, ,$$

$$\langle \psi_0 \rangle = \langle \psi_1 \rangle = \langle \bar{\psi}_0 \rangle = \langle \bar{\psi}_1 \rangle = 0 \, ,$$

$$\langle \psi_0 \psi_1 \rangle = \langle \psi_0 \bar{\psi}_1 \rangle = \cdots = 0 \, , \qquad (3.32)$$

$$\langle \bar{\psi}_0 \psi_0 \psi_1 \rangle = \cdots = 0 \, ,$$

$$\langle \bar{\psi}_0 \psi_0 \bar{\psi}_1 \psi_1 \rangle = 1 \, ,$$

where 1 is a constant which was chosen arbitrarily. This definition of the bracket operation satisfies translational invariance: for example,

$$\langle (\bar{\psi}_0 + \bar{\eta}_0)(\psi_0 + \eta_0)(\bar{\psi}_1 + \bar{\eta}_1)(\psi_1 + \eta_1) \rangle = \langle \bar{\psi}_0 \psi_0 \bar{\psi}_1 \psi_1 \rangle = 1$$

because the terms multiplying the η's are all 0. Note also that the anticommutation rules mean that, for example,

$$\langle \psi_0 \bar{\psi}_0 \bar{\psi}_1 \psi_1 \rangle = -1 \, . \qquad (3.33)$$

(In analogy to the scalar case, one requires $\psi_0 \bar{\psi}_0 = -\bar{\psi}_0 \psi_0$, not $\psi_0 \bar{\psi}_0 = 1 - \bar{\psi}_0 \psi_0$.)

One can now define the Feynman path integral on a lattice for the complete gauge theory including the Dirac fields. For example, the current-current propagator on the lattice is

$$D_{n\mu\nu} = Z_{\text{tot}}^{-1} \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right) \langle \bar{\psi}_n \gamma_\mu \psi_n \bar{\psi}_0 \gamma_\nu \psi_0 e^A \rangle \, , \qquad (3.34)$$

where A is the full action of Eq. (3.12) and

$$Z_{\text{tot}} = \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right) \langle e^A \rangle \, . \qquad (3.35)$$

This formulation of the path integral is different from the formulation discussed in Sec. II. However, one can easily derive a lattice form of the path integrals of Sec. II from the present expression. The procedure is to expand Eq. (3.34) in powers of K, where K is the coefficient of the nearest-neighbor coupling terms $\bar{\psi}_n \gamma_\mu \psi_{n+\hat{\mu}} e^{iB_{n\mu}}$ etc. This nearest-neighbor coupling term can be represented diagrammatically by a line from the site n to the site $n + \hat{\mu}$. The expansion is best described by studying an example of a term from the expansion of the numerator of Eq. (3.34), which will now be discussed. An example of a term in the expansion is represented diagrammatically in Fig. 9. The expression for this term is

$$\left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu}\right) K^4 \langle \bar\psi_{11}\gamma_\mu\psi_{11}\bar\psi_{00}\gamma_\nu\psi_{00}\bar\psi_{00}\gamma_0\psi_{10}e^{iB_{00,0}}\bar\psi_{10}\gamma_1\psi_{11}e^{iB_{10,1}}\,\bar\psi_{11}\gamma_0\psi_{01}e^{-iB_{01,0}}\bar\psi_{01}\gamma_1\psi_{00}e^{-iB_{00,1}}e^{A_0}\rangle = D$$

$$(3.36)$$

(D for diagram), where the four lattice sites involved are $(n_0, n_1) = (0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$; the values of n_2 and n_3 are constant and have been suppressed in the notation of Eq. (3.36). The action A_0 omits the K term, and is

$$A_0 = -c \sum_n \bar\psi_n \psi_n + \frac{1}{2g^2} \sum_n \sum_{\mu\nu} e^{if_{n\mu\nu}} \ . \tag{3.37}$$

The calculation of D has two parts, one being the integration over all gauge fields $B_{m\mu}$, the other being the calculation of the $\psi, \bar\psi$ bracket. These are independent calculations, i.e., D factors into $D_\psi D_B$. The quantity D_B is

$$D_B = \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu}\right) \exp\left[i(B_{00,0} + B_{10,1} - B_{01,0} - B_{00,1}) + \frac{1}{2g^2} \sum_{n\mu\nu} e^{if_{n\mu\nu}}\right] \ . \tag{3.38}$$

This is an example of a guage-field average of the exponential of a line integral over a closed loop, the loop being the loop of Fig. 9. The ψ bracket calculation can be factorized further into separate bracket calculations for each lattice site, since A_0 contains no terms involving ψ or $\bar\psi$ and coupling different lattice sites. Consider only the four lattice sites on the loop, for simplicity. By moving the ψ's around some (using the anticommuting rule) the bracket becomes

$$\langle \bar\psi_{00}\gamma_0\psi_{10}e^{-c\bar\psi_{10}\psi_{10}}\bar\psi_{10}\gamma_1\psi_{11}e^{-c\bar\psi_{11}\psi_{11}}\bar\psi_{11}\gamma_\mu\psi_{11}\bar\psi_{11}\gamma_0\psi_{01}e^{-c\bar\psi_{01}\psi_{01}}\bar\psi_{01}\gamma_1\psi_{00}e^{-c\bar\psi_{00}\psi_{00}}\bar\psi_{00}\gamma_\nu\psi_{00}\rangle \ .$$

To make a product of all possible ψ's and $\bar\psi$'s means one must have products of all possible ψ_{00}'s and $\bar\psi_{00}$'s, all possible ψ_{01}'s and $\bar\psi_{01}$'s, etc. In summary the complete bracket may be written as a product of four separate brackets. Define

$$D_\psi^1 = \langle \psi_{10}e^{-c\bar\psi_{10}\psi_{10}}\bar\psi_{10}\rangle \ , \tag{3.39}$$

$$D_{\psi\mu} = \langle \psi_{11}e^{-c\bar\psi_{11}\psi_{11}}\bar\psi_{11}\gamma_\mu\psi_{11}\bar\psi_{11}\rangle \ . \tag{3.40}$$

Both D_ψ^1 and $D_{\psi\mu}$ are matrices in spin space due to the spinor indices implied for ψ_{10}, $\bar\psi_{10}$, ψ_{11}, and $\bar\psi_{11}$. The full bracket is simply

$$D_\psi = -K^4 \operatorname{Tr}(D_\psi^1 \gamma_1 D_{\psi\mu} \gamma_0 D_\psi^1 \gamma_1 D_\psi{}_\nu \gamma_0) \ . \tag{3.41}$$

The matrices D_ψ^1 and $D_{\psi\mu}$ are easily determined. For example, D_ψ^1 explicitly is

$$D_{\psi\alpha\beta}^1 = \left\langle \psi_{10\alpha} \exp\left(-c \sum_\gamma \bar\psi_{10\gamma}\psi_{10\gamma}\right) \bar\psi_{10\beta}\right\rangle \ . \tag{3.42}$$

The exponential can be expanded in powers of c; assuming the spinors have four components only the c^3 term can produce a product of all four ψ_{10}'s times all four $\bar\psi_{10}$'s; the result is

$$D_{\psi\alpha\beta}^1 = -\delta_{\alpha\beta}(-c)^3 \tag{3.43}$$

(the minus sign comes from the convention that the bracket is positive when $\bar\psi_{10\beta}$ appears to the left of $\psi_{10\beta}$). A similar calculation gives

$$D_{\psi\mu} = c^2 \gamma_\mu \ . \tag{3.44}$$

The results of this example are easily generalized. A term of general order K^l is nonzero only

if the nearest-neighbor couplings combine to form closed loops (the lattice site at the endpoint of an open line would have an extra ψ_n or $\bar\psi_n$ so the bracket at n would give 0). The bracket calculation for a closed loop gives a trace involving K times a γ matrix for each line in the loop and D's for each lattice site in the loop (except the points n and 0 where there are currents). The average over gauge fields involves an exponential of a sum of $B_{n\mu}$'s around each loop. There can be any number of loops.

IV. STRONG-COUPLING APPROXIMATION

The gauge field average $I(P)$ which determines whether quarks are bound was defined on a lattice in Sec. III (Eq. 3.13). There are two limits in which this average can be calculated. The most interesting limit is the strong-coupling limit $g \to \infty$. This is the limit which exhibits quark binding. A strong-coupling expansion will be derived in this section.
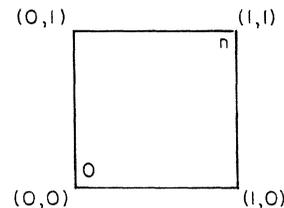

FIG. 9. Elementary square on the lattice.

The strong-coupling expansion will be the basis for a reformulation of the gauge-field theory as a string model. This will also be explained in this section.

Consider specifically the numerator of Eq. (3.13), to be denoted $I_N(P)$:

$$I_N(P) = \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right)$$
$$\times \exp\left[i \sum_P (\pm)B_{n\mu} + \frac{1}{2g^2} \sum_{n\mu\nu} e^{if_{n\mu\nu}} \right].$$

(4.1)

Expanding in powers of $1/g^2$, the zeroth-order term is

$$I_N^{(0)}(P) = \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right) \exp\left[i \sum_P (\pm)B_{n\mu} \right].$$

(4.2)

This term vanishes, since for any $B_{n\mu}$ which appears in $\sum_P$, there is an integral $\int_{-\pi}^{\pi} dB_{n\mu}$ $\times \exp(\pm iB_{n\mu})$ which is zero. Thus one must seek higher-order terms in g^{-2} which cancel the $B_{n\mu}$ in the line integral. The first-order term is

$$I_N^{(1)}(P) = \frac{1}{2g^2} \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right) \sum_{l\pi\sigma} \exp\left[i \sum_P (\pm)B_{n\mu} + if_{l\pi\sigma} \right].$$

(4.3)

The quantity $f_{l\pi\sigma}$ is itself a line integral of the gauge field; it is the line integral around a square originating at the lattice site l of size a (unit square). The integral for $I_N^{(1)}(P)$ will vanish unless it is possible to find a unit square such that $f_{l\pi\sigma}$ cancels completely the line integral $\sum_P (\pm)B_{n\mu}$. This is possible only if the path P is itself a unit square. Otherwise the first-order term vanishes and one must study the terms of order g^{-4} or higher.

The term of order g^{-2k} has the form

$$I_N^{(k)}(P) = \frac{1}{k!} \left(\frac{1}{2g^2} \right)^k \left(\prod_m \prod_\nu \int_{-\pi}^{\pi} dB_{m\nu} \right) \sum_{l_1\pi_1\sigma_1} \cdots \sum_{l_k\pi_k\sigma_k} \exp\left[i \sum_P (\pm)B_{n\mu} + if_{l_1\pi_1\sigma_1} + \cdots + if_{l_k\pi_k\sigma_k} \right].$$

(4.4)

The only nonzero terms in this sum are those for which

$$\sum_P (\pm)B_{n\mu} + f_{l_1\pi_1\sigma_1} + \cdots + f_{l_k\pi_k\sigma_k} = 0.$$

(4.5)

[See Eq. (3.8) for the definition of $f_{l\pi\sigma}$ in terms of $B_{n\mu}$.] This equation can be understood geometrically. Each $f_{l\pi\sigma}$ corresponds to a square of size a on the lattice. For this sum to vanish the set of squares defined by $f_{l_1\pi_1\sigma_1} \cdots f_{l_k\pi_k\sigma_k}$ must combine to make a surface with boundary P. (To be precise, each $f_{l\pi\sigma}$ corresponds to a line integral around a square, and when these squares are joined to make a surface the line integrals must cancel along all internal lines of the surface. The line integrals along the boundary P of the surface must run in the opposite direction to the original path P.) See Fig. 10.

For a given path P the lowest nonzero order in $I_N(P)$ is determined by the minimal area A enclosed by P, the area A being the area of any surface built of unit squares on the lattice with boundary P. Then $I_N(P) \sim (g^2)^{-A/a^2}$, apart from a numerical factor.

This is the result promised in Sec. II: The gauge-field average $I_N(P)$ behaves as $\exp[-A(\ln g^2)/a^2]$, i.e., exponentially in the area enclosed by P. Hence, according to the arguments of Sec. II, quark paths will not separate

macroscopically, and there will be no quarks among the final-state particles.

Consider higher-order terms in the expansion of $I_N(P)$ for given P. There are many such terms because there are many surfaces with boundary P. In particular, there are many ways to combine subsets of f's to add to zero so such subsets can be added to any minimal sum of f's which forms a surface with boundary P. The simplest example of a set of f's which add to zero are the set of f's corresponding to the six faces of a unit cube. Written out, this gives

$$0 = f_{n\mu\nu} - f_{n+\hat{\pi},\mu\nu} + f_{n\nu\pi} - f_{n+\hat{\mu},\nu\pi} + f_{n\pi\mu} - f_{n+\hat{\nu},\pi\mu},$$

(4.6)

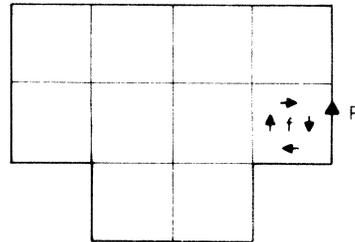which is easily checked using Eq. (3.8). [Equation

FIG. 10. Filling of enclosed area of path P by elementary squares.

(4.6) is the lattice analog of the equation

$$\epsilon^{\mu\nu\pi\sigma}\nabla_\pi F_{\mu\nu}(x) = 0.]$$

Let $A(P)$ be the minimal area as defined above enclosed by P. Since one can place a unit cube anywhere on this minimal area, it means that there are roughly $A(P)/a^2$ more terms of order $g^{-12}(g^2)^{-A(P)/a^2}$ in the expansion of $I_N(P)$ than there are terms of order $(g^2)^{-A(P)/a^2}$. [One can place unit cubes anywhere in space, not just on the minimal surface; but when one divides $I_N(P)$ by Z all disconnected terms cancel, as usual.] This suggests that the $1/g^2$ expansion is not very useful in the limit $A(P) \to \infty$, which is the limit of interest for quark binding. However, experience with related problems suggests that $I_N(P)$ is not the appropriate quantity to expand; instead one should try writing

$$I(P) = Z^{-1}I_N(P)$$

$$= (g^2)^{-A(P)}e^{-c(P,g^2)} , \qquad (4.7)$$

and expand $c(P, g^2)$ in powers of g^{-2} instead. One would expect $c(P, g^2)$ to be dominated by a term proportional to $A(P)$, say

$$c(P, g^2) = A(P)f(g^{-2}) + O(P) \qquad (4.8)$$

(where P is the length of the path P). The crucial question is the nature of the series for $f(g^{-2})$. Past experience with similar types of expansions (namely, the high-temperature expansions of statistical mechanics: see, e.g., Ref. 13) suggests that $f(g^{-2})$ will have a convergent expansion at least for g^{-2} less than a critical value g_c^{-2}. However, no calculations have been done in the gauge-field theory for $f(g^{-2})$ as yet.

Consider the complete expansion of $I(P)$. Each nonzero term in the expansion corresponds to a surface with perimeter P. The complete expansion corresponds to a sum over all possible surfaces with given perimeter P. "All possible" surfaces include surfaces which intersect themselves (to take into account terms where a given $f_{n\mu\nu}$ appears several times in the sum $f_{l_1\pi_1\sigma_1} + \cdots + f_{l_k\pi_k\sigma_k}$). There is a weight factor for each surface, aside from the power of g^{-2} determined by the area of the surface. For a simple surface, the weight factor is 1; the weight is more complicated for self-intersecting surfaces.

Thus, the strong-coupling expansion for the current-current propagator has the same general structure as in string models of hadrons. One is actually dealing here with a double expansion. An expansion in the coefficient K (appearing in the Dirac field action) was needed to define quark loops on the lattice; the sum of the K expansion is a sum over all possible quark loops. The g^{-2} expansion is needed to define surfaces filling in the quark loops. The sum of the g^{-2} expansion is a sum over all such surfaces. This is precisely the structure appearing in string models: combined sums over quark loops and interpolating surfaces. However, the loops and surfaces of the gauge-field theory are defined on a lattice whereas the loops and surfaces of the string models are defined on a continuum. It may not be easy to derive quantitative relations between the two types of surfaces.

V. WEAK-COUPLING APPROXIMATION

The weak-coupling approximation will be discussed briefly, leaving many questions open. Only the pure gauge field will be discussed. Consider again the expression

$$I_N(P) = \left(\prod_m \prod_\nu \int_{-\pi}^\pi dB_{m\nu}\right)$$
$$\times \exp\left[i\sum_P (\pm)B_{n\mu} + \frac{1}{2g^2}\sum_{n\mu\nu} e^{if_{n\mu\nu}}\right]. \qquad (5.1)$$

Suppose the integration variables were $f_{n\mu\nu}$ rather than $B_{m\nu}$. For small g, only small values of $f_{n\mu\nu}$ would be important in the integral, in order that $\mathrm{Re}\,e^{if_{n\mu\nu}}$ be near its maximum value 1. One would then expand:

$$\frac{1}{2g^2}\sum_{n\mu\nu} e^{if_{n\mu\nu}} \simeq \frac{1}{2g^2}\sum_{n\mu\nu}(1 - \tfrac{1}{2}f_{n\mu\nu}^2) . \qquad (5.2)$$

With this approximation one could extend the limits of integration on $f_{n\mu\nu}$ from $\pm\pi$ to $\pm\infty$, with negligible error; one would then have a set of Gaussian integrals to evaluate.

In practice the integration variables are the $B_{m\nu}$, not the $f_{n\mu\nu}$. However, one can make a change of variable from the $B_{m\nu}$ to the $f_{n\mu\nu}$. It is not possible to eliminate all the $B_{m\nu}$ by this transformation, and not all the variables $f_{n\mu\nu}$ are independent. Nevertheless, the transformation is sufficient to make $I_N(P)$ calculable for small g^2.

To make the change of variables precise, consider a system of finite size $(1 \le n_i \le N)$ with periodic boundary conditions. Then one can change variables from the $B_{m\nu}$ to a subset of the $f_{n\mu\nu}$ plus some gauge transformation variables ϕ_n, plus four extra variables ζ_μ, as follows:

(i) For $n_0 \ne N$, n_1, n_2, n_3 arbitrary, $B_{n\mu}$ ($\mu = 1, 2, 3$) is replaced by $f_{n\mu 0}$. For B_{n0}, one writes

$$B_{n0} = \phi_{n+\hat{0}} - \phi_n \qquad (5.3)$$

and replaces B_{n0} by ϕ_n. This is the essence of the transformation from $B_{n\mu}$ to $f_{n\mu 0}$ for $\mu \ne 0$ and from B_{n0} to ϕ_n. To complete the transformation one must discuss the surface $n_0 = N$.

(ii) For $n_0 = N$, $n_1 \neq N$, and n_2 and n_3 arbitrary, $B_{n\mu}$ ($\mu \neq 1$) is replaced by $f_{n\mu 1}$. B_{n1} is replaced by ϕ_n, with

$$B_{n1} = \phi_{n+\hat{1}} - \phi_n . \tag{5.4}$$

(iii) For $n_0 = n_1 = N$, $n_2 \neq N$, and n_3 arbitrary, $B_{n\mu}$ ($\mu \neq 2$) is replaced by $f_{n\mu 2}$. B_{n2} is replaced by ϕ_n, with $B_{n2} = \phi_{n+\hat{2}} - \phi_n$.

(iv) For $n_0 = n_1 = n_2 = N$, and $n_3 \neq N$, $B_{n\mu}$ ($\mu \neq 3$) is replaced by $f_{n\mu 3}$, B_{n3} by ϕ_n, where $B_{n3} = \phi_{n+\hat{3}} - \phi_n$.

(v) For $n_0 = n_1 = n_2 = n_3 = N$ one writes $B_{n\mu} = \zeta_\mu$ with ζ_μ being the new variables. One also sets $\phi_n = 0$ for this value of n.

The variables $f_{n\mu\nu}$ which are not integration variables (for example, $f_{n\mu\nu}$ with $\mu \neq 0$ and $\nu \neq 0$, for $n_0 \neq N$) can be expressed in terms of the independent $f_{n\mu\nu}$ variables using Eq. (4.6); neither the ϕ_n nor ζ_μ appear in these expressions.

When an arbitrary $B_{m\nu}$ is expressed in terms of the new independent variables, one finds (ζ_ν is present only if $m_\nu = N$):

$$B_{m\nu} = \zeta_\nu + \phi_{m+\hat{\nu}} - \phi_m$$

$$+ (\text{linear combination of } f_{n\pi\sigma}) , \tag{5.5}$$

i.e., the ϕ variables define a gauge transformation and ζ_ν represents a translation of some of the B's. It is easily verified that the integrand of $I_N(P)$ involves only the f's: It is independent of both the ϕ's and ζ_ν (the latter does not appear because any closed path P has as many $-B_{n\nu}$ terms as $+B_{n\nu}$ terms on the sublattice $n_\nu = N$). Hence the ϕ_n and ζ_ν integrations can be computed trivially. The f integrations are nontrivial because of the constraint (4.6).

What one wants to accomplish is to reduce the lattice theory for small g to something like a conventional free gauge-field theory. This means restoring the $B_{m\nu}$ as the integration variables, but with infinite limits of integration, and with a gauge-fixing term included. Suppose, for example, one starts with

$$I_N'(P) = \left(\prod_m \prod_\nu \int_{-\infty}^{\infty} dB_{m\nu} \right) \exp\left(i \sum_P (\pm) B_{n\mu} - \frac{\alpha}{2} \sum_n \left[\sum_\mu (B_{n\mu} - B_{n-\hat{\mu},\mu}) \right]^2 - \frac{1}{4g^2} \sum_{n\mu\nu} f_{n\mu\nu}^2 \right) , \tag{5.6}$$

where the α term is a lattice version of a $(\nabla_\mu A_\mu)^2$ gauge-fixing term. This integral can be computed by explicit Gaussian integration methods rather more easily than the $f_{n\mu\nu}$ integrations for $I_N(P)$. In addition, $I_N(P)$ can also be reduced to an integral over a subset of the $f_{n\mu\nu}$, using the same change of variables as for $I_N(P)$. The result is different in this case due to the α term which couples the ϕ's to the f's; also there is no convergence factor for the ζ integration. To make $I_N'(P)$ well defined and equal to $I_N(P)$, one must (a) put in a convergence factor for the ζ integral, i.e., a term $-\frac{1}{2}\beta(\sum_n B_{n\mu})^2$ and (b) add a quadratic form in the f's to compensate for the result of the ϕ integration of the gauge-fixing term. The author has not carried through this calculation; but since the net result is still that $I_N(P) = I_N'(P)$ is a Gaussian integration in the B's, the result will presumably be similar to the conventional free-field calculation reported in Sec. II.

VI. PHASE TRANSITIONS

In the strong-coupling limit ($g \to \infty$, $K \to 0$) the gauge theory is far from being Lorentz-invariant. More precisely, since the action was defined on a Euclidean metric, it is Euclidean invariance that is missing. In the strong-coupling limit, vacuum expectation values decrease rapidly at separations of only a few lattice sites (there is a factor g^{-2} or

K or both for each unit lattice spacing of separation). This corresponds to the existence of masses much larger than the cutoff. [The usual rule is that if a propagator falls as $e^{-x/\xi}$ for x large then the lowest mass intermediate state contributing to the propagator has mass $1/\xi$. If the propagator behaves as g^{-2n} for distances $x = na$, then the corresponding mass is $2(\ln g)/a$. This is larger than the cutoff momentum π/a if g is large.]

Thus, one is interested in practice in values of g and K such that the correlation length ξ is much larger than the lattice spacing a, in order that the corresponding mass is much less than the cutoff. One knows from statistical mechanics that large correlation lengths are associated with second-order phase transitions (critical points). Thus one seeks special values g_c and K_c for g and K at which there is a phase transition.[14]

It has already been argued that there are two distinct phases for the gauge field, a strong-coupling phase for large g which binds quarks, and a weak-coupling phase for small g which does not bind quarks. The arguments given neglected quark vacuum loops, which is reasonable if K is small. There should be a transition between these two phases which would occur at a critical value g_c for any g and any small value of K. This is one possible phase transition; it is this transition which was discussed in Sec. II. But, as will be argued below, this is probably a first-order tran-

sition rather than second order.

Suppose one wishes to construct a model of strong interactions using the lattice theory of this paper with the gauge group separate from ordinary $SU(3) \times SU(3)$ symmetry. Then the gauge fields would all be $SU(3) \times SU(3)$ singlets, while the quark fields would carry $SU(3) \times SU(3)$ quantum numbers as well as gauge-group quantum numbers. A little thought shows that in the strong-coupling limit ($g \to \infty$, $K \to 0$), $SU(3) \times SU(3)$ is an exact symmetry rather than a spontaneously broken symmetry. Varying g does not change this situation; so one must hope that by increasing K one can change the exact $SU(3) \times SU(3)$ into broken $SU(3) \times SU(3)$. If this does not work one is free to introduce additional terms into the quark field action in hopes of forcing a spontaneous breaking of $SU(3) \times SU(3)$. Suppose, for simplicity, that $SU(3) \times SU(3)$ can be broken by increasing K. Then there will be a phase transition at a critical value K_c for K where one changes from exact $SU(3) \times SU(3)$ to spontaneously broken $SU(3) \times SU(3)$. If this transition is a second-order transition then there will be a large correlation length for K near K_c; in this case the theory might be a realistic model of broken $SU(3) \times SU(3)$ for K slightly greater than K_c (with g large enough to maintain quark binding).

In summary, the transition of real interest is a transition in K (or some other parameter introduced into the quark action) rather than g.

Apart from special limits ($g \to \infty$ and $K \to 0$, or g small) it is very difficult to solve the lattice theory. It is especially difficult to solve the lattice theory near a critical point with a large correlation length. Various methods have been developed by statistical mechanicians to deal with this problem. In the remainder of this section these methods will be discussed briefly. There are essentially three approaches to consider: (1) mean-field techniques, (2) series expansions, and (3) the renormalization-group approach.

Mean-field techniques[15] are the simplest and crudest methods for studying a critical point; invariably they are the methods one uses first in studying a new situation. They are used to determine if there is a phase transition, whether it is first or second order, and to give rough estimates of the behavior near the transition. None of the results of a mean-field calculation are entirely trustworthy. Examples of mean-field calculations will be given later.

An example of a series expansion would be the expansion of the current-current propagator for small momentum (momentum $\ll 1/a$) in powers of g^{-2} and K, to high order in g^{-2} and K. One then uses Padé-approximant techniques to look for singularities in either g or K that would be as-

sociated with a mass approaching 0. In simple statistical-mechanical problems one can generate 12 terms or so in analogous expansions. The expansion for the lattice theory of this paper is more complicated, but one could hope to generate maybe 6 or 7 orders with some practice. Series expansions require considerably more effort than mean-field calculations; they apply mainly to propagators, being very awkward to perform on three- and four-point functions, and one must have a clear idea of what one is trying to learn before attempting such calculations. See Ref. 16 for one of the best series-expansion formalisms; see Ref. 13 for a general review.

The renormalization-group approach is potentially the most powerful and accurate method for studying lattice theories near a critical point, but at present the renormalization-group techniques are too limited in scope to be applicable to the present problem. See Refs. 6 and 17.

Return to mean-field ideas.[15] The prototype mean-field calculation is a calculation of the magnetization as a function of the external field for an Ising ferromagnet. Let s_n be the spin at site n with values ± 1 only; let the interaction be

$$\frac{-H}{kT} = K \sum_n \sum_{\hat{\mu}} s_n s_{n+\hat{\mu}} + h \sum_n s_n \ , \qquad (6.1)$$

where K is related to the spin-spin coupling and h is proportional to the external field. Then

$$M = Z^{-1} \left\langle s_0 \exp\left(K \sum_n \sum_{\hat{\mu}} s_n s_{n+\hat{\mu}} + h \sum_n s_n \right) \right\rangle \ , \qquad (6.2)$$

where $\langle \cdots \rangle$ means a sum over all configurations of all spins, and

$$Z = \left\langle \exp\left(K \sum_n \sum_{\hat{\mu}} s_n s_{n+\hat{\mu}} + h \sum_n s_n \right) \right\rangle \ . \qquad (6.3)$$

In the mean-field approximation, one assumes that the spins $s_{\hat{\mu}}$ coupled to s_0 can be replaced by their average value M. As a result, the formula for M simplifies to a sum over s_0 only, namely

$$M = Z_0^{-1} \sum_{s_0 = \pm 1} s_0 e^{(2dKM + h) s_0} \ , \qquad (6.4)$$

with d being the dimensionality (3 usually) and

$$Z_0 = \sum_{s_0 = \pm 1} e^{(2dKM + h) s_0} \ . \qquad (6.5)$$

The result is

$$M = \tanh(2dKM + h) \ . \qquad (6.6)$$

If $2dK < 1$ this equation has a unique solution for M as a function of h; in particular, $M = 0$ for $h = 0$. For $2dK > 1$ the solution is multiple-valued; stability considerations show that one must choose a solution with $M \neq 0$ when $h = 0$.

In this approximation one has actually replaced $\sum_\mu s_{\hat{\mu}}$ by $2dM$, which is a good approximation if d is large. This is generally true of mean-field theories.

An analogous mean-field calculation can be performed for the lattice gauge theory. In this case a simple external field term has the form

$$h\sum_{n\mu} (e^{iB_{n\mu}} + e^{-iB_{n\mu}})$$

(this is to be added to the gauge-field action), and M can be defined to be the expectation value of $e^{iB_{0\mu}}$. The question is whether M is zero in the limit $h \to 0$. In the gauge-field theory $e^{iB_{0\mu}}$ couples to a product of three other exponentials; as a mean-field approximation one replaces this product by M^3. The result of this is that

$$M = Z_0^{-1} \int_{-\pi}^{\pi} dB_{0\mu} e^{iB_{0\mu}}$$

$$\times \exp\left[\left(\frac{2(d-1)}{2g^2} M^3 + h\right)(e^{iB_{0\mu}} + e^{-iB_{0\mu}})\right] ,$$

$$(6.7)$$

where d is the space-time dimensionality, and

$$Z_0 = \int_{-\pi}^{\pi} \exp\left[\left(\frac{2(d-1)}{g^2} M^3 + h\right)(e^{iB_{0\mu}} + e^{-iB_{0\mu}})\right] .$$

$$(6.8)$$

The result of this is that

$$M = f\left(\frac{(d-1)}{g^2} M^3 + h\right) ,$$

$$(6.9)$$

where f is a ratio of Bessel's functions. If g is large the solution to this equation is unique and

$M = 0$ for $h = 0$. If g is small then there are solutions with $M \neq 0$ for $h = 0$, and stability considerations show again that the $M \neq 0$ solutions are preferred.

In the magnetic case, one finds that the spontaneous magnetization M goes to zero for $2dk \to 1$ [from Eq. (6.6)]. However, the gauge-field case never has a solution for $h = 0$ with M small but nonzero. Thus there is a first-order transition at the value of g for which M changes from zero to being nonzero.

A nonzero value of M in the limit $h \to 0$ means one has spontaneous breaking of the gauge-field symmetry. So for small g the theory shows spontaneous breaking.

A much more thorough discussion of the mean-field approximation has been given by Balian, Drouffe, and Itzykson.[18] A Hamiltonian formulation of the lattice gauge theory has been given by Kogut and Susskind.[19] A clear review of quark confinement in the lattice theory is given in Ref. 20. Another formulation of the connection between strongly coupled gauge theories and string models is given in Ref. 21.

ACKNOWLEDGMENTS

---

*Work supported in part by the National Science Foundation.

[1]See, e.g., J. M. Cornwall and R. E. Norton, Phys. Rev. D 8, 3338 (1973); R. Jackiw and K. Johnson, ibid. 8, 2386 (1973); J. Kogut and L. Susskind, ibid. 9, 3501 (1974); D. Amati and M. Testa, Phys. Lett. 48B, 227 (1974); G. 't Hooft, Nucl. Phys. B (to be published); P. Olesen, Phys. Lett. 50B, 255 (1974); A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, Phys. Rev. D 9, 3471 (1974).

[2]J. Schwinger, Phys. Rev. 125, 397 (1962); 128, 2425 (1962).

[3]J. H. Lowenstein and J. A. Swieca, Ann. Phys. (N.Y.) 68, 172 (1971).

[4]A. Casher, J. Kogut, and L. Susskind, Phys. Rev. Lett. 31, 792 (1973); Phys. Rev. D 10, 732 (1974).

[5]A. suri, Ph.D. thesis, Cornell University, 1969 (unpublished).

[6]K. G. Wilson and J. Kogut, Phys. Rep. 12C, 75 (1974). Sec. X.

[7]S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973).

[8]Y. Nambu, in Symmetries and Quark Models, proceedings of the International Conference on Symmetries and Quark Models, Wayne State Univ., 1969, edited by Ramesh Chand (Gordon and Breach, New York, 1970); L. Susskind, Nuovo Cimento 69A, 457 (1970); G. Konisi, Prog. Theor. Phys. 48, 2008 (1972); P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, Nucl. Phys. B56, 109 (1972); J.-L. Gervais and B. Sakita, Phys. Rev. Lett. 30, 716 (1973); S. Mandelstam, Nucl. Phys. B64, 205 (1973).

[9]R. P. Feynman, Phys. Rev. 80, 440 (1950).

[10]See, e.g., E. S. Abers and B. W. Lee, Phys. Rep. 9C, 1 (1973).

[11]See, e.g., F. A. Berezin, The Method of Second Quantization (Academic, New York, 1966), pp. 52ff.

[12]K. G. Wilson, Phys. Rev. D 6, 419 (1972).

[13]M. E. Fisher, Rep. Prog. Phys. 30, 615 (1967).

[14]For the relation of relativistic field theory to critical

points see Ref. 6, Sec. X, and references cited there-
in.

[15]For a review see R. Brout, *Phase Transitions* (Benja-
min, New York, 1965), p. 8.

[16]D. Jasnow and M. Wortis, Phys. Rev. <u>176</u>, 739 (1968).

[17]K. G. Wilson, Cargèse (1973) Lecture Notes, in
preparation.

[18]R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev.

D (to be published).

[19]J. Kogut and L. Susskind, Phys. Rev. D (to be pub-
lished).

[20]K. Wilson, Cornell Report No. CLNS-271 (to be pub-
lished in the proceedings of the conference on Yang-
Mills Fields, Marseille, 1974).

[21]H. B. Nielsen and P. Olesen, Nucl. Phys. <u>B61</u>, 45
(1973); L. J. Tassie, Phys. Lett. <u>46B</u>, 397 (1973).

Elementary and composite particles in asymptotically free gauge theories

T. P. Cheng

Department of Physics, University of Missouri—St. Louis, St. Louis, Missouri 63121

Ng Wing-Chiu*†

Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11790

K. Young

Department of Physics, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong

(Received 3 June 1974)

We demonstrate that in asymptotically free gauge theories the wave-function renormalization constants
for spin-1/2 fields do not vanish. (The scalar fields, if incorporated, also have this property.) However,
there exists a subclass of such theories where the Z's for the gauge fields themselves tend to zero in
the limit of infinite cutoff. These features are shown to be gauge-independent. This suggests the
potentiality of constructing asymptotically free strong-interaction theories in which the only elementary
fields are "quarks" and all other hadrons are bound states.

I. INTRODUCTION

Gauge-invariant quantum field theories based on
non-Abelian groups are being actively investigated.
A number of rather unique properties have been
discovered about such theories. These features,
when interpreted optimistically, indicate that non-
Abelian gauge theories may provide the framework
within which theories of strong interactions (or
event unified theories of all elementary-particle
forces) may be constructed.

Gauge theories are renormalizable.[1] It has been
shown, by way of the renormalization-group equa-
tion,[2] that the origin of the coupling-constant space
is an ultraviolet-stable fixed point[3] only in non-
Abelian gauge theories.[4] This asymptotically free
nature of the theory provides us with a field-the-
oretical explanation of Bjorken scaling—rather
the explanation of *how* Bjorken scaling is ap-
proached in the deep Euclidean limit.[5] This same
property indicates that the effective couplings can
be large in the infrared limit—it just may provide
the desired quark-confinement mechanism.[6]

There are also a number of works suggesting
intriguing connections of gauge theories to dual
models and relativistic string models of hadrons.[7]

This confluence of field-theoretical and S-matrix
approaches to strong-interaction physics is also
indicated by the works of Grisaru, Schnitzer, and
Tsao.[8] These authors have demonstrated that
vector mesons and spin-$\frac{1}{2}$ fermions in such field
theories satisfy the usual criteria of Reggeization:
factorization of Born amplitudes and Mandelstam
countings. However, as possible candidates for
strong-interaction theories, the class of gauge
theories investigated in their works may have
some drawbacks: These gauge theories are not
asymptotically free and while spin-$\frac{1}{2}$ particles
(quarks?) lie on Regge trajectory, scalar fields
do not. (It would seem an unattractive picture of
having spin-$\frac{1}{2}$ fermions composite, but not all
other particles.) In this paper we shall use an-
other criterion for the compositeness of fields
appearing in a Lagrangian field theory, i.e., the
vanishing of the wave-function renormalization
constants. Our results suggest the possibility of
constructing strong-interaction field theories
which are asymptotically free and in which the
only elementary fields are quarks and all other
hadrons are composite.

In Sec. II we shall demonstrate, through a
straightforward exercise of solving the renormal-