Nuclear Physics B169 (1980) 103–125

CHIRAL SYMMETRY BREAKING IN CONFINING THEORIES*

T. BANKS and A. CASHER

Department of Physics and Astronomy, Tel Aviv University, Ramat Aviv, Israel

Received 20 November 1979
(Final version received 18 February 1980)

The behaviour of the chiral order parameter in gauge theories is analysed. It is shown that in the confining phase chiral symmetry is spontaneously broken. We employ a new lattice version of the Dirac equation which avoids spectral multiplication.

1. Introduction

In the past two years a qualitative picture of confinement in quantum chromodynamics (QCD) has emerged [1]. The relevant field configurations are the $Z(N)$ fluxons introduced by 't Hooft [2, 3]. The "knots" where N fluxons meet are monopoles and confinement is a "dual Meissner effect" as first suggested by 't Hooft and Mandelstam [4]. There is no doubt in our minds that this constitutes a qualitatively correct explanation for one of the central facts of hadron dynamics. A similar claim cannot be made about our understanding of chiral symmetry breaking (χSB) in QCD. No convincing explanation of this phenomenon has been presented.

Recently one of us [5] has argued that χSB is an inevitable concomitant of confinement. It is the purpose of the present paper to demonstrate this connection in the context of the fluxon mechanism. Detailed analysis of the way in which fluxons cause χSB leads us to an heuristic proof of the general conjecture that χSB occurs in any field theory if and only if there are fermion–antifermion bound states.

The qualitative argument of ref. [5] is extremely simple. Assume that QCD is a confining theory and consider the evolution of the bare $m = 0$ fermion vacuum under the QCD hamiltonian. Inevitably, massless fermion–antifermion pairs will be created. Consider an isolated pair. The spin-independent confining force acting between the pair will act to force them into a bound state. Semiclassically such an (s-wave) bound state can be described by a trajectory in which the fermions oscillate back and forth on a straight line. Such a trajectory is, however, inconsistent with chiral symmetry. At the point at which the fermion turns around it changes its momentum without changing its spin and thus reverses its helicity. But for a massless particle *satisfying the Dirac equation* helicity equals chirality.

* Work supported in part by the Israel Commission for Basic Research and the US–Israel Bi-National Science Foundation (BSF).

In ref. [5] it was argued that the dual requirements of lowering the vacuum energy by putting all vacuum pairs into bound states while at the same time preserving chiral symmetry could only be satisfied by constructing a fermion vacuum out of correlated pairs with density equal to the inverse cube of the size of the bound state. In the semiclassical picture, whenever a fermion reaches its turning point it is either annihilated by an antifermion or replaced by a fermion from a nearest neighbor pair. Both these mechanisms preserve chirality *locally* but lead to a non-zero value for $\langle\bar{\psi}\psi\rangle$ which can be calculated in terms of the Regge slope. Note that the argument did not really require confinement but only the existence of a fermion–antifermion bound state.

To prove the conjecture that the existence of a bound state implies χSB, ref. [5] employed the Bethe–Salpeter equation. The proof goes through for all static chirally invariant Bethe–Salpeter kernels. A generalization to non-static kernels can probably be found but such a mathematical exercise would not expose the essential mechanism involved. In addition, in QCD the Bethe–Salpeter kernel is not gauge invariant and this will further obscure the meaning of the theorem.

We turn therefore to an entirely different method of attack. In sect. 2 we review the relevant parts of the theory of Z(N) fluxons in QCD and show that their infrared effects can be summarized by a Z(N) lattice gauge theory in its confining phase. We further argue that for the study of χSB we can replace the Z(N) theory (for any N) by a U(1) theory, a result which is rigorously true in the large-N limit.

In sect. 3 we introduce a new formalism for coupling fermions to lattice gauge theories. The formalism is based on the second-order Dirac equation and preserves nearest-neighbor couplings and the correct fermion spectrum in the continuum. We find that we must introduce a new dimensionless parameter K which measures the ratio between the spin and orbital terms in the second-order Dirac equation. K must be correctly renormalized in order to reproduce the continuum Dirac equation. Thus, the intuitive argument of ref. [5], which uses the Dirac relation between helicity and chirality, will only be valid for a particular value of K, $K_c(g^2)$ (where g^2 is the gauge coupling).

In sect. 4 we investigate χSB in the strongly coupled lattice gauge theory. We use a Hartree–Fock-like (one fermion loop) approximation which is rigorously justified for large N and qualitatively correct for any N. In this approximation $\langle\bar{\psi}\psi\rangle$ can be written as a superposition of Wilson loops [in the continuum theory they are "spinning loops" $\operatorname{tr}(\exp\{i\oint A\cdot dx-\frac{1}{2}\int d\tau\sigma_{\mu\nu}F_{\mu\nu}\})_+$]. We then show that for sufficiently large coupling the gauge theory is equivalent to a strongly coupled generalized Nambu–Jona-Lasinio [6] (NJL) model and has a non-zero $\langle\bar{\psi}\psi\rangle$ if $K=K_c(g^2)$. We argue (but do not prove rigorously) that this is true as long as g^2 is in the confining region.

Sect. 5 comprises an heuristic investigation of χSB in a general context. We begin by pulling apart the strong coupling calculation to see what makes it work. Our arguments here are rigorously valid for any strongly coupled lattice gauge theory

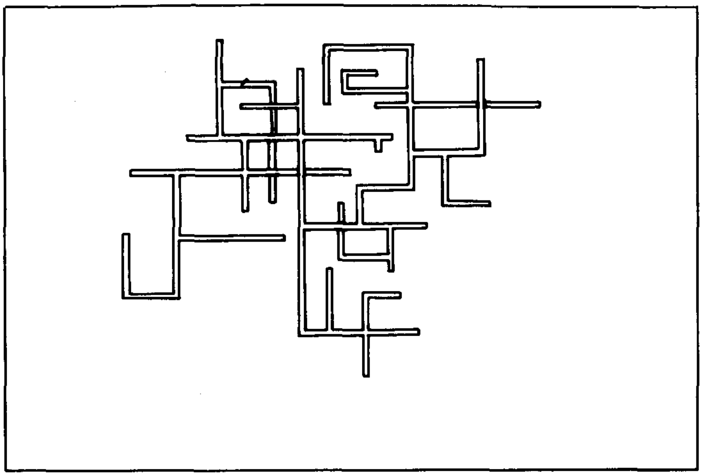

Fig. 1. A typical tree.

(abelian or non-abelian) and do not depend on the fluxon approximation. We show that a necessary and sufficient condition for χSB is that

$$\langle 0T|00\rangle_{T\to\infty}\, T^{-1/2}\,,$$

where $\langle 0T|00\rangle$ is the amplitude for a fermion to return to the origin in euclidean proper time T (averaged over gauge-field fluctuations). We show that this is achieved in the strong coupling limit by a combination of two mechanisms.

(i) K is correctly chosen equal to $K_c(g^2)$.

(ii) The path integral defining $\langle\bar{\psi}\psi\rangle$ is dominated by a set of tree-like paths (fig. 1). This expresses the fact that for strong coupling fermions and antifermions must stay close together in order to screen the strong constant electric field between them.

Domination by trees is too strong a condition however; $\langle 0T|00\rangle$ will have the correct $T^{-1/2}$ behavior as long as tree-like paths contribute a finite fraction to the path integral. We argue that this will be the case whenever the theory has a bound state, and thus, *a fortiori*, in any confining theory.

Sect. 6 is devoted to concluding remarks and to pointing out directions for further study. In particular we give a brief discussion of how f_π may be computed within our formalism.

In appendix A we study certain aspects of the second-order fermion formalism.

2. Z(*N*) fluxons*

Z(N) fluxons in an SU(N) gauge theory are defined relative to other configura-

* This section is based mostly on unpublished work of the present authors. Our ideas about fluxons were described at length in a series of seminars presented at the Université Libre de Bruxelles in September, 1979.

tions. It is easiest to visualize them in the Schrödinger picture at a fixed time. A fluxon is added to an eigenstate of the field A (in the $A^0 = 0$ gauge) by performing a singular gauge transformation $A \to A^{\Omega}$. Ω has a singularity on a closed contour C and for every closed curve Γ which links with C, the Wilson loops of A and A^{Ω} are related by

$$W_\Gamma(A^\Omega) = \left(\exp\frac{2\pi i n}{N}\right) W_\Gamma(A)\,, \tag{1}$$

where $n = 1, \dots N-1$. The Schrödinger picture fluxons may be generalized to define euclidean $Z(N)$ fluxons by following the (imaginary) time history of the closed curve $C(t)$. The singularities thus occupy two-dimensional sheets in space-time. The surface belonging to a given fluxon of strength n/N [eq. (1)] must be either closed or infinite in order to avoid a singularity in the Wilson loop. Note, however, that a set of fluxons which satisfy

$$\sum n = 0(\mathrm{mod}\, N)$$

may annihilate on a space–time curve (which again is either closed or infinite but may not terminate!).

Fluxons are stable only if they exist on a non-trivial background. In particular, if the classical vacuum ($A = 0$) is subjected to the transformation Ω, the resulting "pure" fluxon configuration is unstable [3, 8]. However, the *typical* configuration which contributes to the path integral of a non-abelian gauge theory is not $A = 0$ (which has a vanishing measure). Rather, it is an inherently non-abelian (namely $[A_\mu, A_\nu] \neq 0$) position-dependent fluctuation with a non-vanishing average action density extending all the way to infinity (an example is a finite-density instanton–anti-instanton gas and its accompanying short-distance quantum fluctuations). It will now be argued that on such a background fluxons will be stable.

For simplicity, consider a fluxon defined on a planar closed curve C. Choose a (singular) gauge in which $A^\Omega = A$ everywhere except on the planar surface bounded by the curve C, where a δ-function singularity is added to A in order to change the Wilson loop according to eq. (1). As long as the singularity is quantized in units of $Z(N)$, it will not cost energy since any infinitesimal Wilson loop which pierces the singular surface twice is unaffected by its presence. Thus for quantized fluxons $F^2_{\mu\nu}(A^\Omega) = F^2_{\mu\nu}(A)$ outside the curve C. If the background field was purely abelian over the whole surface, then the same would be true for any discontinuity. In particular we could continuously reduce the discontinuity from $\Omega \in Z(N)$ to $\Omega = 1$. Since this would leave the surface energy unchanged but reduce the energy along the curve C we would have found a direction of instability for the fluxon. However, if the background is non-abelain any jump on the surface which does not commute with the background will cost an energy proportional to the area of the surface. Since the direction in group space of the background varies in space and since we are summing over backgrounds with different directions, the only discontinuities which will not cost surface energy will be those of quantized $Z(N)$ fluxons. Thus a small variation

δa of the fluxon will lead to an energy change

$$\delta E \propto (\delta a)^2[KA - CL],$$

where A is the surface area and L the length of C. For large enough fluxons $KA > CL$, δE will be positive and the fluxon stable. Note that the presence of non-abelian fluctuations is important for the argument. QED fluxons need not be quantized in the absence of charged fields.

A second source of instability is the ultraviolet-divergent energy of the singular core C. The obvious remedy is to convert the line C into a tube of finite average radius λ. The configuration inside the tube should be determined by minimizing the action of the thick space–time surface C subject to the condition that the core merge continuously with A^{Ω} at the boundary. For a given background A there may be a degenerate set of such configurations (especially for large N) and summing over them will tend to increase the entropy of the fluxon. The radius λ should finally be determined by a variational argument (the interaction between the fluxon and the background is not scale invariant due to the scale dependence of the effective coupling). Alternatively, λ may be treated as a collective coordinate and its most probable value determined from the ensuing measure. This value will not be ∞ since the entropy of an infinitely thick fluxon is negligible.

To summarize: fluxons on a non-trivial background are stable sheet-like euclidean configurations which affect the Wilson loop even when the latter does not intersect the sheet. The world surfaces of the fluxons have finite thickness and action per unit area. They are either infinite or "closed modulo N". In fact, all their properties are periodic functions of their flux which may take on the discrete value $2\pi n/N$ $(1 \leq n \leq N-1)$.

In order to proceed, we now assume that a typical fluxon does not fluctuate too much on the scale λ. This is actually a rather strong assumption which may be justified only by a complete dynamical calculation. In fact, we should not exclude *a priori* the possibility that the dominant fluxons have an irregular fluctuating shape rather than a smooth curve such as a circle or a square of size λ. If this is the situation, then the boundaries of different fluxons would interfere with one another and a rather tangled mess would result. It is probable that the considerations which follow would still be qualitatively correct but somewhat unreliable. The smoothness assumption implies that a reasonable approximation for treating phenomena whose scale is larger than λ is to replace the continuum SU(N) theory by an effective lattice Z(N) gauge theory [9].

Each fluxon configuration can be put into 1–1 correspondence with a Z(N)-valued antisymmetric tensor $G_{\mu\nu}(x) = (2\pi/N)n_{\mu\nu}(x)$ where x is a lattice coordinate. $G_{\mu\nu}(x)$ is the flux which crosses the oriented plaquette bordered by the links $(x, x+\hat{\mu})$ and $(x, x+\hat{\nu})$. The "modulo-N" conservation of the flux leads to

$$\Delta_\mu G_{\mu\nu}(x) \equiv \sum_\mu [G_{\mu\nu}(x+\hat{\mu}) - G_{\mu\nu}(x)] = 0 (\mathrm{mod}\, N) . \qquad (2)$$

Eq. (2) is solved by

$$G_{\mu\nu}(x)=\tfrac{1}{2}\varepsilon_{\mu\nu\alpha\beta}(\Delta_\alpha\theta_\beta-\Delta_\beta\theta_\alpha)\equiv\tilde{\theta}_{\mu\nu}\,, \tag{3}$$

where $\theta_\mu(x)$ is a Z(N) variable which resides on the link $(x, x+\hat{\mu})$. The importance of θ_μ stems from the fact that matter fields interact with fluxons through it. Thus, the effective action of a classical spinless particle of N-ality p due to its interaction with fluxons is

$$\mathrm{e}^{-S[x(t)]}=\sum_{\{\theta_\mu\}}\mathrm{e}^{-F[\theta_{\mu\nu}]}\,\mathrm{e}^{ip\Sigma\theta_\mu J_\mu}\Big/\sum_{\{\theta_\mu\}}\mathrm{e}^{-F(\theta_{\mu\nu})}\,. \tag{4}$$

Here $J_\mu(x)$ is the convective current of the particle,

$$J_\mu(x)=\sum_t\delta(x-x(t))(x_\mu(t+1)-x_\mu(t))\,, \tag{5}$$

while $F(\theta_{\mu\nu})$ is the free energy (action minus entropy+short-distance quantum fluctuation) of the fluxon configuration specified by $\{\theta_\mu\}$ as calculated from the SU(N) continuum theory.

Eq. (4) is the expression for the Wilson loop in a Z(N) gauge theory [10]. This class of theories can be described in terms of an ordinary (non-compact) electromagnetic field interacting with magnetic monopoles of strength $2\pi e^{-1}$ and charges of magnitude Ne, where e is essentially the coupling constant of the action functional F [eq. (4)]. Note that in our case e would be determined by the continuum theory. For sufficiently large N ($N\geqslant5$) this system has three phases: electric confinement ($e^2>0(1)$), magnetic confinement ($e^2<(2\pi/N)^2$) and a Coulomb phase with a massless photon in between. For $N<5$ the Coulomb phase disappears.

Using the methods of ref. [11], the Z(N) theories can be cast in a variety of forms. The most useful for our purpose is that of a compact U(1) gauge theory coupled to charges of magnitude Ne. The Wilson loop then takes the form

$$\mathrm{e}^{-S[x(t)]}=Z(J)/Z(0)\,,$$

$$Z(J)=\int_0^{2\pi}\prod\mathrm{d}\theta_\mu\,\mathrm{e}^{-F(\theta_{\mu\nu})}\sum_{j_\mu}\delta(\Delta_\mu j_\mu)\,\mathrm{e}^{i\Sigma\theta_\mu(j_\mu N+J_\mu)}\,. \tag{6}$$

For large N, the charged currents j_μ acquire large electromagnetic self-energies (proportional to N^2) and do not affect eq. (6) very much (unless $e^2<N^{-2}$). Furthermore, for any N the primary role of the currents j_μ in the confining phase is to provide the screening mechanism which enables baryons to have a finite mass. We do not expect virtual baryon-antibaryon states to play a significant role in χSB and may thus neglect the effects depending on Nj_μ.

Our contention then is that the chiral properties of quarks in an SU(N) gauge theory dominated by fat fluxons are simiar to those of massless charged fermions in a Z(N) gauge theory in its confining phase. The latter, in turn, will be similar to the properties of massless fermions in a U(1) gauge theory. The term "similar" means specifically that the behavior of the chiral order parameter as a function of e in these theories differs only quantitatively.

Having motivated at some length our preoccupation with an abelian U(1) gauge theory we now turn to the fermions and their coupling. In sect. 3 we shall introduce a new formulation designed to avoid some of the notorious troubles of lattice fermions.

3. Second-order lattice fermions

The generating functional for the Green functions of all gauge-invariant local Dirac bilinears in a non-abelian gauge theory is given by

$$Z[M(x)] = \int dA\, e^{-S_{\mathrm{YM}}(A)} \frac{\det[(1/i)\not{\partial} - \not{A} + M]}{\det[(1/i)\not{\partial}]} . \tag{7}$$

$M(x)$ is an arbitrary x-dependent matrix in the Dirac and flavour indices which commutes with the gauge-group generators. $S_{\mathrm{YM}}(A)$ is the pure gauge-field action including gauge-fixing terms, Fadeev–Popov determinant and counter terms.

In order to regularize eq. (7) by a lattice, a consistent discretization procedure of the Dirac determinant should be developed. The naive procedure which replaces the derivatives in the Dirac operator by nearest-neighbour finite differences fails, since it leads to

$$\det\left(\frac{1}{i}\not{\partial}_{\text{discrete}}\right) \to \det{}^8\left(\frac{1}{i}\not{\partial}_{\text{continuum}}\right) \tag{8}$$

when the lattice constant tends to zero. Since the logarithmic derivatives of the determinant are equal to those of the spectral density, we conclude that the naive latticization generates 8 degenerate Dirac fermions. This is the problem of spectral mutliplication which plagues lattice fermion theories [12]. It originates from the fact that the Dirac equation is first order.

Observe now that

$$\begin{aligned}\det\left(\frac{1}{i}\not{\partial} - \not{A} + M\right) &= \det \gamma_5\left(\frac{1}{i}\not{\partial} - \not{A} + M\right)\gamma_5 \\ &= \det\left(-\frac{1}{i}\not{\partial} + \not{A} + M_5\right),\end{aligned}$$

where $M_5 \equiv \gamma_5 M \gamma_5$. Thus,

$$\begin{aligned}\det\left[\frac{1}{i}\not{\partial} - \not{A} + M\right] = \det{}^{1/2}\Big\{\left(\frac{1}{i}\partial_\mu - A_\mu\right)^2 &+ \tfrac{1}{2}\sigma_{\mu\nu}F_{\mu\nu} + (M\gamma_5)^2 \\ &+ \left[\left(\frac{1}{i}\not{\partial} - \not{A}\right)\gamma_5, M\right]\gamma_5\Big\} .\end{aligned} \tag{9}$$

The discrete γ_5 symmetry we have used is not plagued by anomalies and is always valid. In fact, γ_5 is a product of two 180° O(4) rotations so that a breakdown of this symmetry would imply a breakdown of Lorentz invariance (this is of course, just the *CPT* theorem).

The differential operator in the curly brackets on the r.h.s. of eq. (9) is second order; in fact, the covariant derivatives appear in the form appropriate to a scalar field. The finite-difference laplacian is known to converge to the continuum version; we are thus assured that the determinant of a gauge invariant finite-difference approximation of this operator will have the correct formal continuum limit. In appendix A we check the consistency of the above statement in perturbation theory by calculating the lowest-order vacuum polarization in QED using the quadratic lattice Dirac operator.

As in all lattice theories, the choice of the action is somewhat arbitrary. The allowed class of lattice lagrangians should be gauge invariant, nearest neighbour, respect all the symmetries of the continuum theory and possess the correct formal continuum limit for weak coupling.

Our choice for the second-order lattice Dirac operator is

$$D=\sum_{\mu}\left[\delta_{x,x'}-U_{\mu}(x)\delta_{x,x'-\hat{\mu}}\right]+\frac{K}{8i}\sum_{\mu\nu}\left[U_{\mu\nu}(x)-U^{\dagger}_{\mu\nu}(x)\right]\sigma_{\mu\nu}\delta_{x,x'}\,. \tag{10}$$

The summation over μ is carried over all eight links connected to the point (recall that $U_{-\mu}(x)=U^{\dagger}_{\mu}(x-\hat{\mu})$. The summation over $(\mu\nu)$ includes all the plaquettes which contain the point x. This means that for every pair $(\mu\nu)$ four terms appear in the sum.

In eq. (10), $U_{\mu}(x)=U(x,x+\hat{\mu})$ is the standard lattice gauge theory link operator, $U_{\mu\nu}(x)$ is the plaquette operator and $\sigma_{\mu\nu}$ are the Dirac spin matrices:

$$U_{\mu\nu}(x)=U^{\dagger}_{\nu}(x)U^{\dagger}_{\mu}(x+\hat{\nu})U_{\nu}(x+\hat{\mu})U_{\mu}(x)\,,$$

$$\sigma_{\mu\nu}=\frac{1}{2i}\left[\gamma_{\mu},\gamma_{\nu}\right]. \tag{11}$$

The summation over all four plaquettes of each pair $(\mu\nu)$ insures cubic and parity invariance. Some care is required in defining the orientation of the four plaquettes which enter into the sum. Also, in a non-abelian theory the $U_{\mu\nu}(x)$'s, which are not gauge invariant, should be defined as in eq. (11): they should start and end at the point x.

The correct formal continuum limit is attained if the constant K is set equal to 1. This is precisely the value picked out by the perturbative calculation of appendix A. The significance and choice of K will be discussed at length below.

An ideal lattice fermion formalism should avoid the spectral multiplication problem while retaining nearest neighbor couplings and should exhibit $U_F\times U_F$ chiral symmetry (where F is the number of flavors). In particular $2F^2-1$ local conserved currents should be identified. The singlet axial current should have an anomaly. Finally, expressions for the scalar and pseudoscalar densities $\bar{\psi}\lambda_f\psi$, $\bar{\psi}\lambda_f\gamma_5\psi$ should be found and a lattice version of Goldstone's theorem proved. Once this is done we can concentrate our attention on the value of the order parameter $\langle\bar{\psi}\psi\rangle$. Goldstone's theorem will guarantee that a non-zero value for this object implies the existence of F^2-1 massless bosons.

We do not yet know whether our second-order formalism preserves continuous chirality. Our analysis of this question will be reserved for a future publication [7]. We certainly preserve (by construction) the discrete mass reversal symmetry and therefore $\langle\bar{\psi}\psi\rangle$ will be a valid order parameter. In this respect our formalism is similar to that of Susskind [12]. Note, however, that we have only one fermion (per flavor) in the continuum limit.

The most important virtue of our formalism is that it proves a dynamical explanation of how χSB occurs in strongly coupled lattice theories. This picture can be immediately generalized to the continuum and constitutes a heuristic proof of our contention that fermion–antifermion bound states can occur only when chiral symmetry is broken (sect. 5).

To obtain a convenient representation for the order parameter we note that in the continuum limit we have [from eq. (7)]

$$\int \mathrm{d}^4x\langle\bar{\psi}\psi\rangle = -\frac{\partial}{\partial m}\ln Z(m)\,, \tag{12}$$

$$Z(m) = Z(M(x)=m)\,. \tag{13}$$

On the lattice

$$\frac{\partial}{\partial m}\ln Z(m) = \left\langle \mathrm{Tr}\,\frac{m}{D+m^2}\right\rangle, \tag{14}$$

where the brackets have the same meaning as those in eq. (12), and Tr is performed over coordinates, color flavor, and spin. The rest of the paper will be devoted to an approximate evaluation of eq. (14).

It remains to discuss the parameter K. A full discussion is relegated to sect. 5. For the moment, observe that K is a dimensionless coupling constant and thus, is subject to renormalization: its bare value should be a function of the cutoff in order to derive fixed renormalized Green functions. Moreover, in order to obtain the correct Dirac theory the K counter terms should be tuned and the bare K becomes a definite function of the gauge coupling constant g. The classical value of K [eq. (9)] is 1, and in perturbation theory this value will be corrected. The lowest-order correction could be computed from the one-loop vertex and wave-function renormalization graphs. On the lattice the above remarks translate into the requirement that K be a function of $g^2(K(g^2))$ when g approaches the relevant fixed point (zero for QCD). We shall, of course, be interested in lattice theories with a g which lies above the confinement critical point.

4. χSB and Z(N) fluxons

It was argued in sect. 2 that the effects of confinement on matter fields may be approximated by a U(1) lattice gauge theory in the strong coupling phase. Designating the link phase variable by θ_μ we thus obtain for the vacuum amplitude of the fermion-gauge field system:

$$Z(m)=\int_0^{2\pi}\prod \mathrm{d}\theta_\mu(x)\,\mathrm{e}^{-F(\theta_{\mu\nu})}\det{}^{1/2}[D(\theta_\mu)+m^2]\,, \tag{15}$$

where $F(\theta)$ is the effective U(1) action [see eqs. (2)–(6) and the accompanying discussion] and $D(\theta)$ is the abelian second-order lattice Dirac operator [eq. (10) with $U_\mu(x)=\exp i\theta_\mu(x)$]. The determinant is a guage-invariant periodic function of the variables $\{\theta_\mu(x)\}$ and may therefore be expanded in a multiple Fourier series:

$$\det{}^{1/2}[D(\theta)+m^2]=\sum_{\{J_\mu\}}\delta[\Delta_\mu J_\mu]d[J_\mu]\,\mathrm{e}^{i\Sigma\theta_\mu J_\mu}\,, \tag{16}$$

where $d[J_\mu]$ are the expansion coefficients. $\{J_\mu\}$ are all integer-valued conserved currents on the lattice. Note that $d(J)$ contains the effects of the spin-dependent interactions and depends on g^2, K and m. We now expand also the gauge-field functional in a Fourier series:

$$\mathrm{e}^{-F(\theta_{\mu\nu})}=\sum_{\{f_{\mu\nu}\}}\mathrm{e}^{-\tilde F(f_{\mu\nu})}\,\mathrm{e}^{i\Sigma f_{\mu\nu}\theta_{\mu\nu}} \tag{17}$$

where $\{f_{\mu\nu}(x)\}$ are integer-valued antisymmetric fields defined on plaquettes. [$\tilde F(f_{\mu\nu}))=g^2f^2_{\mu\nu}$ for the Villain model; $\tilde F(f_{\mu\nu})=-\ln I_{f_{\mu\nu}}(g^{-2})$ where I_f is a Bessel function, for the Wilson–Polyakov model.] Substituting eqs. (16), (17) into eq. (15) we may perform the θ_μ integration:

$$\begin{aligned}\int_0^{2\pi}&\frac{\mathrm{d}\theta_\mu}{2\pi}\exp[i\theta_\mu(J_\mu-\Delta_\nu f_{\nu\mu})]\\ &=\prod_{x,\mu}\delta(J_\mu-\Delta_\nu f_{\nu\mu})\\ &=\lim_{\gamma\to 0}\int_{-\infty}^{\infty}\mathrm{d}a_\mu(x)\exp\left[-\frac{1}{2\gamma^2}\Sigma a_\mu^2\right]\exp\left[i\Sigma\, a_\mu(J_\mu-\Delta_\nu f_{\nu\mu})\right]\delta(\Delta_\mu a_\mu)\,.\end{aligned} \tag{18}$$

Hence,

$$\begin{aligned}Z(m)=\lim_{\gamma\to\infty}&\int_{-\infty}^{\infty}\mathrm{d}a_\mu\delta(\Delta_\mu a_\mu)\\ &\times\exp\left[-\frac{1}{2\gamma^2}\Sigma\, a_\mu^2-F(\Delta_\mu a_\nu-\Delta_\nu a_\mu)\right]\det{}^{1/2}[D(a)+m^2]\,,\end{aligned} \tag{19}$$

where $D(a)$ and $F(a)$ are now regarded as functions of the non-compact field a_μ (the term $\gamma^{-2}a_\mu^2$ breaks the periodicity of the functional). Eq. (19) describes the interaction of our fermions with a massive vector field a_μ (for small g^2 the mass is given by $\mu^2 = g^2\gamma^{-2}$). The exchange of the vector field induces an attractive non-local vector–vector four-Fermi interaction whose strength is a function of the gauge coupling constant g and the parameter γ. The nature of this interaction may be best understood if the vector field action is modified and chosen to be simply $g^{-2}(\Delta_\nu a_\mu - \Delta_\mu a_\nu)^2$, namely, that of a free field. Other terms in the action have higher derivatives and will effect the large-distance behavior only by renormalizing g^2; in that case the a_μ propagator is

$$\Delta_{\mu\nu}(p) = \left(\delta_{\mu\nu} - \frac{K_\mu K_\nu}{K^2}\right)\left(\gamma^{-2} + 4g^{-2}\sum_\mu \sin^2 \tfrac{1}{2}p_\mu\right)^{-1}, \qquad K_\mu = e^{ip_\mu} - 1 \tag{20}$$

(recall that $-\pi < p_\mu < \pi$ on the lattice). If both g^2 and γ^2 are sufficiently large the effective four-Fermi interaction would exceed the Nambu–Jona-Lasinio [6] (NJL) critical strength G_{NJL} and chiral symmetry would be broken. In the standard local one-loop Hartree–Fock approximation where $\Delta_{\mu\nu}(p)$ is replaced by $\Delta_{\mu\nu}(p_\mu = \pi)$, the condition becomes

$$\frac{4N_{\mathrm{F}}}{\gamma^{-2} + 4g^{-2}} \int_{-\pi}^{\pi} \frac{\mathrm{d}^4 p}{(2\pi)^4} \frac{1}{4\sum_\mu \sin^2 \frac{1}{2}p_\mu} \geqslant 1 \tag{21}$$

where N_{F} is the number of fermion species. We thus conclude that if the gauge coupling exceeds the NJL critical value, our $\gamma \to \infty$ theory exhibits χSB since it becomes essentially equivalent to an NJL model. One important point was ignored in the above argument, namely, the K-dependence. In order that the model defined by eq. (19) be really the 2nd-order version of the non-compact vector–fermion theory, $K(g^2)$ has to be adjusted correctly as a function of γ^2. It is, however, clear [eq. (20)] that for a sufficiently large γ^2 the low momentum behavior of the fermions will not be affected much by γ so that the original curve $K(g^2)$ will be corrected only to order $\mathrm{O}(\gamma^{-2})$.

We have thus shown that if g^2 resides sufficiently high inside the confinement phase, chiral symmetry will be broken. Note, however, that this does not constitute a complete proof of our conjecture that confinement implies χSB. What is missing is a proof that the χSB region of the phase diagram covers at least the whole confinement phase down to the transition line to the Coulomb region. We do not yet have such a proof, although the considerations of sect. 5 will be seen to provide a heuristic basis for this assertion.

5. General theory of χSB

We now wish to generalize the considerations of sect. 4 and investigate χSB in a wider context. While we believe that the Z(N) approximation adequately describes the physics of confinement and indeed provides an explicit model, we would like to formulate and prove χSB in a model-independent way. To this end, the only assumption we shall make concerning confinement is that the Wilson area law is satisfied: in the confining phase of the lattice theory (which for YM is presumed to extend down to $g_0 \to 0$) we have

$$\mathrm{Tr}\,\langle U_{\mu_1}(x_1)\cdots U_{\mu_T}(I_T)\rangle = \mathrm{e}^{-\mu^2 A(\mathrm{C})}\,, \tag{22}$$

where $(x_1, \mu_1)\cdots(x_T, \mu_T)$ form a closed contour C, A(C) is the minimal area spanned by C and μ^{-1} is the confinement scale. Observe that the minimal area law automatically provides for screening as may be seen by considering loops of complicated shapes.

Let us now return to the exact expression for $\langle\bar{\psi}\psi\rangle_m$ [eqs. (13), (14)] and recast it in a form which will be a functional of Wilson loops. The Dirac operator D [eqs. (10), (11)] contains a purely local part proportional to $\delta(x, x')$ and a nearest neighbour hopping term $U_\mu(x)\delta(x, x'-\hat{\mu})$. Hence the trace of the inverse of (m^2+D) may be represented as

$$\mathrm{Tr}\,(m^2+D)^{-1} = \mathrm{tr}\sum_{\{\mathrm{C}\}} \Delta_K^{-1}(x_1)U_{\mu_1}(x_1)\cdots U_{\mu_n}(x_n)\Delta_K^{-1}(x_n)\cdots U_{\mu_T}(x_T)\,, \tag{23}$$

where Δ_K is a finite-dimensional matrix,

$$\Delta_K(x) = (8 + K\sigma\cdot F(x) + m^2)\,, \tag{24}$$

and C designates a sum over all closed paths. The notation $\sigma\cdot F$ is a shorthand for the terms in D which involve $\sigma_{\mu\nu}U_{\mu\nu}(x)$. The continuum version of eq. (23) is gotten by parametrizing the path $(x_1, \ldots, x_T)$ by a continuous parameter τ and converting the sum into an integral while putting $K=1$. The result is Schwinger's formula [15] for the Dirac propagator expressed as a path integral:

$$\mathrm{Tr}\,(m^2+D(A))^{-1} = \mathrm{Tr}\int_0^\infty \mathrm{d}T\,\mathrm{e}^{-m^2T}\int_{x(0)=x(T)} \mathrm{d}x(\tau)\exp\left[-\int_0^T \mathrm{d}\tau\tfrac{1}{4}\dot{x}^2\right]$$
$$\times\left(\exp\left[i\oint \mathrm{d}x_\mu A_\mu - \frac{1}{2}\int_0^T \mathrm{d}\tau\sigma\cdot F\right]\right)_+\,. \tag{25}$$

Note that the free particle propagator $\exp[-\frac{1}{4}\int \mathrm{d}\tau\dot{x}^2]$ is the counterpart of the term 8^{-T} of eq. (23). The latter is the correct weighting of the four-dimensional random walk on a cubic lattice which describes the propagation of a massless particle. Before proceeding, we shall make on simplification. The exact formula for $\langle\bar{\psi}\psi\rangle$ contains also the determinant of $D^{1/2}$. The latter may be expressed as the exponential of (minus) the sum over the closed loops (23), (24) weighted with a factor T^{-1} to avoid double counting. It is generally believed (and proved in the limit $N\to\infty$) that the meson spectrum and χSB are insensitive to the neglect of closed fermion loops. More precisely we do not expect multi-loop corrections to change the nature of the phases and the critical behavior of the theory. This type of generalized Hartree–Fock approximation is also employed in the NJL model [6]. We shall thus work in the approximation for one fermion loop. Note that in this approximation a Goldstone boson will occur also in the singlet channel since the anomaly can influence the meson spectrum only in the presence of closed fermion loops [13]. Some quantitiative justification for the approximation will be provdided in sect. 6.

The condition for χSB is that $\lim_{m\to 0}\langle\bar{\psi}\psi\rangle_m \neq 0$. An examination of eqs. (23), (25) now provides us with a precise condition for this to happen. Since these expressions should be multiplied by m, their large T behaviour controls $\langle\bar{\psi}\psi\rangle_0$, and we find

$$\langle\bar{\psi}\psi(x)\rangle_{\rm cont}\neq 0 \Rightarrow \mathrm{tr}\int_{x(0)=x(T)=x} \mathrm{d}x(\tau)\exp\left[-\frac{1}{4}\int_e^T \mathrm{d}\tau\dot{x}^2\right]$$
$$\times\left\langle\left(\exp\left[i\oint \mathrm{d}x_\mu A_\mu - \frac{1}{2}\int_0^T \mathrm{d}\tau\sigma\cdot F\right]\right)_+\right\rangle \underset{T\to\infty}{\sim} T^{-1/2}\,. \quad (26)$$

The corresponding condition on the lattice is

$$\langle\bar{\psi}\psi(x)\rangle_{\rm lat}\neq 0 \Rightarrow \sum_{C(T|x)} \mathrm{tr}\, 8^{-T}U(1)\cdots(1+\tfrac{1}{8}K\sigma\cdot F)^{-1}$$
$$\times\cdots U(n)\cdots U(T) \underset{T\to\infty}{\sim} T^{-1/2}\,. \quad (27)$$

In other words, the class of contours $C(T|x)$ which contributes to $\bar{\psi}\psi(x)$ should contain a sufficiently large number of long ($T\to\infty$) paths such that the large-T behaviour of the propagator is $O(T^{-1/2})$. For comparison, we observe that the corresponding behaviour of free particles in d dimensions is $O(T^{-d/2})$:

$$\int_{x(0)=x(T)=0} \mathrm{d}x(\tau)\exp\left[-\frac{1}{4}\int_0^T \mathrm{d}\tau\dot{x}_\mu^2\right] = (\pi T)^{-d/2}\,. \quad (28)$$

On the lattice, free massless propagation is represented by random walks weighted by the factor $(2d)^{-1}$. We find for the amplitude to return to the origin

$$\langle 0T|00\rangle_{\rm lat} = \int_{-\pi}^{\pi}\frac{\mathrm{d}^d p}{(2\pi)^d}\left(1-\frac{4}{2d}\sum_\mu \sin^2(\tfrac{1}{2}p_\mu)\right)T \underset{T\to\infty}{\sim} T^{-d/2}\,. \quad (29)$$

Eqs. (26), (27) mean that the condition for χSB is that the effective propagation of the fermions in the gauge field vacuum be one-dimensional. $\langle m^{-1}\bar{\psi}\psi\rangle$ represents the Laplace transformed amplitude that a propagating second-order fermion return to the origin; for this to diverge as m^{-1} it is necessary that the interaction of the fermion loop with the gauge field be sufficiently strong to reduce the number of available paths so that the probability of returning to the origin increases from the four-dimensional value T^{-2} to the one-dimensional value $T^{-1/2}$.

Consider first the strong coupling limit ($g^2\to\infty$) in which the confinement scale μ^{-1} becomes zero. If we neglect the terms $\sigma\cdot F$ in eq. (27) and replace Δ by 8, we conclude that the only closed paths C which contribute in this limit are of zero width.

In line with our generalized HF (or large-N) approach we shall concentrate only on those configurations (fig. 1) which form trees and neglect loop corrections. The sum over C may be performed in three stages. First specify the particular tree and its total length L, then sum over the random walk along each tree and finally sum over the trees. It is, however, readily verified that the internal random walk merely results in a small renormalization of the factor 8^{-1} and an overall numerical factor. For simplicity these effects will be neglected. We thus have approximately

$$\langle\bar{\psi}\psi\rangle_m \sim m \sum_{L=0}^{\infty} N_t(L) 8^{-2L}, \tag{30}$$

where $N_t(L)$ is the number of trees of length L (with a fixed initial point) in four dimensions. We therefore have to count the trees and extract the large-L behaviour of their number. We shall accomplish this in two stages. Let $N_t^0(L)$ be the number of trees of total length L whose initial point ($x=0$) lies at the tip of some branch. Designate the Laplace transform of $N_t^0(L)$ with respect to $L(\equiv\sum_L \rho^L N_t^0(L))$ by $\tilde{N}_t^0(\rho)$ and the corresponding quantity for the number of unrestricted random walks by $\tilde{N}_r(\rho)$. Any tree contributing to N_t^0 starts out from the origin as a random walk (in fact the random walks are a sublcass of trees). It then branches into two or more walks each of which may develop into the most general tree consistent with the constraint that the total length of the entire structure is L. Thus $N_t^0(L)$, will satisfy a non-linear convolution equation. $\tilde{N}_t^0(\rho)$ which is a Laplace transform of N_t^0 then satisfies a non-linear algebraic equation. The details of this equation depend on the number of branches we allow at each point which in turn depends on the type of lattice and other unphysical aspects of our approximation. In general we could expect a probability P_n for emitting n branches, with P_n determined by the lattice. Fortunately the asymptotic behaviour of $N_t^0(L)$ is independent of these details. It is a sort of a critical index like the $L^{-d/2}$ behaviour of the free propagator. In fact it is even more universal since it is dimension independent.

For simplicity we write the equation for $\tilde{N}_t^0$ in the case where an indefinite number of branches are allowed with equal *a priori* probability

$$\tilde{N}_t^0 = \tilde{N}_r + \tilde{N}_r \sum_{n=2}^{\infty} \frac{1}{n!} (\tilde{N}_t^0)^n = \tilde{N}_r (e^{\tilde{N}_t^0} - \tilde{N}_t^0), \tag{31a}$$

$$\tilde{N}_r = \frac{2\,d\rho}{1-2\,d\rho}. \tag{31b}$$

An asymptototic behaviour of $N_t^0(L)$ as

$$L^p \rho_c^{-L} \tag{32a}$$

leads to a singularity in N_t^0 of the type

$$(\rho-\rho_c)^{-p-1}. \tag{32b}$$

The solution of (31a) is singular at $\rho_c^{-1} = 2de$. The singularity comes from the confluence of two roots and is of square root type. This mechanism is general. Other assumptions about the number of branches lead to an equation of the form

$$\tilde{N}_{\mathrm{t}}^{0}=\frac{\tilde{N}_{\mathrm{r}}}{1+\tilde{N}_{\mathrm{r}}}f(\tilde{N}_{\mathrm{t}}^{0})\,,$$

with concave f. Thus the singularity in $\tilde{N}_{\mathrm{t}}^{0}$ always comes from the confluence of exactly *two* roots and is of square root type.

We conclude that the large-L behaviour of N_{t}^{0} is

$$N_{\mathrm{t}}^{0}(L)\underset{L\to\infty}{\sim}L^{-3/2}\rho_{\mathrm{c}}^{-L}\,. \tag{33}$$

We note that the quantity we are after is not N_{t}^{0}, but N_{t}, which differs from N_{t}^{0} by virtue of the fact that the initial point ($x=0$) need not lie at the tip of a branch. Since the number of tips is some fraction $(1-f)$ of the total length T we have:

$$N_{\mathrm{t}}(L)\underset{L\to\infty}{\sim}fL^{-1/2}\rho_{\mathrm{c}}^{-L}\,, \tag{34}$$

where in four dimensions

$$8<\rho_{\mathrm{c}}^{-1}=8e<8^{2}\,. \tag{35}$$

Note that ρ_{c} lies between 8 and 8^2 which means that N_{t} is bounded (below) and above by the number of random walks of length (L) and $2L$ respectively. Eqs. (33)–(35) show that $N_{\mathrm{t}}(L)$ exhibits the needed $L^{-1/2}$ behaviour at large L; however, the exponential is clearly "wrong" and we have in this case ($K=0$):

$$\langle\bar{\psi}\psi\rangle_{\substack{K=0\\ g=\infty}}\propto m\sum_{L}\left[\frac{8e}{((2+m^{2})^{2}}\right]^{L}L^{-1/2}\xrightarrow[m\to 0]{}0\,. \tag{36}$$

Before proceeding, we emphasize that the value $\frac{1}{8}e<1$ for the exponent is not just a numerical fluke of the particular approximations we have used. As noted, the geometrical factor ρ_{c}^{-1} will always be smaller than $(2d)^2=64$ so that the series (36) always converges thus leading to $\langle\bar{\psi}\psi\rangle=0$. At the same time, the $L^{-1/2}$ behaviour is also very general and is a consequence of the square root singularity inherent in the counting of trees, and of the necessity of moving the initial point of the loop along the tree.

It would seem that we arrived at a paradox. The neglect of the spin forces has led to a confining theory with no spontaneous chiral symmetry breaking. This fact is actually very general. An examination of the general expression for $\langle\bar{\psi}\psi\rangle$ [eqs. (26), (27)] immediately reveals that for $K=0$ a strict inequality obtains (for both the lattice and continuum)

$$\langle\bar{\psi}\psi\rangle_{K=0}<\langle\bar{\psi}\psi\rangle_{\text{free field}}\xrightarrow[m\to 0]{}0\,. \tag{37}$$

The reason is that in the interacting theory each free lop is multiplied by the average of a Wilson loop and the latter is bounded by a constant:

$$\langle W(\mathrm{C})\rangle < \mathrm{tr}\, 1\,. \tag{38}$$

This seems to contradict our contention that confinement implies χSB. Note, however, that the qualitative argument of ref. [4] explicitly uses the identity of helicity and chirality which obtains only in the Dirac theory and not for a general value of K. The role of the confining force is to inhibit the quark loops from opening up and enforce the tree structure. We shall now show that the short-range spin interaction supplies the enhancement of the exponent in eq. (36) which is needed in order that the series diverge as m^{-1}. The effect of the spin term on $\langle\bar{\psi}\psi\rangle$ is manifested through the insertions of the matrices $\Delta^{-1}=(1+\frac{1}{8}K\sigma\cdot F)^{-1}$ into the Wilson loops [eq. (27)]. These matrices are hermitian, and positive definite (for $K<\infty$). Moreover, for any configuration of the U's some of the eigenvalues of $\Delta^{-1}(U)$ will be larger than 1 due to the possibility of rotating the spin matrices. The matrix $\Delta(x)$ is a local function of the plaquettes $U_{\mu\nu}(x)$ [eq. (11)] and in the confining phase, because of the area law, the correlations between such objects are necessarily exponentially damped. In the extreme strong coupling limit we are considering these correlations extend over a single lattice spacing. These facts imply that the effect of the spin term on the loops may be summarized by a finite K-dependent renormalization of the coefficient of L in eq. (36):

$$8^{-2L}\to(\gamma(K)8^{-2})^{L}\,. \tag{39}$$

$\gamma(K)=1$ for $K=0$ and is an increasing function of K. Recall now that we are ultimately interested in the infrared correlations of the continuum theory, so that the parameter K should be driven to its critical value K_c, where the fermion correlations become long range. This value is attained precisely when $\gamma(K)$ cancels the exponential damping of the $\langle\bar{\psi}\psi\rangle$ (or any other Green function) series. Taking into account the tree counting factor ρ_c^{-1} we have

$$\gamma(K_c)=8^2\rho_c=8e^{-1}\,. \tag{40}$$

At $K=K_c$ the large-L behaviour of the loops is precisely $\mathrm{O}(L^{-1/2})$ and we have then

$$\langle\bar{\psi}\psi\rangle_{K_c}=s(d)\,, \tag{41}$$

where $s(d)$ is a calculable function of the dimensionality (clearly $d=4$ in our case). Before we discuss the meaning of this remarkable effect of the $\sigma\cdot F$ term, let us extend the result away from $g^2=\infty$. When $g^2\neq\infty$, an expansion of the gauge-field term $\exp g^{-2}\Sigma\,\mathrm{tr}\,U^2_{\mu\nu}$ should be carried out. Observe that $\mathrm{tr}\,U_{\mu\nu}$ is a short-range interaction, so that terms of order $(g^2)^{-n}$ couple at most $[z(d)]^n$ points on the lattice, where $z(d)$ is a geometrical dimension-dependent number. Moreover, the variables U and their phase space are bounded. Hence, the strong coupling expansion of any observable is bounded by a geometric series of the form $\Sigma_n(Uz(d)g^{-2})^n$, whose radius of convergence is finite. Within this radius the tree structure of dominant contributions will not change. The effects of turning on g^{-2} are summarized by a set of finite renormalizations:

$$\begin{aligned}\text{area law: } \mu^2 &\to \mu^2(g^2)\,,\\ K_c &\to K_c(g^2)\,.\end{aligned}\tag{42}$$

As a result, the tree expansion of $\bar{\psi}\psi$ retains its form except that the scale of L has to be renormalized since the effective trees are now of finite width. We thus have:

$$\langle\bar{\psi}\psi\rangle_{g^2} \sim s\xi^{-3}(g^2)\,,\tag{43}$$

where $\xi(g^2)$ is the correlation length in units of the lattice spacing and the ξ dependence arises because the counting has to take into account the multiplicity of paths due to the appearance of trees of thickness $\sim\xi$. The above discussion is certainly rigorous when g^{-2} is inside the radius of convergence of the strong coupling power series. Extending the argument to the full domain of analyticity – the whole confinement phase – is not completely straightforward. The point is that while the area law continues to operate and damp open large loops, highly "ergodic" loops which are neither tree-like nor "free field" in nature might contribute. The chiral order parameter will, however, continue to be non-zero so long as long trees contribute a non-vanishing fraction at $K = K_c(g^2)$. We do not have an argument which precludes the complete disappearance of trees at some g_c^2. We believe however that this would signify a phase transition which would destroy the continuum limit of the confinement phase. In any case, a dominance (with measure = 1) of non-tree-like configurations would mean that the spectrum could not be described in terms of bound quark–antiquark pairs. We conclude the discussion of $\langle\bar{\psi}\psi\rangle$ by stressing that the necessary condition for χSB is the presence of (but not saturation by) long trees. This implies that χSB will occur whenever the forces are sufficiently strong to induce binding, even if they are non-confining. This is in line with the intuitive arguments of ref. [5] which were described in sect. 1. A brief qualitative discussion of the evaluation of f_π in our approach will be given in sect. 6.

Finally, we wish to return to the spin term and its effect on the theory. What we have seen is the term $K\sigma \cdot F$ induces a negative contribution to the squared mass of the propagating fermions. By turning on the coupling constant K the total "renormalized" mass may be forced to vanish and the system then reaches a critical line $K_c(g^2)$ wherein the fermionic correlations are long range. This phenomenon is not limited to the strong coupling phase but occurs also in perturbation theory. Consider a continuum theory, which may be called NQQED, where the second-order lagrangian has $K \neq 1$. In this theory, which is not quite QED, the mass counter term may be calculated by ordinary Feynman graphs. In second order there will be three contributions (fig. 2). The graphs 2a, b are those of scalar electrodynamics and contribute together a positive quadratically divergent self-(mass)^2. The graph (2c) is the spin effect and contributes a negative (mass)^2 proportional to K^2. When $K = 1$

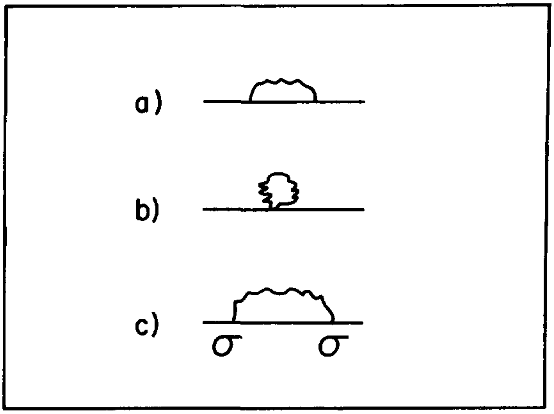

Fig. 2. Contributions to the second-order fermion self-mass in NQQED.

the three contributions exactly cancel leaving the fermion mass equal to zero. In higher orders K would have to be renormalized in order that the quadratically divergent self-(mass)^2 continue to vanish. Note now that our condition for determining K in the strongly coupled lattice theory was precisely the cancellation of the self-(mass)^2. Thus, our critical line $K_c(g^2)$ represents the unique choice of the bare K which guarantees the equivalence of the continuum limit of the second-order lattice theory and the original Dirac theory. This perturbative argument thus justifies our choice of K in the lattice theory to precisely cancel the exponentially falling (with L) contributions to the proper-time fermion propagator. We have determined K in the strong coupling theory by the same qualitative criterion that guarantees equivalence of the second-order formalism to the Dirac equation in the continuum limit. To be honest, however, we should note that we have not given an analytical proof that the two critical lines $K_c(g^2)_{\rm strong}$ and $K_c(g^2)_{\rm weak}$ coincide in lattice QCD. Since they are determined by the same criterion it would seem that a disagreement between them would indicate some sort of a first-order phase transition and would cast doubt on the analytic continuation of the confining phase into the region of $g \sim 0$.

6. Summary and discussion

(i) Throughout this paper we concentrated on the chiral order parameters $\langle\bar{\psi}\psi\rangle$. Once $\langle\bar{\psi}\psi\rangle \neq 0$, the Nambu–Goldstone mechanism insures the emergence of a zero-mass multiplet of pseudoscalar mesons: the pion and its friends. The main chiral property of the pion is f_π ($=95$ MeV) which should be calculable in QCD. We wish to indicate briefly a way in which this calculation may be performed*, by passing the highly complicated problem of solving for the pion wave function. If a small (current!) quark mass term m is turned on, standard PCAC and current algebra

* We owe this beautiful explanation of why $m_\pi^2 \propto m$ (current quark) to Shmuel Nussinov.

manipulations lead to the relation:

$$m_\pi^2 f_\pi^2 = m\langle\bar{\psi}\psi\rangle\,. \tag{44}$$

Having calculated $\langle\bar{\psi}\psi\rangle$, we thus have to find mm_π^{-2} (and prove that m_π^2 is actually proportional to m!). Observe now that according to our analysis, the typical length $L(m)$ of a tree is [eq. (36)]

$$L(m) \sim m^{-2}\,. \tag{45}$$

In the presence of a chiral breaking term $m\bar{\psi}\psi$, the rate at which the chiral correlations are damped as the quark propagates *along the tree* is m. m^{-1}, however, is vanishingly small compared to the length $L(m) \sim m^{-2}$; hence, the path covered by the propagating quark after m^{-1} steps is still essentially a random walk in coordinate space. The corresponding correlation length in coordinate space is therefore

$$\xi_{\text{chirality}} \sim m^{-1/2} \tag{46}$$

$\xi_{\text{chirality}}$ is, however, by definition m_π^{-1}, so that the missing proportionality constant in eq. (46) determines f_π.

(ii) In accordance with a time-honored tradition we disregarded the effects of fermion loop corrections and used the one-loop Hartree–Fock approximation. The justification for this procedure is an estimate based on the reasoning which leads to the $1/N$ expansion. Consider the replacement of a given tree of length L by all possible multi-loop configurations of total length L. Each loop carried a factor $(-4N)$, due to the trace over Dirac, colour and flavour indices. The connected terms however will be damped by the same factor. We may thus estimate the effects of the fermion loop corrections by an expression of the form:

$$\left(\frac{e\gamma(K)}{8}\right)^L \to \left(\frac{e\gamma(K)}{8}\right)^L \sum_n \binom{L}{n}\left(-\frac{1}{4N}\right)^n, \tag{47}$$

so that the fermion determinant merely induces a renormalization of K_c whose determination in the strong coupling limit is now

$$\left(1-\frac{1}{4N}\right)\gamma(K_c) \sim 8e^{-1}\,. \tag{48}$$

K_c clearly is increased in accordance with the extra screening provided by the fermionic loops which reduces the effectiveness of the area law. One further effect of the fermion loops is of course to introduce the chiral singlet anomaly as a dynamical effect. In our formalism the anomaly can be proved [7] to exist so that while $\langle\bar{\psi}\psi\rangle \neq 0$ is not affected, the singlet conservation law is destroyed and the Nambu–Goldstone boson becomes massive by the mechanism elucidated in ref. [16].

(iii) The second-order lattice formalism has the advantage of having a correct formal continuum limit. The performance of euclidean lattice calculations in this formulation is no more (and no less) complicated than Wilson's chiral breaking first-order lagrangian. It is worth noting in this connection that the Wilson lagrangian also has an arbitrary coupling K which should be sent to a critical line $K_c(g^2)$ [17] in the continuum limit. There is, however, a deep conceptual issue which we believe should be resolved in order to fully understand the theory of Fermi fields. For $K \neq K_c(g^2)$ the second-order lagrangian cannot have a consistent continuum limit even in weak coupling since it is not equivalent to a Dirac lagrangian. In particular, for $K = 0$ we recover the euclidean theory of four scalar fermions! The spin-statistics theorem thus precludes a consistent hamiltonian interpretation. It should be instructive to trace the connection between the spin-statistics theorem, the existence of a Dirac hamiltonian, and the tuning of K to its critical value.

(iv) We have concentrated in our development on the approximation of the large-distance behaviour of the Dirac operator by means of a lattice. This was done mainly because the effects of postulating confinement can be systematically analysed through the strong coupling expansion. The problem may however be attacked through a direct analysis of the continuum Dirac operator provided a correct class of confining gauge field configurations is identified. The point is that eq. (14) for $\langle\bar{\psi}\psi\rangle$ may be expressed in terms of the eigenvalue spectrum of the Dirac operator

$$\int \mathrm{d}^4x\langle\bar{\psi}\psi\rangle = \lim_{m\to 0}\left\langle\int \mathrm{d}\mu\rho(\mu\,;A)\frac{m}{\mu^2+m^2}\right\rangle, \tag{49}$$

where $\rho(\mu\,;A)$ is the spectral function of $D(A)$. Since the r.h.s. of eq. (49) is positive definite, the condition that $\langle\bar{\psi}\psi\rangle$ be non-vanishing is that the r.h.s. be non-zero for a finite fraction of the configurations $A_\mu(x)$ which contribute to the average. Clearly we need

$$\rho(\mu\,;A) \underset{\mu\to 0}{\sim} \text{const}\,, \tag{50}$$

which precisely corresponds to the "one-dimensional" propagation already encountered. Thus, a viable approach would be to investigate $\lim_{\mu\to 0}\rho(\mu\,;A)$ in the presence of a "typical" distribution of one's favorite confiners (monopoles, $Z(N)$ fluxons and any combination thereof). We emphasize however, that K should still be renormalized as a function of the relevant infrared scale (Λ) since this renormalization represents the effects of smaller scale gauge-field fluctuations. The problem thus is whether a $K(\Lambda)$ exists which drives $\rho(\mu\,;A)$ to the χSB behaviour. Our lattice theory may be regarded as an approximate calculation of $\rho(\mu\,;A)$, where A has a scale $\Lambda \sim \xi(g^2)$, so that we believe a continuum approach will yield similar results.

(v) Having analysed the connection between confinement and χSB, it is worth returning for a moment to our starting point and inquiring whether we have really proved that χSB is the result of quark confinement. Paradoxically, our theory is actually incapable of answering that question completely. The point is that we have proved that while confinement is sufficient for χSB, a necessary condition is merely the existence of $q\bar{q}$ bound state. In order to supply a complete theoretical answer to our question, QCD has to be solved. In particular, the Gell-Mann–Low function should be computed for the range which interpolates between the asymptotic freedom region and the confinement scale. It is not theoretically inconceivable that the "medium distance" force would generate a "small" bound state [18] which has nothing to do with confinement and thus be the "real" cause of χSB. Having mentioned this possibility we observe however that it is a very remote one [at least for SU(3)]; the crude theoretical estimates which were attempted [3, 9, 18] indicate that the effective scale reaches the strong coupling region at moderately small values of the coupling constant. More to the point, phenomenologically, the spectrum of the light mesons and precocious scaling would seem to preclude the existence of any hypothetical small scale bound state in QCD. We thus conclude that quark confinement indeed induces the breakdown of chiral symmetry.

We have benefited from innumerable discussions with S. Nussinov. A conversation with A. Neveu on the subject of forestry is gratefully acknowledged. This paper was completed at the ULB and we thank R. Brout and F. Englert for their hospitality and for several critical discussions of our theory.

Appendix A

In this appendix we calculate the lowest-order vacuum polarization for 2nd-order lattice fermions coupled to non-compact electrodynamics. The second-order Dirac operator is

$$D=\sum_{\mu}\left[\delta_{x,x'}-U_{\mu}(x)\delta_{x,x'-\mu}\right]+\frac{K}{8i}\sum_{\mu\nu}\left(U_{\mu\nu}(x)-U^{+}_{\mu\nu}(x)\right)\sigma_{\mu\nu}\delta_{x,x'} \tag{A.1}$$

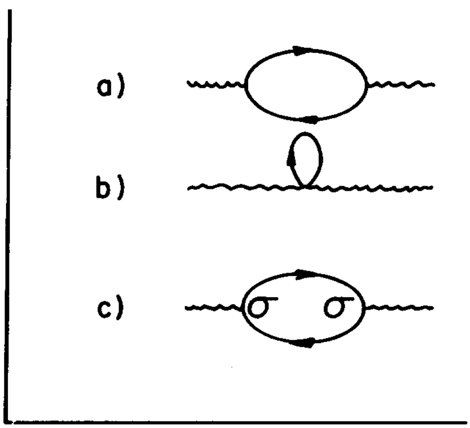

Fig. 3. Contributions to the vacuum polarization for second-order lattice fermions.

expanded out to second order in e this gives

$$D=\sum_{\mu}[\delta_{x,x'}-\delta_{x,x'-\mu}-ieA_{\mu}(x)\delta_{x,x'-\mu}+e^2A_{\mu}^2(x)\delta_{x,x'-\mu}]$$

$$+\tfrac{1}{4}Ke\sum_{\mu\nu}(\Delta_{\mu}A_{\nu}-\Delta_{\nu}A_{\mu})\sigma_{\mu\nu}+O(e^3)\,. \tag{A.2}$$

The vacuum polarization is the term of order A^2 in $\frac{1}{2}\operatorname{tr}\ln D$ and can be written in terms of Feynman diagrams. To this order the only diagrams are those of fig. 3. The term in square brackets in (A.2) gives rise to diagrams 3a and 3b. It is identical in form to the lattice version of scalar electrodynamics. Thus diagrams 3a, b will converge to twice the vacuum polarization of scalar electrodynamics in the continuum limit (remember we are calculating $\frac{1}{2}\operatorname{tr}\ln D$!). Thus we need only check that the spin-spin term converges correctly (the spin-orbit term of order e^2 vanishes because $\operatorname{tr}\sigma_{\mu\nu}=0$).

$$\frac{\delta^2}{\delta A_{\mu}\delta A_{\nu}}(\tfrac{1}{2}\operatorname{tr}\ln D^{s-s})|_{A=0}=-\tfrac{1}{16}e^2K^2\frac{\delta^2}{\delta A_{\mu}(x)\delta A_{\nu}(0)}\operatorname{tr}G_0\Delta_{\lambda}A_{\kappa}G_0A_{\alpha}A_{\beta}\sigma_{\lambda\kappa}\sigma_{\alpha\beta}\,, \tag{A.3}$$

where G_0 is the Green function of $-\Delta^2$, $G_0(p)=1/\sum_{\mu}2(\cos p_{\mu}-1)$. This is the same as the continuum expression for fig. 3c (when $K=1$) except that $1/p^2$ is replaced by G_0 and derivatives by finite differences. We must therefore check that

$$\sum_{y}e^{ipy}\Delta_{\lambda}G_0(y)\Delta_{\alpha}G_0(y)(\delta_{\alpha\lambda}\delta_{\beta\kappa}-\delta_{\alpha\kappa}\delta_{\beta\lambda}) \tag{A.4}$$

"converges" to the correct continuum limit. Of course, the result will be logarithmically divergent and contribute to the charge renormalization. To take the continuum limit we multiply by $1/a^2$ to give the correct dimensions and take $a\to 0$. The external momentum p/a is kept finite in the limit. The result is

$$\int_{-\pi/a}^{\pi/a}\frac{d^4q}{(2\pi)^4}\frac{(p+q)_{\lambda}q_{\alpha}}{(p+q)^2q^2}(\delta_{\alpha\lambda}\delta_{\beta\kappa}-\delta_{\alpha\kappa}\delta_{\beta\lambda})\,, \tag{A.5}$$

which is just the continuum integral.

References

[1] G. 't Hooft, Weizmann Institute seminar (March, 1979);
S. Samuel, Nucl. Phys. B154 (1979) 62.
[2] G. 't Hooft, Nucl. Phys. B138 (1978) 1.
[3] Y. Aharonov, A. Casher and S. Yankielowicz, Nucl. Phys. B146 (1978) 256.
[4] S. Mandelstam, Phys. Reports 23 (1976) 307;
G. 't Hooft, unpublished.
[5] A. Casher, Phys. Lett. 83B (1979) 395.
[6] Y. Nambu and G. Jona-Lasinio, Phys. Rev. 122 (1961) 345.

[7] A. Casher, in preparation.
[8] T. Yoneya, Nucl. Phys. B144 (1978) 195.
[9] D. Forster, Phys. Lett. 78B (1978) 473;
F. Wegner, J. Math. Phys. 12 (1971) 2259.
[10] R. Balian, J. Drouffe and C. Itzykson, Phys. Rev. D11 (1978) 2098;
S. Elitzur, R. Pearson and J. Shigemitsu, Phys. Rev. D19 (1979) 3698;
D. Horn, M. Weinstein and S. Yankielowicz, Phys. Rev. D19 (1979) 3715;
A. Guth, A. Ukawa and P. Windy, Cornell preprint (1979).
[11] T. Banks, J. Kogut, and R. Myerson, Nucl. Phys. B129 (1977) 493.
[12] K. Wilson, Erice Lecture Notes 1975;
L. Susskind, Phys. Rev. D16 (1977) 3031;
S. Drell, M. Weinstein and S. Yankielowicz, Phys. Rev. D14 (1976) 1627.
[14] H. Pagels, Phys. Reports 16 (1976) 219.
[15] J. Schwinger, Phys. Rev. 82 (1951) 664.
[16] J. Kogut and L. Susskind, Phys. Rev. D11 (1975) 3594;
G. 't Hooft, Phys. Rev. D14 (1976) 3432.
[17] N. Kawamoto, Amsterdam preprint (1979).
[18] C.G. Callan, R.F. Dashen and D.J. Gross, Phys. Rev. D17 (1978) 2717;
D. Caldi, Phys. Rev. Lett. 39 (1977) 121.