Lattice theories of chiral fermions

Helen R. Quinn and Marvin Weinstein

Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305

(Received 4 April 1986)

We present a Hamiltonian formulation for gauged nonanomalous chiral theories on a spatial lattice. The formulation has the desirable properties that it provides exact chiral symmetry in the massless limit, with a correct fermion spectrum *and* correct weak-coupling perturbation theory for QED. It can be generalized to non-Abelian theories. The disadvantage of the formalism is that it is complicated and nonlocal (on the lattice) and hence not easily implemented in practical calculations. However, it does provide an existence proof for a satisfactory chiral lattice gauge theory in the Hamiltonian formulation.

I. INTRODUCTION

Lattice methods are an important nonperturbative tool for analyzing gauge theories such as quantum chromodynamics. Considerable progress in understanding quantum chromodynamics has been made using lattice techniques. Unfortunately the extension of these techniques to theories involving truly chiral fermions, for example, all grand unified theories, remains problematic.

There is a widespread belief, principally due to a theorem proven by Nielsen and Ninomiya,[1] that it is impossible to define a local lattice theory *even for free chiral fermions* which does not exhibit spectrum doubling, and which has a sensible continuum limit. In fact, this belief is based on a misinterpretation of the results of Nielsen and Ninomiya. We will show for the case of free fermions that there exists a broad class of counterexamples which do not violate the Nielsen-Ninomiya theorem but which do contradict the broader interpretation which it has been given. The Nielsen-Ninomiya theorem, however, is not the only reason for the belief that there can be no satisfactory formulation of lattice theories of chiral fermions. In addition to their result, there is the work of Karsten and Smit[2] which showed that a gauge theory formulated with a long-range SLAC derivative may not satisfactorily reproduce continuum perturbation theory. Their result was that Green's functions computed in naive perturbation theory are not Lorentz covariant even after all the usual renormalizations are performed. This result was interpreted to be a consequence of the way in which the free-fermion derivative was introduced. Combined with the Nielsen-Ninomiya theorem this calculation has been taken as sufficient evidence that no satisfactory gauge theory of chiral fermions can be defined. We find that while the calculations of Karsten and Smit indicate the existence of a problem, the identification of the origin of this problem is not correct. This is easy to demonstrate for the case of noncompact lattice QED as was shown by Rabin,[3] but no comparable treatment of the compact case has been presented to date.

This paper is devoted to a reexamination of the problem of lattice chiral fermions focusing on an analysis of these earlier difficulties in order to discover at least one way of avoiding them. We explain why spectrum doubling is a misnomer for the consequences of the Nielsen-Ninomiya theorem and exhibit a class of *short-range* derivatives on lattices of finite extent for which there is no doubling. In fact we establish the stronger result that for this class of free field theories the only low-lying states are those whose momenta lie in the region near $\mathbf{k}=0$. While these examples of free field theories prove that the broadest interpretation of the Nielsen-Ninomiya theorem is incorrect, they encounter their own difficulties when one introduces interactions. In $2+1$ dimensions and higher such theories have problems similar to those found by Karsten and Smit for the long-range derivative in a gauge theory, namely, Green's functions that are not rotationally invariant even after all of the usual subtractions required by continuum renormalization are performed. In analyzing these difficulties we are lead to the conclusion that interacting chiral theories based upon short-range derivatives do not exist; however the reason for this is quite separate from the question of whether or not the theory exhibits spectrum doubling. In understanding the true origin of the problems encountered in the interacting theory we achieve a better understanding of how it is that the long-range SLAC derivative manages to avoid these problems. Turning to the case of a lattice gauge theory we find once again that the problems encountered by Karsten and Smit are not rooted in the doubling problem, but are instead due to the way in which the gauge field coupling to fermions is introduced. Once the true cause of these difficulties is identified it is not difficult to see that they can be avoided; however one pays a price in that the resulting theory is considerably more complicated and hence much less easy to compute with than the usual lattice gauge theories. Nevertheless, it is possible to write a lattice version of nonanomalous Abelian gauge theories which has all the following desirable properties: (1) Chiral symmetry is exact for zero-mass cases, with undoubled spectrum; (2) QED continuum weak-coupling perturbative theory is reproduced in the limit $a \to 0$ (g held small[4]), after the usual subtractions have been made; (3) the prescription can readily be generalized to non-Abelian theories.

It is unfortunate that the prescription which we have

found is very unwieldy. While it does reduce to the usual gauged version of the SLAC long-range derivative at strong coupling it is even more nonlocal (on the lattice) at weak coupling. It will be obvious from the final form of our Hamiltonian that carrying out lattice calculations using this formalism will be an even more formidable task than in other approaches. We do not offer this Hamiltonian as the basis of a simple computational procedure, but rather to show that the problem of constructing chiral gauge theories on the lattice has a solution, albeit an ugly one. In the concluding section of this paper we discuss the question of making approximations to this Hamiltonian which can render computations more feasible while at the same time retaining the essential physics of continuum chiral symmetry.

The plan of this paper is as follows. Section II presents a study of various free-fermion lattice field theories which avoid the doubling problem. Section III studies local current-current correlation functions in these free field theories and shows why problems of rotational noninvariance arise in Green's functions for short-range derivatives and how they can be avoided if the range of the derivative is taken to infinity. In Sec. IV we turn to a variant of noncompact lattice QED which shows that a *gauged* version of the long-range derivative can be introduced in a way which explicitly reproduces standard weak-coupling perturbation theory. (This section essentially reproduces results of Rabin,[3] although the approach is somewhat more intuitive.) In Sec. V we turn to the compact theory and suggest a gauging which preserves the good results of Sec. IV. This generalization to the compact case is important because it suggests that a comparable treatment for non-Abelian gauge theories can be considered. Finally in Sec. VI we review our results and discuss some open questions.

II. FREE FIELDS

If one attempts to formulate a Hamiltonian theory of free fermions using the nearest-neighbor lattice derivative one finds that, for a theory in d spatial dimensions, there are 2^d times as many states at a given energy as one expects in the continuum limit. This repetitive *doubling* of the spectrum as one goes up in spatial dimension seems to be a quite general problem. The theorem of Nielsen and Ninomiya[1] has been discussed in several publications; it is widely interpreted to show that any finite-range lattice derivative suffers from this *spectrum doubling*. While the theorem is correct, spectrum doubling is a misnomer for the phenomenon which occurs. We will demonstrate that it is simple to write Hamiltonian *free field* theories based on finite-range derivatives which do not exhibit spectrum doubling and which have entirely sensible continuum spectra.

A. Notation

In what follows we consider a lattice whose points are labeled by d-tuples of integers $\mathbf{j}=(j_1, \ldots, j_d)$ where d is the dimension of space. The physical spacing between adjacent points is given by the parameter a, a quantity with the dimension of length. In addition, we assume that the integers j run in the range $-N \leq j_m \leq N$ for

$m = 1, \ldots, d$, so that our theory is set on a lattice of finite physical volume, L^3, with $L = (2N + 1)a$.

We use dimensionless variables, except where otherwise noted. Since the physical Hamiltonian H has units of energy, we introduce the dimensionless Hamiltonian $\mathscr{H}$ defined by

$$H = \frac{1}{a}\mathscr{H} .$$

In addition all fields are scaled by the appropriate powers of a so as to render them dimensionless and yield simple lattice commutation and anticommutation relations.

In particular, lattice fermion fields are denoted $\psi(\mathbf{j})$ with

$$\{\psi^{\dagger}(\mathbf{j}'),\psi(\mathbf{j})\} = \prod_{m=1}^{d} \delta_{j_m,j_m'} . \tag{2.1}$$

The lattice Fourier transforms of the fields $\psi(\mathbf{j})$ are given by

$$\psi(\kappa_p) = \frac{1}{(2N+1)^{d/2}} \sum_{\mathbf{j}} e^{-i\mathbf{k}\cdot\mathbf{j}}\psi(\mathbf{j}) , \tag{2.2}$$

where $\kappa_p = (\kappa_{p_1}, \ldots, \kappa_{p_d})$ are the dimensionless lattice momenta $\kappa_n = 2\pi n/(2N+1)$, for $-N \leq n \leq N$.

The relationship between dimensionless variables and their dimensioned counterparts is

$$\mathbf{k}_p = a^{-1}\kappa_p , \tag{2.3}$$

$$\Psi(\mathbf{k}_p) = a^{-d/2}\psi(\kappa_p) . \tag{2.4}$$

The operators which converge weakly to the continuum field $\Psi(\mathbf{x})$ are

$$\Psi(\mathbf{x}) = L^{-d/2} \sum_{\mathbf{k}_p} e^{i\mathbf{k}_p\cdot\mathbf{x}}\psi(\mathbf{k}_p) , \tag{2.5}$$

where $L = (2N+1)a$.

We are careful to work in finite volume because this allows us to count energy levels and thus explicitly check whether or not *spectrum doubling* occurs.

B. 1 + 1 dimensions: Short-range derivatives

Let us begin by studying the simplest case, namely, 1 + 1 dimensions. In this case we use two-component Dirac fields. Consider the Hamiltonian

$$\mathscr{H}(\mu) = -i \sum_{\kappa_p} \left\{ \frac{(1+\mu)}{2\mu}[\psi^{\dagger}(j)\sigma_z\psi(j+1)] \right.$$

$$\left. + \frac{(\mu-1)}{4\mu}[\psi^{\dagger}(j)\sigma_z\psi(j+2)] \right\} + \text{H.c.} , \tag{2.6}$$

where we have adopted periodic boundary conditions; i.e., we have implicitly assumed that $\psi(N+1) = \psi(-N)$. Since (2.6) is translationally invariant it is diagonal in terms of the Fourier transforms of the $\psi(j)$ and has the form

$$\mathscr{H} = \sum_{\kappa_p} \mathscr{E}(\kappa_p)[\psi^{\dagger}(\kappa_p)\sigma_z\psi(\kappa_p)] , \tag{2.7}$$

where $\mathscr{E}(\kappa_p)$ is defined by

$$\mathscr{E}(\kappa_p) = \frac{(1+\mu)}{2\mu}\sin(\kappa_p) + \frac{(\mu-1)}{4\mu}\sin(2\kappa_p) \ . \qquad (2.8)$$

Note that the minimum and maximum values of κ_p are $\pm 2\pi N/(2N+1)$ and κ_p changes in increments of $1/(2N+1)$.

First, let us examine the nearest-neighbor case, $\mu=1$. Figure 1(a) shows the spectrum for this case. In order to
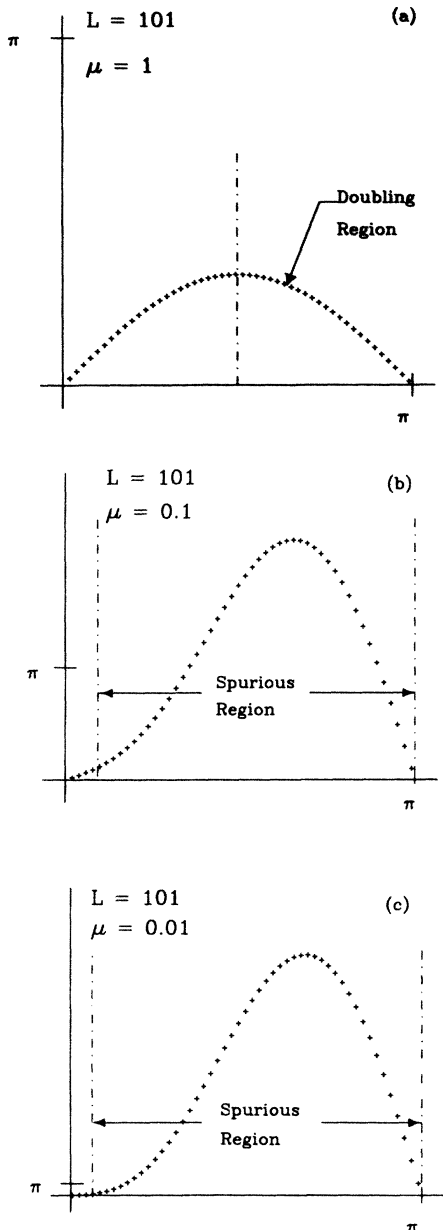

FIG. 1. (a) The spectrum of the nearest-neighbor derivative ($\mu=1$) showing true spectrum doubling ($L=2N+1=$number of sites). (b) The spectrum of the next-nearest-neighbor derivative for $\mu=0.1$. Note the larger slope and correspondingly lower density of states for κ near π. The energy of the last state is approximately $\pi/\mu L$. (c) The spectrum of the next-nearest-neighbor derivative for $\mu=0.01$. (Note the change of vertical scale.)

more clearly exhibit the behavior of all derivatives considered we only show the plots of $\mathscr{E}(\kappa)$ for the region $0 \leq \kappa \leq \pi$ since $\mathscr{E}(\kappa)$ is always an antisymmetric function of κ. The important thing to notice about the nearest-neighbor case is that for every low-lying state near $\kappa_p=0$ there is a degenerate state near $\kappa_p=\pi$. This is true spectrum doubling. Figures 1(a) and 1(b) which plot the same function for $\mu \neq 1$ demonstrate that this is not a general feature of a short-range Hamiltonian. Here, for $\mu=0.1$, 0.01, we see that the energy as a function of κ is not symmetric about $\kappa=\pi/2$.

To understand the important features of this spectrum expand (2.8) for κ near zero and π. Expanding in κ_p for $|p| \ll N$ we obtain

$$\mathscr{E}(\kappa_p) \approx \kappa_p + \frac{1}{12\mu}(5-3\mu)\kappa_p^3 + O(\kappa_p^5) \ , \qquad (2.9)$$

whereas for κ_{N-p} and $|p| \ll N$ we have

$$\mathscr{E}(\pi-\delta_p) \approx \frac{\delta_p}{\mu} + O\left[\frac{\delta_p^3}{\mu}\right] \ , \qquad (2.10)$$

where δ_p is defined to be

$$\delta_p = \frac{\pi(2p+1)}{2N+1} \ . \qquad (2.11)$$

If we restore dimensions of energy to the problem by defining

$$k_p = \frac{1}{a}\kappa_p \qquad (2.12)$$

then we see that the physical energy $\mathscr{E}(\kappa_p)/a$ for k_p near zero is

$$\frac{1}{a}\mathscr{E}(k_p a) = k_p + O\left[\frac{a^2 k_p^3}{\mu}\right] \qquad (2.13)$$

and for k_p near π/a

$$\frac{1}{a}\mathscr{E}(\pi-\delta_p) = \mu^{-1}\frac{\delta_p}{a} + O\left[a^2\mu^{-1}\frac{\delta_p^3}{a}\right] \qquad (2.14)$$

since $(2N+1)a=L$, Eq. (2.14) implies that the lowest state for $\kappa \to \pi$ occurs at an energy $E_{\min} = \pi/\mu L$. Hence, if we take $\pi/\mu L \to \infty$ as $L \to \infty$ the spurious states move off to infinite energy, and the only states which remain at finite energy are a single species of chiral fermions with an ordinary relativistic dispersion law for their energy as a function of momentum.

At this juncture it behooves us to ask "What has happened to the Nielsen-Ninomiya theorem?" Despite appearances to the contrary the formal result stated in their theorem has not been violated. What the theorem really says is that the function $\mathscr{E}(\kappa)$ is a periodic function with period 2π when considered as a function of a continuous variable $-\infty \leq \kappa \leq \infty$. The function (2.8) clearly satisfies this property. However, in finite volume κ is not a continuous variable. The continuum result depends upon how the limits $L \to \infty$ and $\mu \to 0$ are taken. Notice that even if $L \to \infty$ first, so that κ becomes a continuous variable, there is no true spectrum doubling. This is because the density of states for the branch $\kappa \approx \pi$ is μ, whereas it

is unity for $\kappa \approx 0$. What one has are two species of particles with very different speeds of light.

Having established that there are loopholes in the Nielsen-Ninomiya theorem we now hasten to point out that this does not mean we are out of the woods. While the next-nearest-neighbor theory successfully defines an undoubled theory of free chiral fermions, it does not provide a satisfactory interacting theory for $d > 1$. The reason for this is that once interactions are introduced one has to do loop integrations (i.e., sums over intermediate states). Although the contributions to loop integrals coming from the states corresponding to $\kappa \approx \pi$ are damped by energy denominators which are on the order of $\pi/\mu L$, there may be so many states at this energy that even taking the energy denominators into account the sum over these states diverges. When this happens, as it does in more than $1 + 1$ dimensions, it can yield noncovariant contributions to Green's functions which survive even after the usual continuum subtractions have been performed.

The real problem with short-range derivatives has nothing to do with doubling of the low-energy spectrum but rather with the number of states in the region having energies on the order of π/a. For the next-nearest-neighbor case one sees clearly from Fig. 1 that there are, in the limit $\mu \to 0$ many more states at infinite energy than at finite energy; moreover, these states have features which are intrinsically not rotationally invariant.[5] To establish this fact observe that for the Hamiltonian (2.6) the region of the spectrum which is linear in κ corresponds to $|\kappa| < \sqrt{\mu}$. The remaining spectrum, which we refer to as the *spurious region*, is all at high energy in the limit $a \to 0$ and $\mu L \to 0$, and the number of states in this region is of order $(\pi - \sqrt{\mu})L/a$. It follows that if one takes the continuum limit to mean $a \to 0$ and $\mu L \to 0$, then the majority of the states will lie in the spurious region.

In Sec. III we will exhibit the exact nature of the noninvariant contributions to current-current correlation functions in more than $2 + 1$ dimensions and show that these spurious states do indeed cause a problem. Furthermore, such effects are unavoidable for any finite-range derivative. While it is true that the width of the region in which $\mathscr{E}$ is a linear function of κ can be increased somewhat by taking additional terms, it is only by including terms of infinite range that we can make the spurious region shrink rather than grow as $\mu \to 0$. We will show that this shrinkage of the spurious region as $\mu \to 0$ is the crucial ingredient in obtaining a satisfactory continuum limit.

C. Longer-range derivatives

One way to study the width of the region in which the spurious states contribute is to adopt a damped version of the SLAC derivative. Recall that the SLAC formulation of the chiral fermion theory gives a Hamiltonian of the form

$$\mathscr{H} = i \sum_{j,l} \psi^\dagger(j)\sigma_3\psi(l)\delta'(j-l) \, , \tag{2.15}$$

where the function $\delta'(j-l)$ is defined to be

$$\delta'(j-l) = \frac{1}{\sqrt{2N+1}} \sum_{\kappa_p} i\kappa_p e^{i\kappa_p(j-l)} \, . \tag{2.16}$$

In the limit $N \to \infty$ this function has infinite range, falling off like $(-1)^j/j$. For finite N, note that

$$\delta'(j-l+2N+1) = \delta'(j-l) \tag{2.17}$$

which means that the SLAC derivative describes fermions on a periodic lattice, i.e., on a ring containing $2N+1$ points. We observe that (2.15) can be rewritten as a sum of terms with range $1 \le r \le N$:

$$\mathscr{H} = -i \sum_{j=-N}^{N} \sum_{r=1}^{N} \psi^\dagger(j)\sigma_3\psi(j+r)\delta'(r) + \text{H.c.} \tag{2.18}$$

As in the nearest-neighbor case we have implicitly assumed that $\psi(j+r) = \psi(j')$ where j' is the integer in the range $-N \le j' \le N$ obtained by reducing $(j+r)$ modulo $(2N+1)$.

We will now introduce a damped form of this Hamiltonian:

$$\mathscr{H}(\mu) = -i \sum_{j=-N}^{N} \sum_{r=1}^{N} \psi^\dagger(j)\sigma_3\psi(j+r)e^{-\mu r}\delta'(r) + \text{H.c.}$$

$$\tag{2.19}$$

As before, $\mathscr{H}(\mu)$ is diagonal in κ_p with an energy-momentum dispersion given by a function $\mathscr{E}(\kappa_p)$, of the form shown in Fig. 2 for various values of μ. The important difference between this case and the next-nearest-neighbor case is that the width of the spurious region is only on the order of $\sqrt{\mu}$, and so in the limit $\mu \to 0$ the region of physical momenta with a dispersion $E(k_p) = k_p$ grows instead of shrinking. Eventually, for $\mu \to 0$, the dispersion law becomes the correct relativistic formula for massless particles in $1 + 1$ dimensions.[6,7]

D. Higher dimensions and mass terms

The generalization of the preceding formulas to $d + 1$ dimensions, with fermion mass terms, is straightforward. The generic Hamiltonian is of the form

$$\mathscr{H} = \sum_{j,p,\nu} -i[\psi^\dagger(j+r\hat{n}_\nu)\alpha_\nu\psi(j)D_\nu(r) - \text{H.c.}]$$

$$+ m \sum_j \psi^\dagger(j)\beta\psi(j) \, , \tag{2.20}$$

where m stands for the dimensionless mass $m = ma$. The functions $D_\nu(r)$ for $\nu = 1, \ldots, d$ can be chosen either as in (2.6) or (2.19). The matrices α and β are $\alpha_\nu = \gamma_0\gamma_\nu$ and $\beta = \gamma_0$, where the γ's are the Dirac matrices for $d + 1$ dimensions. Because we wish to study QED perturbation theory we consider four component massive fermions. Clearly, this formulation can also be used for truly chiral gauge theories by introducing the projections $(1 \pm \gamma_5)/2$ and setting $m = 0$. The chiral symmetry is manifest in the $m = 0$ case and so the restriction to purely left-handed fermions is not a problem.

The Hamiltonian (2.20) is easily diagonalized in momentum space. The energy-momentum dispersion relation is

$$\mathscr{E}(\kappa_1, \ldots, \kappa_d) = \pm[\tilde{D}_1(\kappa_1)^2 + \cdots + \tilde{D}_d(\kappa_d)^2 + m^2]^{1/2} \, ,$$
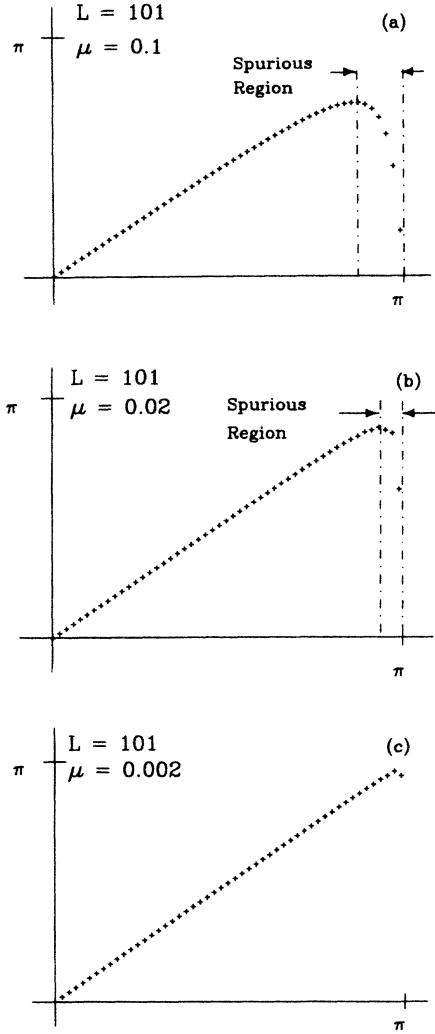
$$\tag{2.21}$$

FIG. 2. (a) The spectrum of the damped SLAC derivative for the case $\mu = 0.01$ ($L = 2N + 1 =$ number of sites). (b) The spectrum of the damped SLAC derivative for the case $\mu = 0.002$. Note the shrinkage of the spurious region. (c) The spectrum of the damped SLAC derivative for the case $\mu = 0.002$.

where $\tilde{D}_l(\kappa_l)$ is the discrete Fourier transform of the function $D_l(r)$.

The 2^d doubling of the nearest-neighbor theory has been lifted in (2.21). In the region of momenta where all components κ_ν are close to zero (2.21) describes the spectrum of a massive fermion with speed of light 1. When the form (2.6) is used for D the states for which all components κ_ν are near π look like a second species which propagates with speed of light $1/\mu$. However, there are $2^d - 2$ regions where some components of κ are near zero and others near π which do not correspond to any sort of relativistic spectrum. For example, for κ_x and κ_y near zero but $\kappa_z = \pi/a - \hat{\kappa}_z$ we have a low-lying region of the spectrum which behaves approximately as

$$\mathscr{E} \approx \left[\kappa_x^2 + \kappa_y^2 + \frac{\hat{\kappa}_z^2}{\mu^2} + m^2 \right]^{1/2} \tag{2.22}$$

which can at best be interpreted as a species of fermion which propagates freely in the x and y directions but pays a price of at least $\pi/\mu L$ for motion in the z direction. It is the existence of such intrinsically noncovariant regions of the spectrum which leads to noncovariant behavior of Green's functions. While the discrete rotational invariance of the lattice is retained when one sums over the contributions from all such regions, we find they lead to terms of the form $\sum_i q_i^2 \ln q_i^2$, which are not invariant with respect to small rotations even in the limit $a \to 0$.

For the damped SLAC derivative, (2.19), we will show that no such terms survive, provided we choose $\mu = O(1/N^2)$ in the limit $N \to \infty$.

III. CURRENT-CURRENT CORRELATION FUNCTIONS

In general, for a free-field theory there is an ambiguity as to which lattice operators should be chosen to represent the continuum currents $\Psi^\dagger(x)\gamma^\mu\Psi(x)$. We will choose an operator whose Green's functions mimic the divergence structure usually encountered in the discussion of weak-coupling lattice perturbation theory. The operator we study is the charge-density operator

$$\rho(j) = \frac{1}{2}[\psi^\dagger(j), \psi(j)] , \tag{3.1}$$

which has the Fourier transform

$$\rho(\kappa_p) = \frac{1}{2} \sum_{\kappa_l, \kappa_m, n} \delta(\kappa_p - \kappa_l - \kappa_m + 2\pi n)$$

$$\times [\psi^\dagger(k_l), \psi(\kappa_m)] , \tag{3.2}$$

where $n = (n_1, \ldots, n_d)$ is a vector whose components are integers. The operator $\rho(j)$ is the time component of a four-vector operator which satisfies the lattice equation

$$\frac{\partial \rho}{\partial t}(i) - \sum_{\alpha,p} \delta_\alpha(p\hat{n}_\alpha)j_\alpha(i + p\hat{n}_\alpha) = 0 , \tag{3.3}$$

where $\alpha = 1, \ldots, d$ and p is such that $-N \leq i_\alpha + p \leq N$. The spatial components of j are intrinsically nonlocal. (Explicit expressions for them are given in Ref. 8; however, since it is sufficient to study the behavior of the time-time components of the current we will not need them here.) The correlation functions of interest will be the time-ordered products of the free field currents:

$$\Delta_{\rho\rho} = \langle T(\rho(j',t')\rho(j,t)) \rangle , \tag{3.4}$$

where the operator $\rho(j,t)$ is defined by

$$\rho(j,t) = e^{-i\mathscr{H}t}\rho(j)e^{i\mathscr{H}t} . \tag{3.5}$$

As defined, the operators $\rho(j,t)$ are normal-ordered products of free fields, and so have vanishing vacuum expectation values. Such we are working in a Hamiltonian formalism these Green's functions can be calculated using familiar Feynman rules with the following exceptions.

(1) Time is continuous, so κ_0 integrations run from $-\infty$ to ∞.

(2) Space is discrete and finite, so momentum sums range over the discrete variables $\kappa_p = 2\pi p /(2N+1)$ for $-N \leq p \leq N$.

(3) The free-fermion propagator is given by

$$\frac{1}{\widetilde{\not{D}}(\kappa_0,\kappa_p)-m} \; , \tag{3.6}$$

where

$$\widetilde{\not{D}}(\kappa_0,\kappa_p)=\gamma^0\kappa_0+\sum_\nu \gamma^\nu \widetilde{D}_\nu(\kappa_\nu) \; , \tag{3.7}$$

where $\widetilde{D}_\nu(\kappa_\nu)$ is the general function appearing in the def-

inition of the Hamiltonian (2.20).

(4) The overall momentum-conserving δ functions at each vertex are ordinary δ functions of the zero components of momenta, but *periodic* δ functions for spatial components of momenta.

A. The problem with loops

We want to examine the Fourier transform of the time-ordered product of two charge densities. (See Fig. 3.) For d spatial dimensions the unsubtracted expression for the charge-density—charge-density correlation function is given by

$$\Delta^d_{p,p}(\mathbf{q},q_0)=\frac{1}{(2n+1)^d a^d}\int dk_0 \sum_{p,r,n} \delta(\mathbf{r}-\mathbf{p}-\mathbf{s}-(2N+1)\mathbf{n})I(\kappa_p,\kappa_r,\kappa_s,k_0,q_0) \; , \tag{3.8}$$

where

$$\kappa_p = \frac{2\pi\mathbf{p}}{2N+1}, \quad \kappa_r = \frac{2\pi\mathbf{r}}{2N+1}, \quad qa=\kappa_s=\frac{2\pi\mathbf{s}}{2N+1} \tag{3.9}$$

and

$$I = \mathrm{Tr}\left[\gamma^0 \frac{1}{(q_0+k_0)\gamma^0+a^{-1}\gamma_i\widetilde{D}_i(\kappa_r)-m} \times \gamma^0 \frac{1}{k_0\gamma^0+a^{-1}\gamma_i\widetilde{D}_i(\kappa_p)-m}\right] . \tag{3.10}$$

Our problem is to estimate the contributions of the spurious regions to (3.8) in the limit $a \to 0$. The detailed discussion of this calculation, while straightforward, is tedious and is relegated to the appendixes. Here we will simply summarize results and note certain salient features.

For the case of $1+1$ dimensions the expression corresponding to (3.8)—(3.10) is naively logarithmically divergent in the continuum limit; however, current conservation guarantees the vanishing of that divergence and no subtractions are needed. Table I lists the dependence of the most singular contributions from various regions of the κ sums on the parameters μ, L, and a for the case of the next-nearest-neighbor derivative (2.6) for $0 < q < \sqrt{\mu}/a$. If we take the limit
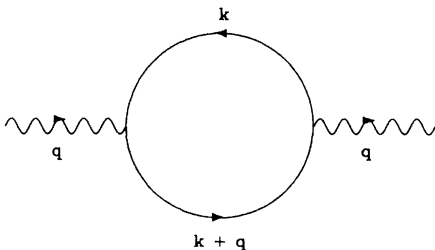
$$\mu = \delta \left\lfloor \frac{a}{L} \right\rfloor = \frac{s}{2N+1} \; ,$$

$$a \to 0, \quad L \to \infty \; , \tag{3.11}$$

i.e., $N \to \infty$ with $\delta = \mathrm{const}$,

then all spurious contributions vanish, as required. It is clear also that the restriction $|q| < \sqrt{\mu}/a$ is satisfied for all finite-physical external momenta in the limit (3.11), and so the inequality holds for all cases of interest. The results for negative q are similar.

For the damped SLAC derivative we see in Fig. 2 that the function $\widetilde{D}_\mu(\kappa)$ can be well approximated by the linear behavior $\widetilde{D}(\kappa)=\kappa$ up to κ of order $\pi(1-\sqrt{\mu})$. The analytic behavior of the function in the spurious region beyond this point is complicated; however we show in Appendix A that the minimum value of $\widetilde{D}(\kappa)$ in this region occurs at $\kappa_N=2\pi N/(2N+1)$, where it is bounded by

$$|\widetilde{D}(\kappa_N)| \geq \pm \frac{2\pi N}{2N+1}[1-O(\mu)] \; . \tag{3.12}$$

FIG. 3. Fermion-loop graph encountered in the computation of the current-current correlation function.

TABLE I. Contributions to the current-current correlation function in $1+1$ dimensions given by the next-nearest-neighbor derivative for $Q < qa < \sqrt{\mu}$. The contributions listed are the dominant terms in the limit $a \to 0$, $L \to \infty$ with $\mu=(\delta a /L)$ and δ held fixed.

Region	Contribution
$-\left[\dfrac{\mu}{a}\right]^{1/2} < k < \left[\dfrac{\mu}{a}\right]^{1/2}$	Usual continuum result $+O\left[\left[\dfrac{qa}{\sqrt{\mu}}\right]^2\right]$
$-\dfrac{\pi}{a}+\dfrac{\pi}{L} < k < -\dfrac{\sqrt{\mu}}{a}$	$O\left[\left[\dfrac{qa}{\sqrt{\mu}}\right]^2\right]+O(\mu)$
$\left[\dfrac{\mu}{a}\right]^{1/2} < k < \dfrac{\pi}{a}-\dfrac{\pi}{L}-q$	from $\sqrt{\mu}/a$ limit
$\dfrac{\pi}{a}-\dfrac{\pi}{L}-q < k < \dfrac{\pi}{a}-\dfrac{\pi}{L}$	$O(\mu)$

Thus, using this derivative in $1 + 1$ dimensions, we can also obtain satisfactory results in the limit (3.11) provided $\tilde{\delta} \lesssim 1$.

In higher dimensions the current-current correlation function is divergent and requires subtraction. In the spurious regions for the next-nearest-neighbor derivative the slope of the function $\tilde{D}(\kappa)$ is of order $1/\mu$; therefore, each subtraction introduces a factor (qa/μ). These inverse powers of μ are dangerous; they can serve to enhance the spurious contributions so that no satisfactory choice of limit can be found. The most singular contributions come from the regions where one component of κ, say κ_i, is in the spurious region near π, and the remaining components are in the linear region near $\kappa = 0$. In this case, for the next-nearest-neighbor derivative, we obtain contributions to the Green's function which behave like

$$\int d\alpha\, \alpha(1-\alpha)\frac{q_i{}^2}{\mu}\ln[1+(1-\alpha)q_iL] \ . \tag{3.13}$$

Clearly, there is no way to take the limit $\mu \to 0$ so these contributions disappear.

How can this problem be avoided? It may appear that the damped SLAC derivative suffers from a similar problem; however, one must be careful. In this case the spurious region has a width of order $\sqrt{\mu}$. Thus we can choose to take the limit

$$\mu = \delta \left[\frac{a}{L}\right]^2 \quad \text{for } a \to 0$$

$$\text{and } L \to \infty \ \text{ with } \delta = \text{const} \ . \tag{3.14}$$

In this case the number of states in the spurious region remains finite as $a \to 0$ and so their contributions to Green's functions must be treated as discrete sums rather than as integrals. Furthermore, as can be seen in Fig. 2 the slope of (κ) is never of order $1/\mu$ in the allowed region of κ. Instead, the function develops a discontinuity at $\kappa = \pi$ which must be carefully treated. Appendix B details our analysis of the $(3 + 1)$-dimensional current-current correlation function, including all of the subtraction terms required by continuum renormalization. Table II defines the regions of momentum space which must be treated separately and Table III shows the dominant behavior of

contributions from various regions. (All other contributions can be found by exploiting the symmetry of the theory under discrete lattice rotations.) In the limit (3.14) only the usual contributions survive.

In this section we have examined the quantity $\Delta_{\rho\rho}$ using a definition of ρ such that the coupling to all fermion modes is the same. In the interacting theory in $A_0 = 0$ gauge we will be interested in the quantities such as π_{ii}, which are calculated in the same way as given here for $\Delta_{\rho\rho}$ except that the numerator factors of γ_0 in (3.10) must be replaced by the $\Gamma_i a_i$, the vertex functions for the photons in the perturbation theory.

IV. NONCOMPACT QED

The previous section proves there exists a current whose correlation functions, calculated using a damped SLAC derivative, behave satisfactorily in the limit $a \to 0$. Yet Karsten and Smit found unsatisfactory results for the SLAC derivative in weak-coupling perturbation theory. How can we reconcile these seemingly contradictory results? The answer is that the current whose correlation functions are well behaved is not the one to which the photon couples in the version of lattice QED considered by Karsten and Smit. The current which we discussed couples equally to all κ modes of the fermion, whereas the vertex obtained by Karsten and Smit behaves like

$$\Gamma_\mu(k,q) = g\gamma_\mu \left[\frac{\tilde{D}_\mu(k+q) - \tilde{D}_\mu(k)}{\sin(q_\mu{}^2 a/2)} \right] \ . \tag{4.1}$$

Equation (4.1) describes a coupling to photons which is proportional to the slope of $\tilde{D}(k)$ and so, in the limit $\mu \to 0$, the coupling to fermionic states with $k+q > \pi/a$ but $\kappa < \pi/a$ becomes arbitrarily strong. Thus, the contributions from the spurious regions to loop integrals are enhanced by factors of order $1/\mu$ and survive the limit $a \to 0$, $\mu \to 0$. It is this behavior and not spectrum doubling which lies at the root of the problems first discussed by Karsten and Smit.

As with spectrum doubling, there is a way out of this problem. It is suggested by the work of Rabin, who gave

TABLE II. Regions of momentum space treated separately in calculating the current-current correlation function in $3 + 1$ dimensions using the damped SLAC derivative for $\pi\sqrt{\mu} < qa$.

	Label		Region	Number of states
S	Spurious	$n = 0$	$-\frac{\pi}{a} + \frac{\pi}{L} \leq k \leq -\frac{\pi}{a} + \frac{\pi\sqrt{\mu}}{a}$	$\delta - 1$
N	Normal	$n = 0$	$-\frac{\pi}{a} + \frac{\pi\sqrt{\mu}}{a} < k \leq \frac{\pi}{a}(1 - \sqrt{\mu}) - q$	$O(L/a)$
S	Spurious	$n = 0$	$-\frac{\pi}{a}(1 - \sqrt{\mu}) - q < k \leq \frac{\pi}{a} - \frac{\pi}{L} - q$	$\delta - 1$
S	Spurious	$n = -1$	$\frac{\pi}{a} + \frac{\pi}{L} - q \leq k \leq \frac{\pi}{a}(1 + \sqrt{\mu}) - q$	$\delta - 1$
U	Umklapp (normal)	$n = -1$	$\frac{\pi}{a}(1 + \sqrt{\mu}) - q < k \leq \frac{\pi}{a}(1 - \sqrt{\mu})$	(qL/π)
S	Spurious	$n = -1$	$\frac{\pi}{a}(1 - \sqrt{\mu}) < k \leq \frac{\pi}{a} - \frac{\pi}{L}$	$\delta - 1$

TABLE III. Dominant contributions to the current-current correlation function in 3 + 1 dimensions (three times subtracted) given by the damped SLAC derivative in the limit $a \to 0$, $L \to \infty$ with $\mu = \delta(a/L)^2$, and δ held fixed.

Type of region			
k_x	k_y	k_z	Order of dominant contribution
N	N	N	$q^2 \int d\alpha\, \alpha(1-\alpha)\ln\{[\alpha(1-\alpha)q^2+m^2]/m^2\}$
N	N	U	$q^3 a$
N	U	U	$q^3 a$
U	U	U	$q^3 a$
N	N	S	$\sqrt{\mu}\, q^3 a$
N	S	S	$\mu q^3 a$
S	S	S	$(\mu)^{3/2} q^3 a$
N	S	U	$\sqrt{\mu}\, q^3 a$
S	S	U	$\sqrt{\mu}\, q^3 a$
S	U	U	$\mu q^3 a$

an answer which is applicable only to the case of noncompact QED. Let us review the substance of Rabin's solution and then turn to a generalization of it which is more suitable for a compact gauge theory and thus for generalization to a non-Abelian theory.

It is helpful to begin with a review of some elementary properties of noncompact lattice QED. The Hamiltonian version of noncompact QED is defined in terms of link fields A_l and their conjugate variables E_l which satisfy the equal-time commutation relations

$$[E_l, A_{l'}] = -i\delta_{l,l'} , \tag{4.2}$$

where l and l' stand for two links on the lattice. For the pure gauge theory the Hamiltonian can be chosen to have one of two forms. The first form of the Hamiltonian is the one which trivially reproduces the ordinary free-field theory of photons simply transcribed to a lattice:

$$H = \tfrac{1}{2}\sum_l E_l{}^2 + \tfrac{1}{2}\sum_{\mathscr{P}} B_{\mathscr{P}}{}^2 . \tag{4.3}$$

Here, l denotes a link variable and $\mathscr{P}$ denotes a plaquette variable; i.e., E_l is the electric field variable associated with the link l, and $B_{\mathscr{P}}$ is the magnetic variable associated with the plaquette $\mathscr{P}$. If we denote the links bounding the plaquette $\mathscr{P}$ by l_1, l_2, l_3, and l_4, respectively, then $B_{\mathscr{P}} = A_{l_1} + A_{l_2} - A_{l_3} - A_{l_4}$. The second form of the Hamiltonian is more complicated and produces a large number of slightly different[9] copies of compact QED; i.e.,

$$H = \sum_l \frac{g^2}{2} E_l{}^2 - \sum_{\mathscr{P}} \frac{1}{2g^2} \cos(B_{\mathscr{P}}) . \tag{4.4}$$

The important difference between (4.3) and (4.4) is that (4.4) defines a theory of interacting degrees of freedom for $g^2 \neq 0$ whereas (4.3) is a free field theory. For this reason the theory based upon (4.4) has some of the features of a non-Abelian theory (at least for $g^2 \gg 1$).

Both theories have an invariance under time independent gauge transformations wherein

$$\delta E_l = 0, \quad \delta A_l = \nabla_l \lambda , \tag{4.5}$$

where $\lambda(\mathbf{j})$ is a function defined on points on the lattice.

If l is the link joining the points $\mathbf{j}$ and $\mathbf{j'} = \mathbf{j} + \hat{\mathbf{u}}$ then $\nabla_l \lambda = \lambda(\mathbf{j'}) - \lambda(\mathbf{j})$. The generator of this transformation is the unitary operator

$$U(\lambda) = \exp\left[i \sum_l E_l \nabla_l \lambda \right] \tag{4.6}$$

which can be rewritten, using the lattice analogue of integration by parts, as

$$U(\lambda) = \exp\left[-i \sum_{\mathbf{j}} (\nabla \cdot E)(\mathbf{j})\lambda(\mathbf{j}) \right] . \tag{4.7}$$

The lattice divergence of the link variables E_l is defined as the directed sum of E_l's over the links touching the point $\mathbf{j}$. Equation (4.7) shows that $\nabla \cdot E(\mathbf{j})$ can be identified as the generator of local, time-independent gauge transformations. From this it follows that imposing the Maxwell equation

$$\nabla \cdot E(\mathbf{j}) = 0 \tag{4.8}$$

is equivalent to the statement that physical states are invariant under arbitrary gauge transformations. The generalization of this result to include charged matter fields is straightforward. In this case a general time-independent gauge transformation not only shifts the fields A_l by the gradient of a function $\lambda(\mathbf{j})$, but it also multiplies the matter fields by a phase factor $e^{iq\lambda(\mathbf{j})}$. If we let $j_0(\mathbf{j})$ be the time component of the conserved electromagnetic current, then an arbitrary gauge transformation is generated by the unitary transformation

$$U(\lambda) = \exp\left[i \sum_{\mathbf{j}} [-\nabla \cdot E(\mathbf{j}) + j_0(\mathbf{j})]\lambda(\mathbf{j}) \right] . \tag{4.9}$$

We identify $-\nabla \cdot E(\mathbf{j}) + j_0(\mathbf{j})$ as the generator of arbitrary gauge transformations, and once again the statement that physical states are gauge invariant corresponds to imposing as a state condition that Maxwell equation which is lost when quantizing in $A_0 = 0$ gauge.

The next question which arises is "What does the pure gauge theory Hamiltonian look like when one restricts attention to the sector of gauge-invariant states?" One way to answer this question is to observe that even on a lattice the operators E_l and A_l can be written as a sum of a gradient and a curl; i.e.,

$$E_l = E_l^L + E_l^T , \tag{4.10}$$

where E_l^L is a function whose lattice curl vanishes and E_l^T is a function whose divergence vanishes. (For details of this decomposition see Ref. 9 and papers cited therein.) Obviously, a similar decomposition holds for A_l. Furthermore, it is clear that the transverse part of A_l must be a function of $B_{\mathscr{P}}$ alone.

With this decomposition in hand it is straightforward to write a Hamiltonian for noncompact QED which has a well-behaved continuum limit:

$$H = H_0 - i \sum_{r,\mu} \Psi^\dagger(\mathbf{j}) a_\mu \Psi(\mathbf{j} + r\hat{\mu}) D_\mu(r) \exp\left[ig \sum_{l'} A_{l'}^L \right]$$

$$+ H_T ; \tag{4.11}$$

$\sum_l'$ stands for the set of links along the line joining $\mathbf{j}$ and $\mathbf{j} + r\hat{\mu}$, where

$$H_0 = \frac{g^2}{2} \sum_l (E_l^{L^2} + E_l^{T^2}) - \frac{1}{2g^2} \sum_{\mathscr{P}} \cos(B_{\mathscr{P}}) \tag{4.12}$$

and H_T is given by

$$H_T = \sum_{\mathbf{j}+\hat{\mu}} \Psi^\dagger(\mathbf{j}) e^{igA_l^T} \Psi(\mathbf{j}+\hat{\mu})(e^{igA_l^T} - 1) . \tag{4.13}$$

It is now easy to answer the question just posed: "What does the Hamiltonian look like when projected into the sector of gauge-invariant states?" Since a gauge-invariant state is one for which the equation

$$\nabla \cdot E^L(\mathbf{j}) - j_0(\mathbf{j}) \mid \phi \rangle = 0 \tag{4.14}$$

is satisfied and E^L, by definition, can be written as

$$E_l^L = -\nabla_l \lambda , \tag{4.15}$$

for gauge-invariant states

$$E_l^L = -\frac{1}{\nabla^2} \nabla j_0 \tag{4.16}$$

so we can rewrite (4.12) as

$$H_0 = -\frac{g^2}{2} \sum_{\mathbf{j},\mathbf{j}'} j_0(\mathbf{j}) \frac{1}{\nabla^2} j_0(\mathbf{j}')$$

$$+ \frac{g^2}{2} \sum_l E_l^{T^2} - \frac{1}{2g^2} \sum_{\mathscr{P}} \cos(B_{\mathscr{P}}) . \tag{4.17}$$

The first term is easily recognized to be the Coulomb interaction between two charges at the sites $\mathbf{j}$ and $\mathbf{j}'$ (a quantity which rapidly passes to its continuum form) and the remaining terms in the H_0 are explicitly gauge invariant. As for the remaining terms in the Hamiltonian (4.11) we see that the factor of e^{iA^L} can be dropped since the only effect of this operator is to change the eigenvalue of E^L

for a gauge-invariant state and we have now chosen to rewrite this change completely in terms of the change in the charge density $j_0(\mathbf{j})$ which is caused by the operators $\Psi^\dagger(\mathbf{j})$ and $\Psi(\mathbf{j})$. Hence, without ever fixing the gauge, beyond making the choice $A_0 = 0$, we have obtained the Coulomb-gauge form of the Hamiltonian *in the sector of physical states*. The important features of this Hamiltonian are the following.

(1) The free-fermion part of the Hamiltonian is arbitrary; in particular, $D_\mu(r)$ can be taken to be the damped SLAC derivative.

(2) The coupling to transverse photons, which together with the Coulomb interaction constitutes the entire interaction Hamiltonian, is strictly local on the lattice; that is, it is independent of the form of $D_\mu(r)$.

It is easy to verify that with a coupling of this form the results of the previous section for the dependence of the current-current correlation function on μ, L, and a are retained and the limit (3.14) is satisfactory.

V. COMPACT QED

The major objection to the treatment presented in the previous section is that it introduces fractional flux on links of the lattice and hence cannot be used for a compact theory. However an important physical point has been learned—gauge invariance does not dictate a relationship between the form of the derivative $D_\mu(k)$ and the form of the vertex Γ_μ. In (4.1) we have seen that it is this relationship which leads to a problem when we gauge the SLAC derivative by introducing a straight line of unit flux between the fermion fields.

Although the discussion of the noncompact theory does not directly tell us how to modify the Lagrangian of the compact theory, it points out the fact that there need be no tight connection between the gauge-invariant fermionic derivative terms and interaction between the fermions and the transverse gauge fields. Consider then the most general form for the gauge-invariant derivative in a compact gauge theory:

$$\bar{\psi} \not{D} \psi \to \sum_{j,n,\mu} \psi_j^\dagger \alpha_\mu \psi_{j+n\hat{\mu}} D_\nu(n)$$

$$\times \frac{\sum_{\text{paths}} \omega_{\text{path}} \prod_{l \in p} U_l}{\sum_{\text{paths}} \omega_{\text{path}}} , \tag{5.1}$$

where the sum over paths runs over all possible paths from j to $j + n\hat{\mu}$ and ω_{path} is any suitably chosen weighting factor. The usual gauged derivative corresponds to the choice $\omega_{\text{path}} = 1$ for the straight line path and $\omega_{\text{path}} = 0$ for all other paths. Let us instead consider the weighting

$$\omega_{\text{path}} = \sum_{l \in p} e^{-g^2 n_p^2(l)} , \tag{5.2}$$

where

$n_p(l)$ = number of times the path P traverses the link l in a positive direction

— number of traverses in a negative direction . $\tag{5.3}$

At strong coupling this weighting is clearly dominated by the shortest path—however for weak coupling the string becomes very nonlocalized. Now let us examine the coupling to any plaquette variable $B_{\mathscr{P}}$ given by this prescription. Consider any plaquette and isolate the contributions to the sum over paths coming from terms which carry flux $(p,q,r,-p-q-r)$ into the four corners of that plaquette. The coupling to $B_{\mathscr{P}}$ is

$$\Gamma_{\mathscr{P}} = \frac{-ig}{Z^{-1}} \sum_{(p,q,r)} \omega(p,q,r) \sum_{n=-\infty}^{\infty} (4n+3p+2q+r)e^{-(4n+3p+2q+r)^2 g^2} , \tag{5.4}$$

where

$$Z = \sum_{(p,q,r)} \omega(p,q,r) \sum_{n=-\infty}^{\infty} e^{-(4n+3p+2q+r)^2 g^2} . \tag{5.5}$$

The factor $\omega(p,q,r)$ is the sum of the weightings of the parts of the paths which do not touch the reference plaquette $\mathscr{P}$ (modified by an n-independent term from the reference plaquette weighting). For small g^2 it is instructive to rewrite this sum using the identity

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{m=\infty}^{\infty} \int_{-\infty}^{\infty} dx\, e^{2\pi mxi} f(x) . \tag{5.6}$$

One then finds

$$\Gamma(q) = \frac{1}{g} \frac{\sum\limits_{p,q,r} \omega(p,q,r) \sum\limits_{m=1}^{\infty} (2\pi m)\sin\pi m \left[\dfrac{3p+2q+r}{2}\right] e^{-m^2\pi^2/g^2}}{\sum\limits_{p,q,r} \omega(p,q,r) \left[1 + \sum\limits_{m=1}^{\infty} \cos\pi m \left[\dfrac{3p+2q+r}{2}\right] e^{m^2\pi^2/g^2}\right]} . \tag{5.7}$$

We see explicitly that every term in the numerator is smaller by at least a factor of $e^{-\pi^2/g^2}$ than the equivalent term in the denominator. Hence, as $g \to 0$, Γ vanishes faster than any power of g. This is then a compact equivalent of the result given in the previous section; i.e. it is a prescription for rendering the derivative gauge invariant which does not introduce a coupling to the plaquette variables $B_{\mathscr{P}}$ to any finite order in g.

Since the plaquette variables $B_{\mathscr{P}} (=\nabla \times \mathbf{A})$ are the transverse degrees of freedom of the gauge field, this calculation shows that any term in the fermion derivative which is rendered gauge invariant by the introduction of such a weighted sum of strings does not introduce any coupling of the fermions to the transverse photons of the theory. Thus we can obtain a theory which has *both* a long-range derivative *and* an acceptable coupling for transverse photons by the following prescription:

$$\sum \bar{\psi} \slashed{D} \psi = \sum_{j,\mu} \psi_j^{\dagger} \alpha_{\mu} U_{j,\hat{\mu}} \psi_{j+\hat{\mu}} + \sum_{\substack{j,n,\mu \\ n>1}} \psi_j^{\dagger} \alpha_{\mu} \psi_{j+n\hat{\mu}} D_{\nu}(n) \frac{\sum\limits_{\text{paths}} \omega_{\text{path}} \prod\limits_{l \in p} U_l}{\sum\limits_{\text{paths}} \omega_{\text{path}}} , \tag{5.8}$$

where the weighting (5.2) is used for all $n > 1$. This gives the propagator

$$D_i(k) = k_i \tag{5.9}$$

while the vertex for the *transverse* photons arises only from the nearest-neighbor term in (5.8) and thus has the form

$$gA_i(q)\Gamma_i(k,q)\psi^{\dagger}(k')\alpha_i\psi(k)\delta_p(k'-k-q) , \tag{5.10}$$

where

$$\Gamma_i(k,q) = \cos(k_i a) . \tag{5.11}$$

(Here transverse photons are simply chosen to be the A_i for those directions $\hat{i}\perp q$.)

The important feature of the coupling (5.11) is that the magnitude of the coupling to fermions with k near π/a is not enhanced relative to that to fermions with k near zero. Hence the calculation for π_{ii} in this theory parallels that given for $\Delta_{\rho\rho}$ in Sec. III. One simply replaces the

numerator factors γ_0 in (5.10) by the quantity $\Gamma_i \alpha_i$. The arguments given in the appendixes can all be repeated with this vertex structure and the conclusions are unchanged. Hence this theory gives the usual weak-coupling perturbation theory results when the continuum limit is taken as described in Sec. III.

Although this is a very attractive result, the generalization of the calculation just given to a non-Abelian theory is nontrivial. Our counting made use of the Abelian nature of the gauge field. Each string was weighted by $\exp(-g^2E^2)$ where E is the field created by the string operator acting on the flux-free state. The state created by the action of successive U operators on a link is not unique in the non-Abelian case so an equivalent weighting is not readily achieved. Fortunately, we believe that the strong result achieved above is not necessary in the non-Abelian case, because such theories are asymptotically free. Consider instead a weighting

$$\omega_{\text{path}} = e^{-q^2 L_{\text{path}}} , \tag{5.12}$$

where L_{path} is the length of the path in question. A calculation of the coupling to a given plaquette variable $B_{\mathscr{P}}$, similar to that given above for the weighting (5.12) shows that Γ vanishes at least as fast as g^5. In an asymptotically free theory such nonleading contributions to perturbation theory are presumably irrelevant in the $a \rightarrow 0$ limit.

VI. CONCLUSIONS

Although the calculations of the preceding sections of the paper were carried out for a theory of massive fermions it is obvious that we could have dealt with a chiral theory from the outset by setting the parameter $m = ma = 0$ and projecting all fermion fields onto fields of a single definite chirality. Hence, we assert that we established that there do exist schemes for defining purely chiral gauge theories on a lattice.

Unfortunately, we learned that in order to get results free of lattice artifacts not only does the fermionic derivative have to have infinite range but also the introduction of gauge fields has to be written so that for each term in the fermion derivative we sum over an infinite number of string configurations. Obviously, for strong coupling the string configurations which do not correspond to the straight line routing of the flux between fermions are severely damped; hence, whether or not one uses the Hamiltonian which we have proposed, the strong-coupling results will be unmodified. The problem, as always, is what happens as we try to approach the weak-coupling region of the theory.

We will now argue that for couplings on the order of unity or less we expect there to be a significant difference between the behavior of the theory we have written down and a straight-line gauging of a theory based upon a SLAC-type of derivative. Furthermore, we will argue that the necessity for summing over multiple string configurations reflects an aspect of the problem which transcends the desire to get weak-coupling perturbation theory to work out correctly; namely, the desire for a renormalization-group-invariant formulation.

A. Exponential damping and the renormalization group

We will show that starting from a theory based upon a SLAC derivative simple renormalization-group ideas force us to gauge the theory in the manner just discussed. To see this, imagine that we begin calculating for a small value of the lattice spacing and a small value of the coupling. Since we are dealing with an asymptotically free theory, we can imagine writing a trial ground-state wave function which is essentially a Gaussian free-field wave function for gluon and fermion degrees of freedom corresponding to momenta near π/a. Integrating out these degrees of freedom we obtain an effective Hamiltonian for the remaining degrees of freedom: i.e., a theory with a new coupling and a larger lattice spacing. While carrying out such a calculation for a non-Abelian theory is somewhat problematic, for compact QED the process is entirely straightforward.

The question we wish to ask is how the gauged fermion kinetic term changes under such renormalization-group transformations. It is easy to see[3] that one effect of integrating out the high-momentum modes of the gauge field is to modify the weighting factor by an amount $\exp(-g^2\tau L_{\text{path}})$ where τ is some calculable constant. For a choice of ω_{path} which is unity for straight lines and zero otherwise this result means that in one renormalization-group transformation the fermion kinetic term goes over to the form of the exponentially damped SLAC derivative with a damping factor $\mu = g^2\tau$. This, however, is a disaster since although in a non-Abelian theory $g^2 \rightarrow 0$ as $a \rightarrow 0$ it does so only logarithmically; thus, given the results obtained in the previous sections, in this sort of theory the fact that lattice artifacts will survive the continuum limit would appear to be unavoidable.

In the case of a more general ω_{path} the situation can be quite different. The reason for this is that now we are faced with the question of what is the mean length of a path joining terms in the fermionic kinetic term. This is not a trivial question since in computing this quantity one has to take into account the competition between the damping factor $e^{-g^2\tau L}$, which tends to favor the shortest paths, and density of entropy factor, the number of paths of length L joining a given pair of points, which grows with the length of the path and favors longer paths. Clearly, it is quite possible that as one varies g^2 the entropy factor can win out and the mean path length can grow rapidly. In fact, in a moment we will argue that this will happen for $g^2 \ll 1$; however, let us first take a moment to see why this is important.

Suppose for a moment that for points separated by a distance less than some fixed value L_0 the sum over paths joining these two points is dominated by paths of length L_0, whereas for points separated by more than this distance the sum is dominated by the straight line joining the points. Now consider a single renormalization-group transformation of the form just discussed. As before, the gauge field factor is rescaled by a factor of $e^{-g^2\tau L}$, but now this means that all terms in the kinetic part of the Hamiltonian which involve fermion fields separated by a distance smaller than L_0 are rescaled by a common factor. Hence, in this case one once again gets an exponentially damped SLAC derivative; however, now the damping scale is set by L_0 and not $1/g^2\tau$. This is important since this provides a way of decoupling the factor μ appearing in our formulation from g^2 and thus invalidating the general renormalization-group argument that indicates that lattice artifacts must survive the continuum limit. In fact we will now argue that there is a value of g^2 below which the mean path length joining two fixed points diverges, which puts us back into the situation of the undamped SLAC derivative.

Now let us see why there is a dramatic change in the mean length of a path joining two fixed points as we vary g^2. The mean length of the path for a term in H involving fermions separated by a distance r is given by

$$\langle L \rangle_r = \frac{\sum\limits_{\text{paths}} \omega_{\text{path}} L_{\text{path}}}{\sum\limits_{\text{paths}} \omega_{\text{path}}} \; . \tag{6.1}$$

We can obtain a crude estimate of this quantity by underestimating the number of paths of length $L > r$ as fol-

lows: any path joining two points separated along some lattice direction by distance r must have at least r links along that direction. Of the remaining $(L - r)$ links, $(L - r)/2$ can be chosen in any of $2(d - 1)$ directions without any risk of retracing the same path. The remaining $(L - r)/2$ links must be chosen to bring the path back *on axis* and so are not completely random. It follows that there are at least $2(d - 1)^{(L - r)/2}$ distinct paths of length L between points separated by a distance r. This is a gross undercounting of the number of long paths but it is sufficient to show that the average path length diverges at some finite g. For a weighting such as (5.12) we find

$$\langle L \rangle_r \geq \frac{\sum\limits_{L > r} L e^{-g^2 \tau L} 2(d - 1)^{(L - r)/2}}{\sum\limits_{L > r} e^{-g^2 \tau L} (2d)^{(L - r)/2}} . \tag{6.2}$$

Clearly for

$$g^2 < (1/2\tau) \ln[2(d - 1)] \tag{6.3}$$

this expression diverges for all r. This is of course the result we wished to establish.

B. Physical scales and confinement

For a non-Abelian gauge theory we believe that confinement occurs at some fixed physical length scale. This suggests that the fact that $\langle L \rangle_r \to \infty$, which was important to reproduce weak-coupling perturbation theory in QED, may in fact be modified in such a theory. It is possible, even intuitively plausible, that the average length of the flux strings does not diverge in physical units but is in fact some finite multiple of the physical confinement scale. This notion is suggested by the discussion of the previous section.

Imagine a recursive procedure of successively thinning degrees of freedom to increase the lattice spacing. Starting with a form such as (5.12) at small lattice spacing we evolve to stronger coupling as the lattice spacing increases over many recursive steps. Eventually, at sufficiently strong coupling, we expect that confinement means that there is an energy cost for having gluon flux stretch over large distances in units of the confinement length. Presumably, this means that at this point the argument given in deriving (6.2) applies but with a value of g^2 above the critical value and so we expect $\langle L \rangle_r \approx L_{\text{confinement}}$; hence we expect the effective form of the fermion derivative to go over to the damped form on a scale set by the confinement length. Since we believe that this form of the Hamiltonian will apply to the study of the motion of quarks within a hadron our earlier analysis would suggest that the turning down of the spectrum for momenta near $\kappa \approx \pi$ means the appearance of states with masses on the order $\pi/L_{\text{confinement}}$.

Although we referred to states of this sort as *spurious* in our earlier discussions, because they did not appear to have a rotationally (or at least approximately) invariant spectrum, the states of the recursed Hamiltonian need not be lattice artifacts—they can be finite-mass hadronic states of QCD. The reason this can be true is that many recursive steps are needed to go from weak to strong cou-

pling and the SLAC derivative will, along the way, evolve into a form which involves off-axis terms as well as on-axis terms and which is essentially rotationally invariant. We raise this rather speculative point because of results which we obtained in an earlier paper where we analyzed chiral-symmetry breaking in the context of a strong-coupling version of the SLAC derivative.[10] Our results did not require an infinite range for the SLAC derivative and would be valid also with a weighting such as (5.12). The result of the calculations described in that paper were that the chiral symmetry of the theory is spontaneously broken, implying the existence of a massless multiplet of pseudoscalar Goldstone bosons and also massive multiplets of particles which we identified as a ρ, ρ', and a possible π' multiplet. The relevance of this calculation to the present one is that if these states are written in terms of the "free" quark modes of the coarse-grained lattice then the massless states contain no quarks from the region $\kappa \approx \pi$ but the massive ones all contain at least one quark from this region.

The combination of these earlier results with our admittedly speculative argument suggest that those features which are essential for obtaining the correct free spectrum and weak-coupling perturbation theory may be less important when it comes to the study of the hadron spectrum.

C. Practical calculations

Given the discussion of the previous sections we would argue that any theory which attempts to deal with truly chiral theories, must eventually deal with a gauging of the fermion derivative term which contains a sum over arbitrarily long strings. Obviously, this is not the sort of term which lends itself readily to carrying out practical calculations. The question then arises as to how to implement our results within the framework of a practical computational scheme.

First, we would like to point out that the formalism we have introduced is explicitly Hamiltonian in nature and we do not know how to generalize it to a Euclidean path-integral approach, except in the obvious way as a theory of a finite-spatial lattice with continuous time. It makes no sense to take our prescription for a spatial lattice and use the same prescription to write down a Euclidean formulation. The introduction of long range in the time direction in $\mathcal{L}$ leads to a non-Hermitian Hamiltonian which would seem to be extraordinarily difficult to understand.

One can, of course, introduce an asymmetric formulation of a Euclidean theory which takes a finer mesh for the time direction than the space direction, and use only nearest-neighbor terms for the time direction. Naively, this would appear to lead to a single doubling of all of the propagators. We have not studied this theory so we cannot make any statements about decoupling of the spurious states in the continuum limit. A more attractive alternative is to adopt the Monte Carlo techniques of Blankenbecler and Sugar[11] who work directly with the transfer matrix computed from the Hamiltonian.

Yet another possibility is to eschew Monte Carlo techniques and directly attack the Hamiltonian theory by

strong-coupling perturbation theory,[12] nonperturbative methods like the t expansion,[13] or whatever other method comes to hand.

In any such approach it will be necessary to truncate the full Hamiltonian and study one's results as a function of the truncation scheme. For example, one could work with the nearest—next-nearest-neighbor fermion derivative and include sums over strings of length up to some length L_0. One could then study quantities which are expected to scale early, i.e., ratios of masses, and see how they depend upon the parameters μ and L_0. Reliable answers will be those which are insensitive to the choice of these parameters. The virtue of this sort of approximation scheme is that we know the specific form of the Hamiltonian being approximated and presumably can obtain control over the dependence of the results obtained on the approximations. Furthermore, we have reason to believe, from the strong-coupling calculations which have been studied previously, that the masses of the true physical states *will* indeed be relatively insensitive to μ and L_0.

D. A word about anomalies

Up to now we have studied the question of how to write a satisfactory lattice gauge theory for the case of a nonanomalous chiral symmetry. This is because we know that in the continuum it does not make sense to gauge an anomalous chiral symmetry. Presumably, although it is not apparent from what we have said in this paper, the same is true for our lattice models; even though one can carry out the entire construction we have just described for theories of this type. Probably, what goes wrong is that in such a theory it is not possible to find a range of couplings which will allow us to take the limit $a \to 0$ and define a satisfactory continuum theory. Analysis of this question requires a thorough study of the dynamics of models of this type and goes beyond the scope of this paper.

There is another question related to anomalies which usually comes up in the discussion of fermions in lattice gauge theories, and that is that there is a widespread belief that such theories cannot exhibit anomalies in any global symmetries. It is therefore believed that the U(1) problem cannot be solved correctly in a lattice theory. This belief is based upon a collection of notions which our discussion has shown to be fallacious. The first fact is that for any theory of massless lattice fermions there is, even for a current which would be conserved in the continuum theory, an exactly conserved lattice chiral charge. The argument then goes that this means there must exist an exactly conserved chiral current corresponding to this charge in the continuum limit of the lattice theory. In an earlier paper by one of us[14] it was shown that this argument was fallacious for a lattice version of the $(1 + 1)$-dimensional Schwinger model based upon a SLAC derivative. While there is an exactly conserved chiral charge in this lattice theory it is not the integral of a current density which has a finite continuum limit; moreover, the only axial-vector current which does have a continuum limit is not conserved. Applying our present analysis to a $(1 + 1)$-dimensional theory shows that the effects of damping by a scale factor proportional to g^2 do not pro-

duce lattice artifacts in the continuum limit of such a theory; hence, our results provide additional supports for the arguments presented in Ref. 14. Obviously, a parallel discussion of theories in higher dimension will require more care and remains to be given. Nevertheless, there is no reason to believe that there are any nonperturbative effects which will change the fact that computations in our present formulation of lattice gauge theories of massless fermions go over to their continuum form if the limits $a \to 0$ and $g^2 \to 0$ are taken properly.

Another misconception which leads to the conclusion that one can have no anomalies in theories with lattice fermions, is that species doubling produces contributions which cancel the ordinary fermion contribution to the anomaly calculation. Obviously, from what has been shown in this paper this argument applies only to the case of the nearest-neighbor derivative. Even in the pathological case of the nearest—next-nearest-neighbor derivative we have seen that the density of states in the *doubled region* is very different from that in the normal region and no detailed cancellation of contributions is possible for general values of μ.

E. Final remarks

We would like to conclude by emphasizing that the arguments which we have presented in this paper show that it is in fact quite possible to formulate lattice gauge theories which allow the introduction of chiral fermions in a straightforward manner. In addition, this same method allows us to discuss vectorlike gauge theories of massless fermions in a way which preserves global chiral symmetries and allows a direct attack on the problem of spontaneous symmetry breaking and the dynamical origin of nonelementary Goldstone bosons. Admittedly, the Hamiltonians which we propose appear more difficult to deal with than the more familiar versions of lattice gauge theories; however, this may be illusory. It could well be that the fact that this class of theories preserves all chiral symmetries allows for a more direct attack on questions of dynamical symmetry breaking than is possible in theories in which one explicitly breaks these symmetries and then attempts to discover how to take the continuum limit in such a way that the symmetry is restored.

In addition to this remark we would also like to point out that while we believe on physical grounds that the complexity of our Hamiltonian is unavoidable, that aspect of our discussion is heuristic and should not be taken at face value. Since this paper shows that something which was heretofore widely believed to be impossible is, in fact, possible, we encourage others to have another look at the problem in the hope that they will find something we have overlooked and arrive at a simpler formulation of a satisfactory theory.

APPENDIX A:
1 + 1 DIMENSIONS—NEXT-NEAREST NEIGHBOR

The results of Table I are obtained from the current-current correlation function given in (3.8). In $1+1$ dimensions $(d=1)$ this quantity requires no subtractions in the continuum theory because the apparently logarithmically divergent terms cancel by current conservation. Hence we make no subtractions on our lattice expressions.

After evaluating the traces, Feynman parametrizing, and performing the κ_0 integration, (3.8) yields

$$\Delta_{00}{}^1(\mathbf{q},q_0) = \prod_{2N+1} \sum_{p,r,n} \delta(\kappa_r - \kappa_p - \kappa_s - 2\pi n)$$
$$\times \int_0^1 d\alpha \, F(\kappa_r,\kappa_p) \,, \tag{A1}$$

where

$$F(\kappa_r,\kappa_p) = \frac{[\widetilde{D}(\kappa_r) - \widetilde{D}(\kappa_p)][(1-\alpha)\widetilde{D}(\kappa_p) - \alpha\widetilde{D}(\kappa_r)]}{[\alpha\widetilde{D}^2(\kappa_r) + (1-\alpha)\widetilde{D}^2(\kappa_p) + m^2 - \alpha(1-\alpha)q_0^2 a^2]^{3/2}} \tag{A2}$$

and

$$m = ma, \quad qa = \kappa_s = \frac{2\pi s}{2N+1} \,. \tag{A3}$$

We wish to study fixed finite external momenta q. For the sake of discussion we examine $q>0$, which means $0<s\ll N$. Then the delta function restricts the n sum to contributions from $n=0$—the nonumklapp terms, and an $n=-1$ contribution when $p+s>N$. Thus we can rewrite the sums in (A1) as

$$\sum_{p,n} \delta_p(\kappa_r - \kappa_p - \kappa_s - 2\pi n)$$
$$= \sum_{p=-N}^{N-s} \delta(r-p-s)$$
$$+ \sum_{p=N-s+1}^{N} \delta(r-p-s+2N+1) \,. \tag{A4}$$

We can further divide the contribution of the first term in (A4) by looking at the behavior of $\widetilde{D}(\kappa)$. We have, for the case of the next-nearest-neighbor derivative,

$$\widetilde{D}(\kappa_r) = \frac{(1+\mu)}{2\mu}\sin\kappa_r + \frac{(1-\mu)}{4\mu}\sin2\kappa_r \,. \tag{A5}$$

In the region of small κ_r this function is well approximated by the linear behavior

$$\widetilde{D}(\kappa) = \kappa \quad \text{for } \kappa < \sqrt{\mu} \,. \tag{A6}$$

We will study the limit

$$\mu = \delta\left[\frac{a}{L}\right] = \frac{\delta}{2N+1} \quad \text{for finite } \delta, \quad N \to \infty \,. \tag{A7}$$

In this limit any finite physical q satisfies

$$q \ll \left[\frac{\sqrt{\mu}}{a}\right] \quad \text{or } \kappa_s \ll \sqrt{\mu} \,. \tag{A8}$$

Thus we can further divide the sum over p into the regions

$$|p| \gtrless \frac{\sqrt{\mu}(2N+1)}{2\pi} = \frac{\delta}{2\pi}\sqrt{2N+1} \,.$$

In each region of the sum there are an infinite number of terms in the limit $a \to 0$ and hence we can make the replacement

$$\frac{1}{(2N+1)} \sum_p \to \frac{1}{2\pi} \int d\kappa_p \,. \tag{A9}$$

The three contributions are thus

$$\Delta_{\text{region I}} = \frac{1}{2} \int_{-\sqrt{\mu}}^{\sqrt{\mu}} d\kappa \, F(\kappa+\kappa_s,\kappa) \,, \tag{A10}$$

$$\Delta_{\text{region II}} = \frac{1}{2\pi} \left[\int_{-2\pi N/(2N+1)}^{-\sqrt{\mu}} d\kappa \, F(\kappa+\kappa_s,\kappa) \right.$$
$$\left. + \int_{\sqrt{\mu}}^{(2\pi N-s)/(2N+1)} d\kappa \, F(\kappa+\kappa_s,\kappa) \right] \,, \tag{A11}$$

and finally the umklapp contribution

$$\Delta_{\text{region III}} = \frac{1}{2\pi} \int_{2\pi(N-s+1)/(2N+1)}^{2\pi N/(2N+1)} d\kappa \, F(\kappa+\kappa_s - 2\pi,\kappa) \,. \tag{A12}$$

All that remains to obtain the results given in Table I is to estimate the dominant contributions from each of these expressions in the limit (A7). To do this, it is useful to note that

$$\widetilde{D}(\kappa+\kappa_s)$$
$$= \widetilde{D}(\kappa) + \frac{\kappa_s[(1+\mu)\cos\kappa - (1-\mu)\cos2\kappa]}{2\mu} + O\left[\frac{\kappa_s^3}{\mu}\right]$$
$$= \widetilde{D}(\kappa) + \kappa_s B(\kappa) + O\left[\frac{\kappa_s^3}{\mu}\right] \,. \tag{A13}$$

In region I we use the linear approximation (A6). This gives

$$\Delta_{\text{region I}} = \frac{1}{2} \int_0^1 d\alpha \int_{-\sqrt{\mu}}^{\sqrt{\mu}} d\kappa \frac{(qa)[(1-2\alpha)\kappa - \alpha qa]}{[(\kappa+\alpha qa)^2 + \alpha(1-\alpha)(q^2a^2 - q_0^2a^2) + m^2]^{3/2}} \,. \tag{A14}$$

Shifting the integration variable and exploiting the $\kappa \leftrightarrow -\kappa$ symmetry this can be rewritten as

$$\Delta_{\text{region I}} = \int \frac{d\alpha}{2\pi} \left[\int_0^{\sqrt{\mu} - \alpha qa} d\hat{\kappa} \frac{2\alpha(1-\alpha)q^2a^2}{[\hat{\kappa}^2 - \alpha(1-\alpha)(q_0^2 - q^2)a^2 + m^2a^2]^{3/2}} \right.$$

$$\left. + \int_{\sqrt{\mu} - \alpha qa}^{\sqrt{\mu} + \alpha qa} d\hat{k} \frac{qa[(1-2\alpha)\hat{\kappa} - 2\alpha(1-\alpha)qa]}{[\hat{\kappa}^2 - \alpha(1-\alpha)(q_0^2 - q^2)a^2 + m^2a^2]^{3/2}} \right]$$

$$= 2\int d\alpha \frac{\alpha(1-\alpha)q^2}{m^2 - \alpha(1-\alpha)(q_0^2 - q^2)} + O\left(\left[\frac{qa}{\sqrt{\mu}}\right]^2\right), \tag{A15}$$

where the second integral has been estimated by setting $\hat{\kappa} = \sqrt{\mu}$ over its entire range. The first term represents the usual continuum contribution. The remaining contributions vanish in the limit (A7).

The contributions of region II can be treated using (A13). We find

$$\Delta_{\text{region II}} = \frac{1}{2\pi} \int d\alpha \left[4q^2a^2\alpha(1-2\alpha) \int_{\sqrt{\mu_\alpha qa}}^{2\pi N/(2N+1) - (1-\alpha)qa} d\kappa \frac{B^2(\kappa)}{R^{3/2}(\kappa, ga)} \right.$$

$$+ \int_{\sqrt{\mu} - \alpha qa}^{\sqrt{\mu} + \alpha qa} d\kappa \frac{qaB(\kappa)[(1-2\alpha)D(\kappa) + 2\alpha(1-\alpha)qaB(\kappa)]}{R^{3/2}(\kappa, qa)}$$

$$\left. + \int_{2\pi N/(2N+1) - (1-\alpha)qa}^{2\pi N/(2N+1) - \alpha qa} d\kappa \frac{qaB(\kappa)[(1-2\alpha)D(\kappa) + 2\alpha(1-\alpha)qaB(\kappa)]}{R^{3/2}(\kappa, qa)} \right] + O\left(\frac{(qa)^3}{\mu}\right), \tag{A16}$$

where

$$R(\kappa, qa) = \tilde{D}^2(\kappa) + \alpha(1-\alpha)(q^2a^2B^2 - q_0^2a^2) + m^2a^2 .$$

The second two integrals are estimated by taking the maximum value of the integrand within the range multiplied by the range. The contribution of the second integral is bounded by a term of order q^2a^2/μ^2 while the third integral gives contributions of order μ.

The first integral is best treated by dividing it into the regions

$$\kappa < \frac{\pi}{2} \quad \text{and} \quad \frac{\pi}{2} < \kappa .$$

In the first of these regions we set $\cos\kappa = 1 - y$, the resultant integral is then bounded by

$$\int_0^1 dx \frac{2\alpha(1-\alpha)q^2a^2\mu}{\pi} \int_{1-\cos(\sqrt{\mu} - qa)}^1 dy \frac{[-3y + 2y^2 + \mu(2-y)(1-2y)]^2}{y^2(2-y)^2[y + \mu(2-y)]^3} < \frac{q^2a^2\mu}{3\pi} \int_\mu^2 dy \left[\frac{9}{4y^3}\right] + \text{higher order in } \mu$$

$$= O\left[\frac{q^2a^2}{\mu}\right] . \tag{A17}$$

In the second region we set $\cos\kappa = -1 + y$. The contribution of this region is then

$$\int d\alpha \frac{4\pi^2 s^2\mu}{(2N+1)^2} \int_\epsilon^1 dy \frac{[2(2-y)^2]^2}{y^2(1-y)^5} \to O(\mu) , \tag{A18}$$

where

$$\epsilon = 1 + \cos\left[\frac{2\pi[N - (1-\alpha)s]}{2N+1}\right] = O\left[\left[\frac{2\pi[s(1-\alpha) + \frac{1}{2}]}{(2N+1)}\right]^2\right] = O((qa)^2) .$$

Finally, in region III, the umklapp region, we can expand both $\tilde{D}(\kappa)$ and $D(\kappa + \kappa_s - 2\pi)$ about their values at $\kappa = \pi$. This gives

$$\tilde{D}\left[\kappa = \pi\left[1 - \frac{\sigma}{2N+1}\right]\right] = \frac{\sigma}{\mu}\left[\frac{\pi}{2N+1}\right], \quad \tilde{D}(\kappa + \kappa_s - 2\pi) = \frac{(-2s + \sigma)}{\mu} \frac{\pi}{2N+1}, \quad 1 < \sigma < 2(s-1) . \tag{A19}$$

Thus the contribution of the umklapp region is

$$\Delta_{\text{region III}} = \mu \int_0^1 d\alpha \int_1^{2(s-1)} d\delta \frac{(-s)[(1-\alpha)\sigma + \alpha(s-\sigma)]}{[\alpha(s-\sigma)^2 - (1-\alpha)\sigma^2]^{3/2}} \to O(\mu) .$$

Hence we have shown that only the usual continuum contribution survives in the limit (A7).

APPENDIX B:
THE DAMPED SLAC DERIVATIVE

In order to evaluate the loop contributions using the damped SLAC derivative we need first to derive certain properties of the spectral function $\widetilde{D}(\kappa_p)$. The Hamiltonian defined in (2.19) gives

$$\widetilde{D}(\kappa_p) = \frac{C(\mu)}{2N+1} \sum_{t=-N}^{n} \kappa_t e^{-\mu m}(e^{i(\kappa_p - \kappa_t)m} + e^{-i(\kappa - \kappa_t)m}) ,$$

(B1)

where $\kappa_p = 2\pi p/(2N+1)$ and the normalization $C(\mu)$ is chosen so that the slope of $\widetilde{D}(\kappa)$ is 1 at $\kappa = 0$. We are going to study the limit

$$\mu = \delta \left[\frac{a}{L} \right]^2 = \frac{\delta}{(2N+1)^2}, \quad N \to \infty \quad \delta \text{ fixed} .$$

(B2)

In this limit μN is of order $1/N$, and hence vanishes as $N \to \infty$. A little algebra gives

$$\widetilde{D}(\kappa_p) = C\kappa_p \left[1 - \mu N + \frac{2\mu}{2N+1} \sum_{r=1}^{n} F(r) \right]$$
$$- \frac{C\mu\pi}{2N+1} \sum_{s-1/2}^{p-1/2} F(N + \tfrac{1}{2} - s) + O(\mu, (\mu N)^2) ,$$

(B3)

where

$$F(r) = \frac{1 - (-1)^r \cos\kappa_r/2}{2\sin^2\kappa_r/2} .$$

(B4)

Now

$$\frac{1}{2N+1} \sum_{r=1}^{N} F(r) = \frac{2N+1}{\pi^2} + O(1)$$

(B5)

and

$$\frac{1}{2N+1} \sum_{s=1/2}^{p-1/2} F(n + \tfrac{1}{2} - s) = \frac{2N+1}{\pi^2(N+1-p)} + O(1) .$$

(B6)

Hence

$$\widetilde{D}(\kappa) = \kappa - \frac{2\mu}{\pi(\pi - \kappa)} + O(\mu, \mu^2 N^2) .$$

(B7)

Thus

$$\frac{d\widetilde{D}}{d\kappa} = 1 - \frac{2\mu}{\pi(\pi - \kappa)^2} - O(\mu, \mu^2 N^2) .$$

(B8)

The maximum occurs at

$$(\pi - \kappa) = \left[\frac{2\mu}{\pi} \right]^{1/2} .$$

(B9)

For κ greater than $\pi - \sqrt{2\mu/\pi}$ the slope of $\widetilde{D}(\kappa)$ is negative—this spurious region is of width $1/N$ and hence contains only a finite number of momenta $\kappa_m = 2\pi m/(2N+1)$ even in the limit $N \to \infty$. The

minimum value of $\widetilde{D}(\kappa)$ in the spurious region occurs at $\kappa_N = 2\pi N/(2N+1)$ and is given by

$$\widetilde{D}(\kappa_N) = \kappa_N - \frac{4\mu(2N+1)}{\pi^2} + O(\mu, \mu^2 N^2) > \kappa_N - 2\mu N .$$

(B10)

The current-current correlation function of interest, Eq. (3.8), is quadratically divergent in the continuum theory and hence requires three subtractions. We will make the same three subtractions and study the resulting expression in the limit (B2). In order to define precisely what we mean by the various contributions in Table III, we adopt the following procedure: (1) evaluate the traces, Feynman parametrize, and perform the (finite) κ_0 integrations; (2) divide the sums over each of the spatial momenta into the regions given in Table II; (3) treat each such region as a separate expression to be three times subtracted at $\langle q_0, \mathbf{q} \rangle = 0$.

This yields the results of Table III. (Because the boundaries of the various regions are q dependent the sums in the subtraction terms may run over a different range than those in the original expression.) Of course, provided the subtraction is done correctly, the answer does not depend on how the contributions are distributed between entries on our table. However it is important to note that a correct subtraction procedure treats the umklapp contribution separately. Let us outline a one-dimensional example of a twice-subtracted sum just to make all of this clear.

Consider the unsubtracted expression

$$B^0(q) = \sum_{p,n} F(\kappa_l, \kappa_\mu) \delta_p(\kappa_l + \kappa_s + 2\pi n - \kappa_r) ,$$

where $q = 2\pi s/(2N+1)$,

$$B^0(q) = \sum_{-N}^{N-s} F(\kappa_p, \kappa_p + \kappa_s)$$
$$+ \sum_{N=s+1}^{N} F(\kappa_p, \kappa_p + \kappa_s - 2\pi) .$$

(B11)

Now subtract twice. The contribution of the first sum in (B11) is

$$(B_{\text{region I}})^2 = \sum_{-N}^{N-s} F(\kappa_p, \kappa_p + \kappa_s) - \sum_{-N}^{N} F(\kappa_p, \kappa_p)$$
$$- \frac{2\pi s}{2N+1} \sum_{-N}^{N} \frac{d}{d\kappa_r} F(\kappa_p, \kappa_p + \kappa_r) \bigg|_{r=0}$$
$$+ \frac{2\pi s}{2N+1} F(\kappa_N, \kappa_N) ,$$

(B12)

while the second sum gives

$$(B_{\text{region II}})^2 = \sum_{N-s+1}^{N} F(\kappa_p, \kappa_p + \kappa_s - 2\pi)$$
$$- \frac{2\pi s}{2N+1} F(\kappa_N, \kappa_{-N}) .$$

(B13)

Since the umklapp region is of width q the contribution of region II is explicitly of order q before any subtraction. The second subtraction introduces a boundary term which

is different for the umklapp and nonumklapp contributions and hence does *not* cancel out when the two regions are added.

The results given in Table III are then obtained by a straightforward step-by-step procedure. In the spurious regions the denominator $\widetilde{D}(\kappa)$ is replaced by the bound (B10) while numerator factors of $\widetilde{D}(\kappa)$ can be replaced by the upper bound ($|\widetilde{D}(\kappa)| < \pi$). In both the normal and the umklapp regions, after subtractions the sums can be replaced by integrals using the prescription

$$\frac{1}{2N+1} \sum_{p=m}^{m'} \rightarrow \frac{1}{2\pi} \int_{2\pi m/(2N+1)}^{2\pi m'/(2N+1)} d\kappa_p \; . \tag{B14}$$

However, the number of states in the spurious region remains finite in the $N \rightarrow \infty$ limit; hence, these regions must be treated as discrete sums even in this limit.

[1]H. B. Nielsen and Masao Ninomiya, Phys. Lett. **130B**, 389 (1983); **105B**, 219 (1981); Nucl. Phys. **B193**, 173 (1981); **B185**, 20 (1981).

[2]L. R. Karsten and J. Smit, Phys. Lett. **85B**, 100 (1979).

[3]J. M. Rabin, Phys. Rev. D **24**, 3218 (1981).

[4]As usual, in this statement we ignore the fact that the renormalization group applied to QED implies that $g \rightarrow \infty$ for $a \rightarrow 0$; that is, we are talking of a naive $a \rightarrow 0$ limit. This is because we are actually interested in formulating asymptotically free theories for which $a \rightarrow 0$ and $g \rightarrow 0$ is indeed the relevant limit.

[5]Note that in the limit $a \rightarrow 0$ any finite value of $\mathscr{E}(\kappa)$ corresponds to an infinite value of the dimensioned energy $\mathscr{E}/a$; hence, in this limit states in what we call *the spurious region* indeed have finite energy.

[6]Paolo Nason, Nucl. Phys. **B260**, 269 (1985).

[7]An alternative procedure for making the range of the Hamiltonian finite is to cut the sum off sharply for $|j - l| \geq L_0$. This is undesirable because any sharp cutoff causes oscillations which have to damp out. This procedure was discussed by Nason in his treatment of the lattice Schwinger model (Ref. 6). For our purposes the smoother, exponentially damped derivative is easier to deal with and we will limit all further discussion to this case.

[8]S. D. Drell, M. Weinstein, and S. Yankielowicz, Phys. Rev. D **14**, 1627 (1976).

[9]S. D. Drell, H. R. Quinn, B. Svetitsky, and M. Weinstein, Phys. Rev. D **10**, 619 (1979).

[10]S. D. Drell, H. R. Quinn, B. Svetitsky, and M. Weinstein, Phys. Rev. D **22**, 490 (1980); **22**, 1190 (1980); H. R. Quinn, S. D. Drell, and S. Gupta, *ibid.* **26**, 3689 (1982).

[11]R. Blankenbecler and R. L. Sugar, Phys. Rev. D **27**, 1304 (1983).

[12]T. Banks, J. Kogut, S. Raby, D. Sinclair, and L. Susskind, Phys. Rev. D **15**, 1111 (1977).

[13]D. Horn, M. Karliner, and M. Weinstein, Phys. Rev. D **31**, 2589 (1985); D. Horn and M. Weinstein, *ibid.* **30**, 1256 (1984); C. P. Van Den Doel and D. Horn, *ibid.* **33**, 3011 (1986).

[14]Marvin Weinstein, Phys. Rev. D **26**, 839 (1982).