SLAC-PUB-2897
March 1982
(T)

Anomalies and the Lattice Schwinger Model: Paradigm Not Paradox. *

Marvin Weinstein

Stanford Linear Accelerator Center, Stanford California 94305

submitted to Physical Review D

* work supported by the Department of Energy, contract DE-AC03-76SF00515.

Abstract

This paper shows that contrary to statements extant in the literature, it is possible to introduce fermions into a lattice gauge theory in such a way as to preserve the continuous chiral symmetries of the massless theory and the physics of the axial anomaly. The particular model discussed is the lattice Schwinger model, and the methods used are based upon the non-perturbative gauge-invariant variational techniques introduced by Horn and Weinstein. It is demonstrated that the physics of the anomaly, and its relation to the angles appearing in the exact solution to the continuum model appears in a simple and elegant way. The generalization of the model to several sets of independent fermions is discussed at the end of the paper. Some brief remarks are made about what happens if one attempts to gauge an anomalous current. These results are of interest, since the DWY-prescription is the only known way of writing down purely lattice gauge theories with only left-handed fermions.

1. INTRODUCTION

This paper demonstrates that one can understand the structure of the continuum Schwinger model, including its axial anomaly, by studying a chirally invariant lattice version of the theory. The analysis to be presented is structured along the lines suggested in a recent paper by D. Horn and M. Weinstein[1] where a formalism for carrying out gauge invariant variational calculations for lattice gauge theories was introduced. In the case of the Schwinger model we will need only the most general aspects of this method, and no detailed assumptions about the nature of the variational wavefunctions will be required.

One reason for studying the structure of the lattice Schwinger model is that in the past few years several authors have argued it is impossible to formulate lattice gauge theories which possess both the continuous chiral symmetries and chiral anomalies of the continuum limit.[2-4] The arguments presented to buttress this position have ranged from technically correct, but unnecessarily restrictive[3-4], to incorrect[2] . A discussion of the relationship of this work to the material to be presented here appears in the concluding section of this paper.

Another reason for studying the Schwinger model is that it has many fascinating properties. In addition to the axial anomaly which gives the photon a mass, it possesses two angles which label the exact solutions to the model but which do not appear in the Hamiltonian, and a Goldstone mode which seizes.[5] Our goal to show that not only are these features of the theory understandable, but that the physics behind these phenomena appears in an elegant and simple manner. Furthermore, we will show that this physics only emerges when one formulates the chirally symmetric lattice theory using the derivative introduced by Drell, Weinstein and Yankielowicz, (DWY).[6,7] We hope this discussion will clarify the physics and lay to rest the spurious argument which says that because a lattice theory defined using the DWY-derivative has a conserved chiral charge it cannot have any of the physics of the axial anomaly.

2. REPRISE OF KNOWN FACTS

The continuum Schwinger model describes the electrodynamics of massless fermions in 1+1-dimensions. As a consequence of the axial anomaly the photon, or more precisely the plasma oscillation, becomes a massive excitation. Exact solution of the continuum model[8] reveals that the exact solutions of the equations of motion are labeled by two angles, say $\phi(\epsilon)=2\pi\epsilon$ and Θ, but that only one linear combination of these angles has a physical significance. Discussion of the massive Schwinger model by Coleman et al.[9] revealed that one of the parameters represents a background electric field; however, the significance of the second angle remained less clear.

Actually, the question of whether physics depends on any linear combination of the angles defining the exact solutions can only be answered if one carefully specifies the algebra of observables along with the Lagrangian or Hamiltonian of the system. In fact, if one specifies the algebra of observables to contain all gauge invariant operators which are invariant with respect to global chiral transformations, then all of the different solutions to the continuum model are unitarily equivalent. If, however, one allows all gauge invariant observables without regard to their properties under chiral transformations, then the set of continuum solutions form a 1-parameter family of inequivalent solutions. (Allowing general gauge invariant observables amounts to permitting mass operators as observables even if you do not add them to the Hamiltonian.)

3. FORMALISM OF THE LATTICE SCHWINGER MODEL

3.1 SOME DEFINITIONS

The Hamiltonian of the lattice Schwinger model is

$$H = \frac{g^2}{2}\sum_{\ell} E_\ell^2 - i\sum_{\{j_1,j_2\}} \Psi^+(j_1)\, \sigma_3\, \delta'(j_1-j_2)\, \Psi(j_2)\, \exp(i{\sum}' A_\ell) \tag{3.1}$$

where the directed sum, $\sum'$, is the sum over all links joining j_1 to j_2, and is taken with a plus sign if j_1 is to the left of j_2 and with a minus sign if j_2 is to the left of j_1. The operators E_ℓ and A_ℓ are taken to be conjugate harmonic oscillator variables associated with each link ℓ, and the field $\Psi(j)$ is a two component fermi field associated with each vertex of our one dimensional lattice. The function $\delta'(j_1-j_2)$ is the DWY-derivative and is given by

$$\delta'(j_1-j_2) = \frac{1}{2N+1}\sum_k ik \exp[\, ik(j_1-j_2)\,]$$

where the variable k runs over

$$k = \frac{2\pi n}{2N+1}\,, \quad -N \leq n \leq N \tag{3.2}$$

and the gamma matrices for the one dimensional theory are 2×2 Pauli spin matrices chosen so that

$$\gamma_0 = \beta = \sigma_1$$
$$\gamma_1 = -i\sigma_2$$
$$\alpha = \gamma_5 = \gamma_0\gamma_1 = \sigma_3 \qquad (3.3)$$

It is convenient to write the two component spinor field $\Psi(j)$ in terms of fermion annihilation and creation operators as

$$\Psi(j) = b(j) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + d^{+}(j) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (3.4)$$

and its Fourier transform is conventionally written as

$$\Psi(k) = \frac{1}{\sqrt{2N+1}} \sum_j \exp[-ikj]\ \Psi(j)$$
$$= b(k) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + d^{+}(-k) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (3.5)$$

The factor of (2N+1) stands for the number of sites on the lattice and is related to L, the length of the lattice in dimensionful units, and to the maximum momentum cutoff, Λ, by the condition

$$L = \frac{(2N+1)}{\Lambda} = (2N+1)a \qquad (3.6)$$

where $a = \Lambda^{-1}$ is the lattice spacing.

Adopting these definitions one can rewrite the Hamiltonian for the free fermi field (see (3.1) without the exp [$i\sum A_{\mathfrak{X}}$] factors) as,

$$H = \sum k\ [\ b^{+}(k)b(k) - d^{+}(k)d(k)\ ] \qquad (3.7)$$

We also observe that the Hamiltonian commutes with the number operators

$$N_b = \sum_k b^{+}(k)b(k) = \sum_j b^{+}(j)b(j)$$

and

$$N_d = \sum_k d^{+}(k)d(k) = \sum_j d^{+}(j)d(j) \qquad (3.8)$$

If one defines the total electric charge operator to be

$$Q = \tfrac{1}{2}\sum_j \left[\Psi^+(j),\Psi(j)\right] \tag{3.9}$$

and the total chiral charge as

$$Q_5 = \tfrac{1}{2}\sum_j \left[\Psi^+(j),\gamma_5\Psi(j)\right] \tag{3.10}$$

then it is easy to rewrite these operators in terms of N_b and N_d as

$$Q = N_b - N_d$$

and

$$Q_5 = N_b + N_d \tag{3.11}$$

The last piece of general information which we need for our discussion is that the groundstate of the free fermion theory is

$$|\Phi_{vac}\rangle = b^+(0)d^+(0)b^{(}(-k_1)d^+(k_1)\dots b^+(-k_n)d^+(k_n)|0\rangle \tag{3.12}$$

where $|0\rangle$ is the state annihilated by all of the operators b_k and d_k, and where the momenta k_n are defined to be $k_n = 2\pi n/(2N+1)$. All other charge zero states having finite energy are obtained by adding or removing pairs of b's and d's having momenta small on the scale of the cutoff Λ.

3.2 GAUGE INVARIANCE

In defining the Hamiltonian, (3.1), we have implicitly defined the theory in $A_0=0$ gauge. Hence, the Maxwell equation $\nabla\cdot E-\rho=0$ is not one of the operator equations of motion. However, as defined, the theory possesses an invariance with respect to arbitrary time independent gauge transformations wherein

$$A(j) \rightarrow A(j) + \alpha(j+1) - \alpha(j)$$

and

$$\Psi(j) \to \exp(-i\alpha(j))\Psi(j) \qquad (3.13)$$

where we use the notation A(j) to denote the field associated with the link joining the points j and j+1. In order to have the missing Maxwell equation hold as an operator equation of motion one must restrict attention to the subspace of gauge invariant states, ie., those states $|\phi\rangle$ for which $[G(j) \equiv -\nabla \cdot E(j) + \rho(j)]|\phi\rangle = 0$. Actually, this point needs more careful discussion and we will return to it in a moment.

3.3 THE GIVM IDEA

While, in order to carry out a physically meaningful variational (or for that matter, perturbative) calculation, one must restrict attention to gauge invariant states, constructing wavefunctions which satisfy these constraints is a truly formidable task for any but the simplest gauge theories. Horn and Weinstein[1] suggested that one can avoid this difficulty by choosing for a trial state any arbitrary function $|\Phi(..,A(j),..)\rangle$ and then projecting it onto its gauge invariant part, $P|\Phi\rangle$. The projection operator, P, is defined as

$$P \equiv \prod_j P_0(j)$$

where

$$P_0(j) = N_0 \int d\alpha(j) \exp(i\alpha(j)G(j)) \qquad (3.14)$$

where N_0 is a normalization operator defined so that $P^2 = P$. Since $|\Phi\rangle$ is an arbitrary wavefunction depending upon a set of variational parameters, $\{\lambda_a\}$, $P|\Phi\rangle$ will in general be a gauge invariant trial wavefunction depending upon the same set of parameters. However, in general the norm of $P|\Phi\rangle$ will not be unity. For this reason the variational parameters are to be determined by minimizing the functional

$$\mathcal{E}(\lambda_a) \equiv \frac{\langle\Phi|PHP|\Phi\rangle}{\langle\Phi|P^2|\Phi\rangle} = \frac{\langle\Phi|HP|\Phi\rangle}{\langle\Phi|P|\Phi\rangle} \qquad (3.15)$$

This is a simple idea which at first glance seems to be impossible to carry out. The gist of the arguments presented in Ref.1 is that, surprisingly, the necessary manipulations can be carried out in detail for a large class of interesting wavefunctions $|\Phi\rangle$. In the discussion to follow, we will only use this idea in order to provide us with and organizing principle for what follows, the reader will not need any more information than that which has already been presented.

3.4 GAUGE PROJECTING: A SIMPLE PHYSICS QUESTION

In the preceding sections we noted that in order to have $\nabla\cdot E-\rho = 0$ hold as an operator equation of motion we have to restrict attention to gauge invariant states. Actually, one has to be a little more careful in arriving at this conclusion. The point is that one wants this equation to be true because one is quantizing the classical theory of electromagnetic interactions, and this equation is true for the classical theory. Note however, that in discussing the classical theory one does not assume that this equation holds at the boundaries of the system because the classical equations of motion do not come with boundary conditions. Boundary conditions must be arrived at from other physical considerations. In this case choosing boundary conditions amounts to deciding whether there are classical charges at the walls of the system. Since gauge invariance is assumed to be a superselection rule (ie., not only is the Hamiltonian assumed to be gauge invariant, but one also assumes that all physical observables are also gauge invariant) it follows that to uniquely define the theory one must first diagonalize all of the G(j)'s. We have already noted that the eigenvalues of G(j) must be zero for all j's in the interior, but for $j=\pm N$ any eigenvalues are possible. In other words, one is free to restrict attention to states for which

$$G(\pm N)|\Phi\rangle = \epsilon(\pm N)|\Phi\rangle \ . \tag{3.16}$$

The generators of gauge transformations for $-N < j < N$ are the operators

$$G(i) = - E(i) + E(i-1) + \rho(i) \tag{3.17}$$

where, $\rho(i)$ is the local charge density operator

$$\rho(i) = n_b(i) - n_d(i) \ . \tag{3.18}$$

The gauge generators for the two endpoints, $i = -N$ and $i = N$ are given by

$$G(-N) = - E(-N) + \rho(-N)$$

and

$$G(N) = E(N-1) + \rho(N) \ . \tag{3.19}$$

Hence, the most general projection operator we can use is

$$P(\epsilon(-N),\epsilon(N)) = \int d\alpha(-N)\int d\alpha(N) \ \exp(i\alpha(-N)(G(-N)-\epsilon(-N))$$
$$\exp(i\alpha(N)(G(N)-\epsilon(N)) \prod_{\{int\}} P_0(j) \tag{3.20}$$

where the product of the operators $P_0(j)$ is taken only over the interior points of the lattice. Also, since we are interested in the sector for which the total electric charge vanishes, then

$$\sum_j G(j) = Q \tag{3.21}$$

implies that for the operator $P(\epsilon(-N),\epsilon(N))$ to project onto a state of $Q = N_b - N_d = 0$ one must choose

$$\epsilon = \epsilon(-N) = - \epsilon(N) \tag{3.22}$$

For this reason our two parameter family of projection operators is reduced to a one parameter family. The label ϵ signifies the existence of a background electric field which enters from the left hand boundary of the lattice and leaves from the right.

In order to make the correspondance between the theory formulated on an open lattice and that formulated on a closed lattice with an equal number of sites and links (ie., a ring) as simple as possible, it is convenient to define the one parameter family of projection operators as follows

$$P(\epsilon') = \int d\delta_t \exp(i\epsilon'\delta_t)\, S(\delta_t)$$

where the operator $S(\delta_t)$ can be written as

$$S(\delta_t) = \exp\left[i\frac{2N}{2N+1}\,\delta_t\,\tfrac{1}{2}(G(N) - G(-N)\right]$$

$$\times\int d\alpha' \exp(\tfrac{1}{2}i\alpha'(G(N)+G(-N)))\prod_{\{int\}} P_0(j) \qquad (3.23)$$

If we think of carrying out the indicated $\alpha',\alpha(j)$ integrations as summing over configurations $\alpha(j)$ with fixed boundary conditions, in order to make things look as they would in the continuum path integral formulation, then we can write

$$S(\delta_t) = \sum_{\{\alpha(j)\}} U(\alpha(j)) \qquad (3.24)$$

where $U(\alpha(j))$ is the unitary transformation

$$U(\alpha(j)) = \exp(i\sum \alpha(j)G(j)\ ) . \qquad (3.25)$$

and the sum over configurations $\{\alpha(j))\}$ in (3.24) runs over functions $\alpha(j)$ such that

$$\alpha(N) - \alpha(-N) = \frac{2N}{(2N+1)}\,\delta_t \qquad (3.26)$$

It is straightforward to verify that with this definition the variable ϵ' differs from the background field by a factor of 2N/(2N+1), ie.,

$$\epsilon = \frac{(2N+1)}{2N} \epsilon' \tag{3.27}$$

and so, it can be identified with the background field only in the limit $L = (2N+1)a \to \infty$. Since this is the limit of interest we will ignore this distinction in what follows.

3.5 BACKGROUND FIELDS AND PAIR PRODUCTION

We will now recapitulate an argument due to Coleman which shows that in the infinite volume limit there is no discernable difference in the local physics in sectors of the theory corresponding to background fields ϵ_1 and ϵ_2 if $(\epsilon_2-\epsilon_1) = n$ where n is an arbitrary integer. To see why this is the case let us focus attention on a theory without dynamical fermions and ask what happens to the energy of a pair of massive external sources when one has a constant background field.

In the theory without dymanical fermions the Hamiltonian is

$$H = \frac{g^2}{2} \sum E(j)^2 \tag{3.28}$$

and so the energy in the presence of a background field ϵ is simply $\frac{1}{2}g^2\epsilon^2 L$. If one now introduces a pair of unit charges (ie., charge one in units of the coupling g) so that the field between the charges is reduced to $(\epsilon - 1)$, then letting the distance between the charges be s and the length of the world L, the energy of this static configuration is

$$\mathcal{E}(s) = \tfrac{1}{2}g^2\epsilon^2(L-s) + \tfrac{1}{2}g^2 s(\epsilon -1)^2 \tag{3.29}$$

Differentiating $\mathcal{E}(s)$ with respect to s we obtain

$$\frac{d\mathcal{E}}{ds} = \tfrac{1}{2}g^2 (1 - 2\epsilon) \tag{3.30}$$

which is a positive quantity if $|\epsilon| < \frac{1}{2}$. It follows from (3.30) that if $|\epsilon|$ is greater than $\frac{1}{2}$ it becomes energetically favorable for the pair of charges to separate and move to the end of the lattice reducing the strength of the background field by one unit. When dynamical fermions are introduced this pair production process continues until the background field penetrating to the center of the lattice has been reduced until it lies within the range $-\frac{1}{2} \leq \epsilon \leq \frac{1}{2}$. At this point any further screening of the background field is due to an effective dielectric constant of the vacuum, but the pair production mechanism has shut off.

With this in mind we see that except for charges at infinity, which play no role in determining the local physics, there is no difference between sectors corresponding to values of the background field ϵ which differ by integers. This says that for the purpose of carrying out a variational calculation it suffices to work in a linear combination of states whose ϵ-values differ by an integer. To be precise, one can restrict attention to $-\frac{1}{2} < \epsilon < \frac{1}{2}$ and define the periodic projection operators

$$P_{per}(\epsilon) \equiv \sum_n P(\epsilon + n) \tag{3.31}$$

Substituting this into (3.24) and carrying out the sum on n we obtain

$$P_{per}(\epsilon) = \sum_m \exp(2\pi\epsilon m)\, S(2\pi m) \tag{3.32}$$

It will be convenient in what follows to observe that every function $\alpha(j)$, for which $\alpha(N) - \alpha(-N) = 2N\delta_t/(2N+1)$, can be written as

$$\alpha(j) = \alpha'(j) + \kappa j \tag{3.33}$$

where $\kappa \equiv \delta_t/(2N+1)$. Hence, we can rewrite the formula for $S(2\pi m)$ as

$$S(2\pi m) = S_0\, U_m$$

where the unitary operator U_m is defined to be

$$U_m = \exp(i\sum \kappa_m j G(j)) \tag{3.34}$$

and where κ_m is defined to be

$$\kappa_m \equiv \frac{2\pi m}{(2N+1)} \tag{3.35}$$

This allows us to rewrite $P_{per}(\epsilon)$ in the form which will be most useful to us, namely

$$P_{per}(\epsilon) = S_0 \sum_m \exp(i\phi(\epsilon)m)\, U_m \tag{3.36}$$

where we have defined the angular variable $-\pi \leq \phi(\epsilon) = 2\pi\epsilon \leq \pi$.

This completes our discussion of the ϵ parameter and it should come as no surprise at this point that $\phi(\epsilon) = 2\pi\epsilon$ is one of the angles appearing in the exact solution to the Schwinger model

Before concluding this section it is worth pointing out that if one had worked on a periodic lattice with an equal number of links and sites we would have arrived at exactly the same formulae. In this case a gauge transformation is really defined by giving a link function $\nabla\alpha$ and its associated fermionic phase factor

$$z(j) \equiv \exp(i\sum_{j=j_0}^{j-1} \nabla\alpha(l)) \tag{3.37}$$

In this case the only requirement one has on the link function $\nabla\alpha(j)$ is that (3.37) defines a real periodic function $z(j)$, which means that $\sum \nabla\alpha = \delta_t = 2\pi m$. Hence, for the periodic lattice we can rewrite the any gauge function as the sum of one for which $\delta_t = 0$ plus one for which $\nabla\alpha = 2\pi m/(2N+1)$. It is easy to see that this leads to the same projection operator formalism.

4. CHIRAL CHARGES, REGULATION AND ANOMALIES

4.1 Q_5 ON THE LATTICE AND THE PROBLEM OF NORMAL ORDERING

In Eq. (3.10) we defined the global chiral charge as a sum over sites of the local chiral charge density $q_5(j)$, where

$$q_5(j) \equiv \tfrac{1}{2} \left[\psi^{+}(j),\gamma_5\psi(j)\right] \tag{4.1}$$

If one Fourier transforms this expression into momentum space one obtains

$$Q_5 = \sum_{k} \left[n_b(k) + n_d(k) \right]$$

where the operators $n_b(k)$ and $n_d(k)$ are defined to be the number operators

$$n_b(k) \equiv b^{+}(k)b(k)$$

and

$$n_d(k) \equiv d+(k)d(k) \tag{4.2}$$

Although these operators appear to be normal ordered, they are not. This is true even for the case of free field theory, since the groundstate of the free field theory is not the state $|0\rangle$, which is annihilated by all of the b's and d's, but rather the state $|\Phi_{vac}\rangle$ defined in (3.12) . Hence, in the limit of the cutoff going to infinity the charge Q_5 as defined in (4.2) does not converge to a finite operator. As we will see, it is the attempt to define a finite operator in this limit which explains the structure of the anomaly. Before proceding with formal considerations, let us as what happens to a general trial ground state for the system when we perform a gauge transformation. In particular, let us focus attention upon what happens when we make the gauge transformation defined by the function $\alpha(j) = 2\pi nj/(2N+1)$ for any integer n.

The discussion to follow will apply to variational functions of the general form

$$|\Psi_{trial}\rangle = \sum_n |\Psi_{gauge}\rangle_n \times |\phi_{fermion}\rangle_n \tag{4.3}$$

where the states $|\Psi_{gauge}\rangle_n$ are taken to be arbitrary functions of the gauge fields, and the states $|\phi_{fermion}\rangle_n$ are restricted in two ways. First, it is assumed that they have total electric charge Q=0; and second, it is assumed that they are arbitrary linear combinations of massless free field eigenstates, whose energy differs from the free fermion vacuum energy by an amount which stays finite as the lattice mass, Λ, is taken to infinity. From the point of view of perturbation theory these restrictions allow for an essentially complete set of physical states. In order to simplify our presentation we will present the arguments to follow for the case of a variational wavefunction which consists of a single product state of the form

$$|\Phi_{var}\rangle = |\Psi_{gauge}\rangle \times |\Phi_{vac}\rangle \tag{4.4}$$

where the gauge part of the wavefunction is arbitrary. Having chosen such a trial state the gauge invariant variational problem is reduced to operating upon this state with the projection operator $P(\epsilon)$. Since the Hamiltonian is gauge invariant the expectation value of the Hamiltonian in the state $U_n|\Psi_{trial}\rangle$ is the same as the expectation value of the Hamiltonian in the state $|\Psi_{trial}\rangle$, where the operator U_n is the unitary transformation defined in (3.34) . Although we will focus in what follows on what happens to the fermionic part of the trial function, it pays to remember that the concomitant transformation of the gauge field part of the function is always understood to be taking place so as to keep the mean energy of the state unchanged.

What happens to the state $|\Phi_{vac}\rangle$ when one operates upon it with the gauge transformation U_n? In particular, what happens when we operate with U_1, U_1 being the gauge transformation generated by the function $\alpha(j)=2\pi j/(2N+1)$? The easiest way to see what happens is to observe that the fermion field transforms under U_1 as follows;

$$U_1 \Psi(j) U_1^+ = \exp(i2\pi j/(2N+1)) \Psi(j) \tag{4.5}$$

If we rewrite

$$\Psi(j) = \frac{1}{\sqrt{2N+1}} \sum \exp(ikj)\, b(k) \begin{pmatrix}1\\0\end{pmatrix} + \exp(-ikj)\, d^+(k) \begin{pmatrix}0\\1\end{pmatrix} \tag{4.6}$$

we see that (4.5) can be rewritten as

$$U_1 b(k) U_1^+ = b(k+2\pi/(2N+1))$$

and

$$U_1 d(k) U_1^+ = d(k-2\pi/(2N+1)) \ . \tag{4.7}$$

Recalling that the ground state of the 1+1-dimensional massless free fermion is the state

$$|\Phi_{vac}\rangle = \prod_{(n=0,N)} b^+(-k_n) d^+(k_n)\ |0\rangle \tag{4.8}$$

where the momentum $k_n \equiv 2\pi n/(2N+1)$, it follows from (4.7) that the state $|0\rangle$ is left invariant under the action of U_n. Hence, we see that

$$U_1|\Phi_{vac}\rangle = b^+(k_1)d^+(-k_1) \prod_{(n=0,N-1)} b^+(-k_n)d^+(k_n)\ |0\rangle \tag{4.9}$$

Examination of (4.9) reveals that the new state has one extra essentially zero momentum pair (ie., $k_1=2\pi/(2N+1)$ which vanishes in the infinite volume limit) and which has lost a pair at essentially infinite momentum, ie., $k=2\pi N/(2N+1)$. Since the total number of pairs is unchanged, we see that the state $U_n|\Phi_{vac}\rangle$ has exactly the same Q_5 value as the state $|\Phi_{vac}\rangle$. This is as it should be since the charge Q_5 as defined is gauge invariant and non-anomalous. However, the structure of (4.9) already contains the answer to the question of whether one can understand the anomaly in a lattice gauge theory. Note, that if one returns to dimensionful units and removed the momentum cutoff, then one would be in an ambiguous situation. In that case, we still have to fill the negative energy sea, however there is now confusion about what happens when we apply U_1 (or U_n) to such a state. The reason for this is that U_1 promotes each pair $b^+(k)d^+(k)$ by a unit of momentum, creating an additional low momentum pair; however, the question arises of whether or not one has a vacant state at infinite momentum. By cavalierly taking the cutoff to infinity and ignoring divergences one has wound up with the problem of the Hilbert hotel with an infinite number of rooms. In this case, even if the hotel is filled one can always make room for more guests by simply moving everyone up one room; if a guest leaves, the hotel is still filled if one simply moves everyone down one room. From this point of view, applying U_1 to the vacuum state of the theory defined by naively taking the cutoff to infinity creates only a low momentum pair and so changes Q_5 by two units.

The key to understanding what is going on is the observation that our definition of Q_5 does not yield a finite operator in the continuum limit. The simplest way to deal with this problem is to copy what is done in the Casimir problem by using an energy cutoff and defining a regulated chiral charge, $q_5(\gamma)$, which will be finite in the limit of the cutoff going to infinity. For example,

$$q_5(\gamma) \equiv \sum_k \exp(-\gamma E(k)^2)\ [n_b(k) + n_d(k)\ ]\ , \tag{4.10}$$

where E(k)=k for the DWY-gradient, and E(k)=sin(k) for the typical doubling prescription. The virtue of this regulated charge is that it is finite if $\gamma/(2N+1)^2$ is held fixed as the ultraviolet cutoff, and hence N, is taken to infinity. Using this definition one can take differences between the expectation value of $q_5(\gamma)$ in a state $|\Psi\rangle$ and in the state $U_1|\Psi\rangle$. This is guaranteed to be a finite quantity. One can then ask how this difference behaves as the regulator is taken away. It doesn't take much to convince oneself that this leads to the same conclusion as the Hilbert hotel analysis; namely, that if one defines

$$\delta q_5 \equiv \lim_{\gamma\to 0} \delta q_5(\gamma) \qquad (4.11)$$

then, when one applies U_1 to a state $|\Psi\rangle$, δq_5 is two. If one applies U_n to a state $|\Psi\rangle$, δq_5 is 2n. It follows from this result that $q_5(\gamma)$ is not a gauge invariant operator for any value of γ, since it changes under the gauge transformation U_n. Let us now establish the relationship between an energy regulator and point splitting in the continuum. Before proceeding to this discussion however, we should note that the conclusion about the way q_5 behaves under a gauge transformation is only true if one adopts a definition of the fermionic gradient which is of the DWY form. Clearly, an energy cutoff only removes the high-momentum states from consideration if there is no spectrum doubling. If there is spectrum doubling the change in occupation number at k near 0 and k near π are both counted; in which case the total change in both Q_5 and q_5 is zero and the theory is anomaly free.

4.2 ENERGY CUTOFFS AND POINT SPLITTING

To begin, let us define the quantities appearing in (4.9) more carefully, so as to make to passage to dimensionful units explicit. As we have written (4.9) the quantities appearing have all been scaled by appropriate powers of the lattice spacing so as to be dimensionless. Hence, the factor 2N+1 which stands for the number of lattice sites, is equal to the length of the lattice, L, divided by the lattice spacing, a. Furthermore, the momenta, $k_p = 2\pi p/(2N+1)$, are equal to the dimensionful momenta $k'_p a$; ie.,

$$k'_p = \frac{k_p}{a} = \frac{2\pi p}{L} \tag{4.12}$$

In the same way, the limits in the dimensionful momenta are $\pm 2\pi N/(2N+1)a$, and if one defines the continuum limit as taking $a \to 0$ with L held fixed we see that in this limit $N \to \infty$ as a^{-1}. Hence, in the limit of small a, or large $\Lambda = a^{-1}$, the bounds on the k'_p sums go to $\pm \pi\Lambda$. Finally, in order for the energy cutoff to defined a finite operator $q_5(\gamma)$ as $a \to 0$, it is clear that γa^2 must be held fixed as $a \to 0$. With this set of definitions we see that for the tranformation U_1, in the limit $a \to 0$, $\delta q_5(\gamma a^2)$ is

$$\delta q_5(\gamma a^2) = 2\exp(-\gamma a^2 4\pi^2/L^2) - 2\exp(-\gamma a^2 4\pi^2/L^2 a^2) \tag{4.13}$$

and so we see that first taking the continuum limit $a \to 0$ and then taking $\gamma a^2 \to 0$ we obtain $\delta q_5(\gamma a^2 \to 0) = 2$.

With this in mind Fourier transform (4.10) and rewrite everything in dimensionful variables in configuration space. The result of this exercise is

$$q_5(\gamma a^2) = \int dx \int dy \ (1/4\pi\gamma a^2)^{1/2} \exp(-(x-y)^2/4\gamma a^2) \ \tfrac{1}{2}[\Psi(x)^+, \gamma_5 \Psi(y)] \tag{4.14}$$

It is immediately clear from the fact that operators at different points x and y appear in the definition of $q_5(\gamma a^2)$ that the regulated operator is not invariant under an arbitrary gauge transformation which maps $\Psi(x)$ into $\exp(i\alpha(x))\Psi(x)$. (The difference in phase factor here is due to the fact that in order to compute q_5' we have taken $U(\alpha)$ off a state and applied it to the operator as $U^+\Psi U$ not $U\Psi U^+$.) In fact, under such a transformation $q_5(\gamma a^2)$ goes to

$$q_5'(\gamma a^2) = \int dx\int dy\ (1/4\pi\gamma a^2)^{1/2}\exp(-(x-y)^2/4\gamma a^2)\ \exp[i(\alpha(x)-\alpha(y)]\ \tfrac{1}{2}[\Psi(x)^+,\gamma_5\Psi(y)] \tag{4.15}$$

Now what happens to $q_5'(\gamma a^2)$ in the limit $\gamma a^2\to 0$.

In order to answer this question we observe that the point split operators $\tfrac{1}{2}[\Psi(x),\gamma_5\Psi(y)]$ are finite for x different from y, but become singular in the limit $x\to y$. However, as is well known, for a two dimensional theory this singularity is simply a normal ordering term. In other words, the normal ordered operator has finite matrix elements in all finite energy states. Hence, it is possible to rewrite (4.14) as

$$q_5'(\gamma a^2) = \int dx\int dy\ (1/\pi\gamma a^2)^{1/2}\ \exp[i(\alpha(x)-\alpha(y))]\ \exp(-(x-y)^2/\gamma a^2)\left\{\ \tfrac{1}{2}N([\Psi(x),\gamma_5\Psi(y)]) + s(x-y)\ \right\} \tag{4.16}$$

where, $N(\ [\Psi(x),\gamma_5\Psi(y)]\ )$ is the finite part of the commutator and $s(x-y)$ is the singular c-number function obtained by taking the Φ_{vac}uum expectation value of the commutator. A trivial manipulation gives

$$s(x-y) = \frac{1}{4\pi}\int dk_1\int dk_2\ \left\{\exp(ik_1x)\exp(-ik_2y)\ \langle\Phi_{vac}|[b^+(k_1),b(k_2)]|\Phi_{vac}\rangle\right\}$$

$$- \exp(-ik_1x)\exp(ik_2y)\langle\Phi_{vac}|[d(k_1),d^+(k_2)]|\Phi_{vac}\rangle\ \} \tag{4.17}$$

Remembering that the vacuum state has a half-filled sea, we see that

$$s(x-y) = \frac{2}{i2\pi(x-y)}(1 - \exp(-i\pi\Lambda(x-y)) - \delta(x-y) \qquad (4.18)$$

where the term $\exp(-i\pi\Lambda)$ reflects the momentum cutoff which is to be taken to infinity. Since $s(x-y)$ is to be integrated against a smoothly varying function of x and y, $\sin(\pi\Lambda(x-y))/\pi(x-y)$ is a representation of the delta function, and the term proportional to $\cos(\pi\Lambda(x-y))$ oscillates away as $\Lambda\to\infty$, we will forget it in what follows and use

$$s(x-y) = \frac{1}{i\pi(x-y)} \qquad (4.19)$$

If we substitute this form for $s(x-y)$ and expand the phase factor $\exp(i(\alpha(y)-\alpha(x))$ as a power series in $(x-y)$ we see that as $\gamma a^2\to 0$ $q_5'(\gamma a^2\to 0)$ becomes

$$q_5'(\gamma a^2\to 0) = q_5(\gamma a^2\to 0) + \frac{1}{\pi}\int dx \frac{d\alpha}{dx}(x) \qquad (4.20)$$

Hence, we see that if the total change in the function $\alpha(x)$ is zero then $q_5(\gamma\to 0)$ does not change under the corresponding gauge transformation. If the function $\alpha(x)$ generated one of the transformations U_n, so that $\int dx\alpha'(x)=2\pi n$ then it follows that $q_5(\gamma a^2\to 0)$ changes by 2n. This is exactly the result which we obtained from our momentum space arguments. If one defines $q_5(\gamma a^2\to 0)$ to be a finite operator by subtracting the function $s(x-y)$ from the point split expression, the same result holds since the change in q_5 is a finite shift coming from the phase factor multiplying the singular operator, clearly no such factor multiplies the c-number subtraction term.

This concludes our discussion of the relationship between δq_5 under a gauge transformation as computed with an energy cutoff and as computed from a point-splitting prescription. It should be clear to the reader that these arguments go beyond the use of the free field vacuum state as a trial wave-function. All that we have said for this state applies equally well to any state which has a finite energy difference from the free field ground state or any arbitrary linear combination of such states times arbitrary boson states.

Before concluding this discussion it is worth pointing out the factor of $1/\pi$ appearing in (4.20) is identifiable with the factor of g/π appearing as the coefficient of the anomaly in the continuum theory; where by the coefficient of the anomaly I mean that in the continuum theory the equation of motion of the axial current is $\partial_\mu j^{5\mu}=2(g/2\pi)E$. This can be done if one observes that we have adopted standard lattice conventions and absorbed a factor of g into the definition of the field A(j). If one undoes this transformation, the the gauge transformation on the fermionic part of the wave function is given by $\exp(ig\sum \rho(j))$, and the missing Maxwell equation becomes $\nabla\cdot E=g\rho$; so, (4.20) picks up a factor of g. The remaining identification comes from the fact that $d\alpha/dx$ is conjugate to the field variable E.

4.3 THE RELATION BETWEEN THE PARAMETER ϵ AND Q_5

In the preceding sections we discussed the formal question of defining an operator q_5 which could have a finite continuum limit. Let us now imagine that this has been done in such a way that both the q_5 and Q_5 value of our trial fermion state is zero. We will conclude this section with a discussion of what happens to a state $P(\epsilon)|\Psi_{trial}\rangle$ when one acts upon it with the operator $U = \exp(\frac{1}{2}i\theta q_5)$. Recalling that U_n

applied to an arbitrary trial state of definite q_5 changes its q_5 value by 2n, we see that $P(\epsilon)$ applied to the state $|\Psi_{trial}\rangle$ can be written as

$$P(\epsilon)|\Psi_{trial}\rangle = P_0 \sum \exp[i2\pi\epsilon(q_5/2)]\ |\Psi_{trial}\rangle \qquad (4.21)$$

It follows immediately from (4.21) that $\exp[i\theta(q_5/2)]$ applied to $P(\epsilon)|\Psi_{trial}\rangle$ is simply the state $P(\epsilon+\theta)|\Psi_{trial}\rangle$. In other words, making a global rotation by the gauge non-invariant chiral charge q_5 is equivalent to changing the parameter ϵ appearing in the projection operator. This result has as its immediate corollary one of the properties of the exact solution to the Schwinger model; namely, that if one restricts the algebra of observables of the model to the set of Q_5 conserving gauge invariant operators, then physics is independent of the value of the parameter ϵ.

Up to this point we have identified one angle, $\phi(\epsilon)$, of the continuum Schwinger model and the truly conserved operator Q_5, which has integer eigenvalues. However, continuum model is usually solved in terms of two angles and it is shown that the physics of the massless theory depends only upon the difference of these angles. In the next section of this paper we will introduce the second angle and how this continuum result comes about.

5. MASS TERMS AND Θ-PARAMETER IN MASSLESS FREE FERMION THEORY

In order to understand the meaning of the second angle appearing in the solution to the continuum Schwinger model it is necessary to go back and discuss the physics of massless free fermion theories with a little more care. Recall that the Hamiltonian of the lattice Schwinger model can be written as

$$H = \sum_p k_p \; (n_b(k_p) - n_d(k_p)) \tag{5.1}$$

where, in dimensionless units, $k_p = 2\pi p/(2N+1)$. We have already discussed the fact that in finite volume (ie., N finite) the groundstate of the theory corresponds to a half-filled sea. More precisely, we defined the groundstate of the theory to be the state $|\Phi_{vac}\rangle = \prod[b^{+}(-k_p)d^{+}(k_p)]|0\rangle$, where the product is taken over p=0,N. Actually, we should have noted that even for finite volume there is a two-fold ambiguity in the definition of the zero-charge state of lowest energy, since one does not change the energy of a state by adding or subtracting a zero-momentum pair. At first glance this would seem to be making much ado about nothing, but the situation becomes more interesting when one takes the limit $L=a(2N+1)\to\infty$ with a held fixed. The reason for this is that the energy of the state $|\Psi\rangle = b^{+}(k_p)d^{+}(-k_p)|\Phi_{vac}\rangle$ differs from the energy of the state $|\Phi_{vac}\rangle$ by $\delta E = 2\pi pa/L$, which vanishes in the limit $L\to\infty$ with pa held fixed. Note that the since this state has one extra fermion pair, it has a Q_5 value which differs from that of the state, $|\Phi_{vac}\rangle$, by two. Generalizing this discussion we see that in the infinite volume limit one can add any finite number of these effectively zero energy pairs to the groundstate without changing its energy. Moreover, with each addition of a zero energy pair one obtains a state with a Q_5 value which has increased by two units. One could also subtract an arbitrary finite number of zero energy pairs from the groundstate $|\Phi_{vac}\rangle$ and produce a state whose Q_5 is is negative with respect to the groundstate by any multiple of two units. From this it follows that in the infinite volume limit there are an infinite number of degenerate states with different eigenvalues of the operator Q_5.

This result suggests that we sum over a subset of these states to form states labelled by a parameter, θ, in the following manner:

$$|\theta\rangle = \sum \exp(i\theta n)\ |\Phi_{vac},\ Q_5=2X_{vac}+2n\rangle$$

$$= \sum \exp[i(\theta/2)Q_5]\ |\Phi_{vac},\ Q_5=2X_{vac}+2n\rangle \qquad (5.2)$$

where by $2X_{vac}$ we mean the eigenvalue of Q_5 for the massless free field vacuum state. Actually, this way of introducing the states $|\theta\rangle$ is ambiguous since, for exactly the same reasons, there are an infinite number of states of a given Q_5 which have the same energy as the state $|\Phi_{vac}\rangle$ in the limit $L\to\infty$.

In order to achieve a better understanding of what is happening in the limit $L\to\infty$ we will turn our attention to massive free fermion theory and study the limit in which the mass is taken to zero as the volume is taken to infinity. We will see that these limits do not commute with one another and this will allow us to make the definition in (5.2) more precise.

If one adds a term $m \sum \bar{\Psi}(j)\sigma_1\Psi(j)$ to the Hamiltonian (5.1) we obtain

$$H = \sum [k_p(b^+(k_p)b(k_p) - d^+(k_p)d(k_p))$$
$$+ m \sum (b^+(k_p)d^+(-k_p) + d(-k_p)b(k_p))] \qquad (5.3)$$

and so it is no longer true that the state $|\Phi_{vac}\rangle$ is the lowest eigenstate of this Hamiltonian. In order to find the true eigenstates of the Hamiltonian one must observe that there is no mixing among operators corresponding to different values of $|k_p|$ and so one must, for each value of $\pm k_p$ diagonalize the sub-Hamiltonian

$$(b^+(k),b^+(-k),d(k),d(-k))\ M(k,-k) \begin{pmatrix} b(k) \\ b(-k) \\ d^+(k) \\ d^+(-k) \end{pmatrix} \qquad (5.4)$$

where the matrix $M(k,-k)$ is the quadratic form

$$\begin{bmatrix} k & 0 & 0 & m \\ 0 & -k & m & 0 \\ 0 & m & k & 0 \\ m & 0 & 0 & -k \end{bmatrix} \tag{5.5}$$

Clearly the eigenvalues of this matrix are $\pm(k^2+m^2)^{1/2}$ and the ground state of the theory is obtained by filling the negative energy sea. In this case that amounts to forming the appropriate linear combinations of $b(k)$ and $d^{+}(-k)$ corresponding to the eigenvalues $-(k^2+m^2)^{1/2}$ and operating with the creation operators associated with these combinations upon the state $|0\rangle'$, which is the state annihilated by the new operators. The resulting state will not be an eigenstate of the operator Q_5 and therefore will have a projection

$$|\Phi_{vac}, Q_5=2X_{vac}+2n\rangle_m \equiv N_n \int d\theta \exp[i\theta(Q_5-2X_{vac}-2n)/2]\ |\Phi_{vac}\rangle_m \tag{5.6}$$

where $|\Phi_{vac}\rangle_m$ stands for the vacuum state of the theory with mass m, and N_n is a normalization factor. It will be these states which we will use in the (5.2) in the limit $m\to 0$.

It should be noted at this point that one will obtain different answers for $|\Phi_{vac}\rangle_m$ in the limit $m\to 0$ depending upon whether one has taken the limit $L\to\infty$ first. If one holds L (ie., N) fixed and takes $m\to 0$ one gets a linear combination of the two degenerate the groundstates of definite $Q_5=2X_{vac}$ and $Q_5=2X_{vac}-2$, which we have already discussed for the massless theory. This is because, in this case the momenta k appearing in the matrix $M(k,-k)$, except for the single value $k = 0$, have non-vanishing values $2\pi p/L$ and so for m small on the scale of these values the effect of m is purely perturbative and vanishes as $m\to 0$. However, if one takes the limit $L\to\infty$ first, then for any m, no matter how small, there are an infinity of small k-values for which the effects of the mass term dominate. This is the non-uniformity of limits which we alluded to before.

With this discussion behind us, we introduce Θ as in (5.2), defining the states $|\Phi_{vac}, Q_5 = 2X_{vac} + 2n\rangle$ as the $m \to 0$ limit of the states $|\Phi_{vac}, Q_5 = 2X_{vac} + 2n\rangle_m$, where the limit $L \to \infty$ has been taken first. Note, an equivalent definition of the state $|\Theta\rangle$ is to define it as the limit as $m \to 0$ of the groundstate of the theory for which the mass term has been chosen to be of the form

$$H_{mass}(\Theta) = m \sum [\ \cos(\Theta)\Psi^{+}(j)\sigma_1\Psi(j) + i\sin(\Theta)\Psi(j)(-i\sigma_2)\Psi(j)\ ] \tag{5.7}$$

Another way of writing $H_{mass}(\Theta)$ is

$$H_{mass}(\Theta) = \exp(i\Theta Q_5/2)\ H_{mass}(0)\ \exp(-i\Theta Q_5/2) \tag{5.8}$$

which explains the factor of $\Theta/2$ which appears in our definitions.

Having managed to introduced the parameter Θ as the dual to the integer eigenvalues of the operator Q_5, we can now use it to construct trial wave functions which depend upon both ϵ and Θ as follows;

$$|\Psi_{trial}\rangle \equiv P(\epsilon)\ |\Psi_{gauge}\rangle \times |\Theta\rangle \tag{5.9}$$

It only remains for us to show that physics can only depend upon the difference of these two parameters.

5.1 THE RELATIONSHIP BETWEEN THE ANOMALOUS AND CONSERVED CHIRAL CHARGES

At this point in our discussion it is relatively easy to understand why only the difference of the angles $\phi \equiv 2\pi\epsilon$ and Θ matters in the continuum Schwinger model, even when the algebra of observables is enlarged to include chirality changing gauge invariant operators. The crux of the matter lies in the fact that the states appearing in $|\Theta\rangle$

(ie.,(5.2)), and therefore in the trial state (5.9), differ from one another by the addition of zero momentum pairs. For each such pair the eigenvalue of Q_5 is increased (or decreased) by two units. It now pays to ask what happens to the value of $q_5(\gamma\to 0)$ when one adds a low momentum pair to a state? Since the definition of $q_5(\gamma a^2\to 0)$ only affects the counting of infinite momentum pairs one sees immediately that the addition of a low momentum pair to a state also increases $q_5(\gamma a^2\to 0)$ by two units. From this it follows that, up to a possible overall phase factor which has to do with the Q_5 value of the state we started with, applying the transformation $U \equiv \exp(-i\Phi q_5/2)$ to the state $P(\epsilon)|\Psi_{gauge}\rangle\times|\Theta\rangle$ changes both ϵ and Θ by the same amount. We have already seen the under this transformation $\epsilon \to \epsilon-\Phi$, and since the difference in Q_5 and q_5 between the states appearing in the definition of $|\Theta\rangle$ are the same, both operators rotate the Θ-variable in the same way too. Hence, we see that since in the limit of volume and cutoff going to infinity $q_5(\gamma=0)$ commutes with the Hamiltonian, one sees that computing expectation values in the states $|2\pi\epsilon,\Theta\rangle$, $|2\pi\epsilon-\Theta,0\rangle$ and $|0,\Theta-2\pi\epsilon\rangle$ must give the same results for the appropriately transformed operators.

6. LOCAL CHIRAL TRANSFORMATIONS AND THE SEIZING OF THE VACUUM

The next question which we will discuss relates to the issue of why the Schwinger model doesn't have a Goldstone boson, even though it has a spontaneously broken global Q_5 which is generated by a local, gauge-invariant, charge density $\rho_5(j)\equiv(n_b(j) + n_d(j))$. The answer, first proposed by Kogut and Susskind[5] is that the vacuum of the theory seizes. By seizing, they meant that although a global chiral rotation could be performed on the system, any local chiral transformation costs infinite energy (or at least energy on the order of the cutoff Λ). What we wish to do in this section of the paper is to show that one can easily understand the mechanism underlying this process.

Let us begin by considering an unprojected trial state

$$|\Psi_{trial}\rangle = |\Psi_{gauge}\rangle \times |\phi_{fermion}\rangle \qquad (6.1)$$

where the gauge field state is arbitrary and, for the sake of argument, the fermionic state $|\phi_{fermion}\rangle$ is the finite volume zero mass groundstate defined in (4.8) . The problem of interest relates to what happens to the expectation value of the Hamiltonian in the state

$$|\Psi'\rangle = \exp(i\sum\rho_5(j)\beta(j))\ |\Psi_{trial}\rangle \qquad (6.2)$$

for the case $\beta(j) = 2\pi pj/(2N+1)$. In particular we will be interested in the case p=1 since in the limit $L\to\infty$ ($N\to\infty$) this should generate the longest wavelength Goldstone excitation. This can be analyzed in the same way as we did for the case of a gauge transformation, since under this transformation the operator $\Psi(j)$ goes to $\exp(i\beta(j)\sigma_3)\Psi(j)$. Since $\beta(j) = \kappa j$ Fourier transforming this result tells us that

$$b(k)\to b(k+\kappa)$$

and

$$d(k)\to d(k+\kappa) \qquad (6.3)$$

Hence, the fermionic state

$$|\Phi_{vac}\rangle = \prod b^{\dagger}(-k_n)d^{\dagger}(k_n)|0\rangle$$

goes over to the transformed state

$$|\phi'\rangle = \prod b^{\dagger}(-k_n+2\pi/L)d^{\dagger}(k_n+2\pi/L)|0\rangle \qquad (6.4)$$

which one can readily convince oneself creates a charged pair at low momentum and a balancing oppositely charged pair at $k \propto \Lambda$). This is exactly the same sort of thing which we discussed when we analyzed what happened under the gauge transformations U_n, except that it is charge and not q_5 which has changes at low momentum balanced by pairs at high momentum. It follows, as it must, from (6.4) that the operators Q_5 and q_5 are left unchanged by this transformation. What then is different about this case?

The important point about this transformation is that it is not a gauge transformation. In the preceding discussion we were dealing with gauge transformations and so although the $q_5(\gamma)$ properties of the states appearing in the gauge projected trial state changed, the expectation value of the Hamiltonian in any one of these transformed states was always the same. That, in fact, was the justification for superimposing these states in order to form a trial state. If one kept the cutoff fixed, one could in fact show that the cross terms between the different terms in the sum lowered the energy of the state $\epsilon=0$. The reason why the expectation value of the Hamiltonian was the same in both the transformed and untransformed state was because the part of the trial wavefunction involving the gauge field transformed in such a way as to cancel the effects on the fermionic part of the trial state. Returning to the case at hand we see the significance of the fact that the local chiral rotation is not a gauge transformation; namely, that it does not cause any transformation of the gauge field part of the wavefunction and so under the local chiral rotation the expectation value of the Hamiltonian of the transformed state is changed. It is easy to estimate this change for the case of very small g, since it will simply be the expectation value of the free fermion Hamiltonian in a state which has two extra high momentum excitations. The difference in energy between the expectation value for the transformed state and the original state is on the order of the cutoff Λ. Hence, in the limit of $\Lambda\to\infty$ we see that except for the case $\kappa=0$, the energy of any long wavelength excitation created by the local chiral current density is infinite and so the vacuum seizes.

7. ANOMALY CANCELLATION AND GAUGING ANOMALOUS CURRENTS

In the previous sections of the paper we discussed the lattice version of the ordinary Schwinger model and showed how one could understand the properties of the continuum theory from general arguments. We will conclude our discussion with a brief treatment of what happens if one attempts to gauge a variant of this model wherein the current one gauges is anomalous. This happens, for example, if one gauges the current whose charge density is $\rho_+(j)$

$$\rho_+(j) \equiv \frac{c}{2}[\Psi^+(j),(1+\gamma_5)\Psi] \qquad (7.1)$$

where we have, for the sake of generality, assumed that the the field Ψ has charge c. Such a theory can be obtained from the Schwinger model if one rewrites the Hamiltonian as

$$H = \sum \frac{\delta'(j-l)}{i}\left\{b^+(j)b(l)\exp(ic{\textstyle\sum}'A_{\ell}) - d^+(j)d(l)\right\} + \frac{1}{2}\sum E(j)^2 \qquad (7.2)$$

In this case, only the operators b(k) transform under a large gauge transformation and they transform as before, ie.,

$$U_1^+\, b(k)\, U_1 = b(k+2c\pi/L) \qquad (7.3)$$

It follows immediately from (7.3) that the transformation U_1 creates c low momentum particles (not pairs in this case), and therefore from (7.1) we see that $\delta q_+(\gamma a^2 \to 0)$ is c^2. Hence, unlike the case of the Schwinger model the current which has for its time component the fermionic part of the generator of local gauge transformations is anomalous, and can only give rise to a conserved charge if one allows infinite momentum states to compensate for low momentum states under large gauge transformations.

Obviously, this model can be made back into a Schwinger model by adding to it a field which transforms with the same charge c, but which has a Hamiltonian

$$H_2 = - \sum \frac{\delta'(j-l)}{i} \left\{ B^+(j)B(l)\exp(ic{\textstyle\sum}'A_{\ell}) - D^+(j)D(l) \right\} . \quad (7.4)$$

From the point of view of gauge transformations the operators d(j) and D(j) are window-dressing and can be ignored. The minus sign in front of H_2 is significant, in that it requires that the neqative energy sea for the B's must be filled for positive instead of negative k_m. There is, however, another way to achieve cancellation of the anomaly. That is to introduce c^2 fields of unit charge. In that event the change in the total charge of the system under a U_1 transformation will be $+c^2$, coming from the fields b_j, and -1 coming from each of the fields of unit charge. Hence, it takes c^2 of the unit charge fields to cancel the anomaly produced by the charge c field. This is of course a well known result for the continuum theory.

The question which arises at this point is what happens if one does not cancel the anomaly in the $(1+\gamma_5)$ current by either of these two mechanisms? At first glance it would seem that the lattice theory is perfectly well defined. Is there any clue as to why it might not have a satisfactory continuum limit? After all, no one said that we had to consider regulated charges.

The obvious conjecture is that for theories where one has gauged an anomalous current the connection between the behavior of states at infinite momentum and zero momentum is made so strong by the requirement that one consider gauge invariant states, that the Lorentz invariance properties of the continuum limit cannot come out correct. Unfortunately, there is no simple way to make this argument. There is, however, one observation about the relationship of gauge invariance and the space-time properties of the Green's functions of the theory which can be made. Suppose one considers, for the case of $L=\infty$, or the case of a periodic lattice, the unitary transformation T_a which translates all operators by one lattice unit. In other words, for an arbitrary operator $O(j)$

$$T_a^{+}\, O(j)\, T_a = O(j+1) \tag{7.5}$$

Let us then consider the state $P(\epsilon)|\Phi\rangle$, for an arbitrary translation invariant wavefunction $|\Phi\rangle$. What is the state $T_a P(\epsilon)|\Phi\rangle$?

Recalling that $P(\epsilon)$ is

$$P(\epsilon) = S_0 \sum_n \exp\Big(i\sum_j \frac{2\pi n j}{L}\, G(j)\Big) \tag{7.6}$$

we see that $T_a P(\epsilon)|\Phi\rangle$ is equal to

$$\begin{aligned} T_a P(\epsilon)|\Phi\rangle &= S_0 \sum \exp\Big(i\sum \frac{2\pi n j}{L}\, G(j+1)\Big)\, |\Phi\rangle \\ &= \exp\Big(i\sum G(j)\Big)\, P(\epsilon)\, |\Phi\rangle \\ &= \exp(iQ_+)\, P(\epsilon)\, |\Phi\rangle. \end{aligned} \tag{7.7}$$

Hence, we see that in order for the gauge projected state to be translation invariant (up to a phase) the charge Q_+ must be conserved. If the gauged charge is anomalous, no regulated version of can be used to generate translations. This is because the conservation of the anomalous charge requires a detailed cancellation between momenta at the scale of the cutoff and zero momentum. On the other hand, if one cancels out the anomalies by either of the mechanisms discussed in this section, then the conservation of the non-anomalous charge comes from cancellations which are local in momentum space. Hence, simultaneous gauge and translation invariance no longer requires a relationship between the highest and lowest momenta in the theory. In this case one expects that one can proceed as in perturbation theory and remove the effects of the high momentum states with impunity.

At present I know of no stronger result which I can state which relates the Lorentz properties of the lattice gauge model, gauge invariance of the wavefunction and anomalies which does not require a detailed calculation to support this contention. However, this is an interesting question which merits further study.

This concludes our discussion of the way in which a chirally invariant form of the lattice Schwinger model can be used to derive properties of the continuum theory. Although we have focused on questions related to taking cutoffs to infinity, we hope that the reader has realized that this was because this limit is useful for explicitly evaluating certain expressions; the physics of the theory changed smoothly as one removed all cutoffs. In particular, the question of the ϵ and Θ parameters had nothing to do with taking the ultra-violet cutoffs to infinity It is also true that the massive photon is a property of the lattice theory with cutoffs in place, although it achieves a truly relativistic energy spectrum only in the limit $\Lambda \to \infty$. It is my hope that the simplicity and straightforward nature of this discussion will make it clear that not only is the physics of the chiral anomaly alive and well in the chirally symmetric theories formulated by means of the DWY-derivative, but that it can be highly illuminating to study such questions in this way.

8. WHAT ABOUT NO-GO THEOREMS?

We noted in the introduction to the paper that many authors have claimed to prove no-go theorems about the possibility of introducing fermions onto a lattice in such a way as to preserve the continuous chiral symmetries of the massless theory and the physics of the chiral anomaly. We hope that the discussion up to this point convinces the reader that, if one adopts the DWY-prescription for introducing fermions, then these claims are not true. The simplest of the arguments put forth to support these claims says that since the theory formulated using the DWY-derivative possesses an exactly conserved chiral charge, the physics of the Adler anomaly cannot be correctly obtained by taking the continuum limit of the theory. As we have shown, the anomaly is not a property of the operator, Q_5, which has no continuum limit; but rather, of the regulated and subtracted operator, $q_5(\gamma a^2 \to 0)$, which does.

What about the relevance of what we have referred to as technically correct no-go theorems?. These discussions prove a theorem which says that chirally symmetric fermionic Hamiltonians with only nearest neigbor (or more generally, finite range) couplings always exhibit spectrum doubling. While one cannot take exception to this result, the assumption that one should restrict attention to theories with only nearest neighbor, or finite range, couplings is an unnecessary one. It has already been shown[6] that there exists a natural transcription of free fermion theories to a lattice, making use of a fermionic gradient term which introduces long range couplings. This procedure allows one to avoid spectrum doubling and preserve the continuous chiral symmetries of the massless theory. Furthermore, it is easy to show that by exploiting the techniques introduced by Drell, Weinstein and Yankielowicz[6] one can do the same for fermions coupled to scalar and pseudo-scalar fields. In particular, it is possible to transcribe continuum theories which allow for spontaneous breaking of chiral symmetries, such as the SU(2) sigma model, to the lattice in such a way as to preserve the usual perturbation theory expansion. The only issue not discussed in Ref.6 was the question of how this prescription works for lattice gauge theories.

The work of Karsten and Smit[2] claiming that the DWY-derivative[5] causes problems when used for lattice quantum electrodynamics, has generated a great deal of confusion about the question of introducing fermions into lattice gauge theories. It has led many people to believe that there are grave problems where in fact none exist. In a recent paper Rabin[10] discussed the computations of Karsten and Smit[2] and showed that the troubles they encountered come from making the expansion, exp(igL)=1+igL+..., (where L stands for the linear dimension of the lattice in question) and then keeping only the first two terms, which is an obviously incorrect procedure in the limit $L\to\infty$. Although Rabin does not discuss the point in detail, the interested reader can quickly convince himself that it is possible to define an infinite number of lattice gauge theories using the DWY-prescription which preserve the structure of weak coupling perturbation theory. This set of counterexamples are sufficient to show that no-go theorems about the use of the DWY-derivative are not possible. Moreover, they show that the problem has to do with the way in which couplings to transverse photons are introduced, and not the way in which the fermionic gradient terms are handled. Rabin does discuss the more intricate question of how one can redefine the expansion of Karsten and Smit so that it is valid. He shows that by normal ordering the Hamiltonian and adding a set of gauge invariant counter-terms, of the same form as were already present in the original Hamiltonian, one can develop a convergent perturbation expansion which does agree with that of the continuum theory in the limit of zero lattice spacing. (The reader is referred to his paper for details.)

The situation can therefore be summarized as follows: since the DWY-derivative involves infinite range couplings in the Hamiltonian, it avoids the no-go theorems based upon the assumption that one only has finite range couplings in the Hamiltonian (or Lagrangian); and in addition, the criticism of Karsten and Smit based upon perturbation theory arguments is not relevant, since the perturbation expansion which they used can be trivially seen to be incorrect.

ACKNOWLEDGEMENTS

I would like to thank L. Susskind and J. Polchinski for many helpful discussions.

REFERENCES

1. David Horn and M. Weinstein, SLAC-PUB-2864; submitted to Phys. Rev. D.
2. Luuk H. Karsten and Jan Smit, Nucl. Phys., B183, 103, 1981; Luuk H. Karsten and Jan Smit, Phys. Lett. 85B, 100, 1979; Luuk H. Karsten and Jan Smit, Nucl. Phys. b144, 536, 1978.
3. H. B. Nielsen and M. Ninomiya, Phys. Lett. 105B, 219, 1981; H. B. Nielsen and M. Nincmiya, Nucl. Phys., B193, 173, 1981; H. B. Nielsen and M. Ninomiya, Nucl. Phys., B185, 20, 1981.
4. Paul H. Ginsparg and Kenneth G. Wilson, CLNS-81/520, Dec 1981 and references cited therein.
5. This concept is introduced for the first time in J. Kogut and Leonard Susskind, Phys. Rev. D11, 3594, 1975.
6. S. D. Drell, Marvin Weinstein and S. Yankielowicz, Phys. Rev. D14, 1627, 1976.
7. The fact that a long range gradient is a natural result of integrating out high momentum degrees of freedom is discussed in Helen R. Quinn and Marvin Weinstein, SLAC-PUB-2795, Sep 1981, submitted to Phys. Rev. D.
8. J. H. Lowenstein and J. A. Swieca, Annals Phys., 68, 172, 1971; earlier work on this subject is contained in J. Schwinger, Phys. Rev., 128, 2425, 1962; L. S. Brown, Nuovo Cimento, 29, 617, 1963; B. Zumino, Phys. Lett., 10, 224, 1964; C. R. Hagen, Nuovo Cimento, B51, 169, 1967.
9. Sidney Coleman, R. Jackiw and Leonard Susskind, Annals of Phys., 93, 267, 1975; Sidney Coleman, Annals of Phys., 101, 239, 1976.
10. Jeffrey M. Rabin, SLAC-PUB-2676, Feb 1981, submitted to Phys. Rev. D.