Undoubling Chirally Symmetric Lattice Fermions

Laurence Jacobs
Institute for Theoretical Physics, University of California at Santa Barbara, Santa Barbara, California 93106
(Received 18 April 1983)

A new formulation of lattice gauge theory with fermionic degrees of freedom is presented and analyzed. Arguments are given which imply that, unlike all previous local prescriptions, the present one allows the implementation of chiral symmetry without leading to the doubling of fermion species. This permits the lattice regularization of theories with handed fermions.

PACS numbers: 11.15.Ha, 11.30.Rd, 12.10.Ck

The lattice regularization of quantum field theory, introduced[1] by Wilson and by Polyakov, has made possible an impressive advance in our understanding of the underlying dynamics of gauge theories. The main reason for this is that, unlike all other methods devised to regularize the ultraviolet divergences of realistic field theories, the lattice provides a momentum cutoff which is both nonperturbative as well as gauge invariant. Although perturbation theory has led to important results in theories like quantum electrodynamics its usefulness in the study of strong-interaction phenomena is rather limited. For quantum chromodynamics (QCD), the current model for the strong interactions, a nonperturbative treatment is essential.

Once a quantum field theory has been defined through a lattice-regulated path integral, nonperturbative Monte Carlo methods can be successfully used to analyze its properties.[2] Indeed, in the last few years, the use of such methods has uncovered many important aspects of quantum chromodynamics.[3]

Although lattice theories describing the dynamics of gauge and scalar fields are by now rather well understood, the very definition of fermion fields on the lattice presents serious difficulties. Several clever techniques have been invented to solve the notorious problem of numerically simulating these theories,[4] but the conceptual problems associated with spinor lattice fields remain. Arguments have been produced which trace these difficulties to the geometrical nature of fermions. Because scalars and vectors carry tensor representations of the rotation group, a well defined prescription exists for writing down lattice theories for these fields which will have the right properties in the continuum limit; no such prescription exists for spinor representations of the rotation group.[5] These rules tell us that scalars are naturally associated with lattice sites and vectors with lattice links. They also tell us what kind of finite-difference operator will be the appropriate lattice version of a partial derivative acting on these kinds of fields. It seems clear that a truly satisfactory state of affairs for lattice fermions must await a deeper understanding of their geometric role.

The problem is how to transcribe the standard action for a fermionic field $\psi(x)$ of mass m minimally coupled to a gauge field $A_\mu(x)$. The continuum action reads

$$S = \int \bar{\psi}(iD + m)\psi, \tag{1}$$

where $D = \sum_\mu \gamma^\mu(\partial_\mu + igA_\mu)$ is the Dirac operator. The simplest lattice action corresponding to (1) is obtained by associating fermions with sites. To maintain the properties of (1) under Hermitian conjugation, $\partial_\mu\psi(x)$ is transcribed by use of the symmetric finite-difference operator,

$$\partial_\mu\psi(x) \to \Delta_\mu{}^s\psi_n = (2a)^{-1}(\psi_{n+\mu} - \psi_{n-\mu}), \tag{2}$$

where a is the lattice spacing, n is a site in the lattice, and $n+\mu$ is the next site in the μ direction. With use of the link variables[1] $U_{n\mu}$, the lattice form corresponding to (1) is given by

$$S = \sum_{n,\mu} \bar{\psi}_n\gamma_\mu\delta_\mu{}^s\psi_n + m\sum_n \bar{\psi}_n\psi_n, \tag{3}$$

where $\delta_\mu{}^s$ is the gauge-covariant version of $\Delta_\mu{}^s$ given by

$$\delta_\mu{}^s\psi_n = (2a)^{-1}(\psi_{n+\mu}U_{n\mu} - U_{n\mu}{}^\dagger\psi_{n-\mu}). \tag{4}$$

Notice that, when $m=0$, both (1) and (2) are invariant under the global chiral transformation

$$\psi \to \exp(i\gamma_5\theta)\psi, \quad \bar{\psi} \to \bar{\psi}\exp(i\gamma_5\theta). \tag{5}$$

This implies the existence of a conserved chiral current $j_\mu{}^5$. An immediate conflict arises. In the continuum theory, this current has an anomalous divergence: Whereas formal manipulations using the equations of motion imply that it is conserved, careful quantization shows that it is not.[6] On the lattice, on the other hand, the equations of motion are rigorously valid and can be used to prove the conservation of $j_\mu{}^5$ for any value of the lattice spacing. The lattice resolves this conundrum by creating extra states which cancel the anomaly. This degeneracy can be seen most easily by diagonalizing (3) in momentum space. With $U_{n\mu}$ set equal to 1, the inverse propagator is given by

$$K^{-1}(p)=a^{-1}\sum_\mu \gamma_\mu \sin p_\mu a+m, \quad -\pi/a< p_\mu < \pi/a. \quad (6)$$

As $a\to 0$ for fixed p_μ, $K^{-1}(p)$ has a nonvanishing limit when $ap_\mu \to 0$ as well as when $ap_\mu \to \pm\pi$ in any direction. This is the standard fermion-doubling problem. Careful perturbative analysis[7] shows that this multiplicity (2^d in d dimensions) indeed cancels the chiral anomaly. The degeneracy can be traced to a symmetry of (3) under $\psi\to T\psi$, $\bar\psi\to\bar\psi T^{-1}$, where $T=1$, $\gamma_\mu\gamma_5(-1)^{n_\mu}$, the so-called doubling symmetry. In momentum space, these transformations correspond to a shift in momentum by π/a (mod $2\pi/a$) in the μ direction. But the existence of the anomaly is crucial. In QCD, for example, it is responsible for the observed mass splitting between the η and π mesons. This difficulty was circumvented by Wilson[8] by adding to (3) a term which gives (momentum-dependent) masses to the extra states in such a way that the correct anomaly is recovered in the continuum limit.[7,8] However, in doing this, the lattice theory is no longer chirally invariant. This seems to preclude the lattice regularization of theories with handed fermions, like the Glashow-Weinberg-Salam model.

Species doubling is not an exclusive property of the particular form chosen for (3). Recently, in fact, a theorem has been proven[9] which shows that, in accordance with the above arguments, it is impossible to solve the species-doubling problem in a chirally invariant fashion. Because this theorem does not rely on quantization, it provides an argument which is more powerful than that based on the chiral anomaly given above. The proof of this result, usually called the Nielsen-Ninomiya theorem, depends, of course, on a set of assumptions. It is required that the interaction operator be local (falling off at infinity fast enough to define a continuous Fourier transform), translationally invariant over a finite number of lattice spacings, and Hermitian (a real Fourier transform). These are certainly reasonable assumptions. However, they are by no means necessary to define a sensible theory in the continuum limit. Recall that the lattice is merely a useful device to define the quantum theory when the lattice spacing is taken to zero. It is clear, though, that the only way to avoid the consequences of the theorem is to modify the physics of the lattice system by violating one or more of the assumptions in such a way that the resulting model may still have the right continuum limit. The purpose of this Letter is to propose one such scheme.

Possible solutions to the doubling problem which differ from that suggested here have been presented in the literature. However, these proposals either require nonlocal interactions,[10] or are only directly applicable in theories without dynamical gauge fields.[11] I will therefore concentrate on the assumptions of Hermiticity and translation invariance.

The action presented below is constructed by introducing a dimensionless, real scalar field $\varphi(x)$ whose interaction with the fermion fields breaks translation invariance as well as Hermiticity of the interaction operator. That this may be done follows from the freedom, afforded by the lattice, of introducing nonrenormalizable interactions (operators which become irrelevant as $a\to 0$). Indeed, the implementation of exact gauge invariance on the lattice requires an infinite number of such terms.[1] Specifically, I consider the action

$$S_\varphi=\sum_{n\mu}\bar\psi_n\gamma_\mu M_\mu(\varphi)\psi_n+m\sum_n\bar\psi_n\psi_n, \quad (7)$$

where

$$M_\mu(\varphi)=\tfrac{1}{2}[(1+\varphi_n)\delta_\mu{}^+ +(1-\varphi_n)\delta_\mu{}^-]; \quad (8)$$

$\delta_\mu{}^+$ and $\delta_\mu{}^-$ are, respectively, the forward and backward covariant first-difference operators,

$$\delta_\mu{}^+\psi_n=a^{-1}(\psi_{n+\mu}U_{n\mu}-\psi_n),$$
$$i\delta_\mu{}^-\psi_n=(i\delta_\mu{}^+)^\dagger\psi_n. \quad (9)$$

Notice that $[iM_\mu(\varphi)]^\dagger=iM_\mu(-\varphi)$. In the classical continuum limit, the φ-independent part of the interaction operator, $\frac{1}{2}(\delta_\mu{}^+ +\delta_\mu{}^-)=\delta_\mu{}^s$, is just ∂_μ, whereas the φ-dependent piece, $\frac{1}{2}\varphi_n(\delta_\mu{}^+ -\delta_\mu{}^-)$, is proportional to the lattice spacing times the covariant Laplacian. Thus, the classical continuum limit of (7) is just (1).

To define the quantum theory in such a way that expectation values of Hermitian operators are real, I choose $\varphi(x)$ to be a symmetrically distributed random field. A convenient choice is a Gaussian distribution of width σ,

$$\langle\varphi_n\rangle=0, \quad \langle\varphi_n\varphi_m\rangle=\sigma\delta_{nm}, \quad (10)$$

and the normalized integration measure for φ is

$$D\varphi=\prod_n\frac{d\varphi_n}{(2\pi\sigma)^{1/2}}\exp\left(-\sum_n\frac{\varphi_n{}^2}{2\sigma}\right). \quad (11)$$

With

$$Z_\varphi=\int DU\,D\bar\psi\,D\psi\,\exp[-S_\varphi(U,\bar\psi,\psi)], \quad (12)$$

the expectation value of an operator $O(U,\overline{\psi},\psi)$ is defined as the quenched average

$$\langle O\rangle = \int D\varphi\, Z_\varphi^{-1} \int DU\, D\overline{\psi}\, D\psi\, O(U,\overline{\psi},\psi) \times \exp(-S_\varphi). \qquad (13)$$

Because φ is symmetrically distributed, it follows that (13) is real if O is Hermitian. I emphasize that φ is a fully quenched, classical field, independent of the lattice spacing and any other parameters of the theory. Introducing a long-range random interaction affects the correlation length of the dynamical fields. In a sense then, a physical effect of defining the theory through (13) is that, for fixed, finite a, the functional integral over φ introduces an average over finite correlation lengths, affecting primarily large-momentum (short-distance) modes. The action given by (7) is explicitly invariant (when $m=0$) under the global chiral rotation (5). However, because of the local nature of φ, it is no longer invariant under the doubling symmetry (or generalizations of such local transformations). Furthermore, although the chiral symmetry of (7) implies the existence of a conserved current $j_{\mu\varphi}{}^5$, the matrix elements of this current, given through (13), need not be divergenceless. This observation provides a plausible way out of the anomaly-cancellation argument. In fact, the appearance of an anomalous divergence of $\langle j_{\mu\varphi}{}^5\rangle$ closely resembles the continuum mechanism: It is quantization (in our case, functional averaging) which bars the naive conservation of the current.

I shall now argue that the spectrum of this theory does not have the unwanted degeneracy present when $\varphi_n=0$ everywhere. For arbitrary φ_n, momentum space will no longer diagonalize $M_\mu(\varphi)$. Nonetheless, some understanding of the effect of φ_n on the spectrum of states can be attained by analyzing a particular set of piecewise-constant configurations. Consider first a situation where φ_n assumes a constant value φ_0 throughout the lattice. For simplicity, consider the case $m=0$. The poles of the propagator are located at momenta satisfying the lattice dispersion relations

$$\begin{gathered}\textstyle\sum_\mu \sin^2\frac{1}{2}\theta_\mu[1-(1+\varphi_0{}^2)\sin^2\frac{1}{2}\theta_\mu]=0,\\ \varphi_0\textstyle\sum_\mu \sin^2\frac{1}{2}\theta_\mu \sin\theta_\mu=0, \quad \theta_\mu=ap_\mu.\end{gathered} \qquad (14)$$

It can be readily shown that $\theta_\mu=0$ is an isolated solution of (14) for any finite φ_0: (14) has no other solutions within a sphere of radius $2/(1+\varphi_0{}^2)^{1/2}$ centered at the origin. Further, it is also easy to see that, for any φ_0, $\theta_\mu=0$ is the only solution with this property. Assume now that two configurations, φ_0 and φ_0', lead to a common pole at θ_μ. It then follows from (14) that either $\varphi_0=\pm\varphi_0'$ or $\theta_\mu=0$. That is, two different constant configurations have no common excitations surviving the continuum limit other than those around $p_\mu=0$. If one expands the dispersion relations for φ_0 around an excitation ($p_\mu\neq 0$) corresponding to φ_0' one finds that the square of the energy $E^2(\varphi_0)$ is shifted from zero by a negative amount proportional to the square of the inverse lattice spacing. In φ_0, such an excitation would correspond to a state with a large imaginary mass. A rough approximation to a set of configurations satisfying (10) can be constructed by dividing the lattice into N sublattices, L_i, with φ_n assuming a constant value $\varphi_0{}^i$ in each of the L_i. If one neglects boundary terms and applies the above analysis to the resulting superposition, it becomes apparent that the only excitations which have a finite probability as $a\to 0$ are those which would correspond to $p_\mu\approx 0$. It seems reasonable to conjecture that the same result applies to arbitrary configurations φ_n obeying (10).

Of course, the above are only plausibility arguments. There are, however, stringent numerical tests of these ideas. The simplest thing one can do is to calculate the two-point function as defined by (13). Asymptotically, the ratio of the two-point function obtained with the naive action given by (3) to that defined through (13) should equal 2^d in d dimensions if the present theory works.

To perform this calculation using standard Monte Carlo techniques, I have studied an auxiliary scalar model. The reasoning is as follows. The correct interaction operator for a scalar field on the lattice is just the lattice Laplacian $-\sum_\mu \Delta_\mu{}^+\Delta_\mu{}^-$. However, a scalar model with precisely the same degeneracy as the naive fermion model can be constructed by using instead the interaction operator $-\sum_\mu \Delta_\mu{}^s\Delta_\mu{}^s \equiv M_0{}^2$, whose dispersion relation coincides with that of the naive fermion model. Likewise, a scalar model with an interaction defined through the square of $M_\mu(\varphi)$ will have the same degeneracy as the fermion model given by (7). The results of a first Monte Carlo analysis of the scalar models, shown in Fig. 1, are consistent with the expected value of 2^d for the ratio of the two-point function of the doubled model to that obtained from the present theory.

I have performed this analysis only for $d=2$ and restricted myself to distributions for φ_n with $\sigma<0.1$. Because of possible peculiarities of dis-

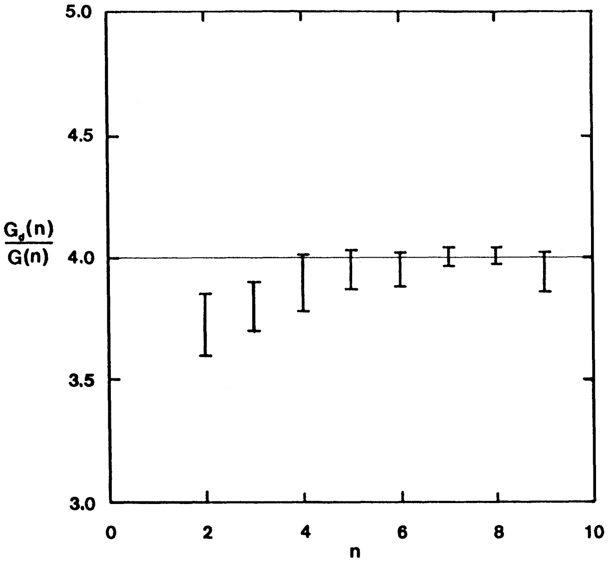

FIG. 1. $G_d(n)$ is the two-point function for the scalar model with interaction operator M_0^2; $G(n)$ is the corresponding correlation for $M^2(\varphi)$ defined through (13). These results are for a 20^2 lattice with $\sigma = 0.05$. The expected ratio is 4.

ordered systems of low dimensionality, a careful analysis of the theory as a function of σ as well as its study for $d > 2$ should be done. Also, the direct application of these ideas in a well understood fermionic model should be tried. This work is currently in progress.

I am grateful to H. Nielsen for getting me interested in this problem, to J. Richardson for his collaboration during the early stages of this work, and to K. Wilson for discovering a fatal flaw in an earlier idea. I also thank my colleagues at Saclay, Tel-Aviv, and the Weizmann Institute, particularly D. Horn, E. Marinari, H. Orland, G. Parisi, A. Schwimmer, and S. Yankielowicz, for their hospitality and useful comments as these ideas developed. Finally, I thank N. Manton for his comments and J. Kuti for very useful conversations and a critical reading of the manuscript. This research was supported in part by the National Science Foundation through Grant No. PHY77-27084.

Note added.—Since this Letter was submitted, the numerical analysis has been extended to three and four dimensions. Also, a more refined study of the two-dimensional model as a function of σ has been performed. Within the statistical accuracy of these simulations, the new results agree with those reported here.

[1]K. G. Wilson, Phys. Rev. D 10, 2445 (1974); A. M. Polyakov, unpublished.

[2]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. 42, 1390 (1979), and Phys. Rev. D 20, 1915 (1979); K. G. Wilson, in *Recent Developments in Gauge Theories*, Proceedings of the Cargese Summer Institute, 1979, edited by G. t'Hooft *et al.* (Plenum, New York, 1980).

[3]For a review, see M. Creutz, L. Jacobs, and C. Rebbi, University of California, Santa Barbara Report No. NSF-ITP-83-07 (to be published).

[4]See, e.g., F. Fucito, E. Marinari, G. Parisi, and C. Rebbi, Nucl. Phys. B180 [FS2], 369 (1981); D. Weingarten and D. Petcher, Phys. Lett. 99B, 333 (1981); J. Hirsch, D. Scalapino, R. Sugar, and R. Blankenbecler, Phys. Rev. Lett. 47, 1628 (1981); A. Hasenfratz and P. Hasenfratz, Phys. Lett. 104B, 489 (1981); J. Kuti, Phys. Rev. Lett. 49, 183 (1982), and references therein.

[5]P. Becher, Phys. Lett. 104B, 221 (1981); J. M. Rabin, Nucl. Phys. B201, 315 (1982); T. Banks, Y. Dothan, and D. Horn, Phys. Lett. 117B, 413 (1982); P. Becher and H. Joos, Z. Phys. C 15, 343 (1982).

[6]S. Adler, Phys. Rev. 177, 2426 (1969); J. S. Bell and R. Jackiw, Nuovo Cimento A60, 47 (1969).

[7]L. H. Karsten and J. Smit, Nucl. Phys. B183, 103 (1981).

[8]K. G. Wilson, in *New Phenomena in Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1977).

[9]H. B. Nielsen and M. Ninomiya, Phys. Lett. 105B, 219 (1981), and Nucl. Phys. B185, 20 (1981), and B193, 173 (1981); L. H. Karsten, Phys. Lett. 104B, 315 (1981); Rabin, Ref. 5.

[10]S. D. Drell, M. Weinstein, and S. Yankielowicz, Phys. Rev. D 14, 487 (1976); J. Rabin, Phys. Rev. D 24, 3218 (1981).

[11]A novel approach to the doubling problem has been pursued by P. H. Ginsparg and K. G. Wilson, Phys. Rev. D 25, 2649 (1982). Although this analysis successfully incorporates the correct anomaly structure of pure fermion theories, it is not known how to extend it to theories with a local gauge symmetry.