PROCEEDINGS OF SCIENCE

Chiral symmetry and lattice fermions

Michael Creutz[*] [†]
Brookhaven National Laboratory
E-mail: creutz@bnl.gov

Lattice gauge theory and chiral perturbation theory are among the primary tools for understanding non-perturbative aspects of QCD. I review several subtle and sometimes controversial issues that arise when combining these techniques. Among these are one failure of partially quenched chiral perturbation theory when the valence quarks become lighter than the average sea quark mass and a potential ambiguity in comparisons of perturbative and lattice properties of non-degenerate quarks.

From quarks and gluons to hadronic matter: A bridge too far?
2-6 September, 2013
European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT), Villazzano, Trento (Italy)*

[*]Speaker.

[†]This manuscript has been authored under contract number DE-AC02-98CH10886 with the U.S. Department of Energy. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. I'm greatful for my Alexander von Humboldt Research Award which partially supported my attendence to this meeting.

PoS(QCD-TNT-III)009

1. Introduction

As is well known, understanding QCD at low energies requires a non-perturbative treatment. While several tools exist, here I will concentrate on two of the most successful, chiral symmetry and lattice gauge theory. Combining these two has a tortuous history, marked with several bitter controversies that continue to this day. In this talk I will fan these fires further.

In the next section I review some rather straightforward predictions of chiral symmetry for the meson spectrum in two flavor QCD with non-degenerate quarks. I restrict myself to two flavors for simplicity, none of the basic points depend on this in any essential way. Everything in Section 2 is well known and I believe not controversial in any way. Then in Section 3 I become less conventional and explore what would happen if the up quark mass were to become negative while maintaining the down quark mass at a positive value. At a certain point the neutral pion can become massless, and beyond that one should expect a pion condensation into what is known as the Dashen phase [1].

Section 4 takes the picture from the previous sections and draws a few general messages which indicate non-trivial consequences for a Euclidean path integral, normally the starting point for lattice gauge simulations. In particular, I will point out some of the peculiarities appearing in the structure of the fermionic operators entering these path integrals. Section 5 I discuss how a generic lattice fermion action automatically has a sort of chiral symmetry. This symmetry is associated with the requirement of doublers, which are pushed to high energy with Wilson or overlap fermions. Section 6 summarizes the main messages and some of the controversial consequences thereof.

2. Chiral symmetry and two flavor QCD

Throughout this discussion I will concentrate on two flavor QCD with non-degenerate quarks, labeled as usual as up (u) and down (d). The restriction to two flavors is for simplicity, and the basic ideas go through for more species. This section is meant to remind you of some of the standard consequences of chiral symmetry for the light pseudo-scalar masses. In this theory one can construct four fermion bilinears that transform as pseudo-scalars

$$i\overline{u}\gamma_5 d \sim \pi_+ \qquad i\overline{d}\gamma_5 u \sim \pi_- \qquad i\overline{u}\gamma_5 u \qquad i\overline{d}\gamma_5 d. \tag{2.1}$$

The first two of these create the charged pions. Usual chiral symmetry arguments give these mesons squared masses proportional to the average of the quark masses

$$M^2_{\pi_\pm} \sim (m_u + m_d)/2. \tag{2.2}$$

The other two pseudo-scalar combinations, both electrically neutral, conceal some rather interesting issues. Gauge theories generically conserve helicity in their interactions. Separating the helicity combinations of these operators naively suggests that the mixing of

$$\overline{u}\gamma_5 u = \overline{u}_L u_R - \overline{u}_R u_L \tag{2.3}$$

with

$$\overline{d}\gamma_5 d = \overline{d}_L d_R - \overline{d}_R d_L \tag{2.4}$$

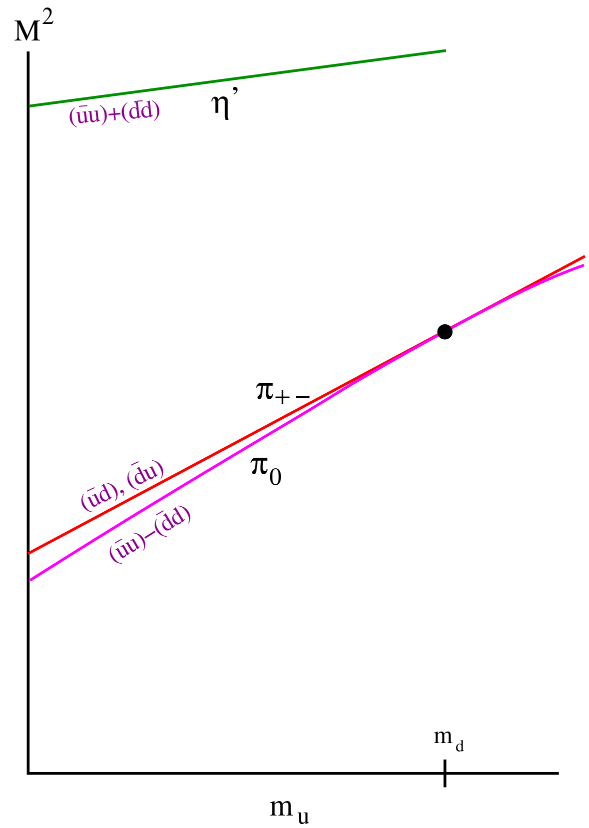

Figure 1: A schematic representation of the pseudo-scalar mass dependence on the up quark mass with a fixed non-vanishing down quark mass. The theory maintains a mass gap even at vanishing up quark mass.

would be suppressed by the product of the up and down quark masses. This would further suggest that there are two light neutral pseudo-scalars, one with $M^2_{\overline{u}\gamma_5 u} \sim m_u$ and a second with $M^2_{\overline{d}\gamma_5 d} \sim m_d$.

Of course it is well known that this prediction is wrong. The chiral anomaly strongly mixes the operators $\overline{u}\gamma_5 u$ and $\overline{d}\gamma_5 d$. This mixing occurs through what is sometimes called the effective "t'Hooft vertex" [2], which for this problem takes the form

$$\overline{u}_L u_R \overline{d}_L d_R + \overline{u}_R u_L \overline{d}_R d_L. \tag{2.5}$$

This mixing breaks naive chiral symmetry and has dramatic consequences for the pseudo-scalar spectrum. In particular, the symmetric combination

$$\eta' \sim i\overline{u}\gamma_5 u + i\overline{d}\gamma_5 d \tag{2.6}$$

is not a pseudo-Goldstone boson. Its mass does not become small in the chiral limit, but rather remains of order the QCD scale

$$M_{\eta'} \sim \Lambda_{qcd} + O(m_u + m_d). \tag{2.7}$$

After this mixing the orthogonal combination remains to represent the neutral pion

$$\pi_0 \sim i\overline{u}\gamma_5 u - i\overline{d}\gamma_5 d. \tag{2.8}$$

Since this is made up of half up and half down quarks, the neutral and charged pions are degenerate up to quadratic order in the quark masses

$$M_{\pi_0}^2 = M_{\pi_\pm}^2 - O((m_u - m_d)^2) - \text{e.m. effects.} \tag{2.9}$$

Of course, in the physical world the electromagnetic effects dominate this mass difference, but here I ignore this for simplicity.

As these points are crucial to the later discussion, it is perhaps useful to sketch the picture graphically. Consider fixing the down quark mass and study the meson spectrum as the up quark mass is varied. As sketched in Fig. 1, the three pions have a leading mass dependence going as $M_\pi^2 \propto \frac{m_u+m_d}{2} + O(m_q^2)$ while the eta prime is heavier. The effect of isospin breaking is to give a small separation of the neutral and charged pions as in Eq. (2.9). The sign of this splitting is not determined by symmetry alone, but it is natural to expect the neutral pion to be lighter because of mixing with the heavier eta prime and any pseudo-scalar gluon bound states. One important observation from this picture is that a mass gap persists even when the up quark is massless.

3. The Dashen phase

The existence of a mass gap at $m_u = 0$ as shown in Fig. 1 suggests that there is no physical singularity as the up quark mass is further varied into the negative region. Such an extrapolation is expected to behave smoothly up to a point where the square of the neutral pion mass goes to zero. Beyond that, and this occurs sigma models, one should expect the neutral pion will spontaneously condense and acquire an expectation value. Since this is a CP odd field, the resulting phase will spontaneously break CP. The resulting extension of Fig. 1 is sketched in Fig. 2.

This possibility of a spontaneous breaking of CP in the strong interactions was suggested in 1971 by Dashen [1]. This was before the understanding that QCD was the likely underlying theory; Dashen's discussion was based on the current algebra ideas of the day. This behavior is also natural in sigma models for chiral symmetry breaking, both in the linear [3] and the non-linear [4] forms. Those models also suggest that the transition is Ising like, with the order parameter being the expectation of the neutral pion field. This transition formally occurs where the strong CP angle Θ takes the value π; it occurs in a region where the product of the quark masses is negative. The Θ parameter as usually defined is the phase of this product.

Fig. 3 shows another way of looking at this phase structure. On varying the two physical parameters m_u and m_d, a CP violating phase covering a portion of the region where their product is negative. The symmetries of this figure are quite instructive. In particular the picture is symmetric in the sign of the average quark mass $\frac{m_d+m_u}{2}$ as well as in the quark mass difference $\frac{m_d-m_u}{2}$. However there is no symmetry between these two quantities. While the vanishing of either of these quantities is protected by the symmetry, non-perturbative effects are expected to renormalize them differently.

4. Messages and consequences for the path integral

The picture presented in the previous sections is quite simple and should not be particularly controversial. However when this physics is thought about in terms of the path integral approach, it raises several interesting points. These are at the heart of several of the controversies dividing the lattice community.

PoS(QCD-TNT-III)009

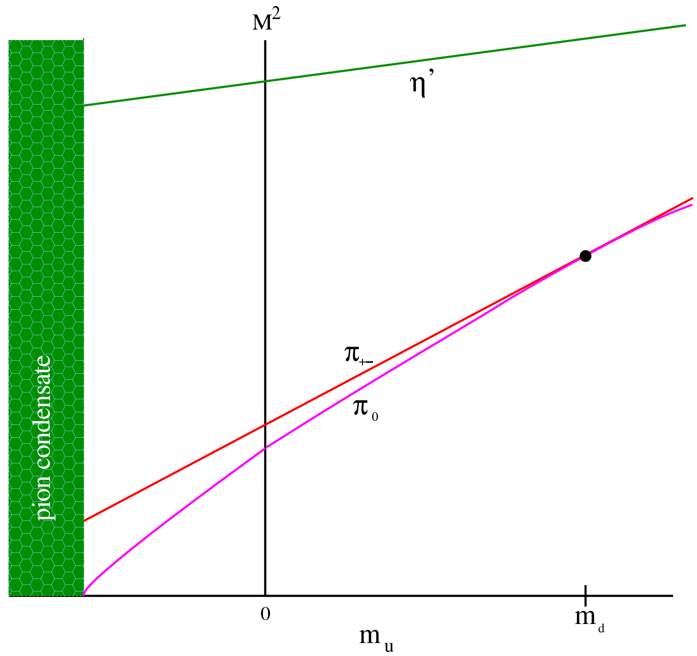

Figure 2: As the up quark mass is varied into the negative mass region, the pions will continue to get lighter until the neutral pion mass vanishes. Beyond that point one expects a CP violating phase where the neutral pion field has an expectation value.

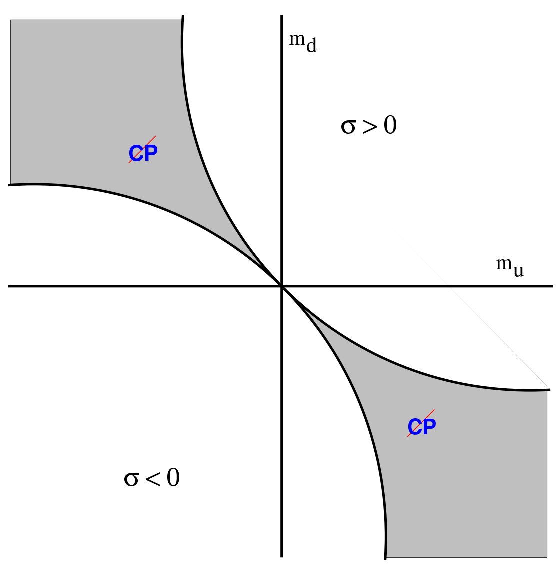

Figure 3: Varying both the up and down quark masses, a CP violating phase lies in a portion of the region where the product of these masses is negative.

4.1 Message 1: A mass gap persists when only one quark is massless

This point is connected to several issues. For one, it means that QCD with only one massless quark has no infrared issues. Of course perturbative QCD requires care with infrared divergences, but these are an artifact of perturbation theory. For any gauge invariant correlator, long distance physics will fall exponentially in the distance with the mass of the lightest state that can be exchanged; that is,

$$\langle \phi(x)\phi(0)\rangle_c \ <\sim \ e^{-m_{\pi_0}|x|}. \tag{4.1}$$

This may be a bit unexpected for the path integral. While there can in principle be small Dirac eigenvalues in the quark matrix, no long distance physics arises. Of course this is special to the one flavor theory; with 2 or more massless quarks the pions become massless.

Note that the prediction from the sigma models does not show any non-analyticity as the up quark mass passes through zero. This raises the question of whether there is any experimental signature of a vanishing up quark mass. Most discussions assume that this point is well defined, but this has never been proven non-perturbatively. The fact that the average quark mass and the quark mass difference have no symmetry between them indicates that there is no symmetry to protect the point of vanishing up quark mass when the down quark mass doesn't vanish. More formally, it is unclear whether different non-perturbative regularization schemes, such as the use of different overlap operators, will have a universal continuum limit for the theory at vanishing up quark mass.

The absence of a singularity at $m_u = 0$ has further consequences for the spectrum of the Dirac operator appearing in the path integral. Banks and Casher [5] long ago pointed out a relation between the density of small Dirac eigenvalues and the jump in the chiral condensate that occurs on passing through the chiral point. Since the non-degenerate quark system being explored here has no jump in physics in the $m_u = 0$ region, there cannot be a finite density of small eigenvalues for the up quark operator.

This point raises serious issues for a popular technique known as "partially quenched chiral perturbation theory." This approach considers adding "valence" quarks to the theory which have no feedback on the gauge fields. The physical dynamical quarks are then referred to as "sea" quarks. The usual assumption is that multiple valence quarks will undergo a chiral condensation and give rise to massless valence pions as their mass goes to zero. It is known that this approximation fails in the pure gauge theory without dynamical quarks [6, 7, 8]. With a massless dynamical quark, however, one might expect the theory to be better behaved. Nevertheless, as a valence quark mass approaches the up quark mass, the sea and valence propagators become identical. If the sea quark mass does not have small eigenvalues, then the valence quarks cannot either. Thus the usually assumed chiral condensation cannot occur if only the up sea quark is massless. The basic conclusion is that the partially quenched theory cannot be trusted for

$$m_{valence} < \langle m_{sea} \rangle. \tag{4.2}$$

4.2 Message 2: The sign of a quark mass is physically relevant

It should be immediately clear from Fig. 2 that the meson spectrum is different between a small positive up quark mass and a negative value of the same magnitude. This is a crucial property of

QCD that cannot be seen in perturbation theory. Indeed, in any perturbative diagram the sign of the mass appearing in the quark propagators can be reversed by a chiral rotation. This rotation, however, is anomalous.

The reason the sign of the mass becomes relevant is most easily understood as arising from gauge configurations of non-trivial topology. In particular, when one mass is negative, the path integral receives a weighting factor of $(-1)^\nu$, where ν is the winding number of the background gauge configuration. This is a special case of the more general result that if the up quark mass is given a phase $m_u \to e^{i\Theta\gamma_5} m_u$ then the theory describes the so called "Theta vacuum." This is a physically different theory for each Θ. I will return to the effects of this rotation later in the discussion.

The relevance of the sign of the quark mass has important consequences for attempts to relate lattice parameters to those obtained from perturbation theory in the continuum. This means that attempts to match lattice and $\overline{MS}$ masses are particularly dangerous for non-degenerate quarks.

4.3 Message 3: A divergent correlation length is possible when no quarks are massless

Fig. 2 indicates that at the point where the Dashen phase begins the neutral pion mass will vanish. In the effective chiral models this is an Ising like second order transition. This occurs at a point where neither the up nor down quark masses vanish, although they are of opposite sign.

In terms of the Dirac operators in the path integral, this shows that the mass gap can vanish without having any small Dirac eigenvalues. The location of this transition is a dynamical issue, requiring the tuning of the up quark mass to the edge of the Dashen phase.

At a somewhat deeper level, this picture is connected to whether or not there is a first order transition at $\Theta = \pi$. When the lightest quarks are multiply degenerate, it is fairly straightforward to argue that there must be a first order transition at such a point [9]. But in the case where the lightest quark is non-degenerate, there is a transition only once one enters the Dashen phase. In that phase as one crosses $\Theta = \pi$, the order parameter $\langle \pi_0 \rangle$ undergoes a jump. But before one reaches this phase, that is when the up quark mass is negative but small, the behavior at $\Theta = \pi$ is smooth. Thus whether there is a transition or not is a delicate dynamical issue. The corresponding question of a possible transition at $\Theta = \pi$ in the zero flavor theory remains open.

Another interesting feature of the boundary of the Dashen phase is that at this point the topological susceptibility diverges to minus infinity [10]. This comes about since the operator $F_{\mu\nu}\tilde{F}_{\mu\nu}$ is expected to have a finite amplitude to create a neutral pion. This gives an infrared divergence to the susceptibility as the neutral pion mass goes to zero. Reflection positivity [11] forces this divergence to be negative.

5. Lattice Fermions

I now turn to some generic properties of lattice Dirac operators and the connection to the chiral anomaly. These points are closely related to the Nielsen-Ninomiya theorem [12], but my approach will be a bit different.

Suppose I am given some arbitrary lattice Dirac operator D to use for a lattice simulation. To proceed, I assume gamma five hermiticity

$$\gamma_5 D \gamma_5 = D^\dagger. \tag{5.1}$$

This is a property of the majority of the operators currently used in practice. (One notable exception is the twisted mass approach which adds a chiral rotation to this property.)

Now consider dividing the operator D into Hermitean and antihermitean parts $D = K + M$ with

$$\begin{aligned} K &= (D - D^\dagger)/2 \\ M &= (D + D^\dagger)/2 \end{aligned} \tag{5.2}$$

This immediately implies

$$\begin{aligned} [K, \gamma_5]_+ &= 0 \\ [M, \gamma_5]_- &= 0. \end{aligned} \tag{5.3}$$

On a lattice everything is fully regulated and finite; so the naive equation $\mathrm{Tr}\gamma_5 = 0$ still holds. An immediate consequence is that

$$M \to e^{i\theta\gamma_5} M \tag{5.4}$$

is an exact symmetry of the determinant. Explicitly I have

$$|K + M| = |e^{i\gamma_5\theta/2}(K+M)e^{i\gamma_5\theta/2}| = |K + e^{i\theta\gamma_5}M| \tag{5.5}$$

This leads to the next message.

5.1 Message 4: Any lattice action is symmetric under the chiral rotation $M \to e^{i\theta\gamma_5}M$

This looks rather peculiar in the context of the earlier remark that a chiral rotation of the mass term yields an inequivalent theory with a non-vanishing Θ parameter. Indeed, where did the anomaly hide? The answer is that this particular symmetry can only be a flavored chiral symmetry, which is allowed by the anomaly. But for this to be a flavored symmetry, all lattice actions must bring in some extra structure.

Some actions, such as naive, staggered, and minimally doubled fermions resolve this by having doublers. Half of them use γ_5 and half $-\gamma_5$ for their chiral rotations. In this way the naive chiral symmetry above is actually a flavored symmetry.

Wilson and overlap fermions contain this symmetry in a more subtle manner. In these cases the Hermitean part M ceases to be a constant. Instead, it produces heavy states that cancel the anomaly. To see this more explicitly, consider Wilson fermions. For the free theory the Dirac operator takes the momentum space form

$$D_W = \frac{1}{a}\sum_\mu (i\gamma_\mu \sin(p_\mu a) + 1 - \cos(p_\mu a)) + m. \tag{5.6}$$

The corresponding eigenvalue structure takes the form of a set of nested ellipses and is sketched in Fig. 4.

For small momentum the "mass" term takes the form

$$m + \frac{1}{a}(1 - \cos(p_\mu a)) = m + O(a) \tag{5.7}$$

however for momentum components near π the eigenvalues are of order $1/a$. Note that as seen in Fig. 4, there is no symmetry if m changes sign. As noted earlier, this is a necessary feature of QCD, and with Wilson fermions it is manifest from the outset. This action also breaks naive chiral symmetry explicitly, *i.e.* $[D_w, \gamma_5]_+ \neq 0$ even when $m = 0$.

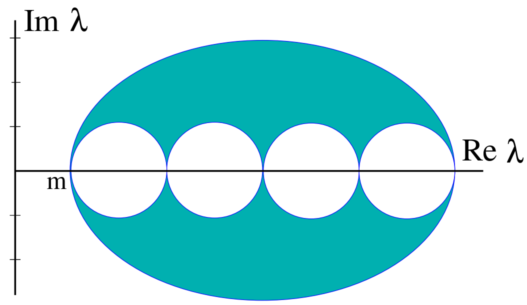

Figure 4: The eigenvalue spectrum of the free Wilson fermion operator forms a set of nested ellipses.

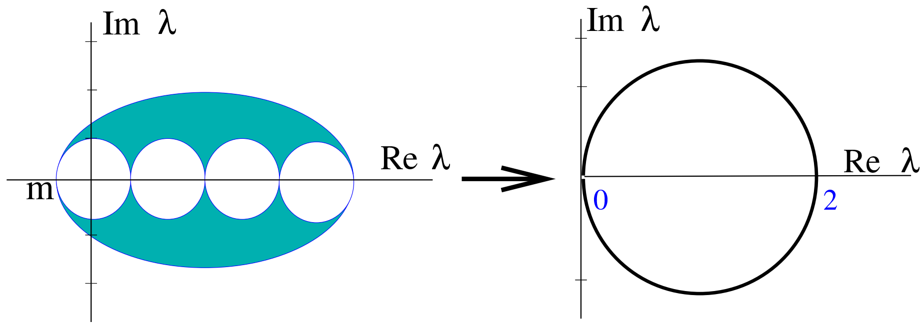

Figure 5: The overlap operator is obtained by projecting eigenvalues of the Wilson fermion operator onto a unit circle.

The overlap operator [13] is obtained by projecting the Wilson operator onto a circle to obtain a unitary matrix V which is then shifted to give small eigenvalues near the origin.

$$D_W \to D_o = 1 - V, \qquad V^\dagger V = 1. \tag{5.8}$$

This process is sketched in Fig. 5.

This projection process is computationally demanding, but it leaves an operator with rather elegant properties. First, it maintains gamma five hermiticity $\gamma_5 D_o \gamma_5 = D_o^\dagger$. Second, unlike in the Wilson case, it is a normal operator; *i.e.* it commutes with its adjoint, $[D_o^\dagger, D_o] = 0$. But probably most important, it retains an exact variant of chiral symmetry

$$e^{i\theta\hat{\gamma}_5} D_o e^{i\theta\gamma_5} = D_o \tag{5.9}$$

where

$$\hat{\gamma}_5 = V\gamma_5. \tag{5.10}$$

As with γ_5, $\hat{\gamma}_5$ is Hermitean and $\hat{\gamma}^2 = 1$. All its eigenvalues are ± 1 and allow one to define an index.

$$\nu = \frac{1}{2}\mathrm{Tr}\hat{\gamma}_5 = \mathrm{Tr}\frac{\gamma_5 + \hat{\gamma}_5}{2}. \tag{5.11}$$

This definition agrees with the continuum index for smooth fields.

With these preliminaries, it should now be clear where the anomaly went with Wilson and overlap fermions. Effectively the doublers are still there, but have been given masses of order the cutoff. The chiral rotation $M \to e^{i\theta\gamma_5} M$ not only rotates the light quark masses, but also those of the heavy doubler states. This emphasizes an old observation of Seiler and Stamatescu [14] that the physical Θ is the relative angle occurring under independent rotations the fermion mass and the Wilson term.

With the overlap operator, any zero eigenmodes corresponding to topology will have heavy counterparts on the opposite side of the unitary circle. Rotating the Hermitean part of this operator rotates these heavy modes as well.

This hiding of the anomaly is not just a lattice artifact, but is relevant to the continuum theory as well. In general the physical Θ can be moved around and placed on any flavor at will. In the standard model, any non-zero Θ can be entirely moved into the top quark phase. Since the value of Θ is physical, *i.e.* it gives a neutron an electric dipole moment, some properties of the top quark remain relevant to low energy physics! In other words, the naive decoupling theorems for heavy quarks do not apply non-perturbatively.

6. Summary

The main messages of this talk are:

1. A mass gap persists when only one quark is massless.

2. The sign of a quark mass is physically relevant.

3. A divergent correlation length can occur even when no quarks are massless.

4. Any lattice action is symmetric under the chiral rotation $M \to e^{i\theta\gamma_5}M$.

These are all rather straightforward consequences of effective chiral Lagrangians, but are less obvious when thought of in terms of fermion determinants in a Euclidean path integral. Some of the more controversial consequences are

1. There is no proof of universality between non-perturbative schemes for the point where the up quark mass vanishes when the other quarks are massive.

2. The usual assumptions of partially quenched chiral perturbation theory can fail if $m_{valence} < m_{sea}$.

3. Matching lattice results to $\overline{MS}$ can miss non-perturbative effects when the quarks are non-degenerate.

4. Decoupling theorems do not apply non-perturbatively.

There are other related controversial points that I have not covered in this talk. One is that the topological susceptibility also has an ambiguity from rough gauge fields. This is closely tied to the above non-universality of the point for vanishing up quark mass. Finally there is the infamous rooting issue for staggered quarks. In the previous editions of this conference [15, 16] I explained in some detail as to why that approach is incorrect; I will not go into it further here.

References

[1] Roger F. Dashen. Some features of chiral symmetry breaking. *Phys.Rev.*, D3:1879–1889, 1971.

[2] Gerard 't Hooft. Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys.Rev.*, D14:3432–3450, 1976.

[3] Michael Creutz. Quark masses and chiral symmetry. *Phys. Rev.*, D52:2951–2959, 1995.

[4] Michael Creutz. Spontaneous violation of CP symmetry in the strong interactions. *Phys. Rev. Lett.*, 92:201601, 2004.

[5] Tom Banks and A. Casher. Chiral symmetry breaking in confining theories. *Nucl.Phys.*, B169:103, 1980.

[6] Benjamin Svetitsky, Yigal Shamir, and Maarten Golterman. Localization of lattice fermions: Lessons for overlap. *PoS*, LAT2005:129, 2006.

[7] Maarten Golterman, Yigal Shamir, and Benjamin Svetitsky. Localization properties of lattice fermions with plaquette and improved gauge actions. *Phys.Rev.*, D72:034501, 2005.

[8] Maarten Golterman, Yigal Shamir, and Benjamin Svetitsky. Mobility edge in lattice QCD. *Phys.Rev.*, D71:071502, 2005.

[9] Michael Creutz. Anomalies and chiral symmetry in QCD. *Annals Phys.*, 324:1573–1584, 2009.

[10] Michael Creutz. Quark masses, the Dashen phase, and gauge field topology. 2013.

[11] Erhard Seiler. Some more remarks on the Witten-Veneziano formula for the eta-prime mass. *Phys.Lett.*, B525:355–359, 2002.

[12] Holger Bech Nielsen and M. Ninomiya. Absence of neutrinos on a lattice. 1. proof by homotopy theory. *Nucl.Phys.*, B185:20, 1981.

[13] Herbert Neuberger. Exactly massless quarks on the lattice. *Phys.Lett.*, B417:141–144, 1998.

[14] E. Seiler and I.O. Stamatescu. Lattice fermions and Theta vacua. *Phys.Rev.*, D25:2177, 1982.

[15] Michael Creutz. Anomalies and discrete chiral symmetries. *PoS*, QCD-TNT09:008, 2009.

[16] Michael Creutz. Quark masses in two-flavor QCD. *POS*, QCD-TNT-II, 2011.

PoS(QCD-TNT-III)009