CHAPTER 39

Uses of Effective Field Theory in Lattice QCD

Andreas S. Kronfeld

USES OF EFFECTIVE FIELD THEORY IN LATTICE QCD

Andreas S. KRONFELD

Theoretical Physics Department, Fermi National Accelerator Laboratory [a]
Batavia, Illinois, USA

Several physical problems in particle physics, nuclear physics, and astrophysics require information from non-perturbative QCD to gain a full understanding. In some case the most reliable technique for quantitative results is to carry out large-scale numerical calculations in lattice gauge theory. As in any numerical technique, there are several sources of uncertainty. This chapter explains how effective field theories are used to keep them under control and, then, obtain a sensible error bar. After a short survey of the numerical technique, we explain why effective field theories are necessary and useful. Then four important cases are reviewed: Symanzik's effective field theory of lattice spacing effects; heavy-quark effective theory as a tool for controling discretization effects of heavy quarks; chiral perturbation theory as a tool for reaching the chiral limit; and a general field theory of hadrons for deriving finite volume corrections.

Contents

1 Introduction

The idea to use lattice gauge theory to study quantum chromodynamics (QCD) was introduced in 1974 in a seminal paper by Kenneth Wilson.[1] It was an exciting time for the strong interactions: it had become clear that quarks are partons,[2] the Lagrangian for the gauge theory of quarks and gluons had just been published,[3] and the discovery of asymptotic freedom was new.[4,5] Soon afterwards an experiment at Brookhaven[6] observed a new resonance, the J, in proton-Beryllium collisions, and an experiment at SLAC[7] also observed a new resonance, the ψ, in e^+e^- collisions. This bound state of a charmed quark and its anti-quark, now called the J/ψ, and its excitation spectrum has gone on to play a role to similar to that of positronium in quantum electrodynamics. Even today, the charmonium system plays an important role in lattice QCD.

The emergence of QCD as the theory of the strong interactions means that high-energy scattering of partons and bound-state properties of hadrons have a common explanation. Owing to asymptotic freedom,[4,5] the gauge coupling in QCD becomes weaker at short distances, making perturbative cross sections more accurate at higher energies. The flip side, however, is that the coupling becomes stronger at long distances: perturbation theory breaks down and the bound-state problem in QCD is intrinsically non-perturbative. Thus, most analysis of non-perturbative QCD has a general nature, relying on, for example, symmetries, analyticity, unitarity, and the renormalization group. At its most successful, this kind of analysis yields quantitative relationships between experimentally measurable quantities. Perhaps the most striking example is to obtain parton densities from deeply inelastic scattering, and then to use them to predict jet cross sections in $p\bar{p}$ collisions.

Despite such successes, there is a need for general purpose tools to calculate properties of the hadrons from the QCD Lagrangian, from first principles. One would like to see, quantitatively, that QCD can explain both jet cross sections *and* the proton mass, only by adjusting the QCD coupling and the quark masses. One would like to gain insight into the mechanisms of confinement, of the type provided by detailed calculations. Finally, to understand interactions of quarks at the shortest distances, where new phenomena may be at play, it is usually necessary to have calculations of certain hadronic matrix elements, with controlled, comprehensible uncertainties.

Lattice gauge theory provides a mathematically well-defined framework for non-perturbative QCD. Before Wilson, Wegner[8] had studied a version of the Ising model, on a two-dimensional lattice, with a discrete gauge symmetry. Wilson realized how to implement the continuous SU(3) gauge symmetry of QCD and other gauge theories and, more importantly, that lattice field theory provides a non-perturbative definition of the functional integral. To make any sense of quantum field theory, an ultraviolet cutoff must be introduced, and the
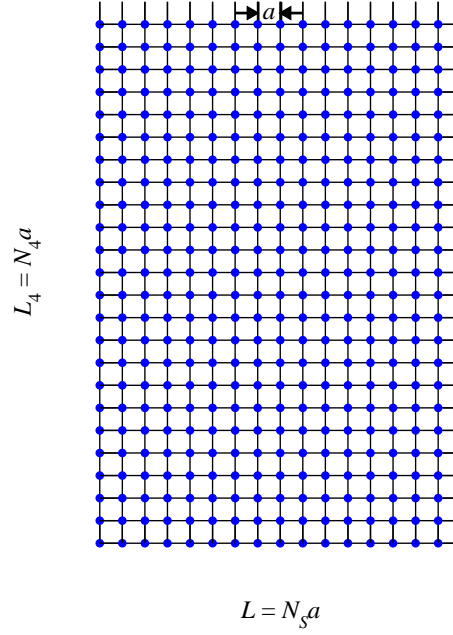
Figure 1: Two-dimensional slice of a lattice with spacing a, spatial size $L = N_S a$, and temporal extent $L_4 = N_4 a$.

key feature of lattice field theory is that it introduces the cutoff at the outset. With the lattice cutoff, it is possible to derive rigorously many properties of field theory, as one can see, for example, in the textbook of Glimm and Jaffe.[9]

The basic idea is to replace continuous spacetime with a discrete lattice, usually hypercubic, as sketched in Fig. 1. The spacing between sites is usually denoted a. For simplicity we consider the spatial volume to have length $L = N_S a$ on each side, and the temporal extent to be $L_4 = N_4 a$. From a theoretical point of view, the lattice and finite volume provide gauge-invariant ultraviolet and infrared cutoffs, respectively. Fermion fields $\psi(x)$ and $\bar{\psi}(x)$ live on sites x. Gauge fields live on links through the variables

$$U_\mu(x) = \text{P} \exp \int_0^a ds\, A_\mu(x + se_\mu) \tag{1}$$

where P denotes path ordering, and e_μ is a unit vector in the μ direction. In quantum field theory, information is obtained from correlation functions, which have a functional integral representation. In lattice QCD the correlation

functions are

$$\langle O_1 \cdots O_n \rangle = \frac{1}{Z} \int \prod_{x,\mu} dU_\mu(x) \prod_x d\psi(x) d\bar{\psi}(x)\, O_1 \cdots O_n\, e^{-S_{\mathrm{QCD}}} \qquad (2)$$

where Z is defined so that $\langle 1 \rangle = 1$, and $S_{\mathrm{QCD}} = -\sum_x \mathcal{L}_{\mathrm{QCD}}(x)$ is the (lattice) QCD action. The O_i are operators for creating the hadrons of interest and also terms in the electroweak Hamiltonian. With quarks on sites and gluons on links, it is possible to devise lattice actions that respect gauge symmetry. As in discrete approximations to partial differential equations, derivatives in the Lagrangian are replaced with difference operators. A simple Lagrangian, introduced by Wilson,[1,10] is

$$\mathcal{L}_{\mathrm{W}} = \mathcal{L}_{\mathrm{Wg}} + \mathcal{L}_{\mathrm{Wq}} \qquad (3)$$

$$\mathcal{L}_{\mathrm{Wg}} = \frac{1}{g_0^2 a^4} \sum_{\mu\nu} \mathrm{Re\,tr} \left[U_\mu(x) U_\nu(x + ae_\mu) U_\mu^\dagger(x + ae_\nu) U_\nu^\dagger(x) - 1 \right] \qquad (4)$$

$$\mathcal{L}_{\mathrm{Wq}} = -m_0 \bar{\psi}(x)\psi(x) - \sum_\mu \bar{\psi}(x) \left[P_{+\mu} D_{\mu\,\mathrm{lat}}^- - P_{-\mu} D_{\mu\,\mathrm{lat}}^+ \right] \psi(x), \qquad (5)$$

where g_0^2 is the bare gauge coupling, and m_0 the bare quark mass. (The literature often uses $\beta = 6/g_0^2$ and $\kappa = (8 + 2m_0 a)^{-1}$.) In Eq. (5)

$$P_{\pm\mu} = \tfrac{1}{2} \left(1 \pm \gamma_\mu \right), \qquad (6)$$

$$D_{\mu\,\mathrm{lat}}^+ \psi(x) = a^{-1} \left[U_\mu(x)\psi(x + ae_\mu) - \psi(x) \right], \qquad (7)$$

$$D_{\mu\,\mathrm{lat}}^- \psi(x) = a^{-1} \left[\psi(x) - U_\mu^\dagger(x - ae_\mu)\psi(x - ae_\mu) \right]. \qquad (8)$$

The lattice clearly breaks spacetime symmetries, and we shall have to confront that issue below. But it preserves SU(3) gauge symmetry, and that is the most important feature of Wilson's formulation of field theory.

It is a simple exercise to show that Wilson's Lagrangian reproduces the Yangs-Mills Lagrangian in the naive continuum limit, $a \to 0$ with g_0^2 and m_0 fixed. The bare couplings are, of course, unphysical, and the real continuum limit must be taken with physical quantities (e.g., hadron masses) held fixed. Many discretizations have the correct naive continuum limit. As long as the lattice Lagrangian is local,[b] they are all expected to share the same quantum continuum limit (with physical masses fixed). The argument for such "universality" is based on the renormalization group, which indicates that physics depends only only the gauge couplings and quark masses.[11] These arguments have been tested experimentally in condensed matter systems, and although a

---

[b] Here "local" means that couplings for interactions of fields separated by a distance r fall off exponentially with r.

rigorous proof has not been achieved, there is a lot of mathematical evidence that lattice field theory defines quantum field theory.[9]

The study of QCD, which a theory of the natural world, is less concerned with rigorous theorems than it is with practical results. From this point of view, the breakthrough of the lattice formulation is that Eq. (2) turns quantum field theory into a mathematically well-defined problem in statistical mechanics. Condensed matter theorists and mathematical physicists have devised a variety of methods for tackling such problems. In addition to weak-coupling perturbation theory, the toolkit includes non-perturbative versions of the renormalization group, strong coupling expansions, and numerical integration of the functional integral by Monte Carlo methods. The first two were pursued with great vigor in the decade following Wilson's original paper. The strong coupling limit is especially appealing, because confinement emerges immediately, cf. Sec. 8. The last is the most widely used today, especially for problems motivated by particle physics, nuclear physics, or astrophysics. Indeed, when most physicists speak of lattice QCD, they mean numerical lattice QCD.

It is difficult to decide what aspects of lattice QCD to cover in a single chapter of a handbook of all QCD. There are several textbooks on lattice gauge theory,[12–14] with emphasis on lattice QCD. These books, as well as many review articles[15–19] and summer school lecture series,[20–27] cover the foundations well. Nevertheless, many colleagues—theorists whose research is in continuum QCD and experimenters whose measurements need non-perturbative QCD to be interpreted—tell me that papers on lattice calculations are impenetrable.

One complaint is that the explanations of numerical techniques are written in an unfamiliar jargon, which is unfortunate but hard to rectify here. A more serious complaint surrounds the uncertainties that arise in moving the idealized problem of mathematical physics to a practical problem of computational physics. Many non-experts know that these arise from Monte Carlo statistics, non-zero lattice spacing, finite spacetime volume, and unphysical values (in the computer) of the quark masses. But, nevertheless, the methods to deal with them are (evidentally) not transparent. This is a shame. Everyone has a feel for statistical errors, even without knowing how to write a Monte Carlo program to evaluate the right-hand side of Eq. (2). Similarly, the other uncertainties are controlled and understood with effective field theories, so a basis for a common language should be possible.

Two examples illustrate why a common understanding of the uncertainties is needed. The first comes from flavor physics, where a wide variety of B, D, and K decays are studied, to test whether the standard Cabibbo-Kobayashi-Maskawa (CKM) mechanism adequately explains flavor and CP violation. The quark-level CKM interpretation of many of these processes is obscured by the uncertainty in hadronic matrix elements, of the type $\langle f|\mathcal{H}|H\rangle$, where H is a strange, charmed, or b-flavored hadron, and $\mathcal{H}$ is a term in the electroweak

Hamiltonian. $\mathcal{H}$ arises from integrating out W, Z, t, and (possibly) other more massive particles. Experimental measurements are, or soon will be, very precise. Without reliable theoretical calculations, including a transparent error analysis, it will be much more difficult tell whether $\mathcal{H}$ is solely electroweak in origin or, perhaps, has non-standard contributions. Lattice calculations are most straightforward when the final state f has leptons and either one hadron or none. [c] Such matrix elements are needed for leptonic, radiative, and semi-leptonic decays, as well as for neutral meson mixing.

A second example, though less often mentioned, is the search for new particles at the Tevatron and, later, at the LHC. If a new particle carries color, then its decays always contain jets, and observation depends on the extra jet production standing out against normal QCD jet production. The QCD uncertainty has, in the past, been estimated by comparing various fits to the parton densities. A closer look,[28] however, reveals that the parton densities are well-constrained over a limited range of x (x: fraction of the proton momentum taken by a parton). Uncontrolled uncertainties in predictions of cross sections arise because there are only meager constraints on the parton densities outside this range. In lattice gauge theory, however, it is possible to calculate the moments of the parton densities: they are related via the operator-product expansion to matrix elements of local operators. By the time the LHC experiments run, it should be possible to obtain the first few moments with an uncertainty that is not only small, but also well justified.[29−32] The direct calculation of moments and the direct measurement over a finite interval in x will provide complementary information, making possible signals of new physics more persuasive.

Suppose the comparison of theory and experiment, in either flavor physics or highest-energy collisions, leads to a hypothesis of new, non-standard phenomena. Then the stakes become very high, and the reliability of error bars becomes the central concern. It seems, therefore, worthwhile to discuss how to control and estimate uncertainties in numerical calculations, and this issue is the central theme of this chapter. The origin of statistical errors in numerical lattice QCD is reviewed in Sec. 2. Then the origin of systematic uncertainties is reviewed in Sec. 3, motivating the main tools for controlling them. In most cases the tool is an effective field theory, which allows us to control the extrapolation of artificial, numerical data to the real world, provided the data start "near" enough. Most particle physicists are familiar with the logic and utility of effective field theories, and know how to judge their range of validity. Thus, effective field theories should also provide a common language for experts and non-experts to discuss the error bars, without requiring the non-experts to repeat all the steps of the numerical analysis. The effective field theories needed

---

[c]When there are two hadrons in the final state, there are additional complications, cf. Sec. 7.2.

to analyze the controllable uncertainties are discussed in Secs. 4–7.

An uncontrolled systematic effect of many lattice calculations has histori-
cally been the quenched approximation. It is reviewed in Sec. 2. It is hard to
estimate the associated error, and only in isolated cases can one argue that it
is a subdominant error, let alone that it is under control.

Section 8 contains a "top ten" list of the trends and developments in
lattice QCD, which warrant appreciation. These have been chosen for their
broad interest, and because they should influence one's thinking about QCD.

2   Overview of Numerical Techniques

This section gives a brief overview the numerical methods. These are covered
in more detail in some of the texts, reviews, and summer schools cited above,
as well as in a set of lecture notes aimed at experimenters.[33]

The foremost issue is that there are very many variables. Continuum field
theory has uncountably many degrees of freedom. Field theory on an infinite
lattice still has an infinite number of degrees of freedom, but the infinity is
now countable, *i.e.*, it is as infinite as the integers. This makes the products
over x in Eq. (2) well-defined. For a computer (with finite memory), the
number of degrees of freedom must be kept finite. To do so, one must also
introduce a finite spacetime volume. This may seem alarming, but what one
has done is simply to introduce an ultraviolet cutoff (the lattice) and an infrared
cutoff (the finite volume). All calculations in QCD, except trivial ones, require
an ultraviolet cutoff, and many require an infrared cutoff, although physical
predictions are cutoff independent. In a sense, removal of the cutoffs is the
subject of Secs. 4 and 7.

Even with a finite lattice, the number of integration variables is large.
For QCD on a $N_S^3 \times N_4$ lattice (cf. Fig. 1) there are $(4 \times 8)N_S^3 N_4$ variables
for gluons and $(4 \times 3)N_S^3 N_4$ for quarks. If one only demands a volume a
few times the size of a hadron and also several grid points within a hadron's
diameter, one already requires at least, say, 10 points along each direction. In
four-dimensional spacetime this leads to $\sim 32 \times 10^4$ gluonic variables.

With so many variables, the only feasible methods are based on Monte
Carlo integration. The basic idea of Monte Carlo integration is simple: gener-
ate an ensemble of random variables and approximate the integrals in Eq. (2)
by ensemble averages. Thus, calling all variables ϕ,

$$\langle O_1 \cdots O_n \rangle = \frac{1}{Z} \sum_{z=1}^{N_Z} w(\phi^{(z)}) O_1(\phi^{(z)}) \cdots O_n(\phi^{(z)}) \qquad (9)$$

with weights w to be specified below, and Z defined so that $\langle 1 \rangle = 1$.

Quarks pose special problems, principally because, to implement Fermi statistics, fermionic variables are Grassmann numbers. In all cases of interest, the quark action can be written

$$S_{\mathrm{q}} = -\sum_x \mathcal{L}_{\mathrm{q}}(x) = \sum_{\alpha\beta} \bar{\psi}_\alpha M_{\alpha\beta} \psi_\beta, \qquad (10)$$

where α and β are multi-indices for (discrete) spacetime, spin and internal quantum numbers. The matrix $M_{\alpha\beta}$ is some discretization of the Dirac operator $\not{D} + m$, such as Eq. (5). Note that it depends on the gauge field, but one may integrate over the gauge fields after integrating over the quark fields. Then, because the quark action is a quadratic form, the integral can be carried out exactly:

$$\int \prod_{\alpha\beta} d\bar{\psi}_\alpha \, d\psi_\beta \, e^{-\bar{\psi} M \psi} = \det M. \qquad (11)$$

Similarly, products $\psi_\alpha \bar{\psi}_\beta$ in the integrand $O_1 \cdots O_n$ are replaced with quark propagators $[M^{-1}]_{\alpha\beta}$ using the familiar rules of Wick contraction. The computation of M^{-1} is demanding, and the computation of $\det M$ (or, more precisely, changes in $\det M$ as the gauge field is changed) is very demanding.

With the quarks integrated analytically, it is the gluons that are subject to the Monte Carlo method. The factor weighting the integrals is now $\det M e^{-S_{\mathrm{g}}}$, where S_{g} is the gluons' action. Both $\det M$ and $e^{-S_{\mathrm{g}}}$ are the exponential of a number that scales with the spacetime volume. In Minkowski spacetime the exponent is an imaginary number, so there are wild fluctuations for moderate changes in the gauge field. On the other hand, in Euclidean spacetime, with an imaginary time variable, S_{g} is real. In that case (and assuming $\det M$ is non-negative) one can devise a Monte Carlo with *importance sampling*, which means that the random number generator creates gauges field weighted according to $\det M e^{-S_{\mathrm{g}}}$. With importance sampling the weights on the right-hand side of Eq. (9) are independent of the fields, so one can set $w = 1$. Because importance sampling is necessary to make lattice QCD numerically tractable, all numerical work is done in Euclidean spacetime.

Importance sampling works well if $\det M$ is positive. For pairs of equal-mass quarks, this is easy to achieve. With the Wilson action, Eq. (5), $\gamma_5 M \gamma_5 = M^\dagger$. Since this is a similarity transformation, M and $M^\dagger$ have the same physical content. With M for one flavor and $M^\dagger$ for the other (of same mass), the fermion determinant is $\det(M^\dagger M)$, which is obviously non-negative. The same argument holds for Neuberger's discretization,[34] which is computationally more demanding, but has better chiral symmetry (cf. Sec. 6). For the Kogut-Susskind quark action,[35] the matrix M_{KS} is non-negative, but each Kogut-Susskind field creates 4 fermion species in the continuum limit.

Thus, most calculations of $\det M$ are for 2 or 4 flavors. The physically desirable situation with three flavors, with the strange quark's mass different from that of two lighter quarks, is difficult to achieve. One way is to cope with occasionally negative weights.[36] Some algorithms for generating the gauge fields set up a guided random walk with a finite step size ϵ.[37-39] In a widely used scheme for introducing the fermions,[39] they can generate weights such as $(\det M^\dagger M)^{1/2}$ or $(\det M_{\mathrm{KS}})^{1/4}$, which formally give a single flavor. There is some evidence that there may be subtleties associated with non-zero ϵ in large systems with small quark masses.[40] These potential problems, and also the physical interpretation of the fractional powers, could be monitored by looking at the pattern of spontaneously broken chiral symmetry, cf. Sec. 6.

The choice of imaginary time has an important practical advantage. Consider the two-point correlation function

$$C_2(t) = \langle 0|\Phi_H(t)\Phi_H^\dagger(0)|0\rangle, \tag{12}$$

where Φ_H is an operator with the quantum numbers of the hadron of interest, H. For simplicity, assume $t > 0$ and take the total spatial momentum to vanish. Inserting a complete set of eiegnstates of the Hamiltonian between Φ_H and $\Phi_H^\dagger$,

$$C_2(t) = \sum_n \langle 0|\Phi_H|H_n\rangle\langle H_n|\Phi_H^\dagger|0\rangle e^{im_{H_n}t}, \tag{13}$$

where m_{H_n} is the mass of $|H_n\rangle$, the nth radial excitation with H's quantum numbers. For real t it would be difficult to disentangle all these contributions. If, however, $t = ix_4$, with x_4 real and positive, then one has a sum of damped exponentials. For large x_4 the lowest-lying state dominates and

$$C_2(x_4) = |\langle 0|\Phi_H|H\rangle|^2 e^{-m_H x_4} + \cdots, \tag{14}$$

where $|H\rangle$ is the lowest-lying state and m_H its mass. The omitted terms are exponentially suppressed. It is straightforward to test when the first term dominates a numerically computed correlation function, and then fit the exponential form to obtain the mass.

This technique for isolating the lowest-lying state is also essential for obtaining hadronic matrix elements. For a transition from one hadron H to another H', one must compute the matrix element $\langle H'|Q|H\rangle$, where Q is the operator inducing the transition. In flavor phenomenology, Q is a term in the electroweak Hamiltonian; for moments of parton densities, Q is a local operator appearing in the operator product expansion of two currents. One uses a three-point correlation function

$$C_{H'QH}(x_4, y_4) = \langle 0|\Phi_{H'}(x_4 + y_4)Q(y_4)\Phi_H^\dagger(0)|0\rangle, \tag{15}$$

where only the Euclidean times of the operators have been written out. Inserting complete sets of states and taking x_4 and y_4 large enough,

$$C_{H'QH}(x_4, y_4) = \langle 0|\Phi_{H'}|H'\rangle\langle H'|Q|H\rangle\langle H|\Phi_H^\dagger|0\rangle e^{-m_{H'}x_4 - m_H y_4}. \qquad (16)$$

The amplitudes $\langle 0|\Phi_{H'}|H'\rangle$ and $\langle H|\Phi_B^\dagger|0\rangle$ and the masses $m_{H'}$ and m_H are obtained from two-point correlation functions C_2, leaving $\langle H'|Q|H\rangle$ to be determined from $C_{H'QH}$. To compute amplitudes for a transition from H to the vacuum (as in a leptonic decay), one can simply replace Φ_H in C_2 with the charged current.

Equations (12)–(16) assumed $t > 0$, $x_4 + y_4 > y_4 > 0$, *and* $L_4 \to \infty$. With finite L_4 other time orderings lead to terms with $e^{-m(L_4 - t)}$. They are straightforward to derive and to incorporate into fits. Thus, these details do not alter the basic paradigm for computing masses and matrix elements.

These methods are conceptually clean and technically feasible for calculating masses and hadronic matrix elements with at most one hadron in the final state. The procedure for computing correlation functions is as follows. First generate an ensemble of lattice gluon fields with the appropriate weight. Next form the desired product $O_1 \cdots O_n$, with quark variables exactly integrated out to form propagators M^{-1}. Then take the average over the ensemble. Finally, fit the Euclidean time dependence of Eqs. (14) and (16). With two hadrons in the final state, correlation functions can be obtained in more or less the same way, but the interpretation of the energies and amplitudes is more complicated, as discussed in Sec. 7.2.

Within the same ensemble, there are correlations in the statistical fluctuations of the quantities calculated. Methods, such as bootstrap and jackknife, that propagate correlations through the analysis are well understood and widely used. So, these days, statistical errors rarely lead to controversy. As discussed in Sec. 3, it is not practical to carry out the Monte Carlo calculations at very small lattice spacings or at very small quark masses. To gain control over these effects (using effective field theory to guide extrapolations to the physical limit) requires small statistical errors on the raw output of the Monte Carlo calculation.

As mentioned above, the computation of the factor $\det M$ in Eq. (11) is very demanding. The determinant generates sea quarks inside a hadron. It is thus tempting to replace $\det M$ with 1 *and* compensate the corresponding omission of the sea quarks with shifts in the bare couplings. This approximation is most often called the quenched approximation.[41] A more vivid name is the valence approximation,[42] which stresses that the valence quarks (and gluons) in hadrons are treated fully, and the sea quarks merely modeled. The idea is analogous to a dielectric approximation in electromagnetism, and it fails under similar circumstances. In particular, if one is interested in comparing

two quantities that are sensitive to different energy scales, one cannot expect the same dielectric shift to suffice.

It is not easy to estimate quantitatively the effect of quenching. The quenched approximation can be cast as the first term in a convergent expansion,[43] providing a method to compute the shifts in the couplings, and further corrections. The computed shift agrees with the empirical one, but it is about as difficult to compute the next term as to restore the fermion determinant. For some quantities one can estimate the short-distance contribution to the quenching shift. Examples include the strong coupling α_s[44] the quark masses[45–47], and $\mathcal{F}(1)$, which is a form factor needed to determine the CKM matrix element V_{cb}.[48] It is the long-distance part which is harder to fathom.

The quenched approximation is going away. In heavy quark physics the CP-PACS[49,50] and MILC[51] collaborations have unquenched calculations of the heavy-light decay constants f_B, f_{B_s}, f_D, and f_{D_s}. Both groups have results at several lattice spacings, so they can study the a dependence. Their results are about 10–15% higher than the most mature estimates from the quenched approximation. In addition, the Rome group has an unquenched calculation of the b quark mass, which agree well with their quenched calculation.[52] There are also unquenched calculations of moments of parton densities.[31]

3   Why Effective Field Theories?

In this section we discuss why it is necessary, as a practical matter, to consider effective field theories. The first clue is that the physical problem has many scales, so one should think of renormalization-group strategies to tackle them. Effective field theory is one of the most powerful such tools. The central idea is to introduce a separation scale (in energy units) μ. Effects from distances shorter than μ^{-1} are lumped into the couplings (or short-distance coefficients) of the effective field theory, whereas effects from distances longer than μ^{-1} are described by operators in the effective field theory. The degrees of freedom in the effective field theory are those required to reproduce the singularities of thresholds,[53] and so on, of the underlying theory, when only processes of energy $E < \mu$ are considered. By demanding that physical results do not depend on μ, and by matching the effective to the underlying theory, one can avoid over- or under-counting contributions at the interface of "long" and "short."

In QCD, the energy scale characteristic of non-perturbative gluonic effects is called Λ, and it lies in a range from the asymptotic freedom parameter $\Lambda_{\overline{\text{MS}}} \approx 250$ MeV through the ρ meson mass $m_\rho = 770$ MeV to the scale of chiral symmetry breaking $m_K^2/m_s \approx 2500$ MeV. For concreteness, we shall think of $\Lambda \approx 750$ MeV, allowing leeway of a factor of three where appropriate. QCD also has quarks, whose masses range widely. Light quarks are those with $m_q < \Lambda$ or even $m_q \ll \Lambda$; the mass of the strange quark is about 100 MeV—

seven or eight times smaller than Λ, and the masses of the up and down quarks
are about 25 times smaller still. Heavy quarks are those with $m_Q > \Lambda$ or even
$m_Q \gg \Lambda$; the mass of the bottom quark is 4–5 GeV—roughly six times larger
than Λ, and the mass of the top quark is about 40 times larger still. The mass
of the charmed quark is slightly larger than Λ, $m_c \approx 1.3$ GeV. Whether, or
under what circumstances, the charmed quark can be treated as heavy is an
open question. Systems with valence charmed quarks seem to enjoy some of
the simplifications of analogous systems with valence b quarks. On the other
hand, a sea of $c\bar{c}$ pairs could play a role in the high-momentum tail of hadronic
wave functions, where not much is known.

In numerical calculations, one also has cutoffs. The lattice has a non-zero
spacing a, corresponding to a ultraviolet cutoff $\sim \pi/a$. The spacetime volume
has a finite size $L^3 L_4$, corresponding to a infrared cutoff $(L^3 L_4)^{1/4}$. With
arbitrarily large computer memory and processing power, one could imagine
taking these scales in the hierarchy

$$L^{-1} \ll m_q \ll \Lambda \ll m_Q \ll a^{-1}, \tag{17}$$

with the quark masses adjusted to their physical values. (Because quarks are
confined, the adjustment is made by tuning one hadron mass for each flavor of
quark.) In that case one would only need to know a little about cutoff effects:
just enough to be confident that the cutoffs introduce only small deviations
from the limits $a \to 0$ and $L \to \infty$.

In practice, however, this idealized situation cannot be achieved. Finite
computer memory and processor power limit both a and L. To get a feel
for the tradeoffs, one needs only a few simple scaling laws for the computer
algorithms. First, the amount of memory needed grows as

$$\texttt{memory} \propto N_S^3 N_4 = L^3 L_4 / a^4. \tag{18}$$

Increasing either the physical volume L^3 or decreasing the lattice spacing a
puts great demands on the memory. The large exponents in Eq. (18) are
unavoidable, because they stem from the fact that we live in $3 + 1$ dimensions.
Second, the amount of CPU time needed to create statistically independent
lattice gauge fields (with L fixed) grows as

$$\tau_g \propto a^{-(4+z)}. \tag{19}$$

The 4 in the exponent arises because the number of variables to process grows
as a^{-4}. In addition, the update algorithms slow down as $a \to 0$, because they
update in a region of size a, but must propagate these changes over physical
regions of size Λ^{-1} to get a statistically independent gauge field. Thus, the
exponent $z > 0$, and available algorithms have z around 1 or 2.[54]

In addition to creating gauge fields, one must compute quark propagators in the background gauge field. These propagators are needed for the valence quarks in any hadron. The numerical problem is to solve

$$MG = S \tag{20}$$

for G, given some source S, where M is the discretized Dirac operator. M, G, and S are $\mathbb{N} \times \mathbb{N}$ matrices, where $\mathbb{N} \propto N_S^3 N_4$ is number of quark degrees of freedom. The matrix M is sparse (meaning that most of the entries vanish), because it is practical only to put the most local interactions into the lattice Lagrangian. The CPU time needed to solve for G, even with the best algorithms, is

$$\tau_q \propto (\lambda_{\max}/\lambda_{\min})^p \sim \min\left\{1/(m_q a)^p, (L/a)^p\right\}. \tag{21}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of M, and m_q is the quark mass. The exponent p depends on the algorithm and is typically 1 or 2. Equation (20) is also needed in algorithms to incorporate sea quarks correctly,[37−39,55−57] where the effective exponent p is 2 or 3. At fixed a it is, therefore, costly to reduce m_q. For suitable boundary conditions, the volume term in Eq. (21) can take over when $m_q < L^{-1}$, but this regime has significant finite size effects, so it is not suited to general-purpose hadron phenomenology. In addition to difficulties with finite-size effects and with the chiral slowdown of the algorithms, statistical uncertainties increase as the quark mass decreases. The bottom line is that it is not practical to run the computer at masses as small as those of the up and down quarks in nature.

In typical calculations, these days, $N_S = 16$–32, with N_4 the same or a few times larger. To balance the infrared and ultraviolet cutoffs, one ends up with $a^{-1} \sim 1$–4 GeV (so $\pi/a \sim 3$–12 GeV) and $L \sim 1$–4 fm. In typical calculations the "light" quarks have mass in the range 0.2–$0.4 < m_q/m_s < 1.2$. Thus, in practice, the idealized hierarchy (17) becomes

$$L^{-1} < m_q < \Lambda \ll m_b \sim a^{-1}, \tag{22}$$

which is sketched in Fig. 2. Although the various scales are not as well separated as in the idealized hierarchy (17), there is still some separation. Thus, one has a chance of using effective field theory to go from sequences of calculations roughly in the hierarchy (22) to continuum, infinite-volume QCD with the quark mass of the real world.

The large masses of the bottom and charmed quarks make the numerical solution of Eq. (20) relatively easy, but heavy-quark discretization effects become a difficult problem, and a topic of some debate. Instead of studying $m_Q \sim a^{-1}$ directly, some groups set $m_Q < m_b$ (and, indeed, $m_Q \lesssim m_c$),
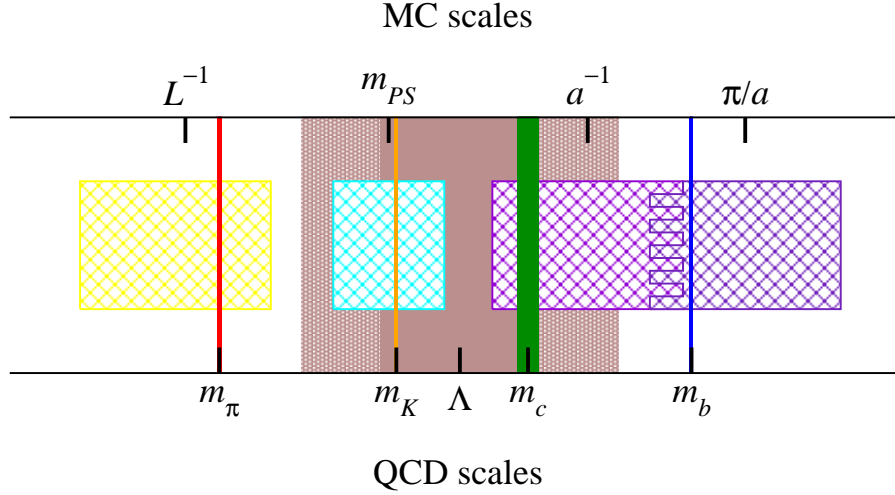
Figure 2: Scales in QCD, and in numerical lattice calculations. Physical scales of QCD are labeled on the bottom, and are indicated by solid colors. Scales from practical Monte Carlo calculations are labeled at the top, and indicated by cross-hatched patches.

leading to the hierarchy

$$L^{-1} < m_q \lesssim \Lambda < m_Q < a^{-1}. \tag{23}$$

Another approach for heavy quarks is the static approximation, $m_Q \to \infty$. Methods for heavy quarks are discussed further in Sec. 5.

In summary, practical limitations of computers constrain the lattice spacing, the quark masses, and the box size to the hierarchy (22) or (23) instead of the idealized hierarchy (17). Nevertheless, these parameters can all be varied over a certain range, providing numerical lattice QCD with one of its most important strengths. The paradigm is as follows: the computer generates numerical data, varying each of a, m_Q, m_q, and L. These data must then be analyzed to extract QCD, at least as long as the data start close enough to the real world. With sound theoretical guides, this paradigm is practical, and allows propagation of errors.[58] In each case, the guide comes from an effective field theory. Moreover, one can test the functional form anticipated from an effective field theory and then—assuming the test succeeds—the extrapolation to the physical limit is justified. Indeed, this paradigm uses limited computer power much more effectively than when relying on brute force alone.

For reference, the most important effective field theories are listed in Table 1. Quark mass dependence relies on methods that are also used in continuum QCD: chiral perturbation theory (χPT) for light hadrons, and heavy-

Table 1: Effective field theories used in controlling and quantifying systematic uncertainties of lattice calculations. (NRQCD is used for bound states of a heavy quark and a heavy anti-quark, called quarkonium, where the small parameter is the relative velocity ν between the constituents.)

scale	EFT	small parameter
a	Symanzik effective field theory	Λa
m_Q	heavy-quark effective theory (HQET)	$\Lambda/2m_Q$
	non-relativistic QCD (NRQCD)	ν
m_q	chiral perturbation theory (χPT)	$(m_K/4\pi f_\pi)^2 \sim m_s/\Lambda$
L	Lüscher effective field theory	$e^{-m_\pi L}$, $(L\Lambda)^{-1}$

quark effective theory (HQET) and non-relativistic QCD (NRQCD) for heavy quark systems. A glance at Fig. 2 shows that in most cases the scales are not very far apart. The leading term in each effective theory may not be enough to give a good guide to the extrapolation, but it is possible to work out higher-order expressions analytically and use them.

In some cases it makes sense to combine two or more of the tools. As mentioned above, on typical lattices the b quark satisfies $m_b a \sim 1$, so an effective theory treating both a and m_b^{-1} as short distance should be useful. This idea is developed in Sec. 5. Most formulations of lattice fermions explicitly break some of the (chiral) flavor symmetries of QCD. To study the interplay of the continuum limit and spontaneous chiral symmetry breaking, one can modify the chiral Lagrangian to parametrize the lattice's explicit chiral symmetry breaking. We shall return to this point in Sec. 6. Finally, when masses of pseudo-Goldstone bosons (π, K, η in nature) become small compared to the volume, it becomes necessary to study their finite-size effects with chiral perturbation theory. This topic is mentioned briefly in Sec. 7.

4 Lattice Spacing Effects: Symanzik Effective Field Theory

The most obvious difference between lattice gauge theory and continuum QCD is the non-zero lattice spacing. It is often (correctly) said that lattice field theories *define* quantum field theories. But the definition entails taking a sequence of lattice theories, with varying lattice spacing, and taking the limit $a \to 0$ with physical quantities held fixed.[11] To interpret computer calculations at non-zero a, however, what one really needs is a description of cutoff effects. This section discusses such a description, based on an effective field theory invented by Symanzik.[59,60] It provides simple semi-quantitative estimates of lattice spacing effects. More interestingly, it provides strategies for eliminating them, both by parametrically reducing their size, and by giving a framework

for combining results from several lattice spacings.

One can get some feeling for lattice spacing effects by looking at the quark-gluon vertex. With Wilson's action for lattice fermions, Eq. (5), the vertex is

$$
\begin{aligned}
\Gamma_\mu(p, p') &= -g_0 t^a \left\{ \gamma_\mu \cos[\tfrac{1}{2}(p + p')_\mu a] - i \sin[\tfrac{1}{2}(p + p')_\mu a] \right\} \\
&= -g_0 t^a \left\{ \gamma_\mu - \tfrac{i}{2} a (p + p')_\mu + O(a^2) \right\}.
\end{aligned} \tag{24}
$$

In a hadron at rest, the typical momenta are of order Λ. So, if Λa is around 0.2, the quark-gluon interaction deviates by about 20% from the continuum.

Recall, however, that the lattice action is not unique. In 1985, Sheikholeslami and Wohlert[61] suggested adding another interaction to the Wilson action, namely a lattice approximant to $i \bar{q} \sigma \cdot F q$, with coefficient $c_{\text{SW}}/4$. Then the quark-gluon vertex becomes

$$
\begin{aligned}
\Gamma_\mu(p, p') &= -g_0 t^a \left\{ \gamma_\mu \cos[\tfrac{1}{2}(p + p')_\mu a] - i \sin[\tfrac{1}{2}(p + p')_\mu a] \right. \\
&\qquad \left. + \tfrac{1}{2} c_{\text{SW}} \sigma_{\mu\nu} \cos[\tfrac{1}{2} k_\mu a] \sin[k_\nu a] \right\} \\
&= -g_0 t^a \left\{ \gamma_\mu - \tfrac{i}{2} a \left[(p + p')_\mu + c_{\text{SW}} i \sigma_{\mu\nu} k^\nu \right] + O(a^2) \right\},
\end{aligned} \tag{25}
$$

where $k = p' - p$, $\sigma_{\mu\nu} = i[\gamma_\mu, \gamma_\nu]/2$. On the mass shell, an easy application of the Gordon identity shows that the two sets of $O(a)$ terms cancel if $c_{\text{SW}} = 1$.

The analysis of Eqs. (24) and (25) is essentially classical. For quantum field theory, one would rather have a formalism that allows for renormalization, and also gives a concept of "on shell" that holds for hadrons. A few years after Wilson's paper, Symanzik[59,60] introduced a local effective Lagrangian for analyzing discretization effects. This work grew out of his earlier studies cutoff dependence in field theories, a line of thinking that had led to the Callan-Symanzik equation.[62,63] The idea is that lattice gauge theory, at given a, can be described by a local effective Lagrangian. Short-distance effects, in particular discretization effects, are lumped into short-distance coefficients, whereas long-distance physics is described by local effective operators. The idea is analogous to the usage of effective field theories to parameterize short-distance phenomena whose microscopic dynamics is unknown.

For the Lagrangian of any lattice field theory, Symanzik says to write[59]

$$
\mathcal{L}_{\text{lat}} \doteq \mathcal{L}_{\text{Sym}}, \tag{26}
$$

where the symbol $\doteq$ can be read "has the same on-shell matrix elements as." The left-hand side of Eq. (26) is the lattice field theory inside the computer, whose output must be analyzed to obtain a continuum result. The lattice action can be complicated, with many free parameters. For lattice QCD with Wilson fermions

$$
\mathcal{L}_{\text{lat}} = \mathcal{L}_{\text{W}} + \sum_O a^{\dim O - 4} c_O O_{\text{lat}}, \tag{27}
$$

where $\mathcal{L}_\mathrm{W}$ is the Wilson (gauge and quark) Lagrangian, Eqs. (3)–(5). The Wilson action contains a bare gauge coupling g_0^2 and a (dimensionless) bare mass $m_0 a$. In Eq. (27), the interactions O_lat are gauge-invariant composite operators of lattice gauge and lattice fermion fields. The couplings c_O can be chosen in the same spirit as c_SW in Eq. (25), and we shall explain how the effective field theory provides a systematic framework for making a choice.

The right-hand side of Eq. (26) is a local effective Lagrangian (LE$\mathcal{L}$) used for analyzing the computer output. Its ultraviolet behavior is regulated and renormalized completely separately from the lattice of the left-hand side. It is convenient to think of $\mathcal{L}_\mathrm{Sym}$ as a continuum field theory. The LE$\mathcal{L}$ is the Lagrangian of the corresponding continuum field theory, plus extra terms to describe discretization effects. For lattice QCD

$$\mathcal{L}_\mathrm{Sym} = \mathcal{L}_\mathrm{QCD} + \mathcal{L}_I, \tag{28}$$

where $\mathcal{L}_\mathrm{QCD}$ is the renormalized, continuum QCD Lagrangian,

$$\mathcal{L}_\mathrm{QCD} = \frac{1}{2g^2}\,\mathrm{tr}[F^{\mu\nu}F_{\mu\nu}] - \bar{q}\,(\slashed{D} + m)\,q. \tag{29}$$

The renormalized gauge coupling g^2 and renormalized mass m depend on the bare gauge coupling g_0^2, the bare mass m_0, the c_O in Eq. (27), and the chosen renormalization scheme:

$$g^2 = g^2(g_0^2, m_0 a; c_O; \mu a), \tag{30}$$
$$m = m_0 Z_m(g_0^2, m_0 a; c_O; \mu a), \tag{31}$$

where μ is the renormalization point. The continuum limit is taken for fixed g^2 and m. Lattice artifacts are described by operators of dimension $\dim \mathcal{O} > 4$:

$$\mathcal{L}_I = \sum_\mathcal{O} a^{\dim \mathcal{O} - 4} K_\mathcal{O}(g^2, ma; c_O; \mu a)\,\mathcal{O}_R(\mu), \tag{32}$$

where $a^{\dim \mathcal{O} - 4} K_\mathcal{O}$ is a short-distance coefficient, written with factors of a so that $K_\mathcal{O}$ is dimensionless. As with the renormalized couplings, these short-distance coefficients depend on all couplings of the lattice action.

The renormalized operators $\mathcal{O}_R$ are sensitive to long distances only. In particular, they do *not* depend on the short distance a. Multiplying matrix elements of $\mathcal{O}_R$ with their coefficients, one finds terms of order $(pa)^{\dim \mathcal{O} - 4}$, where p is a typical momentum, and $\dim \mathcal{O} - 4 > 0$. For hadrons consisting of quarks and gluons, $p \sim \Lambda$. By assumption Λa is small, so one can treat the lattice artifacts in $\mathcal{L}_I$ as perturbations. To do so, one can pass to the interaction picture driven by $\mathcal{L}_\mathrm{QCD}$, and in this way develops a series, with all matrix elements taken in the (continuum) eigenstates of $\mathcal{L}_\mathrm{QCD}$.

In the interaction picture, one can simplify $\mathcal{L}_I$ by omitting redundant interactions. Let us focus on the quark part of the Lagrangian, and consider the field redefinitions

$$q \mapsto q + a^{\dim X} \varepsilon_X X q, \quad \bar{q} \mapsto \bar{q} + a^{\dim X} \bar{\varepsilon}_X \bar{q} X, \tag{33}$$

where X is an arbitrary gauge-*co*variant operator, and ε_X and $\bar{\varepsilon}_X$ are free parameters. Equation (33) simply changes the integration variables of the functional integral. On-shell matrix elements, which are the integrals themselves, do not change.[64] Since the interaction picture is being driven by $\mathcal{L}_{\rm QCD}$, the mass shell in question is that of QCD, even though we have not yet solved for the hadron masses.

A trivial example is if X is a constant. Then Eq. (33) changes the normalization of the fields q and $\bar{q}$, so one concludes that on-shell matrix elements do not depend on how the field is normalized. In other cases, the field redefinition acting on $\mathcal{L}_{\rm QCD}$ induces higher-dimension terms, like those in $\mathcal{L}_I$. Thus,

$$\mathcal{L}_I \mapsto \mathcal{L}_I + \sum_X a^{\dim X} \left[\bar{\varepsilon}_X \bar{q} X (\slashed{D} + m_q) q + \varepsilon_X \bar{q} (-\overleftarrow{\slashed{D}} + m_q) X q \right]. \tag{34}$$

Similarly, the redefinition acting on terms in $\mathcal{L}_I$ changes terms of even higher dimension. In summary, Eq. (33) amounts to changing certain coefficients in $\mathcal{L}_I$. Since the changes are arbitrary, the corresponding operators have no effect on on-shell matrix elements. When using the effective field theory to describe the underlying theory, their coefficients may be set according to convenience, so they are called redundant. They are easy to identify, because they vanish by the equations of motion of QCD.

Let us illustrate with dimension-five operators. There are no pure gauge operators, but there are two linearly independent quark operators, namely

$$\mathcal{O}_5 = i\bar{q} \sigma_{\mu\nu} F^{\mu\nu} q, \tag{35}$$
$$\mathcal{O}_5' = 2\bar{q} D^2 q. \tag{36}$$

The second of these can be re-written as

$$\mathcal{O}_5' = \mathcal{O}_5 + 2\bar{q}\slashed{D} \left(\slashed{D} + m \right) q - 2m\bar{q}\slashed{D}q. \tag{37}$$

These three terms are, in order, the other dimension-five operator, a redundant operator, and something proportional to the kinetic term in $\mathcal{L}_{\rm QCD}$. The last can be absorbed into the field normalization of q and a redefinition of m. Thus, $\mathcal{O}_5$ suffices to describe all on-shell dimension-five effects:

$$\mathcal{L}_I = a K_{\sigma \cdot F} \, \bar{q} i \sigma_{\mu\nu} F^{\mu\nu} q + \cdots, \tag{38}$$

where the ellipsis denotes terms of dimension six and higher.

The vector and axial vector currents can be described along a similar lines. Consider, for example, the flavor-changing transition $s \to u$. The lattice currents take the form

$$V_{\text{lat}}^{\mu} = \bar{\psi}_u i\gamma^{\mu}\psi_s - ac_V \partial_{\nu\,\text{lat}}\bar{\psi}_u \sigma^{\mu\nu}\psi_s + \sum_{O_V} a^{\dim O_V - 3} c_{O_V} O_V^{\mu}, \qquad (39)$$

$$A_{\text{lat}}^{\mu} = \bar{\psi}_u i\gamma^{\mu}\gamma_5 \psi_s + ac_A \partial_{\text{lat}}^{\mu}\bar{\psi}_u i\gamma_5\psi_s + \sum_{O_A} a^{\dim O_A - 3} c_{O_A} O_A^{\mu}, \qquad (40)$$

which are general expressions with the right quantum numbers. In the Symanzik effective field theory these currents are described by[65]

$$V_{\text{lat}}^{\mu} \doteq \bar{Z}_V^{-1}\mathcal{V}^{\mu} - aK_V \partial_{\nu}\bar{u}\sigma^{\mu\nu}s + \cdots, \qquad (41)$$

$$A_{\text{lat}}^{\mu} \doteq \bar{Z}_A^{-1}\mathcal{A}^{\mu} + aK_A \partial^{\mu}\bar{u}i\gamma_5 s + \cdots, \qquad (42)$$

where the ellipsis denotes operators of dimension four and higher, $\bar{Z}_J^{-1}$ and K_J are short-distance coefficients, and

$$\mathcal{V}^{\mu} = \bar{u}i\gamma^{\mu}s, \qquad (43)$$

$$\mathcal{A}^{\mu} = \bar{u}i\gamma^{\mu}\gamma_5 s, \qquad (44)$$

are the vector and axial vector currents in continuum QCD. Like the short-distance coefficients in the effective Lagrangian, $\bar{Z}_J$ and K_J are functions of g^2 and ma, the lattice couplings c_O, and the parameters c_{O_J}. Further dimension-four operators may be omitted from Eqs. (41) and (42), because they are linear combinations of those listed and others that vanish by the equations of motion.

Like the terms of dimension five and higher in $\mathcal{L}_I$, the dimension-four currents can be treated as perturbations. Thus, the effective field theory says that

$$\langle f_{\text{lat}}|\bar{Z}_A A_{\text{lat}}^{\mu}|i_{\text{lat}}\rangle = \langle f|\mathcal{A}^{\mu}|i\rangle + a\bar{Z}_A K_A \langle f|\bar{u}i\gamma^{\mu}\gamma_5 s|i\rangle \qquad (45)$$

$$+ aK_{\sigma.F}\int d^4x \,\langle f|T\,\mathcal{O}_5\mathcal{A}^{\mu}|i\rangle + O(a^2),$$

and similarly for the vector current. The crucial feature of Eq. (45) is that, while the states on the left-hand side are eigenstates of lattice gauge theory, those on the right-hand side are eigenstates of continuum QCD.

The Symanzik effective field theory can be justified in to all orders in perturbation theory (in the gauge coupling). In the late 1980's Reisz proved a power counting theorem[66] that enabled him to set up a version of BPHZ (Bogoliubov-Parasiuk-Hepp-Zimmermann) renormalization tailored to the Feynman diagrams of lattice perturbation theory. With mild assumptions

on the gluon and quark propagators,[d] he showed that loop integrals with external momenta $\{p\}$ and loop momenta $\{k\}$ can be written[67]

$$\int \prod_{i=1}^{l} \frac{d^4 k_i}{(2\pi)^4} \, \mathcal{I}(\{p\}, \{k\}) = I_R + I_U(\{p\}) + O(ap, am), \tag{46}$$

where I_R denotes parts that are absorbed when renormalizing the gauge coupling and the quark masses. The other term $I_U(\{p\})$ does not depend on a or on the coefficients c_O of higher-dimension lattice interactions. Thus, after renormalization, lattice perturbation theory is universal:[68] the continuum limit does not depend on the discretization. The remainder terms can be developed further with BPHZ oversubtractions of the loop integrands. In this way one can develop any amplitude's renormalized perturbation series, including contributions suppressed by powers of a, to any order desired. Similarly, one can develop the renormalized perturbation series generated by Symanzik's LE$\mathcal{L}$. To lend rigor to the Symanzik theory, one could imagine defining the LE$\mathcal{L}$ on an infinitesimally fine lattice, and using Reisz's methods to extract the universal part, and also to define the operator insertions of $\mathcal{L}_I$. But in practice any method of regulating and renormalizing the ultraviolet will do.

Although it is only fully justified in perturbation theory, the Symanzik effective field theory is believed to hold at a non-perturbative level as well. At short distances, small instantons make contributions that should be lumped into the short-distance coefficients. But they come in with a high power of a and are presumably negligible. Long-distance phenomena, such as confinement, are more mysterious. But this mystery is bundled into the QCD term in the LE$\mathcal{L}$, not $\mathcal{L}_I$. Indeed, if the Symanzik formalism were to fall apart, it would be hard to understand why other short-distance methods of QCD work so well to explain measurements of deeply inelastic scattering, jet cross sections, or B decays.

By calculating on-shell matrix elements on both sides of, say, Eq. (45) in renormalized perturbation theory, one can read off the short-distance coefficients in $\mathcal{L}_I$. In perturbation theory, let

$$K_O(g^2, ma; c_O; \mu a) = g^{2n} \sum_{l=0}^{\infty} g^{2l} K_O^{[l]}(ma; c_O; \mu a), \tag{47}$$

where $n \geq 0$: in some cases the coefficients vanish at the tree level. From the justification of the LE$\mathcal{L}$ given above, it is natural to use the renormalized coupling in this series. An especially transparent choice of g^2 is something physical, at least in perturbation theory, for example the coupling that appears

---

[d]Note that Wilson fermions satisfy the assumptions, but staggered fermions do not.

in scattering of two quarks of different flavor. Then from familiar properties of perturbation theory

$$K_{\mathcal{O}}^{[l]}(ma; c_O; \mu a) = \sum_{r=0}^{l} K_{\mathcal{O}}^{[l,r]}(ma; c_O) \, (\ln \mu a)^r. \tag{48}$$

The μ dependence must arise in the right way to cancel the μ dependence of the operators $\mathcal{O}_R$ in $\mathcal{L}_I$.

In the discussion so far, the Symanzik effective field theory is simply a theory of cutoff effects.[59] It applies no matter how the lattice couplings c_O are chosen. The default is $c_O = 0$: most higher-dimension lattice interactions are omitted from the Lagrangian inside the computer. At this level, the Symanzik theory yields, with great sophistication, the result that on-shell matrix elements suffer lattice artifacts suppressed by powers of a.

A lattice gauge theory with smaller short-distance coefficients in the LE$\mathcal{L}$ would yield better estimates of continuum QCD. This is where the Symanzik formalism becomes especially powerful, because one can demand $K_{\mathcal{O}} = 0$, and solve for the lattice couplings c_O.[60] This is called the Symanzik improvement program. A key feature is that if a coefficient $K_{\mathcal{O}}$ (or its expansion coefficient $K_{\mathcal{O}}^{[l]}$) is made to vanish for one observable, then the effective field theory set up shows that it vanishes for all observables.

Of course, it is impractical to solve the equations $K_{\mathcal{O}} = 0$ in any kind of generality. At the very least, one must truncate in (scaling) dimension. For QCD, asymptotic freedom guarantees that the anomalous dimensions are small, so the scaling dimension is not much different from the classical dimension. At dimension five there is only one term in $\mathcal{L}_I$, cf. Eq. (38), and with the Wilson action, $K_{\sigma \cdot F} \neq 0$. But, following Sheikholeslami and Wohlert,[61] one can write down a discretization of $\mathcal{O}_5$,

$$O_5 = i\bar{\psi}\sigma_{\mu\nu}G^{\mu\nu}\psi, \tag{49}$$

where $G^{\mu\nu}$ is a combination of SU(3) U matrices, defined in Eq. (1), such that the naive $a \to 0$ limit gives $F^{\mu\nu}$. The Sheikholeslami-Wohlert Lagrangian

$$\mathcal{L}_{\mathrm{SW}} = \mathcal{L}_{\mathrm{W}} + \tfrac{1}{4}c_{\mathrm{SW}}O_5. \tag{50}$$

The Symanzik coefficient $K_{\sigma \cdot F}$ is, of course, very sensitive to the coupling c_{SW}. At the tree level

$$K_{\sigma \cdot F}^{[0]}(ma) = \tfrac{1}{4}(1 - c_{\mathrm{SW}}). \tag{51}$$

Thus, as in Eq. (25), setting $c_{\mathrm{SW}} = 1$ in the computer calculations reduces the leading lattice artifact.

In the foregoing discussion, the short-distance coefficients are written as functions of ma, to emphasize that the primary goal of the effective field theory

is to separate long- and short-distance scales. For light quarks, however, it is safe to expand the short-distance coefficients in powers of ma. One can then use this expansion and the ordering of $\mathcal{L}_I$ by dimension to develop asymptotic expansions in powers of a. This is the most widely recognized application of the Symanzik effective field theory. Indeed, many reviews of the formalism skip over the scale-separation aspect and focus entirely on the lattice-spacing expansion.

To eliminate all terms of order a, the improvement program is relatively straightforward. One needs only the first two terms in the small a expansion of the normalization factors

$$\bar{Z}_J(ma) = Z_J \left[1 + b_J \tfrac{1}{2}(m_u + m_s)a \right] + O(m^2 a^2). \tag{52}$$

For heavy quarks, the expansion in $m_Q a$ only makes sense if $m_Q a \ll 1$. As discussed in Sec. 3, this situation is not easy to attain for the b quark. We shall return to further aspects of the Symanzik LE$\mathcal{L}$ for heavy quarks in Sec. 5. But for light quarks Eq. (52) is sensible and useful. In the same vein, it is consistent to replace $K_{\sigma.F}$ and K_J with their values at $m_q a = 0$. This set of choices is called $O(a)$ improvement.

At the tree level, for currents given by just the first two terms of Eqs. (39) and (40), respectively, one finds Eq. (51)

$$Z_J^{[0]} = 1, \tag{53}$$

$$b_J^{[0]} = 1, \tag{54}$$

$$K_J^{[0]} = c_J^{[0]}. \tag{55}$$

The one-loop corrections to the $O(a)$ terms have been calculated, with the improved lattice Lagrangian $\mathcal{L}_{\mathrm{SW}}$, for the Lagrangian itself[69] and for the currents.[70,71] The results are not especially informative, except[71]

$$K_V^{[1]} = c_V^{[1]} + C_F 0.01225, \tag{56}$$

$$K_A^{[1]} = c_A^{[1]} + C_F 0.00568. \tag{57}$$

Thus, one can obtain the desired improvement condition $K_V = K_A = 0$ by setting $c_V = 0 - g^2 C_F 0.01225 + O(g^4)$, $c_A = 0 - g^2 C_F 0.00568 + O(g^4)$.

An important development of recent years are methods for computing the short-distance coefficients, through $O(a)$, non-perturbatively. There are two key ingredients. The first is chiral symmetry. In 1985 it was observed[72] that, as $ma \to 0$, both Z_V and Z_A can be computed non-perturbatively by imposing the chiral Ward identities. For flavor conserving currents, $\bar{Z}_V$ is simply the factor that normalizes the flavor charge. Then, because chiral Ward identities relate vector and axial vector correlation functions, they fix Z_A. The second key

ingredient is the Symanzik effective field theory, which provides a framework for imposing chiral symmetry through $O(a)$.[65,73]

To compute the $O(a)$ terms, one can exploit the partially conserved axial current (PCAC).[73] In continuum QCD the PCAC relation is

$$\partial_\mu \mathcal{A}^\mu - (m_s + m_u)\mathcal{P} = 0, \tag{58}$$

where the pseudoscalar density $\mathcal{P} = i\bar{u}\gamma_5 s$. The implication of writing Eq. (58) as an operator equation is that it holds for all combinations of initial and final states. A short manipulation with the Symanzik effective theory shows that

$$\langle f | \bar{Z}_A \partial_{\mu\,\mathrm{lat}} A^\mu_{\mathrm{lat}} - (m_s + m_u)\bar{Z}_P P_{\mathrm{lat}} | i \rangle = \tag{59}$$
$$a\bar{Z}_A K_A (m_u + m_s)^2 \langle f | \mathcal{P} | i \rangle + 2a K_{\sigma \cdot F} \langle f | \bar{u}\sigma \cdot F\gamma_5 s | i \rangle + O(a^2),$$

for any initial and final states. By choosing three combinations of states, one may use one to eliminate $(m_u + m_s)\bar{Z}_P/\bar{Z}_A$, and then use the other two to adjust c_A (introduced in Eq. (40)) and c_{SW} so that the right-hand side vanishes, as desired for continuum QCD. Because the renormalized mass drops out, the resulting conditions on c_A and c_{SW} do not depend on the renormalization scheme of the effective theory. In a non-perturbative calculations, however, the $O(a^2)$ terms are ever present, so c_A and c_{SW} are determined only with an accuracy of order Λa, coming from the $O(a^2)$ matrix elements on the right-hand side of Eq. (59). The improvement coefficient of the vector current, c_V, and the normalization constants Z_J and b_J are then determined by imposing Ward identities.

The non-perturbative calculation of c_{SW} is straightforward,[74] and the value obtained is relatively insensitive to details such as the states used in Eq. (59). It also agrees well with perturbation theory.[69] On the other hand, the non-perturbative calculation of c_A is not so straightforward—two different groups find marginal agreement,[74,75] and it has been found[76] to depend on the difference operator $\partial^\mu_{\mathrm{lat}}$. These difficulties are, perhaps, to be expected, since in Eq. (59) K_A is multiplied with the small quantity $(m_u + m_s)^2$. The non-perturbative results also do not agree well with perturbation theory. The situation for c_V is similarly unsettled.[75] All estimates for c_A and c_V yield small values, but they multiply large matrix elements. For example,[76] in computing f_π the small correction is $c_A a m_K^2/m_s$, and $m_K^2/m_s = 2.5$ GeV.

In the non-perturbative improvement program, symmetries of continuum QCD were imposed to determine improvement coefficients of the lattice action and currents. Another example of the interplay of symmetry and Symanzik's theory comes in the lattice calculation of the kaon bag parameter B_K, which arise in the theory of K^0-$\bar{K}^0$ mixing. B_K is defined by

$$\langle \bar{K} | \mathcal{Q} | K \rangle = \tfrac{8}{3} m_K^2 f_K^2 B_K \tag{60}$$

where the $\Delta S = 2$ four-quark operator

$$\mathcal{Q} = \bar{s}\gamma^\mu(1-\gamma_5)d\,\bar{s}\gamma_\mu(1-\gamma_5)d, \tag{61}$$

and the kaon decay constant $f_K \approx 160$ MeV is defined through

$$\langle 0|\mathcal{A}^\mu|K\rangle = ip_K^\mu f_K. \tag{62}$$

To calculate f_K and B_K one starts with A_{lat}^μ and a lattice approximant to $\mathcal{Q}$, which we will call $Q^{(6)}$. In the Symanzik effective field theory, the lattice axial vector current is described by Eq. (42) and $Q^{(6)}$ by

$$Q_i^{(6)} \doteq Z_{ij}^{-1}\left[\mathcal{Q}_j^{(6)} + aC_{jk}\mathcal{Q}_k^{(7)}\right], \tag{63}$$

where Z_{ij}^{-1} and C_{jk} are short-distance coefficients. The chiral flavor symmetry group in lattice gauge theory is usually smaller than $\mathrm{SU}(n_f) \times \mathrm{SU}(n_f)$, so one must allow for mixing between the target operator $\mathcal{Q}$ and several other operators, indexed by the subscripts i, j, k. Similarly, the leading lattice spacing effects are described by several operators $\mathcal{Q}_k^{(7)}$. As usual, the operators on the right-hand side of Eq. (63) are defined in continuum QCD. Thus, in analogy with Eq. (45), the effective theory says

$$\begin{aligned}
\langle \bar{K}_{\mathrm{lat}}|Z_{ij}Q_j|K_{\mathrm{lat}}\rangle &= \langle \bar{K}|\mathcal{Q}_i^{(6)}|K\rangle + aC_{ik}\langle \bar{K}|\mathcal{Q}_j^{(7)}|K\rangle \\
&\quad + aK_{\sigma\cdot F}\int d^4x\,\langle \bar{K}|T\,\mathcal{Q}_i^{(6)}\sum_{q=d,\,s}\mathcal{O}_5(x)|K\rangle + O(a^2),
\end{aligned} \tag{64}$$

where the states on the left-hand side are eigenstates of the lattice theory, whereas the states on the right-hand side are continuum QCD eigenstates. For other four-quark operators the effective theory description also contains operators of dimension less than six, multiplied by $1/a$. Such "power-law divergences" are absent here, because all operators must have $\Delta S = 2$.

Equation (64) holds for all formulations of lattice fermions, but it becomes interesting for Kogut-Susskind (or staggered) fermions. One Kogut-Susskind field produces four flavors in the continuum limit, so the lists of operators $\mathcal{Q}_i^{(s)}$ contain various flavor combinations. The drawback of four flavors is ameliorated by a remnant of exact chiral symmetry as $m \to 0$, cf. Sec. 6 The symmetry requires $K_{\sigma\cdot F}$ to vanish as $m_s a$. On the other hand, the lattice artifacts of the operator do not vanish: the short-distance coefficients C_{ik} are non-zero. The matrix elements $\langle \bar{K}|\mathcal{Q}_j^{(7)}|K\rangle$ appearing in Eq. (64) *do* vanish, however, by the flavor symmetry of (four-flavor) continuum QCD.[77,23,78] Thus, the leading lattice artifacts in the lattice calculation of $\frac{8}{3}m_K^2 f_K^2 B_K$ is of order a^2. A similar argument applies to the matrix elements of the axial vector current giving f_K,

so the leading lattice artifact in B_K, when calculated with staggered quarks, is also of order a^2.

In the last few years, there has been spectacular development in the understanding of chiral lattice gauge theories.[79−81] For vector-like gauge theories like QCD, these developments have led to methods with either very small violations of chiral symmetry[82] or essentially no violations of chiral symmetry.[83,84,34] When these methods are used, the leading lattice spacing errors are of order a^2. Indeed, one of the methods—domain-wall fermions[82,85,86]—has been applied to B_K.[87,88] These calculations also seem to have much smaller $O(a^2)$ effects than state-of-the-art calculations[89,90] with staggered fermions.

With any of several methods—$O(a)$ improved Wilson fermions, staggered fermions, domain-wall fermions, overlap fermions[34]—the leading cutoff effect is of order a^2. To reduce them further, one is confronted with many, many operators of dimension six in the Lagrangian, of dimension five in the currents, etc. It may not be feasible to eliminate all these effects with non-perturbative methods—especially in light of the difficulties with c_A. On the other hand, it is possible to compute these coefficients in perturbation theory.

The technical details of lattice perturbation theory are more cumbersome that in continuum QCD, as seen in the Feynman rule in Eq. (25). But, since the ultraviolet cutoff is built in, the problem is well suited to computer algebra. One can generate vertices and propagators and combine them into diagrams automatically.[91,92] Although the lattice renders the integrals ultraviolet finite, difficulies can arise in the infrared. But these effects are universal, essentially by Reisz's theorem. Thus, once the infrared has been understood for a simple lattice Lagrangian, the improvement interactions merely add algebraic complexity that a computer program can handle. These automated methods were first introduced for pure gauge theory at the one-loop level,[91] but are now being extended to QCD at the two-loop level.[92]

A more controversial issue is the viability of perturbation theory in this context. Because lattice gauge theory is a theoretically consistent whole, it seems at first glance natural to use the bare coupling g_0^2 as the expansion parameter of perturbative series. This choice ignores the fact that one wants to connect lattice gauge theory (at moderate lattice spacing) to continuum QCD. The set up of the Symanzik effective field theory, especially as underpinned by Reisz's work, encourages the use of a renormalized coupling. Furthermore, from a practical point of view, the bare coupling in the Wilson gauge action is a disastrous choice: if one compares perturbative and Monte Carlo calculations of short-distance quantities such as small Wilson loops, the agreement is very poor. On the other hand, renormalized perturbation theory usually describes short-distance properties well, especially when the renormalized coupling is chosen according to the Brodsky-Lepage-Mackenzie (BLM) prescription.[93,94] For example, the differences between one-loop BLM perturbation theory for
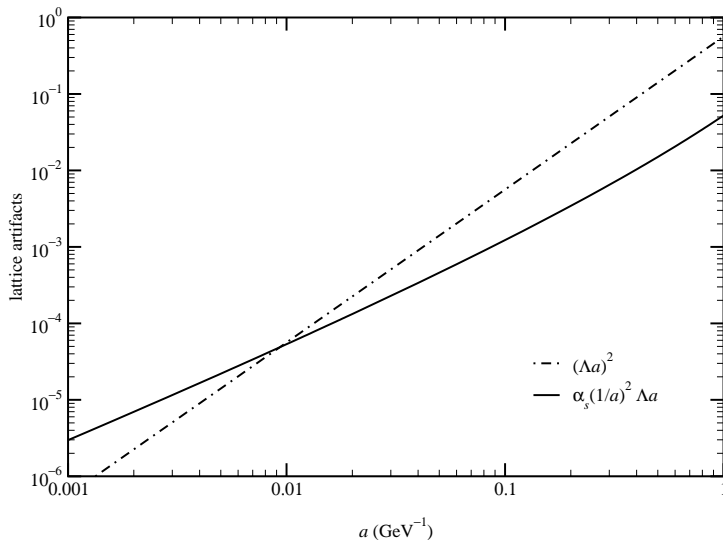
Figure 3: Comparison of the "dominant" lattice artifact $\alpha_s(1/a)^2\Lambda a$ with the "sub-dominant" $(\Lambda a)^2$. Except at unrealistically small a, the latter is larger.

the renormalization and improvement of the vector and axial vector currents is easily attributed to (as yet uncomputed) two-loop effects.[95]

Even assuming perturbation theory is accurate, one is faced with a formal issue. To illustrate it, let us assume that the we have $O(a)$ improvement of Wilson fermions at the one-loop level. Then the leading lattice spacing effects (for light hadron masses) are of order $\alpha_s(1/a)^2\Lambda a$ and $(\Lambda a)^2$. Formally, the former dominates as $a \to 0$. Figure 3 shows $\alpha_s(1/a)^2\Lambda a$ and $(\Lambda a)^2$ as a function of a for $\Lambda = 750$ MeV and $\alpha_s(2$ GeV$) = 0.2$. In the range where lattice calculations can be done, $a > 0.25$ GeV^{-1}, the "sub-dominant" effect is an order of magnitude larger than the "dominant" one. Ideally, one would have enough data to fit to both contributors. Otherwise, one is faced with the choice of introducing a bias, by ignoring $\alpha_s(1/a)^2a$ and fitting to a^2, or introducing an error, by ignoring a^2 and fitting to $\alpha_s(1/a)^2a$ (or a). Figure 3 suggests that the uncertainty stemming from the first Ansatz is larger that the uncertainty stemming from the second.

In the last few years, Symanzik improvement has been a major focus of research in lattice gauge theory. In light of this renaissance of Symanzik's work, it is surprising that many papers still report calculations at only one lattice spacing. The improvement program is based on a theory of cutoff effects, which clearly demonstrates the utility of repeating the calculation at several lattice spacings. To see the benefits, let us consider a simple example. Suppose one

has 100 (arbitrary) units of CPU available. Instead of spending all 100 units on the finest possible lattice spacing a_0, one could consider coarser lattices of spacing $a_1 = a_0/\sqrt[4]{2}$ and $a_2 = a_0/\sqrt{2}$. Spending 65, 25, and 10 units of CPU at a_0, a_1, and a_2 would, according to Eq. (19), yield comparable statistical errors. Compared to putting all 100 units at a_0, the statistical error would be a bit larger, but only by a factor of 1.25 ($= 1/\sqrt{0.65}$). But, if all 100 units are spent at a_0, then one has an uncertainty, of order $(\Lambda a)^2$ say, which requires a guess for the appropriate Λ: 250 MeV or 2.5 GeV? A calculation based on three (or more) lattice spacings does not require a guess, because the added information is tantamount to a calculation of the discretization effect. The slightly larger statistical error seems a small price to pay.

5 Heavy Quark Effects: Heavy-Quark Effective Theory

Some of the most interesting applications of numerical lattice gauge theory arise in heavy quark physics. The main aim of experimental B physics is to study flavor and CP violation precisely enough to test the CKM mechanism. To connect the CKM matrix or, indeed, other short-distance mechanisms of flavor and CP violation, one is faced with theoretical formulae of the form

$$\begin{pmatrix} \text{measured} \\ \text{quantity} \end{pmatrix} = \begin{pmatrix} \text{kinematic} \\ \text{factors} \end{pmatrix} \begin{pmatrix} \text{short-distance} \\ \text{factor} \end{pmatrix} \begin{pmatrix} \text{QCD} \\ \text{factor} \end{pmatrix}, \qquad (65)$$

where the measured quantity is a (differential) rate, and the kinematic factors consist of measurable momenta and (hadron) masses. Here the short-distance factor includes wavelengths less than $(100 \text{ GeV})^{-1}$; in the Standard Model, it consists of well-determined parameters (like the Fermi constant G_F) and the less well-determined CKM matrix. The QCD factor, as a rule, boils down to a hadronic matrix element. In the case of $|V_{ud}|$ and $|V_{us}|$, isospin and SU(3) symmetries bring the QCD under control, but in other cases lattice calculations are essential.[96,97] In some B decays heavy-quark symmetry provides some control, but hadronic matrix elements still appear in contributions at the $1/m_Q$ level, which, generically, are 10% effects.

Measured in lattice units, the bottom and charmed quarks' masses can easily be large. Even if $\Lambda a \sim 0.1$–0.3, so that the Symanzik effective field theory works well for gluons and light quarks, then $m_b a \sim 1$–2 and $m_c a$ about a third of that. For this reason it is frequently (but incorrectly) stated that heavy quarks cannot be directly accomodated by a lattice. This observation overlooks the physical fact that the heavy-quark mass scale is far removed from the QCD scale and that, consequently, the dynamics of heavy-quark systems simplify. Nevertheless, it is fair to say that lattice spacing effects are more challenging for heavy quarks than for light quarks. So, in this section, we

discuss how to classify, control, and minimize discretization uncertainties of heavy quarks.

The key is to make use of effective field theories for heavy-quark systems. Indeed, from the inception of non-relativistic QCD (NRQCD)[98–100] and heavy-quark effective theory (HQET),[101–103] these effective field theories have been used to treat heavy quarks in lattice gauge theory. Indeed, the early papers[98,99,101,102] inspired the development of HQET and NRQCD with continuum ultraviolet regulators, as methods for understanding heavy-light hadrons[104–106] and quarkonia.[107] More recently it has been shown how to use the continuum effective field theories to understand the heavy-quark discretization effects of Wilson fermions.[108–111]

To make a connection with Sec. 4, let us start with Wilson QCD and examine how the Symanzik theory breaks down when $ma \not\ll 1$. A key to the Symanzik LE$\mathcal{L}$ is that the leading dynamics are those of $\mathcal{L}_{\mathrm{QCD}}$, while the corrections $\mathcal{L}_I$ are small. When $ma \not\ll 1$, this split into large+small no longer holds. First, the expansion of short-distance coefficients in small ma is no longer admissible. Furthermore, $\mathcal{L}_I$ contains terms that scale as ma to some power. Consider, in particular,

$$\mathcal{L}_I = \cdots + \sum_X a^{\dim X - 1} \sum_{n=3}^{\infty} K_X^{(n)} \, \bar{q} X \sum_{\mu=1}^{4} (-\gamma_\mu D_\mu a)^n q + \cdots, \qquad (66)$$

which describe deviations from Lorentz (or Euclidean) invariance. (They do respect hypercubic rotations.) At each n, the term with $\mu = 4$ is not small, because $(-\gamma_4 D_4 a)^n \sim (ma)^n$. But one can cull $\gamma_4 D_4$ from $\mathcal{L}_I$ by applying the equation of motion,[112]

$$-\gamma_4 D_4 q = (\boldsymbol{\gamma} \cdot \boldsymbol{D} + m)q. \qquad (67)$$

Repeated application of the equation of motion to yields $a^n (\boldsymbol{\gamma} \cdot \boldsymbol{D} + m)^n$ and (nested) commutators of D_4 and $\boldsymbol{D}$. The commutators do not lead to the heavy-quark mass, but to the gluon field strength and derivatives thereof. Thus, all large terms come from expanding $a^n (\boldsymbol{\gamma} \cdot \boldsymbol{D} + m)^n$ and collecting $(ma)^{n-r} \bar{q} X (\boldsymbol{\gamma} \cdot \boldsymbol{D})^r q$ into a new coefficient for $\bar{q} X (\boldsymbol{\gamma} \cdot \boldsymbol{D})^r q$. If $X = 1$ and $r = 0$ or 1, the coefficients in $\mathcal{L}_{\mathrm{QCD}}$ are modified, and the LE$\mathcal{L}$ takes the form

$$\mathcal{L}_{\mathrm{Sym}} = \mathcal{L}_{\mathrm{gauge}} + \bar{q} \left(\gamma_4 D_4 + \sqrt{\frac{m_1}{m_2}} \boldsymbol{\gamma} \cdot \boldsymbol{D} + m_1 \right) q + \mathcal{L}_I', \qquad (68)$$

where the coefficients m_1 and $\sqrt{m_1/m_2}$ result from coalescing all terms multiplying $\bar{q}q$ and $\bar{q}\boldsymbol{\gamma} \cdot \boldsymbol{D}q$, respectively. The notation is taken from the energy of a quark with small momentum $\boldsymbol{p}$:

$$E(\boldsymbol{p}) = m_1 + \frac{\boldsymbol{p}^2}{2m_2} + O(\boldsymbol{p}^4). \qquad (69)$$

Below m_1 is called the rest mass, and m_2 the kinetic mass. With these rearrangements, the operators in $\mathcal{L}'_I$ now all yield powers of Λa, not ma, so they still can be treated as operator insertions.

Equation (68) rests on the same foundation as Eq. (28); only the split between large and small is different, reflecting the new situation $ma \not\ll 1$. For terms in Eq. (66) with $X \neq 1$, the rearrangement can be absorbed into the short-distance coefficients of $\mathcal{L}'_I$. For the Wilson fermion action, none of these coefficients (except m_1) is unbounded as $ma \to \infty$.[101,108,109] It is important, when developing an improvement program for heavy quarks, not to sacrifice this property. In fact, the currents used in the $O(a)$ improvement program discussed in Sec. 4 do not work well for heavy quarks. The foregoing analysis even suggests a suitable improvement program: the large-mass behavior of $\mathcal{L}'_I$ remains well-behaved if one mimics it and, hence, omits from Eq. (27) operators with extra time difference operators.[108] Similarly, the corrections to the currents should not have any time derivatives at all.[110,111]

Unfortunately, unless $m_1 = m_2$ the LE$\mathcal{L}$ is no longer "QCD plus small corrections," because the normalization of the spatial kinetic energy is incorrect by the factor $\sqrt{m_1/m_2}$. In free field theory, for Wilson fermions,

$$\frac{m_1}{m_2} = 1 - \tfrac{2}{3}m_1^2 a^2 + \tfrac{1}{2}m_1^3 a^3 + \cdots, \tag{70}$$

with no term of order ma. This feature persists to all orders in perturbation theory.[113,110] The deviations from the desired $m_1/m_2 = 1$ can be sizable. For $ma = 0.8$ the two terms shown are -0.46 and $+0.26$. Such strong deviations from continuum QCD remain when the gauge interaction is turned on. The small ma expansions of other short-distance quantities, for example the normalization factors of the currents, also break down as soon as $ma \not\ll 1$.

There are three remedies to this problem. First, one can take $ma \ll 1$, so that the rearrangement discussed above is unnecessary. Second, one can introduce another parameter to the lattice fermion action, so that $\bar{q}\gamma_4 D_4 q$ and $\bar{q}\boldsymbol{\gamma} \cdot \boldsymbol{D} q$ can be normalized separately. Third, one can note that it is not lattice gauge theory that breaks down when $ma \not\ll 1$ but the Symanzik effective field theory. This leaves open the possibility that other tools can be used to control the discretization effects of heavy quarks.

Let us start by considering $ma \ll 1$. In principle, this is fine, because one can expand the short-distance coefficients in ma, and Eq. (68) reverts to "QCD plus small corrections" as in Sec. 4. In practice there are serious difficulties. As discussed in Sec. 2 it will not be possible to reduce a enough to make $m_b a \ll 1$ for many, many years. Another way to reduce ma is to reduce the heavy quark mass. But if $m < m_b$, one must use the heavy-quark expansion to extrapolate back up to m_b.[114,115] The simultaneous requirements $ma \ll 1$ and $\Lambda/m \ll 1$ fight against each other, making it hard, on accessible lattices,

to control both systematic errors, not to mention the crosstalk between them. To obtain $ma < 0.5$, one must take $m < m_c$, sometimes as small as 500 MeV. It is not clear whether this regime can be connected back to m_b through the $1/m$ expansion. To reach larger quark masses, $m \sim 1.5$ GeV, ma is sometimes as large as 0.7–0.8. Then discretization effects of order $(ma)^2$ and $(ma)^3$ are large. Finally, the $(ma)^n$ errors are amplified in the $1/m$ extrapolation, but no one has a solid idea for estimating how much. Concerns of this kind have been voiced before; Sommer[116] and Wittig[117] have insisted on taking $ma \to 0$ before carrying out any extrapolation in $1/m$. Then lattice-spacing and heavy-quark effects are decoupled, and the main drawback of their analyses (of data in the literature), is that the heavy quark mass is too small.

A variation on this theme is to set up a lattice gauge theory with different temporal and spatial lattice spacings, a_t and a_s. Such lattices are called anisotropic. The hope[118] is that the heavy-quark mass appears in short-distance coefficients as ma_t, but not as ma_s. Then one could take $a_t/a_s \sim \Lambda/m_b$, expand the short-distance coefficients in ma_t, and determine the improvement coefficients non-perturbatively.[118] Unfortunately, there is no proof that ma_s does not arise, and, for Klassen's choice of the lattice couplings, it does.[119] When ma_s does appear in coefficients, one cannot take $a_t/a_s \sim \Lambda/m_b$, and another remedy is needed.[e]

The second remedy is to modify the lattice gauge theory so that the temporal and kinetic terms are separately adjustable. A simple way to this was introduced by El-Khadra et al.[108] Then one could adjust the underlying lattice parameters so that in the LE$\mathcal{L}$, $m_1 \equiv m_2$. The adjustment can be made non-perturbatively, by forcing the rest mass and the kinetic mass of a hadron to be the same. It works well,[121] but has not been widely used in heavy-quark phenomenology. In this approach the ma dependence of the short-distance coefficients is not as simple as for light quarks. But at least it is possible to set $m = m_b$, circumventing the heavy-quark extrapolation.

The third remedy is to set up the calculation so that heavy-quark methods (HQET or NRQCD) and lattice gauge theory work together. There are two ways to go about this. One is to derive the heavy-quark theory *a priori* in the continuum, and then replace the derivatives with difference operators.[99–103] These methods are called lattice NRQCD and lattice HQET.[f] The other is to note[108,109] that Wilson fermions possess the same heavy-quark symmetries[124,125] as continuum QCD. Thus, correlation functions computed

---

[e] Even if they do not tame heavy-quark cutoff effects, anisotropic lattices still can be useful for reducing statistical uncertainties, by providing more timeslices in the region of Euclidean time where one state saturates Eq. (14).[120]

[f] In lattice HQET one treats the $1/m$ corrections as insertions.[103] The statistical errors are smaller, however, if one puts the kinetic term into the quark propagator.[122,123] This is still called lattice NRQCD, even when HQET counting is used to classify the $1/m$ expansion for heavy-light systems.

with lattice gauge theory can be described *a posteriori* by HQET (or NRQCD), with a logic and structure parallel to the heavy-quark theory for continuum QCD.[109–111] Heavy quark symmetry persists for all ma, so the HQET description can be developed for all ma, with minor modifications that are discussed below. These ideas give a systematic procedure for matching lattice gauge theory to QCD. It is sometimes called the non-relativistic interpretation of Wilson fermions, and sometimes called the Fermilab method.

For CKM phenomenology, hadrons with one heavy quark are of greatest interest. The most important scales are Λ and the heavy quark mass m. In HQET, as used to describe continuum QCD, one separates these two scales and, then, treats higher dimensional operators as perturbations, to develop a systematic expansion in powers of Λ/m. For quarkonium—bound states of a heavy quark and a heavy anti-quark—there are three important scales, m, $m\nu$, and $m\nu^2$, where ν is the relative velocity between the heavy quark and heavy anti-quark. When m is large enough to probe the Coulombic part of the potential, $\nu^2 \sim \alpha_s(m)$ is small. Each operator in NRQCD must be assigned a power of ν, and the effective theory is used to develop and expansion in ν^2 and α_s, which are treated as commensurate.

Here we would like to use HQET and NRQCD to understand lattice gauge theory with heavy quarks (and moderate lattice spacings). As long as $m_Q \gg \Lambda_{\text{QCD}}$, one can write[109]

$$\mathcal{L}_{\text{lat}} \doteq \mathcal{L}_{\text{HQ}}, \tag{71}$$

which means that the lattice gauge theory inside the computer can be described by a heavy quark effective Lagrangian $\mathcal{L}_{\text{HQ}}$. The philosophy is in some ways similar, but in other ways different from, Eq. (26). The similarity is that we would like to use a continuum field theory to describe lattice gauge theory, with an eye to understanding and controlling discretization effects. The difference is that, for a heavy quark, the descriptive field theory is built from heavy-quark fields, not from QCD quark fields. The latter describes both quarks and anti-quarks.[126] The heavy quark field, on the other hand, satisfies a constraint, so it corresponds either to quarks, or anti-quarks, but not both. The arguments supporting Eq. (71) are both concrete, studying the large mass limit of lattice gauge theory,[108] and abstract, noting (as above) that the degrees of freedom and symmetries are right.[109]

HQET and NRQCD share the same effective Lagrangian,

$$\mathcal{L}_{\text{HQ}} = \sum_n \mathcal{C}_n^{\text{lat}}(m_Q, g^2, m_Q a; \mu/m_Q)\mathcal{O}_n(\mu), \tag{72}$$

where the $\mathcal{C}_n$ are short-distance coefficients and the operators $\mathcal{O}_n$ encode the long-distance behavior. The operators do not depend on the short distance scales $1/m_Q$ or a. It is useful to think of them, as with Symanzik's LE$\mathcal{L}$,

as being defined with a continuum ultraviolet regulator, and some convenient renormalization scheme. Compared to the HQET/NRQCD description of continuum QCD, the main difference is that there are two short distances, $1/m_Q$ and a. Because the change is at short distance, the short-distance coefficients $\mathcal{C}_n^{\text{lat}}$ must be modified: they depend on $m_Q a$, the ratio of short-distance scales.

Let us recall some aspects of heavy-quark theory. One has

$$\mathcal{L}_{\text{HQ}} = \mathcal{L}^{(0)} + \mathcal{L}^{(1)} + \mathcal{L}^{(2)} + \cdots. \tag{73}$$

For HQET $\mathcal{L}_{\text{HQET}}^{(s)}$ contains terms of dimension $4 + s$; for NRQCD $\mathcal{L}_{\text{NRQCD}}^{(s)}$ contains terms of order ν^{2s+2}. In the following, we shall use HQET counting, but the discussion could be repeated in NRQCD, with straightforward modifications. The leading, dimension-four term is

$$\mathcal{L}_{\text{HQET}}^{(0)} = \bar{h}_v(iv \cdot D - m_1)h_v, \tag{74}$$

where h_v is a heavy-quark field satisfying the constraint

$$i\slashed{v}h_v = h_v, \quad \bar{h}_v i\slashed{v} = \bar{h}_v. \tag{75}$$

The choice of the velocity v is somewhat arbitrary. If v is close to the heavy quark's velocity, then $\mathcal{L}^{(0)}$ is a good starting point for the heavy-quark expansion, which treats the higher-dimension operators as small. The most practical choice is the containing hadron's velocity.

The mass term in $\mathcal{L}^{(0)}$ is often omitted. By heavy-quark symmetry, it has an effect neither on bound-state wave functions nor, consequently, on matrix elements. It does affect the mass spectrum, but only additively. Including the mass obscures the heavy-quark flavor symmetry, but only slightly.[109] For two flavors, let $\theta = (m_{1c} - m_{1b})v \cdot x$; then the generators

$$\tau^1 = \frac{i}{2}\begin{pmatrix} 0 & e^{i\theta} \\ e^{-i\theta} & 0 \end{pmatrix}, \quad \tau^2 = \frac{i}{2}\begin{pmatrix} 0 & -ie^{i\theta} \\ ie^{-i\theta} & 0 \end{pmatrix}, \quad \tau^3 = \frac{i}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{76}$$

satisfying the SU(2) algebra $[\tau^d, \tau^e] = \varepsilon^{dfe}\tau^f$. When the mass term is included, higher-dimension operators are constructed with $\mathcal{D}^\mu = D^\mu - im_1 v^\mu$.[127] To describe on-shell matrix elements one may omit operators that vanish by the equation of motion, $-iv \cdot \mathcal{D}h_v = 0$, derived from Eq. (74). Higher-dimension operators are, therefore, constructed from $\mathcal{D}_\perp^\mu = D_\perp^\mu$ and $[\mathcal{D}^\mu, \mathcal{D}^\nu] = [D^\mu, D^\nu] = F^{\mu\nu}$.

The dimension-five interactions are

$$\mathcal{L}_{\text{HQET}}^{(1)} = \mathcal{C}_2^{\text{lat}}\mathcal{O}_2 + \mathcal{C}_{\mathcal{B}}^{\text{lat}}\mathcal{O}_{\mathcal{B}}, \tag{77}$$

where $\mathcal{C}_2$ and $\mathcal{C}_\mathcal{B}$ are short-distance coefficients, and

$$\mathcal{O}_2 = \bar{h}_v D_\perp^2 h_v, \tag{78}$$

$$\mathcal{O}_\mathcal{B} = \bar{h}_v s_{\alpha\beta} B^{\alpha\beta} h_v, \tag{79}$$

with $s_{\alpha\beta} = -i\sigma_{\alpha\beta}/2$ and $B^{\alpha\beta} = \eta_\mu^\alpha \eta_\nu^\beta F^{\mu\nu}$. In NRQCD, $\mathcal{O}_2$ scales as ν^2, and must be treating as a leading term: $\mathcal{L}_{\mathrm{NRQCD}}^{(0)} = \mathcal{L}_{\mathrm{HQET}}^{(0)} + \mathcal{C}_2 \mathcal{O}_2$. In NRQCD, $\mathcal{O}_\mathcal{B}$ scales as ν^4, as do several operators of dimension six and seven, and this collection of operators of order ν^4 gives the next-to-leading correction.[100]

At dimension six and higher, many operators arise. The dimension-seven Lagrangian contains the first term to parametrize the absence of full rotational symmetry;

$$\mathcal{L}_{\mathrm{HQET}}^{(3)} = \cdots + \mathcal{C}_{D^4}^{\mathrm{lat}} \bar{h}_v \sum_\mu (D_\perp^\mu)^4 h_v. \tag{80}$$

It is helpful to think of this operator as appearing in the description of continuum QCD too, but with $\mathcal{C}_{D^4}^{\mathrm{cont}} = 0$ enforced by symmetry.

One can also develop an effective field theory description of the vector and axial vector currents.[110,111] The details have been worked out for decays of a heavy quark into a light quark,[110] and for decays of a heavy quark into another heavy quark,[111] which is useful for $b \to c$ transitions.

We are now in a position to discuss uncertainties in practical calculations. The target is continuum QCD, which can be described along the lines given above, with different short-distance coefficients. The coefficients are

$$m_1^{\mathrm{cont}} = m, \tag{81}$$

$$\mathcal{C}_2^{\mathrm{cont}} = \frac{1}{2m}, \tag{82}$$

$$\mathcal{C}_\mathcal{B}^{\mathrm{cont}} = \frac{z(\mu)}{2m}, \tag{83}$$

where m is a renormalized quark mass, and z is a non-trivial function of g^2 with an anomalous dimension. At the tree level, $z = 1$. In mass independent renormalization schemes, the renormalized mass that appears in Eqs. (81)–(83) is the (perturbative) pole mass.

The description lattice gauge theory with HQET is useful comparing and contrasting the lattice-spacing uncertainties arising in the various heavy-quark methods. Since the dependence on $m_Q a$ is isolated into the coefficients, heavy-quark lattice artifacts arise only from the mismatch of the $\mathcal{C}_n^{\mathrm{lat}}$ and their analogs $\mathcal{C}_n^{\mathrm{cont}}$ in the description of continuum QCD. For brevity, we shall focus on the three most widely used methods, namely the extrapolation method, lattice NRQCD, and the Fermilab method. We shall discuss lattice NRQCD first, and then turn to the other two, which both use Wilson fermions.

For lattice NRQCD and lattice HQET, the Lagrangian is

$$\mathcal{L}_{\text{lat}} = \mathcal{L}_{\text{gauge}} - \Psi^\dagger D_{4\ \text{lat}}^+ \Psi + \sum_n c_n O_n \qquad (84)$$

where Ψ is a two-component lattice fermion field, the O_n are disretizations of the higher-dimension $\mathcal{O}_n$, and the c_n are free parameters. The c_n are chosen so that

$$\mathcal{C}_2^{\text{lat}} = \frac{1}{2m}, \qquad (85)$$

$$\mathcal{C}_{\mathcal{B}}^{\text{lat}} = \frac{z(\mu)}{2m}, \qquad (86)$$

and so on to the desired order. Equation (85) identifies what one means by renormalized quark mass; it is obtained implicitly, by adjusting a meson's kinetic mass to m_B (or m_Υ). Equation (86) is matched in perturbation theory. (The rest mass is ignored, because it does not affect matrix elements or mass splittings.) Solving for the lattice couplings c_n one finds, in many cases, power-law divergences as $a \to 0$.[99] Therefore, lattice NRQCD calculations must keep $a \sim 1/m_Q$, and discretization errors are reduced by keeping more and more terms in $\mathcal{L}_{\text{lat}}$. One must improve the light quark Lagrangian to the same order. The restriction on a is not of much practical importance, because the computing challenges discussed in Sec. 2 restrict it to the same range anyway.

The Fermilab method uses the lattice Lagrangian in Eq. (27) and adjusts the free parameters of the lattice action according to Eqs. (85) and (86). The solution of these conditions gives the lattice couplings c_O in Eq. (27) as a function of $m_Q a$. In practice, these relations are obtained in perturbation theory. The key difference to lattice NRQCD is that, as $a \to 0$, the conventional Symanzik LE$\mathcal{L}$ also applies. Consequently, the short-distance coefficients of the Fermilab method satisfy

$$\lim_{a \to 0} \mathcal{C}_n^{\text{lat}} = \mathcal{C}_n^{\text{cont}}. \qquad (87)$$

Moreovew, the corrections to the limiting behavior are related to the short-distance coefficients in the Symanzik effective field theory.[110] The pattern of uncertainties in the Fermilab method depends on $m_Q a$. For $m_Q a > 1$, discretization effects follow a pattern similar to NRQCD, whereas for $m_Q a < 1$ the Symanzik theory also is valid. On general grounds, one expects the crossover region to be smooth, and this expectation has been explicitly verified in several cases at the one-loop level.[128,113,110,111]

In the extrapolation method, one (artificially) sets $m_Q a < 1$, and assumes that Symanzik improvement is adequate. This leaves

$$\mathcal{C}_n^{\text{lat}} - \mathcal{C}_n^{\text{cont}} \sim (m_Q a)^2 \qquad (88)$$

Table 2: Parametric uncertainties for heavy quarks. For each method (and heavy-light system) the nth row gives the relative uncertainty on Q from Q_n. No estimates are given for the B system in the extrapolation method, because it is not clear how $1/m$ extrapolations amplify the uncertainties. Numerical estimates, in percent, are made taking $\alpha_s = 0.2$, $\Lambda = 500$ MeV, $m_c = 1.25$ GeV, $m_b = 4$ GeV. For illustration, we have taken $a^{-1} = 2.5$ GeV.

method	$O(a)$	extrap	Fermilab		latNRQCD
system	D	B	D	B	B
Q_0	$\alpha_s(m_c a)^2$		α_s^2	α_s^2	α_s^2
	5		4	4	4
Q_1	$m_c a \Lambda a$		$\alpha_s \Lambda a$	$\alpha_s \Lambda/m_b$	$\alpha_s \Lambda/m_b$
	10		4	2.5	2.5
Q_2	$(\Lambda a)^2$		$\alpha_s \Lambda^2 a/m_c$	$(\Lambda/m_b)^2$	$(\Lambda/m_b)^2$
	4		1.6	1.6	1.6

for non-perturbative $O(a)$ improvement. Mass splittings suffer from mismatches of order $(m_Q a)^n \Lambda/m_Q = (m_Q a)^{n-1} \Lambda a$. For matrix elements, one must also look at the normalization factor. Typical recent calculations use a normalization factor based on Eq. (52), supplemented with an Ansatz to incorporate full tree-level mass dependence from the Fermilab method.[129−132] This leaves an uncertainties of order $\alpha_s(m_Q a)^n$. As discussed above, this would be fine if $m_Q a$ were small enough. For the charmed quark, a preliminary study shows that these effects are under control for the spectrum, if one takes the continuum limit.[133]

To get a semi-quantitative feel for the uncertainties, let us consider a generic quantity Q with heavy quark expansion

$$Q = Q_0 + \frac{Q_1}{m_Q} + \frac{Q_2}{m_Q^2}. \tag{89}$$

Table 2 lists the relative uncertainty on Q from each term. The heavy quark expansion is useful here, because in all methods the physical heavy-quark effects are intertwined with lattice spacing effects. We shall assume that Eq. (89) is adequate for charmed hadrons, but the conclusions do not really depend on the assumption. For lattice NRQCD and the Fermilab method, we imagine that most of the normalization of Q_0 is non-perturbative,[134,135,48] but the rest is available at the one-loop level, and that Q_1 is normalized at the tree level. For the extrapolation method, we imagine full $O(a)$ improvement, and neglect the difficulties with c_A and c_V mentioned in Sec. 4. Table 2 also includes numerical estimates, made taking $\alpha_s = 0.2$, $\Lambda = 500$ MeV, $m_c = 1.25$ GeV, $m_b = 4$ GeV.

For B physics, lattice NRQCD and the Fermilab method lead to the same estimates. With lattice NRQCD, heavy quark propagators require negligible

computing; with the Fermilab method, they require more computing, but the overhead is negligible compared to generating gauge fields. The main advantage of the Fermilab method is that it has no restriction on $m_Q a$. No estimate is given for the discretization errors for B physics from the extrapolation method. One has to understand how errors of order $(ma)^n$ propagate through the $1/m$ expansion. One crude method is to compare different fits; this only tests whether the function is smooth in the region where data are available and is, thus, an underestimate.

For D physics, the $O(a)$ method and the Fermilab method have similar uncertainties, because $m_c a < 1$. As soon as $m_c a < \alpha_s$, non-perturbative improvement becomes valuable. It could be implemented for the Fermilab currents without any conceptual difficulty. Lattice NRQCD is rarely used for charmed hadrons, because $a^{-1} \sim m_c$ and then perturbation theory in α_s, needed to improve the NRQCD Lagrangian to interesting accuracy, does not converge well.

To reduce the uncertainty from heavy-quark discretization effects over the short term, one probably requires (automated[92]) perturbation theory. In lattice NRQCD and the Fermilab method, one more loop would reduce each uncertainty by about one fifth. Attacking B physics in this way, one could direct increases in computing at removing the quenched approximation. In the extrapolation method, the program to reduce uncertainties is to reduce a until there is a window with $m_Q a \ll 1$ and $\Lambda/m_Q \ll 1$ simultaneously. Unfortunately, this choice postpones unquenched calculations for at least another generation.

6 Light Quark Effects: Chiral Perturbation Theory

Light quarks bring physical scales $m_q \ll \Lambda$ into QCD. The numerical algorithms for computing the quark propagator slow down for light quark masses, as explained in Sec. 3. This problem makes it impractical, as a rule, to carry out numerical calculations with masses as small as those of the up and down quarks. To reach the physical region, the Monte Carlo is run at a sequence of light quark masses, say in the range $0.2m_s \lesssim m_q \lesssim m_s$, and masses and matrix elements are extrapolated to down to m_d and m_u.

Because of the artificially large light quark masses, hadrons in the computer carry a cloud of light pseudoscalar mesons, all with mass squared $m_{PS}^2 \lesssim 2m_K^2$. In nature, the cloud contains pions, kaons, and η mesons, with a wide range of masses, $m_\pi^2 \ll m_K^2 \approx \frac{3}{4}m_\eta^2$. The relation between the two situations can be understood quantitatively, by studying the interaction of the pseudoscalars with other hadrons. If the numerical data are close enough to the chiral limit, then the machinery of chiral perturbation theory[136,137] (χPT) can be applied. Numerical data can be tested against the leading (or next-to-

leading or next-to-next-to-leading) order prediction of χPT. If the data verify χPT, one has a sound guide to extrapolate in the quark mass.

Chiral perturbation theory is developed by introducing an effective Lagrangian built of pion fields and, where appropriate, other hadrons. In this case, the long distances are the inverse light quark masses, whereas the short distance is the QCD scale Λ. The separation scale μ is, therefore, usually taken to be around 1 GeV. Interactions in the chiral Lagrangian can be classified by the number of derivatives and powers of the quark mass. The momenta of pseudoscalars and the virtuality of other hadrons is also assumed to be small, $q^2 \sim m_K^2 \ll \Lambda^2$. A concrete scale, less fuzzy than Λ, that arises in the expansion is $4\pi f_\pi = 1.65$ GeV.

This chapter is not the place for a full review of χPT. Instead, we shall take a specific example—the calculation of the B meson decay constant—to illustrate how χPT for lattice gauge theory differs from phenomenological applications to low-energy hadron physics. In particular, we will contrast usual χPT with two sitations that arise in lattice calculations. The first is quenched χPT, which is used to describe problems in the quenched approximation. The second is partially quenched χPT, which applies to computer calculations in which the valence quarks have a different mass than the sea quarks in loops. Finally, we discuss how to modify the chiral Lagrangian to take into account the fact that lattice QCD usually has less chiral symmetry than continuum QCD.

The classic example of a chiral extrapolation is shown in Fig. 4. The pseudoscalar mass is related to the quark mass, in leading order, by

$$m_{PS}^2 = (m_q + m_{\bar{q}})B, \tag{90}$$

where B, from the point of view of chiral perturbation theory, is an unknown "low energy constant." From the point of view of lattice QCD, it is calculable. One simply computes m_{PS} for a variety of (light) quark and anti-quark masses and, if the behavior in Eq. (90) is verified, fits to obtain B. Figure 4 shows that Eq. (90) is a good description out to remarkably large quark masses. (This may be a fluke of the quenched approximation.) Indeed, the lattice calculation of $\hat{m} = \frac{1}{2}(m_u + m_d)$ is nothing but $\hat{m} = m_\pi^2/B$, with $m_\pi = 140$ MeV from experiment and B from Fig. 4.

The mass of the strange quark is determined in a similar way. One can set $m_q = m_{\bar{q}}$ and identify m_s as the quark mass giving $m_{PS}^2 = 2m_K^2$. Alternatively, on can extrapolate in the anti-quark mass, until $m_{\bar{q}} = \hat{m} \ll m_s$ and identify m_s when $m_{PS}^2 = (m_{K^+}^2 + m_{K^0}^2)/2$. As a rule, in lattice QCD we neglect electromagnetic effects and isospin violation—other uncertainties, even statistical errors post-extrapolation—are larger. Note that methods have been developed to include these effects, when necessary.[139]

Of course, the renormalized quark mass (and, hence, B) depends on the renormalization scheme. It is important to take this scheme depen-
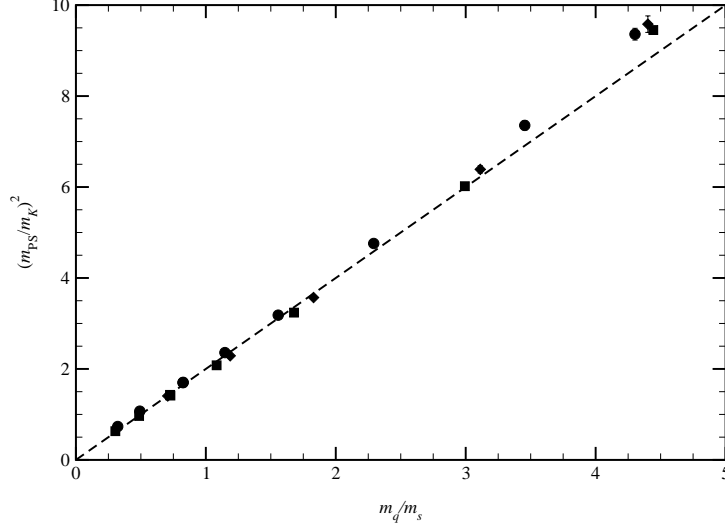
Figure 4: Plot of the squared pseudoscalar mass m_{PS}^2 vs. quark mass m_q, with numerical data in the quenched approximation.[138]

dence into account before quoting a numerical value. The two most widely used conventions are the renormalization-group invariant mass[140] and the $\overline{\text{MS}}$ mass $\bar{m}_q(2\text{ GeV})$. In mass ratios the scheme dependence cancels, and for this reason it is convenient to refer to light quark masses in units of the strange quark mass, as defined in the previous paragraph.

For other masses and matrix elements it makes more sense to eliminate the scheme-dependent quark masses in favor of pseudoscalar meson masses. The meson masses are what one computes numerically, and they are what appear in χPT. For example, let is consider f_{B_q}, where q is the flavor of the light quark in the B meson. We shall neglect $1/m_b$ corrections and write

$$f_{B_q} = \frac{\Phi}{\sqrt{m_{B_q}}}\left[1 + \Delta f_{B_q}\right], \tag{91}$$

where Φ is independent of both heavy and light quark masses, and Δf_{B_q} denotes the (one-loop) contribution of the light meson cloud.

In QCD, the one-loop correction to the decay constants are (neglecting isospin breaking)[141,142]

$$\Delta f_{B_s} = -\frac{1+3g^2}{(4\pi f)^2}\left[m_K^2 \ln(m_K^2/\mu^2) + \tfrac{1}{3}m_\eta^2 \ln(m_\eta^2/\mu^2)\right]$$
$$+ c_1(\mu)(m_K^2 + \tfrac{1}{2}m_\pi^2) + c_2(\mu)(m_K^2 - \tfrac{1}{2}m_\pi^2), \tag{92}$$

$$\Delta f_{B_d} = -\frac{1+3g^2}{(4\pi f)^2} \left[\tfrac{3}{4}m_\pi^2 \ln(m_\pi^2/\mu^2) + \tfrac{1}{2}m_K^2 \ln(m_K^2/\mu^2) + \tfrac{1}{12}m_\eta^2 \ln(m_\eta^2/\mu^2) \right]$$
$$+ \; c_1(\mu)(m_K^2 + \tfrac{1}{2}m_\pi^2) + \tfrac{1}{2}c_2(\mu)m_\pi^2, \tag{93}$$

where f and g are (the chiral limit of) the pion decay constant and B-B^*-π coupling The "low-energy" constants c_i encode QCD dynamics from distances shorter than μ^{-1}, whereas the logarithms are long-distance properties constrained by chiral symmetry. The dependence on μ cancels in the total.

These formulae illustrate how to derive useful formulae from χPT. The light quark masses in numerical lattice calculations are all about the same size, say $m_q \gtrsim 0.2m_s$, so it is not helpful to neglect m_π^2 relative to m_K^2. Because each quark mass is an adjustable parameter in lattice calculations, general formulae with non-degenerate quark masses are needed. If such formulae are available, the combination of lattice calculations and χPT is powerful. Chiral symmetry constrains the coefficient of the logarithmic terms, tying them to quantities that can be calculated by other means. This is a key, because it is difficult, from fitting, to distinguish $m^2 \ln m^2$ from a polynomial, unless the range of m^2 is very wide. With the "chiral logs" constrained, however, a straightforward fit the determines the low-energy constants.

Quenching changes Eqs. (92) and (93) drastically. Figure 5 shows the quark flow of several virtual processes that take place in a meson. The quenched approximation includes Fig. 5(b), but not (a) or (c). As a consequence, some pion loops are omitted, and η' loops are mistreated.[144] In a partially quenched calculation, Fig. 5(a) and (c) are restored, but the masses of quarks in loops are not the same as those on the valence lines. In the following, we shall examine how the omission or modification of these processes changes the quark mass dependence of f_B.

Recall that the quenched (or valence) approximation replaces the fermion determinant in Eq. (11) with 1. This replacement omits closed fermion loops (and absorbs some of the omission into the bare couplings). Instead of omitting the loops, one could imagine canceling them with bosonic loops with the same
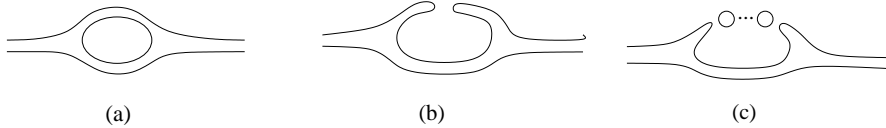

(a)   (b)   (c)

Figure 5: Quark line configurations that lead to meson loops.[143]

action. Written as functional integral[145]

$$\frac{\text{Det}(\slashed{D} + m)}{\text{Det}(\slashed{D} + \tilde{m})} = \int \mathcal{D}q \mathcal{D}\bar{q} \mathcal{D}\tilde{q} \mathcal{D}\bar{\tilde{q}} e^{-\bar{q}(\slashed{D}+m)q - \bar{\tilde{q}}(\slashed{D}+\tilde{m})\tilde{q}} \qquad (94)$$

where $\tilde{q}$ is a *bosonic* field, called scalar quarks or ghost quarks. We want to understand partially quenched calculations with n_f flavors of sea quarks and n_v flavors of valence quarks. The mass matrices are

$$m = \text{diag}(m_1, m_1, \ldots, m_{n_f}, M_1, M_2, \ldots, M_{n_v}), \qquad (95)$$
$$\tilde{m} = \text{diag}(M_1, M_2, \ldots, M_{n_v}), \qquad (96)$$

so the ghosts cancel the determinants of the valence quarks. When studying mesons, one wants $n_v = 2$; baryons, $n_v = 3$. The quenched approximation is then the special case $n_f = 0$.

The advantage of Eq. (94) is that it immediately suggests the hadronic content of the quenched approximation.[145] In addition to the usual $\bar{q}q$ mesons, one has bound states of ghosts $\bar{\tilde{q}}\tilde{q}$ and bound states of quarks and ghosts $\bar{q}\tilde{q}$ and $\bar{\tilde{q}}q$. States with an odd number of ghosts have negative metric and some spooky consequences.

If m and $\tilde{m}$ are both small compared to Λ, there is an approximate chiral symmetry. It is not so straightforward to identify the group, because of differences between bosonic and fermionic integrals, and one must insist that the determinants cancel.[146,147] It turns out that, for $n_f \geq 1$, that the symmetry group has the same number of generators as

$$G_{\text{PQ}} = \text{SU}(n_f + n_v | n_v) \times \text{SU}(n_f + n_v | n_v) \times \text{U}(1), \qquad (97)$$

and the Ward identities derived with this "intuitively obvious" group are the same as for the correct group. Here $\text{SU}(n_2 | n_1)$ denotes a graded Lie group, a mathematical beast that also appears in supersymmetry. The SU groups are spontaneously broken to their diagonal subgroup, just as in QCD.

For the quenched approximation, $n_f = 0$, the result is slightly different[147]

$$G_{\text{Q}} = [\text{SU}(n_v | n_v) \times \text{SU}(n_v | n_v)] \bowtie \text{U}(1), \qquad (98)$$

where denotes a *semi*-direct product. The semi-direct product $\bowtie$ may be less familiar that the direct product $\times$. Let A and H be subgroups of G, and suppose A is normal. (A is normal if, for all $a \in A$, $gag^{-1} \in A$ for all $g \in G$.) If, furthermore, $A \cap H = \{1\}$, then the semi-direct product $A \bowtie H$ consists of all ah, $a \in A$, $h \in H$. This seemingly technical issue has a crucial physical implication: in quenched χPT the flavor singlet mesons (η' in QCD), as well as singlet ghost-quark and ghost-antighost mesons, do not decouple.

Consequently, for quenched χPT, one must construct the effective Lagrangian for pseudoscalar mesons with a formalism that includes the η'. This can be achieved[148] along lines similar to normal χPT.[137] The main difference is that the η' is not a normal particle. Its propagator has a double-pole structure, because Fig. 5(c) is absent. On the other hand, although it was not obvious until the technical details were sorted out, the flavor singlets do decouple once there are closed quark loops to restore Fig. 5(c).[147]

Let us now quote the results from quenched and partially quenched χPT for Δf_B. In the quenched approximation[149,142]

$$
\begin{aligned}
\Delta f_{\bar{b}v}^{\mathrm{Q}} = {} & -\frac{1}{(4\pi f)^2}\left[\frac{1+3g^2}{6}m_0^2 + \gamma m_{vv}^2\right]\ln(m_{vv}^2/\mu^2) \\
& + c_0^{\mathrm{Q}}(\mu)m_0^2 + c_2^{\mathrm{Q}}(\mu)m_{vv}^2,
\end{aligned}
\tag{99}
$$

where γ is a combination of low energy constants of the quenched theory, and m_{vv} is the mass of the light pseudoscalar made from valence quark v. The constant m_0^2 is the residue of the η'-like double pole. Finally, c_0^{Q} and c_2^{Q} are counter-terms, whose dependence on μ renders $\Delta f_{\bar{b}q}^{\mathrm{Q}}$ independent of μ. In general, the low-energy constants of the quenched approximation cannot be related in any rigorous way to those of QCD.

The terms proportional to m_0^2 diverge in the chiral limit $m_{vv}^2 \to 0$. This behavior shows in detail that the quenched approximation cannot be forced to account for a wide range of scales. The common practice to obtain f_B in the quenched approximation does *not* follow from a fit to Eq. (99). Instead one assumes that there is little error from quenching when $m_v \sim m_s$, and then extrapolates linearly in m_{vv}^2. Thus, common practice circumvents the clearly unphysical quenched chiral log, but it must be admitted that the uncertainties from quenching and chiral logs become intertwined.

In the partially quenched case, the chiral logs behave better. If all n_f sea quarks have the same mass[142]

$$
\begin{aligned}
\Delta f_{\bar{b}v}^{\mathrm{PQ}} = {} & -\frac{1+3g^2}{(4\pi f)^2}\left[\frac{n_f m_{vf}^2}{2}\ln(m_{vf}^2/\mu^2) + \frac{m_{ff}^2 - 2m_{vv}^2}{2n_f}\ln(m_{vv}^2/\mu^2)\right] \\
& + c_1(\mu)m_{ff}^2 + c_2(\mu)m_{vv}^2,
\end{aligned}
\tag{100}
$$

which is qualitatively like Eqs. (92) and (93). A simple argument relates the partially quenched low-energy constants to QCD.[150] Consider any matrix element to be a function of $n_f + n_v$ sea and valence mases. If the valence quark masses are set equal to the sea quark masses, then one recovers QCD. Thus, QCD lies on a hyperplane in the space of all masses. If all masses are sent to zero together, there is a unique chiral limit. Thus, the low energy constants of the partially quenched theory and QCD are the same.

The foregoing analysis discussed chiral symmetry as it appears in the continuum limit of QCD, and its quenched and partially quenched theories. The two most widely used formulations of lattice fermions—Wilson fermions and staggered fermions—have a smaller symmetry group. Inspection of Eq. (5) shows that two sources break $\mathrm{SU}(n_f) \times \mathrm{SU}(n_f)$, namely the mass term and the part of the difference operator proportional to Dirac matrix 1. Wilson added the extra terms (or, equivalently, chose such a peculiar difference operaror) to circumvent the so-called doubling problem.

The first action that Wilson tried (and, independently, Jan Smit) was

$$\mathcal{L}_{\mathrm{Nq}} = -m_0 \bar{\psi}(x)\psi(x) \tag{101}$$
$$- \frac{1}{2a} \sum_{\mu} \bar{\psi}(x)\gamma_{\mu} \left[U_{\mu}(x)\psi(x + ae_{\mu}) - U_{\mu}^{\dagger}(x - ae_{\mu})\psi(x - ae_{\mu}) \right]$$

If $m_0 = 0$ this action is chirally symmetric. Unfortunately, the propagator has many poles

$$S(p) = \frac{a}{i \sum_{\mu} \gamma_{\mu} \sin(p_{\mu} a)} \tag{102}$$

where components of the momentum four-vector lie in the range $-\pi/a < p_{\mu} \leq \pi/a$. There is a pole at all 2^d combinations of $p_{\mu}a = 0, \pi$. These correspond to 2^d species of physical particles: vacuum polarization, for example, gets a factor 2^d. Moreover, the axial symmetries are exact: the different species have a pattern of axial charges such that the anomaly cancels. To avoid these problems, Wilson decided it was less drastic to break the axial symmetries explicitly, leaving $\mathrm{SU}(n_f)$ as the flavor group.

The staggered formulation of lattice fermions starts with Eq. (101), but notes that similarity transformation[151,152]

$$\psi(x) \mapsto T(x)\psi(x) \tag{103}$$
$$\bar{\psi}(x) \mapsto \bar{\psi}(x)T^{\dagger}(x) \tag{104}$$
$$T(x) = \gamma_1^{x_1}\gamma_2^{x_2}\gamma_3^{x_3}\gamma_4^{x_4} \tag{105}$$

diagonalizes all Dirac matrices. One then has a sum over four equivalent terms. One can three of them, leaving $2^4/4 = 4$ species in the continuum limit. The remaining species have a remnant axial U(1) symmetry.[152] On the other hand, the flavor group is smaller than $\mathrm{SU}(4)$, and the notion of flavor is tied up with spacetime symmetries—hence the name "staggered." The SO(4) Euclidean invariance is broken by the lattice down to the semi-direct product $\Gamma_4 \bowtie \mathrm{SW}_4$, where Γ_4 is the Clifford group of 4×4 Dirac matrices, and SW_4 is symmetry group of a hypercube.[153,154] The factor Γ_4 is interpreted as a flavor group, and SW_4 as a spacetime symmetry.

Table 3: Pattern of chiral symmetry breaking for various formulations of lattice fermions.

formulation	$G \to H$
Wilson	$\mathrm{SU}(n_f) \to \mathrm{SU}(n_f)$
staggered	$\mathrm{U}(1) \times \Gamma_4 \to \Gamma_4$
Ginsparg-Wilson	$\mathrm{SU}(n_f) \times \mathrm{SU}(n_f) \to \mathrm{SU}(n_f)$
continuum QCD	$\mathrm{SU}(n_f) \times \mathrm{SU}(n_f) \to \mathrm{SU}(n_f)$

Before coming back to the implications of the reduced chiral symmetry, let us review the nature of the problem. The Nielsen-Ninomiya theorem states that it is impossible to maintain chirally symmetry and avoid species doubling with ultralocal interactions.[155,156] Here *ultra*local means that the lattice couplings vanish if the fields are separated by more than a few lattice spacings. The way out, only recently appreciated, is to forgo ultralocality in favor of locality, which only requires that the couplings fall off exponentially with separation. Then, if the lattice Dirac operator $\slashed{D}$ satisfies[157]

$$\gamma_5 \slashed{D} + \slashed{D} \gamma_5 = a \slashed{D} \gamma_5 \slashed{D}, \tag{106}$$

which is called the Ginsparg-Wilson relation, then correlation functions are chirally symmetric. There are two known solutions to Eq. (106), one based on renormalization group ideas related to the original derivation,[84] and the other related to the Narayanan-Neuberger formulation of chiral lattice fermions.[34] Both are local, so they are the basis of a well-behaved field theory, but not ultralocal, so they avoid the hypothesis of the Nielsen-Ninomiya theorem. The solution of Eq. (20) for these formulations is much more computationally demanding than for Wilson or staggered fermions. Consequently, they are only beginning to make their way into numerical calculations.

Table 3 summarizes the flavor symmetry group for lattice fermions. For staggered fermions, the flavor group comes in a semi-direct product with the symmetry group of the hypercube, SW_4; for the others it comes in a direct product. The short-distance coefficients of the chiral Lagrangian must be modified to depend on $a\Lambda$, the ratio of two short distances. Because the Symanzik effective field theory shows that the violations of chiral symmetry come in through higher dimension operators, an expansion in $a\Lambda$ presents no difficulty, except to introduce new low-energy constants. The details depend on how the symmetry is broken. They have been worked out for Wilson quarks[158,159] and staggered quarks.[160]

Now that many groups are acquiring the computer resources needed for unquenched QCD, the topics discussed in this section will gain in importance. Checks of chiral behavior will be important not only for extracting physics from the numerical calculations, but also for testing the algorithms. For example,

Sec. 2 noted that odd numbers of flavors are sometimes incorporated into the Monte Carlo by generating a weight proportional to a fractional power of a determinant. For example, with staggered fermions one has $(\det M_{\mathrm{KS}})^{n_f/4}$, because M_{KS} corresponds to four flavors. It is conjectured,[161] that the appropriate graded group for the continuum limit of this theory is $\mathrm{SU}(4|4 - n_f)$, like a partially quenched theory with $n_v = 4 - n_f$. This conjecture, if true, has important consequences for testing the physical content of these algorithms, which are otherwise a bit mysterious.

7 Finite Spacetime Volume Effects

In this section we address systematic effects stemming from the infrared cutoff imposed by the finite spacetime volume. In Euclidean space, the temporal direction starts out on the same footing as the spatial directions. Even so, the physical interpretation of finite temporal extent, on the one hand, and finite spatial volume, on the other, is different. It turns out that both kinds of effects are valuable, providing tools to extend the range of problems to which Euclidean, numerical lattice QCD can be applied.

In Sec. 7.1, we shall address the implications of the finite temporal extent, and the relation to QCD thermodynamics. In Sec. 7.2, we turn to the effects of the finite spatial volume. Several issues arise here. Generic corrections to masses and $1 \to 0$ and $1 \to 1$ matrix elements are suppressed by a factor $e^{-\mu L}$, where μ is a mass related to the mass of the lightest hadron (assumed massive). The effects are larger, suppressed by powers of L instead of an exponential, when pseudo-Goldstone bosons satisfy $L m_\pi \lesssim 1$. The most intriguing effect of finite volume is on scattering states, whose allowed energies are connected in a model-independent way to final-state phase shifts.

7.1 Finite temporal extent (non-zero temperature)

Recall that numerical lattice calculations are formulated in Euclidean space, namely with imaginary time. In the computer, the extent of the imaginary time is finite, $L_4 = N_4 a$. In Sec. 2, starting with Eq. (12), we blithely assumed a correspondence between correlation functions and vacuum expectation values of time-ordered products, viz.,

$$\frac{1}{Z} \int \prod_{x,\mu} dU_\mu(x) \prod_x d\psi(x) d\bar{\psi}(x) \; O_1(t_1) O_2(t_2) \cdots O_n(t_n) \; e^{-S} \to$$

$$\langle 0 | T \; O_1(t_1) O_2(t_2) \cdots O_n(t_n) \; | 0 \rangle. \qquad (107)$$

When $L_4 = N_4 a$ is finite, however, this correspondence is not exact.

The essential features are easily seen for continuous time. Then the Euclidean action is simply related to a Lagrangian (of a system with several

degrees of freedom at each lattice site). From this Lagrangian one can straight-forwardly derive a Hamiltonian and set up the equivalence between the path-integral and canonical formulations of quantum mechanics. To be explicit, consider the case with only one operator $O(t)$ in Eq. (107). Also let us start with fixed boundary conditions, which means that fields at $x_4 = 0$ and $x_4 = L_4$ are not integrated over. Then,

$$\langle O(t) \rangle = \frac{1}{Z} \langle f| e^{-\hat{H}(L_4 - t)} \hat{O} e^{-\hat{H}t} |i\rangle, \tag{108}$$

$$Z = \langle f| e^{-\hat{H}L_4} |i\rangle, \tag{109}$$

where $|i\rangle$ and $\langle f|$ are the initial and final states implied by the boundary condi-tions, $\hat{H}$ is the Hamiltonian, and $\hat{O}$ is the Hilbert-space operator corresponding to the function (of quark and gluon fields) O on the left-hand side. Inserting complete sets of eigenstates of $\hat{H}$ into Eq. (108) one sees

$$\langle O(t) \rangle = \frac{\sum_{mn} \langle m| \hat{O} |n\rangle e^{-E_m(L_4 - t)} e^{-E_n t} \langle f|m\rangle \langle n|i\rangle}{\sum_n e^{-E_n L_4} \langle f|n\rangle \langle n|i\rangle}, \tag{110}$$

where the sums range over all eigenstates of the Hamiltonian, and E_n is the energy of state $|n\rangle$. For large L_4 the ground state—or vacuum—dominates the sum in the denominator:

$$Z \to e^{-E_0 L_4} \langle f|0\rangle \langle 0|i\rangle. \tag{111}$$

Similarly, if t and $L_4 - t$ are large

$$\langle O(t) \rangle = \langle 0| \hat{O} |0\rangle + O\left(e^{-(E_1 - E_0) \min(L_4 - t, t)} \right), \tag{112}$$

with the vacuum energy and overlap factors canceling out of the ratio in Eq. (110). One can repeat this analysis for operators at several times and, in this way, derive Eq. (107).

Although fixed boundary conditions are sometimes useful,[162,163] the con-tamination from excited states is unnecessarily large. This is especially so in two- and three-point functions, where one wants several time separations to be large, cf. Eqs. (12)–(16). To reduce the contamination term in Eq. (112) it is helpful to choose boundary conditions so that $|f\rangle = e^{-i\phi_i} |i\rangle$, and then to sum over i. Then Eq. (110) becomes

$$\langle O(t) \rangle = \frac{\sum_i e^{i\phi_i} \langle i| \hat{O} |i\rangle e^{-E_i L_4}}{\sum_i e^{i\phi_i} e^{-E_i L_4}}, \tag{113}$$

and for large L_4 but *any* t

$$\langle O(t) \rangle = \langle 0| \hat{O} |0\rangle + O\left(e^{-(E_1 - E_0) L_4} \right). \tag{114}$$

One has $|f\rangle = e^{-i\phi_i}|i\rangle$ if the fields are periodic up to phases, rotations in flavor, and discrete symmetry operations. For example, to set $|f\rangle = |i\rangle$, with no non-trivial phase, one chooses periodic boundary conditions for bosonic variables and *anti*-periodic for fermionic variables:

$$A_\mu(L_4, \boldsymbol{x}) = +A_\mu(0, \boldsymbol{x}), \quad \psi(L_4, \boldsymbol{x}) = -\psi(0, \boldsymbol{x}). \tag{115}$$

The minus sign for fermions follows from Fermi-Dirac statistics. The sum over initial state $|i\rangle$ is achieved by integrating over the fields at $x_4 = 0$.

Let us focus for a while on the boundary conditions in Eq. (115). Then one can re-write Eq. (113) as

$$\langle O(t)\rangle = \frac{\text{Tr}\,\hat{O}e^{-\hat{H}L_4}}{\text{Tr}\,e^{-\hat{H}L_4}}. \tag{116}$$

Thus, the Euclidean functional integrals give thermal expectation values for a quantum system with temperature

$$T = (k_B L_4)^{-1}, \tag{117}$$

where k_B is Boltzmann's constant. Calculations with non-zero temperatures $T \sim \Lambda/k_B$ are of considerable interest in nuclear physics and astrophysics. When the temperature (and also the chemical potential for baryons) is in this range, one expects a phase transition. The hot, dense phase is called the quark-gluon plasma, and it is thought to exist in heavy-ion collisions[167] and in quark stars,[168] which are similar to neutron stars, but with large enough gravitational attraction to crush the neutrons into the quark-gluon plasma.

Lattice calculations have been used to compute the critical temperature of the phase transition. In the pure gauge theory, this was one of the first large-scale calculations without guidance from experiment.[169,170] The critical temperature for QCD with 2 and 4 flavors has also been computed.[171−173] Over the last several years, many issues in QCD thermodynamics have been studied, such as the equation of state of the quark-gluon plasma.[174] For further reading on this rich application of lattice QCD, there are several recent reviews.[175]

For applications of lattice QCD motivated by particle physics, we are interested in zero temperature. In practice, L_4 is large but not infinite, so T is small but non-zero. From Eq. (114), one sees that the corrections are small: the energy gap between a state and the vacuum is nothing but the energy of a particle, as conventionally defined. In quenched calculations, $E_1 - E_0$ is the glueball mass.[g] With dynamical light quarks, the energy gap will be the "pion"

---

[g]One can consider quenched QCD to contain $\bar{q}q$ mesons (and qqq baryons) in the Hilbert space. But then one must also include $\bar{q}\tilde{q}$ mesons, etc. The contributions of genuine hadrons and the ghosts cancel in Eq. (113).

mass, that is, the mass of the lightest pseudoscalar m_{PS}. The exponential suppression means that to reduce the thermal correction below 0.1% one needs $m_{PS}L_4 > 7$ or $L_4 \gtrsim 4$ fm for $m_{PS} \sim 350$ MeV.

The discussion of Eqs. (108)–(116) assumed continuous (Euclidean) time. For discrete time, the analysis stays the same, but the relation between the Lagrangian and the Hamiltonian is a little different. The essential idea is as follows. There is, perhaps not surprisingly, a natural time-evolution operator, called the transfer operator (or, more often, the transfer matrix).[1,10,164–166] It propagates states in Hilbert space forward one unit of Euclidean time:

$$|\Omega(x_4 + a)\rangle = \hat{\mathbb{T}}|\Omega(x_4)\rangle, \tag{118}$$

where $|\Omega\rangle$ is any state in the Hilbert space. For details on the relationship between the lattice action and the transfer matrix, and how to define the Hilbert space, the reader should consult one of the textbooks.[12–14] The Hamiltonian $\hat{H}$ is then defined by solving

$$\hat{\mathbb{T}} = \exp(-a\hat{H}). \tag{119}$$

Indeed, all the masses considered in Sec. 2 are eigenvalues of this Hamiltonian. The maximal eigenvalue $\mathbb{T}_0$ is simply e^{-aE_0}, where E_0 is interpreted as the vacuum energy. Equation (119) provides an adequate and useful definition of $\hat{H}$, if $\hat{\mathbb{T}}$ is positive and has several eigenvalues close to the vacuum. Then $-a^{-1}\ln(\mathbb{T}_n/\mathbb{T}_0)$ are just the particle masses and energies. The transfer matrix is indeed positive for simple actions, like the Wilson action. For improved actions there can be non-positive eigenvalues; unless the action is pathological (and, thus, not an improvement) such eigenstates have energies near the cutoff and are, therefore, not physical.

7.2   *Finite spatial volume*

This subsection discusses corrections from the finite spatial volume. There are two classes of effects: polarization effects and scattering effects. The former arise because interacting particles are surrounded by a cloud of virtual particles. In a periodic volume, these virtual particles can propagate "around the world." Except when the cloud contains relatively light particles with $Lm \lesssim 1$, these are exponentially suppressed. The scattering effects are much more interesting. The boundary influences the energies of two-particle states in a way that is connected to the phase shifts. As a consequence, calculations of the L dependence of these energies yields valuable information, such as scattering lengths, resonance widths, and so on. We shall aim for a somewhat simplified explanation. A more mathematical review is in van Baal's chapter.[176]

We consider the size L of the box to be large enough that the finite volume does not alter the structure of the hadrons, namely $L\Lambda \gg 1$. Then finite-volume effects are at long distances (by definition), so the appropriate degrees

of freedom, in QCD, are hadrons. As long as the "pion" is not too light, namely $Lm_{PS} \gg 1$, the correct effective quantum field theory includes a massive field for each hadron. Unlike the effective theories of Secs. 4–6, there is no split here into a leading term and small corrections. Instead, the hadronic field theory provides a general parametrization that is consistent with analyticity, unitarity, and symmetries. The utility of such a theory for describing finite-volume effects was first pointed out by Lüscher,[177,178] and most of the results sketched below are due to him.[177-180]

For the polarization effects, Lüscher used all orders skeleton expansion for self-energies and vertex functions. The main features can be appreciated at the one-loop level. For simplicity, let us consider only two particles, a "nucleon" and a "pion", with effective Lagrangian

$$\mathcal{L} = -\bar{N}\left(\not{\partial} + m_{0N}\right)N - \tfrac{1}{2}\left(\partial_\mu \pi\right)^2 - \tfrac{1}{2}m_\pi^2 \pi^2 + y\pi \bar{N}\gamma_5 N. \tag{120}$$

The Yukawa interaction leads to the nucleon self-energy

$$\Sigma_L(p) = y^2 \int \frac{dk_4}{2\pi} \frac{1}{L^3} \sum_{\boldsymbol{\nu}} \frac{i(\not{p} + \not{k}) + m_{0N}}{(k^2 + m_\pi^2)[(p+k)^2 + m_{0N}^2]}, \tag{121}$$

where the spatial momentum is discrete in a finite box. For periodic boundary conditions, $\pi(x + Le_i) = \pi(x)$, components of the spatial momentum satisfy

$$e^{ik_i L} = 1 \Rightarrow k_i = \frac{2\pi}{L}\nu_i, \tag{122}$$

where ν_i is an integer. Our aim is to exhibit the difference between the mode sum in Eq. (121) and an integral over all $\boldsymbol{k}$.

The first step is to use an exponential representation of the propagators. Then the self energy becomes

$$\Sigma_L(p) = y^2 \int_0^\infty d\rho \int_0^1 dx \int \frac{dk_4}{2\pi} \frac{1}{L^3} \sum_{\boldsymbol{\nu}} [i(\not{p} + \not{k}) + m_{0N}]$$
$$\times \rho\, e^{-\rho[k^2 + 2xp\cdot k]} e^{-\rho[(1-x)m_\pi^2 + x(p^2 + m_{0N}^2)]}. \tag{123}$$

Now the three sums over spatial momenta are disentangled to the form

$$\mathcal{I}(r,s) = \frac{1}{L} \sum_{\nu_i = -\infty}^{\nu_i = \infty} e^{-r(\nu_i^2 + 2s\nu_i)}, \tag{124}$$

where, in our case, $r = \rho(2\pi/L)^2$ and $s = s_i \equiv xp_i L/2\pi$ for each i. The terms with $k_i \gamma_i$ in Eq. (123) require

$$\frac{1}{L} \sum_{\nu = -\infty}^{\nu = \infty} \nu\, e^{-r(\nu^2 + 2s\nu)} = -\frac{1}{2r} \left.\frac{\partial \mathcal{I}}{\partial s}\right|_{s = s_i}. \tag{125}$$

The sum $\mathcal{I}$ can be rearranged with the help of the Poisson resummation formula

$$\mathcal{I}(\rho(2\pi/L)^2, s_i) = \frac{1}{\sqrt{4\pi}} \frac{e^{\rho x^2 p_i^2}}{\sqrt{\rho}} \sum_{n=-\infty}^{+\infty} \cos(n x p_i L) \exp\left(\frac{-n^2 L^2}{4\rho}\right). \qquad (126)$$

When L is large, the term with $n = 0$ dominates; it is exactly the same as the the integral over all k_i. The leading finite-volume correction comes from the terms with $n = \pm 1$. Thus,

$$\mathcal{I}(\rho(2\pi/L)^2, s_i) = \frac{1}{\sqrt{4\pi}} \frac{e^{\rho x^2 p_i^2}}{\sqrt{\rho}} \left(1 + 2\cos(x p_i L) e^{-L^2/4\rho}\right) \qquad (127)$$

$$= \int \frac{dk}{2\pi} e^{-\rho(k^2 + 2x p_i k)} \left(1 + 2\cos(x p_i L) e^{-L^2/4\rho}\right) \quad (128)$$

up to terms suppressed by $e^{-L^2/\rho}$.

To derive the finite-volume correction to the nucleon mass, it is enough to set p on shell: $p_4 = i m_{0N} \gamma_4$, $\boldsymbol{p} = \boldsymbol{0}$. Then,

$$\Sigma_L = \frac{y^2 m_{0N}}{(4\pi)^{d/2}} \int_0^\infty d\rho \int_0^1 dx\, x\, \rho^{1-d/2}\, e^{-\rho[(1-x)m_\pi^2 + x^2 m_{0N}^2]}$$
$$\times \left(1 + 2 e^{-L^2/4\rho}\right)^{d-1}, \qquad (129)$$

in d dimensional spacetime. Eq. (129) is remarkable for several reasons. It naturally lends itself to dimensional regularization of ultraviolet divergences, which, when $d \geq 4$, arise from the lower limit of the integration over ρ. But in the finite-volume corrections the exponential factors $e^{-n^2 L^2/4\rho}$ suppress this region, faster than any power of ρ. Therefore, renormalization of ultraviolet divergences is carried out the same way as in infinite volume. In Eq. (129), the divergence is removed by mass renormalization. The asymptotic finite-volume correction can be evaluated using the method of steepest descent to carry out the integration over x and ρ. Assuming $m_\pi^2 < 2m_N^2$,

$$m_N(L) - m_N(\infty) = \Sigma_\infty - \Sigma_L = -\frac{3y^2}{8\pi} \frac{m_\pi^2}{m_N^2} \frac{e^{-L\mu}}{L} \qquad (130)$$

where $\mu^2 = m_\pi^2(1 - m_\pi^2/4m_N^2)$. The finite volume mass shift is less then 0.1% if $L\mu > 7$.

The L dependence holds for any stable particle and to all orders in the interaction.[177] (Resonances, such as the ρ meson and the Δ baryon, are another matter; see below.) If both particles in the self energy have the same mass m, then $\mu = \sqrt{3}m/2$. In the case of the physical pion and nucleon, $m_\pi \ll 2m_N$,

so $\mu \approx m_\pi$. Even if the lightest pseudoscalar has $m_{PS} = 400$ MeV, which corresponds to a light quark mass $m_s/3$, then $\mu \approx m_{PS}$.

For matrix elements with at most one hadron in the final state, one can carry out a similar analysis. The basic structure remains the same: use an exponential representation of the propagators and then evaluate the mode sums with the Poisson formula. In this way, it follows that the finite volume corrections are exponentially suppressed (as long as $Lm \gg 1$ for all particles).

Equation (130) holds only if Lm_π is large. When $Lm_\pi \lesssim 1$, the integration over x in Eq. (129) is qualitatively different. For $m_\pi = 0$, the x integration is elementary, leading to

$$m_N(L) - m_N(\infty) = -\frac{3y^2}{4\pi^2}\frac{1}{L^2 m_N} \quad \text{(for } m_\pi = 0\text{)}. \tag{131}$$

Thus, as one would expect, finite-volume corrections are larger when there is a cloud of massless particles. Equation (131) is a simple example of finite-volume chiral perturbation theory, which can be used rather generally for both finite volume and non-zero temperature.[181−183] Section 6 noted that the logarithms of chiral loops are difficult to demonstrate in numerical data. That is unfortunate, because chiral behavior is a good *a posteriori* check of algorithms for full QCD; indeed, the check may be easier to understand than the algorithms themselves. It may, therefore, prove worthwhile to hold L artificially small, and test for finite-volume Goldstone-boson effects.

Before turning to finite-volume effects in scattering states, let us consider how to extract hadron masses and matrix elements with physical light quark masses and infinite volume, from numerical data with practical light quark masses and finite volume. In practice, the lightest pseudoscalar is indeed massive, because of the chiral slowing down of the numerical algorithms, Eq. (21). For the sake of illustration, let us assume the lightest pseudoscalar has a mass $m_{PS} > 400$ MeV. A tractable situation has $L = 2$ fm, $L_4 = 2L$. The exponential factor $e^{-m_{PS}L}$ then falls below 2% (and $e^{-m_{PS}L_4}$ below 0.05%). Up and down quarks are reached by fitting data with $m_s/3 < m_q < m_s$ to the guide given by infinite-volume chiral perturbation theory. Although the finite-volume corrections from Eq. (130) are present in the data, they induce only a small bias in the fit. The fit parameters inherit the bias, but when reconstituting the desired hadronic property from the fit parameters and physical quark masses, the finite-volume effect remains under control.

Finite-volume effects are most interesting for scattering states, including resonances, such as the ρ meson and the Δ nucleon. Using models Wiese[184] and DeGrand[185] showed that it would be difficult to determine resonance masses in a finite volume. The main difficulty is that a resonance has the same quantum numbers as multi-particle states. Intuitively, the latter depend strongly on L, because the natural unit of relative momenta is $2\pi/L$, but a resonance's

mass should be nearly independent of L. Hence, there are many would-be level crossings. Of course, levels cannot cross as a parameter (here L) varies smoothly; instead the states interact and the levels push each other apart. It is plausible that this phenomenon is sensitive to details of the dynamics, but details are not immediate.

A genuine breakthrough came in Lüscher's 1991 paper,[179] which derived a non-perturbative relation between two-particle energies and phase shifts. In a finite volume, the energy spectrum is discrete. The main insight is that the phase shift is generated at relatively short range, whereas the boundary conditions that render a discrete spectrum are imposed at very long distances. Therefore, the two aspects of the problem can be attacked separately. Furthermore, it is clear from this consideration that it is the infinite volume phase shift that should arise.

The relation between the spectrum and phase shifts has been proven for elastic scattering. In the inelastic region, conceptual problems with the definition of finite-volume n-particle states ($n > 2$) have not been worked out. For pion-pion scattering, the elastic region is $2m_\pi < E < 4m_\pi$; for pion-nucleon scattering, $m_N + m_\pi < E < m_N + 3m_\pi$. Particularly in numerical lattice QCD, where the pion is artificially heavy, these limits apply to the ρ and Δ. Indeed, in many examples we suppose that $m_{PS} \gtrsim 300$ GeV; for ρ and Δ resonance properties, one should reduce the pseudoscalar mass further.

In infinite volume, the scattering phase shift arises in connecting the short- and long-range parts of the two-body wave function. In QCD long range means distances greater than $r_0 \sim 1$–2 fm. Thus, the phase shift encodes the part of the confining interactions of quarks and gluons that is "remembered" after hadrons have scattered. At very large separation, the only boundary condition on the two-body wave function is that its scattered component fall off like a spherical wave. All energies are possible and chosen by the initial conditions. Thus the scattering amplitude (which depends on the phase shifts) is the most interesting object, and experiments measure cross sections.

In a periodic finite volume, several features are radically different. Recall that one should take the time extent L_4 is at least as big as L, to set the temperature to zero. In that case, the two "scattering" particles either never separate very far (if their relative momentum is small) or encounter each other more than once (if large). So, a straightforward scattering problem does not arise; scattering amplitudes and cross sections are not natural observables. On the other hand, the two-body wave function has to be periodic,

$$\psi_k(\boldsymbol{r} + \hat{\boldsymbol{e}}_i L) = \psi_k(\boldsymbol{r}), \tag{132}$$

where $\boldsymbol{r}$ is the relative separation of the two particles, and the label k corre-

sponds to the relative momentum. The energy of the two-body state is

$$E = 2\sqrt{m^2 + k^2}, \tag{133}$$

when both particles are assumed to have the same mass m. For fixed L, Eq. (132) allows only certain values of k. Because the boundary condition is a long-distance effect, the only aspect of the interaction that can enter is the set of phase shifts. The allowed values can be compactly (but impenetrably) summarized as the solutions to[179]

$$\det\left[e^{2i\delta(k)} - U(kL/2\pi) \right] = 0, \tag{134}$$

where the matrix indices are azimuthal and magnetic angular momentum quantum numbers l and m, $U_{lm,l'm'}$. The phase shift matrix is

$$\left[e^{2i\delta(k)} \right]_{lm,l'm'} = e^{2i\delta_l(k)}\delta_{ll'}, \tag{135}$$

where $\delta_l(k)$ is the phase shift in the l-th partial wave, and (in an unfortunate clash of notation) $\delta_{ll'}$ is the Kronecker delta.

The most important feature of Eq. (134) is that $U(q)$ is built from completely kinematical, L independent functions.[179] Moreover, U splits into blocks along the diagonal, with a block for each irreducible representation of the cubic group. Finally, if, in each block, the small phase shifts for higher partial wave, then Eq. (134) reduces to a simple equation.

To illustrate this structure, let us consider the case of the ρ meson, considering the practical situation $m_\rho < 4m_\pi$. First let us neglect all phase shifts with $l > 4$. Then Eq. (134) becomes

$$\left\| \begin{matrix} e^{2i\delta_1(k)} - u_{11} & -u_{13} \\ -u_{31} & e^{2i\delta_3(k)} - u_{33} \end{matrix} \right\| = 0. \tag{136}$$

This is tractable, because the information needed to get u_{ij} has been tabulated.[179] It is even simpler if one can neglect δ_3.[180] Then

$$\delta_1(k) = n\pi - \phi(kL/2\pi), \tag{137}$$

where

$$\tan\phi(q) = -\frac{\pi^{3/2}q^2}{\mathcal{Z}_{00}(1;q^2)} \tag{138}$$

$$\mathcal{Z}_{00}(s;q^2) = \frac{1}{\sqrt{4\pi}} \sum_{\boldsymbol{\nu}} \frac{1}{(\boldsymbol{\nu}^2 - q^2)^s}, \tag{139}$$

with the sum over three-vectors of integers. The zeta function $\mathscr{Z}_{00}$ (and several cousins $\mathscr{Z}_{lm}$) are defined by the sum for $\mathrm{Re}\, 2s > l + 3$ and by analytic continuation for $s = 1$. $\phi(q)$ is a monotonically increasing function of q; as one pieces together the phase shift, the integer n is chosen so that $\delta_1(k)$ is smooth.

To see how this works, let us outline how to obtain the ρ mass and width.[180] The first step is to compute, via Monte Carlo, the $I = 1$ two-pion energies E, as a function of L. Via Eq. (133) each energy yields a momentum $k = \frac{1}{2}\sqrt{E^2 - 4m_\pi^2}$, and then the phase shift $\delta_1(k)$ is obtained from Eq. (137). The ρ resonance appears at momentum k_ρ, such that $\delta_1(k_\rho) = \pi/2$. Then, by studying the L (and hence k dependence) around the resonance, the effective range formula

$$\frac{k^3}{E} \cot \delta_1(k) = \frac{4k_\rho^3(k_\rho^2 - k^2)}{m_\rho^2 \Gamma_\rho} \tag{140}$$

can be used to determine the width. The resulting (m_ρ, Γ_ρ) would correspond to an unphysical case, with too large quark masses. From chiral perturbation theory, we expect m_ρ to depend mildly on quark masses. Similarly, if the decay is modeled with the interaction

$$\mathcal{L}_{\rho\pi\pi} = g_{\rho\pi\pi}\varepsilon_{abc}\rho_\mu^a \pi^b \partial^\mu \pi^c, \tag{141}$$

then χPT suggests that $g_{\rho\pi\pi}$ depends only mildly on the mass. The width derived from $\mathcal{L}_{\rho\pi\pi}$ is

$$\Gamma_\rho = \frac{g_{\rho\pi\pi}^2 k_\rho^3}{6\pi m_\rho^2}, \tag{142}$$

with k_ρ^3 accounting for the broadening of the resonance as phase space opens up. Thus, one could determine the physical ρ meson width by extrapolating $g_{\rho\pi\pi}$ to physical quark mass, and then using Eq. (142).

The relationship between the energy levels of scattering states and phase shifts extends to weak non-leptonic decays, again in the elastic region.[186] Thus, it is not a conceptual problem to calculate the phase shifts in K decays. Unfortunately, for for inelastic decays, such as B meson decays, it remains an unsolved problem.

Often an even bleaker picture is painted, based on a superficial understanding a theorem of Maiani and Testa.[187] The theorem assumes an infinite volume and is, thus, relevant only to extremely large volumes, where there the level spacing is so small that it is not possible, because of statistical errors, to resolve them. Then one simply observes that Euclidean correlation functions are real, so phases cannot appear without analytic continuation to Minkowski space. Although possible in principle,[9] this is again not practical when the correlation functions are computed only at discrete values of time separation.

As with many theorems, the way out is to relax one of the hypotheses: in this case, the assumption of infinite volume.

Although the study of finite volume effects, in the elastic region, is on a sound conceptual footing, there are several technical challenges. First, as mentioned above, the quark masses must be somewhat smaller than usually used. Second, many independent Monte Carlo runs, with varying L but other parameters fixed, are needed. Third, the quenched approximation is not suited to these calculations at all, because the would-be η' produces unphysical terms in the $1/L$ expansion.[188] Finally, small statistical errors are needed, to trace out the level-crossing regions, especially when the resonance is narrow. Further practical aspects have been considered by Lin et al.[189]

8 Top Ten Trends

When the Editor invited me to contribute a chapter for the Handbook of QCD, one possibility that was raised was a list of the top ten results from lattice QCD. One version of this idea, which was quickly discarded, would have been a list of specific numerical results. Such lists are quickly out of date, and are better obtained from recent from lattice conferences.

Nevertheless, it is tempting to list some of the main trends. In contrast to the rest of the chapter, which focused on practical aspects, the following topics should enjoy broad interest and influence one's thinking about QCD.

1. *Hadron spectrum: Ab initio* calculations of the hadron spectrum were one of the early motivations for numerical lattice QCD. The results of a recent calculation[190] are shown in Fig. 6, comparing quenched and $n_f = 2$ calculations. The quenched approximation does surprisingly well, agreeing with experiment within several percent. For hadrons with strange quark, particularly the ϕ and the Ω^-, the unquenched calculations agree even better. For hadrons with only up and down quarks, for example the nucleon, it remains difficult to control finite-size effects.

2. *Glueballs must exist:* When QCD was first put forward, it seemed plausible, but not obvious, that gluons would bind together to form hadrons that lie beyond the classification scheme of the quark model. Such states are called glueballs and hybrids. Although there is still controversy over whether such glueballs have been observed in the laboratory, lattice calculations give strong evidence that their existence is a prediction of QCD. There is even a well-argued case that the $f_0(1710)$ is the lowest-lying scalar glueball, based on calculations of the masses and coupling to pseudoscalar meson pairs.[191]

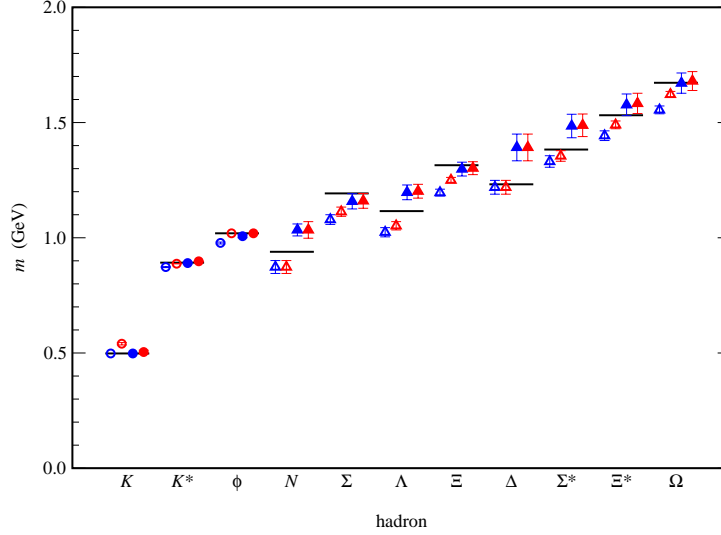3. *Confinement explained:* An early result of the strong coupling expansion

Figure 6: The hadron spectrum.[190] Solid (open) point show the results of unquenched (quenched) lattice QCD. Blue (red) points tune the strange quark's mass to the K (ϕ).

of lattice gauge theory was the so-called "area law" for Wilson loops. A Wilson loop is the line integral of the gauge potential around a closed curve $\mathcal{C}$, or $W(\mathcal{C}) = \mathrm{P}\exp\left(\int_{\mathcal{C}} A \cdot dx\right)$, where P denotes path ordering. For a $R \times T$ rectangle, a short calculation in the strong coupling limit shows[1]

$$W(R,T) \sim \left(\frac{1}{g_0^2}\right)^{RT}, \tag{143}$$

namely, that $\ln W(\mathcal{C})$ is proportional to the area enclosed by $\mathcal{C}$. For large T, $\ln W(R,T) \sim -V(R)T$, where $V(R)$ is the potential energy of the gluon field between two static sources of color. Thus, the area law shows a linearly rising potential, which confines quarks. It is amazing that confinement emerges so simply.

Numerical calculations confirm that the linear rise persists for non-Abelian gauge theories into the weak (bare) coupling regime, *i.e.*, towards the continuum limit. (It goes away for the Abelian U(1) lattice gauge theory.) A recent survey is in the review of Bali.[192] The lattice calculations trace out unmistakably a potential that is Coulombic at short distances (as expected from asymptotic freedom) and linear at large distances (as needed for confinement).

Insight into confinement is not limited to the heavy-quark potential.

From a chromoelectric perspective, studies of the gluonic flux between static sources are starting to paint an even more detailed picture.[193] From a complementary perspective, several ideas for the condensation of chromomagnetic monopoles have been studied.[194]

4. *NRQCD and HQET:* It is worth recalling that both non-relativistic QCD (NRQCD) and heavy-quark effective theory (HQET) were invented, in part, to facilitate lattice calculations of heavy-quark systems.[99,101] With continuum ultraviolet regulators, instead of a lattice, these effective field theories have become powerful tools for studying, for example, quarkonium production in high-energy collisions, or semi-leptonic decays of B mesons. They are important for understanding the CKM matrix, and QCD itself.

5. *The light quarks are very light:* There is now considerable evidence from lattice QCD that the strange quark has a mass lower than previously thought. The up and down quarks are correspondingly lighter, as expected from chiral symmetry: $(m_u + m_d)/2m_s \approx m_\pi^2/2m_K^2 \approx 1/25$. A recent calculation with dynamical quark loops finds the $\overline{\text{MS}}$ mass[190]

$$\bar{m}_s(2 \text{ GeV}) = 88^{+4}_{-6} \text{ MeV}. \qquad (144)$$

This was not an easy conclusion to reach, because lattice spacing effects and quenching effects are both large. If either is not taken into account, a larger estimate of $\bar{m}_s$ is obtained. One implication of this result is that very little of the mass of everyday objects comes from the Higgs mechanism of the electroweak theory. Almost all of it comes from gluons.

6. *QCD has a phase transition:* There are theoretical grounds to expect that pure gauge theory has a deconfinement transition, and that in QCD the high-temperature phase restores chiral symmetry. (See Part 8 of Vol. 3 of the Handbook.) But without numerical lattice QCD it is impossible to compute the critical temperature of the phase transition(s). The calculation of the pure gauge transition temperature was one of the first large-scale numerical simulations to yield, persuasively, a physically interesting result that was not known from experiment.

7. *α_s from hadrons agrees with α_s from high-energy scattering:* One of the entries for α_s in the Review of Particle Properties[195] comes from lattice calculations of charmonium and bottomonium spectra. It lies within the range of the other determinations, which come from high-energy scattering. The agreement shows that the QCD of confinement is also the QCD of partons: QCD works at long and short distances.

8. *Determination of the CKM matrix:* Lattice calculations of matrix elements in B, D, and K decays are now seen as essential to testing the standard CKM picture of flavor and CP violation.[96] The most appealing aspect of the CKM mechanism is that it is very predictive, relating many CP conserving and CP violating processes back to four parameters, only one of which violates CP. With measurements of CP conserving processes, such as semi-leptonic decays and neutral meson mixing, and calculations of the corresponding form factors and mixing amplitudes, one can predict the strength of CP violation. Once unquenched QCD calculations of this quantities mature, this area could well become the most important contribution of lattice QCD.

9. *Lattice chiral symmetry:* The last few years have witnessed remarkable progress of understanding chiral symmetry in lattice gauge theory.[79−81] In the context of QCD, the contribution is significant: a tool for more rigorous derivations of well-known soft-pion theorems,[196] or an ingredient in calculations of kaon matrix elements relevant to CKM phenomenology.[86] The contribution of this development of lattice gauge theory to electroweak physics and to extensions of the Standard Model promises to be much greater.

10. *Perturbation theory is universal:* A somewhat underappreciated, but nevertheless important, achievement is the proof that perturbation theory is universal.[68] After renormalization, the $a \to 0$ limit of perturbative QCD is the same for all lattice actions. Since lattice field theory maintains rigorous control over the ultraviolet at all steps of the analysis, it puts more familiar (i.e., continuum) ultraviolet regulators on a sounder footing, because they are all connected to lattice gauge theory through BPHZ renormalization.

9   Summary: Computational vs. Theoretical Physics

Lattice gauge theory is a broad subject, with many applications beyond QCD, especially now that chiral fermions are better understood. The lattice provides a well-defined mathematical foundation for quantum field theory, which remains one of its strongest attractions. In particular, it permits, in least in principle, real non-perturbative calculations.

Especially in the realm of QCD, one technique for calculation is most prominent, namely computing the functional integral numerically, via Monte Carlo methods. Because we live in 3+1 dimensional spacetime, and because real hadrons have several scales (m_q, Λ, m_Q), the problem is far from easy. When discussing numerical lattice QCD, usually only the computational aspects that are emphasized, perhaps overemphasized.

Computing is necessary, but not sufficient. In addition to the physical scales of hadrons, the numerical technique has an ultaviolet cutoff (from the lattice) and an infrared cutoff (from a finite spacetime volume). A whole branch of theoretical physics has been developed to understand the cutoff effects, and how to get the most out of scarce computer resources. Similarly, for practical reasons one must cope with light quark masses that are not so light, and heavy quark masses near the ultraviolet cutoff. Here tools are available from continuum QCD, but they require some redesign to be usefully deployed.

Even if the numerical solution of QCD were accessible solely by brute force, it would be wasteful to rely on brute force alone. An example, discussed at the end of Sec. 4, is the control of cutoff effects. It is much more efficient to run at several lattice spacings than to put all computer resources onto the finest conceivable lattice. A similar comment applies to the light quark mass, but here, at least, it is completely routine to do so.

In the past decade computing has been essential to test several new ideas, such as faster algorithms, methods for heavy quarks, and non-perturbative renormalization. Most of the progress in lattice QCD has come, however, from the ideas and not from Moore's Law, which observes that computer power per dollar doubles every 18 months. Because we live in four dimensions, and because the algorithms slow down, one can double the lattice spacing (at fixed computer cost) only every 6–8 years. Increases in computing are essential, but not sufficient.

It is likely that this trend will continue. For example, non-perturbative renormalization is helpful for light quarks, but remains immature for heavy quarks. For kaon physics relatively new, chirally symmetric, methods for the quarks are promising. But they are even more computationally intensive. One of the methods, domain wall fermions, adds a fifth dimension to the lattice and to the scaling laws in Sec. 3. The other methods are similarly difficult. Thus, especially here, the need for ideas to reduce the computational burden (for fixed error bar) is great.

Even if no new ideas emerge plenty of traditional theoretical physics is needed to understand numerical lattice gauge theory and, therefore, QCD. Because effective field theories provide tools to reduce the uncertainties, there is always more work to do: calculating the next order in chiral perturbation theory, or programming the next set of operators to reduce discretization effects. This work is necessary, especially when the results of lattice QCD are needed to understand flavor or collider physics. One should assume, however, the there will be new ideas, and that the theoretical and computational sides will continue to inform each other.

Acknowledgments

1. K. G. Wilson, *Phys. Rev.* **D10**, 2445 (1974).
2. J. D. Bjorken and E. A. Paschos, *Phys. Rev.* **185**, 1975 (1969).
3. H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Phys. Lett.* **B47**, 365 (1973).
4. D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973).
5. H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973).
6. J. J. Aubert *et al.*, *Phys. Rev. Lett.* **33**, 1404 (1974).
7. J. E. Augustin *et al.*, *Phys. Rev. Lett.* **33**, 1406 (1974).
8. F. J. Wegner, *J. Math. Phys.* **12**, 2259 (1971).
9. J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View*, 1st. edition (Springer, New York, 1981); 2nd. edition (Springer, New York, 1987).
10. K. G. Wilson, in *New Phenomena in Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1977).
11. K. G. Wilson and J. B. Kogut, *Phys. Rept.* **12**, 75 (1974).
12. M. Creutz, *Quarks, Gluons, and Lattices* (Cambridge University, Cambridge, UK, 1983).
13. H. J. Rothe, *Lattice Gauge Theories: An Introduction*, 1st. edition (World Scientific, Singapore, 1992); 2nd. edition (World Scientific, Singapore, 1997).
14. I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge University, Cambridge, UK, 1994).
15. J. B. Kogut, *Rev. Mod. Phys.* **55**, 775 (1983).
16. A. Hasenfratz and P. Hasenfratz, *Annu. Rev. Nucl. Part. Sci.* **35**, 559 (1985).
17. A. S. Kronfeld and P. B. Mackenzie, *Annu. Rev. Nucl. Part. Sci.* **43**, 793 (1993) [hep-ph/9303305].

18. S. Aoki, in *Proceedings of the XIX International Symposium on Lepton and Photon Interactions at High Energy*, edited by J. A. Jaros and M. E. Peskin, *eConfC* **990809**, 657 (2000) [hep-ph/9912288].

19. R. D. Kenway, in *ICHEP 2000: Proceedings*, edited by C. S. Lim and T. Yamanaka (World Scientific, Singapore, 2001) [hep-ph/0010219].

20. *From Actions to Answers*, edited by T. DeGrand and D. Toussaint (World Scientific, Singapore, 1990).

21. M. Lüscher, in *Fields, Strings, and Critical Phenomena*, edited by E. Brezin and J. Zinn-Justin (North-Holland, Amsterdam, 1990).

22. A. S. Kronfeld, in *Perspectives in the Standard Model*, edited by R. K. Ellis, C. T. Hill, and J. D. Lykken (World Scientific, Singapore, 1992).

23. S. R. Sharpe, in *CP Violation and the Limits of the Standard Model*, edited by J F. Donoghue (World Scientific, Singapore, 1995) [hep-ph/9412243].

24. R. Gupta, in *Probing the Standard Model of Particle Interactions*, edited by R. Gupta, A. Morel, E. De Rafael, F. David (Elsevier, Amsterdam, 1999) [hep-lat/9807028].

25. M. Lüscher, in *Probing the Standard Model of Particle Interactions*, edited by R. Gupta, A. Morel, E. De Rafael, F. David (Elsevier, Amsterdam, 1999) [hep-lat/9802029].

26. G. Münster and M. Walzl, in *Phenomenology of Gauge Interactions*, edited by D. Graudenz and V. Markushin (URL http://ltpth.web.psi.ch/ /zuoz/zuoz2000/zuoz2000proc.htm) [hep-lat/0012005].

27. T. DeGrand, in *Flavor Physics for the Millenium*, edited by J. L. Rosner (World Scientific, Singapore, 2001) [hep-ph/0008234].

28. W. T. Giele, S. A. Keller, and D. A. Kosower, hep-ph/0104052.

29. M. Göckeler *et al.*, *Phys. Rev.* **D53**, 2317 (1996) [hep-lat/9508004];

30. M. Guagnelli, K. Jansen, and R. Petronzio, *Phys. Lett.* **B493**, 77 (2000) [hep-lat/0009006];

31. D. Dolgov *et al.* [LHPC and SESAM Collaborations], hep-lat/0201021.

32. K. Jansen, in *ICHEP 2000: Proceedings*, edited by C. S. Lim and T. Yamanaka (World Scientific, Singapore, 2001) [hep-lat/0010038].

33. M. Di Pierro, "From Monte Carlo Integration to Lattice Quantum Chromodynamics: An Introduction," hep-lat/0009001 [FERMILAB-FN-699].

34. H. Neuberger, *Phys. Lett.* **B417**, 141 (1998) [hep-lat/9707022]; **B427**, 353 (1998) [hep-lat/9801031].

35. L. Susskind, *Phys. Rev.* **D16**, 3031 (1977); T. Banks, J. B. Kogut, and L. Susskind, *ibid.* **D13**, 1043 (1976).

36. S. Aoki *et al.* [JLQCD Collaboration], *Phys. Rev.* **D65**, 094507 (2002) [hep-lat/0112051].

37. S. Duane, *Nucl. Phys.* **B257**, 652 (1985).

38. A. Ukawa and M. Fukugita, *Phys. Rev. Lett.* **55**, 1854 (1985).

39. G. G. Batrouni *et al.*, *Phys. Rev.* **D32**, 2736 (1985).
40. J. C. Sexton, private communication; A. C. Irving and J. C. Sexton, *Phys. Rev.* **D55**, 5456 (1997) [hep-lat/9608145].
41. E. Marinari, G. Parisi, and C. Rebbi, *Nucl. Phys.* **B190**, 734 (1981).
42. D. Weingarten, *Phys. Lett.* **B109**, 57 (1982).
43. J. Sexton and D. Weingarten, *Phys. Rev.* **D55**, 4025 (1997).
44. A. X. El-Khadra, G. Hockney, A. S. Kronfeld, and P. B. Mackenzie, *Phys. Rev. Lett.* **69**, 729 (1992).
45. P. B. Mackenzie, *Nucl. Phys. B Proc. Suppl.* **34**, 400 (1994).
46. C. T. H. Davies *et al.*, *Phys. Rev. Lett.* **73**, 2654 (1994) [hep-lat/9404012].
47. B. J. Gough *et al.*, *Phys. Rev. Lett.* **79**, 1622 (1997) [hep-ph/9610223].
48. S. Hashimoto *et al.*, hep-ph/0110253; J. N. Simone *et al.*, *Nucl. Phys. B Proc. Suppl.* **83**, 334 (2000) [hep-lat/9910026].
49. A. Ali Khan *et al.* [CP-PACS Collaboration], *Phys. Rev.* **D64**, 034505 (2001) [hep-lat/0010009].
50. A. Ali Khan *et al.* [CP-PACS Collaboration], *Phys. Rev.* **D64**, 054504 (2001) [hep-lat/0103020].
51. C. Bernard *et al.* [MILC Collaboration], *Nucl. Phys. B Proc. Suppl.* **83**, 289 (2000) [hep-lat/9909121]; **94**, 346 (2001) [hep-lat/0011029].
52. V. Giménez, L. Giusti, G. Martinelli, and F. Rapuano, JHEP **0003**, 018 (2000) [hep-lat/0002007].
53. S. Coleman and R. E. Norton, *Nuovo Cim.* **38**, 438 (1965).
54. A. S. Kronfeld, *Prog. Theor. Phys. Suppl.* **111**, 293 (1993) [hep-lat/9205008].
55. S. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, *Phys. Rev.* **D35**, 2531 (1987).
56. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett.* **B195**, 216 (1987).
57. M. Lüscher, *Nucl. Phys.* **B418**, 637 (1994) [hep-lat/9311007].
58. G. P. Lepage *et al.*, *Nucl. Phys. B Proc. Suppl.* **106**, 12 (2002) [hep-lat/0110175].
59. K. Symanzik, in *Recent Developments in Gauge Theories*, edited by G. 't Hooft *et al.* (Plenum, New York, 1980).
60. K. Symanzik, in *Mathematical Problems in Theoretical Physics*, edited by R. Schrader *et al.* (Springer, New York, 1982); *Nucl. Phys.* **B226**, 187, 205 (1983).
61. B. Sheikholeslami and R. Wohlert, *Nucl. Phys.* **B259**, 572 (1985).
62. K. Symanzik, *Commun. Math. Phys.* **18**, 227 (1970); **23**, 49 (1971).
63. C. G. Callan, *Phys. Rev.* **D2**, 1541 (1970).
64. M. Lüscher and P. Weisz, *Commun. Math. Phys.* **97**, 59 (1985); (E) **98**, 433 (1985).
65. M. Lüscher, S. Sint, R. Sommer, and P. Weisz, *Nucl. Phys.* **B478**, 365

(1996) [hep-lat/9605038].

66. T. Reisz, *Commun. Math. Phys.* **116**, 81 (1988).
67. T. Reisz, *Commun. Math. Phys.* **117**, 79 (1988).
68. T. Reisz, *Nucl. Phys.* **B318**, 417 (1989).
69. R. Wohlert, DESY 87-069 (unpublished, July 1987).
70. M. Lüscher and P. Weisz, *Nucl. Phys.* **B479**, 429 (1996).
71. S. Sint and P. Weisz, *Nucl. Phys.* **B502**, 251 (1997).
72. M. Bochicchio, L. Maiani, G. Martinelli, G. Rossi, and M. Testa, *Nucl. Phys.* **B262**, 331 (1985).
73. K. Jansen *et al.*, *Phys. Lett.* **B372**, 275 (1996) [hep-lat/9512009]; M. Lüscher, S. Sint, R. Sommer, P. Weisz, and U. Wolff, *Nucl. Phys.* **B491**, 323 (1997) [hep-lat/9609035].
74. M. Lüscher, S. Sint, R. Sommer, and H. Wittig, *Nucl. Phys.* **B491**, 344 (1997) [hep-lat/9611015].
75. T. Bhattacharya, R. Gupta, W. Lee, and S. Sharpe, *Phys. Rev.* **D63**, 074505 (2001) [hep-lat/0009038].
76. S. Collins, C. T. Davies, G. P. Lepage, and J. Shigemitsu, hep-lat/0110159.
77. S. R. Sharpe, *Nucl. Phys. B Proc. Suppl.* **34**, 403 (1994) [hep-lat/9312009].
78. Y. Luo, *Phys. Rev.* **D55**, 353 (1997) [hep-lat/9604025].
79. F. Niedermayer, *Nucl. Phys. B Proc. Suppl.* **73**, 105 (1999) [hep-lat/9810026].
80. H. Neuberger, *Nucl. Phys. B Proc. Suppl.* **83**, 67 (2000) [hep-lat/9909042].
81. M. Lüscher, *Nucl. Phys. B Proc. Suppl.* **83**, 34 (2000) [hep-lat/9909150].
82. Y. Shamir, *Nucl. Phys.* **B406**, 90 (1993) [hep-lat/9303005].
83. U. J. Wiese, *Phys. Lett.* **B315**, 417 (1993) [hep-lat/9306003].
84. P. Hasenfratz, V. Laliena, and F. Niedermayer, *Phys. Lett.* **B427**, 125 (1998) [hep-lat/9801021].
85. D. B. Kaplan, *Phys. Lett.* **B288**, 342 (1992) [hep-lat/9206013].
86. T. Blum and A. Soni, *Phys. Rev.* **D56**, 174 (1997) [hep-lat/9611030].
87. A. Ali Khan *et al.* [CP-PACS Collaboration], *Phys. Rev.* **D64**, 114506 (2001) [hep-lat/0105020].
88. T. Blum *et al.* [RBC Collaboration], hep-lat/0110075.
89. G. Kilcup, R. Gupta, and S. R. Sharpe, *Phys. Rev.* **D57** (1998) 1654 [hep-lat/9707006].
90. S. Aoki *et al.* [JLQCD Collaboration], *Phys. Rev. Lett.* **80** (1998) 5271 [hep-lat/9710073].
91. M. Lüscher and P. Weisz, *Nucl. Phys.* **B266**, 309 (1986).
92. M. A. Nobes, H. D. Trottier, G. P. Lepage, and Q. Mason, *Nucl. Phys. B Proc. Suppl.* **106**, 838 (2002) [hep-lat/0110051].

93. S. J. Brodsky, G. P. Lepage, and P. B. Mackenzie, *Phys. Rev.* **D28**, 228 (1983).

94. G. P. Lepage and P. B. Mackenzie, *Phys. Rev.* **D48**, 2250 (1993) [hep-lat/9209022].

95. J. Harada, S. Hashimoto, A. S. Kronfeld, and T. Onogi, in preparation.

96. M. Beneke, "CP violation and the CKM matrix," hep-lat/0201011.

97. For reviews of the calculations themselves see recent (and future) reviews at lattice conferences, such as S. Hashimoto, *Nucl. Phys. B Proc. Suppl.* **83**, 3 (2000) [hep-lat/9909136]; C. Bernard, *ibid.* **94**, 159 (2001) [hep-lat/0011064]; S. Ryan, *ibid.* **106**, 86 (2002) [hep-lat/0111010].

98. W. E. Caswell and G. P. Lepage, *Phys. Lett.* **B167**, 437 (1986).

99. G. P. Lepage and B. A. Thacker, *Nucl. Phys. B Proc. Suppl.* **4**, 199 (1987); B. A. Thacker and G. P. Lepage, *Phys. Rev.* **D43**, 196 (1991).

100. G. P. Lepage, L. Magnea, C. Nakhleh, U. Magnea, and K. Hornbostel, *Phys. Rev.* **D46**, 4052 (1992) [hep-lat/9205007].

101. E. Eichten, *Nucl. Phys. B Proc. Suppl.* **4**, 170 (1987).

102. E. Eichten and B. Hill, *Phys. Lett.* **B234**, 511 (1990); **B240**, 193 (1990).

103. E. Eichten and B. Hill, *Phys. Lett.* **B243**, 427 (1990).

104. B. Grinstein, *Nucl. Phys.* **B339**, 253 (1990).

105. H. Georgi, *Phys. Lett.* **B240**, 447 (1990).

106. M. E. Luke, *Phys. Lett.* **B252**, 447 (1990).

107. G. T. Bodwin, E. Braaten, and G. P. Lepage, *Phys. Rev.* **D46**, 1914 (1992) [hep-lat/9205006].

108. A. X. El-Khadra, A. S. Kronfeld, and P. B. Mackenzie, *Phys. Rev.* **D55**, 3933 (1997) [hep-lat/9604004].

109. A. S. Kronfeld, *Phys. Rev.* **D62**, 014505 (2000) [hep-lat/0002008].

110. J. Harada, S. Hashimoto, K.-I. Ishikawa, A. S. Kronfeld, T. Onogi, and N. Yamada, hep-lat/0112044.

111. J. Harada, S. Hashimoto, A. S. Kronfeld, and T. Onogi, hep-lat/0112045.

112. A version of this argument is given in S. Aoki, Y. Kuramashi, and S.-I. Tominaga, hep-lat/0107009. This work misstates the origin of the $\sqrt{m_1/m_2}$ in Eq. (68), and it also counts redundant directions incorrectly.

113. B. P. G. Mertens, A. S. Kronfeld, and A. X. El-Khadra, *Phys. Rev.* **D58**, 034505 (1998) [hep-lat/9712024].

114. M. B. Gavela, L. Maiani, S. Petrarca, G. Martinelli, and O. Pene, *Phys. Lett.* **206B**, 113 (1988).

115. C. Bernard, T. Draper, G. Hockney, and A. Soni, *Phys. Rev.* **D38**, 3540 (1988).

116. R. Sommer, *Phys. Rept.* **275**, 1 (1996) [hep-lat/9401037].

117. H. Wittig, *Int. J. Mod. Phys.* **A12**, 4477 (1997) [hep-lat/9705034].

118. T. R. Klassen, *Nucl. Phys.* **B509**, 391 (1998); *Nucl. Phys. B Proc. Suppl.* **73**, 918 (1999).

119. J. Harada, A. S. Kronfeld, H. Matsufuru, N. Nakajima, and T. Onogi, *Phys. Rev.* **D64**, 074501 (2001) [hep-lat/0103026].
120. S. Collins, *et al.*, *Phys. Rev.* **D64**, 055002 (2001) [hep-lat/0101019].
121. Z. Sroczynski, *Nucl. Phys. B Proc. Suppl.* **83**, 971 (2000) [hep-lat/9910004]; *Int. J. Mod. Phys.* **A16S1C**, 1231 (2001) [hep-lat/0011059].
122. G. P. Lepage, *Nucl. Phys. B Proc. Suppl.* **26**, 45 (1992).
123. S. Hashimoto, *Phys. Rev.* **D50**, 4639 (1994) [hep-lat/9403028].
124. M. A. Shifman and M. B. Voloshin, *Sov. J. Nucl. Phys.* **45**, 292 (1987); **47**, 511 (1988) [*Yad. Fiz.* **45**, 463 (1987); **47**, 801 (1988)].
125. N. Isgur and M. B. Wise, *Phys. Lett.* **B232**, 113 (1989); **B237**, 527 (1990).
126. P. A. M. Dirac, *Proc. Roy. Soc. Lond.* **A133**, 60 (1931).
127. A. F. Falk, M. Neubert, and M. Luke, *Nucl. Phys.* **B388**, 363 (1992) [hep-ph/9204229].
128. Y. Kuramashi, *Phys. Rev.* **D58**, 034507 (1998) [hep-lat/9705036].
129. D. Becirevic *et al.*, *Phys. Rev.* **D60**, 074501 (1999) [hep-lat/9811003].
130. K. C. Bowler *et al.* [UKQCD Collaboration], *Phys. Lett.* **B486**, 111 (2000) [hep-lat/9911011].
131. K. C. Bowler *et al.* [UKQCD Collaboration], *Nucl. Phys.* **B619**, 507 (2001) [hep-lat/0007020].
132. A. Abada *et al.*, *Nucl. Phys.* **B619**, 565 (2001) [hep-lat/0011065].
133. J. Rolf and S. Sint, *Nucl. Phys. B Proc. Suppl.* **106**, 239 (2002) [hep-ph/0110139].
134. S. Hashimoto *et al.*, *Phys. Rev.* **D61**, 014502 (2000) [hep-ph/9906376].
135. A. X. El-Khadra, A. S. Kronfeld, P. B. Mackenzie, S. M. Ryan, and J. N. Simone, *Phys. Rev.* **D64**, 014502 (2001) [hep-ph/0101023].
136. J. Gasser and H. Leutwyler, *Ann. Phys.* **158**, 142 (1984).
137. J. Gasser and H. Leutwyler, *Nucl. Phys.* **B250**, 465 (1985).
138. M. Göckeler *et al.*, *Phys. Rev.* **D62**, 054504 (2000) [hep-lat/9908005].
139. A. Duncan, E. Eichten, and H. Thacker, *Phys. Rev. Lett.* **76**, 3894 (1996) [hep-lat/9602005].
140. J. Garden, J. Heitger, R. Sommer, and H. Wittig [Alpha Collaboration], *Nucl. Phys.* **B571**, 237 (2000) [hep-lat/9906013].
141. C. G. Boyd and B. Grinstein, *Nucl. Phys.* **B442**, 205 (1995) [hep-ph/9402340].
142. S. R. Sharpe and Y. Zhang, *Phys. Rev.* **D53**, 5125 (1996) [hep-lat/9510037].
143. W. Bardeen, A. Duncan, E. Eichten, N. Isgur, and H. Thacker, *Phys. Rev.* **D65**, 014509 (2002) [hep-lat/0106008].
144. S. R. Sharpe, *Phys. Rev.* **D41**, 3233 (1990); **46**, 3146 (1992) [hep-lat/9205020].
145. A. Morel, *J. Phys. (France)* **48**, 1111 (1987).

146. P. H. Damgaard, J. C. Osborn, D. Toublan, and J. J. Verbaarschot, *Nucl. Phys.* **B547**, 305 (1999) [hep-th/9811212].

147. S. R. Sharpe and N. Shoresh, *Phys. Rev.* **D64**, 114510 (2001) [hep-lat/0108003].

148. C. W. Bernard and M. F. L. Golterman, *Phys. Rev.* **D46**, 853 (1992) [hep-lat/9204007].

149. M. J. Booth, *Phys. Rev.* **D51**, 2338 (1995) [hep-ph/9411433], hep-ph/9412228.

150. S. R. Sharpe and N. Shoresh, *Phys. Rev.* **D62**, 094503 (2000) [hep-lat/0006017].

151. N. Kawamoto and J. Smit, *Nucl. Phys.* **B192**, 100 (1981).

152. H. S. Sharatchandra, H. J. Thun, and P. Weisz, *Nucl. Phys.* **B192**, 205 (1981).

153. G. W. Kilcup and S. R. Sharpe, *Nucl. Phys.* **B283**, 493 (1987).

154. H. Joos and M. Schäfer, *Z. Phys.* **C34**, 465 (1987).

155. H. B. Nielsen and M. Ninomiya, *Nucl. Phys.* **B185**, 20 (1981); (E) **195**, 541 (1982).

156. D. Friedan, *Commun. Math. Phys.* **85** (1982) 481.

157. P. H. Ginsparg and K. G. Wilson, *Phys. Rev.* **D25**, 2649 (1982).

158. S. R. Sharpe and R. J. Singleton, *Phys. Rev.* **D58**, 074501 (1998) [hep-lat/9804028].

159. G. Rupak and N. Shoresh, hep-lat/0201019.

160. W. J. Lee and S. R. Sharpe, *Phys. Rev.* **D60**, 114503 (1999) [hep-lat/9905023].

161. C. W. Bernard and M. F. L. Golterman, *Phys. Rev.* **D49**, 486 (1994) [hep-lat/9306005].

162. K.-H. Mütter and K. Schilling, *Nucl. Phys.* **B200**, 362 (1982); *Phys. Lett.* **B117**, 75 (1982).

163. M. Guagnelli, J. Heitger, R. Sommer, and H. Wittig [Alpha Collaboration], *Nucl. Phys.* **B560**, 465 (1999) [hep-lat/9903040].

164. M. Creutz, *Phys. Rev.* **D15**, 1128 (1977).

165. M. Lüscher, *Commun. Math. Phys.* **54**, 283 (1977).

166. K. Osterwalder and E. Seiler, *Ann. Phys.* **110**, 440 (1978).

167. H. Satz, *Rept. Prog. Phys.* **63**, 1511 (2000) [hep-ph/0007069].

168. E. S. Fraga, R. D. Pisarski, and J. Schaffner-Bielich, *Phys. Rev.* **D63**, 121702 (2001) [hep-ph/0101143].

169. S. A. Gottlieb *et al.*, *Phys. Rev. Lett.* **55**, 1958 (1985).

170. N. H. Christ and A. E. Terrano, *Phys. Rev. Lett.* **56**, 111 (1986).

171. R. V. Gavai *et al.* [MT_c Collaboration], *Phys. Lett.* **B232**, 491 (1989); *Phys. Lett.* **B241**, 567 (1990).

172. F. R. Brown *et al.*, *Phys. Lett.* **B251**, 181 (1990); *Phys. Rev. Lett.* **65**, 2491 (1990).

173. M. Fukugita, H. Mino, M. Okawa, and A. Ukawa, *Phys. Rev. Lett.* **65**, 816 (1990); *Phys. Rev.* **D42**, 2936 (1990).

174. G. Boyd *et al.*, *Phys. Rev. Lett.* **75**, 4169 (1995) [hep-lat/9506025]; *Nucl. Phys.* **B469**, 419 (1996) [hep-lat/9602007].

175. E. Laermann, *Phys. Part. Nucl.* **30**, 304 (1999) [*Fiz. Elem. Chast. Atom. Yadra* **30**, 720 (1999)]; S. Ejiri, *Nucl. Phys. B Proc. Suppl.* **94**, 19 (2001) [hep-lat/0011006]; S. Hands, *ibid.* **106**, 142 (2002) [hep-lat/0109034].

176. P. van Baal, in *At the Frontier of Particle Physics: Handbook of QCD*, Vol. 2, edited by M. Shifman (World Scientific, Singapore, 2001) [hep-ph/0008206].

177. M. Lüscher, *Commun. Math. Phys.* **104**, 177 (1986).

178. M. Lüscher, *Commun. Math. Phys.* **105**, 153 (1986).

179. M. Lüscher, *Nucl. Phys.* **B354**, 531 (1991).

180. M. Lüscher, *Nucl. Phys.* **B364**, 237 (1991).

181. J. Gasser and H. Leutwyler, *Nucl. Phys.* **B307**, 763 (1988).

182. P. Hasenfratz and H. Leutwyler, *Nucl. Phys.* **B343**, 241 (1990).

183. F. C. Hansen, *Nucl. Phys.* **B345**, 685 (1990); F. C. Hansen and H. Leutwyler, *ibid.* **350**, 201 (1991).

184. U. J. Wiese, *Nucl. Phys. B Proc. Suppl.* **9**, 609 (1989).

185. T. A. DeGrand, *Phys. Rev.* **D43**, 2296 (1991).

186. L. Lellouch and M. Lüscher, *Commun. Math. Phys.* **219**, 31 (2001) [hep-lat/0003023]; See also Sec. 7 of L. Lellouch, *Nucl. Phys. B Proc. Suppl.* **94**, 142 (2001) [hep-lat/0011088], for a simple summary.

187. L. Maiani and M. Testa, *Phys. Lett.* **B245**, 585 (1990).

188. C. W. Bernard and M. F. L. Golterman, *Phys. Rev.* **D53**, 476 (1996).

189. C. J. Lin, G. Martinelli, C. T. Sachrajda, and M. Testa, *Nucl. Phys.* **B619**, 467 (2001) [hep-lat/0104006].

190. A. Ali Khan *et al.* [CP-PACS Collaboration], *Phys. Rev.* **D65**, 054505 (2002) [hep-lat/0105015].

191. J. Sexton, A. Vaccarino, and D. Weingarten, *Phys. Rev. Lett.* **75**, 4563 (1995) [hep-lat/9510022].

192. G. S. Bali, in *Quark Confinement and the Hadron Spectrum IV*, edited by W. Lucha and K. M. Maung (World Scientific, Singapore, 2001) [hep-ph/0010032].

193. C. J. Morningstar, K. J. Juge, and J. Kuti, *Nucl. Phys. B Proc. Suppl.* **73**, 590 (1999) [hep-lat/9809098].

194. J. Fröhlich and P. A. Marchetti, *Nucl. Phys. B Proc. Suppl.* **106**, 47 (2002) [hep-lat/0110193].

195. D. E. Groom *et al.* [Particle Data Group], Eur. Phys. J. **C15**, 1 (2000).

196. S. Chandrasekharan, *Phys. Rev.* **D60**, 074503 (1999) [hep-lat/9805015].