Lattice QCD Simulations beyond the Quenched Approximation

Akira Ukawa†

Theory Division, CERN, Geneva, Switzerland

Review presented at the GIFT International Seminar
on Non-Perturbative Aspects of the Standard Model
Spain, 6–11 June, 1988

Abstract

Present status of lattice QCD simulations incorporating the effects of dynamical quarks is presented. After a brief review of the formalism of lattice QCD, the dynamical fermion algorithms in use today are described. Recent attempts at the hadron mass calculation are discussed in relation to the quenched results, and current understanding on the finite temperature behavior of QCD is summarized.

CERN-TH-5245/88
December 1988

† On leave from Institute of Physics, University of Tsukuba, Ibaraki 305, Japan.
Lattice QCD Simulations beyond the Quenched Approximation

Akira Ukawa\textsuperscript{1}

Theory Division, CERN, Geneva, Switzerland

Review presented at the GIFT International Seminar on Non-Perturbative Aspects of the Standard Model Spain, 6-11 June, 1988

Abstract

Present status of lattice QCD simulations incorporating the effects of dynamical quarks is presented. After a brief review of the formalism of lattice QCD, the dynamical fermion algorithms in use today are described. Recent attempts at the hadron mass calculation are discussed in relation to the quenched results, and current understanding on the finite temperature behavior of QCD is summarized.

Contents

1. Introduction.
2. Review of formalism.
   2.1 QCD on a lattice.
   2.2 Lattice fermions and chiral symmetry.
3. Dynamical fermion algorithms.
   3.1 Source of difficulty.
   3.2 Hybrid Monte Carlo and related algorithms.
   3.3 Other alternatives.
4. Hadron mass spectrum.
   4.1 Preliminary.
   4.2 Recent quenched calculations.
   4.3 Attempts at full QCD.
   4.4 The U(1) problem.
5. Thermodynamics of QCD.
   5.1 Theoretical expectations at high temperatures.
   5.2 Pure gauge simulation results.
   5.3 Phase diagram–Kogut-Susskind quark action.
   5.4 Phase diagram–Wilson quark action.
   5.5 Physics of the high temperature phase.
   5.6 Finite baryon number density.

\textsuperscript{1} On leave from Institute of Physics, University of Tsukuba, Ibaraki 305, Japan.
1. Introduction

Quantum chromodynamics depicts hadrons as made up of the two fundamental building blocks, quarks and gluons. The gluons are responsible for the confining property of the strong force that allows only color singlets as physical excitations. The quarks, on the other hand, are the carriers of the chiral symmetry. The spontaneous breakdown of this symmetry is the second basic characteristic of the strong interactions. The challenge posed to QCD is to explain these features as outcomes of its dynamics and to lead to a quantitative prediction of strong interaction observables from its first principles.

This is not an easy task. The asymptotic freedom of QCD tells us that the dynamics at large distances is strongly coupled, and that the observables sensitive to long wavelength fluctuations have the non-perturbative dependence on the coupling constant $e^{-1/12\alpha_s}$. Studying QCD therefore requires a well-defined framework valid beyond perturbation theory. This has been provided by the formulation of QCD on a space-time lattice in 1974.

A real progress in disentangling the long-distance dynamics of QCD started in 1979 with the application of numerical simulation techniques for evaluating the lattice QCD path integral. Although the earlier analytical studies much helped to our understanding of confinement, they were not powerful enough to answer the crucial question of the existence of the continuum limit nor to extract the quantitative predictions of lattice QCD. On the other hand, the very first calculation of the string tension by a Monte Carlo simulation, though very modest by today's standards, already found an indication of an approximate scaling behavior. This was the first concrete evidence that lattice QCD does have a sensible continuum limit, which clearly demonstrated the power of the numerical simulations.

The study of lattice QCD through numerical simulations has enormously expanded since then. However, the numerical approach has suffered from one serious drawback: it could not handle fermions efficiently since they were represented by anti-commuting Grassmann numbers in a path integral. Numerous proposals have been put forward to overcome the difficulty. It is only since several years, however, that a real step forward has been achieved. This has taken place for two reasons. Firstly, application of non-standard quantization procedures such as the micro-canonical and the Langevin methods has led to the development of a set of new algorithms. These algorithms require the computing power growing only linearly with the lattice volume $V$, as compared to $V^4$ or $V^2$ at best for those known previously. Secondly, the computer technologies have been developing fast and supercomputers such as CRAY-XMP have become widely available for lattice QCD simulations. Thus full QCD simulations including the effects of dynamical quarks on a moderately large lattice such as $8^3$ or $16^3$ have become feasible.

Why do we want to include dynamical quarks in the simulation in spite of the considerable cost it entails? One obvious answer is that there is no apparent reason to believe that the light up and down quarks can be treated as a perturbation to the pure gauge dynamics. There exists, however, a far more specific reason: a number of questions cannot be understood properly without dynamical quarks. Let us list some of those questions to set the stage of this review.

The static properties of hadrons such as their mass spectrum pose the first challenge to lattice QCD. There are a large number of lattice calculation of hadron masses carried out within the "quenched approximation". This is a "valence" approximation which keeps valence quarks but completely ignores the sea quark. The flavor non-singlet hadron masses predicted by this approximation turned out to be quite close to nature. The meson sector is in very good agreement. The baryon sector is less so, but even there the proton mass came out to be about $1.1\text{GeV}$ which is only $15-20\%$ higher than the experimental value of $0.94\text{GeV}$. One also recalls that a variety of quark models, which are less sophisticated versions of the valence approximation, have been fairly successful for hadron spectroscopy. One would like to understand, then, how the pair creation and annihilation of the light quark pairs have only a small effect in the hadron mass spectrum.

The flavor singlet sector, on the other hand, is where we expect the vacuum loops of dynamical quarks to be important and are in fact needed for consistency. The long-standing $D(1)$ problem represents the first case at hand; we want to understand why the $\eta$ and $\eta'$ mesons are so heavy ($m_{\eta}=550\text{ MeV}$ and $m_{\eta'}=960\text{ MeV}$) compared to the pion ($m_{\pi}=140\text{MeV}$). The difference between the flavor non-singlet and singlet mesons comes from the diagrams for the latter which have flavor singlet intermediate states made of gluons and sea quarks. These diagrams carry an opposite sign compared to those having the valence $q\bar{q}$ pair and have a relative weight $N_f$, the number of quark flavors. A large mass splitting is possible only if the two kinds of diagrams have a comparable magnitude. Also, a proper calculation cannot be made in the quenched approximation in which $N_f$ is ill-defined.

A similar problem arises in the glueball mass spectrum. These are the states unique to QCD made out of gluons. They will mix, however, with the flavor singlet $q\bar{q}$ states. The magnitude of mixing largely affects the spectrum and nature of glueballs, and its correct estimate requires full QCD calculations. This problem is also related to the violation of the OZI rule.

Another area where the dynamical quarks are important is the behavior of QCD at high temperature and/or large baryon number density. The arguments going back to Isgur and Cabbibo and Parisi suggest that the confinement will be lost beyond a certain temperature where the normal hadronic phase will be replaced by a novel phase in which the quark and gluon degrees of freedom will become manifest. A simple model for such a phase change goes as follows. At zero temperature a quark-antiquark pair is bound by the color electric string whose energy $E$ increases linearly with the separation $r$, $E = \sigma \cdot r$ with $\sigma = (420\text{MeV})^2$ the string tension. At finite temperatures strings fluctuate thermally. The free energy of a string of length $L$ in such a thermal environment is $F = E - T \cdot S$ with $E = \sigma \cdot L$ the internal energy and $S$ its entropy, related to the number $N(L)$ of configurations of the string by $S = \log N(L)$. Modelling the thermally fluctuating string by a
random walk of basic step size \( l \), one has \( N(L) \approx (2d)^{d/4} \) with \( c = \log 2d \) and \( d = 3 \) the space dimension. We then have \( F \sim (c - cT/L) \) which shows that the effective value of the string tension decreases with the temperature, and eventually vanishes at a critical temperature \( T_c \approx \sigma / c \). Above this temperature quarks no longer feel the linearly rising confining potential, and are therefore liberated. Assuming \( T_c \) to have the typical scale of the strong interactions (1 GeV)\(^{-1} \), one estimates \( T_c \) to be a few hundred MeV. This then is a naive picture of deconfinement that will take place as the temperature is raised. The dynamical quarks might change this picture in a subtle way by allowing breakup and joining of strings through pair creations and annihilations. Quantitative characterization of the deconfinement and the effects of dynamical quarks on it is one of the basic tasks of full QCD studies.

Raising the temperature also affects chiral symmetry. At zero temperature this symmetry is spontaneously broken with pions as the Nambu-Goldstone bosons. This manifests itself in a non-vanishing value of the order parameter of the symmetry \( <\bar{q}q> \neq 0 \). One can then depict the vacuum of QCD as endowed with a "chiral spin" \( S = \frac{\bar{q} 
abla q + \nabla \bar{q} q}{2} \) at each space-time point which is aligned along a fixed direction in the internal flavor space. Raising the temperature will cause the spins to fluctuate. Just as in a ferromagnet, thermal fluctuations will destroy the ordering beyond a critical temperature, leading to the restoration of the symmetry \( <\bar{q}q> = 0 \). One does not expect quarks to play only a passive role in such a phenomenon. Inclusion of dynamical quarks is indispensable for properly understanding the chiral aspects of the finite temperature dynamics of QCD.

In addition to its own intrinsic theoretical interest in fully understanding the dynamics of QCD, the existence of a new phase of matter at high temperatures and/or large baryon number density has concrete physical implications. The possibility that such a new form of matter may be created in the course of heavy ion collisions has spurred much interest, and more than a few experimental programs are being actively pursued. It may also have some astrophysical consequences in phenomena such as type II supernova explosions, the possibility of quark stars, the nucleosynthesis and the density fluctuations in the early universe.

In this review we present the current status of lattice QCD simulations including dynamical quarks. We start with a brief summary of the formalism of lattice QCD in Sec. 2. Since the algorithmic development has been so crucial in the recent progress of this area, we shall describe in some detail why including fermions poses a problem and how it has been overcome (Sec. 3). We then discuss the physics results that have been accumulated by the recent full QCD simulations. The hadron mass spectrum calculation is presented in Sec. 4, while Sec. 5 is devoted to the thermodynamics of QCD. As we shall see, considerable progress has been made in many of the questions we have listed above. Unfortunately some problems, notably the \( U(1) \) problem and the behavior of QCD for large baryon number density, are still only poorly understood. We shall point out where the difficulties lie. We will close with a brief summary and an outlook in Sec. 6.

2. Review of formalism.

2.1 QCD on a lattice.

Lattice QCD is defined on a discrete Euclidean space-time lattice. The choice of the lattice structure should not matter if the mesh is small enough. In practice, however, it is most convenient to use the simple cubic lattice with the lattice sites having integer coordinates \( x = n_a n = (n_1, n_2, n_3, n_4) \in \mathbb{Z}^4 \). The parameter \( a \) denotes the lattice spacing which specifies the distance between the nearest neighbor sites. This is the only parameter in lattice QCD which carries the dimension of length. Dimensionful quantities are calculated in units of an appropriate power of \( a \). For example mass calculation gives the value of \( ma \).

One has to determine a to convert the lattice results into physical units.

The quark fields are assigned on lattice sites and will be denoted as \( q_a \) and \( \bar{q}_a \). The gauge fields denoted as \( A_{\mu} \), on the other hand, are placed on a link connecting the site \( n \) and \( n + \mu \) with \( \mu \) the unit vector in the direction of the \( \mu \)-th axis. The \( SU(3) \) nature of the gauge group is encoded into the assignment that \( U_{\mu} \) takes values in the group \( SU(3) \). The conventional gauge field \( A_{\mu}(x) \) is related to \( U_{\mu} \) by \( U_{\mu} = \exp(iaA_{\mu}(n_{\mu})) \).

The simplest choice of the action for the gauge field is given by

\[
S_g = \frac{\beta}{6} \sum_{\mu \nu \rho \sigma} \text{tr} U_{\mu \nu} U_{\nu \rho} U_{\rho \sigma} U_{\sigma \mu}^\dagger,
\]

with \( \beta = 6/g^2 \) the inverse gauge coupling constant squared. It is an easy exercise to show that this action classically reduces to the conventional form \(-1/2 g^2 \int d^4x \mathcal{F}_{\mu \nu} \mathcal{F}^{\mu \nu} \) in the limit \( a \to 0 \). A naive discretization of the Dirac Lagrangian leads to the quark action of the form

\[
S_q = \frac{1}{2}a^4 \sum_{n_a} (\bar{q}_a \gamma_0 A_{\mu}(n_{\mu}) U_{\mu} q_a - \bar{q}_a \gamma_0 U_{\mu}^\dagger q_a - \bar{q}_a \gamma_0 U_{\mu}^\dagger q_a + m_a a^3 \sum_n \bar{q}_a q_a).
\]

This action actually has a problem related to its chiral property which we will discuss below.

With the fields and actions thus defined, the quantum expectation value of an observable \( \mathcal{O} \) is given by the Feynman path integral:

\[
\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod_{n_a} \prod_{n} d\bar{q}_a d\bar{q}_a e^{S_q + S_g},
\]

with

\[
Z = \int \prod_{n_a} \prod_{n} d\bar{q}_a d\bar{q}_a e^{S_q + S_g}.
\]

For our purpose the lattice is simply a device for defining the theory in a mathematically well-defined manner. In order to obtain the physical predictions of the theory, one has to take the continuum limit of the lattice \( a \to 0 \). This requires a simultaneous change of
the coupling constant. Quantities having a mass dimension $l$ have a functional dependence

$$F = a^{-1} f(g^2, m_q a)$$

which diverges unless $g^2$ is properly adjusted as $a \to 0$. In more physical terms, the correlation length of the system in lattice units $\xi/a$ is a function of the coupling constant. For the physical correlation length $\xi$ to remain finite as $a \to 0$, $g^2$ should be varied toward the point $g^2_0$ where $\xi/a$ diverges to $\infty$, namely the critical point of a second order phase transition. For lattice QCD there is much evidence that this phase transition occurs at $g^2 = 0$ and nowhere else. Thus the continuum limit of lattice QCD means $g^2 \to 0$ (or $\beta = 6/g^2 \to \infty$) simultaneously with $a \to 0$.

Close to the continuum limit one can then use the weak-coupling expansion for calculating the renormalisation group $\beta$ function which dictates how $g^2$ should be varied with $a$. This has the well-known form given by

$$\beta(g) = \frac{dg}{da} = b_0 g^3 + b_1 g^5 + \ldots,$$

where

$$b_0 = \frac{1}{16\pi^2}(11 - \frac{2}{3} N_f), \quad b_1 = \frac{1}{18}\frac{1}{(16\pi^2)^2}(102 \cdot \frac{38}{3} N_f),$$

with $N_f$ the number of dynamical quark flavors. Integrating (5) leads to the lattice $\Lambda_L$ parameter;

$$\Lambda_L = a^{-1}\left(\frac{a}{b_0 g^3}\right)^{b_0/2b_1} \cdot \exp\left(-\frac{1}{2b_0 g^3}\right).$$

This quantity remains invariant under the change of $g^2$ and $a$ following (5) and carries the dimension of mass. It therefore sets the physical scale of QCD, and any dimensionful quantity calculated on a lattice should behave in the continuum limit as

$$F = a^{-1} f(g^2, m_q a) \sim \text{constant} \cdot \Lambda_L^4.$$

This relation tells us that the physical quantities as a function of the coupling constant have a very specific dependence close to the continuum limit. This is known as asymptotic scaling. For example, the hadron mass $m_{HA}$ in lattice units scales as

$$m_{HA} \sim \text{constant} \cdot \left(\frac{a}{b_0 g^3}\right)^{-b_0/2b_1} \cdot \exp\left(-\frac{1}{2b_0 g^3}\right), \quad g^2 \to 0.$$

Apart from exposing the non-perturbative nature of the dynamics, this relation provides a stringent test of the approach to the continuum limit. We also learn that the correlation length in lattice units increases very rapidly; $\xi_H/a = 1/m_{HA} \sim e^{\pi^2/16\pi^2}$. Since the lattice size needs to be larger than the correlation length in a meaningful simulation, one sees here a source of difficulty in pushing it close to $g^2 = 0$.

The lattice $\Lambda$ parameter (7) is numerically different from those defined in the continuum perturbation theory by momentum or minimal subtraction schemes. Their ratio can be determined by a one-loop calculation and depends on the lattice action used. For the single plaquette gauge action (1), one finds

$$\frac{\Lambda_M^2}{\Lambda_L} = \exp\left[\frac{1}{b_0} \left(0.234101 - N_f \cdot c_q\right)\right],$$

where the constant $c_q$ depends on the quark action used. Referring to the discussion below, we record that $c_q = 0.003347$ for the Wilson quark action (12) and $c_q = 0.001324$ for the Kogut-Susskind action (17). The ratio (10) is large and strongly dependent on $N_f$, but does not differ much for the two quark actions mentioned. For the Kogut-Susskind case, we have

$$\frac{\Lambda_M^2}{\Lambda_L} = \begin{cases} 28.81 & (N_f = 0, \text{pure gauge}) \\ 43.38 & (N_f = 2) \\ 56.74 & (N_f = 3) \\ 76.45 & (N_f = 4) \end{cases}$$

For the values for other forms of actions see refs. 20-24.

2.2 Lattice fermions and chiral symmetry.

We have mentioned that the naive quark action (2) has a problem related to chiral symmetry. Since this is the source of complications in the analysis of the chiral symmetry aspects, we elaborate on this in some detail.

Let us consider the free field case ($U_{ab} = 1$). The kinetic term of (2), in momentum representation, has the form $t \sum \gamma_{\mu} p_{\mu}/a$. Taking the limit $a \to 0$ this becomes $i\gamma_{\mu} p_{\mu}$, the free Dirac operator in the continuum. However, shifting $p_{\mu}$ by $\pi/a$ for a given $\mu$ and redefining $\gamma_{\mu} \to -\gamma_{\mu}$, one again obtains the free Dirac operator. One can clearly do this for each component of the momentum $p$. Hence the naive form (2) gives rise to $2^4 = 16$ particles in the continuum limit $a \to 0$. This is the problem of species doubling and is disastrous for at least two reasons. Firstly all 16 particles participate equally in the interaction with gluons. Thus a single field $q_a$ actually means 16 flavors, far more than the real value known and in fact saturating the bound of asymptotic freedom $N_f \leq \frac{33}{2}$. Secondly the redefinition $\gamma_\mu \to -\gamma_\mu$ means that the contribution to the $U(1)$ axial anomaly flips sign under the shift $p_{\mu} \to p_{\mu} + \pi/a$ for each $\mu$. The sum over 16 species therefore exactly cancels out; the theory has no $U(1)$ anomaly and the flavor singlet pseudoscalar will be degenerate with the non-singlet.

This is a general phenomenon arising from the periodicity of the lattice. There is in fact a theorem which states that any bilinear fermion action which is Hermitian, short-ranged, invariant under unit lattice translation, and possesses continuous chiral symmetry gives rise to an even number of species in the continuum limit. Thus there is no satisfactory way of writing down the lattice quark action which is chirally symmetric and does not suffer from the species doubling.
This forces us to resort to a compromise: Either one gets rid of doubling at the cost of chiral symmetry, or allows some doubling but keeps part of chiral symmetry. These choices correspond to the Wilson and Kogut-Susskind (KS or staggered) actions, to which we now turn.

The Wilson action has the following form:

\[ S_W = a^3 \sum_n \bar{q}_n q_n + a^3 K \sum_n (\bar{q}_n \gamma_5 U_{n+q_n+\bar{q}_n} \bar{q}_n - \bar{q}_n - \bar{q}_n) \]

\[ - a^3 K \sum_n (\bar{q}_n U_{n-q_n+\bar{q}_n} \bar{q}_n + \bar{q}_n U_{n-q_n+\bar{q}_n} \bar{q}_n - \bar{q}_n) \]

(12)

The first two terms coincide with (2) except for the normalization of quark fields. The third is the new term which explicitly breaks the chiral symmetry \( q_n \rightarrow e^{i\alpha \gamma_5} q_n, \bar{q}_n \rightarrow e^{-i\alpha \gamma_5} \bar{q}_n \). To see how this “Wilson” term removes the species doubling let us rewrite the coupling matrix in momentum representation for \( U_{n} = 1 \):

\[ D_W(p) = 2 K_i \sum_p \gamma_5 \frac{1}{a} \sin p a + \left( 1 - 2 K \right) \sum_p \cos p a. \]

(13)

We now have a momentum dependent mass term \( m(p) = 1/2 K / (1 - 2 K) \). For \( p \ll 0 \) we have \( m(p) = 1/2 K / (1 - 2 K) \). Hence the hopping parameter \( K \) should be tuned toward \( K \rightarrow 1/2 \) as \( a \rightarrow 0 \). But then if \( p_n \approx \pi / a \) for a given \( n \) we have \( m(p_n) \approx 1/2 K / (1 - 2 K) \) which goes to \( 0 \). This occurs whenever any of the components of \( p_n \) is close to \( \pi / a \). Thus out of the 16 species \( K \) has a mass of the order \( O(1/a) \) and are expected to decouple in the continuum limit, leaving only the state at \( p = 0 \) with the mass \( m = 1/2 K \).

In addition to removing the doubling, the Wilson mass term restores the \( U(1) \) axial anomaly. Since it is not invariant under the \( U(1) \) chiral rotation, it contributes to the divergence of the \( U(1) \) axial current. On a smooth background gauge configuration, this term has been shown to give rise to the correct anomaly as \( a \rightarrow 0 \).

The explicit breaking of chiral symmetry due to the Wilson term exists for any finite value of the lattice spacing. An example is provided by the weak-coupling expansion of \( < \bar{q} q > \) which receives non-vanishing contributions from the Wilson mass term. Since the lattice is a regulator of the ultraviolet divergences, such a breaking may be regarded as an artifact due to the regulator being not chiral symmetric. This prompts the question of how one could separate out such regularization effects to explore the true chiral content. This was examined in detail using the Ward identity. The divergence of an \( SU(N_f) \) axial-vector current \( A_{n \mu} \) receives an additional contribution from the Wilson mass term. This additional operator may be expanded in terms of operators of definite dimension. The essential point of the analysis is to show that the operators of dimension up to 4 which are the pseudoscalar density \( \bar{q} \gamma_5 \lambda \gamma_5 \bar{q} \) and the current divergence \( \nabla_\mu A_{\mu \nu} \) renormalize the coefficients of such operators that already exist in the Ward identity, while higher dimensional ones contribute only short-ranged contact terms in the Green's functions which at on-shell vanish as \( O(a) \) in the continuum limit.

The coefficient of the operator \( \bar{q} \gamma_5 \lambda \gamma_5 q \) in the renormalized Ward identity may be regarded as the bare quark mass. This takes the form \( 1/2a K (1/2K + f) \) with a subtraction constant \( f \) coming from the Wilson term. The critical hopping parameter \( K_c \) that corresponds to a vanishing quark mass is then determined by \( 1/K_c = f(K_c) \). One customarily ignores the slope \( \partial f/\partial K \) and defines the quark mass by

\[ m_q = \frac{1}{2a} \left( \frac{1}{K} - \frac{1}{K_c} \right). \]

(14)

An important question is how one finds the value of \( K_c \) in practice. In the weak-coupling limit \( \beta \rightarrow \infty \), one can evaluate the Ward identity by the weak-coupling expansion. With the single plaquette gauge action (1), this gives

\[ K_c(\beta) \approx \frac{1}{8} + \frac{0.0813}{\beta} + O(\beta^{-2}), \]

(15)

in agreement with the formula \( m_q = 1/4a (1/2K - 4) \) derived above for the free field case. At finite values of \( \beta \), one should expect that there are non-perturbative corrections as well. The two-point functions involving \( \nabla_\mu A_{\mu \nu} \) and \( \bar{q} \gamma_5 \lambda \gamma_5 q \) can still be used to determine \( K_c \) numerically.

The Ward identity analysis shows that the Wilson fermion action possesses chiral symmetry up to \( O(a) \) corrections. Insofar as these corrections are small, one may expect the PCAC relation

\[ (m_q a)^2 \sim m_q a = \left( \frac{1}{2a} \left( \frac{1}{K} - \frac{1}{K_c} \right) \right) \]

(16)

to hold. As a matter of fact, this provides an alternative way of determining \( K_c \): one checks whether the pion mass squared vanishes linearly in \( 1/K_c \) and if it does, defines \( K_c \) to be the value at which \( m_q = 0 \). It is possible to check the validity of the relation (16) explicitly in the strong coupling limit \( \beta = 0 \). One finds \( K_c(\beta = 0) = 1/4 \). Numerical simulations have shown that (16) is well satisfied also away from \( \beta = 0 \) with \( K_c \) moving toward the weak-coupling value \( 1/8 \) as the coupling becomes weaker. In the rest of this article, by \( K_c \) we always mean the value determined by the relation (16). We then define operationally the quark mass by (14).

Let us now discuss the Kogut-Susskind (KS or staggered) action. This is defined by

\[ S_{KS} = m_q a^4 \sum_n \bar{q}_n q_n + a^4 \sum_{n} \eta_n (\bar{q}_n U_{n+q_n+\bar{q}_n} - \bar{q}_n) \]

(17)

where \( q_n \) and \( \bar{q}_n \) are single component fermion fields and \( \eta_n \) is a sign factor such that the product around each plaquette is equal to \(-1\) (the most common choice is \( \eta_n = [-1]^{\eta_1 + \eta_2 + \eta_3 + \eta_4} \)). One way of deriving (17) is to make a transformation

\[ q_n = \gamma_1 \gamma_2 \gamma_3 \gamma_4 \chi_n, \quad \bar{q}_n = \bar{\chi}_n \gamma_1 \gamma_2 \gamma_3 \gamma_4 \]

(18)
Substitution into the naive action (2) gives (17) for each component of $\chi_\alpha$ which are therefore decoupled.

In order to discuss the flavor and chiral content of the Kogut-Susskind fermion, one needs to know how one represents Dirac fermions in terms of the single component $\chi$ fields. This is done by reassembling the 16 $\chi'$s on the vertices of a hypercube into 4 Dirac fermions.\(^{8,45}\) Let $\bar{n}$ denote the hypercube spanned by the 16 vertices $(n_a, n + \bar{\mu}, n + \bar{\nu}, \ldots)$. Define a field $\psi_n, \bar{n}$, for each hypercube $\bar{n}$ by

$$\psi_n, \bar{n} \equiv \frac{1}{64} \left( \delta_{n_a} \gamma_\alpha \sum_{\mu} (\gamma_\mu \gamma_5) \chi_\mu \bar{n} + \sum_{\mu < \nu} (\gamma_\mu \gamma_\nu) \chi_\mu \bar{n} \chi_\nu \bar{n} \right),$$

and similarly for $\bar{\psi}_{\bar{n}}$. Inverting this equation for $\psi_n$ and substituting into (17) then gives

$$(U_{n_\mu} = 1)$$

$$S_{KS} = m_4 (2a)^4 \sum_{\bar{n}} \left[ tr(\bar{\psi}_{\bar{n}} \gamma_5 \psi_n) + (2a)^4 \right] \right]$$

$$\left[ \sum_{\bar{n}} \left[ tr(\bar{\psi}_{\bar{n}} \gamma_5 \psi_n) - tr(\bar{\psi}_{\bar{n}} \gamma_5 \Delta_\alpha \bar{\psi}_{\bar{n}} \gamma_5 n_\alpha) \right] \right],$$

where $\Delta_\alpha \bar{\psi}_{\bar{n}} = (\psi_{n+2\bar{\mu}} - \psi_{n-2\bar{\mu}}) / 4a$, $\Delta_\alpha \psi_n = (\psi_{n+2\bar{\mu}} + \psi_{n-2\bar{\mu}} - 2\psi_n) / 4a$, and the trace is taken regarding $\psi$ as a matrix. From (19) and (20) follows the interpretation of $\psi_n, \bar{n}$; it represents 4 flavors of Dirac fermions with $\alpha$ the Dirac index and $i$ the flavor index. The first two terms in (20) are the mass and kinetic terms on a lattice of spacing $2a$, while the third is a Wilson-like mass term.

One can now discuss the chiral content of the Kogut-Susskind fields. The kinetic term of the action (17) has an $U(1)$ invariance $\chi_n \rightarrow e^{i\theta} \chi_n$, $\bar{\chi}_n \rightarrow e^{-i\theta} \bar{\chi}_n$, with the sign depending on $[\alpha] = n_1 + \ldots + n_4$ is even or odd. Under the transcription (19) this becomes $\psi_n \rightarrow \exp(i \theta) \gamma_5 \psi_n$ with the first $\gamma_5$ acting on the Dirac index and the second on the flavor index. This shows first of all that the Kogut-Susskind action retains the $U(1)$ subgroup of the chiral $SU(4)$ for $m_4 = 0$. Secondly one can regard $m_4$ as the quark mass, in spite of the presence of the Wilson-like term in (20), since the $U(1)$ symmetry protects the point $m_4 = 0$ from renormalization.

The Wilson-like term breaks the flavor $SU(4)$ symmetry. This effect is expected to become small in the continuum limit as in the Wilson case, leading to the restoration of the full $SU(4)$ chiral symmetry. Perhaps this may be examined by the Ward identity technique used for the Wilson action.

The Kogut-Susskind formalism is rigidly tied up with four degenerate flavors. It is not clear whether it allows a modification giving a different mass to each flavor. Another troublesome aspect concerns the $U(1)$ anomaly. One can construct a "current" like operator whose divergence reproduces the correct anomaly in the continuum limit.\(^{14}\) However, there are no group transformation corresponding to this "current" operator. Whether the Kogut-Susskind action has the proper ingredient to resolve the $U(1)$ problem is therefore not clear.

To summarize the Wilson and the Kogut-Susskind formulations, both have potential problems. The Wilson action has the advantage that the spin-flavor content is clear. This is not the case for the Kogut-Susskind action due to the encoding of the two indices into the site label. On the other hand, the Kogut-Susskind action has the $U(1)$ chiral symmetry, while the Wilson action suffers from an explicit breaking and requires an elaborate analysis including the fine tuning of the hopping parameter to extract its chiral content. Rigorously speaking one does not even know whether the two actions lead to the same continuum field theory. Under the circumstance using only one action is clearly not enough, and we will compare the physics results from the two actions in Sec. 4 and 5.

3. Dynamical fermion algorithms.

3.1 Source of difficulty.

From the computational point of view, the problem of lattice QCD amounts to evaluating an average of the form (3). To be more precise, one first calculates (3) on a finite lattice having $N_\mu$ lattice sites in the $\mu$-th axis ($\mu = 1, \ldots, 4$) imposing an appropriate boundary condition on $U_{n_\mu}$, $q_n$ and $\bar{q}_n$. One then takes the infinite volume limit $V = N_1 \times N_2 \times N_3 \times N_4 \rightarrow \infty$, and finally goes to the continuum limit $a \rightarrow 0$ and $\beta \rightarrow \infty$. The initial task therefore is the evaluation of (3) on a finite lattice. The Monte Carlo methods try to do this numerically by generating an ensemble of field configurations distributed according to the weight $\exp(S)$ and estimating the quantum expectation value (3) by an average over the ensemble. The particular difficulty of QCD, and of systems with fermions in general, stems from the fact that fermions are represented not by ordinary numbers but by anti-commuting Grassmann numbers which cannot be directly handled numerically.

The quark action in lattice QCD, however, has the bilinear form $S_q = \sum_\mu \bar{q} D (U_{\mu}, q, \bar{q})$. Hence one can carry out the integral over the quark fields producing the quark determinant $det D = \exp(T \log D)$. One then has an effective action expressed solely in terms of the c-number field $U_{\mu}$, $S_{eff} = S_q (U) + T \log D (U), (21)$ which in principle is amenable for the standard Monte Carlo treatment.

In order to see what this involves in practice, we recall the algorithm of the standard Metropolis procedure.\(^{11}\) Given some configuration of the link variables $U$, one generates a trial configuration $U'$ by some stochastic rule. If this rule is reversible (i.e., $\text{prob} (U \rightarrow U') = \text{prob}(U' \rightarrow U)$), accepting the trial configuration with the probability $\exp(-S(U) - S(U'))$ generates an ensemble of configurations with the distribution $\exp(S_{eff})$. This requires a computation of the ratio of determinants $det D (U') / det D (U)$.

We all know that computing a determinant is a time consuming affair, but to appreciate how true this is for lattice QCD, let us estimate the computer time needed to evaluate a determinant for a given gauge configuration. The dimension of the matrix $D$ is equal to $n = 3(\text{color}) \times 4(\text{Dirac}) \times V$. (For the KS form, the Dirac index is absent.) A standard way to compute a determinant is to eliminate the non-vanishing elements in the lower triangle,
since then the determinant is simply the product of the diagonal elements. This requires
about \(N \approx n^3 = O(V^3)\) arithmetic operations. (The coefficient depends on the specific form
of \(D\), which we ignore for simplicity. We are also ignoring a possible reduction due to the
sparseness of \(D\).) Even for the K5 form, \(n^3\) is roughly \(10^6\) already on a \(4^4\) lattice.
The typical execution speed of supercomputers readily available now such as Cray-XMP is 100
or 200 Mflops, i.e., \((1-2) \times 10^{10}\) floating point operations per second. On such computers,
one would need roughly 1 second to compute a single determinant. Typically one changes
the link variable \(U_n\) one at a time. To obtain a new configuration one should at least cycle
through the entire set of the link variables. Thus the computer time for a single such sweep
is \(N \times 4 \times V\) (number of links) which is about 1,000 seconds for the lattice as small as \(4^4\),
and this grows as \(V^4\) as one increases the lattice volume. Normally one repeats the Metropolis
trial a number of times at each link update to increase the probability that a new value of
\(U_n\) is accepted (>50% in practice). Furthermore thousands of sweeps are generally needed
to achieve a reasonable statistical accuracy. These qualitative considerations show that the
use of the Metropolis algorithm with a direct evaluation of the quark determinant is not feasible in practice.

3.2 Hybrid Monte Carlo and related algorithms.

The primary reason which makes the naive algorithm so time consuming is the \(O(V^4)\)
dependence of the number of arithmetic operations needed per sweep, with \(V^3\) coming
from \(det(D)\) and the additional \(V\) due to cycling through all links. This is in contrast to the
pure gauge case where evaluating the change of action requires only a fixed number
of arithmetic operations for each link update, and hence \(O(V)\) per sweep. The progress
over the last several years has come from the realization that algorithms with an \(O(V)\)
dependence are possible at the cost of restricting the change of configurations to be small.
Historically this progress took place step by step taking the form of a variety of proposals
called microcanonical\(^{52}\), Langevin\(^{63,44}\), hybrid\(^{65,66}\) and hybrid Monte Carlo\(^{67}\) algorithms.
In fact they are all closely related, depending essentially on two key ingredients, and can be
discussed in a unified manner. The two key features are, first, the trading\(^{68}\) of computing a determinant \(det(D(U))\) by solving a linear equation \(D(U)\vec{x} = \vec{Y}\) to obtain \(x = D^{-1}(U)\vec{Y}\) for a given vector \(\vec{Y}\), and second, an updating procedure of the entire gauge configuration involving only a limited number of such inversions independent of the lattice volume. As we shall see the former saves a factor of \(V^3\) and the latter an additional factor of \(V\).

Matrix inversion. Let us start with the first point. The trading is achieved by rewriting
the determinant as

\[
det(D(U)) = \int \prod_n dY_n dY_\varepsilon \exp\left( - \sum_{n,n'} Y_n^\dagger D^{-1}(U)_{nn'} Y_{n'} \right),
\]

using a complex field \(Y\). Strictly speaking the validity of this identity requires positivity of the
matrix \(D(U)\). This is not true in general for Wilson and KS actions. However, they
satisfy an identity \(\frac{1}{4} \Gamma D U D \Gamma = \gamma_{
\]
is controlled by the condition number $c = \lambda_{\text{max}}/\lambda_{\text{min}}$ with $\lambda_{\text{max,min}}$ the maximum and minimum eigenvalue of the matrix $D$. For the conjugate gradient algorithm, for example, it is possible to prove that:

$$\frac{\|x_b - D^{-1}Y\|}{\|x_b - D^{-1}Y\|} \leq 2\left(\frac{\sqrt{c} - 1}{\sqrt{c} + 1}\right)^b.$$  \hspace{1cm} (23)

In lattice QCD applications with $D = D_1D_2$, we have

$$\lambda_{\text{min}} = \begin{cases} \{m_a\}^2, & \text{for KS,} \\ (2Km_a)^2, & \text{for Wilson.} \end{cases}$$  \hspace{1cm} (24)

For the KS action this is easily seen from the anti-hermiticity of the kinetic part (see (17)). For the Wilson case, this is a semi-empirical formula$^{56,51}$, but the additional factor of $2K$ may be ascribed to the normalization of the quark fields. The maximum eigenvalue, on the other hand, is of order $O(1)$. We see that the condition number is not very sensitive to the lattice volume $V$. Therefore the number of steps for solving $Dx = Y$ to a given accuracy remains roughly constant with increasing $V$. This gives us a gain of a factor of $O(V^2)$ over the computation of the determinant.

There is one problem which needs to be discussed. A glance at (23) and (24) shows that for small quark masses the number of steps necessary to achieve a given accuracy increases as $1/m_a$. This is an example of critical slowing down$^{56,59}$. A general procedure for alleviating this problem is to reduce the condition number by preconditioning the matrix $D$. The idea is to prepare a matrix $D_0$ such that $D_0^{-1}D$ is closer to the identity matrix, and solve $D_0^{-1}Dx = D_0^{-1}Y$. The matrix $D_0$ should be easily invertible. For the Wilson case, a successful choice is$^{55}$

$$D_0 = LU,$$  \hspace{1cm} (25)

with

$$L = 1 - cK\sum_{\alpha}(1 + \gamma_a)U_\alpha^\dagger, \quad U = 1 - cK\sum_{\alpha}(1 - \gamma_a)U_\alpha,$$

which are triangular and easily invertible. This is an example of an incomplete LU (lower-upper) decomposition$^{54}$ of $D$. The choice of the parameter $c \sim 1.2$ with the conjugate residual algorithm$^{57,58}$ for the linear solver has led to a factor $\sim 15$ improvement over the standard conjugate gradient in the number of steps$^{55,56}$. Another choice$^{59,60}$, called Fourier acceleration, is $D_0\{p\} = iV_p + m$ in momentum space. The inversion $D_0^{-1}$ is carried out by fast Fourier transform, and this has yielded a similar gain for the Wilson case. On the other hand, no efficient preconditioner has been found for the KS action. The main reason for this is the fact that the condition number is small for the KS action and the standard conjugate gradient already converges quite fast$^{51,61}$.

**Hybrid Monte Carlo algorithm.** Let us now go on to the second key ingredient for the reduction from the $O(V^4)$ to $O(V)$ algorithm, the global update of the field configuration.

We recall that the Metropolis update accepts a trial configuration with the probability $P_{\text{acc}} = \min(1, e^{\delta H})$. If one randomly changes all the link variables simultaneously, the change of the action will be proportional to the volume $V$ and be negative$^{55}$. Thus the individual links cannot be varied too much since the acceptance will be essentially zero otherwise. One needs a scheme of evolving the entire system in which one can control the change of action to a small value.

An elegant way is provided by the microcanonical (or molecular dynamics) method$^{48,47}$. Consider a "Hamiltonian" given by

$$H = \frac{1}{2}\sum_{\mu}\{tr(X_{\mu\mu}^2) - S_{\text{eff}}(U_v)\},$$

$$S_{\text{eff}} = S_q(U_v) - \sum_{\gamma'\gamma} Y_{\gamma'}^\dagger D^{-1}(U_{\alpha\gamma}Y_{\gamma'})^{\dagger},$$  \hspace{1cm} (26)

where $X_{\mu\mu}$ is an(3)-algebra valued momenta conjugate to $U_{\mu\mu}$. Given a configuration $\{X, U, Y\}$ one can use the Hamilton-Jacobi equation

$$-iU^{-1}\frac{dU}{d\tau} = X,$$

$$\frac{dX}{d\tau} = -\frac{\delta H}{\delta U} = -\frac{\delta S_q}{\delta U} + YD^{-1}\frac{\delta D}{\delta U}D^{-1}Y,$$  \hspace{1cm} (27)

to evolve the system with respect to the fictitious time $\tau$ and use the end configuration $\{X_1, U_1, Y\}$ for the Metropolis trial. (Here we have kept $Y$ fixed. One could also introduce momenta conjugate to $Y$ and let them evolve in $\tau$. The modifications in the discussion below are straightforward.) In fact, since generating an ensemble with the distribution $\exp(-H)$ is equivalent to generating that with $\exp(S_{\text{eff}})$, one may equally well apply the Metropolis step with the probability

$$P_{\text{acc}} = \min(1, \exp(-\delta H)),$$  \hspace{1cm} (28)

with $\delta H = H(X_1, U_1, Y) - H(X, U, Y)$. If one follows the evolution (27) exactly, the trial configuration is always accepted since (27) conserves the Hamiltonian $H$. This is not possible numerically. For computer implementation, one has to discretize the fictitious time $\tau$ into discrete units of a size $\Delta \tau$. The crucial point$^{47}$ now is that any discretization generates the correct distribution if it preserves the phase volume and if the motion is reversible (i.e., starting from $\{X_0, U_0\}$ leads back to $\{X_0, U_0\}$). The proof is almost trivial$^{67}$. One draws the starting configuration of $X$ and $Y$ from a Gaussian ensemble $\exp(-1/2\tau X^2)$ and $\exp(-\eta Y^2)$ with $Y = D\eta$ for a given $U$ (recall $D = D_0D_1$). The transition probability for $U$ is then

$$W(U \rightarrow U') = \int dX dY e^{-1/2\tau X^2} P_U((X, U) \rightarrow (X', U')) P_{\text{acc}}((X, U) \rightarrow (X', U'))$$  \hspace{1cm} (29)

where $P_{\text{acc}} = \delta((X', U') - (X_1, U_1))$ represents the transition probability of the discretized molecular dynamics evolution. One can then easily check the chain of identities,

$$e^{S_{\text{eff}}(U)}W(U \rightarrow U').$$
\[= \int dX dX' e^{-H(X,U)} P_H((X,U) \to (X',U')) P_{\text{acc}}((X,U) \to (X',U'))
= \int d(-X')d(-X) e^{-H(-X',-U')} P_H((-X',-U') \to (-X,U)) P_{\text{acc}}((-X',-U') \to (-X,U))
= W(U' \to U) e^{S_{\text{eff}}(U')}
\]

where the reversibility is used at the third line. This is the detailed balance condition which, together with ergodicity, guarantees the unique convergence of the distribution to \(\exp(S_{\text{eff}})\). The ergodicity is in turn ensured by drawing a fresh \(X\) and \(Y\) after each Metropolis trial.

Finding a discretization of (27) satisfying the two conditions given above is not difficult. A well-known example is the leap-frog procedure which assigns \(X\) and \(U\) to integer and half-integer time steps (or vice versa) and move them forward in time in turn. The magnitude of the change of the Hamiltonian \(\delta H \sim \mathcal{O}(\Delta r^2)\) (with \(n = 2\) for the simplest leapfrog) can clearly be controlled through an appropriate choice of the step size \(\Delta r\). Furthermore moving the system one step forward requires only a single matrix inversion for the entire lattice (recall (27)). We have thus achieved the reduction of an extra factor of \(\mathcal{O}(V)\) without losing control over the acceptance.

The method just described is a hybridization of the molecular dynamical evolution through the Hamilton-Jacobi equation and the Metropolis-type global accept/reject step, and is known as the hybrid Monte Carlo algorithm\(^{47,48,44}\).

There are two questions which are important for practical applications. First of all, how much can one increase the step size \(\Delta r\) before the acceptance becomes unacceptably low? Trial runs\(^{47,48,70}\) on a variety of lattice sizes indicate that it can be maintained above 50% level with \(\Delta r \sim 0.02 - 0.04\) on a lattice of size \(4^8\) and the quark mass of \(m_q \sim 0.05\). In addition there exist an asymptotic estimate\(^{44,48}\) that the acceptance behaves as \(\sim \exp\left(-\alpha \Delta r^2 V\right)\) for the number of molecular dynamics steps \(N_{\text{md}}\) between the Metropolis trial chosen to be \(N_{\text{md}} \Delta r \sim 1\). This estimate indicates that the step size need be reduced relatively slowly (\(\sim V^{-1/4}\)). We also note that the acceptance depends only weakly on \(N_{\text{md}}\) for a fixed \(\Delta r\). This is because the Hamiltonian only oscillates around an average value in a phase volume preserving discretization such as the leap-frog procedure\(^{47,48}\).

Another question concerns the precision with which to compute \(D^{-1}Y\). For the global accept/reject step, this should be done precisely enough such that the relative error in the norm \(||D^{-1}Y||\) is much less than \(O(1/V)\). Otherwise the value of \(\delta H\), which is \(O(1)\) for reasonable acceptance, will suffer from a large error, leading to an uncontrollable distortion of the distribution of configurations. More subtle is the effect of imperfect inversion in the molecular dynamics step. At a naive level, this type of error will lead to an irreversibility of the evolution, and hence to a systematic bias. However, this is not necessarily the case\(^{44}\).

An imperfect inversion may be restated as solving exactly for some other matrix \(\hat{D}\) whose precise form depends on the algorithm, the choice of the initial vector and the way one steps the iteration. If one can choose \(\hat{D}\) to be such that \(\hat{D}\) depends only on the configuration of \(U\) at the current molecular dynamics step, then the distribution generated is still exactly \(\exp(S_{\text{eff}})\). This is because the proof (30) actually goes through independent of the form of the Hamiltonian used in the molecular dynamics step\(^{47}\). Since the computer time in full QCD simulation is dominated by the inversion, and hence directly proportional to the number of steps carried out, this point deserves further careful studies. A practical check may be provided by a simple identity\(^{44}\)

\[< e^{-\beta \hat{H}} > = 1, \tag{31} \]

which holds once the system is in equilibrium.

**Hybrid and Langevin algorithms.** The hybrid Monte Carlo algorithm is the most recent of a class of \(\mathcal{O}(V)\) algorithms. For a small enough step size for which the acceptance is close to 100%, one may opt to drop the global accept/reject step. In other words, one simply repeats the process of refreshing the momenta \(X\) and \(Y\) followed by the molecular dynamics evolution. This is the hybrid molecular dynamics\(^{73}\) as applied\(^{44,48,72}\) to lattice QCD and is called the hybrid algorithm.

Suppose one makes a further restriction to a single step in the molecular dynamics evolution. The initial and final configurations \(U\) and \(U'\) are then related by

\[U' = U \cdot e^{-\frac{\delta W_{\text{eff}}}{\Delta r}} \cdot e^{\Delta r X}, \tag{32} \]

Since \(X\) is drawn from a Gaussian ensemble, this is the Langevin equation for a Brownian motion of \(U\) in a potential \(S_{\text{eff}}\), discretized with the step size

\[\Delta r = (\Delta r)^2 / 2. \tag{33} \]

This is the Langevin algorithm\(^{43,44,74,72}\) based on the general proposal of ref. 10.

The absence of the global accept/reject step in the hybrid and Langevin algorithms has an advantage. Since one no longer has to compute the Hamiltonian (29), one might go back to the determinant form (21). Going through the relevant formulism, we find that what we actually need to compute is

\[\text{Tr}(D^{-1} \frac{\delta D}{\delta \xi})], \tag{34} \]

at each update. We now note that the trace operation may be replaced by

\[\xi D^{-1} \frac{\delta D}{\delta \xi}, \tag{35} \]

where \(\xi\) and \(\xi\) are Gaussian noise, because for infinite volume \(V \to \infty\) (35) self-averages to (34). The algorithmic requirement is once again the solution of the equation \(Dx = \xi\). This is known as the bilinear noise scheme\(^{11}\) in the Langevin case, and as the \(R\) algorithm\(^{48}\) for hybrid. Without this modification, every 2 flavors requires the corresponding field \(Y\) (4 for KS action), and one has to evaluate \(D^{-1} Y\) for each \(Y\). With \(\text{det} D\), on the other hand,
one simply raises it to the power \(N_f/2\) (\(N_f/4\) for KS; recall \(D = D_1^2 D_2\)). This leads only to the multiplication of (35) by \(N_f/2\) or \(N_f/4\). One may even take \(N_f\) to be other than those allowed (e.g., \(N_f = 2\) for KS or even non-integer values), since one expects \(det D\) to converge to \(\text{det}(D)^\alpha\) in the continuum limit where \(D\) is the single flavor continuum Dirac operator and \(\alpha = 2\) for Wilson and \(\alpha = 4\) for KS. For the KS action, in particular, this is the only procedure known for simulating the number of dynamical quark flavors less than 4.

For the hybrid Monte Carlo algorithm, maintaining a reasonable acceptance determines the magnitude of the step size \(\Delta r\). The Langevin and hybrid algorithms, not going through the global accept/reject step, may seem to avoid this problem. In fact the problem appears in the form of a \(\Delta r\)-dependent systematic error in the observable averages. This is most easily seen for the Langevin algorithm\(^{13,44,74}\).

Consider the Langevin update (32) in the bilinear noise scheme. Denote the configuration at the Langevin time \(r_0 = n \Delta r_0\) as \(U^{(n)}\). The distribution of configurations at this time is given by

\[
\rho^{(n)}(U) = \delta(U - U^{(n)}) \times L
\]

(36)

One can use (32) to determine the \(n\) dependence of \(\rho^{(n)}\) in a power series expansion in \(\Delta r_0\):

\[
\rho^{(n+1)}(U) - \rho^{(n)}(U) = -\Delta r_0 \cdot H_{FPNP} \rho^{(n)} - \Delta^2 r_0 \cdot H_{FPNP}^2 \rho^{(n)} - \ldots
\]

(37)

with

\[
H_{FP} = \frac{1}{D_0} \left[ \frac{\partial S_{eff}}{\partial U} \right] - \frac{1}{D_0^2},
\]

(38)

and \(H_{FPNP}^2\) a more complicated differential operator. In the limit \(\Delta r_0 \to 0\) (37) becomes the well-known Fokker-Planck equation \(\partial \rho / \partial t = -H_{FPNP}\), which shows the unique stationary distribution to be \(\rho_{eq} = \exp(S_{eff})\). For a finite \(\Delta r_0\), however, the higher order terms in (37) distort the distribution to the form

\[
\rho_{eq}(U) \sim \exp(S_{eff}(U) + \Delta r_0 S^{(1)}(U) + \ldots)
\]

(39)

which in turn leads to a step size dependence of the observables,

\[
F_{\Delta r_0} \equiv \langle O \rangle_{eq} = F_{\Delta r_0 = 0} + \Delta r_0 \cdot F^{(1)} + \ldots
\]

(40)

For the hybrid algorithm the analysis becomes more complicated to take into account the molecular dynamics steps. An elaborate analysis\(^{46}\) shows that the limiting distribution is essentially controlled by the Langevin step (i.e., the refreshing of the moments \(X\) from a Gaussian distribution) with an equivalent step size of \((\Delta r)^2 / 2 = (\Delta r_0)^2\), and hence the systematic error is essentially the same as (39) and (40) (with the substitution (33)). For the R-algorithm, there is a danger that this error is overwhelmed by a larger error from the molecular dynamics steps. However, a careful choice\(^{46}\) of the discretization procedure maintains the latter at the same order of magnitude \(O(\Delta r^2)\).

There are at least four ways to deal with the systematic error coming from a finite step size: (i) Improve the discretization procedure so that the error becomes higher order. (ii) Use small enough step size so that the error is less than the statistical. (iii) Carry out the simulation at several values of the step size and extrapolate to zero. (iv) Examine if the higher order terms in (39) may be absorbed into a redefinition of the coupling parameters up to terms irrelevant in the continuum limit. The first choice generally means an increase in the computer time since it is achieved by dividing the basic time step into finer steps. In the Langevin algorithm, this is nonetheless worthwhile for the pure gauge sector for which an efficient algorithm with \(O(\Delta r^2)\) error has been developed\(^{13,43,46}\). There are also a few attempts\(^{77,78}\) for full QCD which, however, do not appear very practical\(^{46}\). The hybrid algorithm is already second order, and it does not seem worth the computational cost to pursue (i).

Let us note an important point regarding (i) and (ii) for full QCD at small quark mass. Analyses of the higher order term in the limiting distribution (39) show\(^{43,44}\) that the deviation \(S^{(1)}\) has terms involving \(1/D\), and the power increases for higher orders. Thus the magnitude of systematic error is not simply determined by the step size but rather more correctly estimated in terms of \(O(\Delta^2 r_0/\lambda_{min})\) for the Langevin and \(O(\Delta^2 r_0/\lambda_{min})\) for the hybrid algorithm with \(\lambda_{min}\) the minimum eigenvalue of \(D\) given by (24). We see then that the option (i) and (ii) will become increasingly impractical as \(\lambda_{min}\) decreases. Higher order algorithms will be of little help when the ratio given above exceeds a magnitude of \(O(1)\). Reducing the step size to a small enough value will become too costly in terms of the computer time since moving the system over a unit time interval is inversely proportional to the step size. To give an idea on the safe step size, we quote an empirical rule for the Kogut-Susskind action that 1% accuracy in the local quantities such as the 1 x 1 Wilson loop requires \(\Delta r \approx m_q a\) for the hybrid algorithm\(^{52}\) and \(\Delta r \approx (m_q a)^2\) for the Langevin\(^{51}\), both with a coefficient of 1 - 2. For the Wilson action only the Langevin case has been examined\(^{48}\). The systematic error is generally large necessitating an extrapolation in the step size even for heavy quark (\(m_q a \approx 0.5\)). Toward \(K = K_s\), the 1% accuracy again appears to require \(\Delta r / (m_q a)^2 \approx 1 - 2\). In practice the step size cannot be reduced much below \(O(0.01)\). With the Langevin algorithm, the quark mass below \(m_q a = 0.1 - 0.05\) therefore has to be treated by an extrapolation procedure to find a bias-free result\(^{50,51}\). The hybrid algorithm, being second order, can handle \(m_q a\) as low as \(0.01 - 0.02\) with \(\Delta r \approx 0.01\) (The smallest so far tried is 0.0125), but not below.

We finally comment on the option (iv). The point\(^{44}\) is that the form of the action may be modified as in (39) as long as it belongs to the same universality class and therefore has the same continuum limit. If this is true, one can ignore the problem of the systematic error since the dimensionless quantities such as the mass ratio will converge to the correct value in the continuum limit \(\beta \to \infty\), although checking the asymptotic scaling (8) will be difficult. For the pure gauge sector this is easily seen\(^{43,44}\) to be valid to the leading order in
$\Delta r$, and is likely to remain so since higher order terms in (39) will be local. For full QCD, the locality is much less certain because $detD$ and $YD^{-1}Y$ are both non-local quantities. In fact for the Langevin algorithm in the bilinear noise scheme, the Fokker-Planck operator in (37) contains terms\cite{44} which cannot be integrated to a local function of $U$ in the limiting distribution (39). Further study is evidently needed.

**Relaxation rate.** In numerical simulations expectation values of observables are estimated by an average over an ensemble of configurations. To make the resulting values meaningful it is necessary to estimate the statistical error. This requires knowing how many configurations in an ensemble may be regarded as independent. Since the configurations are generated sequentially, this means studying how fast the correlation between successive configurations decays as a function of the simulation time. This problem is particularly acute for the $O(V)$ algorithms for dynamical fermions since they have achieved the computational feasibility at the cost of a small step size, and therefore a smaller change of the configuration per step as compared to the standard Monte Carlo using exact evaluation of the fermion determinant.

A standard apparatus for studying the correlation in time is the autocorrelation function defined by

$$A_0(T) = \frac{\omega_T(T)}{\omega_0(0)}, \quad (41)$$

with

$$\omega_T(T) = \langle O(T)O(T+\tau) - \langle O(T) \rangle \langle O(T+\tau) \rangle \rangle,$$

where $O(T)$ denotes the value of an observable at the simulation time $\tau$, and the bar means average over $\tau$. For a stochastic sequence this function is expected to have an envelope which decreases with $T$. Roughly speaking the two configurations may be regarded as independent after a time $T = \tau$, beyond which $A_0(T) \approx 0$. The precise value of $\tau$, however, depends on a variety of factors: (i) Definition. Some of the ones used are: a) $A_0(\tau) = e^{-1}$, b) $\tau = \int_0^\infty dt \langle T A_0(t) \rangle / \int_0^\infty dt \langle A_0(t) \rangle$ and c) $A_0(\tau) = 1/\sqrt{N}$, i.e., the value expected for a sequence of $N$ samples without correlation\cite{46}. (ii) Observables $O$. (iii) Parameters of the system such as the coupling constant $\beta$ and the quark mass $m_u$ or (iv) Algorithm of simulation.

In practical simulations statistics are often not enough to allow determination of $A_0(T)$ beyond a limited range of $T$. One is frequently forced to a case by case examination of the autocorrelation function and simulation time histories of observables to estimate the relaxation time $\tau$. This is particularly true for detecting long-periodicity fluctuations. For the $O(V)$ algorithms, a relaxation time of much less than unity will not naturally occur and actual simulations tend to give a larger value of $O(1) - O(10)$ and sometimes greater. With the typical step size $\approx 0.01$, $O(1000) - O(1,000)$ sweeps or more are therefore needed to generate an independent configuration. This is typically an order of magnitude larger than the standard Monte Carlo algorithm.

The importance of the dependence on $\beta$ and $m_u$ arises from the critical slowing down. In general one expects the relaxation time $\tau$ to be controlled by the smallest physical mass scale of the theory, and to increase by some inverse power as the latter decreases. One instance of this is the limit $m_u \to 0$ since the pion mass tends to zero. Another is the continuum limit $\beta \to \infty$ where hadron masses scale down as $m_{had} \sim e^{-1/\beta m_{mass}}$. Since these are precisely the limits of physical relevance, methods for overcoming such slowing down will become important. A few possibilities explored are the Fourier acceleration\cite{52,41,55} and application of multi-grid methods\cite{53,49,50,51}. The study is still in its infancy and should be examined further.

A question of practical significance related to the relaxation rate is the choice of step size parameters. One would like to tune them so that the correlation between the configurations separated by a given number of updates $N$ (namely, after a given amount of computer time) is minimized. For the hybrid Monte Carlo, this is achieved by maximizing the effective length of the time interval given by $t_{HMC} = N \Delta T \cdot p(\Delta T)$ the acceptance, which means maximizing the product $\Delta T \cdot p(\Delta T)$. One can try a further tuning by changing the length of the molecular dynamics trajectory $t_{HMC} = N \Delta T$ between the Metropolis trial. Several studies\cite{48,52,53,54} (the first three within the hybrid algorithm) have found $t_{HMC} \approx 1$ to be the optimal value. For the hybrid algorithm, the correlation after $N$ updates will decrease with increasing $\Delta T$. One cannot take $\Delta T$ too large, however, because the trajectory will become unstable and also because of the systematic error. Since this is a second order algorithm one would like the error to be small compared to the statistical error. This requires $\Delta T \approx m_u$. For the optimum length of the trajectory the value $\Delta T_{MC} \approx 1$ applies.\cite{48,53,54}

The Langevin algorithm has only one tunable parameter $\Delta T$. Since the systematic error is large ($O(\Delta T/\langle m_u \rangle)$), an extrapolation is generally needed for its control. Thus the only criterion for the choice of $\Delta T$ is that it be small enough to ensure the linear behavior of the systematic error.

Let us now consider the dependence on the algorithm (iv) and compare the three $O(V)$ algorithms. We note first that the computations carried out at each update is essentially the same and therefore the number of updates for a given amount of computer time is roughly similar. (We are leaving aside the question that the matrix inversion may have to be carried out much more precisely for the hybrid Monte Carlo because of uncertainties involved.) For the hybrid Monte Carlo and hybrid which differ only by the accept/reject step, one may use the same step size parameters $\Delta T$ and $\Delta T_{MC}$. One does not expect much difference in the relaxation as long as the acceptance of the hybrid Monte Carlo is high. For a low acceptance, the hybrid Monte Carlo will have a slower relaxation by the amount of acceptance. Presumably, however, the hybrid algorithm will still suffer from a non-negligible systematic error. With the same precision used for the matrix inversion, the two algorithms will therefore be similar in their performance.

Comparison with the Langevin algorithm is a little more involved. To restrict the
systematic error to a magnitude similar to the hybrid, one has to choose $\Delta \tau = 1/2|\Delta r|^2$. With the typical value $\Delta r \approx 0.02 - 0.05$ the Langevin step size is quite small and the relaxation after the same number of steps will be much slower for Langevin. On the other hand, if one opts to use the extrapolation in the step size for Langevin, the comparison should be made with the choice $\Delta \tau = \Delta r$ since the statistical rather than systematic error is the point of comparison now. In this case the Langevin relaxes much faster than the hybrid by a factor of 3 - 5 both for pure gauge and full QCD. This means that the number of sweeps for the hybrid run has to be an order of magnitude larger to achieve the same statistical accuracy, and even more for the hybrid Monte Carlo due to the acceptance factor. The Langevin run, however, has to be repeated at several values of the step size for the extrapolation, and the total computing cost turns out to be roughly similar.

One of the advantages claimed for the hybrid algorithm is that molecular dynamics evolution leads to a faster relaxation than the Langevin algorithm since it moves the system more rapidly through the phase space than the Brownian motion of the latter. This would be true if the classical trajectory loses correlation over the time interval commonly used $\sim 1$ in the molecular dynamics updates. The results quoted above indicate that this is not the case. The time interval is too short, and the loss of correlation is still controlled by the refreshing of the momenta, i.e., the Langevin step of the simulation.

Summary. The $O(V)$ algorithms described above have all been tested in real simulations and have shown to work on a lattice size of order $8^4 - 10^4$ and the quark mass down to $m_{\pi} \approx 0.02 - 0.05$. Though closely related, they differ in detail each having advantages and drawbacks. The Langevin has a fast relaxation but suffers from a large systematic error. It is opposite for the hybrid algorithm with a small systematic error but a slow relaxation. The hybrid Monte Carlo is probably the most appealing since it is exact. However, this is obtained at the cost of rejecting undesirable configurations leading to an even slower relaxation rate. Furthermore the matrix inversion has to be made more precisely in general, which brings in an additional requirement on the computer time. One should be aware that these problems become increasingly severe as the quark mass decreases.

3.3 Other alternatives.

We have focused our attention on the hybrid Monte Carlo and its relatives. This is because a variety of tests and actual simulations so far carried out show them to be the best available for full QCD. Other alternatives exist, mainly of the $O(V^2)$ category. In particular, an algorithm which combines an efficient way of computing the ratio of the quark determinant $detD(Q(U)/U)$ to $detD(U)$, and matrix inversion to estimate $D^{-1} U$ has been used to some extent in finite temperature simulations, though restricted to a small lattice of $4^4$. Most of the others have not been tested in a practical context. We shall refer the reader to the literature, and here only add a few comments on the two which are $O(V)$ and have been utilized.

Consider extending the time interval of the molecular dynamics steps to $\infty$. If the system is ergodic, the trajectory will cover the entire phase space, and the average over the trajectory will be equivalent to the one over a microcanonical ensemble. For infinite volume, the microcanonical ensemble with the measure $dX dU H(B - E)$ is equivalent with the microcanonical ensemble with the measure $dX dU \exp(-\beta H)$, with the temperature $\beta$ determined by the equi-partition theorem $<trX^2/2> = 8V/2\beta$. This is the microcanonical algorithm. Unfortunately, the ergodicity assumption is suspect for QCD. The theorem of Kolmogorov-Arnold-Moser tells us that a system cannot be fully ergodic if it has integrable limits, and if there is a series of resonances or tori. For full QCD this precisely applies to the weak-coupling limit where it reduces to a free field theory of quarks and gluons. One way to avoid the potential lack of ergodicity is to collect a large number of trajectories with random starting points. This is precisely the hybrid algorithm.

Another possibility is the pseudo-fermion algorithm. This is based on the identity

$$\frac{det D(Q(U))}{det D(U)} = \frac{\int d\phi d\phi^* e^{-\beta SU} e^{\phi^* D(U)\phi}}{\int d\phi d\phi^* e^{-\beta SU} e^{\phi^* D(U)\phi}} = e^{-\Delta S_{pf}}.$$ \hspace{1cm} (42)

where $\Delta S_{pf} = \phi^* D(U)\phi - \phi^* D(U)\phi$. The restriction is that $\Delta S_{pf}$ is small, one then expands $\Delta S_{pf}$, and replaces the right-hand side of (42) by

$$e^{-\phi^* D(U)\phi + \phi^* D(U)\phi}.$$ \hspace{1cm} (43)

The average in the exponent is evaluated by a Monte Carlo procedure with respect to $\phi$ and this is used for the Metropolis update for $U$. This algorithm requires $O(N_{pf} \cdot V)$ operations per sweep, where $N_{pf}$ is the number of sweeps for the update of $\phi$. One usually reduces this to $O(N_{pf} \cdot V)$ by restricting the $\phi$ update to one per sweep of the link variable $U$.

The pseudofermion algorithm has several potential sources of systematic errors. Firstly making only one update of $\phi$ per sweep introduces violation of detailed balance. Secondly one cannot take $N_{pf}$ too large ($O(10 - 100)$ in practice) and so the estimate of $<\phi^* D(U)\phi>$ suffers from statistical errors, especially for small quark mass. Thirdly there is a definite bias due to truncation of the terms of order $O(\Delta S_{pf})$ and higher. This comes from the Jensen's inequality

$$e^{\Delta S_{pf}} < e^{-\Delta S_{pf}} < 1,$$ \hspace{1cm} (44)

which shows that the pseudo-fermion algorithm underestimate the effect of dynamical quark loops. The problem can be cured only at the cost of reverting to the $O(V^2)$ dependence, and this does not appear very practical. The systematic underestimate has been actually seen in several pseudofermion simulations. To reduce its magnitude one has to use a smaller change in the trial configuration, which means a slower relaxation rate. In the end one is forced to carry out an extrapolation in the magnitude of $\Delta S_{pf}$. The lack of good control over this procedure and the other uncertainties mentioned makes this algorithm much less appealing than those we have discussed.
4. Hadron mass spectrum.

4.1 Preliminary.

The standard tool for extracting the hadron mass in lattice QCD is the propagator of hadrons. Consider an operator \( \mathcal{O}_H(n) \) constructed from quark and gauge fields and carrying the quantum number of the hadron in question. Since one is interested in the mass, one takes the projection onto zero spatial momentum \( \mathcal{O}_H(t) = \sum_n \mathcal{O}(n, t) \). The propagator can be rewritten in two ways, either by inserting a complete set of intermediate states or by representing it in terms of path integral;

\[
\mathcal{G}_H(t) = \sum_n \langle \mathcal{O}(0)|\mathcal{O}(t)\rangle = \sum_n A_n e^{-m_n \epsilon_{n, t} t} + \int_{m_n}^\infty \exp \left( -m \right) e^{-m \epsilon_{n, t} t}
\]

The third line represents the propagator as a sum of contributions from the stable particles of mass \( m_H \) below the multiparticle threshold and the continuum including resonances. This shows that the exponential slope of the propagator as a function of the temporal separation \( t \) gives the hadron mass in lattice units. The last line provides a way for calculating the propagator in practice. For numerical purposes, one goes one more step and rewrite it as

\[
\frac{1}{2} \sum_n \langle \mathcal{O}(0)|\mathcal{O}(t)\rangle = e^{m \epsilon_{n, t} t}
\]

where \( \epsilon_{n, t} \) denotes the average over the quark fields. The quantity \( \mathcal{G}_H(t, 0; U) = \sum_n \langle \mathcal{O}(0)|\mathcal{O}(t)\rangle \) represents the hadron propagator on a background gauge configuration \( U \). This is constructed out of the quark propagator \( D_q^{-1}(n, 0; U) \) has a generic form \( \sim D_q^{-1}(n, 0; D_q^{-1}(0, n)) \) for mesons and \( \sim D_q^{-1}(n, 0; U) \) for baryons.

In actual simulations one then proceeds as follows: One first generates a set of independent gauge configurations using one's favorite algorithm. On each configuration one solves the quark propagator \( D_q(U) \) to \( \sim \delta_{n, 0} (D_q^{-1}(0, n)) \), which is obtained automatically as \( \Gamma D_q^{-1}(n, 0; U) \) with \( \Gamma \) given in Sec. 3.2. From this one constructs the hadron propagator \( \mathcal{G}_H(U) \), and then takes an average over the ensemble to produce \( \mathcal{G}_H(t) \). One finally fits the large time asymptotics to the form \( \mathcal{G}_H(t) = A e^{-m \epsilon_{n, t} t} \) to extract the ground state hadron mass in lattice units, and sometimes also the masses of the excited states by including several such terms in one's fitting.

The first practical question is how one constructs \( \mathcal{G}_H \) out of the quark fields. For the Wilson action with one quark field for each flavor, this is not difficult. One can, for example, take over the \( SU(6) \) construction and use, with \( C \) the charge conjugation matrix,

\[
\pi^+ = \bar{u} \gamma_5 d, \quad \bar{p}^+ = \bar{u} \gamma_5 d, \quad p = (u^C, \gamma_5 d)^u, \quad A^H = (u^C, \gamma_5 d)^u,
\]

for \( \pi, \rho, \omega, \phi, \rho, \phi, \) and delta, and similar expressions for other hadrons.

In practical simulation the temporal lattice size \( N_t \) is finite. Normally the time slices \( n_t = 0 \) and \( N_t \) are connected by a periodic boundary condition (either periodic or anti-periodic). The mesons can then propagate both in the positive and negative time directions, and therefore the fitting function takes the form

\[
\mathcal{G}_M(t) = A e^{-m \epsilon_{n, t} t} + A e^{-m \epsilon_{n, t} (N_t - t)}
\]

The baryon propagator has Dirac indices. The positive energy states can be projected out with \( (1 + \gamma_4)/2 \). The particle propagating in the negative time direction has the parity opposite to that in the positive time direction. One therefore uses a fit given by

\[
\mathcal{G}_B(t) \equiv \frac{1}{2} \sum_{n_t} \langle B(t)|\bar{B}(0)\rangle = A e^{-m \epsilon_{n, t} t} - A e^{-m \epsilon_{n, t} (N_t - t)}
\]

The construction of \( \mathcal{O}_H \) is much more involved in the Kogut-Susskind case\textsuperscript{49.109-113}. For simplicity let us restrict ourselves to the operators built out of \( \chi \) and \( \bar{\chi} \) at the same site \( n \). The only combination possible are then \( \bar{\chi}_n \chi_n \) for mesons and \( \epsilon^{abc} \bar{\chi}_a^{n+1} \chi_b^n \chi_c^n \) for baryons with \( a, b, c = 1, 2, 3 \) the color indices. The spin and flavor quantum numbers are coded into the site index of the field \( \chi_n \). Thus they should appear as a site dependent sign factor \( \sigma_n \) multiplying these operators. An easy way to find it is to start with the standard expression (\( \sim \bar{\chi} \gamma_4 \phi \) for mesons with an appropriate gamma matrices \( \Gamma \) and baryons and substitute the expansion [19]. Carrying out this exercise, one finds that the mesons with the same spin but opposite parity have the sign factor related by \( \sigma_{n+1} = -\sigma_n \). The dependence on \( n \) and the \( JPC \) of the corresponding mesons are

\[
\sigma_n = \left\{ \begin{array}{ll}
1 & 0^- (\pi) \quad 0^+ (\rho) \\
(-1)^{n+1} & 1^- (\omega) \quad 1^+ (A_1) \\
(-1)^{n+1} & 1^- (\rho) \quad 1^+ (A_1) \\
0 & 0^- (\pi) \quad 0^+ (\rho) \end{array} \right.
\]

The two possibilities for \( \pi \) and \( \rho \) with and without prime correspond to a different combination of \( SU(4) \) flavors. Whether the masses extracted from those two agree or not therefore provides a way to check the restoration of the \( SU(4) \) flavor symmetry expected in the continuum limit. The meson propagators and their fitting take the following form:

\[
\mathcal{G}_M(t) = \sum_n \sigma_n \langle D_q^{-1}(n, 0; D_q^{-1}(0, n)) \rangle_U = A e^{-m \epsilon_{n, t} t} + A e^{-m \epsilon_{n, t} (N_t - t)}
\]

for \( \pi, \rho, \omega, \phi, \rho, \phi, \) and delta, and similar expressions for other hadrons.
we have discussed above is not the optimal choice to obtain a large value of the coupling, the lattice spacing decreases significantly with the increase of the size of the system. The optimal choice is to use a smaller lattice spacing, which allows for a larger value of the coupling.

In the case of the Wilson action, the hopping parameter $K$ is related to the hopping parameter $K_{\text{Wilson}}$ as

$$K = \frac{K_{\text{Wilson}}}{\sqrt{2}}$$

This allows for a more accurate determination of the coupling constant $\beta$.

The second condition for a good hadron mass estimate is that the large enough temperature must be used. In this case, the temperature $T$ is related to the coupling constant $\beta$ as

$$\beta = \frac{1}{\pi T}$$

This ensures that the hadron mass is well-defined.

The third condition is that the ground state must be well-defined. This is achieved by using a large enough lattice spacing $a$.

In summary, the optimal choice for the parameters $\beta$, $T$, and $a$ is

- $\beta = \frac{1}{\pi T}$
- $T = 2T_{\text{zero}}$
- $a = \frac{1}{\pi T}$

where $T_{\text{zero}}$ is the zero temperature.

These conditions ensure that the hadron mass is well-defined and that the lattice spacing is large enough to obtain a large value of the coupling constant $\beta$.
QCD. This whole process should be repeated at a smaller value of the coupling constant (larger value of $\beta$) to check the stability of the results with respect to a decrease of the lattice spacing. This may be done either by examining how much the mass ratios or the predicted hadron masses change, or more stringently, by comparing with the asymptotic scaling prediction (9) of the renormalization group.

We have mentioned several times that the region of small quark mass suffers from the critical slowing down both in the matrix inversion and in the correlation among successive sweeps. For this reason no simulations have been carried out at the physical up and down quark mass of a few MeV. Thus the fitting (54-55) involves an extrapolation in $m_a$. This necessarily introduces an uncertainty. To avoid this it has become customary to plot the mass ratio $m_a/m_T$ against $m_a/m_T$ (Edinburgh plot)\textsuperscript{114} or $(m_a/m_T)^2$ (Rome plot)\textsuperscript{115}, and examine how the data behave as one moves toward the physical point $m_a/m_T = 0.18$. A smoother approach may be expected for the latter choice because of the different quark mass dependence between the pseudoscalar mesons and other hadrons (54-55).

4.2 Recent quenched calculations.

The attempt at hadron mass predictions from lattice QCD started in 1981\textsuperscript{111-117}. At that time, however, generating gauge configurations including dynamical quarks was not feasible. Thus one resorted to the "quenched approximation"\textsuperscript{113-117} which sets $\text{det} D_0(U) = 1$. Physically this means ignoring the sea quarks. Several arguments were offered to justify this procedure: The vacuum quark loops renormalize the gauge coupling constant. If this is the dominant effect of the sea quark, the effect will be absorbed into a physical scale (e.g., the string tension) and may not have a large effect in the physical predictions\textsuperscript{117}. Empirically the processes that involve sea quarks have a small magnitude for flavor non-singlet hadrons (OZI rule)\textsuperscript{111}. These, however, are a posteriori justifications in the sense that one would precisely like to understand whether these features are dynamical consequences of QCD. Thus we do not have a priori basis for the quenched approximation. At the same time, it is clear that a high quality data of this approximation is very important to examine how dynamical quarks manifest their effect in strong interactions. Before starting the discussion of full QCD hadron mass calculations, we will therefore try to summarize the status of quenched simulations.

In Table I we list the parameters of the recent large scale quenched simulations\textsuperscript{120,124-124} (For an extensive list of references for earlier literature, see ref. 125). This shows a few characteristic features worth keeping in mind: (i) The lattice spacing has reached the value $a \sim 0.08$ fm or $a^{-1} \sim 2.2 - 2.5$ GeV at the coupling $\beta \sim 6.0 - 6.3$. With the typical spatial size of $N_s = 16 - 18$ the current simulations can accommodate hadrons of a radius up to $N_s a/2 \sim 0.7$ fm. This may be large enough for mesons but probably not quite for baryons (We recall that the charge radius is experimentally $0.75$ fm for pion and $0.83$ fm for proton). (ii) The smallest quark mass explored is about $50 - 100$ MeV for the Wilson action and $\sim 20$ MeV for the Kogut-Susskind case. These are still an order of magnitude larger than

<table>
<thead>
<tr>
<th>ref.</th>
<th>$\beta$</th>
<th>$#$ conf.</th>
<th>$a^{-1}$ (GeV)</th>
<th>$m_a$ (MeV)</th>
<th>$m_q$</th>
<th>$m_a/m_T$</th>
<th>$K_a/K_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbot et al.</td>
<td>118</td>
<td>$16^3 \times 32$</td>
<td>6.9</td>
<td>KS</td>
<td>5/200</td>
<td>1.04(13)</td>
<td>17/0.61</td>
</tr>
<tr>
<td>Bowier et al.</td>
<td>120</td>
<td>$16^3 \times 16$</td>
<td>5.7</td>
<td>KS</td>
<td>8/448</td>
<td>0.77(8)</td>
<td>7.7/0.01</td>
</tr>
<tr>
<td>120</td>
<td>$16^3 \times 24$</td>
<td>6.9</td>
<td>KS</td>
<td>32/224</td>
<td>2.12(1)</td>
<td>21/0.01</td>
<td>0.43</td>
</tr>
<tr>
<td>121</td>
<td>$16^3 \times 24$</td>
<td>6.15</td>
<td>KS</td>
<td>32/176</td>
<td>2.51(25)</td>
<td>25/0.01</td>
<td>0.55</td>
</tr>
<tr>
<td>121</td>
<td>$16^3 \times 32$</td>
<td>6.30</td>
<td>KS</td>
<td>32/224</td>
<td>2.62(26)</td>
<td>24/0.01</td>
<td>0.74</td>
</tr>
<tr>
<td>Gupta et al.</td>
<td>123</td>
<td>$16^3 \times 42$</td>
<td>6.20</td>
<td>KS</td>
<td>36/250</td>
<td>2.5(5)</td>
<td>19/0.0075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wilson\textsuperscript{125}</td>
<td>2.5(2)</td>
<td>0/0.036</td>
</tr>
<tr>
<td>Hoh et al.</td>
<td>110</td>
<td>$16^3 \times 48$</td>
<td>2.4\textsuperscript{126}</td>
<td>Wilson</td>
<td>15/130</td>
<td>1.31(6)</td>
<td>71/0.039</td>
</tr>
<tr>
<td>Haas et al.</td>
<td>122</td>
<td>$16^3 \times 28$</td>
<td>6.0</td>
<td>Wilson</td>
<td>9/500</td>
<td>2.24(5)</td>
<td>101/0.045</td>
</tr>
<tr>
<td>Forrillard et al.</td>
<td>112</td>
<td>$24^3 \times 48$</td>
<td>6.3</td>
<td>Wilson\textsuperscript{126}</td>
<td>47/100</td>
<td>4.0(3)</td>
<td>32/0.0081</td>
</tr>
<tr>
<td>Yoshii et al.</td>
<td>124</td>
<td>$16^3 \times 48$</td>
<td>5.85</td>
<td>Wilson</td>
<td>15/130</td>
<td>1.43(4)\textsuperscript{127}</td>
<td>105/0.064</td>
</tr>
<tr>
<td>Ape</td>
<td>109</td>
<td>$12^3 \times 24$</td>
<td>5.7</td>
<td>Wilson</td>
<td>284/360</td>
<td>1.44(5)</td>
<td>105/0.064</td>
</tr>
<tr>
<td>109</td>
<td>$18^3 \times 32$</td>
<td>6.0</td>
<td>Wilson</td>
<td>104/200</td>
<td>2.39(8)</td>
<td>57/0.025</td>
<td>0.79</td>
</tr>
</tbody>
</table>

1) periodically extended from 16.\textsuperscript{15}
2) $\tau = 1/2$ Wilson action.
3) non-standard gauge action.
4) twice $\sqrt{3}$-blocked Wilson action.
5) our estimate from the data given by quadratic fit in $1/K$.\textsuperscript{15}}
the real values of up and down quark masses of a few MeV. Correspondingly the mass ratio $m_u/m_d$ is larger than about 0.5. This limitation is due to the critical slowing down and the associated increase of the statistical fluctuation in the hadron propagators.

Let us look at the data more closely. In fig. 1 we show the inverse lattice spacing $a^{-1}$ from table I determined from $m_\pi = 770$ MeV. The solid line is the asymptotic scaling curve

$$a^{-1} = \Lambda_L \left( \frac{h_0}{g} \right)^{\beta_{12} \beta_{12}} \exp\left( \frac{1}{2g_0^2} \right)$$  \hspace{1cm} (56)

with

$$\Lambda_L = 4.7(5) \text{ MeV}$$  \hspace{1cm} (57)

deduced in ref. 125 at $\beta = 6.0 - 6.2$ from a compilation of data up to June 1987 (refs. 118-120, 122, 123, and preliminary values of refs. 113, 121 were available). As already noted\textsuperscript{124}, the values from the Wilson action are not too far from the curve while those from the Kogut-Susskind action have a much larger slope at $\beta \sim 5.5 - 6.0$, apparently converging together at $\beta \sim 6.0 - 6.2$. The new data at $\beta = 5.7$ (ref. 109) and 5.85 (ref. 124) with the Wilson action are consistent with (57). At $\beta = 6.0$, the recent high statistics data of ref. 109 gives a slightly larger value

$$\Lambda_L = 5.4(2) \text{ MeV},$$  \hspace{1cm} (58)

perhaps indicating that $a^{-1}$ of the Wilson action increases somewhat faster than (56) at this range of $\beta$. The final values of the Edinburgh group\textsuperscript{121} for the KS action ($\beta = 6.15, 6.30$) are consistent with (57).

In ref. 125, a compilation of the string tension $\sigma$ was also made and an estimate

$$\sqrt{\sigma}/\Lambda_L = 90(17)$$  \hspace{1cm} (59)

was given for $\beta = 6.0 - 6.2$. Recent data\textsuperscript{124,137} obtained with the Polyakov line (43) at $\beta = 5.9 - 6.1$ agree, with a smaller error ($\pm 10$). At $\beta = 6.2$, ref. 127 reports a little smaller value $\sqrt{\sigma}/\Lambda_L = 82(2)$. Combined with the estimate of $\Lambda_L$ from $m_\pi$, these give $\sqrt{\sigma} = 90 \cdot \Lambda_L = 490$ MeV at $\beta = 6.0$ (with (58)) and $\sqrt{\sigma} = 82 \cdot \Lambda_L = 390$ MeV at $\beta = 6.2$ (with (57)) also to be compared with $\sqrt{\sigma} = 420$ MeV experimentally. These values and fig. 1 indicate that the current range of highest $\beta = 6.0 - 6.3$ is approaching but has not quite entered the region of scaling behavior. A similar trend has been seen in renormalization group studies\textsuperscript{128}.

Let us examine the question of chiral symmetry, starting with the Kogut-Susskind case. This action keeps the $U(1)$ subgroup of $SU(4)$ chiral symmetry. If the $U(1)$ symmetry is spontaneously broken it will manifest itself in a non-vanishing value of the order parameter $< \bar{\chi} \chi >$ and the behavior of the pion mass $(m_\pi a)^2 \propto m_\pi a$ as the quark mass $m_q a \to 0$. These features have been clearly observed\textsuperscript{118,129,131,132}. Knowing the ratio $m_\pi^2/m_\pi$ and $< \bar{\chi} \chi >$ leads to a determination of the pion decay constant via the PCAC relation,

$$f_\pi = \sqrt{\frac{m_\pi^2 < \bar{\chi} \chi >}{2}}$$  \hspace{1cm} (60)

Fig. 1

The inverse lattice spacing for quenched QCD determined from $m_\pi = 770$ MeV. The filled and open symbols correspond to the Wilson and Kogut-Susskind quark action, respectively. The solid line is the scaling prediction with $\Lambda_L = 4.7$ MeV. The data are taken from Barkai et al\textsuperscript{118} (open inverted triangle at $\beta = 6$), Bowler et al\textsuperscript{120,121} (square), Gupta et al\textsuperscript{132} (triangle), Haan et al\textsuperscript{137} (filled inverted triangle at $\beta = 6$), Yoshi\textsuperscript{134} (diamond), Ape\textsuperscript{139} (filled circle), Fujugita et al\textsuperscript{140,141,147} (inverted triangle at $\beta = 5.5, 5.6$) and Gottlieb et al\textsuperscript{146} (open circle at $\beta = 5.65$).
Using the value of $\alpha^{-1}$ from the rho meson mass has yielded $f_{\pi}(\text{MeV}) = 119(19) (\beta = 6.0)^{121}$, $94(14) (\beta = 6.15)^{121}$ and $89(11) (\beta = 6.2)^{122}$ as compared with the experimental value $f_{\pi} = 93$ MeV.

The restoration of the full SU(4) flavor symmetry has been checked by the masses for the two types of operators for $\pi$ and $\rho$. The data$^{118,120,121,122}$ show that the degeneracy of the masses are better than 5% level for $\beta \geq 6.0$. In particular the $\pi^*$ mass vanishes as $m_{\pi} \rightarrow 0$ together with the $\pi$ mass.

For simulations with the Wilson action, the pion mass has also been found to be well described by the formula $(m_{\pi})^2 \propto m_{\pi}a - 1/2(1/K - 1/K_e)$. To interpret this as a manifestation of the Nambu-Goldstone nature of $\pi$ requires a little care, since the chiral symmetry is explicitly broken by the Wilson mass term. The analysis of the bilinear $\langle q\bar{q} \rangle$ is also complicated because it receives ultraviolet divergent contributions of the same origin.

A systematic way for analyzing the chiral content is provided by the Ward identity briefly sketched in Sec. 2. This analysis implies in particular that the pion to vacuum matrix element of the isospin axial-current divergence $\nabla_{\mu} A_{\mu}^a$, should vanish simultaneously with the pion mass. This has been observed$^{123}$ at $\beta = 6.0$, suggesting that the chiral breaking is already fairly small at this value of $\beta$ with the lattice spacing of $\alpha^{-1} \sim 2.3$ GeV. Some attempts have also been made$^{123,124}$ to estimate the pseudoscalar decay constant. Since $\langle q\bar{q} \rangle$ is contaminated by the short distance divergences, the two point functions of axialvector current and pseudoscalar density have been used. With $m_{\pi}$ as input, one finds$^{129}$ $f_{\pi} = 117(25)$ MeV at $\beta = 6.0$.

We now come to the most problematical issue of the quenched simulation. In fig. 2 we plot the recent data for the ratio $m_N/m_\pi$ against $m_{\pi}^2/m_\rho^2$. We did not draw a curve of the phenomenological mass formula including hyperfine splitting since it has the undesirable feature that $m_N/m_\rho$ is linear in $m_{\pi}/m_\rho$ for small $m_{\pi}$ contrary to chiral symmetry. Looking at fig. 2, we notice that the majority of $m_N/m_\pi$ values lie in the range 1.5–1.8 and that the extrapolation to the physical point appears to fall around 1.4 which is 10-20% higher than the experimental value of 1.22. This trend is common to the Wilson and Kogut-Susskind actions whose values are in fact consistent. There are a variety of effects which might lie behind this discrepancy: (i) Finite spatial size effect. The current lattice size (~0.7 fm) may not be large enough for accommodating a nucleon. Some attempts have been made to check this point by changing the spatial boundary condition of the quark propagator without conclusive results. It will be worthwhile to examine this point further. (ii) Finite temporal size effect. The excited state above nucleon is the Roper resonance (1/2+, 1440 MeV). The nucleon mass data may be overestimated by a contamination from this state$^{118,124}$. (iii) Scaling violation. At the strong-coupling limit $\beta = 0$, approximate analytic calculations of hadron masses are possible$^{123,124,120,60}$. The $N/\rho$ mass ratio lies even higher than the data points in fig. 2 especially for the Wilson action. Actually it monotonically increases with decreasing $m_{\pi}/m_\rho$, reaching $m_N/m_\rho = 2.25$ (Wilson) and 1.78 (KS) at $m_{\pi}/m_\rho = 0.18$.
with the lattice spacing $a^{-1}(GeV) = 0.88$ and 0.43, respectively, from $m_s = 770$ MeV. The current data is improved in that they on the average show a tendency of decrease toward a smaller quark mass and that the Wilson and KS results are consistent. Since the present values of $\beta$ do not yet lie in the scaling regime, however, $m_N/m_0$ ratio may still be suffering from a remnant of such a strong-coupling behavior. (iv) Extrapolation in the quark mass. The existing data with $m_s/m_0 \geq 0.5$ require an extrapolation to the physical point. The rate of decrease of $m_N/m_0$ might accelerate as $m_s$ decreases. Out of these possibilities, (ii) and (iv) now appear less probable in view of the recent high statistics data of the Ape collaboration 106 (filled circles in fig. 2) which have small errors and have a good signal from a small range of $\ell$ due to their smeared hadron operator. If the value $m_N/m_0 \sim 1.4$ stands after further scrutiny, it signifies the importance of the sea quark at the level of 10--20%.

An interesting related question is the magnitude of the nucleon $\sigma$ term defined by

$$\sigma_N = m < N|\bar{u}u + \bar{d}d|N >,$$  \hspace{1cm} (61)

with $m = (m_u + m_d)/2$, which may alternatively be written as

$$\sigma_N = m \frac{\partial m_N}{\partial m}.$$  \hspace{1cm} (62)

We see whether the $N/p$ mass ratio decreases toward the physical value is directly related to the magnitude of the sigma term. Typical values from the quenched simulations using (62) are: $\sigma_N$ (MeV)$=17(4)$ at $\beta = 5.7$, $14(3)$ at $\beta = 6.0$, and $20(4)$ at $\beta = 6.3$. Two direct estimates of the matrix element (61) gave $15(4)$ MeV at $\beta = 6.0$ and $18(6)$ MeV at $\beta = 6.3$.

The experimental value of the sigma term is not completely settled. To leading order of the usual (3, 3) breaking of the flavor $SU(3)$, one can rewrite (61) as

$$\sigma_N = \frac{3}{2} \frac{m}{2m_s - m} \frac{(m_A + m_E) - 2m_N}{1 - y},$$  \hspace{1cm} (63)

with $y = 2 < N|\bar{s}s|N > / < N|\bar{u}u + \bar{d}d|N >$. Assuming $< N|\bar{s}s|N > = 0$, which is certainly true in the quenched approximation, and using $(m_s - m)/m = 2(m_u^2 - m_d^2)/m_s^2 = 24$ MeV from the $K-\pi$ mass difference leads to $\sigma_N = 27$ MeV as an estimate which ignores sea quarks. This appears not too far from the values from the quenched simulations. One cannot use (63) to estimate the physical value since the experimental number for $< N|\bar{s}s|N >$ is not known. Alternatively, current algebra relates (61) to the $\pi-N$ scattering amplitude at a soft pion point, which can be estimated by dispersion techniques. A recent analysis which includes the corrections of non-vanishing quark mass gave $\sigma_N = 56(2)$ MeV. This is one standard deviation smaller than a previous estimate of 64(8) MeV. It is still large compared to the quenched values (either the lattice result or the $SU(3)$ estimate (63) with $y = 0$). If both types of numbers are valid, this implies a surprisingly large contribution from the strange sea quark ($y \sim 0.5$ from (63)). However, the experimental value from the dispersion relation might decrease further due to a sensitive dependence on the $\pi-N$ scattering lengths which are yet to be determined precisely. The lattice estimates should also be improved. It remains to be seen how much effect the dynamical quarks has for the $\sigma$ term, and through it on the nucleon mass.

### 4.3 full QCD simulations

Prior to 1986 there were only sporadic attempts at full QCD calculation of the hadron mass spectrum. Due to the development of the efficient algorithms described in Sec. 2, we now have a number of works in this direction. We shall describe the important features of the dynamical quark loop effect that have come out of these recent attempts.

There are two effects one naturally expect from dynamical quarks. Firstly their vacuum polarization renormalizes the gauge coupling constant. The standard weak-coupling expansion gives its magnitude to be $\beta_{\gamma} = -N_f/24\pi^2 \cdot g^4 \log(ka)^2 + O(g^6)$, showing that the effective coupling decreases. For heavy quark, one can alternatively carry out the integration over the quark field by an expansion in $1/m_Q$, the lowest order term with the quark circulating around a plaquette leads to the shift $\beta \sim N_f/(m_Qa)^3$, which again shows the screening due to quark vacuum polarization. Another effect arises from the fact that the color electric string that binds quarks into hadrons can be split and joined by the creation and annihilation of the quark-antiquark pairs out of the vacuum. This will lead to the flattening of the static $\bar{q}q$ potential at large distances, and to the opening of the decay channels such as $\rho \to \pi\pi$ and $\Delta \to N\pi$ turning them into resonances.

In order to see to what extent the current simulations offer a chance of examining these questions, we show in Table I the list of parameters of the works done up to now on the full QCD hadron mass spectrum, in parallel with the table I for the quenched simulations. We see from this that the scale of simulations is not yet enough to examine the question of hadron decays; the quark mass is still large ($m_q > 20$ MeV for Kogut-Susskind and $> 40$ MeV for Wilson) and consequently the $\rho$ meson is lighter than two pions ($m_\rho/m_p > 0.5$).

There are a few exceptions which has this ratio less than 0.5. However, the non-vanishing relative momentum, which cannot be less than the minimum value $2\pi/m_Q$, substantially overwhelms the decrease. The lattice size is also not large enough to accommodate two pions. Thus the study of effects of decays should be left for future improvement of simulations.

Comparing Table II and I, we notice that the spatial lattice size in full QCD simulations is roughly half of those for the quenched calculations, and the lattice spacing is twice as coarse ($a \sim 0.2 f/m$ as compared to $0.1 f/m$). The latter is in fact the reason for the small value of the smallest quark mass in the two types of simulations. The quark mass in lattice units, which determines the number of steps in matrix inversion, is twice as large in full QCD. The factor of $O(50 - 100)$ that results from the reduction in the lattice volume and the increase of the quark mass in lattice units represents a typical cost in computer time of going from the quenched to full QCD simulations.

Let us now look into some of the details of the hadron mass data. For the Wilson action,
Table II. Full QCD hadron mass calculations.

The inverse lattice spacing is estimated from $m_s = 770$ MeV. The quark mass and the $m_q/m_s$ ratio represent the smallest value reached. For the Wilson action $m_{eq}$ is calculated from (14). The critical hopping parameter and the largest value of $K$ used are also listed ($K_c/K$). The last column shows the algorithm employed: pf=pseudo-fermion, L=Langevin, h=hybrid, HMC=hybrid Monte Carlo.

<table>
<thead>
<tr>
<th>ref.</th>
<th>$N_f$</th>
<th>$\beta$</th>
<th>$a^{-1}$ (GeV)</th>
<th>$m_q$(MeV)</th>
<th>$m_{eq}$</th>
<th>$m_{eq}/m_s$</th>
<th>($K_c/K$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson</td>
<td>50</td>
<td>2</td>
<td>$9^3 \times 18$(x2)</td>
<td>5.5</td>
<td>1.75(12)</td>
<td>44/0.026</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1613(2)/0.16]</td>
</tr>
<tr>
<td>Patel et al.</td>
<td>149</td>
<td>2</td>
<td>$8^3 \times 8$(x3)</td>
<td>5.3</td>
<td>1.43(2)</td>
<td>45/0.028</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1668(3)/0.167]</td>
</tr>
<tr>
<td>Ukawa</td>
<td>70</td>
<td>2</td>
<td>$6^3 \times 12$</td>
<td>5.0</td>
<td>0.99(6)</td>
<td>100/0.10</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.197(1)/0.18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.5</td>
<td>0.73(6)</td>
<td>100/0.1</td>
<td>0.71</td>
<td>HMC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.206(2)/0.186]</td>
</tr>
<tr>
<td>Nambu</td>
<td>102,142</td>
<td>3</td>
<td>$10^3 \times 10$(x3)</td>
<td>5.3</td>
<td>1.13(5)</td>
<td>7.0/0.0062</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1804(2)/0.19]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.4</td>
<td>2.60(20)</td>
<td>63/0.024</td>
<td>0.80</td>
<td>pf</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.1643(3)/0.163]</td>
</tr>
<tr>
<td>Kogut-Susskind</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Billoire et al.</td>
<td>141</td>
<td>2</td>
<td>$8^3 \times 12$(x2)</td>
<td>5.4</td>
<td>0.66(1)</td>
<td>66/0.1</td>
<td>0.58</td>
</tr>
<tr>
<td>Fukagita et al.</td>
<td>143</td>
<td>2</td>
<td>$8^3 \times 18$</td>
<td>5.5</td>
<td>0.89(7)</td>
<td>42/0.05</td>
<td>0.57</td>
</tr>
<tr>
<td>Gottlieb et al.</td>
<td>146</td>
<td>2</td>
<td>$8^3 \times 24$</td>
<td>5.2875</td>
<td>0.57(2)</td>
<td>14/0.025</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$-5.375$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$10^3 \times 24$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.4375</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$-5.525$</td>
</tr>
<tr>
<td>Fucito et al.</td>
<td>139</td>
<td>3</td>
<td>$12^3 \times 8$(x3)</td>
<td>5.4</td>
<td>0.82(11)</td>
<td>41/0.05</td>
<td>0.57</td>
</tr>
<tr>
<td>Camponesti et al.</td>
<td>101</td>
<td>3</td>
<td>$10^3 \times 32$</td>
<td>5.7</td>
<td>1.33(2)</td>
<td>67/0.05</td>
<td>0.76</td>
</tr>
<tr>
<td>Kolke et al.</td>
<td>147</td>
<td>3</td>
<td>$10^3 \times 20$</td>
<td>5.6</td>
<td>1.21(6)</td>
<td>81/0.05</td>
<td>0.72</td>
</tr>
<tr>
<td>Grady et al.</td>
<td>144</td>
<td>4</td>
<td>$8^3 \times 16$(x2)</td>
<td>5.2</td>
<td>1.04(7)</td>
<td>26/0.025</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.3</td>
</tr>
<tr>
<td>Horn et al.</td>
<td>148</td>
<td>4</td>
<td>$12^3 \times 24$</td>
<td>5.2</td>
<td>0.80(2)</td>
<td>8.0/0.61</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.35</td>
</tr>
</tbody>
</table>

The masses in lattice units are largely reduced$^{88}$ from the quenched ones at the same value of $\beta$. The amount of decrease increases with the number of dynamical quark flavors $N_f$. Consequently, the critical hopping parameter $K_c(\beta)$ becomes smaller with increasing $N_f$. The pion mass still follows the PCAC-like relation (54) with respect to $K$ quite well, while the other masses exhibit a trend of bending down toward the critical hopping parameter$^{50,142,149}$. The decrease of hadron masses in lattice units have also been seen$^{141,146,148,147}$ with the Kogut-Susskind action except for the pion. As a function of $m_{eq}$, the pion mass exhibits the Nambu-Goldstone nature $(m_{eq})^2 \propto m_{eq}$, and the quark bilinear $\langle q\bar{q} \rangle$ extrapolates to a non-vanishing value, showing that the chiral $SU(3)$ is spontaneously broken also in full QCD. The pion mass, however, hardly moves from corresponding value in the quenched approximation. This appears to be due to the spontaneous breakdown of the chiral symmetry. For small $m_{eq}$ this forces the behavior $(m_{eq})^2 \propto m_{eq}$ in both the quenched and full QCD. For large $m_{eq}$, one expects little difference between the two. Tied up at both ends of the $m_{eq}$ range it is not very surprising that they roughly coincide. There is one feature where the current full QCD does poorly than the quenched approximation$^{146-148}$. The mass of $\pi'$ appears to stay large with decreasing $m_{eq}$ indicating a substantial violation of the flavor $SU(3)$ symmetry. The $\rho'$ is quite degenerate with $\rho$, however. The origin of the problem is most likely the coarseness of the lattice employed in the current full QCD simulations.

In fig. 3 we transcribe the inverse lattice spacing $a^{-1}$ determined from the $\rho$ meson mass in table II in parallel with fig. 1 for the quenched case. Reflecting the decrease of the hadron masses in lattice units, $a^{-1}$ is generally larger. For $N_f = 2$ a comparison of the Wilson and Kogut-Susskind actions is possible. Just as in the quenched case with $\beta = 7.5 - 6.0$, the Kogut-Susskind values are lower than the Wilson ones and rapidly increasing, while the latter roughly follows the scaling curve over $\beta = 5.0 - 5.5$. With this similarity in mind extrapolating to higher $\beta$, we may estimate from fig. 3 that the lattice spacing of the two actions converge at around $\beta = 5.7$ with $\Lambda_f(\beta) = \Lambda_f(\beta) = 3$ MeV. This corresponds to $a^{-1} \sim 2.3$ GeV, quite similar to the quenched case. We note that $\Lambda_{KSS}$ is also similar; $\Lambda_{KSS} \approx 130$ MeV for $N_f = 2$ as compared to the quenched value $135 - 135$ MeV with (57-58) where use was made of (11). We finally comment that the increase of the scaling slope $1/\theta = 16\pi^2/(1 - 2N_f/3)$ expected in the presence of dynamical quarks is not very dramatic for $N_f = 2 - 4$ and not yet seen clearly.

How much does the physical predictions of full QCD differ from the quenched ones? We now examine one aspect of this question, the effect due to quark vacuum polarization. The first step for this purpose is to compare the hadron masses from the full and quenched QCD at the same value of the coupling $\beta$. We have already mentioned that in lattice units they decrease by a considerable amount. This is a trend expected; the hadron masses in lattice units become smaller as one moves toward the continuum limit $\beta \to \infty$, but the quark vacuum polarization does exactly this by increasing the effective value of $\beta$.

The magnitude of renormalization is generally scale dependent. Whether their effect
The inverse lattice spacing for full QCD determined from $m_\rho = 770$ MeV. The filled and open circles correspond to the Wilson and Kogut-Susskind quark action, respectively. The solid line is the scaling prediction with $\Lambda_L = 3.0$ MeV for $N_f = 2$. The dashed line is copied from fig. 1 for the quenched case. $N_f = 2$; Fukugita et al.\textsuperscript{85,145,76} (circle at $\beta = 5.5, 5.0$), Patel et al.\textsuperscript{149} (filled triangle) and Gottlieb et al.\textsuperscript{144} (open square). $N_f = 3$; Hambler\textsuperscript{131,142} (filled square), Fucito et al.\textsuperscript{151} (open triangle at $\beta = 5.4$), Camposrini et al.\textsuperscript{159} (open triangle at $\beta = 5.7$) and Koike et al.\textsuperscript{147} (circle at $\beta = 5.8$). $N_f = 4$; Grady et al.\textsuperscript{144} (open inverted triangle) and Born\textsuperscript{149} et al. (diamond).

This quantity measures the degree of randomness of the gauge field fluctuation at the scale of the loop $C$. As a function of $\beta$, the Wilson loop expectation value monotonically increases from a small value at strong-coupling $\beta \sim 0$, where the fluctuations are large, toward unity in the weak-coupling regime $\beta \rightarrow \infty$, where they become more and more ordered around $U = 1$. The dynamical quarks effectively increase the value of $\beta$. The full QCD Wilson loop therefore lies above the pure gauge value as a function of $\beta$. Given a full QCD Wilson loop at a coupling $\beta$ and a quark mass, one can then search for the value of the coupling $\beta_{eff}$ in the pure gauge theory which has the Wilson loop matching that of full QCD. The difference $\Delta \beta = \beta_{eff} - \beta$ gives an estimate of the renormalization due to vacuum quark loops at the scale of the loop $C$.

Analyses have been made both for the Wilson and Kogut-Susskind cases (see also ref. 150) and have revealed two interesting features: (i) The magnitude of renormalization increases with the decreasing quark mass, and is generally larger for a larger loop, as expected. (ii) Perhaps surprisingly, the scatter depending on the scale is relatively small compared to the magnitude of the overall shift. One can imagine rewriting the quark determinant $Tr \log D_q(U)$ as a sum of Wilson loops of various sizes. Together with the gauge action $S_g(U)$, this allows a representation of full QCD in terms of a pure gauge theory with, however, a complicated action whose coupling constants have a specific dependence on the loop size. The point (ii) above says in essence that the coefficient of the plaquette term in $Tr \log D_q(U)$ is large compared to those of larger loops. This suggests that the bulk of the effect due to dynamical quark loop may be absorbed by shifting the coupling constant $\beta$; the pure gauge theory at the shifted coupling will reproduce the results of full QCD fairly well. This has been tested for hadron masses in a variety of simulations\textsuperscript{86,101,141,142,147}. It has been found that the quenched estimates at the shifted coupling determined from the matching of large Wilson loops agree quite well with those of full QCD within the error (for example, see fig. 24 in ref. 50).

This raises an obvious question concerning the accuracy of matching and the range of parameters over which it holds. The current data is limited in several respects. The quark mass does not extend well below $\sim 50$ MeV and so relatively heavy. We note, however, that the amount of the shift $\Delta \beta$ observed far exceeds the value reachable by the $1/m_\rho$ expansion\textsuperscript{147,141}. Also the size of the Wilson loop examined reaches $\sim 0.5$ fm, not too small compared to the typical size of hadrons. More important limitation in the present context is the magnitude of error in the mass data. At the moment the error quoted for hadron masses...
is of the order of 5–10% for small $m_\rho$. In view of the various limitations such as statistics and systematic errors, the actual uncertainty may well be 10–20%. A genuine physical difference at this level is allowed, and we recall that this is roughly the magnitude of difference being questioned in the discussion of the $N/\rho$ mass ratio in the quenched approximation. Thus the message one is receiving from current full QCD simulations on hadron mass is that the bulk of the effect of dynamical quarks goes into renormalizing the coupling constant with the net physical result coinciding with the quenched approximation within 10–20%. This may be the reason underlying the success of the "valence" approximation to the hadron mass spectrum.

What is an intuitive picture of the matching and where will it break down? A hint on this question is provided by analyses of the static potential between heavy quarks. The central piece of the potential may be extracted from rectangular Wilson loops of size $r \times t$ via the formula,

$$< W[r \times t] > \sim \exp(-t \cdot V_q(r)), \quad t \to \infty,$$

which is sensitive to the scale-dependent change of the fluctuations. Analyses have shown that the quenched potential, after the shift of the coupling constant, agrees remarkably well with the full QCD potential in the inner region of the potential, and that there are some indications that the latter may be flattening off in the outer region. A similar agreement in the inner region has also been found for the spin-spin and spin-orbit part of the potential, though with a large statistical error. The ground state hadrons have wave function peaked at the origin of the potential. It thus appears natural that their masses do not show much difference after the inclusion of the sea quark contributions. At the same time, we might expect a larger difference for the excited states whose wave function extends over a larger distance and therefore feels the effect of flattening more strongly. Finally there is the question of how the hadronic decays, for which the large distance behavior is certainly relevant, might modify the current understanding, and this is quite open at the moment.

An important issue with the quenched approximation was the value of the $N/\rho$ mass ratio summarized in fig. 2 and the related question of the nucleon $\sigma$ term. As for the $\sigma$ term, full QCD do not show any improvement so far. For example, we estimate with (62) that $\sigma_N = 12$ MeV at $\beta = 5.3^{147}$ and 5.5$^{148}$, both with the Wilson quark action. In fig. 4 we give a summary of $m_N/m_\rho$ for full QCD. Comparing with fig. 2, one sees a disturbing trend that full QCD values are worse than the quenched results; the $m_N/m_\rho$ ratio follows the band of 1.5–1.7, higher than the value 1.5–1.6 for the latter, and extrapolates to $m_N/m_\rho \approx 1.6$. (The two left-most points are below the $\rho \to 2\pi$ threshold. It is not clear whether current simulations with a limited spatial volume provide reliable results in this region.) To understand this, we note that the data for full QCD is dominated by the Kogut-Susskind results with the lattice spacing of $a^{-1} = 1$ GeV, while the quenched data in fig. 2 have $a^{-1} \approx 2.1 - 2.5$ GeV (see table I and II). We also recall that $m_N/m_\rho$ at the strong-coupling limit reaches 1.76 at $m_\sigma/m_\rho = 0.18$ for the Kogut-Susskind action. It is most
likely, then, that a larger $N/p$ mass ratio for current full QCD estimates is a finite lattice spacing effect. In order to see whether the dynamical quarks decrease $m_N/m_p$ below the quenched value, one should halve the lattice spacing of the full QCD simulations with the Kogut-Susskind action and consequently double the lattice size. The numbers are less severe for the Wilson case with $a^{-1} \sim 1.4 - 1.8$ GeV at the moment. The simulation is more time and memory consuming, however, especially for small quark masses. To detect a meaningful difference, it is also important to reduce the error for the mass ratio to the level of a few % (10% at the moment) in both quenched and full QCD simulations. Each of these requires a substantial improvement of the computing power and/or efficiency of the algorithm. There is clearly a long way to go.

4.4. The $U(1)$ Problem.

We close the discussion on hadron mass with the status of the $U(1)$ problem. For simplicity let us consider the case of two flavors. The propagator of the flavor non-singlet $\pi$ and the flavor singlet $\eta$ is given, respectively, by

$$G_{\pi}(n) = G_1(n),$$

$$G_{\eta}(n) = G_1(n) - G_2(n),$$

where

$$G_1(n) = < \langle \gamma_5 D_q^{-1}(n,0) \gamma_5 D_q^{-1}(0,n) \rangle > ,$$

$$G_2(n) = < \langle \gamma_5 D_q^{-1}(n,0) \gamma_5 D_q^{-1}(0,n) \rangle > .$$

Clearly a large mass splitting between $\pi$ and $\eta$ is possible only if there is a substantial cancellation between $G_1$ and $G_2$ in the $\eta$ propagator. This has been studied with the Wilson quark action. The result shows that $G_2$ is negligibly small ($G_2(0)/G_1(0) < 10^{-4}$) for the range of quark mass down to $\sim 50$MeV, and that there is no improvement from dynamical quark loops over this range. It has, however, been found in the quenched approximation that $G_2$ becomes comparable to $G_1$ if the quark mass is reduced to the realistically small value of a few MeV. This enhancement arises from the set of gauge configurations with a non-trivial instanton number because the Wilson quark operator $D_q$ turned out to have zero modes very close to the critical hopping parameter $(\tilde{\kappa} = 0.0001)$ on such gauge configurations. This seems to tie up nicely with the continuum argument for the resolution of the $U(1)$ problem. It is important to check whether the enhancement remains sufficiently large for full QCD where zero modes lead to a suppression of the corresponding gauge configurations. Unfortunately full QCD simulation at the quark mass of a few MeV is not feasible at the moment.

For some attempts on the related question of the glueball-$\bar{q}$-q mixing, see refs. 153,155.

5. Thermodynamics of QCD.

5.1 Theoretical expectations at finite temperatures.

There are several theoretical arguments on the phase diagram of QCD at finite temperatures which serve as a guide in more quantitative analyses based on numerical simulations. These will be summarized here.

We begin with the setup for studying finite temperature behavior of lattice QCD. The Euclidean formulation of finite temperature field theory is well known. One works on a space-time whose temporal extent is finite, imposing the periodic and anti-periodic boundary conditions on bosons and fermions, respectively, in the time direction. The Feynman path integral is then equal to the canonical partition function at a temperature $T$ which is the inverse of the temporal extent of the space-time.

Transcribing this procedure onto a lattice is quite straightforward. One simply follows the usual formulation of Sec. 2, but chooses the temporal lattice size $N_t$ to be finite and uses the periodic and anti-periodic boundary condition for the gauge and quark fields, respectively, to connect the $t = 0$ and $t = N_t$ planes. In principle the lattice spacing in the time and space directions may be different, and this is sometimes useful for theoretical purposes. In practice one takes them to have a common value $a$. The spatial lattice, often chosen to be symmetric $N_x \times N_y \times N_z$, has to be large compared to $N_t$ in this case. The physical temperature $T$ and the spatial volume $V$ are then given by

$$T = \frac{1}{N_t a}, \quad V = (N_x a)^3,$$

and the canonical partition is expressed as

$$Z = \text{Tr} (e^{-H_{QCD}/T}) = \int dU_{n,p} \prod_n dq_a dq_{\bar{a}} e^{S_1 + S_2},$$

with $H_{QCD}$ the physical Hamiltonian of QCD. Since $N_t$ is an integer, it may appear from (70) that the temperature can be varied only discretely. We recall, however, that the continuum limit of QCD is defined as $a \to 0$ and $\beta \to \infty$. Thus for a fixed $N_t$, increasing $\beta$ corresponds to increasing the temperature. We also note the additional condition that $N_t$ should be increased inversely proportional to $a$ to keep the temperature finite.

The dynamics of QCD at zero temperature is characterized by confinement and spontaneous breakdown of chiral symmetry. To study how finite temperature might modify these properties, one needs to know the physical quantities that characterize them. Let us start with the confinement property. Consider a static quark at a spatial site $n$. Looking at the quark action (12) or (17), one sees that its interaction with the gauge field is given by the Polyakov line:

$$\Omega_n = \frac{1}{3} \text{Tr} \left( \prod_{t=1}^{N_t} U_{n,t}(1) \right).$$
Inserting the Polyakov line in (71) corresponds to projecting out the states having a color charge of a quark at site $n$ in the thermodynamic average. This means that

$$< \Omega_n > = e^{-F_n/T},$$

with $F_n$ the free energy of a static quark.

The confinement property of the pure gauge theory is controlled by a $Z(3)$ symmetry\cite{114,115}. This is the group of elements $\{1, e^{2\pi i/3}, e^{4\pi i/3}\}$ forming the center of the gauge group $SU(3)$. The gauge action (1) or generalization of it with more types of Wilson loops is clearly invariant under a global transformation,

$$U_{n\xi} \rightarrow \zeta \cdot U_{n\xi}, \quad \zeta \in Z(3),$$

(74)

executed for the sites on a given time slice. The Polyakov line, on the other hand, transforms as

$$\Omega_n \rightarrow \zeta \cdot \Omega_n,$$

(75)

and is therefore an order parameter of the $Z(3)$ symmetry.Being an order parameter of a global symmetry, the expectation value $< \Omega >$ may be zero or finite depending on whether the symmetry is intact or spontaneously broken. In the former case the free energy of a static quark $F_n = - T \log < \Omega > = + \infty$ and we have confinement. In the latter case, the free energy becomes finite and quarks are liberated. The deconfinement in the pure gauge sector is characterized by a spontaneous breakdown of the center $Z(3)$ symmetry.

An alternative way to see this is provided by the correlation function of the Polyakov line $< \Omega_n \Omega_m^*$ which is related with the free energy $F_{\text{eff}}$ of a static quark-antiquark pair located at $n$ and $m$ by

$$e^{-F_{\text{eff}}/T} = < \Omega_n \Omega_m^* >.$$

(76)

For a large separation $r = |n - m| \rightarrow \infty$, one expects from cluster decomposition that

$$< \Omega_n \Omega_m^* > \sim < \Omega_n > < \Omega_m^* > + O(e^{-r/T}).$$

(77)

Combining the two we have $F_{\text{eff}}(r) \sim r$ if $< \Omega > > 0$, which is the usual confining potential, whereas if $Z(3)$ is broken, we have $F_{\text{eff}}(r) \sim \text{constant} + O(e^{-r/T})$ which does not confine.

There exists a rigorous proof\cite{116} that the $Z(3)$ symmetry is spontaneously broken for sufficiently large $\beta$ for any fixed $N_c$. For small $\beta$, on the other hand, the strong-coupling power series expansion in $\beta$ with a finite radius of convergence shows that $< \Omega > = 0$. Since $\beta \rightarrow \infty$ corresponds to increasing the temperature, the pure gauge theory should undergo a phase transition from the low temperature confined phase to the high temperature deconfined phase.

An interesting question is the order of this transition. Imagine constructing a 3-dimensional effective theory of the Polyakov line $\Omega_n$ by integrating out all the other gauge degrees of freedom. The effective action should have the property that (i) it is $Z(3)$ symmetric under (75), and (ii) short-ranged. The latter comes from the fact that the dynamics of the spatial components $U_{n,i} = x, y, z$, which is responsible for the interaction of $\Omega_n$, is believed to have a mass gap, being that of the $SU(3)$ gauge theory in 3 dimensions. The effective theory is therefore a $Z(3)$ symmetric spin system with the complex spin $\Omega_n$. The $Z(3)$ symmetry allows a cubic term $\Omega^3 + h.c.$ in the Landau free energy. In mean field approximation this leads to a second minimum away from $\Omega = 0$ and hence to a first order phase transition. Furthermore no second order fixed points are known for such a system in 3 dimensions. One thus expects the deconfinement transition to be first order\cite{117,118}.

The dynamical quarks introduces a modification in this picture\cite{119,120}. Looking at the temporal hopping term $\sim g_\sigma U_{n\sigma} a_{n+1}^\dagger a_{n+1}^\sigma h.c.$ of the quark action, we note that no transformation of the quark fields can compensate the $Z(3)$ rotation of $U_{n\sigma}$ given by (74). Thus dynamical quarks explicitly break the center $Z(3)$ symmetry. Since the breaking arises from the kinetic term, its magnitude will be small for a heavy quark but will become larger as the quark mass decreases. If the pure gauge deconfinement transition is first order, it will weaken in strength but will remain first order for heavy quarks because a weak breaking they induce will not wash away the double well structure of the free energy. The distortion of the shape of the free energy will increase with decreasing quark mass, however, and this may eventually lead to a disappearance of the first order transition.

The qualitative arguments above can be illustrated nicely for small $\beta$ and large $m_\sigma$. Taking the Kogut-Susskind action for definiteness, we integrate out the spatial link variables by an expansion in $\beta$ and the quark fields by that in $1/m_\sigma$. To leading order this gives

$$Z = \int \prod_n dU_n e^{-H_{\text{eff}}/T_{\text{eff}}},$$

(78)

where

$$H_{\text{eff}} = - \sum_{n,i} \Omega_n \Omega_n^* - \hbar \sum_n \Omega_n + h.c.,$$

(79)

with

$$T_{\text{eff}} = \frac{1}{\beta} \frac{18}{9} \frac{1}{\beta}, \quad h = \frac{1}{3} \frac{18}{\beta m_\sigma^2}.$$

(80)

We see that a high physical temperature ($\beta \rightarrow \infty$) corresponds to low "temperature" $T_{\text{eff}} \rightarrow 0$ for the effective theory. The spontaneous breakdown of $Z(3)$ at high physical temperature is quite natural when viewed in this way. The effect due to quarks is also explicit in (79), taking the form of a symmetry breaking external field $h$ of increasing strength as the quark becomes lighter.

So far we have treated dynamical quarks as a perturbation to the pure gauge system. This is no longer possible when $m_\sigma$ tends to zero, and the discussions based on $Z(3)$ symmetry cease to be tenable. On the other hand, the chiral symmetry becomes a good symmetry. At zero temperature this symmetry is spontaneously broken. The question then is how this may be modified as the temperature is raised.
Examination of this problem directly based on the QCD Lagrangian is not straightforward, especially on a lattice because of the problem of controlling the chiral symmetry. In the continuum, however, a renormalization group analysis using an effective \(\sigma\) model for mesons provides a reasonable picture\(^{166}\) (see also ref. 164, 165). In this approach one works with the meson field \(M_{ij} \sim q_i, \bar{q}_j\) with \(i, j = 1, \ldots, N_f\) the flavor indices. Under the chiral rotation \(q_i \rightarrow V_{ij} q_i\) with \(V_{ij} \equiv \exp(-it \gamma^5)\) \(j \in SU(N_f)\), the meson field transform as \(M \rightarrow VM^\dagger\). The effective Lagrangian for mesons consistent with the symmetry is

\[
\mathcal{L}_{\text{eff}} = \text{tr}(\partial_\mu M^\dagger \partial_\mu M) + m^2 \text{tr}(M^\dagger M) + \frac{\sigma^2}{3} g_1 \text{tr}(M^\dagger M)^2 + \frac{\eta^2}{3} g_2 \text{tr}(M^\dagger M)^2 \tag{81}
\]

Standard mean field analyses predict that the chiral symmetry, spontaneously broken at low temperature, will be restored through a phase transition at high temperatures. The precise nature of the transition is controlled by the renormalization group fixed points in the \((g_1, g_2)\) plane in \(d = 3\) dimensions. This in turn is given by the zeroes of the \(\beta\) function. The well known \(\epsilon\)-expansion in \(4 - \epsilon\) dimensions gives\(^{168}\)

\[
\frac{d g_1}{d \mu} = -\epsilon g_1 + \frac{N_f^2 + 4}{3} g_1^2 + \frac{4N_f}{3} g_1 g_2 + g_2^2, \tag{82}
\]

\[
\frac{d g_2}{d \mu} = -\epsilon g_2 + 2g_1 g_2 + \frac{2N_f}{3} g_2^2.
\]

For \(N_f \leq \sqrt{3}\) this predicts an infrared stable fixed point, and hence a second order phase transition. On the other hand, there are no such fixed points of \(O(\epsilon)\) for the number of flavors \(N_f > \sqrt{3}\). Thus these flavors do not seem to have a second order phase transition. The renormalization group trajectory in such a case runs off and generally goes into the region of phase space where the classical potential in \((81)\) is unbounded from below. This is saved by the one-loop correction which contributes terms of the form \(\sim M^4 \log M\). This turns the shape of the potential into a double well type. One therefore expects a first order phase transition induced by fluctuations\(^{169}\) for \(N_f \geq 2\).

There is a slight complication\(^{169}\) for \(N_f \leq 2\) because of the necessity of adding a term \(\det M \sim O(M^{N_f})\) to \((81)\) to make it consistent with the \(U(1)\) anomaly. For \(N_f \leq 4\) this is an irrelevant operator, while for \(N_f = 3\) this cubic term by itself will drive a first order transition. Thus the conclusion remains unchanged for \(N_f \geq 3\). For \(N_f = 2\), on the other hand, this is a mass-like term, and if large enough, might turn the order from first to second. Finally, for \(N_f = 1\) this explicitly breaks the chiral symmetry which smooths out the second order transition predicted by \((82)\) independent of its magnitude. The determinantal interaction is typically induced by instantons. Its density is dependent on the temperature. Whether the \(N_f = 2\) case is first or second order may therefore depend on the temperature dependence of the instanton density.

A non-vanishing quark mass explicitly breaks the chiral symmetry. For first order cases \((N_f \geq 3\) and possibly \(2\)), this weakens the transition, and will wipe it out for sufficiently large \(m_q\). The second order transition for \(N_f = 1\) and \(2\) is always smoothed out however small the quark mass.

Whether these predictions apply on a lattice is a delicate question. The Kogut-Susskind action does have chiral symmetry, but it is a \(U(1)\) symmetry. There is not even this symmetry for the Wilson case. We expect, however, that the full \(SU(N_f)\) symmetry will be recovered in the continuum limit for both actions. Thus the \(\sigma\) model prediction provides a useful guide on what to expect in the lattice simulation studies. We also note that for some types of lattice quark actions a rigorous proof exists\(^{170}\) that the chiral symmetry is not spontaneously broken at sufficiently high temperatures. In particular this applies to the Kogut-Susskind case, guaranteeing that its \(U(1)\) chiral symmetry will be restored at a finite temperature.

To summarize the theoretical arguments presented above, one lead is to expect the following phase diagram on the \((T, m_q)\) plane, or equivalently, on the \((\beta, m_q, \epsilon)\) plane. For \(m_q = \infty\), there is a first order deconfining transition. This extends into a finite value of \(m_q\) with a weakening of its strength, and perhaps disappears for small enough quark mass. At \(m_q = 0\), on the other hand, one expects another phase transition restoring chiral symmetry. This may again be first order for \(N_f \geq 2\), and if so, extends into a region of finite \(m_q\).

To find out if this picture is correct and what happens in the middle of the \((T, m_q)\) plane, including the possible relation between the two phase transitions, is the task posed to lattice QCD simulations.

### 5.2 Pure gauge simulation results.

There exists a large number of numerical simulations for the finite temperature behavior of the pure gauge theory (see ref. 125 for an extensive list of references). The lattice size which started out\(^{168, 169}\) with \(N^2 \times N_t = 7^4 \times 3\) (with \(SU(2)\) gauge group) has been expanding, and has reached\(^{170-174}\) \((16 - 21)^3 \times 14\) for the largest. The objective of the study has shifted from a qualitative exploratory calculations of the early period to a precise numerical determination of the phase transition characteristics and the physics of the high temperature phase.

The large amount of data accumulated up to now supports the expectation that the pure gauge theory undergoes a deconfining phase transition which is first order. This is based on the observation of a sharp rise across the transition of physical quantities such as the Polyakov line and the internal energy density, and the signatures for coexistence of confining and deconfined phases close to the critical point. The evidence for the latter came from the flip-flop behavior in which the system jumps back and forth in the simulation time between the confining phase characterized by \(\Omega \approx 0\) and the deconfined phase having \(\Omega \neq 0\) with the phase angle of \(0.2 \pi/3\), or \(4\pi/3\). Related evidences come from the existence of separate peaks in the distribution of \(\Omega\) at these positions on the complex \(\Omega\) plane. The conclusion of the first order transition, however, has been challenged recently\(^{175}\). We will discuss this below.

In numerical simulations one changes the value of the coupling constant \(\beta\) to alter the
temperature. Suppose \( \beta_c \) is the critical coupling of deconfinement on a lattice of temporal size \( N_t \). The critical temperature is given by

\[
T_c = \frac{1}{N_t \sigma(\beta_c)}
\]

with \( \sigma(\beta) \) the lattice spacing at the critical coupling. The critical temperature is a physical quantity with a dimension of mass, which should have a well-defined value in the continuum limit \( \alpha \to 0, N_t \to \infty \). Thus it should become proportional to the \( \Lambda \) parameter (7) evaluated at \( \beta = \beta_c \) as the temporal size \( N_t \to \infty \). To check this scaling and therefore the existence of the continuum limit has been one of the motivations of recent large scale simulations. The result is shown in fig. 5 together with several full QCD results to be discussed in the next section.

The determination of \( \beta_c \) poses a practical problem. As one increases the temporal size \( N_t \) the value of the Polyakov line just above the transition becomes increasingly small. The finiteness of the spatial volume rounds off the transition and also shifts the critical coupling downwards. These latter effects become worse for larger \( N_t \) since it has practically been not feasible to increase the spatial size proportionately. This means that the signal for the critical point becomes less sharp, and one has to specify what one means by \( \beta_c \). The procedure used is to divide the sample of the Polyakov loop into the confined and deconfined subsamples by some criterion, and define \( \beta_c \) to be the value at which the fraction \( f_d \) of the deconfined subsample exceeds a given number \( r \). Now looking back at fig. 5 we see by how much a different choice of \( r = 0.5 \) or \( 0.75 \) gives rise to a change of the slope of \( T_c/\Lambda_{MS} \) with a restricted spatial volume used so far ((16 - 19) \times 10 to (16 - 21) \times 14). For a discussion on which choice of \( r \) might lead to a better estimate for the infinite volume value of \( \beta_c \), see ref. 172.

Allowing for a violation of the order of 10\%, fig. 5 indicates that the critical temperature approaches the asymptotic scaling for \( \beta = 6.1 - 6.5 \) corresponding to \( N_t = 10 - 14 \). If one uses the value of \( \beta_c \) over this range of \( N_t \) reported in ref. 172, one obtains

\[
T_c/\Lambda_L = 51(2), \quad T_c/\Lambda_{MS} = 1.79(7).
\]

(84)

Converting this into a physical value may be done in two ways: If we use the estimate for the string tension over a similar range of \( \beta \) given by (59) one has

\[
T_c = 0.56(10) \cdot \sqrt{\sigma} = 240(40)\text{MeV},
\]

(85)

with \( \sqrt{\sigma} = 420\text{MeV} \). Alternatively if one uses the value \( \Lambda_L = 4.7(5)\text{MeV} \) deduced from the \( \rho \) meson mass of the quenched approximation (see (57)) one obtains

\[
T_c = 240(30)\text{MeV}.
\]

(86)

Let us come back to the question of the order of the deconfining transition. From the asymptotic decay of the Polyakov line correlation function (77), one can define a correlation

\[
T_c/\Lambda_{MS}
\]

$T_c/\Lambda_{MS}$

\[ \begin{array}{cccc}
N_t &=& 0 & 2 \\
& 4 & 6 & 8 \\
& 12 & 14 \\
\end{array} \]

\[ \begin{array}{cccc}
3/4 & 1/2 & r=1/2 & r=3/4 \\
\end{array} \]

The critical temperature in units of \( \Lambda_{MS} \) as a function of \( N_t \). Triangles are for the pure gauge deconfinement, while circles and squares represent the chiral transition temperature in full QCD at zero quark mass. At \( N_t = 4 \) and \( 6 \), the spatial lattice is \( 8^3 \text{ and } 10^3 \), respectively. For the pure gauge case with \( N_t = 8 \), the triangle is from ref. 171, and the inverted one from ref. 174. For \( N_t \geq 10 \), the filled inverted triangles (ref. 174) use \( r=1/2 \) definition. The triangles are from ref. 172, with the open ones determined by \( r=1/2 \) and the filled ones by \( r=3/4 \).
and the order parameter exhibit a discontinuous behavior across the transition. A recent measurement on the critical point, however, has suggested that the correlation length at the critical point is finite, which is consistent with the finite size scaling behavior expected for a first-order transition. The critical point, however, remains elusive for Ising-type systems, as no exact critical point has been found to date.

The order parameter at the transition point exhibits a discontinuous behavior across the transition. A recent measurement on the critical point, however, has suggested that the correlation length at the critical point is finite, which is consistent with the finite size scaling behavior expected for a first-order transition. The critical point, however, remains elusive for Ising-type systems, as no exact critical point has been found to date.

The order parameter at the transition point exhibits a discontinuous behavior across the transition. A recent measurement on the critical point, however, has suggested that the correlation length at the critical point is finite, which is consistent with the finite size scaling behavior expected for a first-order transition. The critical point, however, remains elusive for Ising-type systems, as no exact critical point has been found to date.

The order parameter at the transition point exhibits a discontinuous behavior across the transition. A recent measurement on the critical point, however, has suggested that the correlation length at the critical point is finite, which is consistent with the finite size scaling behavior expected for a first-order transition. The critical point, however, remains elusive for Ising-type systems, as no exact critical point has been found to date.

The order parameter at the transition point exhibits a discontinuous behavior across the transition. A recent measurement on the critical point, however, has suggested that the correlation length at the critical point is finite, which is consistent with the finite size scaling behavior expected for a first-order transition. The critical point, however, remains elusive for Ising-type systems, as no exact critical point has been found to date.
Table III. Finite temperature simulation with Kogut-Susskind action.

For $N_f \geq 2$ an asterisk * attached to the quark mass value means a first order phase transition claimed. For $N_f = 1$ it signifies its absence concluded. For the $N_f = 2 + 1$ case, the quark mass listed is for up-down/strange. The estimates of the critical coupling $\beta_0$ are from the most recent publication listed last. The last column shows the algorithm employed; pf=pseudo-fermion, L=Langevin $[\Delta T_L]$, h=hybrid $[\Delta T]$). HMC=hybrid Monte Carlo $[\Delta T]$, mc=microcanonical $[\Delta T]$, exact=algorithm of ref. 87, 98. When not written, it is the same as the line above.

<table>
<thead>
<tr>
<th>$N_f = 1$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ref.</td>
<td>size</td>
<td>$m_{\pi \pi}$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Fukugita et al.</td>
<td>$8^3 \times 4$</td>
<td>0.05</td>
<td>5.475 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.51 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.57 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4</td>
<td>5.63 (1)</td>
</tr>
<tr>
<td>Gavai et al.</td>
<td>$4^3 \times 4$</td>
<td>0.00125</td>
<td>5.42-5.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025</td>
<td>5.44-5.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.46-5.50</td>
</tr>
<tr>
<td></td>
<td>$8^3 \times 4$</td>
<td>0.00125*</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025*</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.48</td>
</tr>
<tr>
<td>Irbbek et al.</td>
<td>$8^3 \times 4$</td>
<td>0.1</td>
<td>5.53 (1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N_f = 2$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ref.</td>
<td>size</td>
<td>$m_{\pi \pi}$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Gavai et al.</td>
<td>$6^3 \times 2$</td>
<td>0.1</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.15</td>
<td>4.6</td>
</tr>
<tr>
<td>Fukugita et al.</td>
<td>$4^3 \times 4$</td>
<td>0.1</td>
<td>5.32 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.32 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.32 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.45 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4</td>
<td>5.54 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8</td>
<td>5.63 (1)</td>
</tr>
<tr>
<td>Gottlieb et al.</td>
<td>$8^3 \times 4$</td>
<td>0.00125</td>
<td>5.375 (19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025</td>
<td>5.375 (19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.320 (19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.375 (20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.430 (30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.470 (40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.525 (40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N_f = 3$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ref.</td>
<td>size</td>
<td>$m_{\pi \pi}$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Fusco et al.</td>
<td>$6^3 \times 4$</td>
<td>0.1</td>
<td>5.30 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.35 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>5.30 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>5.60 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>5.65 (5)</td>
</tr>
<tr>
<td>Gavai-Karsch</td>
<td>$6^3 \times 4$</td>
<td>0.1</td>
<td>5.2455 (15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.25-5.30</td>
</tr>
<tr>
<td>Gavai et al.</td>
<td>$4^3 \times 4$</td>
<td>0.025</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025*</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.25 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.30 (5)</td>
</tr>
<tr>
<td>Gogu-Sinclair</td>
<td>$4^3 \times 4$</td>
<td>0.025*</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.25 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.30 (5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N_f = 4$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ref.</td>
<td>size</td>
<td>$m_{\pi \pi}$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Fischler et al.</td>
<td>$4^3 \times 4$</td>
<td>0</td>
<td>5.05-5.15</td>
</tr>
<tr>
<td>Podolsky et al.</td>
<td>$8^3 \times 4$</td>
<td>0.08</td>
<td>5.05-5.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.05-5.15</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>$N_f = 4$ (continued)</th>
<th>ref.</th>
<th>size</th>
<th>$m_{\pi}$</th>
<th>$\beta$</th>
<th>algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duane-Kogut 84, 85</td>
<td>$8^3 \times 4$</td>
<td>0.05</td>
<td>5.000-5.025</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.075</td>
<td>5.10-5.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.10-5.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kogut-Sinclair 201, 206</td>
<td>$10^3 \times 6$</td>
<td>0.05</td>
<td>5.15-5.175</td>
<td>h[0.02]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.065</td>
<td>5.20-5.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.20-5.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>5.45-5.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.17-5.26</td>
<td>mc</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.075</td>
<td>5.25-5.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.23-5.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fukugita et al. 203</td>
<td>$8^3 \times 4$</td>
<td>0.1*</td>
<td>5.1(1)</td>
<td>L[0.01/0.002]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.25(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>5.35(5)</td>
<td>L[0.01]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>5.5(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>5.7(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>$8^3 \times 4$</td>
<td>0.1*</td>
<td>5.1(6)</td>
<td>L[0.01]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.26(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05*</td>
<td>5.073(5)</td>
<td>L[0.01]</td>
<td></td>
</tr>
<tr>
<td>Gavai 204</td>
<td>$8^3 \times 4$</td>
<td>0.1</td>
<td>5.1-5.2</td>
<td>pf</td>
<td></td>
</tr>
<tr>
<td>Gupta et al. 205, 216</td>
<td>$4^3 \times 4$</td>
<td>0.025*</td>
<td>4.91(3)</td>
<td>exact</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05*</td>
<td>4.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1*</td>
<td>5.04(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gottlieb et al. 46, 211</td>
<td>$8^3 \times 4$</td>
<td>0.025*</td>
<td>4.96(3)</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>5.06(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>5.13(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>5.25(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4</td>
<td>5.40(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>5.45(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>213</td>
<td>$4^3 \times 4$</td>
<td>0.025*</td>
<td>4.92</td>
<td>h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025*</td>
<td>4.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kovacs et al. 207</td>
<td>$8^3 \times 4$</td>
<td>0.0125*</td>
<td>4.91(6)</td>
<td>h[0.005/0.01]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025*</td>
<td>4.94(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.035*</td>
<td>4.99(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10^3 \times 6</td>
<td>0.025*</td>
<td>5.125(35)</td>
<td>h[0.01]</td>
</tr>
<tr>
<td>Karsch et al. 208</td>
<td>$4^3 \times 4$</td>
<td>0.025*</td>
<td>4.94(4)</td>
<td>h[0.01]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025*</td>
<td>4.96(3)</td>
<td>h[0.01]</td>
<td></td>
</tr>
<tr>
<td>Gavai et al. 210</td>
<td>$4^3 \times 4$</td>
<td>0.025*</td>
<td>5.03</td>
<td>L[0.01]</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
smoothly continues. However, the weakening trend is reversed\textsuperscript{[203,205,211]}; the Polyakov line reverts to a sharp rise. At the same time the chiral order parameter $<\bar{\chi}\chi>$ exhibits an increasingly discontinuous decrease in $\beta$ at the point where $<\Omega>$ has a jump. This strongly indicates the restoration of chiral symmetry above the critical temperature. To confirm this one can follow the behavior of $<\bar{\chi}\chi>$ along the transition toward $m_q = 0$. On the high temperature side $<\bar{\chi}\chi>$ converges to a value consistent with zero, while it extrapolates to a non-vanishing value on the confined side\textsuperscript{[203,211,216]}.

We will call the transition for small $m_q$ a chiral transition since it is the chiral symmetry of the light dynamical quarks that is responsible. It is not a remnant of the pure gauge deconfinement because it can only weaken with a decreasing quark mass. We thus see that dynamical quarks play quite a different role in different regions of the phase diagram: For $m_q$ large they weaken the transition that already exists. For small $m_q$, on the contrary, they induce one. A priori the two transitions are unrelated and may occur at a totally different part of the phase diagram, even giving rise to two transitions for some range of $m_q$. This does not happen in reality. The two transitions together form a single continuous line that runs across the ($\beta, m_q\alpha$) plane for the entire range of the quark mass. One may wonder why the Polyakov line shows an abrupt jump even for small $m_q$ where one expects the $Z(3)$ symmetry to be badly broken. This should be related to the fact that the restoration of chiral symmetry removes much of the free energy of static quark dynamically generated by the spontaneous breakdown in the low temperature phase (recall (73)).

The important issue to settle is the order of the chiral transition since its effect will not stand out in astrophysical applications and heavy ion collisions unless it is first order. This used to be a controversial point\textsuperscript{[197,198,199,200-204]}. However, most of the recent works for $N_f = 2 - 4$ conclude\textsuperscript{[202,203,207-211,216-223]} that it is a first order phase transition for sufficiently small quark mass. (See table III. The quark mass values with an asterisk have been claimed to be first order.) This basis on the signatures for a coexistence of two phases at the critical point; the flip-flop behavior of observables and/or the two peak structure in their distribution. The lattice size studied ranges from $4^8 \times 4$ up to $12^3 \times 4$, and all the algorithms described in Sec. 2. have been used.

There are two points requiring further clarification before the question of the order can be considered settled. First of all different algorithms differ in the largest value of $m_q\alpha$ beyond which the signal becomes too unclear to conclude the presence of a first order transition (see table III). On an $N_f = 4$ lattice with $N_f = 4$, the hybrid simulations see first order signals only at $m_q/T < 0.15\textsuperscript{[207,208,211,213]}$. The Langevin algorithm reports clear evidence up to $m_q/T \approx 0.42$, and the hybrid Monte Carlo sees it as high as $m_q/T \approx 2.0\textsuperscript{[214,215,69]}$. The corresponding values for $N_f = 2$ are: $0.05\textsuperscript{[214]}$ (less clear sign at 0.4$\textsuperscript{215}$) for the hybrid, and 0.4 for Langevin\textsuperscript{[69]}. After correcting for systematic effects such as the step size and spatial volume dependence, the critical coupling $\beta_c$ obtained by different algorithms are in agreement within $0.5 - 1\% (\delta\beta_c \approx 0.02 - 0.04)\textsuperscript{[211,219]}$. The difference therefore appears more to do with the relaxation rate. We note that this is not merely an academic problem. It bears on the question whether the chiral phase transition ceases to be first order beyond a certain $m_q/T$ creating a gap of singularity in the phase diagram. In that case the low and high temperature phases can be continued analytically. In particular this may be used to argue for the existence of real excitation having hadronic quantum numbers in the high temperature phase\textsuperscript{[210]}. The second point concerns the spatial volume dependence. In many simulations the distribution of the observables has been found to exhibit a double peak structure close to the transition. Recent studies\textsuperscript{[215,216,219,221,227]} have found indications that the separation between the peaks shrinks and the valley in between becomes less deep as the spatial volume is increased from $4^4$ to $(10-12)^3$ at $N_f = 4$. This raises a suspicion that the two peak structure may disappear in the infinite volume limit with the consequence that the chiral transition is not first order. The shrinking of the peak separation is not inconsistent with a first order transition, however. Too see this one may invoke a phenomenological model\textsuperscript{[211,69]} which represents the distribution as a sum of two Gaussians corresponding to the two phases. The two peak positions are functions of $\beta$, and the width shrinks as $1/V$. The relative weight is a function of the volume and $\beta$ and converges to the step function at $\beta = \beta_c$ as $V \rightarrow \infty$. At a finite volume the $\beta$-dependence of observables are rounded off with a width of order $O(1/V)$. One may define the critical value $\beta_c(V)$ to be the point where the two Gaussians contribute with an equal weight 1/2. With increasing $V$ this point moves toward $\beta_c(\infty)$. In principle the peaks of the two Gaussians at $\beta = \beta_c(V)$ can equally well become closer or move further apart as $V \rightarrow \infty$. They will, however, eventually stop moving for a large enough volume. Clearly a detailed finite size scaling analysis is needed to make sure that this is what happens for the chiral transition in QCD. Some recent pure gauge simulations\textsuperscript{[217]} indicate that perhaps the lattice size of the range $16^3 - 24^3$ might be needed for the temporal size $N_t = 4$.

The question that comes next to that of order is the critical temperature $T_c$. To determine its value in physical units, one needs the lattice spacing at $\beta = \beta_c$. In full QCD the most natural way for this purpose is to make a hadron mass calculation at zero temperature at $\beta = \beta_c$. After a few preliminary works\textsuperscript{[208,211]}, a detailed calculation\textsuperscript{[213,146]} has been carried out for $N_f = 2$ at $N_t = 4$ and 6. The result after extrapolation to the physical quark mass is

$$T_c(\text{MeV}) = \left\{ \begin{array}{ll}
142(6) & \text{for } N_t = 4, \\
164(25) & \text{for } N_t = 6,
\end{array} \right. \text{ MeV}$$

where the first number for each $N_t$ is from $m_N = 770$ MeV and the second from $m_N = 940$ MeV. The disparity of order 30% between the two is a direct reflection of the problem of large $m_N/m_q$ discussed in Sec. 4.3.

The transition temperature (87) is smaller than the estimates (85-86) for the pure gauge case obtained on a larger lattice ($N_t = 10 - 14$). In order to check whether this is also true
for the same $N_t$, ref. 146 carried out a quenched hadron mass calculation at the critical coupling of the pure gauge deconfining transition for $N_f = 6$. The resulting estimate is $T_c(N_f = 0) = 234(30)$ MeV from $m_{ho}$. One should note, however, that these pure gauge estimates have a larger inverse lattice spacing compared to the full QCD in (87). To examine whether the chiral transition really takes place at a lower temperature, one has to consider the question of scaling. For this purpose, one can collect the reported values of $\beta_c$ for small quark masses (see table III) to make an extrapolation to $m_u = 0$, and from this compute the ratio $T_c/\Lambda_{QCD}$. We find

$$\beta_c = \begin{cases} 
5.67(1) & \text{for } N_f = 0 \text{ (ref. 125)}, \\
5.86(1) & \text{for } N_f = 2 \text{ (ref. 212, 213, 146)}, \\
5.24(1) & \text{for } N_f = 4 \text{ (ref. 206, 207, 211)} 
\end{cases}$$

and therefore

$$T_c/\Lambda_{QCD} = \begin{cases} 
2.55(3) & \text{for } N_f = 0 \\
1.94(8) & \text{for } N_f = 2 \\
2.14(10) & \text{for } N_f = 4 
\end{cases}$$

where the first and second numbers for each $N_f$ are for $N_t = 4$ ($N_t = 8$) and $N_t = 6$ ($N_t = 10$), respectively. These values are plotted in fig. 5. It is somewhat remarkable that the chiral transition temperature expressed in units of $\Lambda_{QCD}$ is quite similar for $N_f = 2$ and 4 for which the ratio $\Lambda_{L}/\Lambda_{QCD}$ differs by a factor 1.74 (see (11)). It is perhaps even more so with the pure gauge critical temperature which is completely different in its origin from the chiral transition.

The asymptotic scaling has clearly not set in for the chiral transition. If the similarity with the pure gauge case seen in fig. 5 is of some guide, we need at least $N_t = 8$ or perhaps 10 for its onset, with a reduction of $\sim 15 - 20\%$ of $T_c/\Lambda_{QCD}$ from $N_t = 6$. In order to see what this implies for the value of $T_c$ for $N_f = 2$, we recall the discussion in Sec. 4.3 that the lattice spacing for the Kogut-Susskind and Wilson actions with $N_f = 2$ might converge at $\beta \approx 5.7$ with $\Lambda_L(N_f = 2) \approx 3$ MeV. With a 20% reduction in $T_c/\Lambda_{QCD}$ from $N_t = 6$, this gives $T_c \approx 200$ MeV with the aid of (11). This is substantially higher than the current value (87) and closer to the pure gauge case (85-86). Although there is a large margin of uncertainty in our estimate above, it provides a warning that the today’s simulations with $a^{-1} \sim 1$ GeV might still involve a non-negligible finite lattice spacing effect. Clearly we need results from a higher $N_t$ to establish the critical temperature of the chiral transition.

The case of $2+1$ flavors. So far we have discussed the case of $N_f$ degenerate flavors. In reality only up and down quarks can be treated as equal but the strange quark is about 20 times heavier.222 A few simulations117,222 have attempted this $2+1$ flavor case on an $8^3 \times 4$ lattice introducing the bilinear noise for up-down and strange quarks separately with the weight 2/4 and 1/4, and choosing $m_{s}/m_{u} = m_{s}/m_{d} = 20 - 25$, $m_{s}/T = 0.05 - 0.1$. With $T_c \sim 140$ MeV as indicated from the $N_f = 2$ result for the same size of the lattice, this corresponds to $m_{s} = m_{d} \sim 10$ MeV and $m_{s} \sim 200$ MeV which is not too far from reality. The result shows that the strange quark does not have much effect on the transition itself compared to the $N_f = 2$ case; it is a first order transition with virtually the same critical temperature.

Unusual number of flavors. We now turn to the case of unusual number of flavors. Let us start with small $N_f \leq 1$. The interest in this case comes from the prediction of the $\sigma$ model that a phase transition will be absent for small $m_s$. One also expects that the transition will look more and more similar to the pure gauge deconfinement over a larger range of the quark mass except near $m_{s} = 0$ as the effect of dynamical quarks diminishes with decreasing $N_f$. The simulations were carried out for $N_f = 1$ (see Table III), 0.5 (ref. 225) and 0.1 (ref. 223) on an $8^3 \times 4$ lattice using the bilinear noise Langevin or R-hybrid algorithm. For $N_f \leq 1$ the transition remains essentially the same as the pure gauge transition down to the quark mass of $m_q/T \sim 1$ except for a downward shift of $\beta_c$ which, however, is allowable to the quark vacuum polarization effect already existing at zero temperature. The transition remains first order at $m_q/T = 0.4$ and 0.2. There are signs of a weakening in that the amount of jump of the Polyakov line becomes smaller and that the range of $\beta$ over which the system exhibits a flip-flop behavior is reduced. At $m_q/T \leq 0.1$ a Langevin simulation223 did not find signs of a first order transition.

The large number of flavors studied so far are $N_f = 8, 10, 12$, and 18 (see table III). Typical lattice size is again $8^3 \times 4$. The dependence of the transition on the quark mass is quite similar to the well-studied cases of $N_f = 2 - 4$. The pure gauge deconfining transition initially weakens with decreasing $m_s$. This stops at $m_q/T \sim 1$ below which the transition becomes stronger again. Contrary to $N_f = 2 - 4$ there is evidence209,11 that $N_f = 10$ that the transition remains first order for intermediate quark mass. This is probably explained by another observation by the large $N_f$ simulations that the chiral transition at a fixed $m_{s} = 0.1$ becomes stronger for a larger $N_f$. With a stronger first order chiral transition a larger quark mass is needed to smooth it out. But then the quark might be heavy enough to cause only a weak perturbation on the deconfining transition of the pure gauge theory.

To summarize the dependence of the chiral transition on $N_f$ the present status of full QCD simulations with the Kogut-Susskind quark action is quite consistent with the prediction of the sigma model: it is first order for $N_f \geq 2$ and there is some evidence that there are none for $N_f = 1$.

We add a comment on the large number of flavors. We have treated the transitions found in simulations as a finite temperature transition. There is, however, another possibility. Looking back at the $\beta$ function (5-8) of QCD at zero temperature we see the well known fact that the first coefficient $b_1$ changes sign at $N_f = 16.5$. We also note that the second coefficient $b_2$ becomes negative already at $N_f = 8.05 \ldots$ where $b_3$ is still positive. This raises the possibility223 that the $\beta$ function of QCD might develop new zeroes away from
the origin at an $N_f$ as small as 9 before the asymptotic freedom is lost at $N_f = 16.5$. If true this means the existence of novel phases of zero temperature QCD; confinement without asymptotic freedom and asymptotic freedom without confinement.

It is possible that the transitions one has seen for large $N_f$ actually belong to this category. A way to examine this point is to change the temperature via an increase of $N_f$, since a critical point of the zero temperature theory should not move with temperature. This was examined by two simulations. The conclusions differ: ref. 51 found that $\beta_1$ changes by an amount expected from scaling both for $N_f = 10$ and 12, consistent with a finite temperature transition, while ref. 218 claims to have found a zero temperature fixed point for $N_f = 12$. Further study appears to be needed to clarify the problem.

5.4 Full QCD phase diagram - Wilson quark action.

There are only a handful of studies\textsuperscript{232, 233, 145, 76} of finite temperature QCD with Wilson action for quarks. This is due to the lack of manifest chiral symmetry. We would like to emphasize, however, that the Wilson action deserves much more attention. Both the Wilson and Kogut-Susskind actions are to recover the full $SU(N_f)$ chiral symmetry in the continuum limit and to produce the same continuum physics, one should understand how this is achieved in spite of a considerable difference of their chiral structures at a finite lattice spacing. We also repeat that the manifest chiral symmetry of the Kogut-Susskind action is only $U(1)$, and that the flavor interpretation is not straightforward, especially for $N_f < 4$.

The phase diagram with the Kogut-Susskind action has two basic features. The deconfining transition becomes weaker as quarks become lighter, but toward $m_q \to 0$ it is replaced by the chiral transition which is first order. The chiral transition on the $(\beta, m_q)$ plane intersects the $m_q = 0$ axis at a finite value of $\beta$, and this value increases as one approaches the continuum limit $N_f \to \infty$. Let us translate this into Wilson language. The quark mass is given by \textsuperscript{(14);} $m_q = 1/(1/K - 1/K_c(\beta))$. At a naive level one may identify the critical line $K = K_c(\beta)$ with the chiral limit $m_q = 0$ of the Kogut-Susskind case. The results with the Kogut-Susskind action, translated this way, imply that the pure gauge transition weakens as the hopping parameter $K$ increases away from zero, but toward $K = K_c$ it becomes a first order chiral transition which intersects the critical line at a non-vanishing value of $\beta$.

The first part of this expectation for small $K$ has been confirmed\textsuperscript{232-235}. However, a problem arose when the line of transition was followed toward the critical line. It was found\textsuperscript{236} that the transition not only kept becoming more and more gradual but also did not seem to cross the critical line. In fact it appeared to run parallel reaching down to the strong-coupling limit where no sign of a first order transition was seen as a function of $K$.

It looked as if the transition is smoothed out beyond a certain value of $K$ and does not turn strong again. The original analysis\textsuperscript{236} on a small lattice $(5^3 \times 3$ and $N_f = 4$) has been extended\textsuperscript{236, 238, 76} up to $8^3 \times 4$ with the conclusion unchanged.

This is a puzzling result in many respects. One might be tempted to say that the Wilson action does not have chiral transition precisely because it does not have chiral symmetry.

This does not help because one then has to contend with the possibility that the Wilson and Kogut-Susskind action define different continuum field theories. The naive identification of $K = K_c$ with $m_q = 0$ may also be doubted. According to the Ward identity analysis\textsuperscript{231}, however, this will become better as the lattice spacing decreases. It is possible that at the range of $\beta \sim 5$ where the transition approaches the critical line in current simulations with $N_f = 4$ the chiral breaking due to the Wilson term is too large to allow a chiral transition. Perhaps one begins to see a chiral transition in the Wilson action only when the temporal size $N_t$ is increased sufficiently. The transition line will move upward toward larger $\beta$ in that case, and therefore the magnitude of chiral breaking at the part of the line $K = K_c$ approached by the transition may become small enough. This should be checked by simulations with a larger temporal extent.

5.5 Physics of the high temperature phase.

We have seen much evidence in the previous sections that the normal phase of hadronic matter undergoes a first order phase transition at a temperature of the order of $T_c \sim 150 MeV$. We know that the chiral symmetry is restored in this phase. We would now like to understand better its thermodynamic properties. At extremely high temperatures $T \to \infty$ this may look treatable by the standard weak-coupling perturbation theory since the effective coupling constant will be given by $g^2(T)$ which decreases as $T \to \infty$. This, however, turned out not to be straightforward\textsuperscript{246}. For the energy density, for example, the $n$-loop ring diagrams of gluons with zero Matsubara frequency contribute an amount $g^{2n} \int d^2q \delta^n (\vec{p}^2 / |q|)^{n-2}$ which is infrared divergent from $3$-loop order. This divergence is an artifact; the $3$-dimensional gauge theory of the spatial components of gluons that describes the $T \to \infty$ limit is believed to have a mass gap $m_M$ (magnetic mass), and this cuts off the integral at $|q| \sim m_M$. However, the magnetic mass is proportional to the coupling constant of the $3$-dimensional theory $g^2$. Thus the order of the expansion actually ceases to increase beyond $3$-loop, and one can not discount the possibility that they add up to a large correction. Even if this turns out to be small the perturbative treatment is not applicable near the phase transition. One thus needs detailed numerical studies over the entire range of the temperature to understand the physics of the high temperature phase.

Extensive quantities. Let us start with some extensive thermodynamic quantities - the energy density $\epsilon$ and pressure $p$. We first need to derive their expressions on a lattice\textsuperscript{241}. We recall that

$$\epsilon = \frac{T^3}{V} \frac{\partial \log Z}{\partial T},$$

$$p = \frac{T^3}{V} \frac{\partial \log Z}{\partial V},$$

with $V$ the spatial volume. To carry out the differentiation it is convenient to use a different lattice spacing for the time $(a_t)$ and space $(a_s)$ directions. One then has $dT = -T da_t / a_t$, and $dV = 3V da_s / a_s$. Now that one has an asymmetric lattice the gauge coupling constant for
the space-like and time-like plaquettes are generally different. For the single plaquette action (1), they will be denoted as \( \beta_s / \xi \) and \( \beta_t / \xi \) with \( \xi = a_s / a_t \). Working out the differentiation and then setting \( \xi = 1 \) we find for the gauge contribution for the single plaquette action,

\[
\begin{align*}
c_t &= 3 d (c_t P_t - c_s P_s), \\
p_t &= \frac{1}{3} c_t + d (P_t + P_s),
\end{align*}
\]

with

\[
c_{ts} = 1 + \frac{1}{\beta} \frac{\partial \beta_{ts}}{\partial \beta}, \quad d = \frac{d \beta}{d \log a},
\]

and \( P_{ts} = 1/3 < \text{tr} U_{ts} > \) the average values of the time-like and space-like plaquettes. For \( g^2 = 6/\beta \) we have the lattice weak-coupling perturbation theory gives the following values of the coefficients \( \nu_{24} - \nu_{24} \) for the single plaquette action:

\[
\begin{align*}
c_t &= 1 + g^2 [-0.13194 - N_f / 0.00026 \quad \text{Wilson}], \\
&= 1 + g^2 [-0.03891 / 0.00391 \quad \text{KS}],
\end{align*}
\]

where the quark contribution is calculated at \( m_q = 0 \) (Kogut-Susskind case) or \( K = 1/8 \) (Wilson case), and

\[
d = -2 \beta (g^3) / g^2 = -0.83590 + N_f / 0.00066 + O(g^4),
\]

with \( \beta (g) \) the beta function (5-6). The energy density is predominantly determined by the difference \( P_t - P_s \). This is not the case for pressure. In fact, ignoring the contribution from the sum \( P_t + P_s \) is not justified since this will lead to the ideal gas equation of state \( \epsilon_t = 3 p_t \) independent of the parameters.

For the quark contribution a similar calculation leads to

\[
\begin{align*}
c_t &= -N_f / 4 < \text{tr} D_t D_s^{-1} >, \\
p_t &= -N_f / 12 < \text{tr} D_t D_s^{-1} >,
\end{align*}
\]

for the Kogut-Susskind action in which \( D_t, D_s \) denotes the temporal and spatial hopping term of \( D_q \), and we have ignored terms depending on \( \partial m_q / \partial \xi \) and \( \partial m_q / \partial a_0 \) which are proportional to \( m_q \) and small for light quarks.

The averages in the formulas for \( \epsilon \) and \( p \) do not vanish at zero temperature. This should be separately calculated on a symmetric lattice and subtracted. For qualitative analyses, this procedure may be skipped for the energy density. For the gluon part its contribution is small. The quark contribution can be rewritten using the cubic invariance:

\[
< \text{tr} D_t D_s^{-1} >_{T=8} \approx -3 - m Q < \text{tr} D_s^{-1} >_{T=0}. \]

For light quarks the second term is small, and hence may be ignored or substituted by the finite temperature value. This shortcut is not applicable for pressure.

In current full QCD simulations at small quark mass, the gluon and quark parts of the energy density exhibit the following typical behavior (see, for example, the figures in ref. 51, 224). As the coupling \( \beta \) is increased, both \( \epsilon_g / T^4 \) and \( \epsilon_q / T^4 \) rise abruptly away from zero at the critical point. For the gluon part, this is followed by a broad bump whose height rapidly increases with \( N_f \), and a subsequent slow decrease. The quark part, on the other hand, is roughly constant above \( \beta \) with a minor increase, and its value (normalized by \( N_f \)) increases relatively mildly with \( N_f \). Compared to the lattice weak-coupling perturbation expansion to \( O(g^2) \) on the same size of the lattice \( 24^3 \), the simulation values for the gluon part are larger even in the high temperature tail, while the quark part is comparable already at just above \( \beta \).

The large overshoot of the gluon part is different from the behavior seen in the pure gauge theory case on a lattice of similar size (for a recent data, see refs. 246-247). There the gluon energy density \( \epsilon_g / T^4 \) for \( T / T_c > 1.1 \) agrees very well with the lattice perturbative value to \( O(g^2) \) and shows only weak variations with \( \beta \). The origin of this difference is not yet completely understood. The weak-coupling expansion, however, gives some hint that the finite size corrections due to dynamical quark loops might be large. On an \( 12^3 \times 4 \) lattice with \( m_q a = 0.025 \), for example, the lattice weak-coupling expansion gives \( 246, 248 \)

\[
\begin{align*}
\epsilon_g / T^4 &= 7.72 + (-0.842 + 0.859 \cdot N_f) g^2 + \ldots, \\
\epsilon_q / T^4 &= -6.13 - 0.455 \cdot N_f g^2 + \ldots.
\end{align*}
\]

where for \( \epsilon_g \) we have included the vacuum subtractions. (\( m_q \) dependence toward \( m_q = 0 \) is already small \( 251 \).) In the continuum perturbation theory only the sum is available \( 249, 248 \):

\[
\epsilon / T^4 = (\epsilon_g + \epsilon_q) / T^4 = 8 \pi^2 / 15 + \frac{7 \pi^2}{20} N_f - \frac{1}{2} \frac{5}{24} N_f g^2 + \frac{2}{\pi} (1 + N_f / 6) \frac{1}{2} g^2 + \ldots
\]

(100)

Comparing (99) and (100) one sees that the lattice gluon loop correction (the first \( O(g^2) \) term in \( \epsilon_g \) in (99)) has the same sign and a similar magnitude as the continuum result. The net quark loop contribution (the sum of terms proportional to \( N_f g^2 \)), however, is positive on the lattice, while the continuum value is negative and small \( (-5/24) \). The difference arises from the large positive coefficient of \( N_f g^2 \) in the gluon part \( \epsilon_g \). This is qualitatively consistent with the behavior of \( \epsilon_q \) observed in simulations; a rapid increase of its value with \( N_f \) and a decrease for larger \( \beta \). The coefficient, however, has to decrease below the corresponding term in \( \epsilon_q \) in the continuum limit to produce a net negative coefficient. This suggests that the overshoot of \( \epsilon_g \) may become less pronounced as the lattice size is increased (see also ref. 252 for a study of other quantities).

In this context, we should recall that the proper limit at finite temperatures is to first go to a large spatial volume \( N \rightarrow \infty \) and then to increase the temporal size \( N_t \). The necessity for a large ratio \( N / N_t \), even for a small \( N_t \), is nicely illustrated by a recent pure gauge simulation on an \( 16^3 \times 4 \) and \( 24^3 \times 4 \) lattice \( 249 \). It has been found that \( \epsilon_g / T^4 \), constant for
\[
\frac{\Delta \epsilon_\beta}{T_c^4} = 2.54(12), \quad \text{for } N_f = 4 \text{ (pure gauge)}, \quad (101)
\]

which is roughly half of the previous estimates from smaller lattices\textsuperscript{232}. With a small \( N_f/N_c \sim 2 \), the transition is too much rounded off and the fluctuation is too large to obtain accurate values close to \( \beta_c \). A large temporal size is also important of course. For example, examination of the lattice size dependence within the weak-coupling expansion\textsuperscript{243,251} shows that the lattice satisfying \( N_f/N_c \sim 3 - 4 \), though sufficiently close to the infinite spatial volume limit, still has a net positive sign for the \( N_f g^2 \) term in the total energy density \( \epsilon \) at \( N_f = 8 \) for \( m_\pi = 0 \).

The saturation of the quark energy density \( \epsilon_\beta \) (and also that of \( \epsilon_\pi \) in the pure gauge case) at the weak-coupling value quite close to the critical temperature might suggest that the high temperature phase is essentially a plasma of weakly interacting quarks and gluons. That this is not so is indicated by the behavior of pressure which is far less than \( \epsilon/3 \) expected for free quarks and gluons near \( T_c \). This has been observed both for the pure gauge case\textsuperscript{244,247,283-284} and for full QCD\textsuperscript{211}. Consequently the entropy density \( s = (\epsilon + p)/T \) is also reduced below the ideal gas value \( 4\epsilon/3T \). They do approach the ideal gas law for \( T/T_c \sim 2 - 3 \) and above, however. An interesting question to clarify is what is the difference in the energy from the \( p = 1/3\epsilon \) relation when the energy density is well approximated by the weak-coupling values. Perhaps this may be explained by the existence of heavy modes in addition to the light quarks and gluons\textsuperscript{284,285}.

**Screening lengths.** We now come to the next quantity - the screening length of the plasma. We have already encountered one of this type in the correlation function of the Polyakov line (76-77). We see from (77) that the \( \bar{q}q \) potential in the high temperature phase behaves as

\[
V_{q\bar{q}} \sim \frac{\alpha}{r^n} e^{-\mu_D r}, \quad (102)
\]

with \( n \) some constant. This is very similar to the Debye-screened potential in the QED plasma, and for this reason \( \xi_D = 1/\mu_D \) is also called the Debye screening length and \( \mu_D \) the Debye mass.

The magnitude of \( \mu_D \) has implications on the production of heavy quark bound states such as \( J/\psi \) in heavy ion collisions. For \( \mu_D \) large, the screened potential will be too short-ranged to bind the quark-antiquark pair. If the high temperature phase is formed during the collisions, this will lead to a suppression of their production\textsuperscript{244}.

There are a fair number of calculations of \( \mu_D \) in the pure gauge theory\textsuperscript{276,287-288}, but only a few in full QCD\textsuperscript{224,285}. The results show several interesting features. Firstly the full QCD value is generally higher than the pure gauge value. This supports the naive expectation that the \( q\bar{q} \) pair creation and annihilation enhance the shielding. Secondly, close to \( T_c \), the Debye mass is already large \( (\mu_D/T \sim 3 - 4 \) for full QCD and \( \sim 2 \) for pure gauge) at \( T/T_c \sim 1.1 \). (This pure gauge value shows a rapid drop below this temperature similar to the behavior of the energy density. This is another aspect of the weakness of the transition discussed in Sec. 5.2.)

The third point concerns the high temperature behavior and a comparison with the weak-coupling perturbation theory. In the continuum the Polyakov line is the line integral of the time component of the gauge potential \( A_4(x) \). A simple calculation shows that the static limit of the propagator \( \langle A_4(x, \omega = 0)A_4(y, \omega = 0) \rangle \) develops a mass given by\textsuperscript{240,244-246}

\[
m^2_0 = \lim_{\omega \to 0} \Pi_{\mu \nu}(k, \omega \to 0) = g^2 T^2 (1 + \frac{N_f}{6}) + O(g^4), \quad (103)
\]

with \( \Pi_{\mu \nu} \) the self-energy tensor for gluons, which is called the electric mass. Using this dressed propagator for the Polyakov line correlation function, one finds that the lowest non-vanishing term is given by the two gluon exchange, with the result \( \mu_D = 2m_\pi \) and \( n = 2 \) in (102). For the pure gauge case, a simulation results\textsuperscript{247} on a \( 24^3 \times 6 \) lattice using a fit with \( n = 2 \) is in reasonable agreement for \( T/T_c > 4 \), approaching the perturbative prediction from above. (The lattice correction to (103) is about \( -10\% \) on this lattice. See ref. 292.) The agreement is lost close to \( T_c \). In fact the power \( n \) appears to change from \( n = 2 \) for large \( T \) to \( n = 1 \) near the critical point\textsuperscript{285-286}. Perturbative computation of higher order terms are controversial for the electric mass\textsuperscript{264-268} as well as for the Polyakov line correlation function\textsuperscript{285,287}. The agreement found with the lowest order result might be a sign that they are not large.

In a QED plasma the spatial components of the electromagnetic potential do not develop a mass. We have already mentioned that the QCD plasma, on the contrary, is expected to possess a magnetic mass of the order \( g^2 T \) which, however, is not calculable in perturbation theory. This is clearly a quantity of considerable interest. Unfortunately one does not know what types of operators allow a reliable extraction of the magnetic mass in numerical simulations.

The screening length is generally dependent on the quantum numbers. For the Debye screening length, one was examining the shielding of the color charge. An alternative choice is the quantum number of hadrons\textsuperscript{289}. This is measured by hadron Green's functions at a spatial separation;

\[
\sum_{i,x,y} < O_H(i,x,y,z) O_H(0) > \sim e^{-\xi_H}, \quad (104)
\]

The calculations of this quantity in full QCD with light Kogut-Susskind quarks have revealed two interesting features\textsuperscript{269-270}. (i) Above the chiral transition, the pion remains massive at \( m_\pi = 0 \), and the parity partners (\( \pi - \sigma \) and \( \rho - A_1 \) for mesons and the nucleon \( N_\pi \) and its counterpart \( N_\pi \) ) become degenerate, while they were completely separate below. (ii) The "mass" \( m_\pi \equiv 1/\xi_H \) is quite large \( (\mu_D - \sigma)/T \sim 3 - 4, \mu_N - N_c/T \sim 10 \). These characteristics have also been found\textsuperscript{79} for the Wilson quark action at \( K \approx K_0 \). The first point shows that the chiral symmetry is no longer spontaneously broken above the critical temperature, which
is consistent with the behavior of $\langle \chi \chi \rangle$ found in the phase structure analyses.

It has been argued that the second feature might imply the existence of heavy real
time excitations with the quantum number of hadrons. If the phase diagram has a gap of
singularity in the intermediate range of $m_q$, one can analytically continue the hadronic states
in the low temperature phase into the high temperature phase. The exponential decay (104)
may then follow by Wick-rotating the real time Green's function into imaginary time, with
$\mu_H$ related to the excitation spectrum in real time. This argument is rather qualitative.
In particular it does not say anything about the life time of the excitations. In fact one might argue
that (104) is merely probing the property of the spatial components of the gauge
field and has nothing to do with real time excitations because the 3-dimensional dynamics of
gauge fields is confining at any temperature and this by itself will lead to the exponential
decay.

This criticism is probably valid at very high temperatures where the temporal size of the
system shrinks to zero. Close to the transition, however, the problem is open at the moment.
Further clarification is not straightforward in numerical simulations of lattice QCD. Being
closely tied up with the Euclidean formulation, the real time dynamics has to be explored
by making the analytic continuation back to real time numerically, and this is not an easy
task (see ref. 272 for a related attempt).

Quark number susceptibility. There is, however, an indirect way to examine the
presence of modes carrying a certain quantum number $Q$. This is through its susceptibility defined by

$$\chi_Q = \frac{1}{V} \frac{\partial \langle Q \rangle}{\partial Q},$$

where $\mu_Q$ is the chemical potential for the quantum number $Q$ which is set to zero in the
final formula. If there is a light mode of mass $M(< T)$ with $Q \neq 0$, the susceptibility will be
large. If the mass is heavy $M > T$, then it will be suppressed ($\chi_Q \propto e^{-M/T}$).

This has been applied in an $N_f = 2$ simulation taking $Q$ to be the quark number
and the isospin. In the low temperature phase one expects the quark number susceptibility
$\chi_Q$ to be close to zero since baryons are heavy, while the isospin susceptibility $\chi_{NS}$ need not
be small due to light pions. The numerical results show that both $\chi_Q$ and $\chi_{NS}$ are close
to zero below the transition, but once the transition is crossed they quickly move up to a
constant value which are roughly equal and similar to those of the free quarks. The small
isospin susceptibility in the low temperature phase may be explained by the artificial feature
of the simulation that the pion mass was still large $m_\pi / T \sim 1.6$. The behavior above the
transition, on the hand, suggests that the quarks and antiquarks constitute a part of the
light excitation spectrum.

5.6 Finite baryon number density.

Up to now we have raised only the temperature and have kept the density of baryon
number at zero. It is very desirable to lift this restriction, especially in the astrophysical
context where the density effect is often more important than that of temperature. This,
however, turned out to be difficult. We shall explain why this is so and also summarize what
little we know about the behavior of QCD at large baryon number density.

The standard method to control the baryon number density is to introduce the chemical
potential $\mu_B$ for quarks. Thus one switches from the canonical partition function (71) to the
grand canonical partition function defined by

$$\Omega = Tr(e^{-\beta (H_{Q}\bar{c} - \mu_{B} N_{Q}) / T}),$$

where $N_{Q}$ denotes the number of quark minus that of antiquark. The quark number density
is

$$\frac{N_{Q}}{V} = \frac{T}{\partial \Omega / \partial \mu_{B}}.$$

and the quark action in the continuum receives an additional term $\mu_{B} Q \bar{c} c$.

To transcend this onto lattice we recall that the role of the chemical potential is to induce
an unbalance between the quark and antiquark. On a lattice the meaning giving different
weights to the hopping term of quarks in the positive and negative time directions. This
could be done by substituting $U_{\alpha \gamma} \rightarrow \eta_{\gamma} U_{\alpha \gamma}$ and $U_{\beta \gamma} \rightarrow \eta_{\gamma} U_{\beta \gamma}$ in the quark action
(12) or (17). Since the states having an equal number of quarks and antiquarks should not
be affected, one can put the constraint that $\eta_{\gamma} = 1$. The factors $\eta_{\gamma}$ are functions of
the chemical potential in lattice units $\mu_{B}$. Demanding that the classical continuum limit
reproduces the continuum form fixes their small $\mu_{B}$ behavior: $\eta_{\gamma} = 1 + \mu_{B} a + O((\mu_{B} a)^2)$. This
actually represents the most general choice within the constraint that the corresponding
energy density converges to the continuum form $\rho_{b} / 16\pi^2$ without encountering quadratic
divergences. The most commonly used form is $\eta_{\gamma} = e^{\mu_{B} a}$.

One can make a few guess as to what might occur with an increasing $\mu_{B}$. For heavy
quark, the breaking of the center $Z(3)$ symmetry, though small at $\mu_{B} = 0$, becomes large
since the time-like hopping term which is the source of its breaking is enhanced by $\mu_{B}$.
This means that the remnant of the pure gauge deconfining transition weakens further and
will disappear for sufficiently large $\mu_{B}$. For light quarks, the chemical potential does not
break chiral symmetry. However, the mismatch of quark and antiquark will disfavor the
formation of the $\bar{q}q$ condensate. One therefore expects a restoration of chiral symmetry as
$\mu_{B}$ is increased.

Numerical attempts to verify this picture run into a serious problem. Because $\eta_{\gamma}$ and
$\eta_{\gamma}$ are real and different, the quark determinant det$D_{\gamma}$ is no longer real for any $\mu_{B} a \neq 0$.
All known methods of simulations are based on a stochastic process governed with a real
transition probability. With a complex action $S = S_{R} + i S_{I}$, one may either compute the
ratio $\langle Q e^{S_{I}} \rangle / \langle e^{S} \rangle$ using $S_{P}$ as the action or may resort to a complex extension of the Langevin algorithm and its hybrid-type modifications. None of these work for QCD. The basic reason is the fluctuation of the phase of the determinant. It may well happen that the phase has a global oscillation over the phase space resulting in a substantial
cancellation of the contributions from its various parts. All known methods, however, are
local in phase space in the sense that they determine the direction of motion by examining
the immediate neighborhood of the current position. Thus these methods are in fact unaware
of the possibility of global cancellations. Theoretical284, as well as numerical284 analyses
have shown that just such a phase cancellation occurs for QCD when the chemical potential
approaches the critical point. It is quite natural, then, that the conventional methods have
failed to produce meaningful results on the phase structure at finite \( \mu_B \). This argument also
shows that the quenched approximation \( det D_4 = 1 \) is in fact wrong285.

It should be clear that progress requires a method which either integrates out the
phase from the partition function or has good control over it. In the strong-coupling limit
\( \beta = 0 \) there have been a few attempts285,286 of the former type using a representation
of the QCD partition function in terms of dimers and monomers. These simulations show
that at \( \beta = 0 \) there is a clear first order phase transition across which the chiral symmetry
becomes restored. The mean field theory turned out to be fairly accurate in predicting the
transition. A possibility for the latter is the application286 of the spectral density method286
which counts the number of configurations with a given value of the phase. Unfortunately
both methods suffer from a severe limitation. The dimer method is difficult to extend beyond
the strong coupling limit, and the spectral density method is limited to small lattices such
as 2\(^4\) or 4\(^4\).


In this review we have surveyed the current status of full QCD simulations in two
major areas of concentrated activity—the hadron mass spectrum and the finite temperature
behavior of QCD. This is an endeavor which seriously started in 1984-85. At that time very
little was known on the effects of dynamical quarks in either of these subjects. Just prior
to this period it was not even clear whether a reasonable full QCD simulation was possible
because of the lack of efficient algorithms at that time.

Progress since then has been rapid particularly on the finite temperature behavior of
QCD. We now have a fairly accurate picture of the phase diagram. There exists a chiral
phase transition at small quark mass which is first order for the realistic case of \( N_f = 2 \sim 4 \).
Current simulations tentatively place the critical temperature at about 150 MeV, though
with a large margin of uncertainty. In addition the physics of the high temperature phase
are being studied using a variety of physical quantities. The progress has been less dramatic
for the hadron mass spectrum. The dynamical quarks do not seem to change the quenched
spectrum beyond the level of 10%—20%.

There are many questions which remain to be clarified. Whether the dynamical quarks
account for the discrepancy between the experimental and quenched value of the \( m_N/m_p \)
ratio is an important one. Equally interesting is the \( U(1) \) problem which has not seen much
progress. The mass spectrum is only one of a variety of hadronic quantities of interest. A
large effort in the quenched calculation today is the evaluation of hadronic weak matrix
elements286. Eventually full QCD simulations should be extended to such quantities and
others. Among the questions at finite temperatures, the order of the pure gauge deconfinement,
though not really a full QCD question, should be settled since it forms one of the basis
of our understanding of the phase diagram. The finite size dependence of the chiral transition
as well as the scaling of its critical temperature also require further studies. In addition
we have emphasized two questions quite problematic at the moment; the chiral transition
for the Wilson quark action and the consistency with the results from the Kogut-Susskind
action, and the phase structure at large baryon number density.

The full QCD simulations are about two orders of magnitude more costly in computer
time compared to the quenched calculations. For this reason the lattice size and the inverse
lattice spacing used for full QCD have been roughly half of that of the most advanced
quenched ones. In most of the remaining questions listed above, further progress will be
difficult unless such restrictions are removed. This depends on the efficiency of the algorithm
and the speed of computers.

Concerning the algorithm, the currently used ones are all \( O(V) \) category and the number
of arithmetic operations at each sweep will be difficult to reduce further. Improvement in
efficiency may be pursued in accelerating the matrix inversion and in reducing the correlation
over the successive sweeps.

The crucial factor, however, will be the speedup of computers. The peak speed of the
supercomputers such as CRAY XMP used in full QCD simulations up to now is 300 Mflops to
1 Gflops. An important development recently is the appearance of custom made computers
dedicated for lattice QCD. The ones already completed (Columbia parallel processor and
Ape computer) have a peak speed of 1 Gflops, comparable to the commercial supercomputers.
They, however, have the advantage of a dedicated machine and have started producing
simulations of a scale125,173,176 not easy to achieve with general purpose computers. There
are several other ongoing projects of dedicated computers, all aiming at a peak speed of 5
to 10 Gflops to be completed within a few years time. (For a review of the status of these
projects, see ref. 291.) The commercial machines are also entering this range of speed. A
ten-fold increase in speed is therefore quite plausible within a year or two. Looking further
ahead the Ape collaboration has announced a project for achieving 100 Gflops in 3 to 4
years from now292. With this range of speed, full QCD simulations will reach and perhaps
surpass the scale of current quenched calculations in the lattice size and the lattice spacing.

We conclude with an optimistic note that our understanding of the dynamics of full
QCD will become one level deeper within the next few years.
Acknowledgments.

I would like to thank M. Fukugita and Y. Oyanagi for close and fruitful collaboration over the years. I have greatly benefited from discussions with F. Katsch on QCD at finite temperatures. Valuable discussions with N. Christ, R. V. Gavai, E. Marinari, G. Martinelli, G. Parisi and A. Patel are gratefully acknowledged. I would also like to thank M. Gavels and A. Gonzalez-Arroyo for inviting me to present this review at the GIF Seminar.

References.

8. See, for example, M. Creutz, Quarks, Gluons and Lattices (Cambridge University Press, Cambridge1983).
14. For recent status, see Proceedings of the Fifth Int. Conf. on Ultra Relativistic Nucleus-Nucleus Collisions, September 1988, to appear in Nucl. Phys. A.
134. See J. Gasser, H. Leutwyler, M. P. Locher and M. E. Sainio, Bern preprint (June 1988) and references cited therein.
152. Y. Koike, Univ. of Tokyo preprint (1988).
179. G. Parisi, private communication.
185. For a review, see F. Y. Wu, Rev. Mod. Phys. 54 (1982) 235.