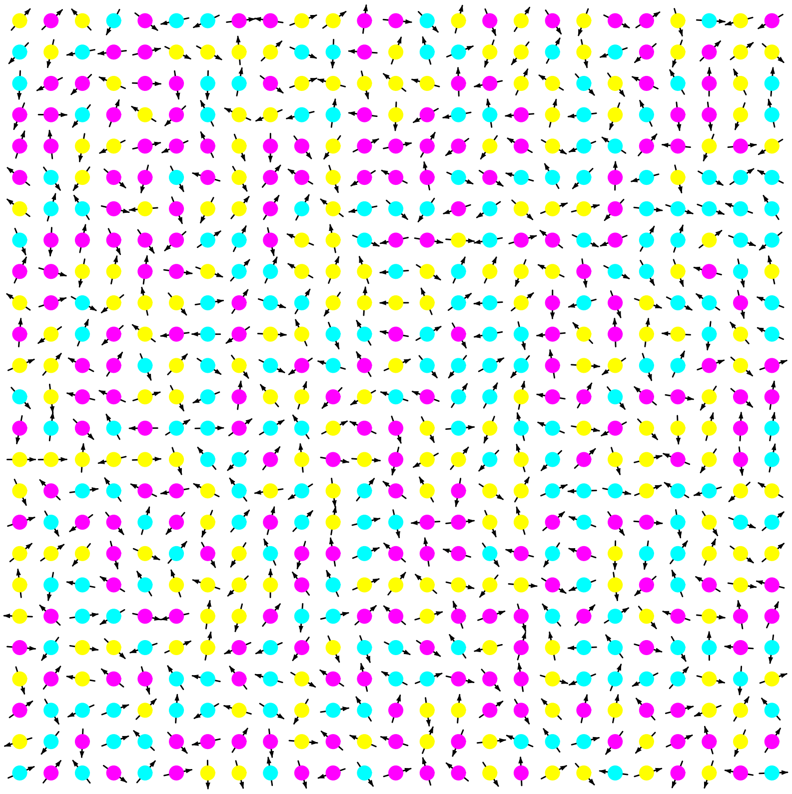
ECFA/99/200

Requirements For High Performance Computing For Lattice QCD: Report of the ECFA Working Panel

F. Jegerlehner (DESY-Zeuthen), R.D. Kenway (Edinburgh),
G. Martinelli (Rome-I), C. Michael (Liverpool), O. Pène (Orsay),
B. Petersson (Bielefeld), R. Petronzio (Rome-II),
C.T. Sachrajda (Southampton – Chairman) and K. Schilling (Wuppertal)

Abstract

This report, prepared at the request of the European Committee for Future Accelerators (ECFA), contains an assessment of the High Performance Computing resources which will be required in coming years by European physicists working in Lattice Field Theory and a review of the scientific opportunities which these resources would open.



Contents

1 Introduction

For the last 15 years or so lattice simulations have been making major contributions to the development of our understanding of subtle details of the standard model of particle physics, to the determination of its parameters and to searches for signatures of "new physics". For many physical processes the major source of systematic uncertainty is a theoretical one, namely our inability to quantify non-perturbative strong interaction effects. Large scale numerical simulations, using the lattice formulation of QCD, allow these effects to be computed from first principles and with no model assumptions or free parameters. The precision of these computations is limited, however, by the available computing resources. In this report we review the status and prospects for lattice simulations across a broad range of processes and fundamental quantities in particle physics. We also review the likely technical developments, both in computing hardware and in algorithms, and discuss the scientific opportunities which these developments will provide.

For illustration we present in Fig. 1 the allowed region for the vertex A of the unitarity triangle (i.e. the vertex which subtends the angle α) from a recent analysis and an illustration of how this region would shrink after about a year of lattice calculations on a 10 TFlops machine. In the latter case we have not included any expected improvements in the precision of experimental measurements.

The European lattice community is in a very strong position to exploit the opportunities which will be opened up by access to the 10 TFlops generation of computers and beyond. The community is very successful and has direct experience in almost all of the wide range of problems which are amenable to numerical simulations (see sections 3 – 6 below). It led the way in the use of parallel computers for non-perturbative QCD and its successes include pioneering studies of hadronic effects in weak decay and mixing processes; the development of an understanding of the chiral structure of lattice QCD; the "improvement" of lattice computations and the "non-perturbative normalization" of lattice operators (which reduce substantially two important sources of systematic uncertainty); understanding the equation of state of QCD and the spectroscopy of glueballs and hybrids. We estimate that the community numbers approximately 150 physicists at junior faculty level or above. Many of the lattice research groups have for over ten years worked in collaborations with groups from other European countries and currently there are two EU supported networks, one in QCD at finite temperature (awarded under the Framework 4 program) and the second in Lattice Phenomenology (awarded under the Framework 5 program in October 1999). These networks support the development and exchange of expertise through the mobility of researchers between different countries and the provision of high quality training. The European lattice community is therefore in an excellent position to coordinate its research activities still further as required for future projects (see below).

The report is organised as follows. In the following section we present a brief introduction to lattice calculations in order to set the scene for the following discussions. Sections 3 – 6 contain reviews of the main physics problems which can be studied using lattice simulations. We start in section 3 with a review of the applications of lattice simulations to hadron spectroscopy, including studies of glueballs, hybrid states and other exotic particles. Computations of the strong coupling constant α_s and quark masses are described in section 4. This section also contains a discussion of the evaluation of hadronic matrix elements which is central

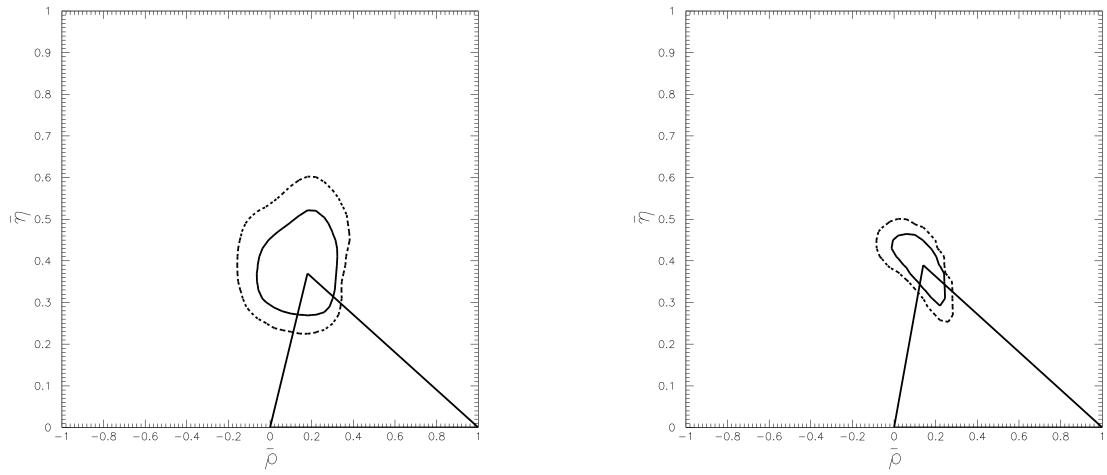


Figure 1 A recent analysis of the allowed region for the Unitarity Triangle (left) and an illustration of how it might shrink after about a year of calculations on a 10 TFlops machine (right). Contours correspond to 65% and 95% confidence levels. For the current (future) analysis we have taken: a) $\hat{B}_K$ = 0.87(13) (0.87(4)); b) $f_{B_d}\sqrt{B_{B_d}}$ = 210(30) MeV (210(15) MeV); c) $\xi = f_{B_d}\sqrt{B_{B_d}}/f_{B_s}\sqrt{B_{B_s}}$=1.14(6) (1.14(6)); d) $\bar{m}_b$ = 4.25(15) GeV (4.25(4) GeV) and e) V_{ub} = 0.00370(70) (0.00370(15)).

to studies of, for example, weak decays and hadronic structure. Simulations of QCD at finite temperature, with its implications for cosmology, the quark-gluon plasma and heavy-ion physics are discussed in section 5. We end our review of the physics opportunities in section 6, where we explore possible applications to simulations of quantum field theories other than QCD, including electroweak theories, supersymmetric gauge theories and quantised theories of gravity. Throughout sections 3 − 6 estimates of the current uncertainties are given and prospects for the reduction of these uncertainties are discussed. In order to make optimal use of the available computing resources it is necessary to continue developing efficient algorithms for the generation of field configurations, and this is discussed in detail in section 7. In section 8 we present a review of Computing Technology discussing today's high-end computers, likely developments in commercial and purpose built systems and possible machine strategies for the European lattice community. Finally we present our conclusions and recommendations to ECFA.

2 A Brief Introduction to Lattice Calculations

We start by outlining very briefly some general features of lattice simulations. This will enable us, throughout this report, to discuss the precision of these computations and how it can be improved. Most lattice simulations in QCD involve the direct computation of the vacuum expectation values of multi-local operators composed of quark and gluon fields (in Euclidean space):

$$\langle 0|\,O(x_1,x_2,\ldots,x_n)|0\rangle = \frac{1}{Z}\int [DA_\mu][D\psi][D\bar{\psi}]\,e^{-S}\,O(x_1,x_2,\ldots,x_n)\ , \qquad (1)$$

where Z is the partition function

$$Z = \int [DA_\mu][D\psi][D\bar{\psi}]\, e^{-S} \ , \tag{2}$$

S is the action and the integrals are over quark and gluon fields at each space-time point (ψ and $\bar{\psi}$ represent quark and anti-quark fields and A_μ the gluon fields). In eq. (1) $O(x_1, x_2, \dots, x_n)$ is a multi-local operator; the choice of O governs the physics which can be studied.

The right-hand side of eq. (1) is obtained by discretizing space-time (hence the word *lattice*), and evaluating the functional integral in eq. (1) by Monte-Carlo integration. Clearly the lattice must be sufficiently large to contain the hadron being studied and sufficiently fine-grained so that the errors due to the finite lattice spacing are small. The number of lattice points in a simulation is limited by the available computing resources and so a compromise generally has to be accepted and the errors due to the finite volume or finite lattice spacing have to be quantified.

The Monte-Carlo algorithms generate field *configurations* with a probability given by the Boltzman factor, $\exp(-S)/Z$, and the integral in eq. (1) is evaluated by averaging O over these configurations. We present a detailed discussion of the scaling behaviour of these algorithms as the lattice parameters (volume, lattice spacing and quark masses) are varied in section 7.

2.1 Eliminating the Quenched Approximation

For each gluon configuration $\{A_\mu(x)\}$ used to sample the integral in eq. (1), the functional integral over the quark fields can be performed formally, involving the determinant of the Dirac operator in the gluon background field corresponding to this configuration. The numerical evaluation of this determinant is possible, but is computationally very expensive since it has to be performed at each step of the iterative process used to generate independent configurations. For this reason, in most simulations performed in the past, the determinant was set equal to its average value, which is equivalent to neglecting virtual quark loops (this approximation is conventionally referred to as *quenching*). Although results for physical quantities obtained in the quenched approximation are typically within about 20% of the experimental values when these exist, it is not possible to quantify the effects of quenching in general. It is therefore necessary to eliminate this approximation in order to control the precision of lattice simulations with confidence. This is now becoming possible and throughout this report, unless explicitly stated to the contrary, we assume that we are discussing unquenched calculations (these are also referred to as *full-QCD* simulations or simulations with *dynamical quarks*).

It is difficult to perform unquenched simulations with sufficiently light u and d-quarks. Not only does the computing cost increase very significantly (see section 7) but light pions propagate across long distances and hence very large lattices are required. In practice therefore one simulates with u and d quarks which have unphysically heavy masses and the physical result is obtained by extrapolation.

3 Hadronic spectroscopy

The hadron spectrum is the benchmark of lattice QCD. It is essential to have excellent control over particle masses to give us confidence in the evaluation of other quantities, such as weak

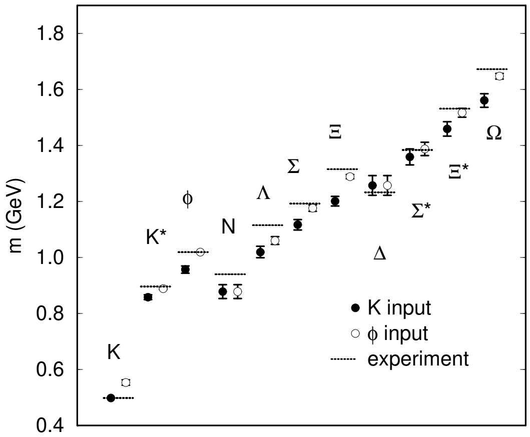


Figure 2 Meson and baryon quenched lattice results from CPPACS Collaboration.

matrix elements. The hadron spectrum also generally provides the input required to establish the values of the bare lattice parameters which give the lattice spacing and quark mass.

Precision studies in the quenched approximation (with sea-quark effects neglected) have been performed and a comparison with experimental values is shown in fig. 2. It is found that the lattice results for hadronic masses typpically differ by $\pm 10\%$ from the experimental values. One feature of this mismatch is that setting the strange mass scale from the K- or the ϕ-mesons gives different results. Note that, unlike the case of a few years ago, the quenched lattice results obtained using small lattice spacings and small valence quark masses do now reproduce the experimental ratio of masses $m(N)/m(\rho)$ to within 10%. This semiquantitative agreement of quenched results with experimental ones is satisfying, and gives us confidence in the use of quenched simulations for other quantities. In order to have full control over all sources of systematic errors, however, it is essential to include sea-quark effects.

As one reduces the masses of the sea quarks in unquenched simulations, one expects the computed spectrum to converge to the experimental one. Current studies provide some evidence of this, but they rely on extrapolations from sea-quark masses that are quite heavy. For full QCD, it is possible that the sea-quark effects turn on non-linearly as the mass of the sea-quarks is reduced. One simple example is that the ρ-meson can only decay if $m(\pi)/m(\rho) < 0.5$, and this condition is not attained in most present studies. From arguments based on Chiral Effective Lagrangians one expects that the extrapolation to physical u and d quarks should be reliable if one can compute directly with light quarks with masses of about 25% of the strange quark mass. This corresponds to a ratio of $m(\pi)/m(\rho)$ of about 0.4.

The main computational bottleneck in reducing the systematic errors is in generating gauge configurations with light sea quarks. In section 7 we discuss the scaling behaviour of algorithms as the lattice parameters are varied and show that in order to keep the spatial volume large enough, with a fine lattice spacing, we will require a tremendous increase in computing power. By comparison, the analysis needed to extract information about the hadronic spectrum from the configurations is computationally relatively straightforward. As we will now discuss in more detail, in many cases a large number of gauge configurations will be needed to reduce the error on the measurement and this will limit the precision which can be attained.

3.1 Flavour-singlet mesons

Within the quenched approximation the flavour-singlet pseudoscalar meson η_1 is degenerate with the flavour non-singlet η_8 (where η_8 and η_1 are admixtures of the physical η and η' mesons). This degeneracy is not present when sea-quark effects are included and obtaining a reasonable description of the η' mass is a major goal of the lattice simulations with dynamical quarks. Other flavour-singlet quantities are also of considerable interest: glueball-meson mixing, the nucleon sigma-term, B meson decays to η', etc.

It is possible to evaluate the relevant quark propagators (hairpin diagrams) stochastically with sufficient precision. However, the intrinsic variance of the signal is large (the error does not decrease with time separation t) so that measurement of many independent gauge configurations is needed to get accurate results. This is in contrast to the case of flavour non-singlet mesons (and baryons) where only connected quark propagators are needed and where data sets of the order of 1000 configurations have proved to be sufficient. The flavour-singlet correlators are more noisy, much like glueball correlators, and more than 10^4 gauge configurations will be needed to get accurate determinations of mass shifts. For light sea-quarks, this is a leading edge challenge.

3.2 Hadronic decay

Using Euclidean lattice methods, a few of the lowest lying discrete energy levels for any set of quantum numbers are measurable. For a state unstable to strong decay, this implies that the mass of the lightest decay product will dominate. For example for the ρ meson this will be the $\pi\pi$ state with non-zero relative momentum (to allow P-wave angular momentum). Thus a precision measurement of the mass of such an unstable particle is potentially difficult. It has not been possible to explore this problem in present studies with dynamical fermions because in most cases the quark masses have been too large (and the minimum lattice momenta is $2\pi/L$) so that the ρ is actually stable on these lattices.

As well as hadronic decays, a considerable and related challenge is to find methods to explore non-leptonic weak decays on the lattice. Within the quenched approximation, hadronic decays can be studied in principle between states of the same mass. This allows a study of mixing between glueballs and scalar mesons or between a string state and two heavy-light mesons. An exploratory low precision study of glueball decay and mixing employed 10 GFlops-year. Extending these studies to full QCD will require analysis of at least 10^4 gauge configurations.

3.3 Glueball and hybrid mesons

Lattice QCD has proved to be an excellent guide to experiment in the search for states which are not present in the quark-model. The quenched results of 1.6 GeV for the scalar glueball and 1.9 GeV for the $J^{PC} = 1^{-+}$ spin-exotic hybrid mesons have been benchmarks in this area.

The glueball is defined in the quenched approximation, whereas with sea quarks included, the glueballs are just flavour-singlet mesons. Thus a detailed measurement of the lightest flavour-singlet mesons will be needed and this is a major computational challenge as described above.

Quantity	Present error	2003-6
$m(\eta') - m(\eta)$	dominated by systematic	30 MeV
0^{++} mixing	dominated by systematic	10%
$m(1^{-+})$	200 MeV	20 MeV
$m(H)$	dominated by systematic	20 MeV

Table 1 Estimates of present and future (assuming 18 TFlops-year) achievable errors on flavour-singlet splitting in pseudoscalar mesons, glueball-scalar meson quenched mass mixing matrix element, spin-exotic hybrid meson, and H dibaryon.

The spin-exotic hybrid meson spectrum has already been measured in full QCD but with rather heavy sea-quarks. A study of mixing with meson-meson channels is important since experiment finds candidate states at 1.4 GeV decaying to $\eta\pi$ and at 1.6 GeV decaying to $\rho\pi$. One should therefore evaluate the $q\bar{q}q\bar{q}$ contributions to the spin-exotic state since this might lead to a better understanding of the disagreement in mass values between theory (1.9 GeV) and experiment (1.4 GeV).

3.4 Other exotic states

The best known quark molecular state is of course the deuteron. A study of nucleon-nucleon forces from first principles using lattice QCD is an important step to de-mystifying nuclear forces. A related problem which is amenable to lattice QCD study is the binding of the dilambda or H di-baryon which has been postulated to be stable in some models and hence of cosmological relevance.

A precision study of these di-baryon systems needs a large spatial volume (since there are two baryons) and, since the binding energies are typically of order 10 MeV, a considerable statistical precision will be needed to study the residual strong force responsible for binding.

4 α_s, Quark Masses and Hadronic Matrix Elements

The determination of the CKM parameters, A, $\bar{\rho}$ and $\bar{\eta}$, the study of CP violation in kaon systems (including the measurements of ϵ and ϵ'/ϵ), and the search for FCNC effects induced by new physics (for example by supersymmetry), all rely on the evaluation of weak hadronic amplitudes. Important improvements have been achieved in recent years in phenomenological analyses, mainly using input from lattice QCD. Future programs will also pursue the calculation of the basic (renormalized) parameters of the QCD Lagrangian, the coupling constant and the quark masses, in order to reach a precision which is competitive with determinations coming from perturbative calculations combined with high-energy experiments. We discuss these and related topics in some detail in the remainder of this section.

For many physical quantities results from lattice simulations have been available for some time and the main issue is the precision which can be achieved in the next few years. We review this question in subsections 4.1-4.3 below. For other quantities, such as two-body hadronic B-decays and the shape function in inclusive semileptonic decays of heavy hadrons, the computations are still at an exploratory stage. In this case the main issues are the theoretical developments which have to be undertaken and the computing resources which are neccessary. A number of such quantities are discussed in subsec. 4.4.

In table 2 we present our assessment of the current errors for a number of important physical quantities. Educated guesses for *short-term* improvements (corresponding to 250–600 GFlops in the years 2000-2003) and *long-term* improvements (corresponding to about 10 TFlops in the years 2003-2006) are also presented.

In estimating the long-term uncertainties in table 2 it is anticipated that with access to a 10 Tflops machine it will be possible to:

- obtain results in the full, unquenched theory, at values of the lattice spacing and quark masses comparable to those presently used in quenched calculations;
- study physical quantities in the quenched approximation at values of the lattice spacing in the range currently used in simulations but with much ligher u- and d-quarks (with quark masses of the order of 20–30 MeV). For quenched simulations we assume lattice volumes of $36^3 \times 96$ in the short term and $72^3 \times 192$ in the longer term with a statistics of $\mathcal{O}(1000)$ configurations;
- study B-physics in quenched simulations at small values of the lattice spacing, with the mass of the heavy quark close to the physical value of the b-quark mass (and charm physics at the physical value of the c-quark with substantially reduced discretization errors).

A further reduction of the systematic errors will come from the combination of the increased computing power with theoretical developments, such as the better control of the normalization of lattice composite operators and from systematic studies of "improvement", as well as from extrapolations to the infinite volume and zero lattice spacing limits (the large range of lattice volumes and spacings allowed by more powerful computers will be particularly valuable).

4.1 Calculation of the strong coupling constant α_s

Several determinations of the strong coupling constant from lattice calculations, obtained using a variety of methods, have already appeared in the literature both in the quenched and in the unquenched case. Errors as small as 2 % have been presented, but these are likely to be underestimated. Indeed, central values from different simulations differ by 2–4 standard deviations when the authors include estimates of the effects of quark loops. These discrepancies indicate that the errors have often been underestimated and that the present uncertainty in the physical value of α_s is most probably of the order of 10 %. It should also be added that present attempts to infer the physical (unquenched) value of α_s from extrapolations based on our perturbative knowledge of the running of the coupling constant with zero quark flavours are instructive, but do not lead to a reliable control of the systematic errors.

Two issues are crucial for an accurate determination of α_s, namely the renormalization of the lattice bare coupling constant and accurate unquenched calculations. In the quenched case, it has been shown that in order to achieve an accuracy of $1-2$ percent in the renormalization

Quantity	Present error	2000-2003	2003-2006
$\alpha_s(M_z)$	10 %	5 %	3 %
$\overline{m}_{u,d,c,s}$	25 %	10 %	5 %
$\overline{m}_b$	5 %	2.5 %	1 %
$\hat{B}_K$	15 %	10 %	5 %
Other $\hat{B}$'s	30 %	10 %	5 %
f_{D,D_s}	15 %	10 %	5 %
f_{B,B_s}, $f_B\sqrt{\hat{B}}$	20 %	15 %	7 %
$F(0)^{D\to M}$	25 %	10 %	5 %
$F(0)^{B\to M}$	35 %	15 %	7 %

Table 2 $\alpha_s(M_z)$, quark masses, $\hat{B}$-parameters, decay constants and form factors at zero momentum transfer from lattice QCD. The errors in the second column are present estimates. In the third and fourth column short- and long-term estimates of the future errors are given.

(and hence in the determination) of the strong coupling constant, a total of 10^{17} floating point operations is necessary. This corresponds to about three to six months of a substained 10 GFlops machine. The corresponding unquenched calculation, with two or four flavours, would exceed 10^{19} operations. In order to reduce the running time to a few months, a machine with a power of a few TFlops is required.

With a few TFlops machine the masses of the quarks in the fermion loops would still be, however, larger than the experimental value, thus leaving a residual systematic error in the determination of α_s. This systematic error may be studied, and possibly reduced, by varying the value of the quark masses in the fermion loops.

4.2 Determination of quark masses

Quark masses are among the fundamental parameters of the QCD Lagrangian. Moreover, the masses of the heavy quarks are important for theoretical predictions of heavy hadron lifetimes and their inclusive semileptonic widths (used to determine the corresponding CKM matrix elements). In order to discuss lattice determinations of quark masses it is convenient to classify them as light ($m_q \ll \Lambda_{QCD}$) or heavy ($m_Q \gg \Lambda_{QCD}$). The up, down and, to some extent, the strange quark belong to the first category; the b quark to the second one. The charm quark is neither light nor really heavy, and for this reason its study requires special attention, as discussed below. In both cases, the bare quark masses are determined from the study of the spectrum of hadrons composed of different quark flavours, with the help of the relevant Ward identities.

For light quarks, non-perturbative procedures have been introduced in order to obtain the short-distance renormalized quark mass. These techniques, together with the improvement of the lattice fermion action and operators, have allowed a determination of the quenched strange quark mass (m_s) with a precision of a few percent, corresponding to an error of about 5 MeV. A similar accuracy has been reached in the determination of the average light-quark mass, $m_u + m_d$. An indication of the errors due to quenching comes from the observation that the values of m_s obtained in quenched simulations using M_K or M_Φ differ by about 20 MeV. Very little is known directly, however, about quenching errors. Although there is some preliminary indication that the values of the masses decrease in the unquenched case, the magnitude of this reduction is controversial. From a comparison of different quenched and unquenched calculations, we estimate that the relative error of the physical quark masses to be about 25 %.

Quenched results for the mass of the b-quark have been obtained using both NRQCD and the HQET at lowest order in the heavy quark expansion. Recently these calculations have been extended to the unquenched case, with two light quarks in the fermion loops. The renormalized quark mass is obtained by using perturbation theory. In the case of the HQET the corrections are known at two loop order and will soon be extended to three loops. The estimated error coming from the truncation of the perturbative series has been estimated to be $\lesssim$ 100 MeV and will become $\lesssim$ 50 MeV when the three loop calculation is completed. To go beyond this level of precision, it will be necessary to include terms of higher orders in the heavy quark expansion which are expected to be of $\mathcal{O}(\lambda_1/m_b^2) \lesssim 30$ MeV, where λ_1 is the matrix element of the kinetic energy operator.

In the charm case, the applicability of the HQET or NRQCD is doubtful, since m_c is of the same order as the strong interaction scale $\Lambda_{\rm QCD}$. On the other hand, calculations performed with fully propagating charm quarks are afflicted by large discretization errors, since the Compton wavelength of the quark is comparable to the lattice spacing. This is true even when using an improved action (for which the errors are of $\mathcal{O}(m_c^2 a^2)$) since $m_c a \sim 0.25 \div 0.30$ at current values of the inverse lattice spacing a. These errors will be reduced very significantly with a 10 TFlops machine, which will allow simulations with values of the lattice spacing which are three times smaller than at present.

4.3 Hadronic Matrix Elements

In this subsection we review the computation of several important quantities of phenomenological interest, and where possible we compare the results with experimental measurements. We restrict the discussion to those quantities for which lattice computations are reasonably well established, and for which the precision will improve as available computing power is increased (see table 2). We postpone the discussion of new and exploratory computations until the following subsection.

1. Lattice predictions for $\hat{B}_K$ and $f_{B_d}\sqrt{\hat{B}_{B_d}}$ are widely used in phenomenological studies intended to constrain the parameters of the Standard Model. With the improved experimental accuracy, it is now even possible to extract their values from the analysis of the unitarity triangle. One typically finds $\hat{B}_K = 0.90^{+0.26}_{-0.16}$ and $f_{B_d}\sqrt{\hat{B}_{B_d}} = 224^{+23}_{-25}$ MeV. These numbers, extracted from experimental data, represent a significant success for lattice calculations since they are in very good agreement with lattice predictions. In the future, in order to further constrain the unitarity triangle, a reduction of the present theoretical uncertainties will be necessary, requiring a substantial increase of computer power (see table 2).

2. Some preliminary information on the effects of unquenching for light-quark masses and for $\hat{B}_K$ has already been obtained. On the other hand very little is known about the

quenching errors in calculations of the decay constants, the $\hat{B}_B$ parameters and the semileptonic form factors. A systematic study of discretization errors in this sector will also help to reduce this source of uncertainty.

3. In supersymmetric extensions of the Standard Model, K^0–$\bar{K}^0$ mixing can receive substantial enhancement from the presence of new $\Delta S = 2$ operators present in the effective Hamiltonian. Recently, the first calculations of the matrix elements of SUSY operators relevant for K^0–$\bar{K}^0$ and B^0–$\bar{B}^0$ mixing have been made. Unquenched calculations of these amplitudes have not started yet and no attempt to extrapolate the values of the matrix elements to the continuum limit has been done. We expect that errors on these matrix elements can be reduced to less than 10 % using a 10 TFlops machine.

4. The most accurate determination of $|V_{ub}|$ at present is that obtained from the semileptonic inclusive differential rate $d\Gamma/dM_X$ where M_X is the final state hadronic mass. One of the main uncertainties (probably underestimated in theoretical calculations) originates from our ignorance of the shape function, which is discussed below. On the other hand, the extraction of $|V_{ub}|$ from exclusive decays is presently limited by the uncertainties on the relevant form factors. For example, one of the main sources of error in the evaluation of the form factors for $B \to (\pi, \rho)\ell\nu_\ell$ semileptonic and $B \to K^*\gamma$ radiative decays is the extrapolation of the results to small values of the momentum transfer q^2. Increased computer power will significantly reduce this source of errors. Exploratory unquenched calculations are also necessary but have not been attempted so far.

5. $B \to (D, D^*)\ell\nu_\ell$ semileptonic decays are used to determine $|V_{cb}|$ from the extrapolation of the differential rate to the point of zero recoil. The calculation of the slope and curvature of the Isgur-Wise function can be improved with respect to previous lattice studies, in particular by using methods which extract directly the slope from lattice correlation functions. In these approaches, the largest systematic error is induced by finite volume effects, and an increase in computer power is very important. A real challenge for lattice calculations is the determination of the form factor normalization at maximum recoil, since a precision of a few percent is required for this quantity. Without a substantial improvement in computer power, systematic effects are likely to remain larger than the required precision.

4.4 New Challenges

In this subsection, we briefly discuss and illustrate some physical quantities for which only preliminary calculations exist, or for which theoretical ideas have been presented but exploratory computations are still to be performed. In many cases, in order to get precise results for the hadronic amplitudes one needs theoretical developments as well as substantial computing power. An accurate determination of the renormalization constants (with non-perturbative methods) and improvement of the operators (or an extrapolation to the continuum) is necessary.

It has been proposed to extract the matrix elements of local operators using lattice T-products of suitable currents (operators) at short distances. This approach avoids the problems related to power divergences and the difficulties in obtaining the renormalized operators present in all effective theories (including the operator-product expansion in deep inelastic scattering, effective Hamiltonians in weak decays and the HQET). With this technique, for each quantity one has a large scale Q which should lie in the window $\Lambda_{\rm QCD} \ll Q \ll a^{-1}$ (where a is the lattice spacing). The main problem is to ensure that such a window exists, and for this a factor of 100 in computer power is crucial.

1. We start by discussing non-leptonic $K \to \pi\pi$ decays which are fundamental for our understanding of CP violation and of the strong interaction dynamics in hadronic weak decays. The comparison of the theory with the experimental measurements of ϵ'/ϵ by the KTeV and NA48 Collaborations is hampered by our ignorance on the relevant operator matrix elements. ϵ'/ϵ is one of the most difficult quantities to compute because of large cancellations between different contributions. In addition to similar difficulties to those present in the calculation of ReA_0 (see following paragraph) with conventional Wilson fermions this quantity demands a calculation in the presence of heavy quarks, thus requiring a smaller lattice spacing. With staggered fermions there are huge one-loop perturbative corrections to the relevant penguin operator Q_6, which make the results unreliable. With domain-wall fermions, the only existing calculation of the matrix element of Q_6 is in contradiction with all other theoretical estimates, and is difficult to accommodate with the experimental measurements, even invoking new physics beyond the Standard Model. Moreover, estimates of the amplitudes obtained by computing the matrix element $\langle \pi|\mathcal{H}_W|K\rangle$ (which is possible as a result of the soft-pion theorems) lead to large uncertainies in the prediction of ϵ'/ϵ because of the significant cancellations between the matrix elements of Q_6 and of the electropenguin operator Q_8. New theoretical ideas and strategies to compute these matrix elements on the lattice have been presented. They all require large computing resources to reduce the present uncertainties significantly.

2. ReA_0, the real part of the $\langle \pi\pi|\mathcal{H}_W|K\rangle$ amplitude with the isospin of the final state $I_{\pi\pi} = 0$, has never been computed in a reliable way or with sufficient numerical accuracy. Several new proposals to compute this quantity exist but have not been attempted yet. Most of these require a very large statistical sample and/or a fine-grained lattice. It is also expected that one needs rather light quark masses for a realistic calculation. A precision of 30 % on the amplitude would already be considered a success. The calculation of the amplitude $\langle \pi\pi|\mathcal{H}_W|K\rangle$ with $I_{\pi\pi} = 2$ is certainly feasible and has been recently computed on large lattices (up to $32^2\times 64$). The results are not fully convincing, however. In particular, the result requires the use of chiral perturbation theory. Some indication of the precision of chiral perturbation theory can be obtained by applying it to obtain the $I_{\pi\pi} = 2$ amplitude from the lattice results for $\hat{B}_K$. In this way one obtains a value for the $K \to \pi\pi$ amplitude which is 40% larger than the experimental value.

3. The evaluation of $K \to \pi\pi$ amplitudes can be attempted using the short distance method mentioned above. The requirement of a short-distance window implies that the lattice spacing a must be small, ($a^{-1} \geq$ 3–5 GeV). This method has not been tried in QCD yet, but encouraging results have been obtained with the non-linear σ-model in two dimensions. The effect of discretization errors still needs to be understood. In any case, the problem of the calculation of complex amplitudes, when two or more particles are present in the final state, is still an open one which demands further theoretical investigation.

4. The computation of the amplitudes for $\langle\pi\pi|\mathcal{H}_W|B\rangle$ decays with $I_{\pi\pi}=2$ and other B-decays without penguin diagrams are feasible, but need an extrapolation in the mass of the heavy quark. These amplitudes can provide very interesting information on factorization and isospin breaking effects.

5. The evaluation of the amplitudes for $\langle\pi\pi|\mathcal{H}_W|B\rangle$ decay with $I_{\pi\pi}=0$, and other B-decays with penguin diagrams have all the difficulties of kaon decays, with in addition the presence of a very heavy quark (representing the effect of the top-quark in the loops).

6. A good testing ground in the calculation of non-leptonic amplitudes is provided by charmed-hadron decays, for which very accurate determinations of the different branching ratios exist and for which the phases have also been measured in some cases.

7. The $\mathcal{O}(1/m_Q^3)$ contributions to heavy-hadron lifetimes is an important issue, related to the validity of the HQET for inclusive decays. In particular it is important to find an explanation of the small experimental value of the lifetime ratio $\tau(\Lambda_b)/\tau(B)\sim 0.78$ which, on the basis of the HQET, is expected to be $\geqslant 0.9$. Calculations of the relevant matrix elements are still in their infancy. A related problem is the width difference of neutral B_s mesons.

8. There is a proposal to compute the "shape function" relevant for inclusive $b\to(u,c)\ell\nu_\ell$ semileptonic and $b\to s\gamma$ radiative decays. The approach relies on the study of weak-current T-products at large momentum transfers. Similar techniques have been proposed for the calculation of light-cone wave functions.

9. Hadronic structure functions have been measured experimentally with remarkable precision, most recently at the HERA electron-proton collider. It remains a major and important challenge for the lattice community to contribute competitively to this field. Lattice computations of structure functions have been extensively performed in the recent past. All the calculations, in spite of the non-perturbative renormalization of the relevant operators and of the use of improved actions and operators, give a value of the momentum carried by valence quarks which is larger than the experimental value. It will be important to confirm that this value is reduced when unquenched calculations are performed. The short-distance method could also be used in this case.

10. There are a number of quantities for which lattice calculations have only been attempted once or twice or which have never been tried yet. These include the evaluation of the electric dipole moment of the neutron and the matrix elements relevant for proton decay and n–$\bar{n}$ oscillations.

5 QCD thermodynamics

In QCD a new state of matter, the quark gluon plasma (QGP), is expected at very high temperature and/or baryon density. The extreme conditions where such a phase exists were realized in the early universe and may exist in the interior of neutron stars. The theoretical description of these states is based on QCD thermodynamics. A considerable experimental effort, particularly at RHIC and at the LHC, is committed to observing the transition into this new phase in relativistic heavy ion collisions. For the interpretation of these experiments one needs predictions from QCD at finite temperature. In the temperature region which can be explored, corresponding to temperatures which are at most a few hundred MeV, perturbation theory has large systematic uncertainties and the same is true for effective models based on QCD. For reliable quantitative information for QCD at finite temperature one has to use lattice simulations.

Among the quantities which have been studied up to now, the most important are

- the equation of state
- the value of the transition temperature
- the order of the phase transition (or cross over)
- the screening lengths in the plasma phase
- thermal effects on hadronic masses, widths and decay constants.

There are some features which are specific to finite temperature lattice calculations. At finite temperature the quenched approximation is not quantitatively a good approximation. Dynamical fermions give contributions, which are similar in magnitude to those from the gluons. Furthermore it is important to keep as far as possible the chiral properties of the continuum action, at least in the neighborhood of the transition. One should also note that several of the interesting thermodynamic quantities scale with the fourth power of the inverse lattice spacing, which demands considerable statistics to obtain a precise determination as one approaches the continuum limit.

Nevertheless, very significant progress has already been achieved. The bulk properties of the pure gluon theory (properties of the transition and the equation of state) have been calculated and extrapolated to the continuum limit with systematic errors estimated to be a few percent. A lot of qualitative information has been obtained on the properties of full QCD at finite temperature, and reliable estimates can be made of the computing power needed to obtain a similar precision to that which one now has in the pure gluon theory ($O(10\ \mathrm{TFlops})$).

Simulations of QCD at finite density is a major challenge and conceptual progress is required to obtain quantitative results. The main problem is that the action becomes complex even in the Euclidean formulation, which means that straightforward Monte Carlo methods do not work. Theoretical studies are in progress, as well as simulations of field theories which are expected to have some common features with QCD (e.g. to have superconducting phase(s) at large densities) but for which the action is real.

5.1 The equation of state

The basic tools needed to study the thermodynamics of strongly interacting matter have been developed over the last 15 years. In particular, the methods to study the equation of state at finite temperature have greatly advanced during this time. In the pure gluon sector, typical lattice problems (finite size effects, finite cut-off effects) discussed above have been studied in detail. Controlled extrapolations to the continuum limit are now possible and have been performed. Today we thus know fairly well, with systematic errors of a few percent, the equation of state of purely gluonic matter and its approach to the equation of state for an ideal gluon gas in the high temperature limit, as well as basic thermodynamic parameters such as the critical temperature T_c (in units of the string tension), the ratio ϵ_c/T_c^4 (where ϵ_c is the energy density at the phase transition of the quark gluon phase), and the latent heat. These quantities will be discussed in more detail in the following two sections.

Exploratory studies of the equation of state have also been performed for QCD with light dynamical fermions. These investigations are most advanced in the staggered fermion formulation, where one has a remnant of chiral symmetry.

One can estimate that in order to obtain the temperature dependence of the energy density, pressure and entropy in the QGP phase in full QCD with a similar precision to that currently achieved in the pure gluon theory, using staggered fermions, will require about one TFlops-year.

5.2 The value of the critical temperature

The value of the critical temperature in physical units is particularly important in order to know if the experiments can produce a sufficient energy density to be in the QGP phase. Since the energy density is proportional to the fourth power of the temperature, a 10% uncertainty in the transition temperature leads to a 50% uncertainty in the energy density. In the pure gluon theory the value of the critical temperature in units of the square root of the string tension is known to within 3% or so (this systematic error is estimated from the difference of the two most precise measurements). If one would put in the physical value of the string tension, one would obtain $T_c = 270(10)\,\mathrm{MeV}$.

In full QCD with two light flavours we are aiming at a similar precision. Up to now, different discretization ansätze lead to values which are barely consistent. However, it already seems clear that the transition temperature is in the range between 150 – 200 MeV, considerably lower than in the pure gluon theory. In order to determine the critical temperature with a 5% accuracy, one requires of the order of 10 TFlops-years. This is because one needs reliable results for zero temperature quantities for normalization and also to check the sensitivity to various lattice formulations of the fermionic action.

5.3 The order of the transition

In the pure gluon theory the phase transition is first order. It is however, rather weak in the sense that the latent heat and the bubble surface tension are small. The ratio of the entropy in the QGP phase above the transition to the same quantity in the confined phase is rather large, however. All these quantities can be determined on the lattice in the pure gluon theory and extrapolated to the continuum with errors of a few percent. If the results were indicative of the corresponding physical values, one would deduce that in heavy-ion experiments, where the sharp change of the entropy is more significant than the actual latent heat because of the finite size of the system, the change in behaviour when entering the quark-gluon phase would be quite clear. In the early universe, however, the transition is too weak to lead to significant effects, in particular the bubble formation does not lead to a significant nonhomogeneity.

In full QCD it is expected from effective models that the order of the transition is crucially dependent on the number of light quark flavours, being first order for three massless flavours and second order for two massless flavours (if the $U(1)_A$ anomaly is still significant at the transition temperature). Simulations up to now indicate a second order transition for two flavours, although the values of the critical exponents do not agree with those from the effective models. It appears that in order to determine the order of the transition in full QCD it is important to preserve as much chiral symmetry in the action as possible and to simulate with very light quarks. Estimates indicate that a computer in the 10 TFlops range is necessary to achieve this.

5.4 The screening lengths in the QGP phase

One of the most interesting possible signals for the existence of the QGP in heavy ion collisions is the suppression of heavy quark bound states, like the J/ψ. These bound states are supposed to dissolve in the QGP because of colour screening. This may not happen at the transition, but at a higher temperature, depending on how strongly they are bound. To predict this temperature, lattice measurements of the relevant screening lengths are needed. They can be obtained as a "by-product" during the determination of the equation of state, discussed above. As discussed in section 3, large statistics are required, because correlation functions with disconnected diagrams have to be calculated.

5.5 Thermal masses and decay parameters

The analysis of the temperature dependence of hadronic masses requires, at least in part, a different set up for the numerical simulations from that used for the quantities discussed above. Hadronic screening masses extracted from the spatial correlation functions can be directly determined in the same project. However, the calculation of pole masses requires the analysis of time-like correlation functions. With increasing temperature these correlation functions, which contain information about all states with given quantum numbers, receive a growing contribution from the excited states. Therefore lattices, which have many points in the Euclidean time direction are needed. A possible approach to this problem is to use anisotropic lattices, where the anisotropy is introduced by different lattice spacings in the discretized actions. The corresponding computing requirements are then similar to those for the determination of the corresponding quantities at zero temperature.

5.6 Finite baryon density

In the last year, several studies of effective models indicate a very interesting and rich phase structure with superconducting phases at low temperature and high density. This should be relevant for the physical properties of neutron stars. As discussed in the introduction, there is a major obstacle to obtaining results on the thermodynamics at finite baryon density in QCD, because the action is not real even in the Euclidean domain. Therefore conceptual progress is needed and it is difficult to estimate quantitatively what can be achieved. There are, however, models where the action is real and these may give insights into the problem.

It may be possible, using present methods, to extrapolate into the region of finite density at high temperature, thereby obtaining results relevant to heavy ion collisions. This will require computing power in the TFlops range.

6 Theoretical issues and Non-QCD Physics

Throughout most of this report we focus on QCD simulations, but lattice studies of other quantum field theories also yield very interesting results. In this section we briefly review a few examples of such applications as well as discussing some theoretical issues in QCD.

6.1 Electroweak physics: QED, SM, MSSM

QED on a lattice is interesting for several reasons. One can study the phase structure of this Abelian gauge theory. In contrast to QCD, it is not asymptotically free so that one can investigate the existence of a Landau pole in the non-perturbative regime. This is an extremely important issue for the high energy behavior of non–asymptotically free sectors of physics. The main problems in this field are conceptual and computational resources are not the limiting issue, so we will not discuss them further.

More generally in electroweak physics there are a number of interesting questions related to

- the electroweak phase transition
- heavy fermions (strong Yukawa coupling sectors)
- vacuum stability and mass bounds
- the heavy (strongly interacting) Higgs sector

For some of these problems, in principle, the proper handling of chiral fermions is crucial and recent progress in understanding chiral symmetry on the lattice is very important in this context (it has recently been shown that the Nielsen-Ninomiya theorem can be obviated by relaxing some of the assumptions).

The issue of the electroweak phase transition in the evolution of the early universe and its relevance for the baryon asymmetry in the universe has initiated a lot of activity into extending the perturbative regime valid for light Higgs masses towards the non–perturbative region of heavier Higgs masses by means of lattice simulations. Recent simulations in the bosonic sector yield a first-order phase transition line in a (T_c, m_H)–plot, which is essentially linear, passing through $(90, 50)$ GeV and ending at about $(110, 75)$ GeV in a second-order endpoint. Simulations with fermions are expected to yield very similar results. Together with the direct experimental bound of $m_H \geq 95$ GeV this result implies that physics beyond the SM is needed to account for the observed baryon asymmetry. As a next step one may ask whether the MSSM would to able to rescue the situation. Due to the large parameter space this question is much more formidable and would require substantially more computational resources. Other aspects of electroweak baryogenesis which may be studied on the lattice are sphaleron rates and spontaneous CP violation.

6.2 Supersymmetric Yang Mills theories

Supersymmetric Yang–Mills (SYM) theories allow one to study very interesting non–perturbative features because of the high degree of symmetry that these theories exhibit. The access to non–perturbative physics by analytical methods is based on the observation that in SYM theories strong coupling properties may be obtained by analytic continuation of weak coupling results (holomorphic structure). This has triggered a revolution in understanding four dimensional gauge theories. By now, many non-perturbatively exact statements have been obtained about various types of SYM theories. These exciting developments were initiated by the work of Seiberg and Witten on the $N = 2$ supersymmetric $SU(2)$ Yang–Mills theory. By exploiting duality symmetries between electric and magnetic, or weak and strong coupling sectors (Seiberg–Witten duality Ansatz) it was possible to write down an exact solution for the non-perturbative low energy effective Lagrangian. The subsequent developments have dramatically deepened our understanding of non–perturbative features of gauge theories, although it is not yet clear what the implications are for non-supersymmetric theories like ordinary QCD.

These developments make the investigation of SUSY theories via other non–perturbative approaches, such as the Monte Carlo simulation of the corresponding lattice model, very exciting. This, however, is a formidable task (because of technical problems with putting SUSY on a lattice, the large number of degrees of freedom etc.). Therefore one has to investigate simpler systems first, like the $N = 1$ supersymmetric Yang–Mills theory. In this case non–perturbative results are not rigorous but fit into a plausible self–consistent picture of low energy dynamics of supersymmetric QCD (SQCD). There are still mysteries which remain to be understood. It is expected that SQCD exhibits confinement and spontaneous chiral symmetry breaking (SχSB) much as in QCD. This requires that a gluino condensate is produced by the dynamics. Even for these simpler theories lattice studies are only just beginning. The simulation of $N = 1$ SUSY requires the implemention of Majorana fermions on the lattice, which can be done by utilizing the multi–boson algorithm. Consequently, the computational requirements are comparable to the simulation of full QCD (with dynamical fermions): e.g. the $SU(2)$ theory on a $12^3 \times 24$ lattice requires of the order of a month's running on a machine with 50 GFlops sustained performance. Such simulations have shown that the model indeed exhibits a spontaneous discrete chiral symmetry breaking caused by a gaugino condensate. The SUSY point is characterized by a vanishing gluino mass. Many questions of interest require small fermion masses and hence are expensive in simulations. Like chiral symmetry, supersymmetry is broken on the lattice in standard approaches and must be recovered somehow in the continuum limit. A scaling up by a factor 2 is needed in order to gain a quantitative insight into these problems and this would require half a year's running at 1 Tflop sustained. With such computing power, one could answer many of the important questions (mass spectrum, SUSY potential, Ward identities). For $N = 2$ SUSY only the first steps are possible. The parameter space is much more complicated and requires a substantial effort to explore, for example, the phase diagram boundaries. In this case no final answers can be expected soon, but as a longer term goal it is very interesting and important.

6.3 Other interesting phase transition problems in QFT

In general phase transition problems are intrinsically non–perturbative in nature and in the absence of analytical methods require simulation by Monte Carlo methods. The existence of a phase transition of order $n \geq 2$ (critical point) is directly related to the existence of a continuum limit. This could be very important in connection with the questions related to extensions of the SM, which could shed new light on parameter patterns like CKM matrix elements, mass spectra and the like.

A major challenge is the simulation of gauge theories with fermions, where phase transitions show up as one increases the number of flavours N_f. Questions arise concerning the existence of a flavour violating phase or a parity violating phase. Aoki in a lattice QCD simulation found a flavour and parity violating phase in QCD with $N_f = 2$ flavours which is in contradiction with the Vafa–Witten theorem. More elaborate simulations which will be possible with Tflops computers will allow us to settle this important question concerning the structure of the QCD vacuum. Absence of spontaneous parity breaking is essential for the viability of the axion approach to the strong CP problem and its phenomenological consequences.

The question of spontaneous violation of parity is particularly interesting and crucial for technicolor and preon models. Simulation with non–standard boundary conditions, the study of the massless spectrum and of the clustering properties are special requirements here.

In QCD with many flavours an infrared stable strong coupling fixed point is expected for $9 \leq N_f \leq 16$. Such simulations are also feasible with the multi–boson algorithm where N_f may be taken to be a continuous variable (exact algorithm). However, the number of required boson fields increases with increasing N_f. For QCD with $N_f = 3$ one requires 64 to 80 fields instead of 24 to 32 in the case of the $N = 1$ supersymmetric $SU(2)$ Yang Mills theory, which we discussed in subsec. 6.2; this results in a slowing down by a factor of 3 to 6. Such studies will lead to a reasonable understanding of how the phase structure of QCD depends on N_f.

6.4 Chiral Fermions

There have recently been very promising developments in implementing chiral fermions on a lattice. For QCD, where the extrapolation to light quarks is crucial, lattice chiral formalisms with Ginsparg-Wilson symmetry offer great potential advantages. This symmetry acts to preserve the good properties of chiral symmetry in the continuum and so there is no additive renormalization of the quark masses. This is a very promising approach with, however, one big drawback: all formalisms that are currently known which implement this approach (perfect action, domain wall, Neuberger action) involve a considerable increase in computation in evaluating the fermion operator (by 20-100 times). This will be offset in part by the improved behaviour at small quark masses, but it is still a significant additional computational effort. As yet no comprehensive studies of hadron spectroscopy and matrix elements using one of these formalisms have been carried out - so a quantitative estimate of computational resources is not possible.

6.5 Chiral condensates in QCD

Chiral symmetry breaking in gauge theories (QCD) imposes strong constraints on the effective low energy behavior of the light degrees of freedom (pions, kaons, etc.). It leads to the framework of effective chiral Lagrangians and chiral perturbation theory (χPT). Both the spontaneous chiral symmetry breaking and the explicit breaking of chiral symmetry due to the $U(1)$ anomaly can be used to make exact predictions for the properties of light quarks. Among the most important phenomenological parameters of the χPT approach is the quark (or chiral) condensate which triggers the SχSB. This non-perturbative quantity can be calculated numerically on the lattice, although reliable calculations were hampered until recently by the difficulties associated with putting chiral fermions on a lattice. With the advent of the Ginsparg-Wilson fermions more reliable simulations are possible now in the quenched approximation. In any case the investigation of the chiral condensates is of fundamental importance both for the structure of χPT and for the chiral extrapolation of lattice results, which have been done on a purely phenomenological basis so far. With the help of a Teraflops machine all relevant questions could be settled in the quenched approximation. The future development in full QCD depends on progress in the implementation of the chiral fermion simulations.

6.6 Quantum gravity

Quantum gravity describes space–time dynamics and therefore requires the consideration of random geometries. In lattice simulations this implies a random distribution of data in the memory and hence differs dramatically from normal lattice field theory simulations on a fixed lattice. The primary problem is the investigation of the phase structure, because the existence of a continuum limit is directly related to the existence of a 2nd or higher order phase transition.

The formidable task of attempting to simulate realistic $4D$ gravity problems is currently approached in steps:

- Simplicial gravity (triangular lattices, no matter): exhibits a simple parameter space and the main issue here is to establish the existence of a critical point and then to perform the construction of the continuum theory.

- More realistic modeling requires the addition of matter which implies a much more complicated parameter space. Exploring criticality is much more tricky in this case and all kind of scaling laws have to be established in order to be able to make contact with real physics.

By its basic structure quantum gravity is not suitable for parallelization! The natural relationship is " one processor $\leftrightarrow$ one copy of the geometry" and hence a few powerful processors are most suitable for present studies. The typical feature needed is dynamical repartitioning of the processors (integer arithmetic) and processors like the Intel Pentium are ideal for these kind of problems.

6.7 Classical electrodynamics (MAFIA package) and accelerator physics

We conclude this section by presenting one example which is not related to quantum field theory, but which is of interest for the particle physics community.

MAFIA is a widely used package for the solution of MAxwell's equations using a Finite Integration Algorithm and implementation of classical electrodynamics on a lattice. It is used heavily in the design of structures for new particle accelerators like TESLA at DESY. A typical application is the simulation of a superconducting cavity for TESLA. However realistic 3D simulations are not feasible presently on clusters of high end workstations. Therefore a project started recently at DESY Zeuthen aiming to implement part of the MAFIA code on the massively parallel APE 100 computers. For systems of about 20×10^6 grid points a gain by a factor 20 was already achieved in comparison with high end workstations. Presently the limitations are dictated by the available memory. Full 3D simulations are possible now for accelerator cavities etc.

Another more challenging problem in accelerator design is the global accelerator structure, beam dynamics calculations, polarization stability investigations and particle tracking in linear accelerators. What is needed is a self-consistent calculation of the beam in the accelerator structure. Corresponding computer codes were ported to massively parallel processor (MPP) computers and MPP's potential was reported to outperform high end workstations by factors of 10^2 to 10^3. These kind of developments are of great interest for the ongoing linear accelerator design.

7 Algorithms and computing requirements

When it comes to the realization of grand challenge lattice computations, such as full QCD simulations in the deep chiral regime, research into improved or novel algorithms is of similar importance to the building of high efficiency dedicated parallel computers. In fact there is a delicate interplay between improvements in the performances of algorithms and machines.

7.1 Some aspects of HMC

Within the stochastic sampling techniques (SST) which produce ensembles of QCD vacuum configurations most of the computational work goes into the numerical representation of long range interactions as effected by the fermionic determinant. At present the preferred sampling technique is Hybrid Monte Carlo (HMC) which is a combination of deterministic and stochastic evolution through the configuration phase space of quantum fields. In this process HMC makes heavy use of quark propagator computations (inversions). Presently, because of the well-known sign problem of the determinant we are limited to simulations with even numbers of degenerate flavours.

Most lattice simulations are performed with Wilson fermions (or improvements thereof). The values of quark masses are determined by tuning a bare parameter of the action (for each quark flavour), called the hopping parameter κ. The value of κ corresponding to massless quarks is called κ_c, the critical kappa. The lattice Dirac operator, M, becomes ill conditioned in the light quark mass regime with the additional problem that κ_c is fluctuating from configuration to configuration, on finite lattices. This not only causes critical slowing down of propagator computations but might induce serious stability problems when approaching the critical point within the characteristic range of these fluctuations. The severity of this problem depends on the choice of action and the numbers of lattice degrees of freedom.

So far, with Wilson fermions, no major full QCD simulation has ever been carried out with sufficiently light quarks so that the ρ-meson can decay into two pions. This would require larger lattices than are feasible so far – the pion correlation length (the inverse of its mass) should not exceed 25 % of the spatial lattice dimension to avoid finite size effects.

Autocorrelations with Wilson fermions. Algorithmic research in QCD is hampered by the fact that reliable verification of the performance of a proposed new algorithm is generally very time consuming. The reason is that the true power of an algorithm can only be understood following some reliable test of its decorrelation efficiency along the Markov process. This requires very long runs. In this sense it is extremely difficult to extrapolate from our experience on the current generation of computers to predict performances on the multi-Tflops machines, without the experience of using intermediate machines such as APEmille. It is not unreasonable to suggest that 20 % or so of computing time should be devoted to algorithmic research as an investment for the future.

At this stage there is some preliminary information on autocorrrelations of the HMC algorithm, obtained with the standard, unimproved, Wilson action. Autocorrelation time depends on the observable being studied and the slowest modes are expected to occur in nonlocal quantities. The autocorrelation study was based on the MC time dependence of the convergence rate of the inverter (BiCGstab with lexicographic SSOR preconditioning). Convergence is of course sensitive to the lowest part of the spectrum of the Dirac operator M, and therefore to the global structure of field configurations.

7.2 Improved actions

"Improvement" is the name given to modifications of the lattice action (and operators) introduced so that the errors due to the finiteness of the lattice spacing (the discretization errors) are reduced. In principle this allows us to operate on coarser lattices than in unimproved simulations. Improved actions tend to be less local, and this naturally affects the performance of the algorithms. A priori, it is not obvious what size of fluctuations is induced into the low lying part of the spectrum of the Dirac operator by improving the action and coarsening the lattice.

None of big projects simulating full QCD that use improved Wilson actions have so far been able to go below the ρ decay threshold even in simulations on coarse lattices. Evidently more research must be done in this direction.

7.3 Improving on the chiral limit

The past 18 months have witnessed very active research in connection with variant forms of the Wilson action that obey the Ginsparg-Wilson (GW) relation and which therefore possess a lattice form of chiral symmetry.

Domain wall fermions constitute a particular realization of this idea which amounts to adding a fifth dimension to the lattice, and this approach is currently being explored in some detail. The advantage is that one has control of the location of the critical point (no additive quark mass renormalization) but at the price of a factor of at least 20 in simulation effort. At this stage it is not clear how this cost factor will rise as one decreases the value of the quark masses.

Another approach to satisfying the GW relation is to adopt a nonlocal form of the Wilson action from the very beginning, one which has all its eigenvalues on a unit circle. Seen as a numerical analyst's problem this approach involves the repeated computation of $M^{-1/2}$ to high precision. At present no efficient way of doing this is known and hence this 'solution' to the problem of chiral symmetry on the lattice is only formal and not yet a practical one.

7.4 Cost predictions for standard Wilson fermions

We now make the brave attempt to extrapolate the costs of simulating full QCD with unimproved Wilson fermions from a recent study of the performance of the HMC algorithm. The runs were restricted to a single value of the lattice spacing, $a = 0.08$ fm (as determined from the mass of the ρ-meson). For other values of the lattice spacing one therefore has to resort to certain scaling assumptions. The following cost function was obtained for the number of operations required to generate an independent configuration, with all dimensionful quantities taken in the same units (e.g. in fm):

$$N^{\rm ops}_{\substack{\rm per\,indep.\\ \rm conf.}} \simeq 1.7 \times 10^7 \cdot (L^3 T)^{4.55/4} \left(\frac{1}{a}\right)^{7.25} \left(\frac{1}{m_{ps}}\right)^{2.7} . \tag{3}$$

The predicted effort required for sampling 100 independent lattice configurations with the HMC algorithm on a lattice of size 4 fm is plotted in fig. 3 in units of sustained Tflops-months. In place of the quark mass we prefer to select the pseudoscalar-vector mass ratio as the independent variable. At lattice spacings other than 0.08 fm there is an additional uncertainty in the expected computing costs (of approximately a factor of 2) due to the unknown a-dependence of the meson trajectories $m_v(m_q)$ and $m_{ps}(m_q)$. Clearly the uncertainties in the cost estimates, which are currently very substantial, will be significantly reduced as soon as the next generation of computers becomes available (e.g. the APEmille machine) to supplement present full-QCD runs.

As an example of how to read fig. 3 we present here one possible strategy for the year 2000. If one wishes to devote 1.5 TFlops months to full QCD simulations, then one could generate 100 independent configurations on a 4 fm lattice, with the rather coarse lattice spacing $a = .12$ fm, and quark masses corresponding to m_{ps}/m_v =.4.

8 Computing Technology

8.1 High-End Computers Currently Used for QCD

Lattice QCD pioneered the exploitation of parallel computers for large-scale scientific simulations in the early 1980s. Since then, cost-effective performance at the highest levels possible for QCD codes has been achieved through a variety of approaches, extending from purpose-

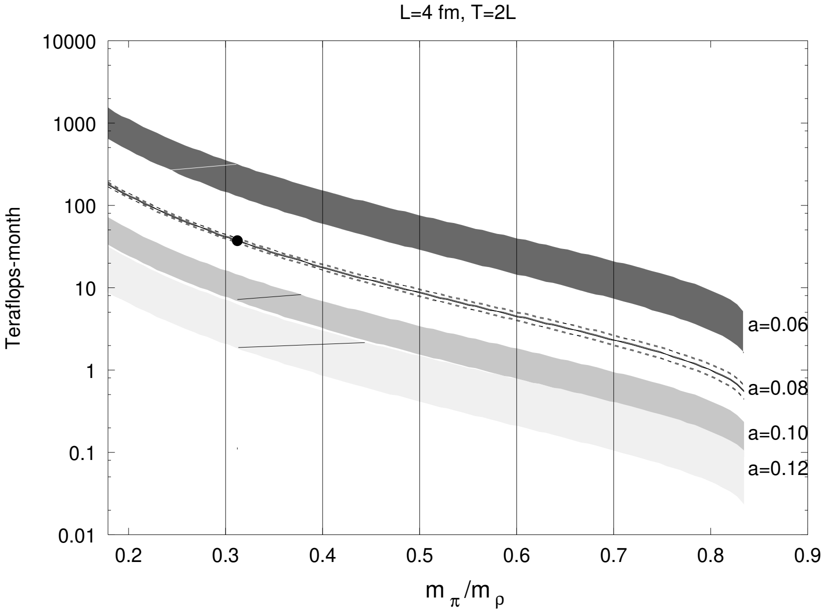


Figure 3 Cost estimate (in Teraflops-months sustained) according to Güsken et al., from the SESAM/TχL collaboration for the generation of 100 independent vacuum configurations vs. the ratio of pseudoscalar to vector masses, m_{ps}/m_v . The plot is for lattice size $L = 4$ fm. On the $a = .08$ fm curve, finite size effects would dominate to the left of the bullet where m_{ps}^{-1} exceeds 1 fm, i.e. a quarter of the lattice (on the remaining three trajectories with different a's, the bullet is replaced by the horizontal error bars).

built machines employing customised processors, to highly-optimised codes on commercial systems. Currently, the following systems are leading the field.

CP-PACS/Hitachi SR2201: This joint academic-industrial project began in Japan in 1991 with the goal of developing a massively parallel computer for research in computational science, with emphasis on QCD. The machine is an $8 \times 16 \times 16$ MIMD array of modified HP PA-RISC microprocessors, connected by a 3-dimensional crossbar network, with a peak speed of 614 GFlops and a distributed memory of 320 GB. It has been operational since 1996. Fortran 90, C, C++ and assembly language programming are available, and optimised QCD codes achieve 40-50% of peak speed.

QCDSP: The QCDSP machine is a four-dimensional mesh with nearest-neighbour communications. Each processing node contains a 32-bit 50 MFlops Texas Instruments digital signal processing chip, 2 MB of memory and a Node Gate Array (NGA) chip. The NGA was designed at Columbia University and provides a buffer to accelerate memory access, controls error detection and correction, and has 8 serial communication units. Software is written in C/C++ with critical routines written in hand-optimised assembler. QCD codes typically achieve around 20% of peak performance. The largest configuration is a 12,288-node machine, which has been operational at the RIKEN Brookhaven Research Centre since 1998. An 8,192-node system was completed earlier that year at Columbia.

Cray/SGI T3E: This is the most widely used commercial machine for lattice QCD. Systems varying in size from a few hundred to 1024 processors have been available in Europe and the US since 1997, usually in multi-user environments. The T3E-900 is based on Digital EV5.6 processors with a peak speed of 900 Mflops. The UKQCD HMC code achieves 30% of peak in 32-bit arithmetic using Fortran and MPI, with optimised assembler routines for SU(3) operations. The latest version is the T3E-1200E, in which the clock speed has been increased so that each processor has a peak speed of 1.2 GFlops.

APE100: This is a synchronous SIMD machine composed of a 3-dimensional cubic mesh of processors. Each node contains a custom 32-bit floating-point processor, 1 MB of memory, and a network interface device. The largest machine consists of 512 nodes in an $8 \times 8 \times 8$ configuration with a peak speed of 25 GFlops. Achievable performance for QCD is in the range $30 - 70\%$ of peak. Originally designed and built as an academic project by INFN and the University of Rome, the manufacture and distribution was undertaken by Alenia Spazio, and subsequently its subsidiary QSW, who marketed the machine under the name 'Quadrics'. Systems have been available since 1993. Outside Italy, the major APE100 installations are at DESY-Zeuthen and Bielefeld.

8.2 The Commercial High-End Computer Market

The computers designed by Seymour Cray, and their clones, dominated supercomputing for 20 years until the early 1990s. These machines were based on fast custom-designed, but esoteric, processors and, as such, attracted premium prices. Despite continual improvements in clock rates, the performance required by users gradually outstripped their ability to deliver, and massive parallelism emerged. Massively Parallel Processors (MPPs) use many standard RISC-based microprocessors connected together with proprietary networks and have been the mainstay of high-end computing for the last $5 - 10$ years. However, despite advances in software standards to improve code portability, they provide a complex programming model and, consequently, an entry-level barrier for new users. The small marketplace for such products, coupled with the still significant research and development costs, probably means that the end is in sight for MPP. They are likely to be replaced by clusters of shared-memory machines and/or some convergence towards using the same technology as in PCs.

8.2.1 US Strategy

The strategy of both US Government Labs and US vendors is to use commodity building blocks for both processors and communications to build distributed memory systems. Today's high-end machines have around 10,000 nodes. MPI is widely used as the software standard and many vector codes are being rewritten from scratch using it. As elsewhere in the world, experimental farms of PCs/workstations using fast ethernet for communications are being built. Bandwidth is a limitation, although these architectures appear to be a cost-effective solution to some problems. The rationale is that economies of scale and rapid improvement of commodity parts, together with the smallness of the high-end market, make a customised approach unprofitable. On the other hand, the commodity parts are designed and optimised for other purposes, particularly PCs and workstations, and their performance in an MPP environment can be disappointing. The following are major factors currently influencing academic high-end computing in the US.

ASCI and Related Programs: The Accelerated Strategic Computing Initiative (ASCI) is probably the major driving force behind the development of high-end computers today. By 2004, the average age of US nuclear weapons will exceed their design lifetimes and almost all design/test-experienced scientists will have retired. Simulation will be needed to certify the functionality of US nuclear weapons. The computational requirement is a balanced machine which can sustain 100 TFlops, with 30 TB of memory. This hundredfold increase in speed over four years requires an accelerated technology programme.

All ASCI machine architectures are hierarchical distributed shared-memory systems, with low-latency high-bandwidth interconnect. The commitment to using entirely commodity parts means that to achieve this goal will require aggregating components on an unprecedented scale. As a consequence, the most powerful systems will cost in the region of $100M. There is much more to ASCI than 100 TFlops machines: data storage, visualisation, tool, and applications development are all important. The Program is currently funded at the level of $500M pa. The Pathforward Program within ASCI is a further injection of R&D funding to accelerate commercial capabilities to achieve 30 TFlops in 2001 and to involve all US vendors, thereby spreading the risk of any one vendor failing.

The effect of ASCI is not only to accelerate R&D in US computing companies, but to establish simulation as a strategic technology of international importance. It, not QCD, is driving the development of high-end computing. ASCI will produce machines which are on an unprecedented scale, enormously expensive and require megawatts of power. It is probable that US physicists will get access to them for QCD.

The current DOE Scientific Simulation Initiative (SSI) is a response to the growing pressure for civilian access to ASCI-scale machines. It is aimed at climate modelling, combustion and basic science. The effect of the SSI funding seems to have been to stimulate computational science work and, with it, increased interest in lattice QCD at the national laboratories.

In 1998, the President's IT Advisory Committee (PITAC) reported that Federal investment in IT was inadequate, and that IT R&D is excessively focussed on the short-term. Concerning high-end computing, PITAC recommended that: software research should be a priority; high-end systems should be provided for academic research; and new technologies should be developed to achieve Pflops by 2010.

US Vendors: All US suppliers are concentrating on providing shared-memory machines for a market which peaks at systems costing around $1M. Their business plans for high-end systems are not well-developed.

IBM is building the ASCI White machine for installation at LLNL in 1Q 2000 at a cost of $85M. This will have 512 16-way Power3-based symmetric multi-processor (SMP) nodes, a total of 4 TB of memory, and a peak performance of 10 TFlops. ASCI Blue Mountain, which was installed at LANL in 4Q 1998, comprises 48 SGI 128-way Origin R10k SMP nodes, connected by a HiPPI-6400 switch, to give 3 TFlops peak. A similar 16-SMP system, with a peak speed of 1 TFlops, is available for QCD and other unclassified work at LANL.

Compaq's new Wildfire system, due in 4Q 1999, is expected to be a cache-coherent cluster of up to 32 Alpha processors connected via memory channel. Sun's HPC10000, due in 1Q 2000, will be a 64-way SMP using Ultra-III processors. The current interconnect technology allows a cache-coherent cluster of up to 4 HPC10000 systems, ie, 256 processors. Pathforward is funding research into an interconnect which will allow up to 16 SMPs to be connected in a cache-coherent cluster.

8.2.2 Japanese Strategy

Japan agrees with the US that commodity-based distributed memory systems are economically attractive, useful for some applications, and will become more attractive in future as software and tools are developed. All major Japanese vendors are developing distributed memory systems, although, so far, such machines have mostly been placed in the research community. However, Japanese vendors believe that, at present, customised vector processors and shared memory are what most users want, despite the fact that they do not scale easily, or inexpensively, and they result in machines which are expensive to run. If real shared memory is not possible, then fast crossbar switches are used for communications within distributed memory systems. It is necessary to customise processors, because vector registers and operations are not yet available on commodity chips. As more sophisticated features are integrated into commercial chips, US and Japanese strategies may converge.

Japanese Vendors: Fujitsu and NEC both have CMOS-based systems with peak performances of a few GFlops per CPU. Fujitsu has built enormous systems with hundreds of nodes for dedicated applications, such as the Numerical Wind Tunnel for CFD. Hitachi licensed HP's workstation processor architecture and modified it to include a "pseudovector" capability for its SR2201.

A speeded-up version of Hitachi's new SR8000 will be leased by KEK from March 2000, at a cost of \$10M pa, primarily to support the lattice QCD groups in Japan. This comprises 100 nodes connected by a 3-dimensional crossbar, with a peak performance of 1.2 TFlops and 448 GB of main memory. Each node is an SMP comprising eight 1.5 GFlops scalar microprocessors, 50% faster than in the original SR8000. The performance expected for QCD is 650 – 700 GFlops, ie twice that of CP-PACS's SR2201. The Japanese lattice QCD community is aiming for a 100 TFlops computer in 2003, although this project is not yet funded.

8.2.3 Commercial Technology Development

Over many years, microprocessor speeds have doubled roughly every 18 months, in what has come to be called Moore's Law. For the foreseeable future, the economic driving force will be the PC market, and the high-end system designer will be forced to exploit a 'dual-use' approach which, as far as possible, employs PC technology. The issue of packaging is important: how chips can be attached to each other, how many pins can be connected, what density, power supply, and cooling rates can be achieved. The state-of-the-art today is roughly-speaking a 1 GHz processor, with a peak performance of a few GFlops, using 0.2μ CMOS, and requiring around 50W of power.

Market forces, rather than technological development, drive the prices and availability of memory. Computing speed is being held back primarily because of the time needed to move data between memory and CPU, and this is the main performance issue for QCD. The low bandwidth is because the two functions are on different chips. The current solution is increased use and sophistication of caches.

Industry is not improving interconnect latency and bandwidth fast enough to meet the ASCI objective. To remedy the situation, Pathforward is focussing resources into the R&D programmes of Compaq, IBM, SGI and Sun. QSW is developing a switching network in conjuction with Compaq. However, the small market for systems employing large interconnection networks means that this technology will continue to be priced at a premium. In addition to bandwidth and latency, minimum packet size and granularity become important issues as we move towards fine-grained parallelism.

ASCI is aiming to reduce significantly the cost and footprint of PB storage systems. Thus, we can anticipate 1 TB benchtop optical tape systems over the next few years. The ASCI target for 2004 is a 200 GB/s data rate to a 100 PB storage system. This substantially exceeds the requirements for QCD, so data storage is unlikely to be a limiting factor.

8.3 Purpose-Built Systems Development

The APE, Columbia University and CP-PACS groups are all engaged in, or planning next-generation machines. Detailed information has been provided by the APE group. The Columbia group has supplied an indication of its aims, whereas the CP-PACS group has not provided any information beyond that mentioned above.

APEmille: This is the latest generation of APE computers, with peak performance of 500 Mflops per node, an order of magnitude greater than APE100. The architecture is a 3-dimensional SIMD grid of customised nodes which are optimised for 32-bit complex arithmetic. The main enhancements compared to APE100 are the availability of local addressing on each node, the support for double precision floating-point arithmetic (with a loss in performance of a factor of $3-4$), and the realisation of a multi-SIMD capability in software, which allows division of the machine into partitions able to run independent programs. Each board contains a 2^3 processor subgrid, and a Linux-based PC controls groups of up to four boards. The I/O bandwidth is 1 MB/s per 1 GFlops. The intention is to construct 64, 128 and 250 GFlops units (with the option of combining two 250 GFlops units into one 500 GFlops system). Each cabinet holds two 64 GFlops units and is air-cooled.

Currently, all the components have been fabricated and tested. A 64 GFlops machine is working in Rome and a 128 GFlops machine is planned by the end of 1999. Tenders have been invited for the manufacture of the components for systems with a combined performance of approximately 1 TFlops (plus about 300 GFlops for DESY). The first machines from this batch will begin to be available for physics in late spring 2000 (the procurement process should be completed in late summer). The cost is expected to be around 3.5 euro/Mflops. While it will be possible to purchase APEmille hardware, it may not be possible to purchase maintenance or support. Software will be available as freeware, and HDL descriptions of the APEmille chip set will be available by negotiation with INFN. It is clear that operating an APEmille system will require more effort from the host than was required for APE100 systems.

apeNEXT: The APE Group submitted a draft proposal for a new machine project to INFN in the summer of 1999. This proposal involves collaboration with DESY and the IT Division at CERN. INFN is expected to support the preparation of a detailed proposal, which could be submitted for approval in 2000.

The apeNEXT target is 3 – 6 TFlops peak in double precision, 1 TB of on-line memory and 1 GB/s I/O to disk. To be competitive, a pan-European project must aim for at least 30 TFlops integrated sustained performance by 2003. A feasible timeframe would involve completing the design in 2000, construction of prototype subsystems in 2001, a one-tenth machine in 2002 and full systems in 2003. The estimated cost is 0.5-1 euro/Mflops.

The proposal is for an architecture which will accommodate two versions of the machine:

- PC nodes within a purpose-built fast point-to-point 3-dimensional interconnection network. Dual-Pentium boards could achieve 500 Mflops per node for QCD, but the power consumption is likely to be an order of magnitude higher than with custom processors;
- the same interconnection network, in which each node is a customised 1.6 GFlops processor.

Currently, the main limitations of commercial hardware are the latencies and bandwidths for local memory access and interprocessor communications. Thus, any design requires a customised interconnection network. The choice of processor is less critical. The proposed network design would allow the construction of PC-based systems for medium-scale applications, and very large fully-customised machines for the most demanding QCD work. In this way, PC-based systems could be available quickly and could track the rapidly evolving PC technology, although they are unlikely to be cost-effective for machines in the multi-TFlops range on the timescale required for QCD; hence the need also to develop a customised processor. The PC-based version might have wider applications than QCD and some aspects of the technology might be relevant to the data-processing requirements of the LHC.

Columbia University's QCDSP II: The Columbia Group has funding for the design of a new machine with a ten-fold improvement in cost performance compared to the QCDSP, or ∼$1/Mflops. It will use a similar interconnection network to that of the QCDSP and a new 64-bit digital signal processing chip. The objective is to have a prototype in 2001 and a large machine, of up to 6 TFlops peak performance, available in 2002.

9 Conclusions

The panel's main conclusions are as follows:

- The future research program using lattice simulations is a very rich one, investigating problems of central importance for the development of our understanding of particle physics. The program includes detailed (unquenched) computations of non-perturbative QCD effects in hadronic weak decays, studies of hadronic structure, investigations of the quark-gluon plasma, exploratory studies of the non-perturbative structure of supersymmetric gauge theories, studies of subtle aspects of hadronic spectroscopy and much much more.

- The European Lattice Community is large and very strong, with experience and expertise in applying numerical simulations to a wide range of physics problems. For over 10 years it has organised itself into international collaborations when appropriate, and these will form the foundation for any future European project. Increased coordination is necessary in preparation for the 10 TFlops generation of machines.
- Future strategy must be driven by the requirements of the physics research program. From the Technology Review in section 8 we conclude that it is both realistic and necessary to aim for machines of O(10 TFlops) processing power by 2003. As a general guide, such machines will enable results to be obtained in unquenched simulations with a similar precision to those currently found in quenched ones.
- It will be important to preserve the diversity and breadth of the physics program, which will require a number of large machines as well as a range of smaller ones.
- The lattice community should remain alert to all technical possibilities for realizing its research program. The panel concludes, however, that it is unlikely to be possible to procure a 10 TFlops machine commercially at a reasonable price by 2003, and hence recognises the central importance of the apeNEXT project to the future of European lattice physics.

10 Recommendations to ECFA

1. The panel seeks ECFA's endorsement of the success and importance of Lattice Field Theory to the European research program in Experimental Particle Physics.
2. The panel asks ECFA to recognise that continued support is required to ensure that the growing range of scientific opportunities discussed in this report are realised and to maintain the vitality and strength of the European Lattice Community. In particular, if it is to remain competitive, the community will require access to a number of 10 TFlops machines by 2003.
3. The panel proposes to organise a workshop of the European Lattice Community in the first half of 2000 to present the conclusions of this report and ECFA's response to it and to discuss the possibilities of a coordinated strategy for obtaining and exploiting 10 TFlops machines from 2003. We seek ECFA's mandate to organise the meeting as an ECFA panel.
4. The panel asks ECFA to recognise the need for a European project and seeks its approval for a coordinated approach by the national agencies to fund it.

Acknowledgements

We gratefully acknowledge the help and encouragement which we have received from our colleagues, and in particular wish to thank Norman Christ, Marco Ciuchini, Yoichi Iwasaki, Karl Jansen, Helmut Satz and Lele Trippicione for their valuable contributions.

A Appendix

Mandate for ECFA Panel on Requirements for High Performance Computing in Lattice QCD

The development of Lattice Quantum Field Theory, together with the availability of powerful parallel supercomputers, is enabling particle physicists to study a wide range of fundamental questions, which were previously intractable. A particularly important class of problems is that which requires control of long-distance, and hence non-perturbative, effects in Quantum Chromodynamics (QCD). The main objective of this study is to assess the High Performance Computing resources which will be required in coming years by European physicists working in this field, and to review the scientific opportunities that these resources would open.

Specifically the panel is asked:

1. to assess the computing power, memory, I/O and storage resources which are likely to be technically possible in the coming years (ip to about 2005);
2. to review the questions in particle physics which could be studied using such resources;
3. to present possible strategies for the future development and organisation of Lattice Quantum Field Theory in Europe.

Among the physics questions the panel is asked to consider are:

1. What level of (statistical and systematic) precision is likely to be possible in the computations of physical quantities including:
   a) leptonic decays constants, form-factors of semi-leptonic and rare radiative decays of beauty hadrons and mixing amplitudes of neutral mesons;
   b) quark masses and the strong coupling constant;
   c) the hadron spectrum, including glueballs and hybrids;
   d) hadronic structure functions;
   e) non-leptonic decays of B and K-mesons, including ϵ'/ϵ and the $\Delta I = 1/2$ rule;
   f) the phase structure of QCD at non-zero temperature and baryon density?
2. What are the prospects for non-QCD lattice simulations, including those in chiral gauge theories and supersymmetric theories?