Numerical Simulations in Lattice Quantum Chromodynamics

URS M. HELLER

CSIT
Florida State University
Tallahassee, FL 32306-4120, USA

March 24, 2000

1 Introduction and Motivation

The twentieth century has brought striking progress in our understanding of the fundamental structure of matter, beginning with quantum mechanics and culminating with the "Standard Model" of particle interactions. The dramatic successes of quantum electrodynamics, verified to an accuracy of better than 1 part in 10^8 in some processes, and of the unification of electromagnetic and weak interactions explain a vast array of physical phenomena. However traditional theoretical tools have proven very limited in the study of quantum chromodynamics (QCD), the component of the Standard Model which describes strongly interacting particles, the hadrons. The rich and complex structure of this highly nonlinear theory arises from the interactions between quarks, the constituents of protons, neutrons and all the other hadrons, and gluons, the carriers of the strong force. QCD is quite similar to quantum electrodynamics (QED), where photons are the carriers of the force between electric charges. However, contrary to photons, gluons interact with each other, leading to the nonlinearities that make QCD so difficult to deal with.

At short distances, which correspond to high energies, the interactions in QCD are relatively weak, allowing for a perturbative treatment, an expansion in the small coupling constant. Most confirmations that QCD correctly and accurately describes the strong interactions come from high energy experiments probing this regime. At longer distances, corresponding to lower energies, the interactions, due to the self-interactions among the gluons, become really strong. Then the nonlinearities become important and the perturbative methods fail. But it is exactly this regime, which is necessary to explain many of the properties of the hadrons, such as their masses, their decay amplitudes and lifetimes, if they are unstable, their sizes, charge radii, etc. As the distance increases, in fact, the interaction becomes so strong that the quarks are permanently confined within hadrons.

Lattice quantum chromodynamics, by way of large scale numerical simulations, provides the only known comprehensive method to compute, with controlled systematic errors, properties of hadrons starting from the simple equations of QCD. Many of these properties, such as the hadron spectrum, are well known experimentally. Lattice QCD then aims to confirm – or disprove – QCD as the theory that explains these properties correctly.

More importantly, precise knowledge of the effects of QCD are needed to complete the determination of the basic parameters of the Standard Model, which is the central focus of current and planned high energy physics experiments, and in the search for new physics beyond the Standard Model. Prime among the basic parameters are those that describe how the weak interactions mix different species of quarks, the elements of the Cabbibo-Kobayashi-Maskawa (CKM) matrix. For heavy quark species those matrix elements remain poorly known. Their determination requires combining experimental measurements with lattice QCD calculations.

While all experimentally known hadrons are made up of quarks and anti-quarks, QCD predicts the existence of "exotic matter" that contains gluons as an essential ingredient, such as so-called glueballs and hybrid mesons. These particles, if found experimentally, would make a dramatic confirmation of QCD as the theory of the strong interactions. Searching for them in experiments, though, is much like finding a needle in a hay stack. Lattice QCD computations of the mass and other properties of these particles would help tremendously in narrowing down in which reactions in experiments, such as being performed and planned at Jefferson National Lab, to search for these exotic particles.

Under normal conditions quarks and gluons are confined inside hadrons. However, at sufficiently high temperatures and/or pressure a new state of matter should appear, a phase best described as a plasma of unconfined quarks and gluons. This quark-gluon plasma filled the entire universe up to roughly 100 microseconds after the Big Bang, and it may still play a role today in the cores of neutron stars. The observation of this plasma is the primary goal of heavy ion experiments at RHIC at the Brookhaven National Lab. Lattice QCD simulations have already provided the best estimate of the temperature, at which the plasma appears, roughly 2 trillion ^{0}C. Further large scale simulations are needed to study the detailed nature of the phase transition and to determine the equation of state for the plasma phase.

2 Lattice QCD simulation setup

Lattice QCD simulations started with the seminal work of Creutz in 1979 [2], who was the first to apply Monte Carlo simulation techniques to lattice QCD and to produce the first numerical evidence of confinement in QCD.

The non-perturbative solution of QCD is in many ways similar to solving fluid dynamics problems (especially when employing molecular dynamics techniques, as described below). One has a simple set of equations that implicitly contain all the information (up to the boundary conditions). But because they are very nonlinear, these equations are extremely difficult to solve. Indeed, the numerical solution of QCD appears to be one of the most challenging computational problems in physics.

The starting point of calculations in QCD is the path integral approach to quantum field theories. To allow for a numerical attack, one first discretizes space-time into a regular 4-dimensional grid, called a lattice [1], with lattice spacing a. The quarks are then described by fields $\psi(x)$, complex 3-vectors, attached to the sites of this lattice, and the gluons by special complex unitary 3×3 matrices $U_\mu(x)$, belonging to the color group $SU(3)$, attached to the links with endpoints x and $x + a\hat{\mu}$. The relation to the gluon fields $A_\mu(x)$ of the continuum field theory is given by $U_\mu(x) = \exp\{iagA_\mu(x)\}$ with g denoting the coupling constant at energy scale $1/a$. The physical observables are extracted from expectation values

$$\langle\Omega\rangle = \frac{1}{Z}\int [d\psi][d\bar{\psi}][dU] \exp\{-S(U) - \bar{\psi}\mathcal{M}(U)\psi\}\ \Omega(U,\psi,\bar{\psi}) \ . \tag{1}$$

The function $S(U)$ in the exponential is the gluon action, containing their kinetic term and local interactions. It consists of the sum of products of the U-matrices around elementary plaquettes of the 4-dimensional lattice, labeled by their lower left-hand corner site x and the directions $1 \le \mu < \nu \le 4$ specifying the plane in which the plaquette lies,

$$S(U) = \frac{1}{2g^2} \sum_{x,\mu<\nu} \mathrm{Tr}\left[U_\mu(x)U_\nu(x+\mu)U_\mu^\dagger(x+\nu)U_\nu^\dagger + h.c.\right] \ . \tag{2}$$

This is the simplest form which is invariant under local gauge transformations

$$U_\mu(x) \to U'_\mu(x) = V^\dagger(x)U_\mu(x)V(x+\mu) \tag{3}$$

and reduces in the limit $a \to 0$ to the continuum action,

$$S \to \int d^4x \frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}^2(x)\right] + \mathcal{O}(a^2) \tag{4}$$

with

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + g[A_\mu(x), A_\nu(x)] \ . \tag{5}$$

$\mathcal{M}(U)$ in (1) contains the kinetic term of the quark fields and their interactions with the gluons. One of the most commonly used forms, know as Wilson's fermion action, is:

$$\begin{aligned} \bar{\psi}\mathcal{M}(U)\psi &= \sum_x \bar{\psi}(x)\Big\{(4+m)\psi(x) \\ &\quad -\frac{1}{2}\sum_{\mu=1}^{4}\Big[(1-\gamma_\mu)U_\mu(x)\psi(x+\mu) + (1+\gamma_\mu)U_\mu^\dagger(x-\mu)\psi(x-\mu)\Big]\Big\} \\ &\to \int d^4x \bar{\psi}(x)\Big[\gamma_m u\left(\partial_\mu + gA_\mu(x)\right)\psi(x)\Big] + \mathcal{O}(a) \ . \end{aligned} \tag{6}$$

Z in (1) is a normalization factor, often referred to as partition function, because of its similarity to partition functions in statistical mechanics. The integration is over the fields at each site and link of the lattice. The fermion fields are somewhat peculiar. They are anti-commuting and therefore not representable on a digital computer. Fortunately their integration is Gaussian and can be carried out analytically. This leads to

$$\begin{aligned} \langle\Omega\rangle &= \frac{1}{Z}\int [dU] \det\mathcal{M}(U) \exp\{-S(U)\}\ \Omega(U,\mathcal{M}^{-1}(U)) \\ &= \frac{1}{Z}\int [dU] \exp\{-S(U) + \mathrm{Tr}\log\mathcal{M}(U)\}\ \Omega(U,\mathcal{M}^{-1}(U)) \ . \end{aligned} \tag{7}$$

The price we have to pay for the peculiarity of the quark fields now consists in the very time consuming computation of $\det \mathcal{M}(U)$, with $\mathcal{M}(U)$ a huge, but sparse, matrix, as can be seen from (6).

We still need to do the integration over the U-field on each link of the lattice. Altogether this is typically an integral over more than 10^6 dimensions. This is an impossible task using any kind of standard integration method. However, due to the exponential factor in eq. (7), the contribution from most field configurations to the integral is negligible and we can use stochastic methods. For pure gauge simulations, *i.e.,* without the fermion determinant $\det \mathcal{M}(U)$ in (7) standard Monte Carlo algorithms, as used in statistical mechanics, like the Metropolis algorithm, heat bath algorithm and (slight) improvements, are adequate, and widely used.

In full QCD simulations, because of the non-locality of $\det \mathcal{M}(U)$ these standard Monte Carlo methods are impractical: for every local change of the gauge fields the change in the fermion determinant would be needed. The best algorithms for the computation of eq. (7) known to date are based on the analogue to micro-canonical ensembles in statistical mechanics. One introduces a fictitious time and momenta conjugate to the U-fields, with Gaussian distribution $\exp(-p^2/2)$ for each. $S(U) - \mathrm{Tr}\log \mathcal{M}(U)$ is now viewed as the potential term and $\sum p^2/2$ as the kinetic term of a Hamiltonian. Then, only a single computation of $\mathrm{Tr}\log \mathcal{M}(U)$, or rather its derivative, is needed to update all the gauge fields simultaneously. This enormous saving in computing effort for QCD simulations makes these kinds of simulation algorithms the most effective in lattice QCD simulations.

The average in (2) is then replaced by an average over $\Omega(U(t), \mathcal{M}^{-1}(U(t)))$ in fictitious time, with $U(t)$ the solution of the equations of motion resulting from the Hamiltonian. This procedure is referred to as the "Molecular Dynamics" algorithm. To ensure ergodicity, the momenta p are, from time to time, after what is referred to as a trajectory, replaced by new random Gaussian variables, *i.e.,* refreshed. This combination is known as the "Hybrid Molecular Dynamics" algorithm (HMD) [3]. The expectation values, eq. (7), are then computed as simple averages

$$\overline{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \Omega(U(t_i), \mathcal{M}^{-1}(U(t_i))) \ , \tag{8}$$

with t_i, $i = 1, \ldots, N$ labeling a set of ending points of trajectories. This amounts to a stochastic estimation of $\langle \Omega \rangle$ and becomes exact only in the limit $N \to \infty$. For a finite number of trajectories, the estimate has a statistical error which decreases, for large N, as $1/\sqrt{N}$.

To solve the equations of motion numerically we have to discretize the fictitious time. This introduces finite step size errors, which have to be kept under control (smaller than *e.g.* the statistical errors). Better still would be to use a few different step sizes and then extrapolate to zero step size. Since each simulation is very time consuming such an extrapolation is rarely possible. Usually a "leap frog" integration scheme is used to solve the discretized equations of motion. It has the advantage of being simple and thus easily implemented, but also that it is (up to round-off errors) explicitly time reversal invariant. This property is important in the variant of the algorithm where the errors coming from the discretization of the equations of motion in fictitious time are compensated by periodically, after each trajectory, performing an accept/reject Metropolis step. This last variant which, contrary to the previous two that suffer from finite step size errors, is an exact algorithm, goes under the name of "Hybrid Monte Carlo" algorithm (HMC) [4]. Unfortunately, for technical reasons, the HMC algorithm can not be used for all systems of interest. In those cases one resorts to the inexact HMD algorithm and tries to control the finite step size errors.

The equations of motion to be solved contain a term like $\mathrm{Tr}\{\partial\mathcal{M}(U)/\partial U\ \mathcal{M}^{-1}(U)\}$ from the derivative of $\mathrm{Tr}\log\mathcal{M}$ in (7). Thus, for each step in the integration, the large, but sparse, matrix $\mathcal{M}(U)$ needs to be inverted. This is still not practical. However, on can avoid this by noting that, for two flavors of fermions, and using that for the fermion action (6) $\det\mathcal{M}^\dagger = \det\mathcal{M}$,

$$\exp\{2\mathrm{Tr}\log\mathcal{M}\} = \det[\mathcal{M}^\dagger\mathcal{M}] = \int[d\phi][d\phi^\dagger]\exp\{-\phi^\dagger[\mathcal{M}^\dagger\mathcal{M}]^{-1}\phi\} \tag{9}$$

with ϕ, $\phi^\dagger$ bosonic fields. During the molecular dynamics evolution of the HMD or HMC algorithm these fields are held constant. At the beginning of each trajectory they are refreshed, like the momenta, by creating them with the distribution of (9). This can be achieved by creating Gaussian random fields χ and setting $\phi = \mathcal{M}^\dagger\chi$. The derivative in the molecular dynamics evolution now becomes $\phi^\dagger[\mathcal{M}^\dagger\mathcal{M}]^{-1}\partial[\mathcal{M}^\dagger\mathcal{M}]/\partial U[\mathcal{M}^\dagger\mathcal{M}]^{-1}\phi$ and involves only the computation $[\mathcal{M}^\dagger\mathcal{M}]^{-1}\phi$, *i.e.* only one row of the inverse. This is still the task that makes lattice QCD simulations so expensive. The inversion is done with an iterative method, usually the Conjugate Gradient (CG) algorithm.

3 Computational requirements

To give an impression of the computational demands: One iteration of the CG algorithm for a lattice size $24^3 \times 64$ requires between about $6 \cdot 10^8$ and $3 \cdot 10^9$ floating point operations, depending on the exact way the fermion - gluon interaction was discretized on the lattice, *i.e.,* depending on the exact form of the matrix $\mathcal{M}$. Most of the operations come from multiplying the complex 3×3 matrices on the links, representing the gluons, with complex 3-vectors on the sites, representing the quarks (see eq. (6)). Convergence of the CG algorithm can take between 500 and, for physically more interesting parameter values, more than 1000 iterations each time. To keep the acceptance rate sufficiently high in the HMC algorithm, above 50%, or the step size errors reasonably small in the HMD case, step sizes as small as $dt = 0.005$ to 0.01 are needed. And finally, to collect reasonable statistics for the observables, computed as fictitious time averages, the equations of motion have to be integrated for a length of 2000 to 5000 time units. Combining all these numbers, we see that a "state of the art" computation requires between $3 \cdot 10^{16}$ and $3 \cdot 10^{18}$ floating point operations, which translates to between about 1 and 100 Gigaflop/s-Years.

These are the resources needed for a simulation with just one set of parameters, *i.e.* one fixed lattice spacing a, one volume and one quark mass value. Several simulations are needed to make sure that finite volume effects are negligible and that an extrapolation to zero lattices spacing, to the continuum limit, can be made with controlled errors. In a typical simulation the lattice spacing is, say, 0.1 fm (1 fm $= 1 \cdot 10^{-15}$ m). A proton has a charge radius of about 1 fm, and a proton therefore should fit nicely into our 24^3 box without finite size effects, while at the same time the lattice spacing should be fine enough to give a good resolution of the proton. Nevertheless, it is known that the results of such a computation can differ from the final continuum limit by as much as 10 – 30%. This difference is referred to as lattice or discretization effects.

In addition, the quark masses in a lattice QCD simulation are typically much larger than those in nature, and therefore simulations at a few different quark masses are needed to allow an extrapolation to the almost massless light up and down quarks of nature.

4 Implementation considerations

From the requirements of a single computation described above, it is clear that the computing power of even high-end workstations is dismally inadequate for lattice QCD calculations. In a lattice QCD code, almost all the time, the same operations have to be performed on all the lattice sites of a regular, fixed grid. And the data needed either reside already on that site or on one of the nearest neighbor sites in one of the directions of the 4-dimensional grid. This is a classic case of a data parallel situation and lends itself to rather straightforward vectorization or parallelization. Indeed lattice gauge theory codes are among the most efficient, both on vector and, more importantly, on massively parallel supercomputers.

The author is a member of a large, partially DOE sponsored collaboration, known under the name "MIMD Lattice Calculation" (MILC) Collaboration. This collaboration is using any parallel computer that they can get time on. For this they have developed a family of portable MIMD codes, which run on a wide variety of scalable parallel computers, from single workstations for code development and testing, to the T3E, SP systems, Origin 2000, and more recently PC clusters. The code is written in C and is highly portable. The only parts of the C code which are machine dependent are the communications routines. These are stored in a single file. A different version of this file exists for each machine or communications library. Standard message passing libraries are especially interesting from the point of view of maintaining portable code, and the MILC collaboration has implemented a version of the communications routines for both PVM and MPI [5]. Older versions were running on Intel Paragons and TMC CM-5's.

5 Recent developments and future prospects

As any computational field, lattice QCD has profited from the fast development of ever more powerful computers. The numerical algorithms used, Hybrid Monte Carlo, and Conjugate Gradient type routines for the very frequent inversions of the large sparse quark matrices, are by now fairly standard. Tremendous progress, can, and already has been achieved in the last few years, in reducing the lattice discretization errors. In the most commonly used lattice formulation of the quarks, the so-called Wilson fermions, the finite lattice spacing errors are of order $\mathcal{O}(a)$. Therefore, to reduce those errors by half, the lattice spacing needs to be decreased by a factor two. The number of lattice sites for a fixed physical 4-dimensional lattice then grows by a factor 2^4 and the actual costs, in CPU time, of a simulation grows by a factor 2^8 to 2^{10}, depending on the details of the simulation algorithm used. Therefore, reducing the discretization errors from $\mathcal{O}(a)$ to $\mathcal{O}(a^2)$ can produce enormous savings. Unfortunately, due to the intrinsic quantum nature of the problem, this is not as easy as just using a better finite difference approximation to a derivative. However, a method to achieve this goal has recently been developed [6].

So far, lattice QCD computations have determined the value of the strong coupling constant at the energy scale given by the Z-boson mass, where all different determinations are usually compared (recall that the coupling decreases with increasing energy), to an accuracy of 3%, which is about the same as the best experimental determinations. The error estimates includes, in quadrature the statistical error form the stochastic Monte Carlo computation, and the systematic errors, such as those from the extrapolation to the continuum, $a \to 0$, limit [7]. The masses of the light hadrons have been computed to an accuracy of 5% to 10%, including those of the lightest exotic states, glueballs and hybrid mesons [8]. Computations of the QCD effects on weak matrix elements, on the other hand, so far typically have errors of 20% to 40%. In many cases this is 5 to 10 times the errors of the experimental measurements with which the computations need to be combined to extract the fundamental parameters of the Standard Model.

This discrepancy has lead to a considerable effort to increase the computing resources available for lattice QCD computations. If successful the errors on the weak matrix element computations, for example, are expected to be reduced by a factor two over the next two years. Teraflop/s scale computations, foreseen for the years 2002 – 2005, are expected to bring the errors down to be comparable with the experimental errors.

ACKNOWLEDGMENTS: This work has been supported in part by DOE contracts DE-FG05-85ER250000 and DE-FG05-96ER40979.

References

[1] K.G. Wilson, Phys. Rev. **D10** (1974) 2445.

[2] M. Creutz, Phys. Rev. Lett. **43** (1979) 553; Erratum-ibid **43** (1979) 890.

[3] S. Duane, Nucl. Phys. **B257** (1985) 652; S. Duane and J.B. Kogut, Nucl. Phys. **B275** (1986) 398.

[4] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, Phys. Lett. **195B** (1987) 216.

[5] The MILC QCD code is described in some detail at, and can be obtained from, the URL `http://cliodhna.cop.uop.edu/~hetrick/milc/`.

[6] M. Lüscher, S. Sint, R. Sommer, P. Weisz, and U. Wolff, Nucl. Phys. **B478** (1996) 365, Nucl. Phys. **B491** (1997) 323, 344; K. Jansen and R. Sommer, Nucl. Phys. **B530** (1998) 185.

[7] Particle Data Group, C. Caso et al, The European Physical Journal C3 (1998) 1; http://www-pdg.lbl.gov/1998/qcdrpp_part1.pdf.

[8] For recent reviews on the status of lattice QCD see, S.R. Sharpe, "Progress in Lattice Gauge Theory", plenary talk at ICHEP98, Vancouver, Canada, September 1998, hep-lat/9811006; http://arXiv.org/abs/hep-lat/9811006; H. Wittig, "Lattice Gauge Theory", plenary talk at the Europhysics Conference on High Energy Physics, Tampere, Finland, July 1999, hep-ph/9911400; http://arXiv.org/abs/hep-ph/9911400.