Fermi National Accelerator Laboratory

FERMILAB-CONF-89/265-T
December 1989

CRITIQUE OF PRACTICAL METHODS FOR COMPUTER EVALUATION OF THE HADRON SPECTRUM *

Paul B. MACKENZIE

Theoretical Physics Group, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA

I review Monte Carlo algorithms for lattice QCD including the dynamical effects of quarks, emphasizing scaling behavior of the algorithms with lattice spacing, volume, and quark mass.

1. INTRODUCTION

In computationally intensive areas of science and engineering such as computational hydrodynamics which have been in existence for a long period of time, it is common for algorithmic improvements to contribute as much as hardware improvements to the efficiency of the calculations. This is to be expected from evolutionary improvements and apart from any revolution in methods (such as might be caused by a reformulation of QCD suitable for methods borrowed from quantum chemistry[1]). Monte Carlo methods for QCD calculations have indeed made enormous progress over the last ten years. However, they still are not adequate for convincing results on today's computers. Current calculations of the hadron spectrum are done with too large a quark mass m_q, and too large a lattice spacing a, and too small a physical volume V. This means that in addition to testing algorithms on the largest lattices possible, it is important to understand how their performance scales with a, V, and m_q.

By critique in the title, I mean the attempt to understand the behavior of methods and approximations using arguments outside the direct application of the methods. An ideal which we can at present only partially achieve is to determine in advance the expected behavior of algorithms and then to test the predictions numerically. An early paper assembling a priori arguments about the behavior of methods and approximations is Parisi's "Prolegomena to Any Future Computer Evaluation of the QCD Mass Spectrum".[2] Parisi's tongue in cheek allusion to Kant's metaphysics is appropriate: in this branch of lattice gauge theory, we are studying the means to knowledge about hadrons, rather than hadrons themselves. (Other ways in which the allusion may be appropriate will not be reviewed.)

By practical, I mean that I will focus on the currently most promising algorithms. For a more encyclopedic classification, see Weingarten's review at Lattice 88.[3] In particular, I will not discuss promising methods for simulation of Bose theories which have not been applied to full QCD, such as overrelaxation algorithms[4] and cluster updating methods[5].

To include the dynamical effects of the determinant of the Dirac matrix M, the small step size simulation algorithms which I will be discussing require the derivative of the fermionic effective action with respect to the gauge field:

$$F \equiv \frac{\partial}{\partial U} \ln \det(M^\dagger M) \qquad (1.1)$$

$$= Tr \frac{1}{M^\dagger M} \frac{\partial M^\dagger M}{\partial U}. \qquad (1.2)$$

Most methods for incorporating the fermionic force in equation 1.2 into QCD simulations introduce scalar fields with the quantum numbers of quarks. Weingarten and Petcher[6] proposed adding a term to the gauge action with the inverse of the sqared Dirac matrix between scalar fields: $S_\Phi = -\frac{1}{2}\Phi^\dagger \frac{1}{M^\dagger M}\Phi$. Path integration over the scalar fields produces the correct fermion determinant. The Cornell group[7] introduced Gaussian random fields η with quark quantum numbers (fermionic noise). The trace in equation 1.2 can be approximated by sandwiching the matrix to be traced between the $\eta^\dagger$ and η. Fucito

*Presented at the 1989 Symposium on Lattice Field Theory, Capri, September 18-21, 1989.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

et al.[8] simulated over bosonic "pseudofermions" χ with the original quark action, rather than its inverse. The inverse matrix required in equation 1.2 is obtained from $< \chi\chi^\dagger >$. The necessity of doing matrix inversions is replaced by the necessity of a Bose simulation at each update step. The first method is applicable only to sets of two equal-mass Wilson fermions or four equal-mass staggered fermions. The other two methods may be used with any number of fermions.

	Scalar field distribution	Fermionic force
Weingarten, Petcher.	$\exp[-\frac{1}{2}\Phi^\dagger \frac{1}{M^\dagger M}\Phi]$	$\frac{1}{4}\Phi^\dagger \frac{1}{M^\dagger M}\frac{\partial M^\dagger M}{\partial U}\frac{1}{M^\dagger M}\Phi$
Cornell noise.	$\exp[-\frac{1}{2}\eta^\dagger\eta]$	$\frac{n_f}{8}\eta^\dagger \frac{1}{M^\dagger M}\frac{\partial M^\dagger M}{\partial U}\eta$
Pseudofermions.	$\exp[-\frac{1}{2}\chi^\dagger M^\dagger M\chi]$	$\frac{n_f}{8} < \chi^\dagger \frac{\partial M^\dagger M}{\partial U}\chi >$

The three approaches are summarized schematically in the above table. The matrix inversion algorithms required for the first two approaches are discussed in section 2. Simulation algorithms used with all three methods are discussed in sections 3 and 4. Pseudofermion Monte Carlo algorithms for estimating the trace in equation 1.2 are discussed in section 5.

2. ALGORITHMS FOR QUARK PROPAGATORS

There is a large literature on iterative methods for the solution of large sparse systems of linear equations. This is the best understood component of QCD algorithms.

In iteratively solving a sparse matrix equation, $Af = b$, the residual vector $r_i \equiv b - Af_i$ is used as a test of the deviation of the approximate solution f_i from the true solution f. Contributions of small eigenvalues to r_i are suppressed, and their associated eigenvectors take longer to converge. The convergence rate of an iterative algorithm may be roughly defined as the inverse of the number of sweeps required for the norm of the residual vector or the error vector ($\equiv f_i - f$) to drop by a factor of e. It is governed by the the "condition number" of A, $\kappa(A) \equiv \lambda_{max}/\lambda_{min}$. ($\kappa$ of the lattice Dirac matrix is of order $1/am_q$.)

Convergence is improved if a preconditioning matrix P can be found such that $\kappa(PA) \ll \kappa(A)$, which is expected if $P \approx A^{-1}$. The solution of $PAf = Pb$ is then much more rapid than the solution of $Af = b$.

Examples which have been tried in lattice gauge theory are:

- Fourier Acceleration.[2, 12] P is taken to be a matrix diagonal in momentum space, such as the free propagator.
- Incomplete Lower Upper (ILU) decomposition.[13] If A can be written as $A = 1 + L + U$, where L is lower triangular and U is upper triangular, take $P = (1 + \omega L)^{-1}(1 + \omega U)^{-1}$, where ω is the overrelaxation parameter.

The two most successful algorithms for quark propagators so far have been the minimum residual (MR) and conjugate gradient (CG) algorithms. The minimum residual algorithm is defined by

$$\begin{aligned} \alpha &= r_i A r_i / |A r_i|^2, \qquad (2.1)\\ \alpha' &= \omega\alpha, \\ f_{i+1} &= f_i + \alpha' r_i, \\ r_{i+1} &= r_i - \alpha' A r_i. \end{aligned}$$

If $\omega = 1$, each step minimizes $|r_i|^2$ along the direction r_i. Choosing $1 < \omega < 2$ may improve convergence. The convergence rate is proportional to $1/\kappa$ of the matrix. Vectors p satisfying $p_i A^\dagger A p_j = 0$ are called conjugate with respect to $A^\dagger A$. Minimizations along conjugate directions are independent, so that the MR algorithm using a complete set of mutually conjugate directions rather than the sequence of residual vectors for the test directions would converge in a finite number of steps in the absence of round off errors. Conjugate residual (CR(k)) algorithms use for the test direction the linear combination of the residual and k previous test directions which is conjugate to those test directions. They converge somewhat faster than the MR algorithm and have similar properties.

The conjugate gradient algorithm also proceeds by advancing in certain direction a distance which

minimizes a quadratic form. It is defined by

$$\begin{aligned} \alpha &= |r_i|^2/p_i A p_i, \qquad (2.2) \\ f_{i+1} &= f_i + \alpha p_i, \\ r_{i+1} &= r_i - \alpha A r_i, \\ \beta &= r_{i+1}^2/r_i^2, \\ p_{i+1} &= p_i + \beta r_i. \end{aligned}$$

α is chosen to minimize the quadratic form $F \equiv \frac{1}{2} fAf - fb$ along the direction p_i. β is chosen to make the new direction conjugate to the old one with respect to A, so that the current minimization does not spoil the previous one. The magic of conjugate gradient is that, if A is a symmetric, positive definite (SPD) matrix, *none* of the earlier minimizations is spoiled. This implies that f_i is the minimum of F on the i-dimensional hyperplane defined by the directions $\{b, Ab, AAb, \ldots, A^{i-1}b\}$. It then follows that the conjugate gradient method converges faster than a steepest descents method with any polynomial preconditioning. Hageman and Young[11] tell us that Chebyshev acceleration improves the convergence rate of iterative methods from $O(1/\kappa)$ to $O(1/\sqrt{\kappa})$, which implies that conjugate gradient should do at least as well. The catch is that the Dirac matrix M is not symmetric, so conjugate gradient must be used with $M^\dagger M$, which brings us back to an expected convergence rate of order am_q.

The expected behavior of the algorithms may be checked on an easy-to-understand toy problem. On a large n-dimensional diagonal matrix with diagonal elements $\{1, \kappa^{1/n}, \kappa^{2/n}, \ldots, \kappa\}$, the expected dependence on κ can be easily seen, and the convergence of both algorithms is very uniform. If a few very small eigenvalues are added, the convergence of minimum residual has the expected dependence on the new condition number. Conjugate gradient, on the other hand, converges in roughly the same number of sweeps as were required without the new eigenvalues. However, a plot of r^2 vs. iteration number shows a long plateau followed by a steep cliff which is similar to behavior sometimes observed in lattice gauge theory.

Conjugate gradient and minimum residual are thus expected (and observed) to converge roughly as $1/\kappa \sim am_q$. However, at small m_q, for Wilson fermions they are observed to behave differently. Conjugate gradient always converges unless one of the eigenvalues is within roundoff of zero. (Presumably this means unless the condition number of M is of order the square root of machine precision or larger.) For small m_q, the convergence exhibits plateaus and cliffs, and on a fixed sized lattice, the length of the plateaus is not very sensitive to m_q. The minimum residual algorithm can fail to converge if $r_i A r_i = 0$ can be obtained when $r_i \neq 0$. This can occur if at least one eigenvalue of M has a negative real part, which is typically the case for Wilson fermions at small m_q.

Other algorithms which are in principle competitive with these two are the successive overrelaxation (SOR) and symmetric successive overrelaxation (SSOR) methods operating on the squared matrix $M^\dagger M$. SOR is the Gauss-Seidel method with an overrelaxation parameter ω. SSOR is the SOR method performed in alternating forward and backward directions. These methods may be considered IL (incomplete lower) and ILU preconditionings of the Jacobi method. The Gauss-Seidel method using the unsquared Dirac matrix M was one of the first methods used for calculating quark propagators. It is rarely used today since it is slower than and fails to converge earlier than the minimum residual method. On an SPD matrix like $M^\dagger M$, Gauss-Siedel is guaranteed to converge, and furthermore converges as the square root of the condition number, like conjugate gradient. (This square root dependence in overrelaxed minimization algorithms motivated the development of overrelaxed Monte Carlo simulation algorithms, which have a similar behavior.[4]) However, there is no theoretical reason to expect SOR and SSOR to be superior to conjugate gradient, and they require careful tuning of the relaxation parameter to achieve peak performance, so these methods have not been much investigated.

2.1. Wilson fermions

ILU preconditioning of the minimum residual and CR(k) algorithms with was proposed for QCD by Oyanagi,[13] who found it to be an order of magnitude faster than unpreconditioned conjugate gradient in the region where minimum residual converged. The method was tested and compared with Fourier acceleration by Rossi, Davies, and Lepage,[15] who

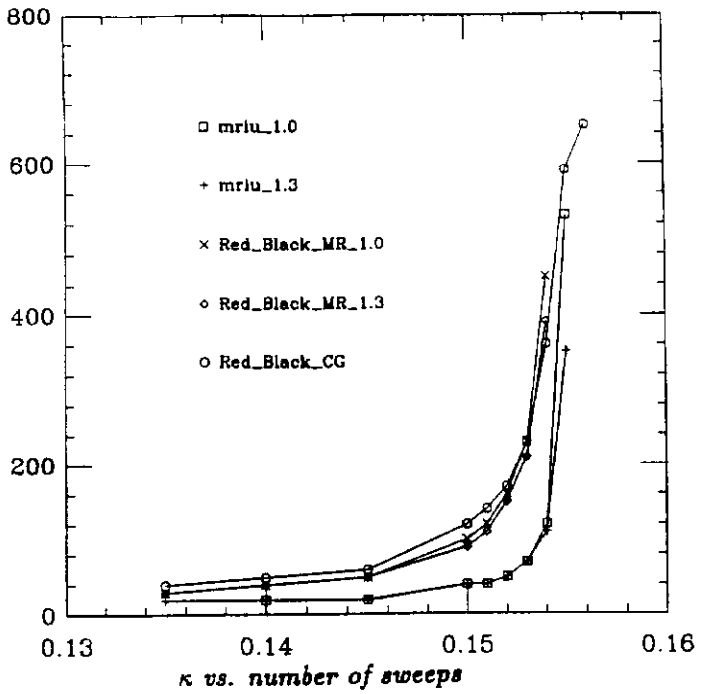

Figure 1: With the MR-ILU method, natural ordering is much more effective than red-black ordering. Data are from Hockney, using Wilson fermions on a 16^4 lattice at $\beta = 6.1$.

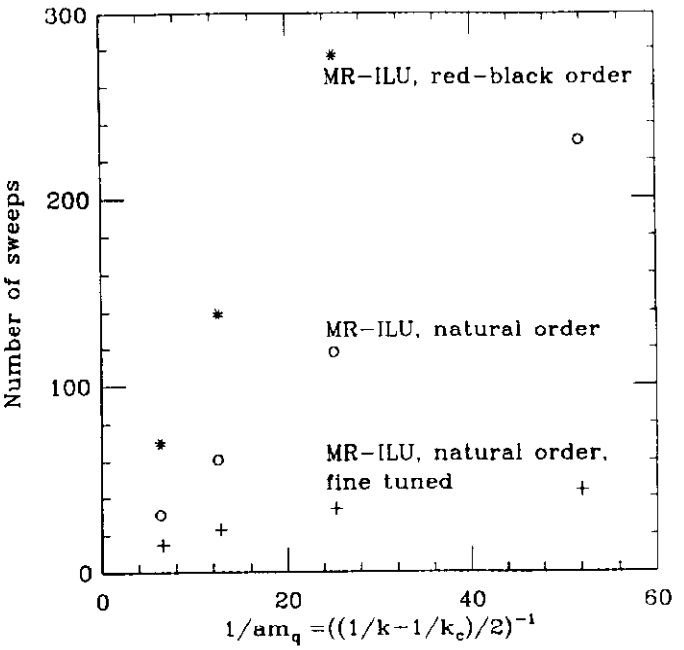

Figure 2: For free wilson fermions, when the hopping parameter in the preconditioning matrix of the MR-ILU algorithm is carefully tuned by trial and error, the number of sweeps required for convergence rises as the square root of the condition number of the matrix, rather than as the linear power.

found much less improvement. This difference was ascribed by Oyanagi to the order used in the two tests.[16] The separation of the off-diagonal matrix elements into upper and lower triangular pieces is defined by an arbitrary ordering of the sites of the lattice. Possibilities include the natural ordering advocated by Oyanagi and defined by $i = (((t-1)n_z + z - 1)n_y + y - 1)n_x + x$, red-black ordering which is possible if interactions are nearest neighbor, the 16 color ordering used by Rossi et al., and many others.

Hockney[17] has tested the MR-ILU algorithm with several different orderings on several lattice sizes and at several coupling constants, and has confirmed Oyanagi's results. As may be seen in figure 1, the ILU method using natural ordering is much more effective than using red-black ordering, and for medium mass and heavy quarks is much more effective than conjugate gradient with any preconditioning.

The dependence on ordering of sites is so large that it is worth remembering one other algorithm in which ordering makes a large difference. This is the SSOR method. Using natural order, but not red-black order, this method converges as $1/\sqrt{\kappa}$ rather than as $1/\kappa$. The theory given in Hageman and Young does not apply to the minimum residual algorithm on a non-symmetric matrix, but the dependence on the condition number may be tested experimentally on free fermions where the condition is known exactly. As may be seen in figure 2, with careful tuning of the relaxation parameter (or equivalently, of the hopping parameter in the preconditioning matrix) MR-ILU shows the same $\sqrt{\kappa}$ behavior as SSOR.[18] The graph shows the number of sweeps required to obtain $\sqrt{r^2/V} < 10^{-7}$ on a 6^4 lattice. For interacting fermions, the sensitive dependence on the relaxation parameter observed for free fermions does not seem to occur, and the degree of speedup does not seem to depend sensitively on the pion mass as might have been expected from the free fermion results. More work is necessary to see if the $\sqrt{\kappa}$ behavior can be obtained in the interacting case.

Since the minimum residual algorithm fails to converge at light quark masses, it would be desirable to find a preconditioning for the conjugate gradient method as effective as this. For the conjugate gradient algorithm, natural order is no better than

red-black order, if the preconditionings are applied to M and $M^\dagger$ individually. Hageman and Young tell us that conjugate gradient and SSOR may be used together on an SPD matrix to produce an algorithm which depends on the fourth root of the condition number. It should therefore be possible in principle to find an algorithm combines the improved dependence on the condition number with the guaranteed convergence of conjugate gradient, but so far one hasn't been found. Applying ILU preconditioning to the squared matrix $MM^\dagger$ would be very complicated to program.

The question of whether it is worth trying to find an algorithm which converges when minimum residual fails is worth considering. Simulations with very low quark masses on small volumes have poor statistical behavior, with some "exceptional" configurations often dominating the statistics. This is presumably due to fluctuations in the critical value of the hopping parameter which fall with increasing volume. Minimum residual fails when some eigenvalues get a negative real part, which is related to the same effect. It may be that when minimum residual starts to fail, it should be taken as a warning that the physics may be sick there also, and that larger volumes are necessary at that quark mass.

Several twists on the minimum residual method have been invesigated. If no overrelaxation is applied in ILU preconditioning, the preconditioning matrix multiplication can be obtained with no addiional floating point operations.[19] A higher order red-black ILU method has been investigated by Gupta et al.[10] which does not improve the speed of the algorithm but has the interesting property that the quark mass at which it fails is half what is was for the ordinary red-black MR-ILU method. It would be interesting to know the effect of k in CR(k) conjugate residual methods on the failure point. They also report that choosing $\omega \sim 1.3$ in equation 2.2 improves the convergence rate by about 30%.

Hockney has tested Fourier acceleration of the conjugate gradient algorithm for several lattice sizes and coupling constants.[17] He finds a factor of 2.5 to 3 improvement on 8^4 lattices at $\beta = 5.7$, consistant with the results of the Cornell group. However, he finds roughly the same factor improvement on all lattices, rather then the hoped for linear dependence of the improvement on the correlation length. He has also tested Fourier acceleration combined with red-black ILU preconditioning. He finds that the combination requires around 30% fewer sweeps than ILU by itself, which is not enough to make up for the computational overheads of the Fourier transforms on most machines. We therefore have the unexpected possibility that ILU preconditioning does more to reduce critical slowing down than does Fourier acceleration. It is possible that multigrid methods may prove useful for reducing critical slowing down, but these are complicated and have not yet been shown to be effective for QCD.

Lanczos algorithms are identical to conjugate gradient algebraically, but have different round-off properties. Henty, Setoodeh, and Davies[14] have investigated a version of the Lanczos algorithm which uses the Hermitian matrix $\gamma_5 M$ rather than positive definite, Hermitian matrix $M^\dagger M$ $(= (\gamma_5 M)^2)$. In their tests it was about 30–40% faster than the conjugate gradient algorithm, and suffered less problems with roundoff error. The last fact may be important, since round off problems in the conjugate gradient algorithm sometimes make it "converge" to an incorrect answer, and since decisions about when to restart it require some artistry. They have also investigated block versions of the two algorithms which calculate the matrix inverse for several sources at once, which will be useful for quenched calculations.

2.2. Staggered fermions

For staggered fermions, no effective preconditioner has been found. This is puzzling, at least to me. Rossi and Davies tried Fourier acceleration[20] with amazing lack of effect. Fukugita and Ukawa tried many preconditioners including several types of ILU, all without positive effect.[21] The squared staggered fermion matrix used in the conjugate gradient algorithm has the property that the red and black sublattices are completely decoupled.[22] A simple factor of two in speed is obtained by using only one of the sublattices, but the convergence rate in sweeps is not affected by this trick.

As long as $m_q \neq 0$, all staggered fermion eigenvalues have a positive real part. Therefore, the failure of the minimum residual algorithm at small m_q

cannot occur for staggered fermions as it does for Wilson fermions. The conjugate gradient and conjugate residual algorithms were compared for staggered fermions by Fukugita et al.[23] They reported that the conjugate residual algorithm was a factor of 1.7 faster than the conjugate gradient algorithm on their machine. They did not report using the trick of decoupled sublattices. If they did not use it, its use would make the performance of the two algorithms approximately equal. The behavior of the two algorithms at very low quark mass was not discussed.

3. SIMULATION ALGORITHMS: CORRELATION TIMES

3.1. Arguments from free field theory

Our best current understand of the behavior of correlation times in simulation algorithms comes from perturbative arguments, but it it not yet clear that these arguments suffice to explain the observed behavior of the algorithms. In free field theory, the action is diagonal in the Fourier transformed fields:

$$\begin{aligned} S(\phi) &= \tfrac{1}{2}\sum_p (\sum_\mu 4\sin^2(\frac{ap_\mu}{2}) + (am)^2)\phi_{-p}\phi_p \\ &\equiv \tfrac{1}{2}\sum_p \omega_p^2 \phi_{-p}\phi_p. \end{aligned} \quad (3.1)$$

The low momentum modes undergo larger fluctuations than the high momentum modes.

The Metropolis algorithm and the heat bath algorithm proceed by varying a small number of degrees of freedom at a time. The size of the steps, which affect all modes equally, is controlled by the high momentum (large ω) modes. The steps therefore produce only small changes in the low momentum modes, which require many sweeps (of order $N_{steps} \sim (\omega_{max}/\omega_{min})^2$) to random walk to a new value. $\omega_{max}/\omega_{min}$ is approximately equal to the correlation length ξ.

If the effects of quarks are to be included in the action, the Metropolis method requires a Dirac matrix inversion $O(N)$ times per sweep of an N site lattice. Most simulations with dynamical quarks therefore use small step size algorithms which require $O(1)$ matrix inversions per sweep. The Langevin algorithm[24] updates a field by taking a small step in a random direction and moving slightly downhill to maintain equilibrium:

$$\phi_x' = \phi_x + \epsilon\eta_x - \tfrac{1}{2}\epsilon^2 \frac{\delta S}{\delta \phi_x}. \quad (3.2)$$

η is a Gaussian random variable satisfying $< \eta_x^{(n)}\eta_y^{(m)} >= \delta_{mn}\delta_{xy}$ for all time steps m and n and all sites x and y. ϵ is related to the usual Langevin time step dt_L by $dt_L = \epsilon^2/2$. The system requires $1/(\epsilon\omega_{max})^2$ steps to random walk to a new value of the fastest modes which the Metropolis method randomizes in $O(1)$ steps, and $N_{steps} \sim 1/(\epsilon\omega_{min})^2$ steps to random walk to new values of the slowest evolving modes.

The scaling exponent 2 in the above expression for N_{steps} may be reduced to 1 by evolving the field configurations according to classical equations of motion rather than by a random walk. We introduce fictitious momenta π_x conjugate to the field variables ϕ_x and consider a classical system with hamiltonian

$$H = \tfrac{1}{2}\pi_x O_{xy}\pi_y + S(\phi). \quad (3.3)$$

O_{xy} is usually, but not necessarily, taken to be the identity matrix. The equations of motion arising from this hamiltonian, using leapfrog discretization, are

$$\begin{aligned} \pi_x(t+\epsilon/2) &= \pi_x(t-\epsilon/2) - \epsilon\frac{\delta S}{\delta\phi_x}(t), \quad (3.4) \\ \phi_x(t+\epsilon) &= \phi_x(t) + \epsilon\, O_{xy}\pi_y(t+\epsilon/2). \end{aligned}$$

One step of the Langevin algorithm may be thought of as two half steps of a classical trajectory using these equations, with $\pi(t)$ is drawn from the distribution implied by equation 3.3. In the microcanonical (or classical dynamics) algorithm,[25] a simulation run consists of one long classical trajectory. The classical dynamics method has poorer statistical behavior than the hybrid (or refreshed classical dynamics) algorithm,[26, 27] in which the momenta are periodically randomized. Duane argues that on long simulation runs, the statistical errors are dominated by the slowest evolving modes, and that therefore the optimal trajectory length is of order $N_{traj} \sim 1/(\epsilon\omega_{min})$, which would give decorrelation times of $N_{steps} \sim N_{traj} \sim 1/(\epsilon\omega_{min})$.

These small step size algorithms suffer from discretization errors which fall as some power of the

step size ϵ. In versions of these algorithms which are exactly reversible and area preserving in phase space (such as the leapfrog discretization, eqn. 3.4, of the classical equations of motion), step size errors may be removed by performing a Metropolis rejection on configurations generated by Langevin steps or hybrid trajectories.[28, 29] In the hybrid Monte Carlo (or corrected hybrid) algorithm,[29] the momenta are chosen from the distribution defined by the first term in eqn. 3.3, the fields (ϕ, π) are evolved some number of time steps to a new field configuration (ϕ', π'), and the new configuration is accepted with a probability

$$P_{acc} = \min(1, \exp(H(\pi', \phi') - H(\pi, \phi)). \quad (3.5)$$

In methods based on classical equations of motion, it is not necessary to take the time scales of the gauge fields and scalar fields to be the same. Gottlieb et al.[30] proposed slowing the evolution of the scalar fields Φ to zero during each classical dynamics trajectory (which they called the "Φ algorithm"). This has the significant advantage that the matrix inversions may be started from a guess derived from an extrapolation of the previous matrix inverse. They found that this reduced the number of conjugate gradient sweeps required by a factor of 3–5. The scaling behavior of this improvement factor is unknown. The fermionic noise method applied to the hybrid algorithm (the "R algorithm" of Gottlieb et al.) may be thought of as an infinitely fast evolution of the Φ fields between each gauge field step.

3.2. Smallscale tests

If the length of the classical trajectory is much shorter than the shortest time scale in the system, $\epsilon N_{traj} \ll 1/\omega_{max}$, the correlation time in sweeps should fall as $1/\sqrt{N_{traj}}$. I do not know of any attempts to check this dependence carefully. However, many tests in two dimensional models and in QCD on small lattices have shown qualitatively that a finite trajectory length in the hybrid method is preferable to either a single step trajectory (the Langevin limit) or an infinite trajectory length (the microcanonical limit). For example, Fig. 3 shows the autocorrelation time as a function of N_{traj} on a 4^4 lattice with $\epsilon = 0.01$, $\beta = 5.9$, and $am_q = 0.1$.[30]

Duane argues that in free field theory, the optimal value of N_{traj} is of order $1/(\epsilon\omega_{min})$.[26] In contradiction to this, numerical tests have usually found N_{traj} to be optimized at something more like $N_{traj} \sim 1/(\epsilon\omega_{max})$.[27, 30, 31] This changes the expected scaling behavior to $N_{steps} \sim 1/(\epsilon\omega_{max})(\omega_{max}/\omega_{min})^2$, which is disappointing. However, most of the tests have been done with a fixed trajectory length rather than the distribution of trajectories originally discussed by Duane for the hybrid. It is easy to see that in free field theory at least this leads to the observed behavior.[32] Fig. 4 shows the integer multiples nT of the periods T of the modes of a free scalar field with $m = 0.2$ on an 8^4 lattice. For any fixed trajectory length $\tau_0 = \epsilon N_{traj}$ larger than $2\pi/\omega_{max}$, there will be some modes with nT very close to τ_0. These modes will change very little under a classical dynamics trajectory of length τ_0. Tests with a free scalar field show that the hybrid Monte Carlo method fails to equilibrate after a large number of trajectories with a fixed long trajectory length $\tau \sim 1/\omega_{min}$, but equilibrates nicely with a random distribution of long trajectory lengths.[32] It is therefore important to use a distribution of trajectory lengths when optimizing the average trajectory length.

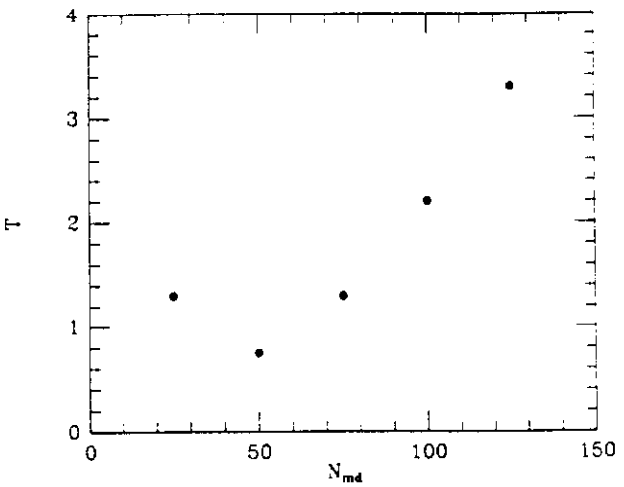

Figure 3: The autocorrelation time T as a function of the number of classical dynamics steps N_d in a trajectory. T has a minimum when N_d is larger than 1 (the Langevin limit) but less than infinity (the microcanonical limit). Data are for the Φ algorithm of Gottlieb et al. on a 4^4 lattice with $\beta = 5.9$ and $m_q = 0.1$.

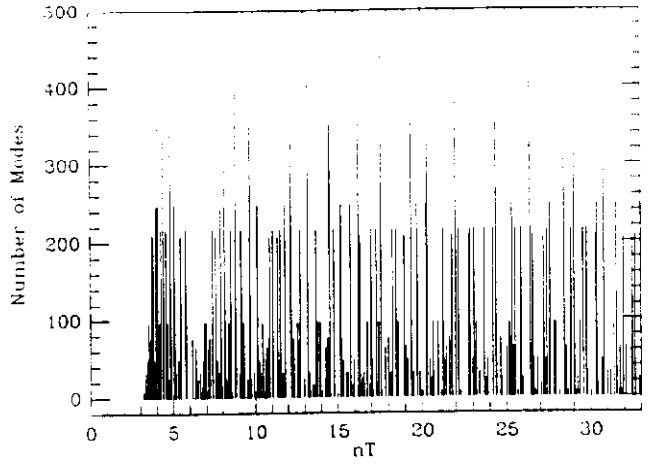

Figure 4: Integer multiples nT of the periods T of the modes of a free scalar field with $m = 0.2$ on an 8^4 lattice. The distribution is dense above a certain T_{min}.

3.3. Largescale tests

Two groups have attempted to compare the efficiencies of the Langevin and hybrid Monte Carlo algorithms in full scale tests. The Langevin algorithm used in the comparisons was a "partial second order" algorithm which removes some but not all of the second order errors.[33] It does not remove errors in the fermionic sector which are thought to dominate at low m_q. The two groups reached seemingly opposite conclusions: Fukugita et al.[34, 35] estimated that Langevin was around five times as efficient as hybrid Monte Carlo, while Gupta et al.[10] estimated a similar factor, but in the opposite direction.

There are quite a few assumptions in the comparisons which may contribute to the apparent discrepancy, but the most important factor which makes comparison difficult at present is poor statistics and lack of statistical information in the published data. When comparing the efficiencies of algorithms, it is the statistical error bars on physical quantities which are compared. This means that to estimate the statistical reliability of the comparisons, we need to estimate the accuracy of the error bars: we need error bars on the error bars. These have almost never been published up to now. Figure 5 shows the evolution of the plaquette as a function of classical dynamics time in a hybrid Monte Carlo run.[36] It shows obvious correlations lasting hundreds of trajectories, the length of the run. This implies

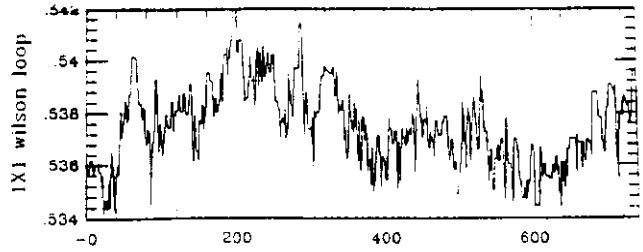

Figure 5: Some data for hybrid and hybrid Monte Carlo simulations show correlations lasting hundreds or thousands of trajectories. Here, the hybrid Monte Carlo data of the LANL collaboration show the evolution of the plaquette as a function of classical dynamics time.

that the relative accuracy of the statistical errors on this quantity must be at least of order 1. Similar plots for the Langevin algorithm do not seem to show such obvious correlations,[35] but it is difficult to guess the significance of such eyeball error estimation.

A reasonable temporary conclusion is that Langevin and the hybrid have been shown to be comparably efficient at the order of magnitude level, but without more detailed understanding of errors, it is not possible to conclude more at present. Quite apart from the question of finding the best algorithm, it is disconcerting that the statistical errors from large full QCD simulations cannot be taken very seriously in the presence of such long and not understood correlations.

3.4. Further work

Free field theory arguments lead to the expectation that that the hybrid and hybrid Monte Carlo algorithms should be far superior to the first order Langevin algorithm:

Algorithm	N_{steps} for decorrelation (free field)
Metropolis	$(\omega_{max}/\omega_{min})^2$
Langevin	$1/(\epsilon\omega_{min})^2$
Hybrid	$1/(\epsilon\omega_{min})$

ω_{max} is of order one in lattice units, $(\omega_{max}/\omega_{min})$ is probably of order 5–10 in QCD simulations, and ϵ is usually around 0.01-0.05.

Although the large scale tests are not fully reliable statistically, it is still fair to say that the order of magnitude superiority of hybrid over Langevin which

might have been expected for small ϵ has not yet shown up. On the contrary, there is evidence for correlations lasting hundreds of classical dynamics time units which are not expected from free field theory.

It is possible that nonperturbative effects are responsible for the very long correlation times in hybrid and hybrid Monte Carlo. On the other hand, there is much unused freedom in the formulation of the algorithms. I have already mentioned that the use of a fixed trajectory length can be catastrophic. I don't know of any experimentation with distribution of trajectory lengths in QCD, but in tests for QED simulations, Kogut and collaborators found that careful choosing of the distribution of trajectory lengths was crucial in obtaining reasonable correlation times for the low momentum photon propagator.[37]

Another example has to do with ergodicity in the short wavelength modes. It is sometimes argued that long trajectory lengths are ineffective in the hybrid and hybrid Monte Carlo algorithms because short wavelength modes with similar frequencies will tend to stay correlated during a long trajectory. The KAM (Kolmogorov-Arnold-Moser) theorem is sometimes invoked to motivate worry about ergodicity. It says that systems of oscillators with sufficiently small perturbations are not ergodic. However, the KAM theorem is unlikely to be rigorously applicable for two reasons. First, it should become inapplicable for large enough perturbations, and it is a reasonable guess that for lattice QCD, this may occur at the deconfining transition. Second, even when applicable, it deals with ergodicity in the full phase space. We require that every point in the configuration space be reachable, but are less concerned with correlations in the space of the fictitious momentum variables. A classical mechanics theorem which may be more relevant for us is the *theorem on the averages.*[38] It states that for a set of harmonic oscillators with independent frequencies (all frequency ratios irrational), the average of a function of the coordinates over time is the same as the average of the function over space. This implies that the motion is dense in configuration space, even if the modes are decoupled. From this point of view, the choice of $O_{xy} \equiv 1$ in equation 3.3 is very bad, in that in free field theory it ensures large numbers of modes with identical frequencies. A better choice would be to take O_{xy} to have diagonal elements randomly distributed in some range, so that even in free field theory, trajectories will be dense in configuration space (though not in phase space).

A similar point may be made concerning the exponent of π_x in equation 3.3.[39] The quadratic form is familiar from real physical systems, but may be much worse than a quartic form or some distribution of exponents for ensuring a rapid randomizing of the configuration. A nonquadratic kinetic energy term would also help to produce a rapid exchange of energy between modes.

It may be that after the most sophisticated tuning of the algorithms along these lines, corrleation times are still much longer expected from free field theory arguments. For example, it may be that long tunneling times are required to move between different topological sectors of the path integral. Some hint that this may be a factor comes from the experience of the Grand Challenge spectrum collaboration, who find that very long correlations are much more obvious in measurements of the topological susceptibility than in measurements of hadron masses.[40] It is known that in compact QED, the presence of nonlocal topological structures such as monopole loops causes extremely long correlation times.[41] Their importance decreases as the volume increases. Such structures may be important in nonabelian gauge theory as well.[42] If topological effects in QCD which do not vanish at large volumes are responsible for very long correlation times, further improvement may be impossible without new physics ideas to encourage tunneling between topological sectors.

3.5. Correlation lengths

In free field theory, the ratio $\omega_{max}/\omega_{min}$ which governs correlation times is of order the correlation length of the system. In this subsection, I discuss complications which occur for full QCD.

The Cornell group has shown in tests with pure gauge theory that Fourier acceleration of the gauge dynamics can improve the performance of the Langevin algorithm.[43] The hybrid and hybrid Monte Carlo algorithms may be Fourier accelerated by taking O_{xy} in equation 3.3 to have off diagonal elements.[44] If fermion loops are included, the gauge

fixing overhead required for Fourier acceleration is masked by the time spent for fermion matrix inversion, so it might be that the method is useful even if it produces a constant factor of improvement rather a change in the scaling exponent. On the other hand, inclusion of fermions presents new problems which cannot be analyzed in weak coupling (which is where we get our intuition about Fourier acceleration). It seems to be the case that the step size in simulation algorithms is limited by the effects of the small, non-perturbative eigenvalues of the Dirac matrix, rather than by the high momentum gauge field modes as in pure gauge theory. The derivatives of the inverses of these small eigenvalues with respect to the gauge field modes is not obvious, so it is not obvious how to choose an acceleration function. Therefore, the appropriate acceleration function may have to be investigated numerically by calculating the momentum spectrum of the force on the gauge field produced by the fermion part of the action, equation 1.2.

The related question of the dependence of correlation times on correlation lengths is similarly more complicated when fermions are included. The usual intuition about the relation between correlation lengths and correlation times may not be valid in theories with nonlocal actions. The most obvious guess is that correlation times are governed by the longest correlation length in the system, the pion Compton wavelength in QCD. The usual folklore that simulation correlation times rise as a power of the correlation length (the linear power in the case of the hybrid family) depends on the fact that the longest wavelength modes produce the smallest force terms in the simulation algorithms. However, as we have seen, the force terms in full QCD are thought to be dominated by the small eigenvalue modes in the Dirac matrix which are related to chiral symmetry breaking and the long wavelength propagation of the pion. It is not obvious that these modes evolve slowly. A numerical investigation of the Fourier transform of the fermionic force, equation 1.2, might shed light on this question, too.

Naively, correlation times are expected to rise linearly with the correlation length in the hybrid algorithm, and rise quadratically with the correlation length in the Langevin algorithm. Whether this is what occurs in practice when the action is nonlocal, and whether relevant correlation length is the pion Compton wavelength or the QCD scale, remain to be demonstrated.

An additional complication is that physical quantities are expected to approach finite limits in QCD as the pion mass approaches zero. Therefore, as pion modes with longer and longer wavelength are added to the system by reducing the quark mass, even if they are poorly averaged they may have little effect on physics.

4. SIMULATION ALGORITHMS: DEPENDENCE OF THE STEP SIZE ON V, a, and m_q

4.1. Volume dependence in hybrid Monte Carlo

4.1.1. The leapfrog method

The V dependence of the step size in hybrid Monte Carlo was derived by Gupta, Kilcup, and Sharpe[45] extending arguments due to Creutz[46]. They show that in order to keep constant acceptance, ϵ must be decreased as ${N_{site}}^{-1/4} \sim aV^{-1/4}$, where V is the physical volume of the lattice. Creutz's version of the argument employs the formula

$$< H' - H >= \tfrac{1}{2} < (H' - H)^2 > + O((H' - H)^3). \tag{4.1}$$

With the leapfrog method, $< (H' - H)^2 >$ is $O(\epsilon^4)$. If the volume is larger than the correlation length, it is an extensive quantity. (By using the formula $aV^{-1/4}$ rather than simply $V^{-1/4}$ I am jumping to the conclusion that correlations between the degrees of freedom do not affect the argument, which may or may not be justified.) This implies an acceptance probability falling as

$$P_{acc} \sim \exp(-C N_{site}\, \epsilon^4), \tag{4.2}$$

where C is an unknown constant.

4.1.2. Higher order hybrid Monte Carlo

Methods for integrating the classical equations of motion which are accurate to a higher order than the ϵ^2 of the leap frog method will yield a hybrid Monte Carlo with computing requirements growing as a weaker function of the volume than $T \sim V^{5/4}$. Kennedy has investigated methods for explicitly dealing with the complicated corrections to the action

generated by finite step size.[47] A simple method for reducing these errors has been suggested by Campostrini and Rossi,[48] and has been extended and explained by Creutz and Gocksch.[49] The method is to take i steps in the forward direction, one larger step back, and then i more steps forward. If a discrete transformation on phase space

$$T_n(\epsilon) = \exp(H\,\epsilon) + \Delta\epsilon^{n+1} + \ldots \qquad (4.3)$$

has errors of order ϵ^{n+1}, then a combined transformation

$$T_{n+2}((2i-s)\epsilon) \equiv T_n(\epsilon)^i T_n(-s\epsilon) T_n(\epsilon)^i \qquad (4.4)$$

has errors of order ϵ^{n+3} if $s = (2i)^{\frac{1}{n+1}}$. If T_2 is taken to be the ordinary leap frog method, the method may be iterated to remove step size errors to any desired order. As the number of degrees of freedom is increased, it becomes profitable to go to higher iterations of the method. Creutz and Gocksch estimate computer time requirement growing as $T \sim V \exp(\sqrt{\log V})$.

Creutz and Gocksch have tested their higher order algorithms on the Hubbard model with a relatively small ($O(10^3)$) number of degrees of freedom. They found that the lowest order leap frog method was "hard to beat". For the much larger lattices used in QCD simulations, the higher order method may still be important, but see the next subsection on the effects of fermions.

4.1.3. Analytic calculation of acceptance in free field theory

For free field theory, acceptance rates in hybrid Monte Carlo and Langevin Monte Carlo may be calculated analytically.[50] The acceptance rate in these calculations has the form $P_{acc} \sim \mathrm{erfc}(C\sqrt{N\epsilon^p\sigma})$, where erfc is the complement of the error function, C is a calculable constant, and N is the number of degrees of freedom. σ is obtained from a sum over the frequencies to some power: $\sigma = \frac{1}{N}\sum\omega^p$.

It should be possible to calculate the acceptance rate for a four dimensional lattice of free gluons, and to determine the lattice volume at which it would become more efficient to perform hybrid Monte Carlo simulations with the second order Campostrini and Rossi method, rather than the leap frog method. This might yield some intuition about the expected behavior of the algorithms in real QCD simulations.

4.2. Dependence on a and m_q in Langevin, hybrid, and hybrid Monte Carlo

Perturbative analysis does not explain the a and m_q dependence of the step size. It is well established numerically that as $m_q \to 0$, the step size ϵ must also approach zero if systematic errors or acceptance are to remain constant in Langevin, hybrid, and hybrid Monte Carlo. For example, Gupta[51] reports that he and his collaborators needed to use $\epsilon = 0.06,\ 0.026,\ 0.011$ at $am_q = 0.1,\ 0.05,\ 0.025$ in order to maintain 70% acceptance. This would indicate that for HMC, ϵ must be reduced as a power of am_q between 1 and 1.5. Similar results have been reported for the Langevin[52] and hybrid[53] algorithms.

The form of the finite step size corrections has been derived for some algorithms. When the action $S = S_{gauge} + \Phi^\dagger \frac{1}{M^\dagger M}\Phi$ is used, they have a relatively simple form. For the Langevin equation, Fukugita, Oyanagi, and Ukawa[33] give for the fermionic corrections to the action induced by the finite step size as

$$\Delta S = \tfrac{1}{2}\epsilon^2\left[-\frac{1}{4}\Phi^\dagger \frac{1}{M^\dagger M M^\dagger M}\Phi + 2Tr\frac{1}{M^\dagger M}\right]. \qquad (4.5)$$

Both terms are of order $\frac{1}{M^\dagger M}$. For the hybrid algorithm using one particular form of the leapfrog method, Duane and Kogut[27] give

$$\Delta S = -\frac{1}{4}\epsilon^2 Tr\frac{1}{M^\dagger M}. \qquad (4.6)$$

Different forms of the leap frog method may give different error terms.[45]

For the fermionic noise method, there are additional errors induced which cannot be expressed as corrections to the action.[54] For hybrid Monte Carlo, Gupta et al.[10] give a schematic argument based on counting powers of $1/M$ that ϵ should be decreased as $(m_q)^{\frac{7}{4}}$ to maintain constant acceptance, if long trajectory lengths are used.

A naive power counting estimate suggests that $Tr\frac{1}{M^\dagger M}$ is dominated by the smallest eigenvalues of M and diverges as $1/(am_q)^2$ as $am_q \to 0$. This has led to the guess that ϵ must approach zero as am_q to maintain constant systematic errors. However, for staggered fermions at least, we can actually evaluate the expressions for the errors in equations 4.5 and 4.6, and find a different answer. The eigenvalues for

the staggered fermion Dirac operator can be written $i\lambda + am_q$. We can write the chiral condensate as

$$a^3 < \bar{\Psi}\Psi > \sim tr\frac{1}{M} = \sum \frac{1}{i\lambda + am_q} \quad (4.7)$$

$$= \sum \frac{am_q}{\lambda^2 + (am_q)^2} \quad (4.8)$$

This means that the trace in the error term, equation 4.6, has the form

$$Tr\frac{1}{M^\dagger M} = \sum \frac{1}{\lambda^2 + (am_q)^2} \sim \frac{a^3 < \bar{\Psi}\Psi >}{(am_q)}. \quad (4.9)$$

If this term dominated the error, in the chirally broken phase ϵ would have to be decreased as $1/\sqrt{m_q}$ rather than as $1/m_q$ to maintain constant errors. In the chirally symmetric phase, where perturbation theory is valid, the error term stays finite as $m_q \to 0$, meaning that ϵ does not have to vanish with m_q. The naive power counting argument is therefore wrong in this case, but the idea that adding a power of $1/M$ adds a power of $1/(am_q)$ is partly right. If it is true that in the power series for the errors, adding a power of ϵ adds a power of $1/M$ and that adding a power of $1/M$ adds a power of $1/(am_q)$, then it could still be true that ϵ must be reduced as am_q to maintain the stability of the power series. The data cited above is for the Langevin and hybrid algorithms with fermionic noise, and for hybrid Monte Carlo. They mostly agree better with $\epsilon \sim m_q$ than with $\epsilon \sim \sqrt{m_q}$, but the statistical accuracy of the results may not be high.

The MT_c collaboration has produced a phenomenological graph showing the effects of several parameters on the the acceptance rate in hybrid Monte Carlo[56] (figure 6). The graph was developed more to guide the choice of parameters in simulation runs than to test theories about hybrid Monte Carlo. It shows that the acceptance rate from all the runs lies on an approximately universal curve when plotted against the phenomenologically obtained function $\epsilon\,(\bar{\chi}\chi)^{2/3}V^{1/4}(m_q)^{-1}$, where the dimensionful quantities are presumably measured in lattice units. In these runs, m_q varied by a factor of five, V by a factor of nearly 4^4. $\bar{\chi}\chi$ varied by an order of magnitude since some of the runs were confined and some deconfined. The quoted dependence on volume is

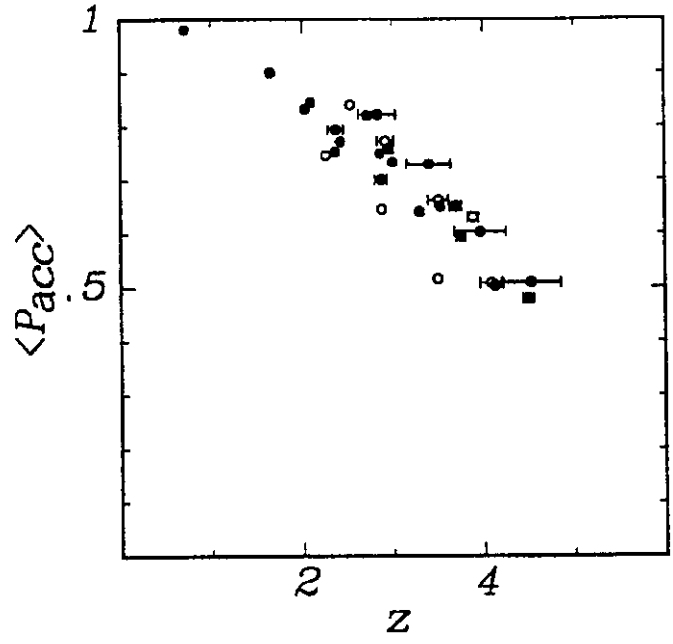

Figure 6: The hybrid Monte Carlo acceptance rates for many runs of the MT_c collaboration lie on an approximately universal curve when plotted against a phenomenologically obtained function $z = \epsilon\,(\bar{\chi}\chi)^{2/3}V^{1/4}(m_q)^{-1}$.

precisely what is expected and the dependence on m_q is in the reasonable range. It is interesting to see a strong dependence on $\bar{\chi}\chi$ as suggested by equation 4.9, although the exponent is different. It would be very interesting to get some guesses as to the reasonable range for this exponent and the others suggested by these data. (This may not be straightforward in practice since the data are not statistically consistent, as they need not be for such a phenomenological function.)

If the right variable in the function for the acceptance rate were really that given by the MT_c collaboration, the factors of a going along with the dimensionful quantities would imply that ϵ did not need to be reduced as $a \to 0$, which would be very nice. It may be more likely, however, that powers of V and $< \bar{\Psi}\Psi >$ do not go hand in hand with powers of am_q, and that the right function is more complicated, for example an unknown function of ϵ/am_q multiplying $V^{1/4}$, and some function of ϵ, $< \bar{\Psi}\Psi >$, and a. This means that it is important to test the a dependence of ϵ separately from the m_q dependence.

The assumption that powers of ϵ in the power series of step size errors are matched by powers of

$1/am_q$ but not by powers of $< \bar{\Psi}\Psi >$ leads to the unfortunate conclusion that higher order integration schemes cannot reduce the unpleasant dependence of the step size on am_q to better than linear power. Explicit calculations of error terms demonstrating this assumption would be useful. A much nicer situation exists for pure gauge theories. The work of Zinn-Justin showing that the theories corresponding to Langevin simulations at finite and zero ϵ are in the same universality class implies that discretization errors for fixed ϵ vanish as $a \rightarrow 0$.[57] The situation for full QCD simulations can be different, since the effective action is nonlocal.

5. PSEUDOFERMIONS

The bulk of algorithmic work in recent years has focused on algorithms based on the Weingarten-Petcher action and the closely related fermionic noise method of the Cornell group. The original pseudofermion algorithm of Fucito et al.[8] used a Metropolis simulation at each time step to estimate the matrix inverse for the force term, and there is very little doubt that this is a far more time consuming method than methods using the deterministic minimum residual and conjugate gradient algorithms. However, just as subsequent work dramatically improved algorithms based on the Weingarten-Petcher action, improvements to the pseudofermion method may keep it competitive.

The most obvious defect in the original pseudofermion method compared with other algorithms is the use of the Metropolis method to estimate the matrix inverse necessary for the fermionic force. This method is expected to require a number sweeps proportional to $\kappa_{Dirac}{}^2$ to include correctly the effects of the effects of the small eigenvalue fermionic modes. This may be remedied by the use of a first order Langevin equation, which is expected to have correlation times linear in κ_{Dirac}. This method has recently been investigated by Hamber,[58] who is optimistic about its prospects. Even more effective may be the replacement of the small step size first order Langevin formula with the large step size overrelaxed method. It is not clear whether to expect deterministic methods to be comparable in efficiency with the large step size or small step size Monte Carlo methods. It is possible that an improved Monte Carlo matrix inverse estimation may be quite competitive with the deterministic methods. If a deterministic method is found converging in $O(\sqrt{\kappa_{Dirac}})$ sweeps (as optimistically suggested above), it is unlikely that any of the known Monte Carlo methods will be competitive.

A second argument sometimes made against the pseudofermion approach has to do with the fact that in its small step size version, it may be thought of as a disguised version of the Langevin algorithm. (See, for example, Toussaint's discussion.[9]) However, although there are good theoretical reasons for expecting the superiority of the hybrid family over the Langevin family, so far they have not been conclusively confirmed in large scale simulations. In addition, algorithm comparisons highly dependent on the effects of systematic errors, which will be different in pseudofermion-Langevin than in the usual Langevin. It might even be possible to develop a hybrid method using a pseudofermionic estimate of the fermionic force instead of a minimum residual or conjugate gradient based estimate, but the buildup during long trajectories of systematic errors from the quark matrix inversion reported by Gupta et al.[10] might be a problem for this approach.

The pseudofermion method has recently been investigated on medium sized lattices by Potvin et al.[59] They emphasized the need for careful control of systematic errors and reached pessimistic conclusions regarding its effectiveness compared that of the hybrid algorithm. However, while there are some good reasons why the pseudofermionic approach has lost popularity lately, it is not definitively established whether or not an improved member of this family can turn out to be competitive.

6. CONCLUSIONS

An ideal approach to the study of algorithms would be:

1. Calculate as many of the exponents and coefficients as possible in the expressions for the computing requirements of algorithms.
2. Test the calculations numerically.
3. Measure the uncalculated coefficients.

4. Predict the ultimate computing requirements as parameters such as a, V, and m_q are scaled to their desired values.

Ranges of reasonable guesses for the scaling behavior of some existing algorithms are given in the following table.

Algorithm	Step size (inverse)	ξ	N_q
Lang.	$(\frac{1}{am_q})^{\frac{1}{2},1}$	$\frac{1}{(am_\pi)^2}, \frac{1}{(a\Lambda_{QCD})^2}$	$(\frac{1}{am_q})^{\frac{1}{2},1}$
Hyb.	$(\frac{1}{am_q})^{\frac{1}{2},1}$	$\frac{1}{am_\pi}, \frac{1}{a\Lambda_{QCD}}$	$(\frac{1}{am_q})^{\frac{1}{2},1}$
HMC	$\frac{V^{\frac{1}{4}}}{a}(\frac{1}{am_q})^{\frac{1}{2},\frac{7}{4}}$	$\frac{1}{am_\pi}, \frac{1}{a\Lambda_{QCD}}$	$(\frac{1}{am_q})^{\frac{1}{2},1}$

The least controversial guess for the exponent of $\frac{1}{am_q}$ in the inverse step size and in the number of matrix inversion sweeps N_q required is 1. The least controversial guess for the correlation length ξ governing critical slowing down is am_π. The guesses regarding the step size and correlation length are particularly uncertain at present.

Current simulations of full QCD typically are done at values of m_π^2 ($\sim m_q$) which are an order of magnitude larger than the physical value. The range of exponents in the table indicates a two order of magnitude uncertainty in the increase of CPU time required per site, were the simulations to be done at the physical value of m_π. If I make the guess that the lattice spacing a needs to be less than 0.1 fermi, and may need to be as small as 0.03 to 0.05 fermi, there is a similar uncertainty due to this parameter, due to both scaling exponents and to the increase in the number of sites.

The price of going to large physical volume may be less severe if the use of lattice sizes larger than the correlation length leads to a commensurate improvement in statistics. (In addition to simply reducing finite volume errors, there are other reasons for doing simulations on the largest volume possible: matrix inversion algorithms are better behaved on large volumes, "exceptional" configurations are rarer at large volumes, the large QED correlation times due to topological effects were reduced at large volumes.)

Although we are not yet in a position to reliably predict computing requirements at the physical quark mass for given values of the volume and lattice spacing, quite a bit is known, both theoretically and numerically. A good understanding of all of the exponents and coefficients is an achievable goal, although perhaps just as difficult as a good understanding of the hadron spectrum itself. It is a desirable goal, both for the practical purpose of understanding computing requirements, and for purpose of understanding our theoretical tools.

ACKNOWLEDGEMENTS

I would like to thank Peter Lepage, Andreas Kronfeld, Doug Toussaint, and George Hockney for useful discussions.

REFERENCES

1. As suggested by Ken Wilson at this conference.
2. G. Parisi in *Progress in Gauge Field Theory*, edited by G. 't Hooft et al. Plenum (1984) 531.
3. D. Weingarten in *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 447.
4. See the review by S. L Adler in *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 437.
5. See the review by U. Wolff in these proceedings.
6. D. Weingarten and D. Petcher, Phys. Lett B 99 (1981) 333.
7. G. G. Batrouni et al., Phys. Rev. D32 (1985) 2736.
8. F. Fucito et al., Nuc. Phys. B180 (1981) 369.
9. D. Toussaint in the proceedings of *New Developments in Hardware and Software for Computational Physics*, Buenos Aires (1988), to be published in Comput. Phys. Commun.
10. R. Gupta et al., Phys. Rev. D 40 (1989) 2072.
11. L. A. Hageman and D. M. Young, *Applied Iterative Methods*, Academic Press (1981).
12. G. Katz et al., Phys. Rev. D 37 (1988) 1589.
13. Y. Oyanagi, Comp. Phys. Commun. 42 (1986) 333.
14. D. Henty, R. Setoodeh, and C. T. H. Davies, Glasgow preprint GUTPA/89/6/1.
15. P. Rossi, C. T. H. Davies, and G. P. Lepage, Nuc. Phys. B297 (1988) 287.
16. Y. Oyanagi, J. Info. Proc. 11 (1987) 32.
17. G. M. Hockney, in these proceedings.

18. P. B. Mackenzie, unpublished.

19. T. A. DeGrand, Comput. Phys. Commun. 52 (1988) 161. Hageman and Young call this technique reduced system conjugate gradient.

20. P. Rossi and C. T. H. Davies, Phys. Lett. B202 (1988) 547.

21. M. Fukugita, private communication.

22. J. Polonyi et al., Phys. Rev. Lett. 53 (1984) 644; O. Martin and S. W. Otto, Phys. Rev. D 31 (1985) 435.

23. M. Fukugita, S. Ohta, Y. Oyanagi, and A. Ukawa, Phys. Lett. B 191 (1987) 164.

24. G. Parisi and Y. S. Wu, Sci. Sin. 24 (1981) 483.

25. D. Callaway and A. Rahman, Phys. Rev. D28 (1983) 1506; J. Polonyi and H. W. Wyld, Phys. Rev. Lett. 51 (1983) 2257.

26. S. Duane, Nuc. Phys. B257 (1985) 652.

27. S. Duane and J. Kogut, Nuc. Phys. B275 [FS17] (1986) 398.

28. R. T. Scaletter, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B34, (1986) 7911.

29. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B195 (1987) 216.

30. S. Gottlieb et al., Phys. Rev. D 35 (1987) 2531.

31. K. Bitar et al., Nuc. Phys. B313 (1989) 348, B313 (1989) 377.

32. P. B. Mackenzie, Phys. Lett. B 226 (1989) 369. See also Weingarten's comments.[3]

33. M. Fukugita, Y. Oyanagi, and A. Ukawa, Phys. Rev. D 36 (1987) 824.

34. A. Ukawa, in *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 463.

35. M. Fukugita and A. Ukawa, Nuc. Phys. B 300 [FS22] (1988) 433.

36. From the talk by R. Gupta at this conference.

37. J. Kogut, private communication.

38. See, for example, V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York.

39. G. Parisi, private communication.

40. D. Toussaint, private communication.

41. V. Grösch et al., Phys. Lett. 162B (1985) 171.

42. See, for example, J. Smit and A. van der Sijs in these proceedings.

43. C. T. H. Davies et al., Cornell preprint CLNS 89/947 (1989).

44. J. B. Kogut, Nuc. Phys. B275 [FS17] (1986) 1; S. Duane et al., Phys. Lett. B 176 (1986) 143.

45. R. Gupta, G. W. Kilcup, and S. R. Sharpe, Phys. Rev. D 38 (1988) 1278.

46. M. Creutz, Phys. Rev. D 38 (1988) 1228.

47. A. D. Kennedy, in *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 457.

48. M. Campostrini and P. Rossi, Pisa preprint IFUP-TH 12/89 (1989).

49. M. Creutz and A. Gocksch, Brookhaven preprint BNL 42601, (1989).

50. A. D. Kennedy and B. J. Pendleton, in preparation; H. Gausterer and M. Salmhofer, Phys. Rev. D 40 (1989) 2723.

51. R. Gupta, in *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 473.

52. M. Fukugita and A. Ukawa, Phys. Rev. D 38 (1988) 1971.

53. Figure 4 in S. Gottlieb et al., Phys. Rev. D 36 (1987) 3797.

54. G. Batrouni, Phys. Rev. D33 (1985) 1815; A. S. Kronfeld, Phys. Lett B 172 (1986) 93.

55. A. S. Kronfeld and P. B. Mackenzie, unpublished.

56. A. Irbäck, in these proceedings.

57. J. Zinn-Justin, Nuc. Phys. B275 (1986) 135.

58. H. Hamber, in these proceedings.

59. J. Potvin et al., *Lattice 88*, Nuc. Phys. B (Proc. Suppl.) 9 (1989) 490.