Fourier acceleration in lattice gauge theories. II. Matrix inversion and the quark propagator

G. Katz,* G. Batrouni,[†] C. Davies,[‡] A. Kronfeld,[§] P. Lepage, P. Rossi,** B. Svetitsky,^{††} and K. Wilson

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853

(Received 1 June 1987)

We present a new algorithm for inverting the quark propagator in lattice QCD that removes the critical slowing down associated with the continuum limit. Both a theoretical analysis and a practical implementation are given, together with results from numerical simulations that illustrate the algorithm. The acceleration speeds convergence by factors of 3-4 on small lattices and should lead to proportionately larger gains on larger lattices.

I. INTRODUCTION

The most significant obstacle to fully realistic simulations of QCD on the lattice is the inclusion of dynamical quarks. Most algorithms involve repeated solution of **Euclidean Dirac equations**

$$M\psi = \eta , \qquad (1)$$

where

$$M = i\partial + gA + m_q , \qquad (2)$$

in the lattice approximation. The cost of solving this system of linear equations generally dominates all other costs inherent in the simulation. Relaxation methods must be used in solving these equations as the systems are generally very large (e.g., $16^4 \times 12$ equations), although quite sparse. Unfortunately this is numerically costly because conventional relaxation algorithms suffer from critical slowing down as the quark mass m_a and lattice spacing a are decreased. The cost usually becomes prohibitive long before realistic values for the uand d quark masses are attained. In this paper we describe a simple acceleration technique that dramatically reduces the cost required for a numerical solution of equations such as the Dirac equation.¹

As is frequently the case for physical equations this equation is badly ill conditioned in the continuum limit, as lattice spacing $a \rightarrow 0$. It has eigenvalues ranging from approximately m_a to $p_{\text{max}} \approx 1/a$. Traditional relaxation schemes have no problem dealing with the shortwavelength, large-eigenvalue modes. However, the physics lies mostly in the long-wavelength modes and these relax at a rate that is typically

$$\frac{p_{\max}}{m_q} \approx \frac{1}{am_q} \tag{3}$$

times longer than the ultraviolet modes. As we will demonstrate this problem can be greatly reduced by preconditioning the matrix M so as to reduce the range of its eigenvalues. For example, one might try replacing *M* by $M_0^{-1}M$, where $M_0 = i\partial + m_{\text{eff}}$ is the free propagator with some effective mass. This procedure does indeed work well but it is not obvious why. The original

operator M contains a random field, the gauge field A_{μ} , and it is not at all clear that the eigenmodes of this random matrix are approximated by the eigenmodes of M_0 , i.e., by Fourier modes. Indeed this preconditioning fails completely unless the gauge for A_{μ} is carefully chosen.

In Sec. II we explore the conditions under which one expects such preconditioning to work, and in particular illustrate the importance of gauge fixing. In Sec. III we illustrate our findings by presenting the results of QCD simulations on an 8⁴ lattice and for a variety of coupling constants.

II. THEORETICAL BACKGROUND

A. Free Dirac theory

To set the stage we illustrate the problem of critical slowing down by applying the simplest of relaxation methods, the Jacobi algorithm, to free Dirac theory. Rewriting the Dirac equation as

$$\psi = \psi + \epsilon \left(\eta - M \psi \right) \tag{4}$$

suggests an iterative algorithm for its solution:

$$\psi^{(n+1)} = \psi^{(n)} + \epsilon (\eta - M \psi^{(n)}) ,$$

$$\psi^{(0)} = 0 .$$
(5)

The *n*th iterate is simply

$$\psi^{(n)} = [1 + (1 - \epsilon M) + \cdots + (1 - \epsilon M)^{n-1}]\epsilon\eta$$
$$= \psi^{(\infty)} - (1 - \epsilon M)^n \psi^{(\infty)}.$$
(6)

Clearly step size ϵ is limited by the convergence criterion $\|1 - \epsilon M\| < 1$, indicating that

$$\epsilon \approx \frac{1}{p_{\max} + m} \tag{7}$$

is optimal for the free Dirac operator. However, for the lowest-momentum modes ϵM is of order ϵm_a $\ll \epsilon p_{\max} \approx 1$ so that $(1 - \epsilon M)^n \approx e^{-\epsilon M n}$ becomes small only after

$$n \approx \frac{1}{\epsilon M_{\min}} \approx \frac{p_{\max}}{m}$$
 (8)

©1988 The American Physical Society

iterations.

The solution is obvious for this case. One wants $\epsilon M \approx 1$ for all modes and so the step-size ϵ should be chosen to be a matrix. As the eigenmodes of M are just Fourier modes the optimal form for the step-size matrix is

$$\epsilon = \hat{F}\epsilon(p)\hat{F}^{-1}, \qquad (9)$$

where in this case

$$\epsilon(p) = \frac{1}{\not p - m_q} \tag{10}$$

and \hat{F} represents a Fourier transform. With this choice $(1-\epsilon M)$ can be made small and convergence fast for all modes. In effect, we are employing the Fourier transform to resolve the different modes of M, allowing us to use larger Jacobi steps for the longer-wavelength modes.

Parenthetically we note that the norm of the residue

$$\|\boldsymbol{r}\| = \|\boldsymbol{\eta} - \boldsymbol{M}\boldsymbol{\psi}\| \tag{11}$$

is a very poor measure of convergence for such illconditioned problems. The residue after n iterations has norm

$$\|r^{(n)}\| = \|\eta - M\psi^{(n)}\|$$

= $\|M(\psi^{(\infty)} - \psi^{(n)})\|$ (12)

A large error in the long-wavelength modes of $\psi^{(n)}$ is completely masked in $||r^{(n)}||$ by the fact that M very nearly vanishes for such modes. As these are the very modes that lead to critical slowing down the residue can be very misleading about the errors. A far better indicator of error is something like

$$\|\delta\psi^{(n)}\| \approx \frac{\|\psi^{(n)} - \psi^{(n-1)}\|}{\|r^{(n-1)}\|} - 1.$$
(13)

This works since from Eq. (6)

$$\psi^{(n)} \to \psi^{(\infty)} - \zeta^n \nu \tag{14}$$

for large *n*, where ζ is the dominant eigenvalue of $(1 - \epsilon M)$ (for the Jacobi algorithm) and v is the corresponding eigenvector. Consequently,

$$\|\psi^{(n-1)} - \psi^{(n)}\| \to \zeta^n \left[\frac{1}{\zeta} - 1\right] \|\nu\| ,$$

$$\|\psi^{(n)} - \psi^{(\infty)}\| \to \zeta^n \|\nu\| ,$$
(15)

while

$$\frac{\|r^{(n-1)}\|}{\|r^{(n)}\|} \to \frac{\zeta^{n-1}\|Mv\|}{\zeta^{n}\|Mv\|} = \frac{1}{\zeta}.$$
 (16)

Equation (13) then follows immediately for *n* large. Notice that in critical situations where $r^{(n)} \approx r^{(n-1)}$ and $\zeta \approx 1$ the actual error is far larger than the change in ψ over a single iteration. Equation (14) is valid for all common relaxation methods and so these comments apply more generally than just to the Jacobi algorithm. In general one might want to use more widely separated iterations to minimize fluctuations:

$$\|\delta\psi^{(n)}\| \approx \frac{\|\psi^{(n)} - \psi^{(n-k)}\|}{\|r^{(n-k)}\|} - 1$$
(17)

although k = 1 proved to be quite reasonable for all our studies using the conjugate-gradient algorithm (because $r^{(n)}$ decreases monotonically for this algorithm).

B. Dirac operator with random fields

Unlike M, the step-size matrix ϵ in the Jacobi algorithm, or more generally the preconditioning matrix M_0^{-1} , is not a sparse matrix. Any preconditioning that will affect long-wavelength modes must necessarily be quite nonlocal. Consequently the cost of conditioning M by forming the product $M_0^{-1}M\psi$ is prohibitive unless some trick is used. [It would require $O(N^2)$ operations in practice where N is the dimension of M.] One such trick is the fast Fourier transform (FFT). If M_0 is diagonal in momentum space, as in the previous section, the multiplication requires only $O(N \ln N)$ operations, roughly the number required to form $M\psi$. Restricting M_0 in this way is fine provided M^{-1} is approximately diagonal in momentum space. Then M_0 can be chosen so that $M_0^{-1}M$ is approximately unity.

In a QCD simulation the gluon fields A_{μ} that appear in the Dirac operator M are just random configurations drawn from a distribution with probability density proportional to $e^{-S[A_{\mu}]}$ where $S[A_{\mu}]$ is the QCD action. To see what effect these fields have on M we examine its inverse in perturbation theory:

$$M^{-1} = \frac{1}{i\partial + gA + m} \rightarrow -\frac{V\delta_{p,q}}{\not p - m} - \frac{1}{\not p - m}gA(p-q)\frac{1}{\not q - m} - \frac{1}{\not p - m}\left[\frac{1}{V}\sum_{k}gA(p-k)\frac{1}{\not k - m}gA(q-k)\right]\frac{1}{\not q - m} + \cdots, \qquad (18)$$

where the system is in a box of volume V. So how diagonal is $M^{-1}(p,q)$? The first term in (18) is just the free propagator and is obviously diagonal. The next term is the first that can be nondiagonal. However, $A_{\mu}(p-q)$ is a random field that vanishes on average, at least in a reasonable gauge such as the Landau gauge $(\partial A = 0)$. So the leading order off-diagonal terms arise only from fluctuations in A_{μ} and these are of order the square root of the gluon propagator:

$$\langle |A_{\mu}(p-q)|^{2} \rangle = \frac{\int DA_{\mu}e^{-S} |A|^{2}}{\int DA_{\mu}e^{-S}}$$
$$\approx \frac{3V\delta_{p,q}}{(p-q)^{2}} \quad \text{(Landau gauge)}. \quad (19)$$

Consequently large fluctuations in A_{μ} that carry large momentum (p-q) are suppressed.

The ratio

$$\frac{\sum_{q \neq p} |M^{-1}(p,q)|^2}{|M^{-1}(p,p)|^2}$$
(20)

is a measure of how diagonal M^{-1} is. Keeping just the leading two terms in (18) one finds that on average this ratio is of order

$$\frac{1}{V^{2}} \sum_{q} \left\langle g A_{\mu}(p-q) \frac{1}{|\not q-m|^{2}} g A^{\mu^{*}}(p-q) \right\rangle$$

$$\approx \frac{3C_{f}}{V} \sum_{q} g^{2} \frac{1}{(p-q)^{2}} \frac{1}{|\not q-m|^{2}}$$

$$\approx \frac{3C_{f} \alpha_{s}(p_{\max})}{2\pi} \ln(p_{\max}/p) , \quad (21)$$

where

$$\alpha_{s}(p_{\max}) = \frac{g^{2}}{4\pi} = \frac{4\pi}{\beta_{0} \ln(p_{\max}^{2}/\Lambda^{2})}$$
(22)

is the QCD running coupling constant with $\beta_0 = 11 - 2n_{\text{flavor}}/3$, $\Lambda =$ the QCD scale constant, and $C_f = \frac{4}{3}$ for QCD. Provided $p_{\text{max}} \approx 1/a$ is large enough this ratio is certainly less than one for all p of order Λ or larger.²

This argument indicates that M^{-1} should be roughly diagonal in momentum space for all momenta down to some fixed physical scale of order Λ (\approx a few hundred MeV). Consequently, all critical slowing down associated with momenta ranging from Λ to p_{max} can be removed using Fourier acceleration with an M_0^{-1} that is diagonal in momentum space. So for fixed quark mass the number of accelerated iterations required to solve Eq. (1) is roughly constant as the lattice spacing $a \rightarrow 0$. However, momenta between Λ and m_q cannot be accelerated in this fashion and one still expects critical slowing down as $m_a \rightarrow 0$.

It is also clear from (18) that the p dependence of M^{-1} might be quite different from that of just the first term. For example, on average the $O(g^2)$ term is just

$$\frac{V\delta_{p,q}}{\not p-m} \left[\frac{1}{V} \sum_{k} \frac{g^2}{k^2} \gamma_{\mu} \frac{g^{\mu\nu} - k^{\mu}k^{\nu}/k^2}{\not p-k-m} \gamma_{\mu} \right] \frac{V\delta_{p,q}}{\not p-m}, \quad (23)$$

which is just the usual quark self-energy correction. At the very least one expects renormalization of the quark mass from such terms. Consequently, $M_0^{-1}(p)$ should be designed so as to contain parameters such as the quark mass that can be tuned to optimize the acceleration.

C. Gauge fixing

The importance of gauge fixing is evident from the preceding analysis. Momentum is not a gauge-invariant quantity. With the gauge unrestricted one would inevitably hit crinkly gauges, e.g., $A_{\mu} \rightarrow A_{\mu} - \partial_{\mu}(e^{i10^{10}x^2})$, for which $A_{\mu}(p-q)$ in Eq. (18) could be arbitrarily large at high momentum. To avoid such problems one must always fix to a smooth gauge such as the Landau or Feynman gauges.

The use of covariant gauges is not capricious. Axial gauges such as the $A^0=0$ gauge are far simpler to implement on the lattice. However such gauges are almost useless for Fourier acceleration. As in Eq. (19) the fluctuation in $A_{\mu}(q)$ is given roughly by the perturbative propagator which in the $A^0=0$ gauge is

$$\langle A_{\mu}(q) A_{\nu}^{*}(q) \rangle$$

= $\frac{1}{q^{2}} \left[-g_{\mu\nu} + \frac{q^{\mu}\eta^{\nu} + \eta^{\mu}q^{\nu}}{q^{0}} - \frac{q^{\mu}q^{\nu}}{(q^{0})^{2}} \right],$ (24)

where $\eta \cdot p = p^0$. Taking $q^2 \rightarrow \infty$ while holding q^0 small we find that the last term in the propagator does not vanish

$$\langle |A_{\mu}(q)|^2 \rangle \rightarrow \frac{1}{\mathbf{q}^2} \left[\frac{-\mathbf{q}^2}{(q^0)^2} \right] \rightarrow \frac{-1}{(q^0)^2} .$$
 (25)

Consequently there can be fluctuations $A_{\mu}(q)$ in such a gauge that carry arbitrary amounts of three-momentum, making $M^{-1}(p,q)$ highly nondiagonal. The same infrared singularities that make such gauges unpleasant to use in continuum calculations destroy their utility for Fourier acceleration.

The problem with axial gauges is easily visualized in terms of their lattice implementation. Consider a lattice



FIG. 1. The links for which the gauge field vanishes (i.e., $U_{n\mu} = 1$) in a timelike axial gauge.

FIG. 2. The links for which the gauge field vanishes in a blocked axial gauge.

configuration that is already in a smooth gauge such as the Landau gauge. To fix to the $A^0=0$ gauge one would introduce color rotation matrices G(x) at each site x formed from a product of contiguous link variables $U_{\mu}(n)$ connecting x to the origin. This product would be comprised of links along the t axis down to the t = 0plane, and then links within that plane to the origin (see Fig. 1). Thus two gauge matrices G(x) and $G(x+\hat{\mu})$ with $\mu \neq 0$, adjacent in three-space and at the same time, would differ on average by a product of order L links for a lattice having L sites per side. As $L \rightarrow \infty$ such a gauge transformation would introduce large fluctuations as one moved transverse to the t axis; i.e., it would introduce large q fluctuations. Furthermore, the links in the t=0plane break translation symmetry so that $M^{-1}(p,q)$ is nondiagonal even after fluctuations are averaged out.

One might wonder whether some simple variation on axial gauge might exist that does not have this disagreeable property. In fact one can connect every point on a lattice to the origin in such a way that the paths to neighboring sites differ only by $O(\ln(L))$ links rather than O(L) links as in $A^0=0$ gauge. A blocked axial gauge that has this property is illustrated in Fig. 2. Such gauges may well be suitable substitutes for covariant gauges though probably only on fairly large lattices.

III. NUMERICAL EXPERIMENTS

To test Fourier acceleration we solved the lattice Dirac equation on an 8⁴ lattice using quenched gauge field configurations at $\beta = 5.8$, 6.2, and ∞ . The details of our lattice formulations of the Dirac equation and the gauge field dynamics are presented in Appendix A. Here suffice it to say that we used Wilson fermions in the Dirac equation and Langevin updating to generate gauge field dynamics (see Ref. 1).

A. A Fourier-accelerated conjugate-gradient algorithm

The details of the conjugate-gradient algorithm we prefer are presented in Appendix B. To Fourier accelerate one simply replaces the Dirac equation

$$M\psi = \eta \tag{26}$$

by

$$\widetilde{M}\psi = \widetilde{\eta}$$
, (27)

where

$$\widetilde{M}\psi = \frac{1}{M_0(p)}\widehat{F}M\psi$$
(28)

is computed by starting with ψ and applying the operators from right to left. The source

$$\tilde{\eta} = \frac{1}{M_0(p)} \hat{F} \eta \tag{29}$$

is computed only once. Consequently one Fourier transform is required for each multiplication of the sort $M\psi$ and two such multiplications are required for each iteration of the conjugate-gradient algorithm. The function $M_0(p)$ can be tabulated once and used repeatedly in all iterations. Using code highly optimized for the FPS-264 attached processor the additional overhead for Fourier acceleration on an 8^4 lattice was only of order 25-30 %. Some large part of this overhead is due to the FFT's and will grow logarithmically with lattice size.

As discussed earlier it is critical that the gauge fields in the Dirac operator M be in a smooth gauge such as Landau gauge. Having chosen the gauge the only remaining ingredient for the algorithm is the conditioning matrix $M_0(p)$. There is little point in giving $M_0(p)$ spinor structure since the conjugate-gradient algorithm always uses \tilde{M} in the combination $\tilde{M}^{\dagger}\tilde{M}$ and the square $M_0^{\dagger}M_0$ of the free propagator is just a unit matrix in spinor space. Thus $M_0(p)$ can be thought of simply as a function of p. One might adopt one of two obvious strategies in choosing $M_0(p)$. On the one hand, $M_0(p)$ can be chosen to be the square root of the square of the free Dirac propagator (Appendix A) but with some effective hopping parameter κ_{eff} tuned to optimize the algorithm. This has the advantage of simplicity, but the disadvantage of requiring tuning. An alternative is to actually measure a small number of the diagonal elements of $M^{T}M$ in momentum space, take the inverse and square root of these elements, and then choose $M_0(p)$ so that it interpolates between these values.³ Such a procedure is attractive because it is self-tuning.

The analysis in Sec. II indicates that Fourier acceleration should be effective for momenta ranging from Λ (or m_q if it is large) to $p_{\rm max} \approx 1/a$. Thus one expects the accelerated conjugate-gradient algorithm to be faster than the unaccelerated algorithm by a factor of order p_{max}/Λ . As the lattice spacing a is decreased, while keeping the physical volume of the lattice fixed, the advantage should grow in proportion to L, the number of sites on a side of the lattice. On very small lattices this advantage is reduced as low-momentum modes $(\geq \Lambda)$ are excluded by the small volume of the lattice.

B. Data

In Fig. 3 we plot the number of iterations required by the conjugate-gradient algorithm, with and without acceleration, to achieve an accuracy of

$$\frac{\|\delta\psi\|}{\|\psi\|} \leq 0.005$$

for QCD coupling $\beta = 5.8$ and for a range of hopping parameters κ . Also plotted is the ratio of the number of iterations without acceleration to that needed with ac-



37



1592



FIG. 3. The number of iterations of the conjugate-gradient algorithm needed to obtain the quark propagator to a relative accuracy of 0.005 at $\beta = 5.8$. This is shown for the algorithm both with and without Fourier acceleration and for various values of the hopping parameter κ . The factor by which the number of iterations is reduced through Fourier acceleration (i.e., the "advantage") is also shown. The quark mass vanishes roughly at $\kappa = 0.167$, the last point on the plot.

celeration. The numbers shown represent averages over 16 independent gauge configurations, all in Landau gauge. Similar plots are given for $\beta = 6.2$ and for the free theory, $\beta = \infty$, in Figs. 4 and 5, respectively. Quenched QCD is deconfined at $\beta = 6.2$ for 8^4 lattices and this might well affect the Fourier acceleration. The error $\|\delta\psi\|$ was estimated in a variety of ways, all more or less equivalent variants of Eq. (13). Several things are obvious from these plots. First the gain due to Fourier acceleration is quite significant for each case, including that with the relatively low $\beta = 5.8$. Second as expected the advantage grows with increasing β and may be significantly larger at the somewhat larger β 's. Furthermore this advantage grows as the quark mass is decreased (i.e., κ is increased) saturating at the point where the quark mass becomes smaller than the QCD scale $(\approx \Lambda)$. So simulations at very near zero quark mass (i.e., $\kappa \approx \kappa_c$) are still quite expensive, though apparently not impossibly so when Fourier acceleration is used. Finally note that the conjugate gradient, both with and without acceleration, seems to work better as β is increased. This bodes well for future simulations, but also indicates that care must be taken, in comparing matrix inversion algorithms, to work at similar values of β .

In Figs. 6 and 7 we plot the residue $||M\psi^{(n)} - \eta||$, the error $||\delta\psi|| / ||\psi||$, and the magnitude of the solution $||\psi^{(n)}||$ as a function of the number of conjugate-gradient iterations. This is done for $\kappa = 0.16$, 0.167 ($\approx \kappa_c$) and $\beta = 5.8$. Nonuniform convergence in early iterations



FIG. 4. The same curves as in Fig. 3 but for $\beta = 6.2$. Note that QCD is deconfined at this β on lattices of the size used here.

seems to be typical when the quark mass is very nearly zero. These plots also illustrate the dangers inherent in using, say, the residue or $\|\psi^{(n)} - \psi^{(n-1)}\|$ as a measure of the error. On the other hand, the estimate (17) as used in these figures gives a useful estimate of the true error, particularly towards the end of the simulation.

In discussing the additional overhead incurred by Fourier acceleration we did not include the cost of gauge fixing the gluon fields. The Landau gauge used in these simulations is obtained by a potentially costly relaxation method. However, this relaxation scheme is readily accelerated using Fourier methods similar to those used here (see Ref. 4). Consequently, relatively few iterations are required to obtain a reasonable gauge fixing, particu-



FIG. 5. The number of iterations of the conjugate-gradient algorithm needed to compute the free quark propagator to a relative accuracy of 0.005. Only one iteration is needed to compute the propagator with the accelerated algorithm, and consequently this plot also shows the advantage gained from acceleration. The quark mass vanishes at $\kappa = 0.125$.



FIG. 6. The residue $||M\psi^{(n)} - \eta||$, the error $||\delta\psi|| / ||\psi||$, and the magnitude of the solution $||\psi^{(n)}||$ as a function of the number of conjugate-gradient iterations used in computing solution ψ of the Dirac equation. The curves here are for $\kappa = 0.16$ and $\beta = 5.8$.



FIG. 7. The same curves as in Fig. 6 but for $\kappa = 0.167$, which is approximately equal to κ_c .



FIG. 8. The number of conjugate-gradient iterations needed for a relative accuracy of 0.005 as a function of the number of (accelerated) iterations of Landau gauge fixing used in preprocessing the gauge field. The data here are for $\kappa = 0.16$ and $\beta = 5.8$.

larly when the configuration is first smoothed by an $A^{0}=0$ or blocked-axial gauge transformation (as was done here). This is illustrated in Fig. 8 where we plot the number of iterations required by the accelerated conjugate-gradient algorithm to achieve an accuracy of 0.005 as a function of the number of iterations of the Landau gauge-fixing algorithm. Gauge configurations were initially in the blocked axial gauge. Landau gauge is obtained by maximizing the average value of Tr[$U_{\mu}(n)$] using gauge transformations, where $U_{\mu}(n)$ is the OCD link variable. The effectiveness of Fourier acceleration seems simply related to the average trace of a link, as is illustrated in Fig. 9. Although we have not done detailed timing it is obvious that the ten or so sweeps of Landau gauge fixing required should not cost more than a single Langevin update of the gauge field. This is a negligible overhead. The blocked axial gauge should become more effective on larger lattices, further reducing the overhead.

A viable alternative to Landau gauge fixing is stochastic gauge fixing, where the Langevin force in the gauge field updating algorithm is augmented by a gauge-fixing force (see Ref. 4). Stochastic gauge fixing seems as effective as Landau gauge fixing in smoothing the gauge



FIG. 9. The number of conjugate-gradient iterations needed for a relative accuracy of 0.005 as a function of the average value of $Tr(U_{n\mu})$. The data here are for $\kappa = 0.16$ and $\beta = 5.8$.



FIG. 10. The number of conjugate-gradient iterations needed for a relative accuracy of 0.005 as a function of the effective hopping parameter κ_{eff} used in preconditioning the Dirac operator for Fourier acceleration.

fields in the Dirac operator, and again costs little. The only drawback to stochastic gauge fixing is that it may be incompatible with Fourier acceleration of the Langevin algorithm.

The need to tune the conditioning function $M_0(p)^{-1}$ is another potentially costly overhead. Almost all of the data shown in Figs. 3-7 for the accelerated algorithm used an $M_0(p)$ interpolating between measured values of M(p,p), as discussed in the previous section. However, almost identical performance is obtained simply by using a free propagator for $M_0(p)$ with an effective hopping parameter $\kappa_{\text{eff}}(<\frac{1}{8})$. Furthermore the peak performance is obtained for any value of κ_{eff} in a rather large interval. This is illustrated in Fig. 10 where the number of conjugate-gradient iterations required for an accuracy of 0.005 is plotted versus the $\kappa_{\rm eff}$ used in M_0 . When $\kappa_{\rm eff}$ is too close to the critical value of $\frac{1}{8}$ the evolution of the long-wavelength modes is destabilized and the algorithm degrades. When κ_{eff} is very small $M_0(p)$ becomes roughly p-independent and acceleration no longer occurs. However, any value in the range $\kappa_{\rm eff} \approx 0.07 - 0.10$ gives excellent results. Furthermore the optimal κ_{eff} is roughly independent of κ , at least for interesting κ 's. Consequently fine-tuning of $\kappa_{\rm eff}$ seems unnecessary.

IV. CONCLUSIONS

In this paper we have presented a new algorithm for inverting the Dirac operator in QCD. Even at relatively low values of β (≈ 6 in quenched QCD) our Fourier accelerated conjugate-gradient algorithm is 3-4 times faster than ordinary conjugate gradient for small quark masses. This advantage should grow as β is increased to more realistic values, roughly in proportion to the QCD correlation length measured in units of the lattice spacing. As matrix inversion is by far the most costly part algorithms such as this will have a large impact on realistic simulations of QCD.

The utility of Fourier acceleration is not restricted to solving the Dirac equation in QCD. The basic acceleration technique presented here can be applied to any linear system of equations for which the eigenmodes can be approximated by Fourier modes. Nor must one restrict oneself to the conjugate-gradient algorithm. Any relaxation method that updates the entire solution vector simultaneously can be Fourier accelerated. We have used conjugate-gradient methods throughout our analysis since in our experience no other common algorithm outperforms the conjugate gradient when quark masses are small, and most are decidedly inferior.

Further development of Fourier accelerated matrix inversion might focus on generalizing the form of M_0^{-1} , perhaps allowing some sort of off-diagonal contributions. Also one might try other gauges—e.g., the covariant gauge for which fluctuations in the off-diagonal elements vanish on average in leading order [see Eq. (21)]. Another possibility is combining Fourier acceleration with other techniques that have been proposed. For example, a method is given in Ref. 5 for partially decomposing the Dirac operator into triangular matrices. Although quite natural in the strong-coupling limit, this decomposition does not much improve the condition number of the operator in the continuum limit and so Fourier acceleration is still needed.

A complementary approach for reducing critical slowing down is the multigrid technique, where relaxation is used in turn on the original problem and on successively blocked versions (as in a renormalization-group transformation) of the problem. Such an approach might allow one to capitalize on renormalization-group studies of quark operators when and as they develop. However, multigrid methods might only become effective on rather large lattices, where many levels of blocking are possible. Fourier acceleration is itself quite similar in spirit to multigrid, using Fourier modes in place of block modes. Insofar as the Fourier modes very nearly are the eigenmodes of the Dirac operator, at least for large momenta, Fourier acceleration may be the natural choice for this particular problem.

ACKNOWLEDGMENTS

This research was supported by the NSF. Numerical work was performed on the NSF supercomputer facility at Cornell University. G.P.L. was partially supported by the Alfred P. Sloan Foundation. B.S. thanks the IBM Corporation for financial support.

APPENDIX A: THE DIRAC OPERATOR AND LATTICE QCD

In our numerical studies we used the Euclidean Dirac operator for Wilson quarks:

$$(M\psi)_{n} = \psi(n) - \kappa \sum_{\mu=1}^{4} \left[(1 - i\gamma_{\mu})U_{\mu}(n)\psi(n+\hat{\mu}) + (1 + i\gamma_{\mu})U_{\mu}^{\dagger}(n-\hat{\mu})\psi(n-\hat{\mu}) \right].$$
(A1)

Here κ is the hopping parameter, γ_{μ} is an anti-Hermitian Euclidean Dirac matrix, and $U_{\mu}(n)$ is the gauge-field operator at site *n* pointing in direction μ . The hopping parameter determines the quark mass, a reasonable definition for the bare mass (in units of a) being

$$m_q \approx \frac{1}{2\kappa} - \frac{1}{2\kappa_c}$$
, (A2)

where zero mass occurs at the critical point $\kappa = \kappa_c$ ($=\frac{1}{8}$ for free quarks).

The link operators $U_{\mu}(n)$ in Eq. (A1) were generated in the quenched approximation using the standard Langevin algorithm.¹ The action was the usual plaquette action:

$$S = -\frac{\beta}{2n_c} \sum_{p} \operatorname{Tr}(U_p + U_p^{\dagger})$$
(A3)

with $n_c = 3$ and U_p is the product of link operators around the perimeter of plaquette p. The Langevin step size was $\epsilon \approx 0.01$ and configurations were drawn from among those generated in several thousand Langevin iterations. Skewed-periodic boundary conditions were used for both gauge fields and fermion fields:

$$(x + L_x, y, z, t) = (x, y + 1, z, t),$$

$$(x, y + L_y, z, t) = (x, y, z + 1, t),$$

$$(x, y, z + L_z, t) = (x, y, z, t + 1),$$

$$(x, y, z, t + L_t) = (x, y, z, t),$$

(A4)

where L_{μ} is the number of lattice sites in direction μ .

Most of our simulations were at $\beta = 5.8$ and 6.2. At $\beta = 5.8$ we performed a complete analysis of the π and ρ masses as functions of κ . The π mass vanished at $\kappa = \kappa_c = 0.167(2)$. Our results agreed well with previous work at this β . The theory is deconfined at $\beta = 6.2$ on 8^4 lattices.

APPENDIX B: THE CONJUGATE-GRADIENT ALGORITHM

The Dirac equation

$$M\psi = \eta \tag{B1}$$

- *Present address: Naval Research Laboratory, Code 6550, Washington, D.C. 20375-5000.
- [†]Present address: Physics Department, Boston University, Boston, MA 02215.
- [‡]Present address: Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, U.K.
- §Present address: Deutsches Electronen-Synchrotron, DESY, Hamburg, Germany.
- **Present address: Department of Physics, University of California at San Diego, La Jolla, CA 92093.
- ^{††}Present address: Center for Theoretical Physics, Laboratory of Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139.
- ¹G. Batrouni, G. Katz, A. Kronfeld, P. Lepage, B. Svetitsky, and K. Wilson, Phys. Rev. D **32**, 2736 (1985); C. Davies, G. Batrouni, G. Katz, A. Kronfeld, P. Lepage, P. Rossi, B.

cannot be solved directly with the conjugate-gradient algorithm since M is not a positive matrix. This can be remedied by rewriting the Dirac equation in the form

$$(\boldsymbol{M}^{\dagger}\boldsymbol{M})\boldsymbol{\psi} = (\boldsymbol{M}^{\dagger}\boldsymbol{\eta}) \tag{B2}$$

for which the ordinary conjugate-gradient algorithm is applicable. A slight variant of this approach achieves the same result but is more stable against round-off errors.⁶ The basic iteration is given by

$$\psi^{(n)} = \psi^{(n-1)} + ap^{(n-1)} ,$$

$$r^{(n)} = r^{(n-1)} - aMp^{(n-1)} ,$$
(B3)

$$p^{(n)} = M^{\dagger} r^{(n)} + \frac{\|M^{\dagger} r^{(n)}\|^2}{\|M^{\dagger} r^{(n-1)}\|^2} p^{(n-1)}$$

where $\psi^{(n)}$, $r^{(n)}$, and $p^{(n)}$ are vectors, and

$$a = \frac{\|M^{\dagger}r^{(n-1)}\|^2}{\|Mp^{(n-1)}\|^2}$$
(B4)

is a scalar. Initially one sets

$$r^{(0)} = \eta - M \psi^{(0)} ,$$

$$p^{(0)} = M^{\dagger} r^{(0)} ,$$
(B5)

where $\psi^{(0)}$ is chosen as one wishes. Note that it is convenient when Fourier accelerating to maintain r in momentum space while ψ and ρ are maintained in coordinate space.

Several features of this algorithm should be noted: each iteration requires two multiplications involving Mand a vector; a minimum of four vectors must be stored at any given point; the residue $\|\eta - M\psi\|$ decreases monotonically with each iteration; the algorithm is gauge covariant when applied directly to the Dirac equation. Consequently, changing the gauge has no effect on performance in the absence of Fourier acceleration.

Svetitsky, and K. Wilson, J. Stat. Phys. 43, 1073 (1986). For more details, see G.R. Katz, Cornell Ph.D. thesis, 1986.

- ²Strictly speaking one should resum perturbation theory to all orders in $\alpha_s(p_{\text{max}})\ln(p_{\text{max}}/p)$. However our basic conclusions are unchanged by such a resummation.
- ³A convenient procedure for interpolating is to treat M_0 as a function of the free propagator, i.e., $M_0 = f[M_{\text{free}}^{\dagger}M_{\text{free}}(p)]$. Then M_0 is a relatively smooth function of only a single variable, rather than four variables, and the interpolation is simple.
- ⁴Here we use the "AL gauge" introduced in C. Davies *et al.*, preceding paper, Phys. Rev. D 37, 1581 (1988); see also P. Rossi *et al.*, Nucl. Phys. (to be published).
- ⁵Y. Oyanagi, Comput. Phys. Commun. 42, 333 (1986).
- ⁶M.R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).