1. INTRODUCTION

The overlap operator $\mathbf{1}$ is the closest known lattice Dirac operator to the continuum operator. With the lattice community currently moving beyond the quenched approximation, one obvious possibility is to use dynamical overlap fermions. The advantages of the overlap operator are well known: it satisfies the Ginsparg-Wilson lattice chiral symmetry exactly; there is an easy non-perturbative renormalisation; there is no operator mixing involving different chiral sectors; there is a well defined index $\langle Q_f \rangle = \frac{1}{2} \text{Tr}(Q)$, equal to the topological charge in the continuum limit; the anomaly is correctly accounted for; and it is essential for studies of topics such as topology, spontaneous chiral symmetry breaking, and the eigenvalue spectrum of the Dirac operator. Since chiral symmetry is so important to low energy QCD, it seems a waste not to use the only Dirac operator which we know of which fully respects this symmetry.

Of course, there are reasons not to use the overlap operator. Firstly, it is considerably slower than (for example) staggered or clover fermions. Secondly, the discontinuity in the overlap operator creates a number of unique problems when trying to implement a Hybrid Monte Carlo algorithm. The first of these issues will not be a problem once we have sufficiently fast computers, which will be in the very near future (we have already run some trajectories on a 16$^3$32 lattice). Now is the perfect time to tackle the second problem, and to create a Hybrid Monte Carlo algorithm for overlap fermions.

In this talk, I shall summarise the work done so far on this issue. Work has been published in this area by Z. Fodor et al, by myself in collaboration with Thomas Lippert and Stefan Krieg, and by T. DeGrand and S. Schaefer.

Section 2 provides a brief introduction to Hybrid Monte Carlo and the overlap operator. Section 3 discusses the problem of topological charge changes. Section 4 mentions some additional problems and advantages concerned with dynamical overlap fermions. Section 5 gives a few numerical results, and our conclusions are presented in section 6.

2. HYBRID MONTE CARLO WITH THE OVERLAP OPERATOR

The overlap Dirac operator is

$$D = (1 + \mu) + \gamma_5 (1 - \mu) \epsilon(Q),$$

where $\mu$ is a mass parameter, $Q$ is the hermitian Wilson Dirac operator, and $\epsilon$ is the matrix sign function. In our numerical simulations, we use a Zolotarev rational approximation to the sign function with the small eigenvalues treated exactly using eigenvalue projection, but for the purposes of this talk, I shall assume that we can calculate all the eigenvalues $\lambda_i$ and eigenvectors $\langle \psi_i |$ of $Q$ and thus treat the matrix sign function exactly using $\epsilon(Q) = \text{sign}(\lambda_i) \langle \psi_i |$. This will simplify the algebra while retaining the important features of the algorithm. We will define $H$ as the
Hermitian overlap operator $\gamma_5 D$.

The Hybrid Monte-Carlo (HMC) algorithm \cite{12} updates the gauge field in two steps: (1) a molecular dynamics (MD) evolution of the gauge field; (2) a Metropolis step which renders the algorithm exact. In the MD step, we introduce a momentum, $\Pi$, which is conjugate to the gauge fields $U$, and a spinor field $\phi$ which is used to estimate the fermion determinant via a heat-bath. We define an Hybrid Monte-Carlo energy

$$E = \frac{1}{2} \Pi^2 + S_g[U] + \phi^\dagger (H)^{-2} \phi.$$  

(2)

$S_g$ is the gauge action, and we will use either the Luscher-Weisz or Wilson plaquette action. We introduce a computer time $\tau$ and integrate over the classical equations of motion to generate the correct ensemble. We cannot perform an exact integration, so we need to use a numerical method such as the Omelyan integration step \cite{13}. This will create a small error in the energy conservation, which we can correct for by including an additional metropolis step, accepting or rejecting the new configuration according to a probability $P_{\text{acc}} = \min(1, \exp(\Delta))$, where $\Delta = E_i - E_f$, $E_i$ is the initial energy and $E_f$ the energy at the end of the MD. It is therefore important that we conserve energy as well as possible during the MD to ensure a high acceptance rate. Note that we do not have to use the classical trajectory: any reversible update which leads to a small $\Delta$ will suffice. Our MD procedure does not even have to conserve area, as long as we can easily calculate the Jacobian $J$. If we have a non-area conserving update, we just have to include the Jacobian in the Metropolis step, using a new $\Delta = E_i - E_f + \log J$ \cite{14}. To have a high acceptance rate, we need $e^\Delta \sim 1$.

The crucial part of the MD procedure is the force used to update the momentum, defined as

$$F_T = -U \partial (S_g[U] + \phi^\dagger (D^\dagger D)^{-1} \phi) \partial U.$$  

(3)

The fermionic part of this force for the overlap operator is

$$F_F \Pi + \Pi F_F^\dagger = -(1 - \mu^2) \phi \frac{1}{H^2} \left( \gamma_5 \frac{d}{d\tau} \epsilon(Q) + \frac{d}{d\tau} \epsilon(Q) \gamma_5 \right) \frac{1}{H^2} \phi.$$  

(4)

We can differentiate the eigenvectors and eigenvalues of $Q$ using a procedure analogous to first order time independent perturbation theory in quantum mechanics \cite{5}. This gives

$$d \frac{d}{d\tau} \epsilon(Q) = \sum_{i,j \neq i} |\psi_i \rangle \langle \psi_i | \frac{dQ}{d\tau} |\psi_j \rangle \langle \psi_j | \frac{\text{sign}(\lambda_i) - \text{sign}(\lambda_j)}{\lambda_j - \lambda_i}$$

$$+ \sum_i |\psi_i \rangle \langle \psi_i | \frac{d}{d\tau} \text{sign}(\lambda_i).$$  

(5)

Note that only mixings between eigenvalues of different signs contribute to the fermionic force, and only mixings between the small eigenvalues are important. The main feature of the fermionic force is the Dirac $\delta$-function coming from the differential of the sign function. We shall discuss how to deal with this in the next section.

### 3. Eigenvalue Crossings

The $\delta$-function in the fermionic force, which occurs whenever one of the Wilson eigenvalues crosses zero (i.e. whenever there is a change in the index of the overlap operator), should, in an exact integration, introduce a discontinuity in the momentum, which will exactly cancel the discontinuity in the pseudo-fermion energy caused by the abrupt change in the matrix sign function. We can visualise this in a classical mechanics picture by picturing a potential wall of height $-2d$ surrounding each topological sector. We can easily calculate the height of the wall (which can be either positive or negative), either by integrating the fermionic force across the $\delta$-function, or by calculating the difference in the pseudo-fermion energy. Both these procedures give

$$d = -(1 - \mu^2) \langle \phi | \frac{1}{(H^\dagger)^2} \left\{ \gamma_5, \epsilon(\lambda^-) |\psi \rangle \langle \psi | \right\} \frac{1}{(H^-)^2} |\phi \rangle.$$  

(6)
\(H^+\) is the Hermitian overlap operator just after the crossing, \(H^-\) the operator just before the crossing. In a classical mechanics picture, the momentum would be updated in a direction parallel to \(\eta\), the normal to the topological sector wall, thus: \((\Pi^+, \eta)^2 = (\Pi^-, \eta)^2 + 4d\). If the momentum is too small (i.e. this procedure can lead to an imaginary momentum) we reflect of the potential wall. However, as remarked earlier, there is no reason why we have to stick to the classical mechanics picture. Below, we shall describe how a general updating procedure can be derived to account for the potential wall.

To calculate the Jacobian, we need to work with the coordinate and momentum vectors \(u\) and \(\pi\). \(\pi\) can be calculated easily from the momentum field \(\Pi_\mu(x) = T_i \pi_\mu(x)\), where \(\mu\) and \(x\) refer to the direction and lattice site, and \(T_i\) are the generators of the gauge group. The gauge coordinate \(u\) is defined so that an update of the coordinate and lattice site, and \(G\) later; (3) We return the gauge field to the original point \(u_c = u^+ - \tau_c \pi^+\). Here \(\tau_c\) is the computer time at which the eigenvalue is zero, i.e.

\[
\tau_c = \frac{(u_c - u, \eta)}{\pi^-, \eta}.
\]

\((\pi^-, \eta)\) refers to the scalar product of the two vectors. Differentiating \(\tau_c\) with respect to \(u\) and \(\pi\) gives

\[
\frac{\partial \tau_c}{\partial \pi_k} = \tau_c \frac{\partial u_k}{\partial \pi_k} = -\tau_c \frac{\eta_k}{\pi^-, \eta}.
\]

We are now in a position to write down the Jacobian. Using the gauge update above gives

\[
J = \begin{vmatrix}
\frac{\partial \pi^+}{\partial u_k} & \frac{\partial \pi^+}{\partial \pi_k} \\
\frac{\partial u^+}{\partial \pi_k} & \frac{\partial u^+}{\partial u_k}
\end{vmatrix}
\]

\[
\frac{\partial u^+}{\partial \pi_k} = \tau_c \delta_{ik} + \frac{\partial \tau_c}{\partial \pi_k} (\pi^- - \pi^+) - \tau_c \frac{\partial \pi^+}{\partial \pi_k}
\]

\[
\frac{\partial u^+}{\partial u_k} = \delta_{ik} + \frac{\partial \tau_c}{\partial u_k} (\pi^- - \pi^+) - \tau_c \frac{\partial \pi^+}{\partial u_k}
\]

We can calculate \(J\) using two quick determinant manipulations. We subtract \(\tau_c\) times the top row from the bottom row. We then subtract \(\tau_c\) times the right column from the left column. These manipulations kill the bottom left hand element. Two lines of algebra later, and we obtain

\[
J = \left(\frac{\partial \pi^+}{\partial \pi_k}\right)_{u_c} \frac{(\eta, \pi^-)}{(\eta, \pi^-)}.
\]

We consider the components of the momentum normal to \(\eta\) and perpendicular to \(\eta\) separately. For example, if we update the momentum normal to \(\eta\), then we use the momentum update above and insert this Jacobian into the condition \(e^\Delta = 1\) needed for a high acceptance rate. This immediately gives us a differential equation for \(G_\eta\):

\[
e^{-G_\eta/2+2d} \frac{1}{\pi^+} \left(\frac{\pi^- + 2}{2 \partial(\pi^-)} \frac{\partial G_\eta}{\partial \pi^-} \right) = 1.
\]

It is trivial to solve this equation to obtain the momentum update

\[
e^{-(\pi^-)^2/2} - e^{-(\pi^-)^2/2-2d} - A(|d|)(1 - e^{-2d}) = 0.
\]

We have written the constant of integration as \(A(1 - e^{-2d})\) to ensure reversibility. \(A\) should lie in the range \(0 \leq A \leq 1\). \(A = 0\) gives us the classical mechanics solution.

We cannot allow the final momentum to be complex. Therefore, if the initial momentum is in the range \(a < \exp(-(\pi^-)^2/2) < b\), where

\[
a = A(1 - e^{2d})
\]

\[
b = e^{2d}(1 - A) + A,
\]

\[
(\pi^-)^2 - A(1 - e^{-2d}) = 0
\]
we can use equation 13 to update the momentum (we call this case transmission). Following 2, we reflect the momentum of the potential wall if the momentum lies outside this range (i.e. we use \( \pi^+ = -\pi^- \), with some additional terms — outlined in 10 — to ensure \( O(\tau^2) \) energy conservation). We need to keep the transmission rate as high as possible in order to reduce the topological autocorrelation. We can calculate the transmission rate by assuming that the momentum is initially distributed according to \( \exp(-\pi^2/2) \). The probability of transmission is

\[
P_t = \int \frac{\sqrt{-2 \log(b)}}{\sqrt{-2 \log(a)}} e^{-\pi_n^2/2} d\pi_n^-
\]

and we can show that \( \partial P_t/\partial A > 0 \) for \( 0 > A > 1 \). Therefore \( A = 1 \) maximises \( P_t \).

We can also update the momentum in directions perpendicular to \( \eta \). The procedure is exactly the same as above, and there are a large number of possible solutions, depending on the number of dimensions and how we combine the gauge fields. The simplest is, perhaps, the two dimensional case. We write \( \pi_1 = r \cos \theta \) and \( \pi_2 = r \sin \theta \). We can now proceed to change \( \pi^1 \), using

\[
e^{-r^2/2} - e^{-(r')^2/2} = - A(|d|)(1 - e^{-2d}) = 0. \]

The good news is that this is (approximately) independent of the volume. However, it is proportional to the inverse square of the mass \( \mu \) (see table 11 for a rough numerical confirmation of this). The probability of transmission has an exponential dependence on \( d \). Therefore, at small masses we are going to have a low acceptance rate. This can partially be solved by introducing multiple pseudo-fermion fields \( 10, 15 \), but it still remains a serious issue that still needs to be resolved. We also expect problems changing topological sector at small lattice spacing.

Zoltan Fodor and his collaborators have recently developed a novel algorithm to avoid the necessity of changing topological sectors [4]. It is easy, by continually reflecting, to keep the simulation fixed within one topological sector. By starting in several different topological sectors, it is possible to calculate the expectation value of an observable with each topological sector. To calculate the total expectation value, one has to find the relative weighting of the various sectors. This can be done by measuring an observable (which is defined only on the topological sector wall) on either side of the wall. Although there are still some

\[\text{If we prefer, we can also change } \theta, \text{ using } \theta \rightarrow e^{2d} \theta.\]
doubts concerning the ergodicity of this method\(^2\), it does represent an interesting possibility to solve
the problem of the topological autocorrelation at small masses.

Another problem is that the fermionic force can be unstable if we have two small eigenvectors of opposite signs — the force is proportional to \(1/(\lambda_1 - \lambda_2)\) (see equation \(5\)). This can be reduced by using stout smearing \([13]\) or an improved kernel operator to reduce the number of small eigenvalues \([9]\), but these do not address the underlying problem. We shall discuss this issue further in a future publication.

One significant advantage of overlap fermions is that the chiral symmetry allows us to factorise the squared overlap operator into the two chiral sectors:

\[
2 + \gamma_5 \epsilon(Q) + \epsilon(Q) \gamma_5 = \\
\left[ \frac{1}{2} (1 + \gamma_5) + \alpha \frac{1}{2} (1 - \gamma_5) + \frac{1}{4} (1 + \gamma_5) \epsilon(Q) (1 + \gamma_5) \right] \times \\
\left[ \frac{1}{2} (1 - \gamma_5) + \alpha \frac{1}{2} (1 + \gamma_5) - \frac{1}{4} (1 - \gamma_5) \epsilon(Q) (1 - \gamma_5) \right] \frac{2}{\alpha}
\]

\(\alpha \neq 0\) is an arbitrary constant. Both of the factors are positive definite, so we can use this decomposition to run single flavour simulations. It is easy to show that these two operators have the same non-zero eigenvalue spectrum as the overlap operator \(H\) (up to an unimportant sign). Zero modes can be included for by introducing additional pseudo-fermion fields to generate the determinant of \(1 - \frac{1 - \mu}{1 + \mu} \epsilon(Q)\). This is only useful at large masses, because we have to use a polynomial or rational approximation to obtain the square root of the chiral projected overlap operator, but we have tested it at approximately the strange quark mass on small lattices (using single flavour simulations) and the results for the plaquette and topological susceptibility agree with the 2-flavour HMC. This method can be used to simulate a 2+1 (or 2+1+1 etc.) flavour theory with little additional effort.

An important question is how well this algorithm scales with the volume. It has been suggested that the algorithm scales as the square of the volume \(V^2\) \([9]\). The work needed to perform a correction step is is proportional to the volume, and the density of small modes of the Wilson operator is also proportional to the volume. If the number of crossings were proportional to the density of small eigenvalues, then this would lead to an \(O(V^2)\) algorithm. However, it is by no means certain that this is the case because small eigenvalues with opposite signs repel during the molecular dynamics. Our numerical experience is that although the number of crossings increases as we increase the volume, it scales considerably better than \(O(V)\). On small lattices we observed a \(V^{1.5}\) scaling for the entire HMC algorithm, although this needs to be checked on larger volumes. More work needs to be done on this area to fully answer this important question.

5. RESULTS

In figure 1 we compare the small eigenvalues of the squared overlap operators for quenched and dynamical ensembles on \(12^4\) lattices at (approximately) the same lattice spacing. The dynamical configurations were generated at a mass \(\mu = 0.1\), although both sets of eigenvalues were plotted with the massless overlap operator. It can be seen that fermion determinant significantly suppresses the small eigenvalues of the Dirac operator. This is, of course, to be expected, because small eigenvalues would reduce the determinant. However, it is significant because we know from the Banks-Casher relation that the small eigenvalues lead to the spontaneous breaking of chiral symmetry. However, although the configurations shown in figure 1 do not possess small eigenvalues, we have seen them during the MD.\(^3\) Figure 2 plots the eigenvalues of the overlap operator during one molecular dynamics trajectory. There

\(^2\)But see the discussion in section 3 of \([4]\).

\(^3\)We discontinued the run which we used to generate figure 1 shortly after the plot was made because the lattice spacing was too large. We do not yet have enough data on our current runs to generate an improved plot.
Table 1
The dependence of the potential wall on lattice size and mass $\mu$. The ensembles were generated using the Luscher-Weisz gauge action.

| lattice size | $\beta$ | $\mu$ | $-<d>$ | $-<d>\mu^2$ |
|--------------|--------|-----|--------|------------|
| $4^4$        | 7.5    | 0.2 | -2.21  | 0.084      |
| $12^4$       | 7.5    | 0.1 | -6.49  | 0.065      |
| $4^4$        | 7.5    | 0.05| -35.92 | 0.090      |

Figure 1. The smallest non-zero eigenvalues for the overlap operator for a $\mu = 0.1$ dynamical ensemble (left), and quenched ensemble (right) on $12^4$ lattices with lattice spacing $\sim 0.18$ fm.

were five topological charge changes during this trajectory. The trajectory started with a topological charge -1, and the eigenvalues of magnitude $\sim 0.21$ are the zero modes (calculated to about a 10% accuracy with a mass $\mu = 0.1$). The non-zero modes at either end of the plot are similar in magnitude to those in figure 11 and these will not be responsible for chiral symmetry breaking. There is a considerably smaller non-zero eigenvalue between the third and fourth topological charge changes. A possible interpretation of this figure is that we create an anti-instanton on the second topological charge change, and an instanton on the third. The small eigenvalue is then be generated by the mixing between the two zero modes. If this interpretation is correct, then this presents further evidence for an underlying topological cause for chiral symmetry breaking.

One area in which we expect dynamical overlap simulations to be particularly important is the measurement of the topological susceptibility, defined as $\chi_t = \langle Q_f^2 \rangle/V$, where $Q_f$ is the index of the overlap operator, and $V$ is the lattice volume, since overlap fermions are the only lattice fermions with a well defined index theorem. In figure 12 we show plot the average value of the squared topological charge $\langle Q_f^2 \rangle$ against quark mass. It proved impossible to calculate the lattice spacing to any accuracy on these small lattices. 

\footnote{Although I am referring to instantons and anti-instantons, recent quenched calculations suggest that topological vacuum is not in fact dominated by instantons, but by long range topological fluctuations. But this argument holds however the zero modes are generated: I only use the terms instantons and anti-instantons for the sake of simplicity.}
tices, so we have no data for $\chi_t$ itself. The lattice spacing will change with the quark mass, but not significantly. Although we are unable to calculate the volume, we do see the expected linear dependence of the topological susceptibility with the quark mass.

6. CONCLUSIONS

Dynamical overlap fermions offer an exiting prospect for lattice QCD at small masses. However, the HMC is considerably harder to implement because of the discontinuity in the Dirac operator. The largest problem, how to deal with the Dirac-delta function in the fermionic force, has been solved. We still have a number of smaller issues to resolve, such as the problem of the topological autocorrelation at small masses and due to mixings between small eigenvectors of opposite signs. However, progress is being made on these issues. Although overlap fermions are still rather slow, we hope to begin large-scale simulations in the near future. We are currently working on lattices with sizes up to $16^332$ and at masses of approximately one third of the strange quark mass, and larger lattices and smaller quark masses should be possible on the next generation of computers.

7. Acknowledgements

I would like to thank in particular my collaborators Th. Lippert and S. Krieg, and also A. Borici, T. DeGrand, G. Egri, Z. Fodor, T. Kennedy, S. Schaefer G. Schierholz, H. Steuben, T. Streuer and K. Szabo for many useful discussions. The computer simulations were carried out on the Wuppertal AliceNext cluster, and the Jülich Blue-Gene, and I would like to thank Norbert Eicker for his help administering these machines. NC was supported by the EU grant MC-EIF-CT-2003-501467 and this work was part of the EU integrated infrastructure initiative HADRONPHYSICS project under contract number RI3-CT-2004-506078 as well as the EU integrated infrastructure project I3HP “Computational Hadron Physics” contract No. RI3-CT-2004-506078.

REFERENCES

1. H. Neuberger, Phys. Lett. B417, (1998) 141 [hep-lat/9707022]
2. Z. Fodor, S. D. Katz and K. K. Szabo, JHEP 0408 (2004) 003 [arXiv:hep-lat/0311010].
3. Z. Fodor, S. D. Katz and K. K. Szabo, Nucl. Phys. Proc. Suppl. 140 (2005) 704
Figure 3. The average value of the squared topological charge $Q_f^2$ on $6^4$ lattices plotted against quark mass with the Wilson plaquette action at $\beta = 5.4$.

4. G. I. Egri, Z. Fodor, S. D. Katz and K. K. Szabo, arXiv:hep-lat/0510117.
5. N. Cundy, S. Krieg, G. Arnold, A. Frommer, T. Lippert and K. Schilling, arXiv:hep-lat/0502007.
6. N. Cundy, S. Krieg, A. Frommer, T. Lippert and K. Schilling, Nucl. Phys. Proc. Suppl. 140 (2005) 841 [arXiv:hep-lat/0409029].
7. N. Cundy, S. Krieg, and T. Lippert, PoS (LAT2005) 107, arXiv:hep-lat/0511044.
8. T. DeGrand and S. Schaefer, Phys. Rev. D 71 (2005) 034507 [arXiv:hep-lat/0412005].
9. T. DeGrand and S. Schaefer, Phys. Rev. D 72 (2005) 054503 [arXiv:hep-lat/0506021].
10. S. Schaefer and T. DeGrand, arXiv:hep-lat/0508025.
11. J. van den Eshof, A. Frommer, T. Lippert, K. Schilling and H. A. van der Vorst, Comput. Phys. Commun. 146 (2002) 203 [arXiv:hep-lat/0202025].
12. S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Phys. Lett. B 195 (1987) 216.
13. T. Takaishi and P. de Forcrand, arXiv:hep-lat/0505020.
14. A. Borici, personal communication.
15. M. Hasenbusch, Phys. Lett. B 519 (2001) 177 [arXiv:hep-lat/0107019].
16. C. Morningstar and M. J. Peardon, Phys. Rev. D 69 (2004) 054501 [arXiv:hep-lat/0311018].