arXiv:hep-lat/0303026 v1 28 Mar 2003

DESY 03-040



# Accelerating the Hybrid Monte Carlo algorithm

A. Ali Khan[a], T. Bakeyev[b], M. Göckeler[cd], R. Horsley[e], D. Pleiter[f], P. Rakow[dg], A. Schäfer[d], G. Schierholz[fh] and H. Stüben[i]
QCDSF Collaboration

[a]Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany

[b]Joint Institute for Nuclear Research, 141980 Dubna, Russia

[c]Institut für Theoretische Physik, Universität Leipzig, 04109 Leipzig, Germany

[d]Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany

[e]School of Physics, The University of Edinburgh, Edinburgh EH9 3JZ, UK

[f]John von Neumann-Institut für Computing NIC, 15738 Zeuthen, Germany

[g]Theoretical Physics Division, Department of Mathematical Sciences, University of Liverpool, Liverpool L69 3BX, UK

[h]Deutsches Elektronen-Synchrotron DESY, 22603 Hamburg, Germany

[i]Konrad-Zuse-Zentrum für Informationstechnik Berlin, 14195 Berlin, Germany

An algorithm for separating the high- and low-frequency molecular dynamics modes in Hybrid Monte Carlo (HMC) simulations of gauge theories with dynamical fermions is presented. The separation is based on splitting the pseudo-fermion action into two parts, as was initially proposed by Hasenbusch. We propose to introduce different evolution time-scales for each part. We test our proposal in realistic simulations of two-flavor $O(a)$ improved Wilson fermions. A speed-up of more than a factor of three compared to the standard HMC algorithm is observed in a typical run.

## 1. Introduction.

The numerical simulations of QCD with Wilson fermions using the HMC algorithm [1] provide a significant challenge as the quark masses become smaller. Any improvement to make the simulations faster will help to increase the overlap between the domain of chiral perturbation theory and lattice QCD, allowing for an extrapolation to physical quark masses [2].

In the standard implementation of HMC one introduces pseudo-fermion fields to take into account the contribution of the fermion determinant. As the quark mass becomes lighter, the force induced by pseudo-fermions produces increasingly large high-frequency fluctuations. One is therefore forced to decrease the step-size of the integration scheme to keep a constant acceptance rate. This problem is addressed in the literature as "ultra-violet slowing-down" [3].

A possible solution of this problem is the introduction of multiple time-scales for different parts of the action in performing the discrete integration of molecular dynamics (MD) equations of motion in the fictitious time. This approach was initially advocated in Ref. [4], where the authors proposed to introduce different time-scales for the Yang-Mills term and the pseudo-fermion action. Since the number of arithmetic operations required to evaluate the pure gauge force is much smaller than the one needed to evaluate the pseudo-fermion force, one can keep a smaller step-size for the pure gauge part of the action. However for light fermions, the highest frequency fluctuations belong mainly to the pseudo-fermion action, so the approach of Ref. [4] gives only a moderate improvement in that case.

In Ref. [5] it was suggested that a multiple time-scale scheme is efficient only if one can split

the action

$$S = S_{UV} + S_{IR} \tag{1}$$

in a way to satisfy the following two criteria simultaneously:

- the force term generated by $S_{UV}$ is cheap to compute compared to $S_{IR}$;
- the splitting (1) mainly captures the high-frequency modes of the system in $S_{UV}$ and the low-frequency modes in $S_{IR}$.

If these criteria are met, one can keep a relatively large step-size for the "infra-red" part of the action $S_{IR}$ (which generates the computationally more expensive force term) and relax the step-size for the "ultra-violet" part $S_{UV}$, while the quark mass is becoming smaller.

Ref. [5] proposed to use a low-order polynomial approximation for mimicking the high-frequency modes of the pseudo-fermion action. The algorithm was tested for the 2D Schwinger model with Wilson fermions, producing a substantial speed-up in comparison with the standard HMC implementation.

In the present work we are testing a different approach. In Ref. [6] it was proposed to split the pseudo-fermion action into two parts, partially separating the small and large eigenvalues of the Dirac matrix. This splitting reduces the condition number of the fermion matrix, allowing for a larger step-size. In Refs. [7–9] the method was developed and successfully applied to four-dimensional lattice QCD with two flavors of Wilson fermions. We propose to further improve this method by putting the two contributions of the pseudo-fermion action from Ref. [6] on different time-scales of the integration scheme.

Our proposal was already considered by Hasenbusch while preparing Ref. [6], but he found no advantage (see Ref. [9]). Since this statement referred to tests within the two-dimensional Schwinger model, we have decided to repeat the tests for lattice QCD to see if one can profit from the "multiple time-scales idea" there. We have found that for two-flavor $O(a)$ improved Wilson fermions the introduction of different time-scales for the splitting chosen as in Ref. [6], indeed gives some speed-up compared to the case where the time-scale is the same for both parts.

In all our runs we have used the educated initial guess (chronological inversion method) proposed in Ref. [10]. This method estimates the trial solution for the matrix inversion as a linear superposition of a sequence of solutions in the recent past while performing the integration along the MD trajectory. (It was always checked that the accuracy of inversion was sufficient to make the solution effectively exact and keep the algorithm reversible.) The smaller the step-size of the MD integration scheme, the more efficient the chronological inversion method is. Hence the new improvements, which increase the effective step-size, may seem less efficient because the chronological guess becomes worse. Therefore, the improvement factor which we gained from splitting the pseudo-fermion action and introducing multiple time-scales is less pronounced than it would be if we had not used the chronological inversion. (We did not switch off the chronological guess because our tests were part of a production run.) One can hope to gain relatively more from the method tested in this paper if one is not using the educated guess. Nevertheless, it would be naive to say that the chronological inversion makes things worse. As the quark mass decreases, the step-sizes for each time-scale inevitably decrease, so the chronological inversion method will realize its potential after all, making the increase of computer power requirements smoother.

The paper is organized as follows. In section 2 we recall lattice QCD with $O(a)$ improved [11], even-odd preconditioned [12,13] Wilson fermions. We also briefly describe the splitting of the pseudo-fermion action proposed in Ref. [6]. In section 3 we discuss the multiple time-scales integration scheme in the HMC algorithm and specify the splitting of the action (the choice of time-scales) for the models which we are going to simulate. In section 4 we give the details of our simulations and present the results. Conclusions follow in section 5.

## 2. The model

We test our proposal by simulating two-flavor QCD with $O(a)$ improved [11], even-odd preconditioned Wilson fermions. One of the possible effective actions for a standard HMC simulation of this theory is given in Refs. [12,13]:

$$S_0[U,\phi^\dagger,\phi] = S_G[U] + S_{det}[U] + \phi^\dagger (Q^\dagger Q)^{-1} \phi \ . \quad (2)$$

Here $S_G[U]$ is the usual Wilson plaquette action, $\phi^\dagger, \phi$ are pseudo-fermion fields,

$$S_{det}[U] = -2\mathrm{Tr}\log(1+T_{oo}) \ , \quad (3)$$

$$Q = (\mathbf{1}+T)_{ee} - M_{eo}(\mathbf{1}+T)^{-1}_{oo} M_{oe} \ , \quad (4)$$

$T_{ee}(T_{oo})$ is the clover matrix (diagonal in coordinate space) on the even (odd) sites

$$(T)_{a\alpha,b\beta}(x) = \frac{i}{2} c_{sw} \kappa \sigma^{\alpha\beta}_{\mu\nu} \mathcal{F}^{ab}_{\mu\nu}(x) \ , \quad (5)$$

and the off-diagonal parts $M_{eo}$ and $M_{oe}$, which connect the even with odd and odd with even sites, respectively, are the usual Wilson hopping matrices (see Ref. [12] for further details).

According to Ref. [6] we start to modify the action (2) by introducing other pseudo-fermion fields $\chi^\dagger, \chi$:

$$S_1[U,\phi^\dagger,\phi] = S_G[U] + S_{det}[U] + \phi^\dagger W (Q^\dagger Q)^{-1} W^\dagger \phi + \chi^\dagger (W^\dagger W)^{-1} \chi \ , \quad (6)$$

where $W$ is some auxiliary matrix. The idea of this modification is that $W$, as well as $QW^{-1}$, have smaller condition numbers than the original matrix $Q$. This reduces the fluctuations of the HMC Hamiltonian at the end of the MD trajectory, allowing for a larger step-size in the HMC simulation at the same acceptance rate.

We consider here only the following choice of the matrix $W$ [9]:

$$W = Q + \rho \ , \quad (7)$$

which depends on one real parameter $\rho$.[1] Up to the multiplication by a constant factor this matrix is equivalent to the original proposal of Ref. [6].

The modification of the pseudo-fermion action (6) can be easily implemented, if a standard HMC program is already available (see Refs. [6,9] for details).

[1] Another possibility, discussed in Refs. [7–9], is

$W = Q + i\gamma_5 \rho$ .

We did not consider this, although the generalization of our approach to this choice of the auxiliary matrix is straightforward.

## 3. Multiple time-scales

The introduction of multiple time-scales for different segments of the action in the HMC method was initially proposed by the authors of Ref. [4]. Following their idea, one constructs a reversible integrator $V_M(\tau)$ for the action (1) by

$$V_M(\tau) = V_{IR}\left(\frac{\tau}{2}\right) \cdot \left[V_{UV}\left(\frac{\tau}{2M}\right) V_Q\left(\frac{\tau}{M}\right) V_{UV}\left(\frac{\tau}{2M}\right)\right]^M \cdot V_{IR}\left(\frac{\tau}{2}\right) , \quad (8)$$

where $M$ is a positive integer, and the effect of $V_Q, V_{UV}, V_{IR}$ on the system coordinates $\{P,Q\}$ is given by

$$\begin{aligned} V_Q(\tau): &\quad Q \to Q + \tau P \ , \\ V_{UV}(\tau): &\quad P \to P - \tau \partial S_{UV} \ , \\ V_{IR}(\tau): &\quad P \to P - \tau \partial S_{IR} \ . \end{aligned} \quad (9)$$

This integrator effectively contains two evolution time-scales, $\tau$ and $\tau/M$. The choice of $M$ is a trade-off between the computational overhead from computing the force $\partial S_{UV}$ more frequently, and the gain from reducing the fluctuations of the HMC Hamiltonian at the end of the MD trajectory. In the case $M = 1$ one gets an ordinary leap-frog integrator.

For testing the efficiency of our approach we performed numerical simulations for three models. The first model is based on the action (2). The other two models differ from each other by a different splitting (1) of the action (6):

- **Model A**

$$\begin{aligned} S_{UV} &= S_G[U], \\ S_{IR} &= S_{det}[U] + \phi^\dagger (Q^\dagger Q)^{-1} \phi \end{aligned} \quad (10)$$

- **Model B**

$$S_{UV} = S_G[U],$$
$$S_{IR} = S_{det}[U] + \phi^\dagger W (Q^\dagger Q)^{-1} W^\dagger \phi + \chi^\dagger (W^\dagger W)^{-1} \chi \qquad (11)$$

- **Model C**

$$S_{UV} = S_G[U] + S_{det}[U] + \chi^\dagger (W^\dagger W)^{-1} \chi,$$
$$S_{IR} = \phi^\dagger W (Q^\dagger Q)^{-1} W^\dagger \phi \qquad (12)$$

Model A is just a standard HMC algorithm for which the original splitting of the time-scale proposed by Sexton and Weingarten [4] is applied. Model B is the modification proposed by Hasenbusch [6], which was numerically studied in Refs. [8,9]. Finally, model C is our proposal for introducing different time-scales for the two parts of the pseudo-fermion action (6).

The splitting (12) is motivated by the hypothesis that most of the high-frequency modes of the pseudo-fermion part of the action (6) are located in $\chi^\dagger (W^\dagger W)^{-1} \chi$. We also put the clover determinant $S_{det}[U]$ on the "ultraviolet" time-scale because the force generated by it is computationally cheap. The computationally expensive term $\phi^\dagger W (Q^\dagger Q)^{-1} W^\dagger \phi$ is put on the "infra-red" time-scale.

## 4. Simulation details and results

We tested the approach (12) in production runs on a $16^3 \times 32$ lattice at $\beta = 5.29$, $\kappa = 0.1355$, $c_{sw} = 1.9192$ done by the QCDSF-UKQCD collaboration [14],[15],[16]. These parameters correspond to $m_\pi / m_\rho \approx 0.7$. The program was executed on the APEmille [17] at NIC Zeuthen.

For the fermion fields we use periodic (antiperiodic) boundary conditions in the spatial (time) directions. A trajectory was composed of $N_{steps}$ consecutive steps (8), with the trajectory length equal to 1:

$$N_{steps} \cdot \tau = 1 \, . \qquad (13)$$

The linear equations appearing in the calculation of the fermionic force and in updating $\phi, \phi^\dagger$ we solve by the conjugate gradient algorithm. In all cases the starting vector for the iterative solution was the linear superposition of $N_{guess}$ solutions from the recent past [10]. In all our simulations we kept the value $N_{guess} = 7$, which was empirically found to be close to the optimum.

We performed one run for model A, two runs for model B, and a few runs for model C with different values for $\rho$ and $M$. All runs had a length of 300 trajectories, which allowed us to get a reasonable estimate of the acceptance rates $P_{acc}$. Our strategy was to try to keep the same acceptance rates for all runs by tuning the step-size $\tau$.

The main goal of this study was to compare the efficiencies of A, B, and C. These efficiencies are determined by the amount of CPU-time $t_{CPU}$ required for estimates of some observables with a given statistical error. Since the computer time in simulations with dynamical fermions is mostly spent in the calculation of the pseudo-fermion force, the CPU-cost is roughly proportional to

$$t_{CPU} \propto (N_Q + N_W) \cdot \tau_{int} \, . \qquad (14)$$

Here $N_Q$ and $N_W$ denote the average number of multiplications by the matrices $Q^\dagger Q$ and $W^\dagger W$, respectively, required for producing one MD trajectory, and $\tau_{int}$ is the integrated autocorrelation time for the observable under study. Unfortunately, our computer resources did not allow us to estimate $\tau_{int}$ reliably. Therfore we base our investigation on the hypothesis that for fixed acceptance rates, the autocorrelation times are the same for the different approaches and different parameter sets considered in this paper, i.e. for all runs

$$\tau_{int} \propto \frac{1}{P_{acc}} \, . \qquad (15)$$

This hypothesis was confirmed by simulations of model B on a $8^3 \times 24$ lattice in Ref. [9].

Therefore, we measured the relative gain in computer time with respect to the standard HMC algorithm (model A) for the approach tested in the $i$-th run by the following formula:

$$D^{(i)}_{gain} = \frac{N^{(A)}_Q}{N^{(i)}_Q + N^{(i)}_W} \cdot \frac{P^{(i)}_{acc}}{P^{(A)}_{acc}} \, . \qquad (16)$$

Table 1
Runs of 300 trajectories each on the $16^3 \times 32$ lattice at $\beta = 5.29, \kappa = 0.1355, c_{sw} = 1.9192$ for the models of Eqs. (10), (11), (12). Here $\rho$ is the parameter in the operator (7), $M$ defines the second time-scale of the integration scheme (8), $N_{steps} = 1/\tau$ is the number of steps of which the trajectory with length 1 was composed. $P_{acc}$ is the acceptance rate, $N_Q$ and $N_W$ denote the average number of multiplications per trajectory by the matrices $Q^\dagger Q$ and $W^\dagger W$, respectively. $D_{gain}$ denotes the speed-up factor with respect to the standard HMC algorithm (model A).

| model | $\rho$ | $M$ | $N_{steps}$ | $P_{acc}$ | $N_Q$ | $N_W$ | $N_Q + N_W$ | $D_{gain}$ |
|---|---|---|---|---|---|---|---|---|
| **A** | 0 | 3 | 140 | 0.601 | 139492 | 0 | 139492 | 1 |
| **B** | 0.5 | 3 | 100 | 0.599 | 65951 | 5233 | 71184 | 1.95 |
| **B** | 0.2 | 3 | 70 | 0.664 | 47214 | 7378 | 54592 | 2.82 |
| **C** | 0.5 | 3 | 50 | 0.547 | 45160 | 7687 | 52847 | 2.40 |
| **C** | 0.2 | 3 | 40 | 0.663 | 32659 | 12373 | 45032 | 3.42 |
| **C** | 0.1 | 3 | 30 | 0.603 | 24932 | 15938 | 40870 | 3.42 |
| **C** | 0.07 | 3 | 30 | 0.640 | 24512 | 20738 | 45250 | 3.28 |
| **C** | 0.1 | 5 | 30 | 0.733 | 24622 | 26235 | 50857 | 3.35 |

The larger the gain $D^{(i)}_{gain}$ for the $i$-th run, the less computer time is required for estimating the observables by using the approach tested in that run.

We present our results in Table 1. The following observations can be made:

- Putting the two contributions of the pseudo-fermion part of the action (6) on different time-scales of the integration scheme (model C) gives some additional gain in computer time compared to the case, where the time-scale is the same for both parts (model B). A speed-up of $\approx 20\%$ is observed for $\rho = 0.5$ and $\rho = 0.2$.

- In agreement with the studies in Refs. [8,9], we observe that for fixed $M$ the performance of the approach C is best for some optimal value $\rho$, which in our case is likely to lie in the interval $\rho \in [0.1, 0.2]$ for $M = 3$.

- In one of the runs we increased the value of $M$ from 3 to 5, while $\rho = 0.1$ was close to the optimal value. We kept the same step-size $\tau$ for both runs. One sees that this change of $M$ increased the acceptance rate $P_{acc}$, but the gain in computer time $D_{gain}$ stayed almost the same (or even slightly decreased) due to the computational overhead coming from calculating the pseudo-fermion force $\partial S_{UV}$ more frequently.

- By using the approach C one achieves a speed-up of more than a factor three compared to the standard HMC algorithm A.

Our computational resources did not allow for a further resolution of the algorithmic performance in the space of the parameters $\rho, M$. Probably, the best improvement factor which we obtained, $D_{gain} = 3.42$, can still be slightly increased by further tuning of the parameters. However, we notice that $D_{gain}$ seems to be quite stable for some range of the parameter $\rho$. Therefore, we expect that not much tuning of the algorithm will be required in the forthcoming production runs.

## 5. Conclusions

In Refs. [6–9] it was suggested to accelerate the HMC simulation of dynamical fermions by splitting the fermion matrix into two factors with smaller condition numbers than that of the original matrix, and introducing pseudo-fermion fields for each of the factors. Inspired by the proposal of Ref. [5], we tested the possibility to further speed up the simulations by putting each part of this new pseudo-fermion action on a separate time-scale. We have found that such a strategy gave

a speed-up of $\approx 20\%$ in comparison to the case, where the time-scale was the same for both parts.

In our simulations, which are a part of the production runs of the QCDSF collaboration, we have found a reduction of the numerical cost of more than a factor three compared to the standard HMC algorithm. One can speculate that this reduction would have been even higher if we had not used the educated chronological guess [10]. One can also expect that the gain in computer time will grow as the quark mass decreases.

Further work in the direction of algorithmic improvement can be done by testing more complicated integration schemes than that of Eq. (8). In Ref. [9] it was shown that the splitting of the pseudo-fermion action (6) provided more computational gain for the partially improved integration scheme suggested in Eq. (6.4) of Ref. [4] than for the standard leap-frog scheme. It may be interesting to check the compatibility of that integration scheme (and higher order integration schemes) with the multiple time-scales approach studied in our paper.

When going to smaller quark masses, a possibility might be to generalize the idea of Ref. [6] by splitting the pseudo-fermion action into three or more parts [9]. One can introduce different time-scales for each part of the pseudo-fermion action in such an approach, profiting even more from the separation of high- and low-frequency modes.

## ACKNOWLEDGEMENTS

The numerical calculations were performed on the APEmille at NIC (Zeuthen). We thank the operating staff for their support. TB acknowledges the hospitality of NIC (Zeuthen) and the Physics Department of Regensburg University. This work has been supported in part by the European Community's Human Potential Program under contract HPRN-CT-2000-00145, Hadrons/Lattice QCD. TB acknowledges support from INTAS grant 00-00111, Heisenberg-Landau program grant 2002-10, RFBR grants 02-01-00126 and 02-01-06064.

## REFERENCES

1. S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Phys. Lett. B **195** (1987) 216.
2. C. Bernard, S. Hashimoto, D. B. Leinweber, P. Lepage, E. Pallante, S. R. Sharpe and H. Wittig, arXiv:hep-lat/0209086.
3. A. Borici, arXiv:hep-lat/0208048.
4. J. C. Sexton and D. H. Weingarten, Nucl. Phys. B **380** (1992) 665.
5. M. Peardon and J. Sexton [TrinLat Collaboration], arXiv:hep-lat/0209037.
6. M. Hasenbusch, Phys. Lett. B **519** (2001) 177 [arXiv:hep-lat/0107019].
7. M. Hasenbusch and K. Jansen, Nucl. Phys. Proc. Suppl. **106** (2002) 1076 [arXiv:hep-lat/0110180].
8. M. Hasenbusch and K. Jansen, arXiv:hep-lat/0210036.
9. M. Hasenbusch and K. Jansen, arXiv:hep-lat/0211042.
10. R. C. Brower, T. Ivanenko, A. R. Levi and K. N. Orginos, Nucl. Phys. B **484** (1997) 353 [arXiv:hep-lat/9509012].
11. B. Sheikholeslami and R. Wohlert, Nucl. Phys. B **259** (1985) 572.
12. K. Jansen and C. Liu, Comput. Phys. Commun. **99** (1997) 221 [arXiv:hep-lat/9603008].
13. X. Q. Luo, Comput. Phys. Commun. **94** (1996) 119 [arXiv:hep-lat/9603021].
[14]
14. D. Pleiter [QCDSF Collaboration], Nucl. Phys. Proc. Suppl. **94** (2001) 265 [arXiv:hep-lat/0010063].
15. H. Stuben [QCDSF collaboration], Nucl. Phys. Proc. Suppl. **94** (2001) 273 [arXiv:hep-lat/0011045].
16. S. Booth *et al.* [QCDSF-UKQCD collaboration], Phys. Lett. B **519** (2001) 229 [arXiv:hep-lat/0103023].
17. A. Bartoloni *et al.* [APE-Collaboration], Nucl. Phys. Proc. Suppl. **106** (2002) 1043 [arXiv:hep-lat/0110153].