PROCEEDINGS OF SCIENCE

Computational Requirements of the Rational Hybrid Monte Carlo Algorithm

Norman Christ*
Department of Physics, Columbia University, New York, NY 10027, USA
E-mail: nhc@phys.columbia.edu

Chulwoo Jung
Brookhaven National Laboratory, Upton, NY 11973, USA
E-mail: chulwoo@bnl.gov

RBC and UKQCD collaborations

The rational hybrid Monte Carlo algorithm and Hasenbusch mass preconditioning provide dramatic reductions in the computer time required to generate an ensemble of lattice configurations which include the effects of dynamical quarks. In fact, the advantage over earlier methods grows rapidly as the mass of the lightest quarks decreases. We present theoretical estimates for the dependence of the computer requirements for simulations using these new methods on the quark mass, lattice spacing and volume. Comparisons are made with the performance found in 2+1 flavor domain wall fermion simulations using lattice spacings of 0.114 and 0.086 fm and quark masses as small as 1/5 of the strange quark mass.

The XXV International Symposium on Lattice Field Theory
July 30 - August 4 2007
Regensburg, Germany

*Speaker.

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike Licence.

PoS(LATTICE 2007)028

Recent algorithmic developments have dramatically reduced the computer resources needed for full QCD calculations which include dynamical light quarks. Specifically Hasenbusch mass preconditioning [1, 2] and the rational hybrid Monte Carlo (RHMC) method of Clark and Kennedy [3, 4] are now essential to the program of the RBC and UKQCD collaborations to generate large ensembles of lattice configurations using domain wall quarks [5, 6].

Combining these two methods with the multi-time-step method of Sexton and Weingarten [7] permits the large molecular dynamics forces associated with the numerous high-momentum modes (which require a small time step) to be separated from the computationally expensive but relative weak light quark forces, giving a significant reduction of the cost associated with the light quarks.

We first recall earlier estimates for the computational cost of the hybrid Monte Carlo (HMC) method. Next, we introduce a more detailed model for the light-quark mass dependence of the molecular dynamics force with which we predict the scaling behavior of the current Hasenbusch and RHMC method. These expectations are then compared with the RBC-UKQCD experience [8].

1. Previous Estimates of HMC Computational Cost

We begin by reviewing the standard estimates of the computational cost required for the HMC algorithm that were discussed, for example, at the Lattice 2001 conference in Berlin [9]. Such an estimate is provided by the following equation:

$$\mathrm{Cost}_{\mathrm{total}} = C_0 \cdot \underset{\text{(A)}}{L^4} \cdot \underset{\text{(B)}}{L} \cdot \underset{\text{(C)}}{\frac{1}{m_\pi}} \cdot \underset{\text{(D)}}{\frac{1}{m_l}} \cdot \underset{\text{(E)}}{\frac{1}{m_l}}. \tag{1.1}$$

Here the linear size of the system (L), the light quark mass (m_l) and the pion mass (m_π) are all expressed in lattice units. If the quantities in this expression are put in physical units, one finds:

$$\mathrm{Cost}_{\mathrm{total}} = C_1 \left(\frac{L}{\mathrm{fm}}\right)^5 \left(\frac{\mathrm{MeV}}{m_l}\right)^{2.5} \left(\frac{\mathrm{fm}}{a}\right)^8 \quad \text{where} \tag{1.2}$$

(A) Total number of lattice sites which appears multiplicatively when counting most of the operations needed to perform an HMC time step.

(B) Weak volume dependence of the step size expected for the leap-frog integrator.

(C) Critical slowing down of the algorithm, assumed to have a critical index of 1.

(D) Dependence of the size of the time step on the light quark mass. This appears because the fermion force is expected to grow as $1/m_l$ as $m_l \to 0$.

(E) Dependence of the number of conjugate gradient iterations on the light quark mass.

2. Fermionic force on the gauge fields

Experience of the RBC and UKQCD collaborations with the generation of domain wall fermion lattice configurations calls into question the factor (D) above. Typically we have found a much weaker dependence of the step size on the light quark mass. Thus, a closer look at the quark mass dependence expected in the fermion force may be warranted. We begin by writing a general expression for the fermion force as a sum over Dirac eigenvalues. The force acting on the momentum

PoS(LATTICE 2007)028

$P^a_\mu(x)$ conjugate to the link $U_\mu(x)$ might be written:

$$F^a_\mu(x) = \Big(t^a U_\mu(x)\Big)_{\alpha\beta} \frac{\partial}{\partial U_\mu(x)_{\alpha\beta}} \ln\Big(\det(D+m)\Big) \tag{2.1}$$

$$= \Big(t^a U_\mu(x)\Big)_{\alpha\beta} \frac{\partial}{\partial U_\mu(x)_{\alpha\beta}} \sum_n \ln(\lambda_n^2(U)+m^2) \tag{2.2}$$

$$= \Big(t^a U_\mu(x)\Big)_{\alpha\beta} \sum_n \frac{2\lambda_n}{\lambda_n^2+m^2} \frac{\partial \lambda_n}{\partial U_\mu(x)_{\alpha\beta}}, \tag{2.3}$$

where t^a is a color generator and a sum over the fundamental representation indexes α and β is understood. In Eqs. 2.2 and 2.3 we have introduced an explicit sum over the Dirac eigenvalues.

Here we have assumed we are sufficiently close to the continuum limit, that we can treat the eigenvalues and eigenvectors of the lattice Dirac operator as possessing continuum properties. Thus, as is suggested in Eq. 2.1, we have separated the Dirac operator into a chiral part, $D = \gamma^\mu D^\mu$ and the mass term m and we have denoted by λ_n and ψ_n the eigenvalues and eigenvectors of D:

$$D\psi_n^\pm = \pm i\lambda_n \psi_n^\pm. \tag{2.4}$$

For simplicity we are ignoring the possibility of zero-modes and recognizing that in the continuum limit the eigenvalues are organized into pairs, $\pm\lambda_n$, with $\psi_n^- = \gamma^5\psi_n^+$.

Next we can use standard, first-order perturbation theory to derive an expression for the derivatives of $\lambda_n(U)$ with respect to the gauge link in the μth direction located at x.

$$\frac{\partial \lambda_n(U)}{\partial U_{\alpha\beta}(x,\mu)} = \frac{1}{a}\overline{\psi}_n(x)_\alpha \gamma^\mu \psi_n(x+\mu)_\beta. \tag{2.5}$$

This expression can be used in Eq. 2.3 and derivatives with respect to $U^\dagger$ included to give:

$$F^a_\mu(x) = \sum_{n,\pm} \frac{2\lambda_n}{\lambda_n^2+m^2}\,\frac{1}{a}\Big\{\overline{\psi}_n^\pm(x)^\dagger \gamma^\mu t^a U_\mu(x)\psi_n^\pm(x+\hat{e}_\mu) + \overline{\psi}_n^\pm(x+\hat{e}_\mu)^\dagger \gamma^\mu U_\mu(x)^\dagger t^a \psi_n^\pm(x)\Big\}. \tag{2.6}$$

The factor of the inverse lattice spacing, $1/a$, appears here because we are expressing the eigenvectors ψ_n in lattice units but the mass m, the Dirac operator and its eigenvalues in physical units.

We will now attempt to use this equation to discuss the possible dependence of the force $F^a_\mu(x)$ on the quark mass m. Clearly $1/m$ behavior is possible from those terms in the sum with $\lambda_n \sim m$. We might attempt to distinguish two types of such small-eigenvalue modes. The first are extended states filling the space-time volume and of order $(a/L)^4$ at the point x where the force is being calculated. (The first-order perturbation theory formula used above requires that the wave functions ψ_n be normalized to unity over the L^4 space-time volume.)

While for large L the contribution of a single such mode is suppressed by $1/L^4$, the number of such modes will grow proportional to L^4. This is easily described by introducing the eigenvalue density per unit space-time volume, per unit energy: $\rho_{\rm ext}(\lambda)$. The total force from those extended modes with eigenvalues in the interval $[\lambda, \lambda+\Delta\lambda]$ will then be $a^3\rho_{\rm ext}(\lambda)\Delta\lambda\,\lambda/(m^2+\lambda^2)$. The factor a^3 appears because we choose to express $\rho_{\rm ext}(\lambda)$ in physical units. As we approach the

PoS(LATTICE 2007)028

continuum limit, this a^3 factor becomes very important, reducing the sensitivity of the force and hence the molecular dynamics step size to the light quark mass.

The second type of mode to consider is a localized mode with the largest contribution to $|\psi_n(x)|^2$ coming from x within a region of radius R of the mode's location, where we will also express R in physical units. This is the sort of fermion mode that would result from gauge configuration containing an instanton-like contribution of radius R. If the place where we are computing the force lies within R of the mode's location, the resulting force will be $\lambda/(m^2+\lambda^2)(a^3/R^4)$. However, the probability of finding this mode within a radius R of the point where we are computing the force is $\rho_{\rm loc}(\lambda)R^4\Delta\lambda$. Thus, the average contribution of such a mode to the force at a specific point x might be estimated as the product of these two quantities: $a^3\rho_{\rm loc}(\lambda)\Delta\lambda\lambda/(m^2+\lambda^2)$, the same result estimated above for extended modes.

However, the probability that a large force $\propto \lambda/(m^2+\lambda^2)(a^3/R^4)$ will appear somewhere in the space-time volume is $\rho_{\rm loc}\Delta\lambda L^4$. Since it is this largest force which determines the largest allowable step size, this latter estimate is likely the most relevant. This force becomes large only if R is sufficiently small to compensate for the a^3 factor in the numerator. The density of such modes, localized on the lattice scale, will depend on details of the gauge action and should not effect physical quantities. In the remainder of this paper, we will assume that these small, localized modes with R of order a are not important and consider only the extended modes.

Thus, we adopt the following simple model to describe the mass dependence of the molecular dynamics force at the site x:

$$F(x) = C\int_0^{\lambda_{\rm max}} \frac{\lambda}{m^2+\lambda^2} a^3\rho(\lambda)d\lambda \tag{2.7}$$

Note the upper limit $\lambda_{\rm max}$ depends on the type of fermions used and will be of order π/a. We can easily estimate the size of the contributions to this integral from two regions: $\lambda\sim\lambda_{\rm max}$ and $\lambda\sim 0$.

For large λ, we can rely on asymptotic freedom and use a free-field estimate for the eigenvalue density $\rho(\lambda)\approx(24\pi^2\lambda^3)/(2\pi)^4$ which implies (assuming three colors and a single flavor):

$$\int_{\lambda_{\rm min}}^{\lambda_{\rm max}} a^3\frac{\rho(\lambda)}{\lambda}d\lambda \approx \int_{\lambda_{\rm min}}^{\lambda_{\rm max}} \frac{24\pi^2\lambda^2}{(2\pi)^4a^3}d\lambda \approx 2.4 + O(\lambda_{\rm min}a)^3 \tag{2.8}$$

where the lower limit $\lambda_{\rm min}$ is chosen sufficiently large that the free field form for $\rho(\lambda)$ is valid for the entire integration range and the 2.4 is an estimate for Wilson fermions. Alternatively, for small λ, we can treat $\rho(\lambda)$ as a constant determined by the Banks-Casher relation, $\rho(\lambda)=\langle\overline{q}q\rangle/\pi$:

$$\int_0^{\Lambda_{\rm max}} a^3\frac{\rho(\lambda)\lambda}{m^2+\lambda^2}d\lambda \approx a^3\rho(0)\ln(\Lambda_{\rm max}/m). \tag{2.9}$$

Here $\Lambda_{\rm max}$ represents the upper limit of the region where $\rho(\lambda)\approx\rho(0)$. Comparing Eqs. 2.8 and 2.9, we expect the largest contribution to the force comes from the much more numerous eigenvalues $\lambda\sim 1/a$ and that the light quark mass dependence is logarithmic and suppressed by a^3.

Thus if we return to the computer cost estimates for the standard HMC algorithm and continue to neglect the possible effects of near zero modes localized on the lattice scale, we conclude that a more realistic estimate for the cost of generating an independent lattice configuration can be obtained by dropping the factor (D) from Eq. 1.1 with the result:

$$\text{Cost}_{\rm total} = C_3(\frac{L}{\rm fm})^5(\frac{\rm MeV}{m_l})^{1.5}(\frac{\rm fm}{a})^7. \tag{2.10}$$

PoS(LATTICE 2007)028

3. Effects of mass preconditioning

We now apply Hasenbusch preconditioning [1] and add an intermediate, Hasenbusch mass M to the evolution replacing:

$$\det(D+m_l) \to \det(D+M)\det\Big(\frac{D+m_l}{D+M}\Big), \tag{3.1}$$

where m_l is the physical light quark mass. Different step sizes are used for the forces coming from each factor. For the first factor each step will require inverting a Dirac operator with the mass M while the force resulting from the second term will require an inversion with the mass m_l.

The cost of generating a trajectory arising from each of the resulting two forces can then be estimated as the product of the cost of the Dirac operator inversion, $\propto 1/M$ or $1/m_l$, times the force estimated in the style of the previous section:

$$\text{Cost} = C_4\left(\frac{L}{a}\right)^5 \frac{1}{m_\pi a}\Big\{\frac{1}{Ma}\int_0^{\lambda_{\max}} \frac{\lambda}{\lambda^2+M^2}a^3\rho(\lambda)d\lambda + \frac{1}{m_l a}\int_0^{M}\frac{\lambda}{\lambda^2+m_l^2}a^3\rho(\lambda)d\lambda\Big\}. \tag{3.2}$$

Here we have included the overall factor of L^4 for the number of operations, L for the volume dependence of the step size and the autocorrelation factor $1/m_\pi$ to make this easy to compare with Eq. 1.1. The force coming from the ratio of determinants has been simplified by treating the extra $\det(D+M)$ factor as simply limiting the integration over λ to the region $0 \le \lambda \le M$ within which the effects of the two determinants do not cancel.

We can now study Eq. 3.2 to determine the optimal choice for the *ad hoc* parameter M. To make this study easier, we will assume a simple model for the Dirac eigenvalue density $\rho(\lambda)$:

$$\rho(\lambda) = \begin{cases} \frac{1}{\pi}\langle\overline{q}q\rangle & \text{for } \lambda \le \Lambda_{\min} \\ 12\frac{2\pi^2\lambda^3}{(2\pi)^4} & \text{for } \Lambda_{\min} \le \lambda. \end{cases} \tag{3.3}$$

in which we simply join a Banks-Casher constant region and the naive free-field behavior at the point $\lambda = \Lambda_{\min}$. If we use $\langle\overline{q}q\rangle = (250\text{MeV})^3$, continuity requires $\Lambda_{\min} = 320$ MeV.

We can now combine this model for $\rho(\lambda)$ and with the cost estimate of Eq. 3.2 to determine the optimal choice of M. If M is chosen in the region $0 \le M \le \Lambda_{\min}$ we find the minimum is at the upper limit, $M = \Lambda_{\min}$, even for a physically light quark mass. For $\Lambda_{\min} \le M \le \lambda_{\max}$, the minimum occurs for $M = (m_l\pi^3/3)^{1/4}$ which is approximately 1 GeV for $m_l = 0.01/a$. For such a value of M, the dependence of the minimum computational cost on the light quark mass comes from a sum of terms behaving as $\langle\overline{q}q\rangle a^2/m_l$ and $1/(m_l a)^{1/4}$ with the later term roughly twice the former for $m_l = 0.01/a$.

4. Comparison with 2+1 flavor DWF simulations

Of course, such a simple treatment of the light quark mass dependence of the computational cost is unlikely to describe an actual situation where other consideration enter as the specific arrangement of terms and algorithms is chosen for a particular calculation. For example, in the case of the RBC-UKQCD, 2+1 flavor, domain wall fermion (DWF) simulations which we are interested

PoS(LATTICE 2007)028

in modeling, choosing the strange quark mass as the Hasenbusch mass saves sufficient operations to more than compensate for the resulting non-optimal value of M.

These 2+1 flavor, DWF simulations are performed using the positive definite operator:

$$\mathscr{D}(m) = D_{\rm DWF}(m)^\dagger D_{\rm DWF}(m). \tag{4.1}$$

Here $D_{\rm DWF}(m)$ is the standard DWF operator with the mass term m connecting the right and left walls. (The five-dimensional Wilson mass M_5, the "domain wall height", is fixed at 1.8 and not explicitly displayed.) An efficient evolution algorithm with a reduced number of force term calculations results if the target ratio of determinants is written:

$$\frac{\det[\mathscr{D}(m_l)]\det[\mathscr{D}(m_s)]^{\frac{1}{2}}}{\det[\mathscr{D}(1)]^{\frac{3}{2}}} = \frac{\det[\mathscr{D}(m_l)]}{\det[\mathscr{D}(m_s)]}\cdot\frac{\det[\mathscr{D}(m_s)]^{\frac{1}{2}}}{\det[\mathscr{D}(1)]^{\frac{1}{2}}}\cdot\frac{\det[\mathscr{D}(m_s)]^{\frac{1}{2}}}{\det[\mathscr{D}(1)]^{\frac{1}{2}}}\cdot\frac{\det[\mathscr{D}(m_s)]^{\frac{1}{2}}}{\det[\mathscr{D}(1)]^{\frac{1}{2}}} \tag{4.2}$$

Here the determinants in the denominator containing unit quark mass are the Pauli-Villars terms needed to cancel the contributions of the naively irrelevant 5-dimensional DWF modes whose numbers are enhanced in the limit that extent in the fifth dimension, L_s grows large. The first factor on the right-hand-side is evaluated using the standard HMC algorithm. The remaining three factors require the RHMC method.

If we apply the style of cost estimate developed earlier to this use of strange quark mass for Hasenbusch preconditioning we find:

$$\text{Cost} = C\left(\frac{L}{a}\right)^5 \frac{1}{m_\pi a}\left\{\frac{3}{m_s a}\int_{m_s}^{\Lambda_{\rm max}} a^3\rho(\lambda)d\lambda + \frac{2}{m_l a}\int_0^{m_s}\frac{\lambda}{\lambda^2+m_l^2}a^3\rho(\lambda)d\lambda\right\} \tag{4.3}$$

$$\approx \left(\frac{L}{fm}\right)^5 \frac{MeV}{m_\pi}\left(\frac{fm}{a}\right)^6\left\{C_0 + C_1\left(\frac{MeV}{m_K}\right)^2\left(\frac{fm}{a}\right) + C_2\left(\frac{MeV}{m_\pi}\right)^2\left(\frac{a}{fm}\right)^2\right\} \tag{4.4}$$

where the factors of 3 and 2 in Eq. 4.3 reflect the three factors of the strange and two factors of light quark determinant in Eq. 4.2.

To obtain Eq. 4.4, we substitute the ansätz for $\rho(\lambda)$ given in Eq. 3.3, ignore a slowly varying $\ln(m_s/m_l)$ factor and add the constant C_0 to account for those parts of the evolution algorithm which don't involve Dirac matrix inversion and are hence mass independent. We now use this result to describe the performance found in the RBC-UKQCD $24^3\times 64$ ($1/a = 1.73$ GeV) and $32^3\times 64$ ($1/a \approx 2.3$ GeV) simulations.

We compare the predictions of Eq. 4.4 with the observed behavior in Table 1. We have chosen the parameters C_0, C_1 and C_2 to reproduce the computation times listed in the third row of the table. While the agreement between Eq. 4.4 and the actual performance of the code is not perfect, the calculated total times agree with those observed to within about 20%.

It is interesting to use the values determined for C_0, C_1 and C_2 to predict the computer resources needed for a DWF simulation at realistic pion masses and a $(6\text{ fm})^3$ volume. Using $a = 0.86$ fm, a $64^3\times 128$ volume and a 137 MeV pion we find a predicted total time of 150 Petaflops days to collect 20,000 molecular dynamics time units: a possible goal for Lattice 2011?

We would like to thank our RBC and UKQCD collaborators with whom this work was performed as well as the University of Edinburgh, PPARC, RIKEN, BNL and the U.S. DOE for providing the QCDOC computers and other facilities which were essential for this work. This research was also supported by DOE grants DE-FG02-92ER40699 and DE-AC02-98CH10886.

PoS(LATTICE 2007)028

Table 1: Execution times for the three parts of the RBC-UKQCD evolution code for the two lattice sizes and lattice spacings studied. Here "Const" labels those parts of the program not involving Dirac matrix inversion, "MInv" the time taken by the multimass inverter used for ratios of strange and Pauli-Villars determinants and "CG" the inversions evaluating the ratio of light and strange determinants. For each double row, the top row shows the observed times and the lower row the prediction of Eq. 4.4. The third double row is used to determine $C_0 - C_2$. Times are given in Teraflops days need to generate 50 molecular dynamics time units.

Volume	a (fm)	m_l	ml/m_s	m_π (MeV)	Const (days)	MInv (days)	CG (days)	Tot (days)
$24^3 \times 64$	0.114	0.005	0.186	315	0.21	0.79	0.37	1.4
					0.17	0.70	0.37	1.2
$24^3 \times 64$	0.114	0.01	0.302	402	0.18	0.70	0.21	1.1
					0.17	0.70	0.23	1.1
$24^3 \times 64$	0.114	0.02	0.535	534	0.17	0.70	0.13	1.0
					0.17	0.70	0.13	1.0
$24^3 \times 64$	0.114	0.03	0.767	620	0.17	0.71	0.10	1.0
					0.17	0.70	0.09	1.0
$32^3 \times 64$	0.086	0.004	0.150	276	0.60	2.5	1.2	4.3
					0.50	2.9	1.1	4.5
$32^3 \times 64$	0.086	0.006	0.216	330	0.71	2.5	0.90	4.1
					0.49	2.9	0.75	4.1

References

[1] M. Hasenbusch, *Speeding up the hybrid-monte-carlo algorithm for dynamical fermions* , *Phys. Lett.* **B519** (2001) 177–182, [`hep-lat/0107019`].

[2] C. Urbach, K. Jansen, A. Shindler, and U. Wenger, *Hmc algorithm with multiple time scale integration and mass preconditioning*, *Comput. Phys. Commun.* **174** (2006) 87–98, [`hep-lat/0506011`].

[3] M. A. Clark and A. D. Kennedy, *The rhmc algorithm for 2 flavors of dynamical staggered fermions*, *Nucl. Phys. Proc. Suppl.* **129** (2004) 850–852, [`hep-lat/0309084`].

[4] M. A. Clark and A. D. Kennedy, *Accelerating dynamical fermion computations using the rational hybrid monte carlo (rhmc) algorithm with multiple pseudofermion fields*, *Phys. Rev. Lett.* **98** (2007) 051601, [`hep-lat/0608015`].

[5] **RBC and UKQCD** Collaboration, D. J. Antonio *et al.*, *Production and properties of 2+1 flavor dwf ensembles*, *PoS* **LAT2006** (2006) 188.

[6] **RBC and UKQCD** Collaboration, C. Allton *et al.*, *2+1 flavor domain wall qcd on a (2 fm)*^{3} *lattice:light meson spectroscopy with* l_s *= 16*, *Phys. Rev.* **D76** (2007) 014504.

[7] J. C. Sexton and D. H. Weingarten, *Hamiltonian evolution for the hybrid monte carlo algorithm* , *Nucl. Phys.* **B380** (1992) 665–678.

[8] **RBC-UKQCD** Collaboration, C. Jung, *Status of (2+1) flavor,* 32^3*x*64 *domain wall fermion simulations*, *PoS* **LAT2007** (2007) 037.

[9] C. Bernard *et al.*, *Panel discussion on the cost of dynamical quark simulations* , *Nucl. Phys. Proc. Suppl.* **106** (2002) 199–205.

PoS(LATTICE 2007)028