

Note on Grid implementation of RM-HMC and LA-HMC

Guido Cossu, ...

Edinburgh University, UK

Abstract

Description of the tests on two modifications of the Hybrid Monte Carlo (HMC) algorithm, the Riemannian Manifold Hamiltonian Monte Carlo, the Look Ahead HMC. Both algorithms are designed to travel much farther in the Hamiltonian phase space for each trajectory and reduce the autocorrelations among physical observables thus addressing the problem of the critical slowing down towards the continuum limit. We present a comparison of costs of the new algorithms with the standard HMC evolution for pure gauge fields, studying the autocorrelation times for various quantities including the topological charge.

Keywords: Grid, HMC, Topology freezing

Contents

1	Introduction	2
1.1	Markov Chain Monte Carlo	2
1.2	Hybrid Monte Carlo	2
2	Grid implementation of RMHMC via Duane & Pendleton	4
2.1	Covariant Laplacian	4
2.2	Action	4
2.3	Integration scheme	5
2.4	Full action	6
2.4.1	Full integration scheme	7
2.5	Preliminary correctness runs	7
2.6	Further runs plan	11
3	Look Ahead HMC	12
3.1	Detailed balance and new algorithm	12
4	Methodology	13
5	Reference results and tables	13
6	LAHMC results and tables	16
7	RMHMC tables	18
8	Plots	18
9	Histories for the average energy T_0. $L = 32$.	25
10	Histories for the topological charge Q. $L = 32$.	27
11	References	29

1. Introduction

One of the frontiers of recent Lattice QCD studies is providing precise, sub-percent, Standard Model predictions in the heavy quark sector to compare against the experiments. Accurately including dynamical heavy quarks with mass approaching several GeV in the discretised theory simulations requires fine lattice spacings, below 0.05 fm, approaching the continuum limit.

The continuum limit of a discretised theory is described by a critical point of second order where all the correlation lengths and the correlation times diverge with some power of the inverse lattice spacing, $\tau_{\text{int}} \propto a^{-\varepsilon}$. This problem is typically referred as critical slowing down. In general the (integrated) autocorrelation time is going to depend on the algorithm used to generate the Markov chain and on the observable under study. For the Langevin algorithm it has been proven [1] that this power is universal, $\varepsilon = 2$. For the Hybrid Monte Carlo scheme (HMC, sometimes referred as Hamiltonian Monte Carlo) [2] we do not know the actual exponents and there are claims that these may not be universal since the algorithm is non-renormalisable for the ϕ^4 theory [3].

Critical slowing down depends on how much the observable will couple to the low modes of the transition matrix of the evolution algorithm. The cost to generate ensembles whose sampling is in practice ergodic will increase accordingly.

There are several studies in the literature on how to reduce the cost of configuration generation for theories close to the continuum limit [4, 5, 6, 7]. In these proceedings we discuss two modifications of the HMC algorithm that have the potential to reduce the total cost of the expensive dynamical simulations with fermions. The first modification, called Riemannian Manifold HMC (RMHMC), modifies the kinetic term of the HMC Hamiltonian to speed up the evolution of the slowest modes. The second, referred as Look Ahead HMC (LAHMC), drops the detailed balance requirement in favour of a generalized principle with the effect of an increased trajectory length.

The first step is to show effective reductions of the autocorrelation times and related algorithm cost in the simple cases of SU(3) pure gauge theory that already exhibits critical slowing down [6, 8, 9, 10, 11].

1.1. Markov Chain Monte Carlo

The basic problem of simulating a quantum field theory is that we have to sample field configurations distributed according to some target probability $p(x) \propto \exp(-S(x))$ where $S(x)$ is the corresponding action functional of the field x . If we are able to generate a Markov chain sequence with transition probabilities $T(x'|x)$ from the state x to state x' that satisfies

$$\int p(x)T(x'|x)dx = p(x') \quad (1)$$

i.e. $p(x)$ is a *fixed point* of the transition matrix¹, then the expectation value of an observable can be computed by simply averaging the measurements over the generated ensemble. For this purpose, a sufficient but not necessary condition is for $T(x'|x)$ to satisfy the detailed balance relation:

$$p(x)T(x'|x) = p(x')T(x|x'). \quad (2)$$

i.e. that for the pair x, x' , the probability of going from state x to x' is the same as the probability of reaching state x from x' . Equation (2) automatically implies the fixed point condition, but introduces a random walk component in the HMC. This relation is satisfied by the accept-reject step in the HMC.

1.2. Hybrid Monte Carlo

The Hybrid Monte Carlo scheme [2] for the generation of correctly distributed configurations for a field ϕ extends the space with a fictitious time coordinate and generates new configurations according to the distribution $q(\phi)$:

$$q(\phi) \propto \exp(-H(\phi)) \quad H(P, \phi) = \frac{1}{2}P^T P + S(\phi) \quad (3)$$

where P are the conjugate momenta for ϕ in the Hamiltonian evolution and it enters as a trivial Gaussian factor. An HMC update step can be easily described as an application of a sequence of operators. In order to fix our notation these operators are

- $L(\epsilon, M)$, any integrator that is area preserving and reversible. Runs the integration for M steps of size ϵ .
- F , momenta sign-flip

¹In a discrete state space the target distribution is an eigenstate of the transition matrix.

- $R(\alpha)$, randomization of the momenta by Gaussian noise, $\alpha \in [0, 1]$. The parameter α is the mixing between the new random momenta P_{new} and the old ones P_{old} , i.e. $R(\alpha) = \sqrt{\alpha}P_{\text{new}} + \sqrt{(1-\alpha)}P_{\text{old}}$ it is typical to set $\alpha = 1$, i.e. a complete randomization of momenta.

Using these operators we can write the HMC in a compact form as follows. The $x^{(t,n)}$ stands for the configuration state (P, ϕ) at the point t of the Markov chain and at the step n of the evolution sequence to generate the new Markov chain state. We have 3 sub-states $n = 0, 1, 2$ in the HMC for the generation of a new state.

1. $x' = FLx^{(t,0)}$, where $x^{(t,0)}$ is the initial state. Then compute the acceptance probability using the Metropolis step and $x^{(t,1)} = x'$ if accepted, otherwise $x^{(t,1)} = x^{(t,0)}$.
2. Flip the momenta. $x^{(t,2)} = Fx^{(t,1)}$.
3. Randomize momenta. The new element of the Markov chain is $x^{(t+1,0)} = R(\alpha)x^{(t,2)}$.

The FL step is necessary for the HMC in order to cancel the contribution of forward and backward transition probabilities in the Metropolis-Hastings acceptance probability of state x' , $\pi(x|x')$

$$\pi(x|x') = \min\left(1, \frac{p(x')}{p(x)} \frac{T(x|x')}{T(x'|x)}\right) \quad (4)$$

since $(FL)^{-1} = L^{-1}F = FL$.

2. Grid implementation of RMHMC via Duane & Pendleton

We follow the ideas of Duane et al. [12, 13] on Fourier acceleration of slow modes in the evolution of the Markov chain, that was recently revisited by Girolami and Calderhead [14] proposing an exact algorithm for non separable hamiltonians. This should allow to fight the critical slowing down at small lattice spacing. The algorithm acknowledges that the hamiltonian manifold where the evolution takes place is not flat and modifies the kinetic term with a metric defined by the laplacian operator.

The plan is to investigate the decorrelation properties of observables like the topological charge and large Wilson Loops, or smeared plaquette, using the new algorithm. We will start with pure gauge simulations on modest lattice sizes ($L = 16$ sites per direction) in a region of the parameter space that is known to have large autocorrelation times [6] and then move to full QCD on a second stage.

2.1. Covariant Laplacian

The core of the RM-HMC algorithm is the modification of the kinetic term including the inverse of an operator that tries to capture the slow modes in the configuration. One possible choice, of many, is the covariant laplacian operator for fields ϕ in the adjoint representation of $SU(N)$.

The discretized version on the lattice is written as:

$$\nabla^2 \phi(x) = \sum_{\mu}^d \left[U_{\mu}(x) \phi(x + \mu) U_{\mu}^{\dagger}(x) + U_{\mu}^{\dagger}(x - \mu) \phi(x - \mu) U_{\mu}(x - \mu) - 2\phi(x) \right] \quad (5)$$

that has the correct continuum limit and covariantly transforms in the adjoint representation

$$\Omega(x)[\nabla^2 \phi(x)]\Omega^{\dagger}(x) = \nabla^2 \phi(x) \quad (6)$$

Continuum version

$$\nabla^2 \phi(x) = D_{\mu}^{\dagger} D_{\mu} \phi = \partial^2 \phi(x) + ig[\partial_{\mu} A_{\mu}, \phi] + 2ig[A_{\mu}, \partial_{\mu} \phi] - g^2[A_{\mu}, [A_{\mu}, \phi]] \quad (7)$$

and the last term can be rewritten as $-g^2\{A_{\mu}^2, \phi\}$.

Operator M in the RM-HMC

$$M\phi(x) = (1 - \kappa)\phi(x) - \frac{\kappa}{4d}\nabla^2 \phi(x) \quad (8)$$

and we get maximal decoupling (in the free field theory) for $\kappa \rightarrow 1$.

2.2. Action

Kinetic part of the action, momenta p_{μ} are matrices in the algebra (with generator trace normalisation $\frac{1}{2}$):

$$K(U) = \sum_{x, \mu}^d \text{tr}(p_{\mu}(x)^{\dagger} M^{-1}(U) p_{\mu}(x)) \quad (9)$$

The correct distribution of momenta is generated by having an antihermitian field $\eta_{\mu}(x)$ distributed as $\exp(\text{tr}[\eta^{\dagger}\eta])$ (so the usual distribution of momenta) and then writing $\eta_{\mu} = M^{-\frac{1}{2}}p_{\mu}$, so that p_{μ} will have the correct distribution.

Momenta

$$p_{\mu} = M^{\frac{1}{2}}\eta_{\mu} \quad (10)$$

and p_{μ} is still antihermitian.

The derivative in direction T^{α} (lie algebra generator) of the action eq. 9 is

$$T_{ab}^{\alpha} \frac{\delta K(U)}{\delta \omega_{\mu}^{\alpha}(x)} = -T_{ab}^{\alpha} \sum_{\nu}^d \text{tr} \left[(M^{-1}p_{\nu})^{\dagger} \frac{\delta M(U)}{\delta \omega_{\mu}^{\alpha}(x)} (M^{-1}p_{\nu}) \right] \quad (11)$$

$$= -T_{ab}^{\alpha} \sum_{\nu}^d \text{tr} \left[\xi_{\nu}^{\dagger} \frac{\delta M(U)}{\delta \omega_{\mu}^{\alpha}(x)} \xi_{\nu} \right] \quad (12)$$

where $\xi_{\nu} = M^{-1}p_{\nu}$ and the trace summation over the colour indexes is implicit.

Let us compute the derivative of M with $U_\mu(x) = \exp(iT^\alpha\omega_\mu^\alpha(x))$. I am dropping the prefactor $-\frac{\kappa}{4d}$ from now on. So

$$\begin{aligned} \xi_\nu^\dagger \frac{\delta M(U)}{\delta \omega_\mu^\alpha(x)} \xi_\nu &= \xi_\nu^\dagger(x)(iT_{cd}^\alpha)U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x) \\ &\quad + \xi_\nu^\dagger(x)U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)(-iT_{cd}^\alpha) \\ &\quad + \xi_\nu^\dagger(x+\mu)U_\mu^\dagger(x)\xi_\nu(x)(iT_{cd}^\alpha)U_\mu(x) \\ &\quad + \xi_\nu^\dagger(x+\mu)U_\mu^\dagger(x)(-iT_{cd}^\alpha)\xi_\nu(x)U_\mu(x) \end{aligned} \quad (13)$$

The first term is rewritten as

$$(iT_{cd}^\alpha) \left[U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)\xi_\nu^\dagger(x) \right]_{dc} \quad (14)$$

and the second it is just the adjoint. The third term

$$(iT_{cd}^\alpha) \left[U_\mu(x)\xi_\nu^\dagger(x+\mu)U_\mu^\dagger(x)\xi_\nu(x) \right]_{dc} \quad (15)$$

and the fourth is its adjoint. Notice that first and third terms are the same since $\xi = -\xi^\dagger$.

We get the final force term by contracting with T_{ab}^α and using that $T_{ab}^\alpha T_{cd}^\alpha = \frac{1}{2}[\delta_{ad}\delta_{bc} - \frac{1}{N_c}\delta_{ab}\delta_{cd}]$:

First term

$$\begin{aligned} \frac{1}{2}[\delta_{ad}\delta_{bc} - \frac{1}{N_c}\delta_{ab}\delta_{cd}][U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)\xi_\nu^\dagger(x)]_{dc} &= \\ \frac{1}{2}([U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)\xi_\nu^\dagger(x)]_{ab} - \text{tr}[U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)\xi_\nu^\dagger(x)]) \end{aligned} \quad (16)$$

Since the derivatives wrt U^\dagger are the adjoint (with a minus sign) we can rewrite the total derivative as

$$\frac{\delta K(U)}{\delta U_\mu(x)} = -\frac{\kappa}{4d} \sum_\nu 2[U_\mu(x)\xi_\nu(x+\mu)U_\mu^\dagger(x)\xi_\nu^\dagger(x)]_{\text{TA}} \quad (17)$$

taking the traceless antihermitian part TA.

2.3. Integration scheme

Stormet-Verlet leapfrog integration does not work for non-separable Hamiltonians. It is not reversible and Jacobian of transformations do not have unit determinant.

Leimkuhler and Reich [15] proposed the generalised leapfrog

$$p^{n+\frac{1}{2}} = p^n - \frac{\epsilon}{2} \frac{\delta H}{\delta \theta}(\theta^n, p^{n+\frac{1}{2}}) \quad (18)$$

$$\theta^{n+1} = \theta^n + \frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(\theta^n, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta p}(\theta^{n+1}, p^{n+\frac{1}{2}}) \right] \quad (19)$$

$$p^{n+1} = p^{n+\frac{1}{2}} - \frac{\epsilon}{2} \frac{\delta H}{\delta \theta}(\theta^{n+1}, p^{n+\frac{1}{2}}) \quad (20)$$

that is clearly reversible. In terms of our variables

$$p^{n+\frac{1}{2}} = p^n - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^n, p^{n+\frac{1}{2}}) \quad (21)$$

$$U^{n+1} = \exp \left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^n, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{1}{2}}) \right] \right) U^n \quad (22)$$

$$p^{n+1} = p^{n+\frac{1}{2}} - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{1}{2}}) \quad (23)$$

Attaching blindly more than one of these steps results in an non invertible algorithm:

$$p^{n+\frac{1}{2}} = p^n - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^n, p^{n+\frac{1}{2}}) \quad (24)$$

$$U^{n+1} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^n, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{1}{2}}) \right]\right) U^n \quad (25)$$

$$p^{n+\frac{3}{2}} = p^{n+\frac{1}{2}} - \epsilon \frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{3}{2}}) \quad (26)$$

$$U^{n+2} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{3}{2}}) + \frac{\delta H}{\delta p}(U^{n+2}, p^{n+\frac{3}{2}}) \right]\right) U^n \quad (27)$$

$$p^{n+2} = p^{n+\frac{3}{2}} - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^{n+2}, p^{n+\frac{3}{2}}) \quad (28)$$

because of equation 26. The solution is easily discovered by symmetrising it:

$$p^{n+\frac{1}{2}} = p^n - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^n, p^{n+\frac{1}{2}}) \quad (29)$$

$$U^{n+1} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^n, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{1}{2}}) \right]\right) U^n \quad (30)$$

$$p^{n+\frac{3}{2}} = p^{n+\frac{1}{2}} - \frac{\epsilon}{2} \left[\frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{3}{2}}) \right] \quad (31)$$

$$U^{n+2} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{3}{2}}) + \frac{\delta H}{\delta p}(U^{n+2}, p^{n+\frac{3}{2}}) \right]\right) U^{n+1} \quad (32)$$

$$p^{n+2} = p^{n+\frac{3}{2}} - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^{n+2}, p^{n+\frac{3}{2}}) \quad (33)$$

The equation (33) is general, valid even when fermions are included. In particular notice that the implicit evaluation of the force term would not require any additional computation of the $\frac{\delta H}{\delta U}(U^n)$ force term during the iterative process.

2.4. Full action

The modification of the kinetic term introduces the unwanted determinant of the matrix \mathcal{M} in the path integral. To cancel this term we introduced a new set of auxiliary bosonic fields ϕ_μ in the adjoint representation with action $\text{Tr}[\phi_\mu \mathcal{M} \phi_\mu + \phi^2]$, see [13, 12] and estimated stochastically their contribution.

2.4.1. Full integration scheme

Since the hamiltonian for the auxiliary fields is separable a simple leapfrog is enough. I splitted the update of the auxiliary fields around the update of the original fields in order to make the algorithm invertible.

$$\pi^{n+\frac{1}{4}} = \pi^n - \frac{\epsilon}{4} \phi^n \quad (34)$$

$$p^{n+\frac{1}{2}} = p^n - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^n, p^{n+\frac{1}{2}}, \pi^{n+\frac{1}{4}}) \quad (35)$$

$$\pi^{n+\frac{1}{2}} = \pi^{n+\frac{1}{4}} - \frac{\epsilon}{2} \phi^n \quad (36)$$

$$\phi^{n+\frac{1}{2}} = \phi^n + \frac{\epsilon}{2} M(U^n) \pi^{n+\frac{1}{2}} \quad (37)$$

$$U^{n+1} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^n, p^{n+\frac{1}{2}}) + \frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{1}{2}}) \right]\right) U^n \quad (38)$$

$$\phi^{n+1} = \phi^{n+\frac{1}{2}} + \frac{\epsilon}{2} M(U^{n+1}) \pi^{n+\frac{1}{2}} \quad (39)$$

$$\pi^{n+1} = \pi^{n+\frac{1}{2}} - \frac{\epsilon}{2} \phi^{n+1} \quad (40)$$

$$p^{n+\frac{3}{2}} = p^{n+\frac{1}{2}} - \frac{\epsilon}{2} \left[\frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{1}{2}}, \pi^{n+1}) + \frac{\delta H}{\delta U}(U^{n+1}, p^{n+\frac{3}{2}}, \pi^{n+1}) \right] \quad (41)$$

$$\pi^{n+\frac{3}{2}} = \pi^{n+1} - \frac{\epsilon}{2} \phi^{n+1} \quad (42)$$

$$\phi^{n+\frac{3}{2}} = \phi^{n+1} + \frac{\epsilon}{2} M(U^{n+1}) \pi^{n+\frac{3}{2}} \quad (43)$$

$$U^{n+2} = \exp\left(\frac{\epsilon}{2} \left[\frac{\delta H}{\delta p}(U^{n+1}, p^{n+\frac{3}{2}}) + \frac{\delta H}{\delta p}(U^{n+2}, p^{n+\frac{3}{2}}) \right]\right) U^{n+1} \quad (44)$$

$$\phi^{n+2} = \phi^{n+\frac{3}{2}} + \frac{\epsilon}{2} M(U^{n+2}) \pi^{n+\frac{3}{2}} \quad (45)$$

$$\pi^{n+\frac{7}{4}} = \pi^{n+\frac{3}{2}} - \frac{\epsilon}{4} \phi^{n+2} \quad (46)$$

$$p^{n+2} = p^{n+\frac{3}{2}} - \frac{\epsilon}{2} \frac{\delta H}{\delta U}(U^{n+2}, p^{n+\frac{3}{2}}, \pi^{n+\frac{7}{4}}) \quad (47)$$

$$\pi^{n+2} = \pi^{n+\frac{7}{4}} - \frac{\epsilon}{4} \phi^{n+2} \quad (48)$$

$$(49)$$

We should test the cancelation of the determinant by just introducing the auxiliary momenta π (pseudofermion-like update).

2.5. Preliminary correctness runs

I implemented the implicit version of the Minimum Norm integrator (not described in these notes). These are some performance results.

- Pure gauge Wilson action
- $\beta = 6.2$, lattice size: 16^4
- trajectory length: $\tau = 1.0$
- molecular dynamics steps: $N_{MD} = 30$
- $\kappa = 0.999$
- integrator: modified second order minimum norm with implicit steps
- double precision
- approximation of the exponential function uses Cayley-Hamilton formula..
- Running (20650 configurations, 1000 thermalization steps)
- $\langle \exp(-dH) \rangle = 1.0161 \pm 0.009585$

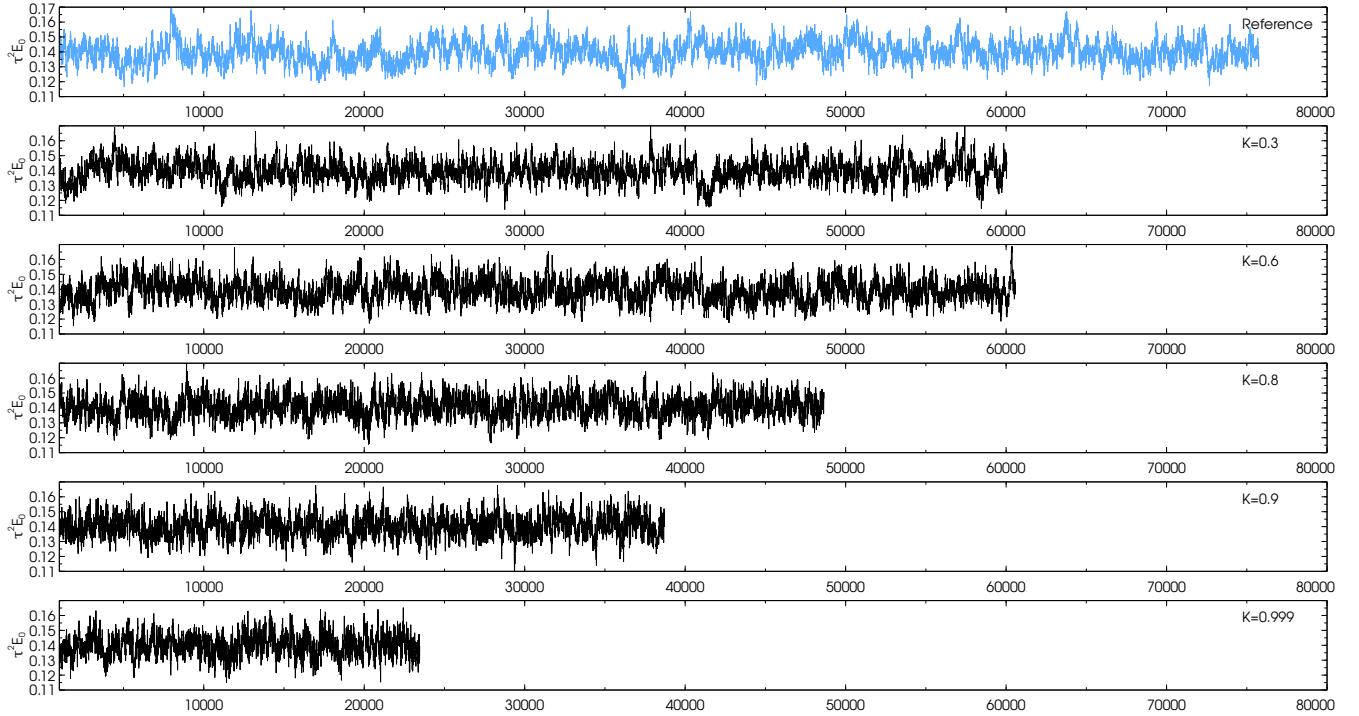


Figure 1: Histories for the flowed energy.

- Plaquette : $0.613619258.367e - 06$ (ref value from [6], with $L = 32$ $p = 0.6136323 \pm 2e - 07$)
- Time per trajectory about 250 seconds (1 KNL)
- Autocorrelations (after $\tau = 2.0$ flow time):
 - Energy, $\tau_E = 11.28 \pm 3.14$
 - Average energy $\langle E \rangle = 0.034941094 \pm 9.468e - 5$
 - Topological charge $\tau_Q = 48.6 \pm 18.72$
- Ref run, Grid $N_{MD} = 17, \tau = 1.0$, 75750 trajectories (acceptance 83%), this is the run used in the plots.
 - Plaquette : $0.61364658 \pm 4.61e - 06$
 - Topological charge $\tau_Q = 148.9 \pm 51$
 - Energy $\tau_E = 36.2 \pm 9.07$, $\langle E \rangle = 0.03493200 \pm 8.65e - 5$
- Ref run, Grid $N_{MD} = 40, \tau = 1.0$, 40200 trajectories (acceptance > 95%),
 - Plaquette : $0.613633 \pm 6.62e - 06$
 - Topological charge $\tau_Q = 162 \pm 87$
 - Energy $\tau_E = 26.9 \pm 8.00$, $\langle E \rangle = 0.0352764152 \pm 9.5e - 5$

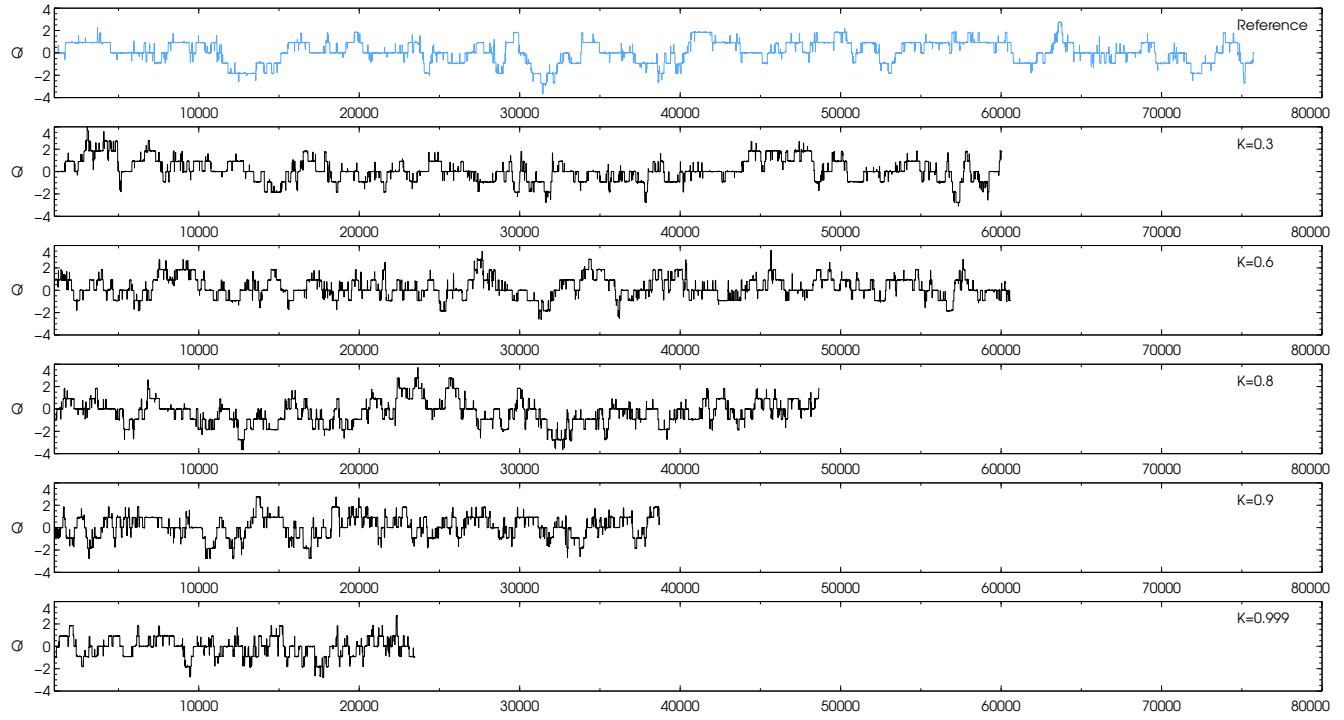


Figure 2: Histories for topological charge.

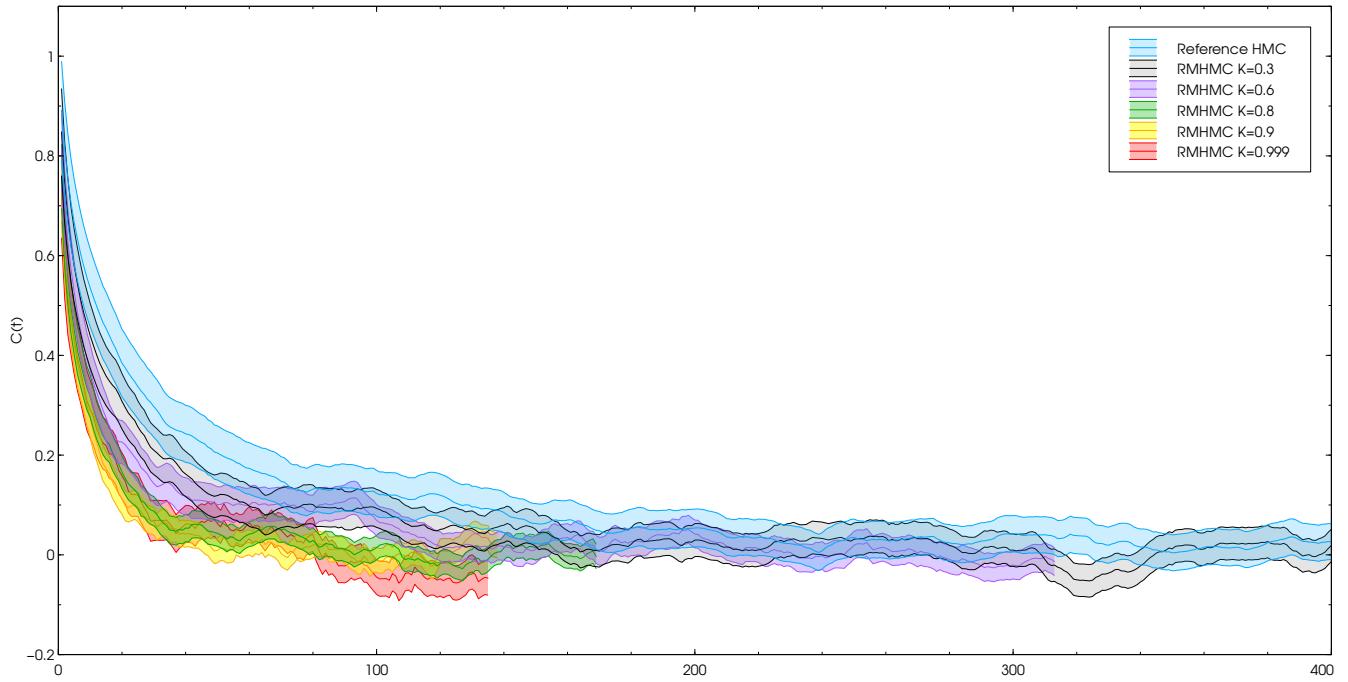


Figure 3: Current analysis for the autocorrelation function of E .

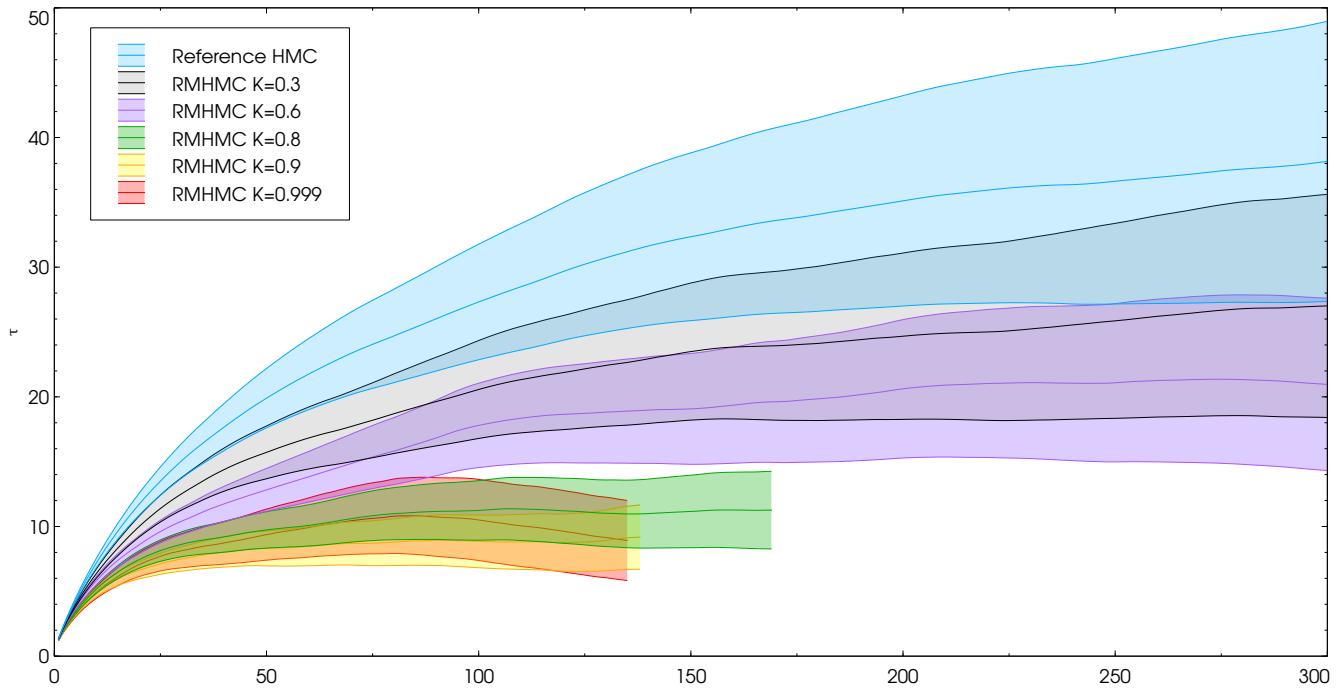


Figure 4: Current analysis for the autocorrelation time of E .

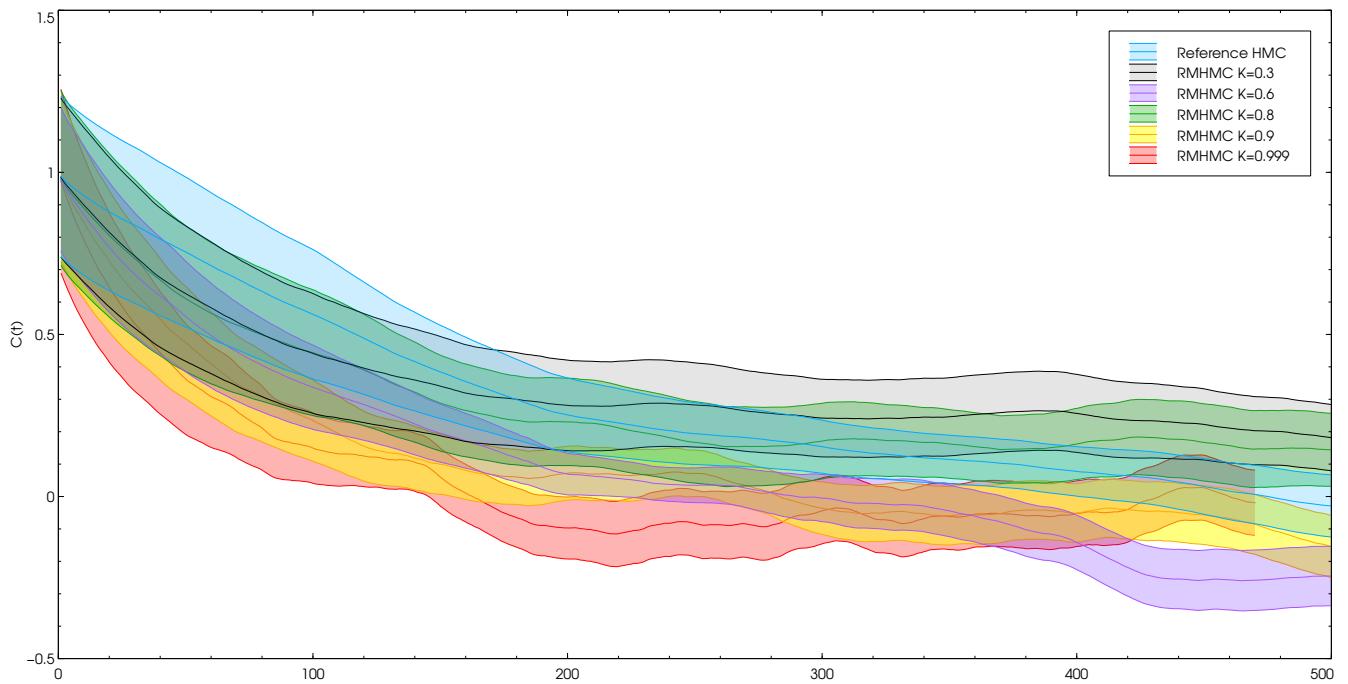


Figure 5: Current analysis for the autocorrelation function of Q .

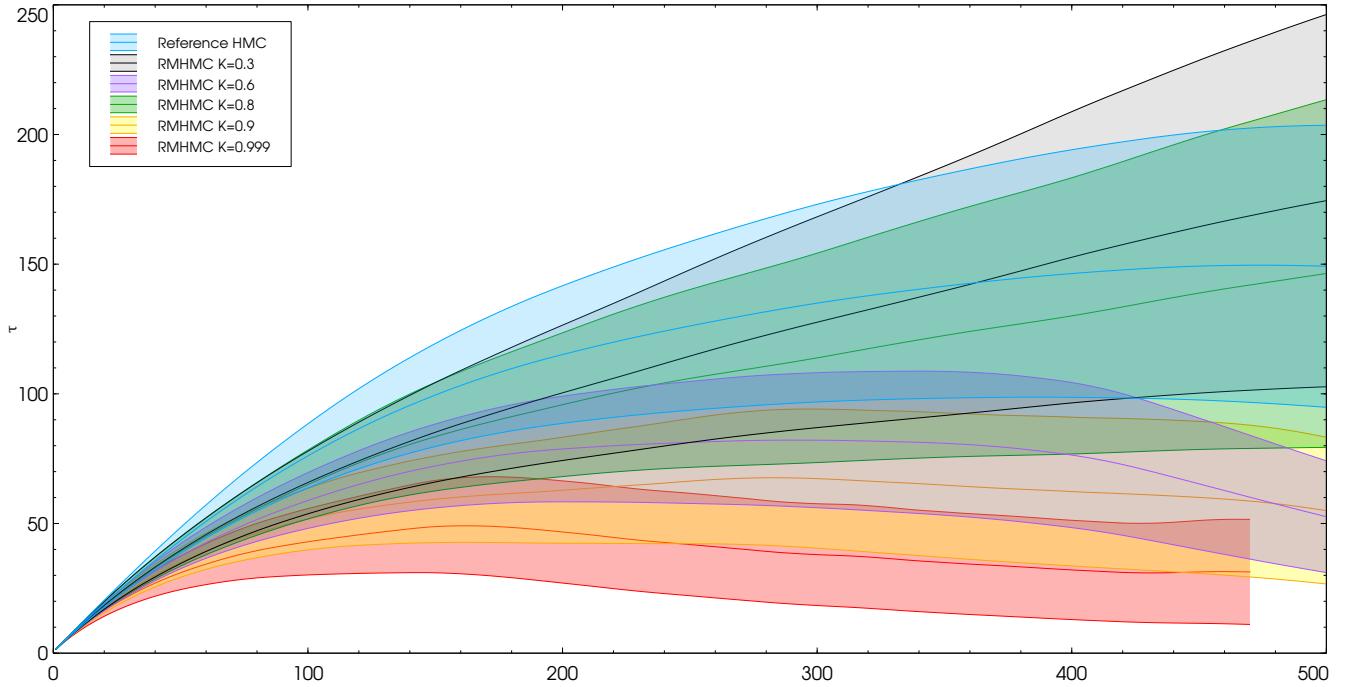


Figure 6: Current analysis for the autocorrelation time of Q .

2.6. Further runs plan

Runs with Wilson pure gauge action

- Reference: S. S. Schaefer paper, [6]
- $\beta = 6.2$, lattice size: 16^4 ($a = 2.9 \text{ Gev}^{-1} = 0.068 \text{ fm}$)
- $\beta = 6.4$, lattice size: 32^4 ($a = 3.8 \text{ Gev}^{-1} = 0.051 \text{ fm}$)
- trajectory length: $\tau = 1.0, 2.0$
- $\kappa = 0.9$
- integrator: modified minimum norm 2nd order with implicit steps
- double precision
- measure E and the topological charge every 5 trajectories (Wilson Flow up to time τ depending on β).

Runs with the DBW2 action to compare with Columbia results (?)

- Pure gauge DBW2 action[5]
- $\beta = 0.9465$, lattice size: $14^4 \times 28$
- trajectory length: $\tau = 1.75$
- molecular dynamics steps: $N_{MD} = 20$ (minimum norm)
- $\kappa = [0.0, \dots, 0.99]$

3. Look Ahead HMC

3.1. Detailed balance and new algorithm

The detailed balance is a sufficient condition for the probability distribution $p(x)$ generated by the Markov Chain of the HMC to be a fixed point of the conditional transition probabilities $T(x'|x)$ of transitioning to the state x' given the state x . It automatically satisfies the fixed point condition, but introduces a random walk component in the HMC.

In order to suppress the random walk behaviour induced by the detailed balance it has been suggested to modify the Metropolis accept-reject step [16, 17]. The resulting probability distribution satisfies the fixed point equation but drops the detailed balance.

Using the notations described in section 1.2 the new LAHMC algorithm [16, 17] consists in the following steps

- Repeat the integration accepting with the modified probabilities $\pi_{L^K}(x^{(t,0)})$ for a maximum number K of times.

$$x^{(t,1)} = \begin{cases} Lx^{(t,0)} & \text{prob } \pi_{L^1}(x^{(t,0)}) \\ L^2x^{(t,0)} & \text{prob } \pi_{L^2}(x^{(t,0)}) \\ \dots & \\ L^Kx^{(t,0)} & \text{prob } \pi_{L^K}(x^{(t,0)}) \\ Fx^{(t,0)} & \text{prob } \pi_F(x^{(t,0)}) \end{cases} \quad (50)$$

- Randomize momenta $x^{(t+1,0)} = R(\alpha)x^{(t,1)}$.

The transition probabilities $x \rightarrow L^a(x)$ are defined in order to satisfy the fixed point equation and are:

$$\pi_{L^a}(x) = \min \left[1 - \sum_{b < a} \pi_{L^b}(x), \frac{p(FL^a x)}{p(x)} \left(1 - \sum_{b < a} \pi_{L^b}(FL^a x) \right) \right] \quad (51)$$

These transition probabilities do not satisfy the detailed balance relation but a more generalized set of identities, called generalized detailed balance [17]. Please have a look at the equations (28)-(33) of [16] for a straightforward demonstration of this fundamental result. Notice that for $K = 1$ the LAHMC reduces to the usual HMC.

The LAHMC has been implemented in Grid [18].

4. Methodology

We are accumulating configurations for standard HMC reference runs at different trajectory lengths and RMHMC and LAHMC. We try to keep the acceptance probability fixed at about 80-83% to avoid introducing any bias in the cost estimate.

We compute the following observables:

- Average Energy T_0 after Wilson Flow of a fixed smearing radius $r = \sqrt{8t}$, t being the Wilson Flow time in this case.

$$T_0 \equiv 2t^2 \frac{\langle S \rangle}{\Omega}. \quad (52)$$

Ω is the 4d volume and S is the total action. Sometimes in this draft I will use the notation E for T_0 .

- Topological charge and susceptibility (after the same Wilson Flow radius smearing). Q is the computed using the simple clover term.

Observables are measured every 5 trajectories. All data presented here are computed in the ensemble of the computed observables, i.e. absolute numbers in the autocorrelation should be multiplied by 5 to compare with other paper/results. Within the notes all results have consistently the same normalization for the autocorrelation.

5. Reference results and tables

The HMC parameters of the runs are:

- L , linear lattice size. Total size $L^4 = \Omega$.
- N_{MD} , number of molecular dynamics steps per trajectory.
- τ , trajectory length.
- Integrator: Minimum norm (Omelyan), 2nd order.

Storing configurations every 5 trajectories.

Name	L	τ	β	N_{MD}	dH	Acceptance
Ref0	16	1.0	6.2	17	0.0877 ± 0.00281	0.83
Ref1	16	1.0	6.2	30	0.0130 ± 0.00282	0.95
Ref2	16	1.0	6.2	40	0.0053 ± 0.00157	0.97
Ref3	16	1.3	6.2	30	0.0244 ± 0.00281	0.91
Ref4	16	1.5	6.2	40	0.0142 ± 0.00119	0.93
Ref5	16	2.0	6.2	40	0.0447 ± 0.00323	0.88
Ref6	16	3.0	6.2	55	0.0693 ± 0.00289	0.85
Ref7	16	4.0	6.2	68	0.0913 ± 0.00134	0.83
Ref8	32	1.0	6.4	35	0.0758 ± 0.00066	0.84
Ref9	32	2.0	6.4	70	0.0774 ± 0.00095	0.84
Ref10	32	3.0	6.4	105	0.0741 ± 0.00302	0.84
Ref11	32	4.0	6.4	140	0.0769 ± 0.00422	0.84

Table 1: Parameters for the reference runs.

The average cost in molecular dynamics steps units per configuration for the observable \mathcal{O} , $C(\mathcal{O})$, is computed as:

$$C(\mathcal{O}) = \frac{N_{traj} N_{MD}}{N_{conf}} \tau_{int}(\mathcal{O}) \quad (53)$$

where $\tau_{int}(\mathcal{O})$ is the integrated autocorrelation time for the observable \mathcal{O} ; N_{traj} and N_{conf} are respectively the number of trajectories generated (in units of τ) and the number of new configurations generated per ensemble. For the HMC case these two numbers are exactly the same. This is not the case for the LAHMC algorithm where for each new configuration we can travel more than one trajectory length. The prefactor before $\tau_{int}(\mathcal{O})$ represents the average number of molecular dynamics steps required to generate a new configuration. Travelling for one molecular dynamics step costs exactly the same for all algorithms, so this is a natural quantity for a fair comparison.

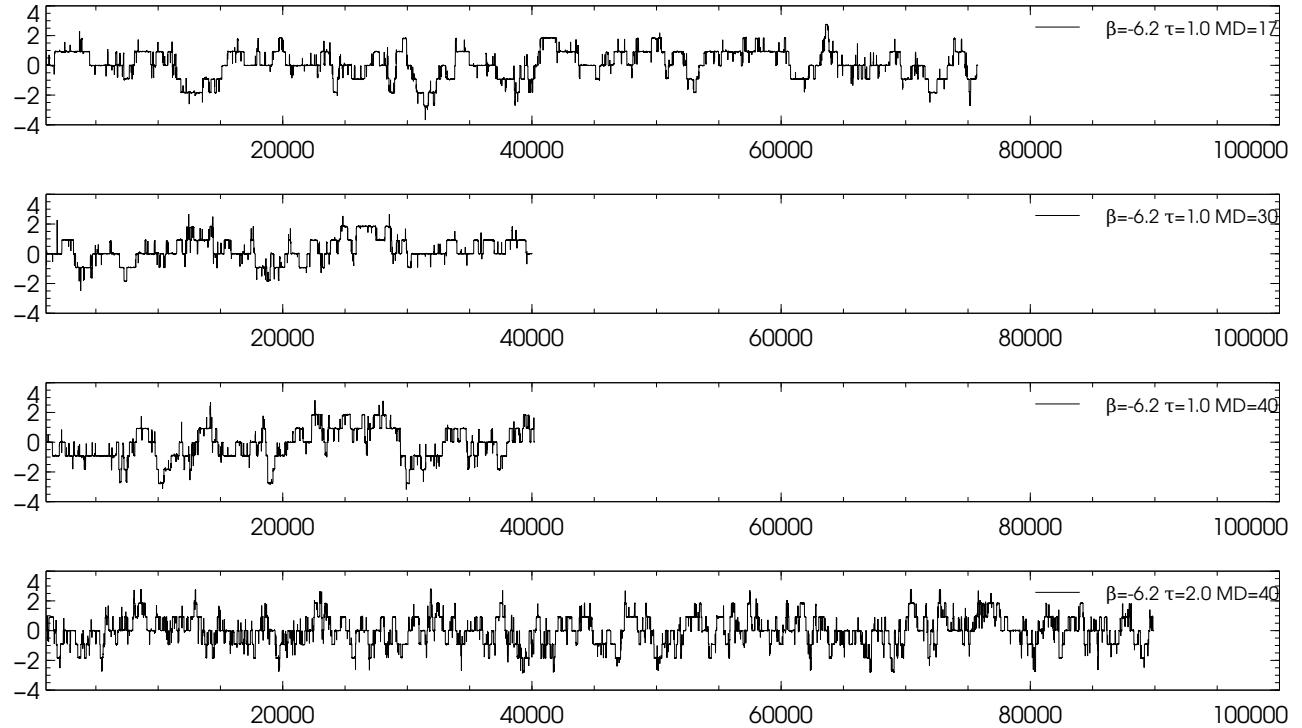


Figure 7: Some topological charge histories for the Ref runs, $L = 16$.

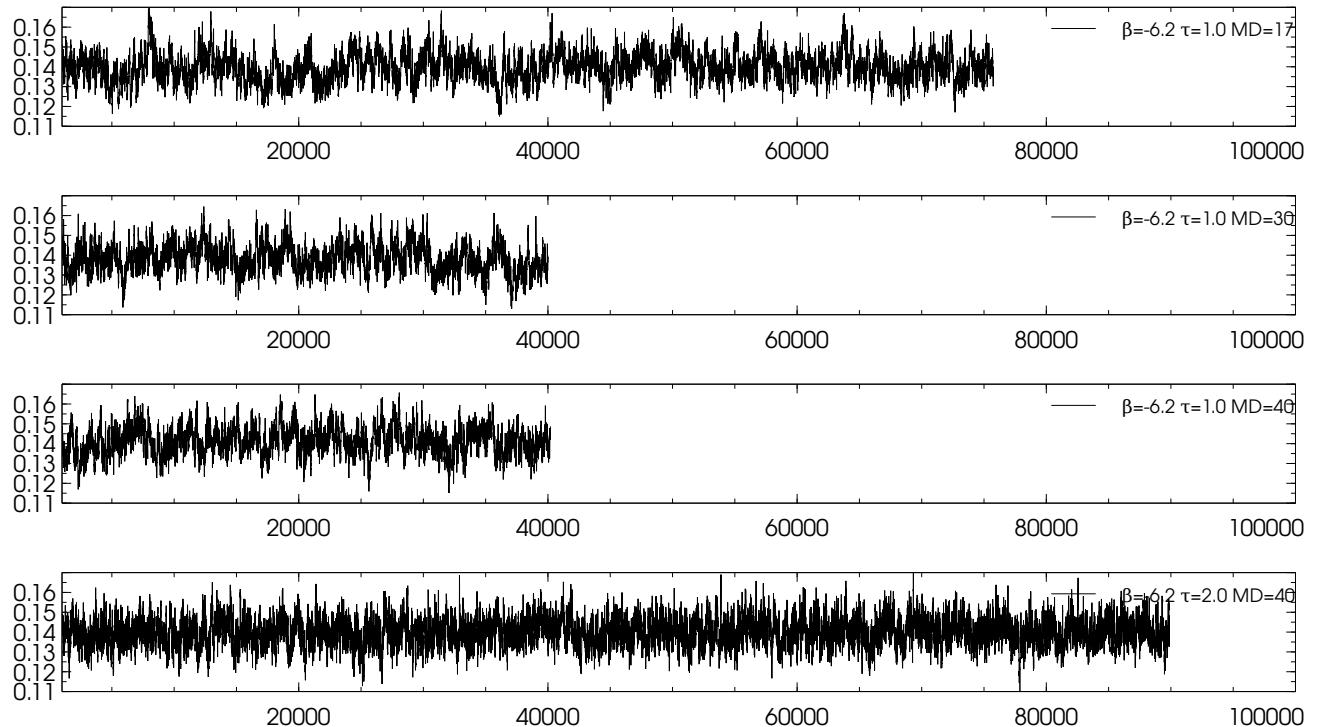


Figure 8: Some E_0 history for the Ref runs, $L = 16$.

ID	N_{conf}	τ_E	Err	τ_Q	Err	τ_{Q^2}	Err
0	75750	36.29	9.071	148.9	51.05	79.26	20.79
1	39995	32.52	9.502	223.7	138.1	128.8	64.31
2	40200	26.98	8.003	162.3	87.71	45.44	11.11
3	98595	18.2	2.775	96.81	21.89	54.4	10.11
4	100500	13.63	1.744	103.2	30.69	38.51	6.653
5	100500	10.3	1.449	50.95	8.654	23.87	3.29
6	100500	4.509	0.3291	29.72	4.161	13.81	1.458
7	100500	4.002	0.3428	20.32	2.418	13.55	1.712
8	405245	76.6721	16.6748	1613.794	873.5266	384.4886	88.5534
9	216775	17.414	2.3878	877.1958	498.8296	196.449	49.4966
10	57130	9.7963	1.0861	282.2526	178.2656	80.8775	21.5634
11	41925	8.466	1.0336	148.6014	69.4152	47.6471	12.7054

Table 2: Reference autocorrelation times in saved-trajectories time (x5 for the total autocorrelation).

ID	C	Cost $_E$	Err	Cost $_Q$	Err	Cost $_{Q^2}$	Err
0	17	616.9	154.2	2532	867.9	1347	353.4
1	30	975.6	285.1	6711	4142	3865	1929
2	40	1079	320.1	6490	3508	1818	444.3
3	30	545.9	83.25	2904	656.7	1632	303.3
4	40	545.1	69.77	4126	1228	1540	266.1
5	40	411.9	57.96	2038	346.1	954.9	131.6
6	55	248	18.1	1635	228.9	759.3	80.2
7	68	272.1	23.31	1381	164.5	921.1	116.4
8	35	2683.5252	583.6172	56482.7898	30573.4305	13457.1012	3099.3673
9	70	1218.9827	167.1429	61403.709	34918.0739	13751.4332	3464.7644
10	105	1028.6119	114.0456	29636.5234	18717.8853	8492.1413	2264.1549
11	140	1185.2389	144.7026	20804.2002	9718.1234	6670.5964	1778.7578

Table 3: Average cost per observable. C defined in eq.53

6. LAHMC results and tables

ID	L	β	N_{MD}	τ	K	α	dH	Acc.	Acc. 1
LA0	16	6.2	10	1.0	5	0.6	0.704 ± 0.0080	0.81	0.55
LA1	16	6.2	10	1.0	5	1.0	0.697 ± 0.0061	0.81	0.55
LA2	16	6.2	15	1.0	5	1.0	0.140 ± 0.0026	0.94	0.79
LA3	16	6.2	20	2.0	5	1.0	0.711 ± 0.0062	0.81	0.55
LA4	16	6.2	30	3.0	5	1.0	0.740 ± 0.0063	0.81	0.54
LA8	32	6.4	10	1.0	5	1.0	1.080 ± 0.0044	0.82	0.55
LA9	32	6.4	20	2.0	5	1.0	1.102 ± 0.0072	0.82	0.55

Table 4: Parameters for the LAHMC runs. The last column "Acc. 1" is the acceptance after the first trajectory, consistent with the acceptance of the HMC with the same number of N_{MD} .

ID	N_{conf}	N_{traj}	Avg. τ	Avg. MD steps	Avg. MD steps $\tau = 1$
0	44095	98173	1.31	22.26	17.05
1	95114	211444	1.3	22.23	17.07
2	40410	60538	1.21	22.47	18.64
3	45805	84335	2.63	44.74	17.04
4	37700	85327	3.94	67.9	17.21
8	39636	88845	1.31	22.42	17.05
9	18395	41131	2.63	44.72	17.02

Table 5: LAHMC runs statistics.

ID	τ_E	Err	τ_Q	Err	τ_{Q^2}	Err
0	15.2	3.067	267.9	177.5	61.73	22.02
1	12.28	1.47	68.86	14.35	40.19	6.959
2	22.42	6.58	95.11	31.7	71.78	22.21
3	4.452	0.6406	28.55	5.79	13.37	1.82
4	2.883	0.31	36.62	12.41	12.54	2.318
8	23.83	9.316	206.9	111.2	120.4	51.13
9	9.016	2.514	125.3	72.17	94.87	46.17

Table 6: Autocorrelation times in saved-trajectories time (x5 for the total autocorrelation). The units are the same for the reference runs.

ID	Cost_E	Err	Cost_Q	Err	Cost_{Q^2}	Err
0	338.3	68.29	5965	3951	1374	490.2
1	273	32.68	1531	319.1	893.4	154.7
2	503.9	147.9	2137	712.4	1613	499.1
3	199.2	28.66	1277	259.1	598	81.45
4	195.8	21.05	2486	842.4	851.7	157.4
8	534.3	208.8	4637	2493	2699	1146
9	403.2	112.4	5605	3228	4243	2065

Table 7: Average cost per observable. $C(\mathcal{O})$ defined in eq.53

7. RMHMC tables

ID	L	β	K	N_{MD}	τ	dH	Accept.	Trajectories
RM0	16	6.2	0.3	30	1.0	0.0133 ± 0.00069	0.93	59908
RM1	16	6.2	0.6	30	1.0	0.0169 ± 0.00062	0.92	88697
RM2	16	6.2	0.8	30	1.0	0.0260 ± 0.00094	0.91	59288
RM3	16	6.2	0.9	30	1.0	0.0368 ± 0.00113	0.89	59635
RM4	16	6.2	0.999	30	1.0	0.0401 ± 0.00276	0.88	14583
RM5	16	6.2	0.9	50	2.0	0.0718 ± 0.00191	0.85	41436
RM8	32	6.4	0.9	40	1.0	0.1209 ± 0.00308	0.80	26510
RM9	32	6.4	0.9	90	2.0	0.1573 ± 0.00670	0.78	9003
RM10	32	6.4	0.9	135	3.0	0.1949 ± 0.00819	0.76	7320

Table 8: Parameters for the RMHMC runs.

ID	N_{conf}	τ_E	Err	τ_Q	Err	τ_{Q^2}	Err
0	60040	23.84	5.558	230.8	130.6	97.39	40.02
1	89760	18.22	3.002	92.11	21.8	60.46	14.65
2	60510	12.34	2.421	149.7	70.97	49.79	11.38
3	60330	8.37	1.048	80.85	25.9	25.34	4.2
4	25210	12.97	3.387	47.03	16.36	26.18	6.711
5	41170	4.748	0.7302	30.93	6.849	14.23	2.407
8	31440	12.06	2.672	276.8	186	117.8	58.85
9	11250	12.12	4.401	122	86.48	63.63	37.24
10	8740	9.261	4.466	55.61	29.85	43.22	21.24

Table 9: RMHMC Reference autocorrelation times in saved-trajectories time (x5 for the total autocorrelation).

ID	C	Cost_E	Err	Cost_Q	Err	Cost_{Q^2}	Err
0	30	715.1	166.7	6924	3918	2922	1201
1	30	546.6	90.07	2763	653.9	1814	439.4
2	30	370.3	72.64	4492	2129	1494	341.4
3	30	251.1	31.44	2425	777.1	760.2	126
4	30	389.1	101.6	1411	490.9	785.5	201.3
5	55	261.2	40.16	1701	376.7	782.4	132.4
8	40	482.6	106.9	11070	7442	4712	2354
9	90	1091	396.1	10980	7783	5727	3351
10	135	1250	602.9	7508	4030	5835	2868

Table 10: RMHMC Average cost per observable. $C(\mathcal{O})$ defined in eq.53

8. Plots

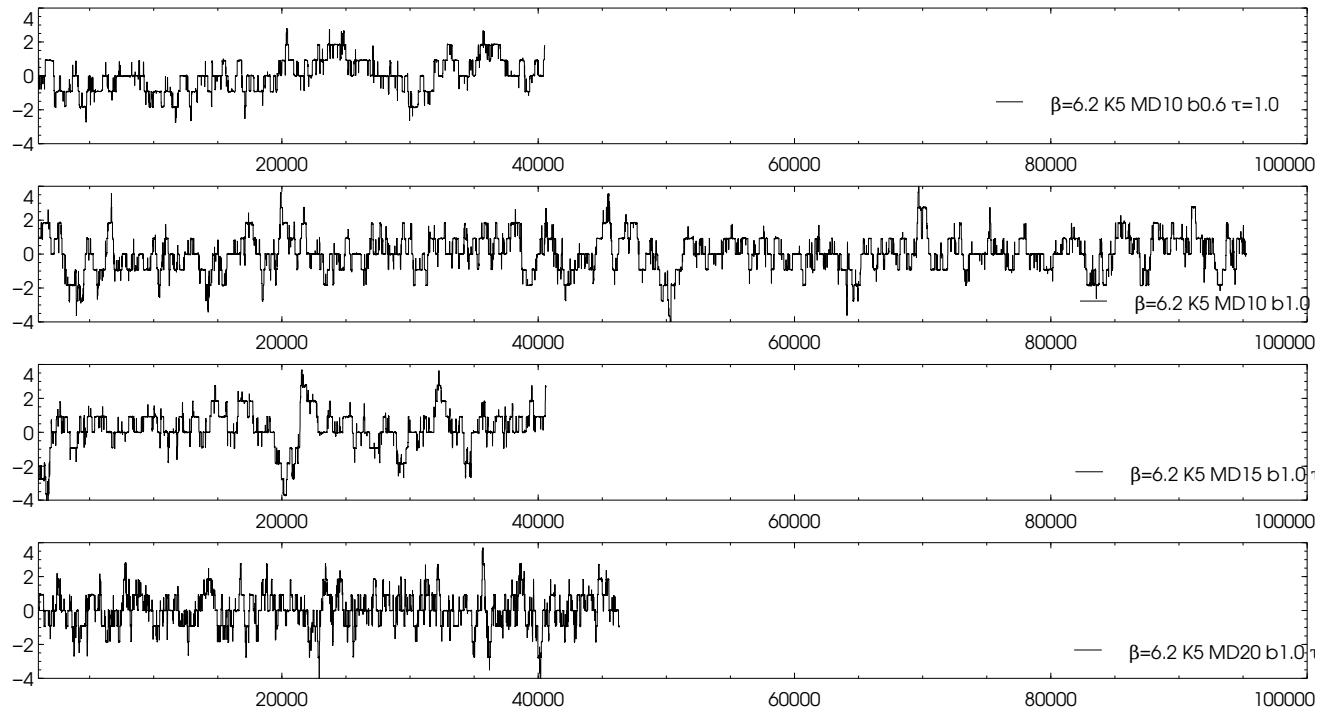


Figure 9: Topological charge history for the LAHMC runs, in the same order as the tables. $L = 16$.

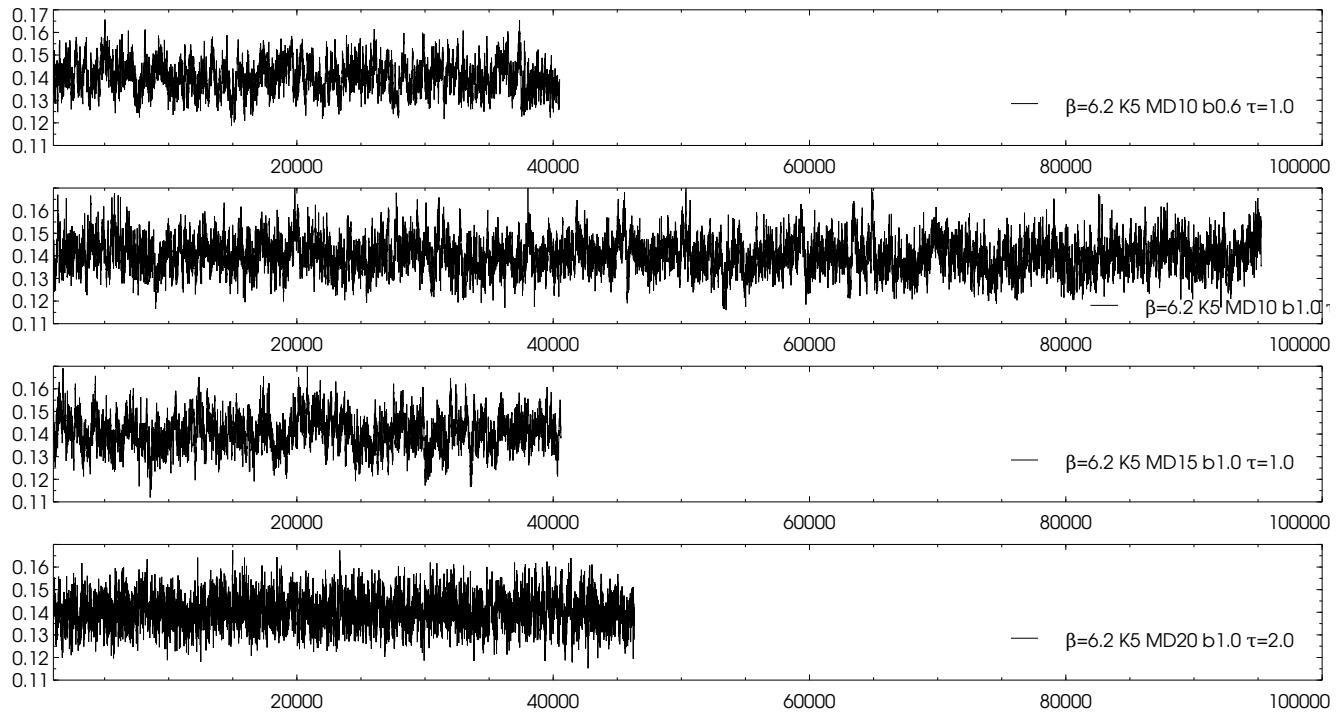


Figure 10: T_0 history for the LAHMC runs, in the same order as the tables. $L = 16$.

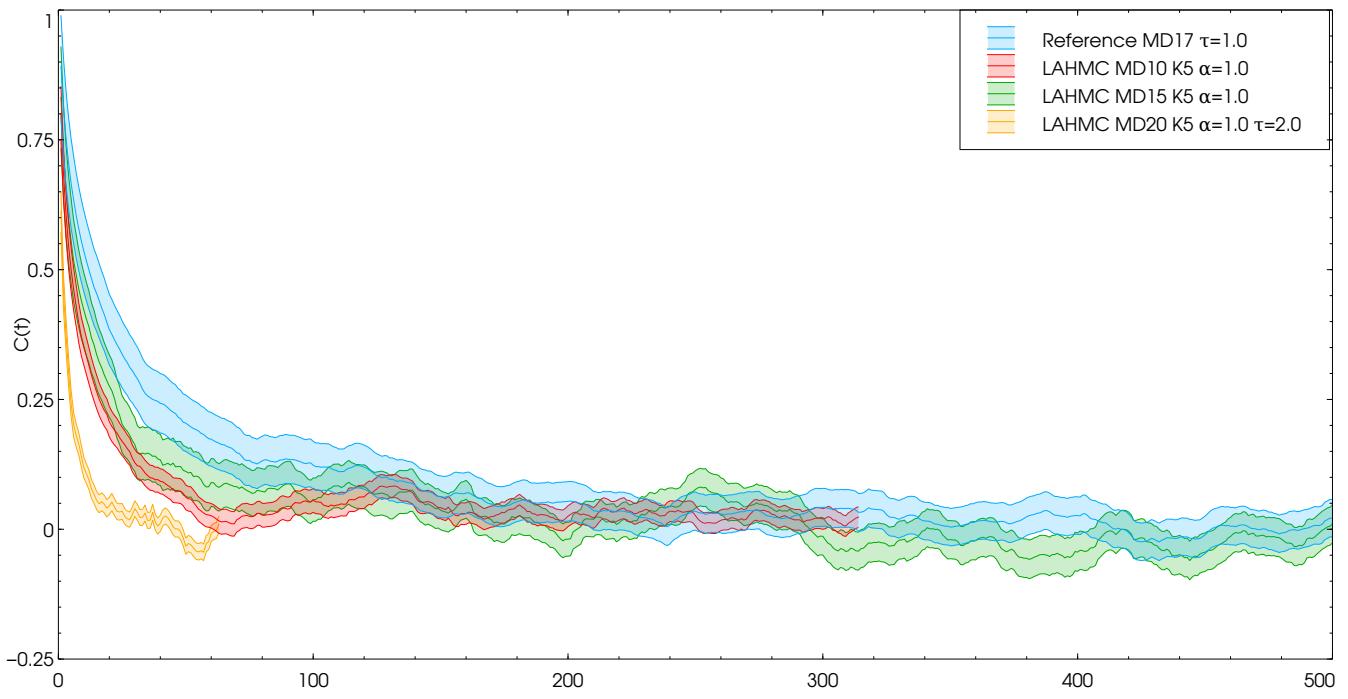


Figure 11: T_0 Autocorrelation functions.

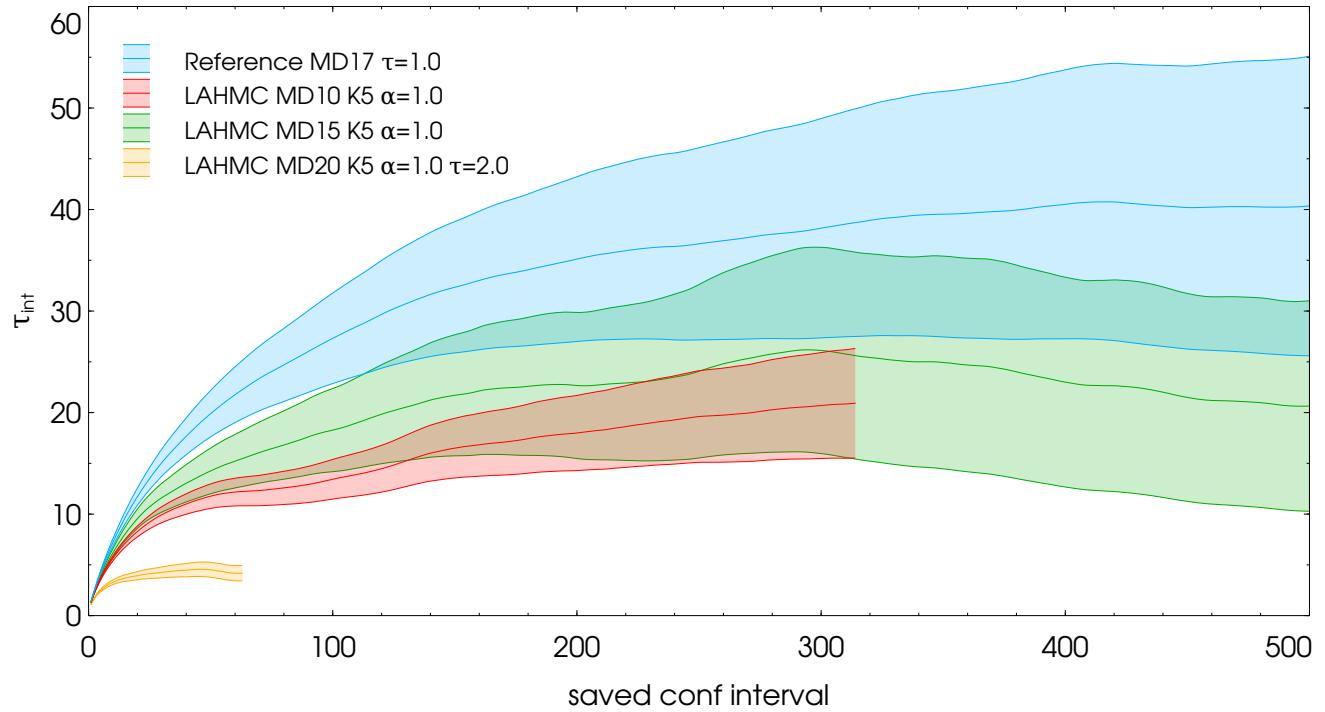


Figure 12: T_0 Autocorrelation times.

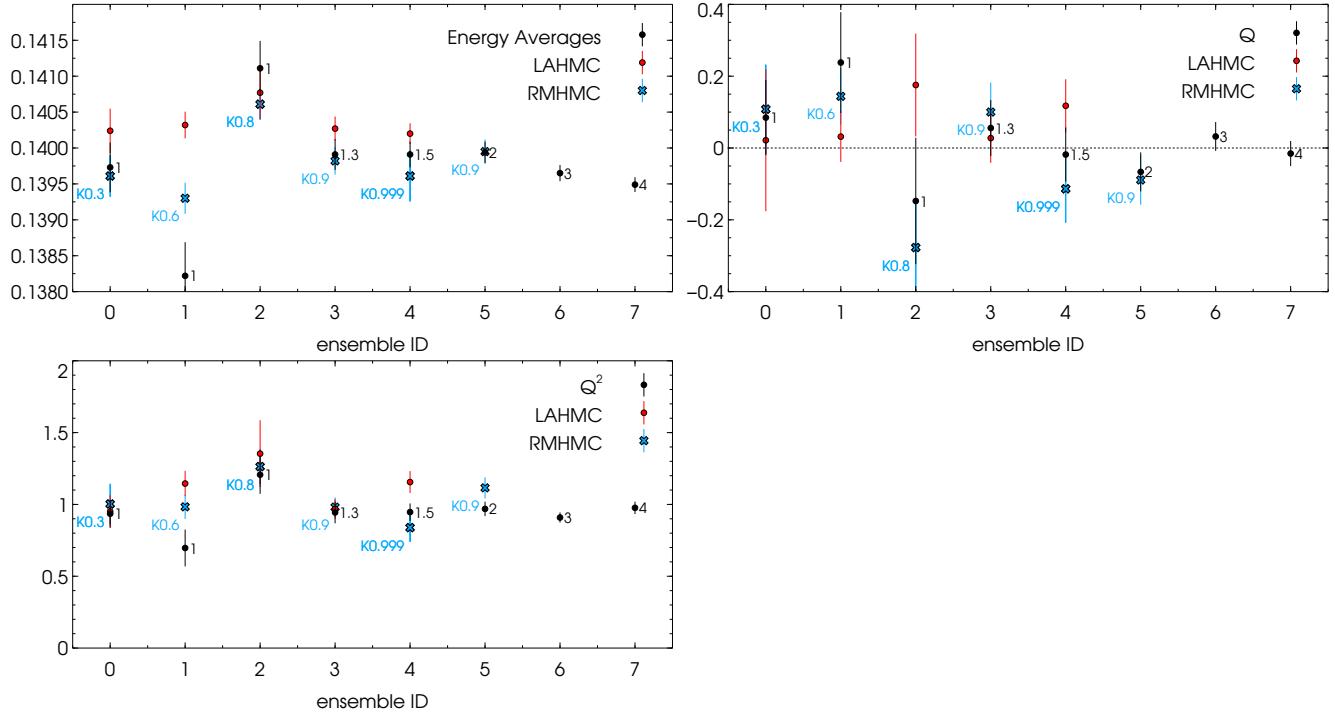


Figure 13: $L = 16$ Averages.

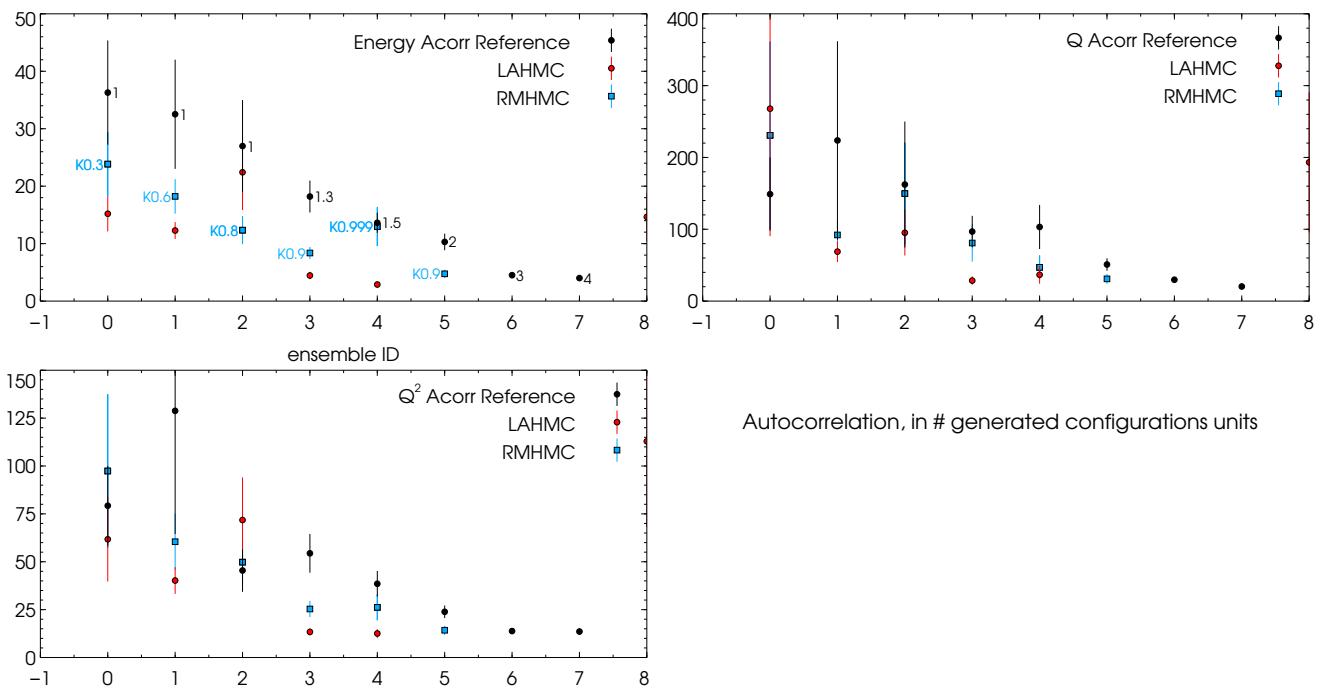


Figure 14: $L = 16$ Autocorrelation times, in number of generated configurations units.

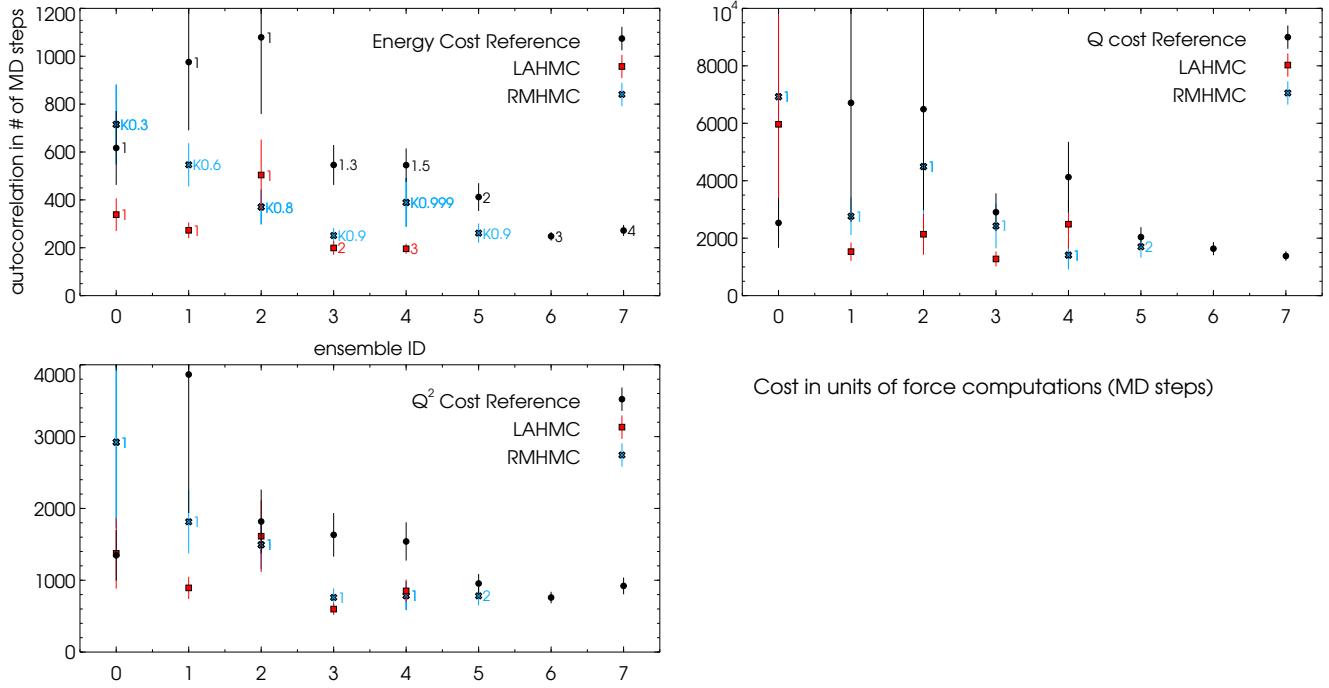


Figure 15: $L = 16$ Cost estimates. The cost C is defined in eq.53

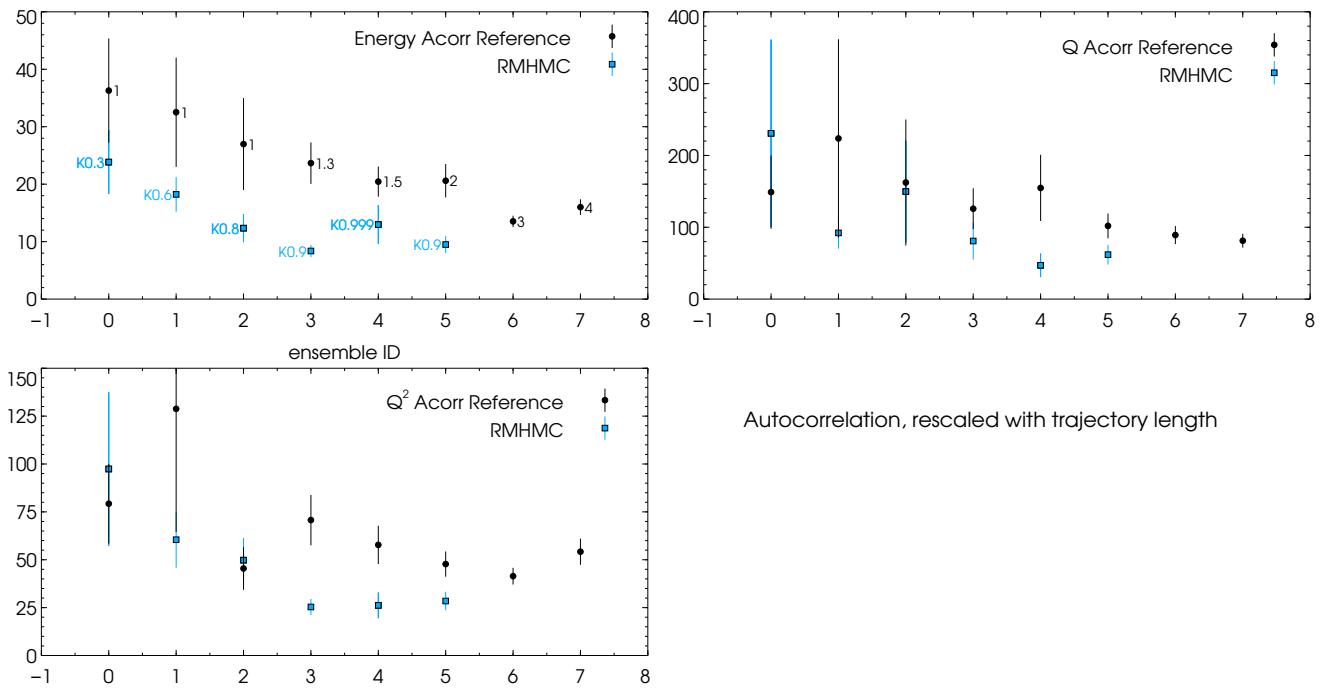


Figure 16: $L = 16$ Autocorrelation times, rescaled by the trajectory length.

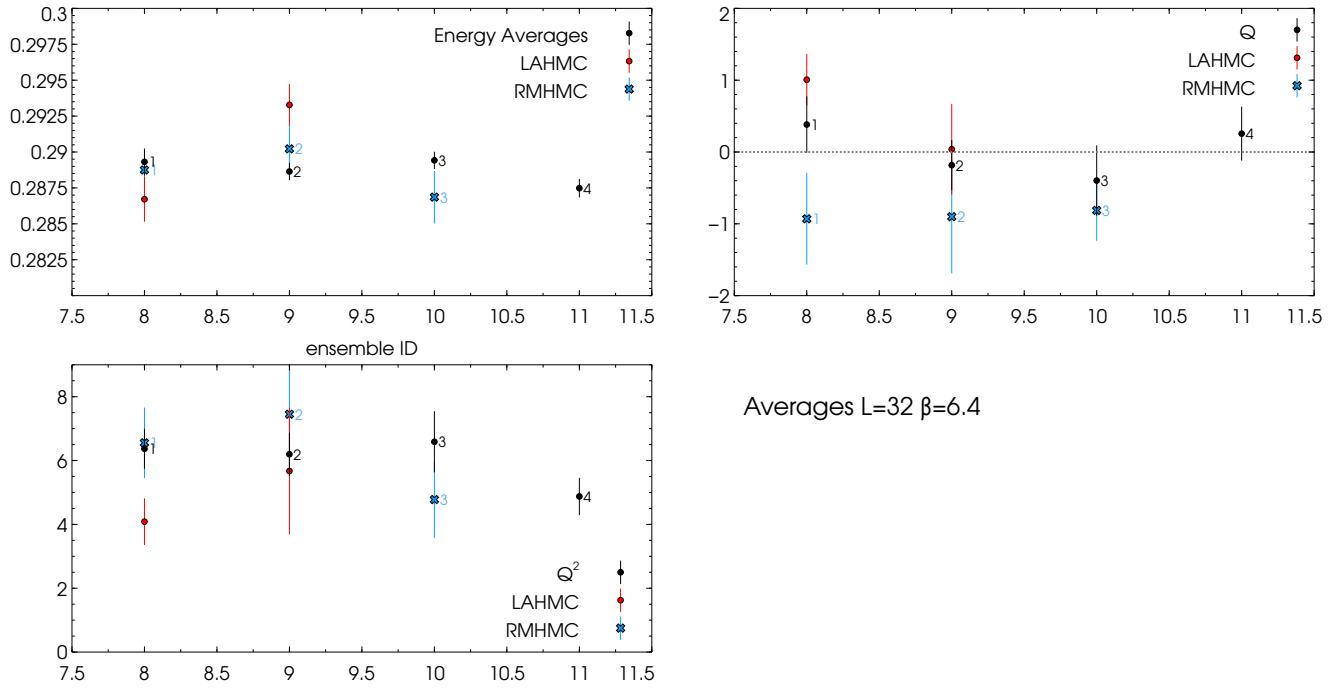


Figure 17: $L = 32$ Averages.

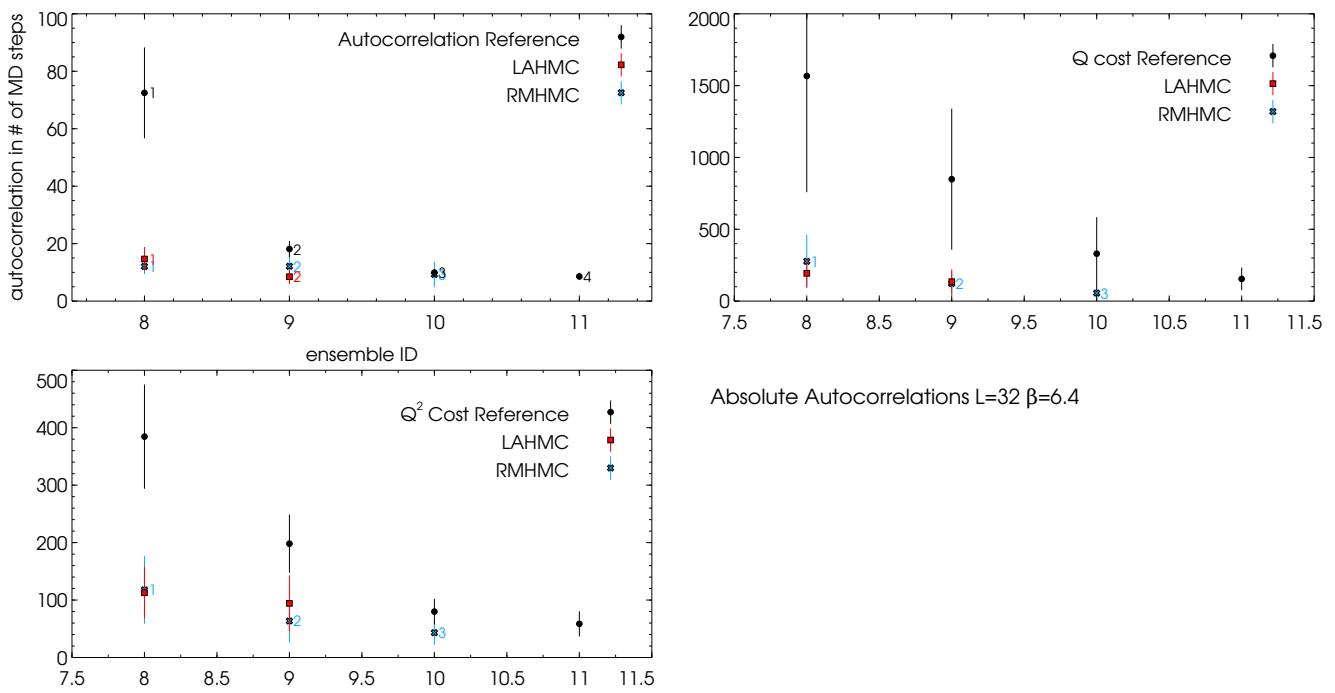


Figure 18: $L = 32$ Autocorrelation times.

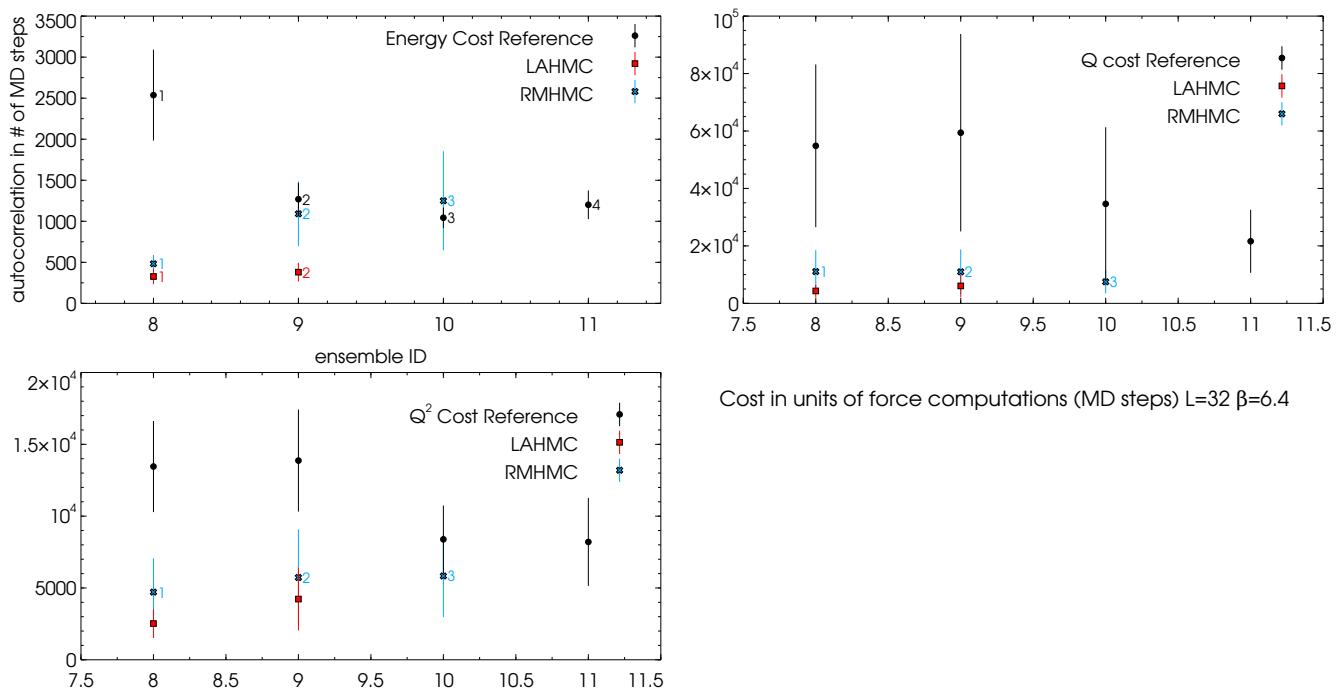


Figure 19: $L = 32$ Cost estimates. The cost C is defined in eq.53

9. Histories for the average energy T_0 . $L = 32$.

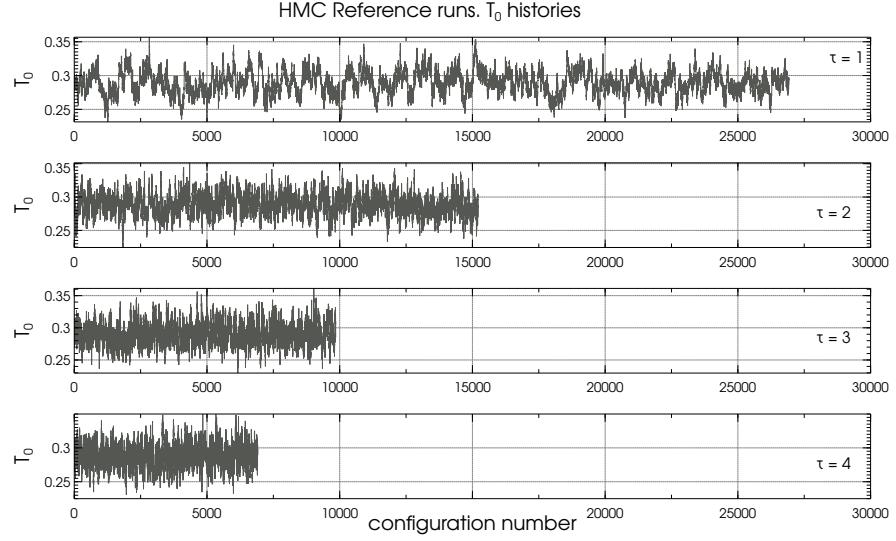


Figure 20: History of T_0 eq. 52, $L = 32$. Standard HMC. Notice that since we switched the smearing radius in the middle of the run we have less data here w.r.t. the total configurations reported in the tables. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

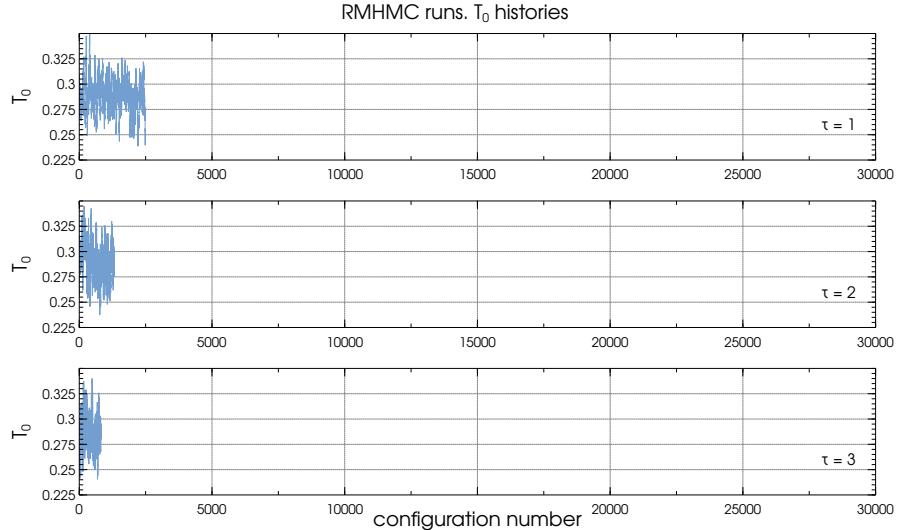


Figure 21: History of T_0 eq. 52, $L = 32$. RMHMC. Notice that since we switched the smearing radius in the middle of the run we have less data here w.r.t. the total configurations reported in the tables. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

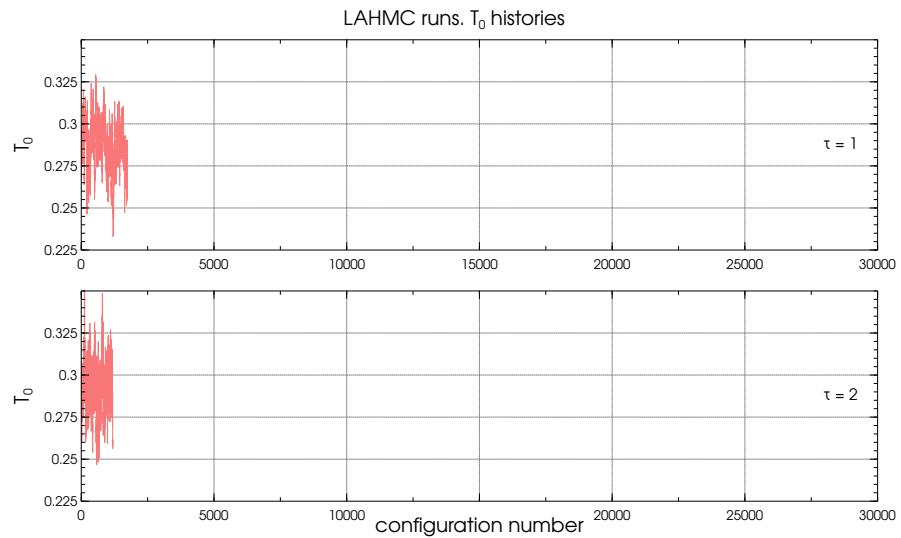


Figure 22: History of T_0 eq. 52, $L = 32$. LAHMC. Notice that since we switched the smearing radius in the middle of the run we have less data here w.r.t. the total configurations reported in the tables. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

10. Histories for the topological charge Q . $L = 32$.

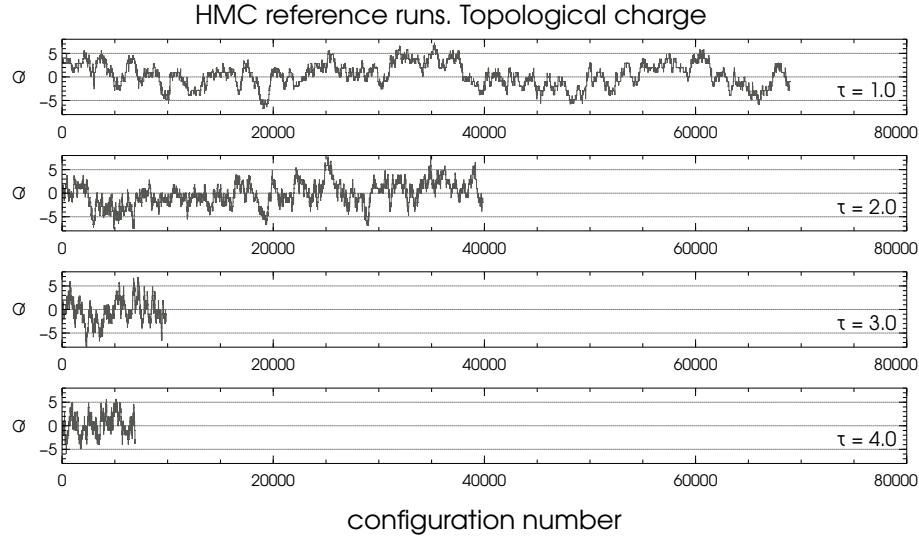


Figure 23: History of Q , $L = 32$. Standard HMC. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

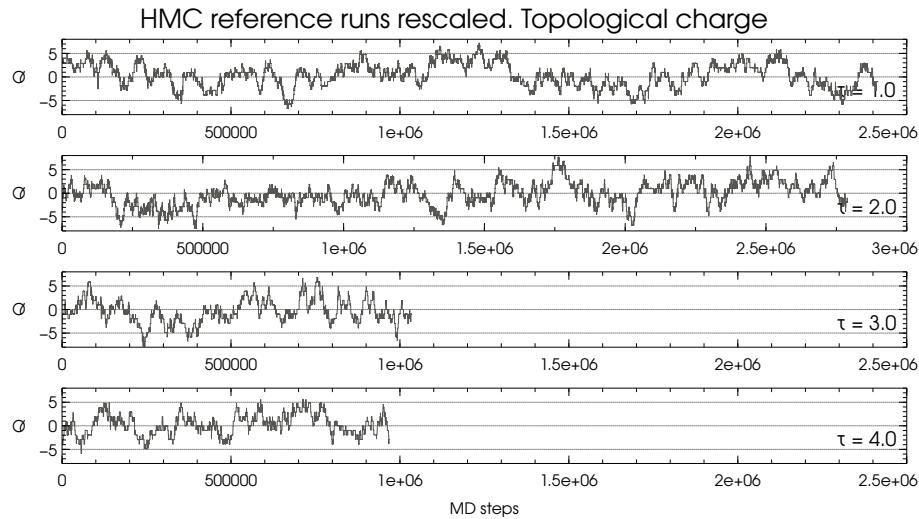


Figure 24: History of Q , $L = 32$. RMHMC. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

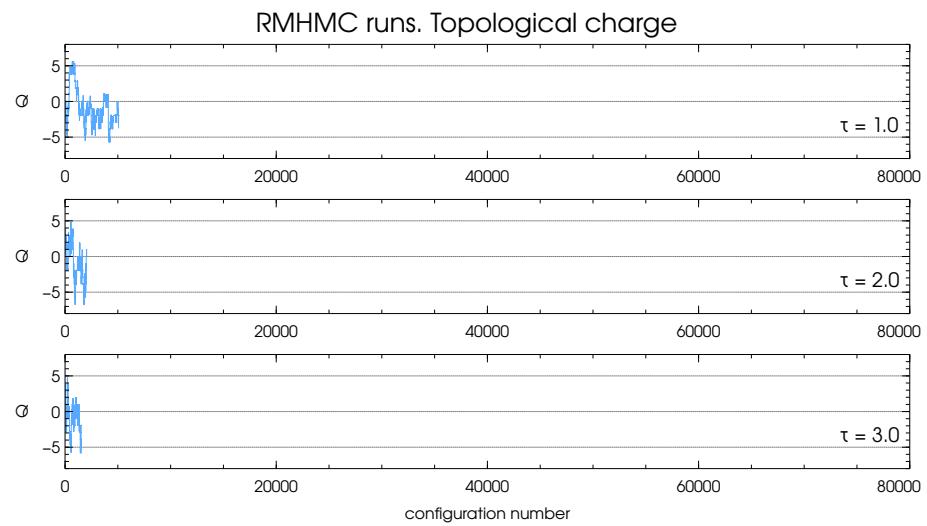


Figure 25: History of Q , $L = 32$. LAHMC. X axis is NOT the trajectory number but a counter for the measurement steps (N.B. I am taking a measurement every 5 trajectories).

11. References

- [1] Jean Zinn-Justin. Renormalization and Stochastic Quantization. *Nucl. Phys.*, B275:135–159, 1986. doi: 10.1016/0550-3213(86)90592-4.
- [2] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett.*, B195:216–222, 1987. doi: 10.1016/0370-2693(87)91197-X.
- [3] Martin Luscher and Stefan Schaefer. Non-renormalizability of the HMC algorithm. *JHEP*, 04:104, 2011. doi: 10.1007/JHEP04(2011)104.
- [4] Martin Luscher and Stefan Schaefer. Lattice QCD without topology barriers. *JHEP*, 07:036, 2011. doi: 10.1007/JHEP07(2011)036.
- [5] Greg McGlynn and Robert D. Mawhinney. Diffusion of topological charge in lattice qcd simulations. *Phys. Rev. D*, 90:074502, Oct 2014. doi: 10.1103/PhysRevD.90.074502. URL <http://link.aps.org/doi/10.1103/PhysRevD.90.074502>.
- [6] Stefan Schaefer, Rainer Sommer, and Francesco Virotta. Critical slowing down and error analysis in lattice QCD simulations. *Nucl. Phys.*, B845:93–119, 2011. doi: 10.1016/j.nuclphysb.2010.11.020.
- [7] Harvey B. Meyer, Hubert Simma, Rainer Sommer, Michele Della Morte, Oliver Witzel, and Ulli Wolff. Exploring the HMC trajectory-length dependence of autocorrelation times in lattice QCD. *Comput. Phys. Commun.*, 176:91–97, 2007. doi: 10.1016/j.cpc.2006.08.002.
- [8] Massimo Campostrini, Paolo Rossi, and Ettore Vicari. Monte Carlo simulation of CP^{**}(N-1) models. *Phys. Rev.*, D46:2647–2662, 1992. doi: 10.1103/PhysRevD.46.2647.
- [9] A. D’Adda, M. Luscher, and P. Di Vecchia. A 1/n Expandable Series of Nonlinear Sigma Models with Instantons. *Nucl. Phys.*, B146:63–76, 1978. doi: 10.1016/0550-3213(78)90432-7.
- [10] Edward Witten. Instantons, the Quark Model, and the 1/n Expansion. *Nucl. Phys.*, B149:285–320, 1979. doi: 10.1016/0550-3213(79)90243-8.
- [11] Luigi Del Debbio, Gian Mario Manca, and Ettore Vicari. Critical slowing down of topological modes. *Phys. Lett.*, B594:315–323, 2004. doi: 10.1016/j.physletb.2004.05.038.
- [12] Simon Duane and Brian J. Pendleton. GAUGE INVARIANT FOURIER ACCELERATION. *Phys. Lett.*, B206:101–106, 1988. doi: 10.1016/0370-2693(88)91270-1.
- [13] Simon Duane, Richard Kenway, Brian J. Pendleton, and Duncan Roweth. Acceleration of Gauge Field Dynamics. *Phys. Lett.*, B176:143, 1986. doi: 10.1016/0370-2693(86)90940-8.
- [14] Mark Girolami and Ben Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2010.00765.x. URL <http://dx.doi.org/10.1111/j.1467-9868.2010.00765.x>.
- [15] Benedict J Leimkuhler and Sebastian Reich. *Simulating Hamiltonian dynamics*. Cambridge monographs on applied and computational mathematics. Cambridge Univ., Cambridge, 2004. URL <https://cds.cern.ch/record/835066>.
- [16] Jascha Sohl-Dickstein, Mayur Mudigonda, and Michael R. DeWeese. Hamiltonian monte carlo without detailed balance. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML’14, pages I–719–I–726. JMLR.org, 2014. URL <http://dl.acm.org/citation.cfm?id=3044805.3044887>.
- [17] C. M. Campos and J. M. Sanz-Serna. Extra Chance Generalized Hybrid Monte Carlo. *Journal of Computational Physics*, 281:365–374, January 2015. doi: 10.1016/j.jcp.2014.09.037.
- [18] Peter A. Boyle, Guido Cossu, Azusa Yamaguchi, and Antonin Portelli. Grid: A next generation data parallel C++ QCD library. *PoS, LATTICE2015:023*, 2016.