Status of dynamical ensemble generation

Chulwoo Jung
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, U.S.A.
E-mail: chulwoo@bnl.gov

I give an overview of current and future plans of dynamical QCD ensemble generation activities. A comparison of simulation cost between different discretizations is made. Recent developments in techniques and algorithms used in QCD dynamical simulations, especially mass reweighting, are also discussed.

The XXVII International Symposium on Lattice Field Theory - LAT2009
July 26-31 2009
Peking University, Beijing, China

*Speaker.
1. Introduction

Recently there has been a remarkable progress in reaching continuum limit via Lattice QCD, made possible by better simulation algorithms and better lattice discretizations to suppress lattice spacing errors in moderate lattice spacings, in addition to steady advances in computing hardware. In many ways, it is more than just a rhetoric to say we are at the cusp of generating realistic lattice QCD configurations.

Since an extensive review of systematic errors, theoretical and practical issues of each lattice discretization schemes was given at the last year’s conference [1], here I will focus more on a survey of ongoing lattice ensemble activities and technical details from different groups presented at the conference and updates since the last year. More comprehensive review of physical quantities from these lattices will be made elsewhere, for example in [2, 3, 4, 5]. Also, a brief and incomplete review of recent progress and trends in dynamical simulation algorithms is given with more details on mass reweighting technique.

It should be noted that while the list of dynamical lattice QCD ensemble generation activities presented here is extensive, it is limited to only zero temperature ($T \geq L$) and "QCD-like" ensemble simulations, with 2 light quarks. There are significant amount of ensemble generation activities for QCD thermodynamics studies and beyond the standard model studies, especially in anticipation of upcoming results from Large Hadron Collider(LHC). These topics are covered in [6] and [7] respectively.

The organization is as follows. Section 2 gives descriptions and summaries of ongoing ensemble generating activities. A comparison of simulation cost for various lattice discretizations is given in section 3. Section 4 summarizes available measurements of autocorrelations, especially topological charge. Summary of advances in simulation algorithms and techniques is given in section 5.

2. Ensembles

2.1 Domain Wall Fermion(DWF)

RBC/UKQCD collaborations have been generating $N_f = 2 + 1$ DWF configurations [8, 9, 10] using DWF formalism from [11] and Iwasaki gauge action[12] which gives a good balance between preservation of chiral symmetry and practicality for the lattice spacings studied so far. Iwasaki action was chosen instead of DBW2 action [13], which was used for previous DWF studies, to ensure sufficient decorrelation in topological charge.

A recent deployment of IBM BG/P machines at Argonne Leadership Computing Facility(ALCF) made the ensembles at the second lattice spacing much earlier than initially anticipated. Currently ensembles in 2 lattice spacings are available:

- $\beta = 2.13, a \sim 0.121 \text{fm}, 16^3 \times 32 \times 16, m_\pi \sim 400, 526, 627 \text{ Mev}$
- $\beta = 2.13, a \sim 0.114 \text{fm}, 24^3 \times 64 \times 16, m_{\text{res}}^{\overline{\text{MS}}} \sim 8.5 \text{ Mev}, m_\pi \sim 328, 417 \text{ Mev}$
- $\beta = 2.25, a \sim 0.084 \text{fm}, 32^4 \times 64 \times 16, m_{\text{res}}^{\overline{\text{MS}}} \sim 2.45\text{MeV}, m_\pi \sim 295, 350, 397 \text{ Mev}$
Status of dynamical ensemble generation  
Chulwoo Jung

Figure 1: Edinburgh plot for dynamical DWF lattices, from [14].

Figure 2: Pseudoscalar decay constant \( f_\pi \) with \( m_{\text{sea}} = m_{\text{val}} \) for 24\(^3\) and 32\(^3\) DWF lattices, from [15].

Figures 1 and 2 Shows the scaling behavior between DWF ensembles with 2 different lattice spacings. The quantities measured within the range of sea quark mass shows the scaling violation is within 2%. This allows fitting both ensembles to NLO SU(2) ChPT in \( a^2 \) and \( m_l \), where only leading order (LO) ChPT low energy constants (LEC’s) have a dependence. Reweighting in dynamical strange quark mass (section 5.1.1) and interpolation in valence mass is used to reach the physical strange quark mass point. Numbers in physical units for the last 2 sets of ensembles are from this SU(2) ChPT global fit using both ensembles. Details of the fitting procedure and the results for pseudoscalar meson and decay constants is reported in [16]. \( B_K \) and detailed scaling study is reported in [15]. Results on hadron masses are in [14].

2.2 DWF with Auxiliary Determinant (Modified Gap DWF)

Lattice dislocations which induce residual chiral symmetry breaking in DWF is currently the biggest obstacle for DWF and overlap fermions at larger lattice spacings. \( (a \geq 0.1 \text{ fm}) \). Hence, controlling residual mass at these lattice spacings is crucial for DWF studies of quantities which requires large lattice volume, such as QCD thermodynamics, nucleon matrix elements, weak matrix elements via direct studies of \( K \rightarrow \pi\pi \) process on the lattice.

Various approaches have been used to suppress dislocations in the past. Different gauge actions such as DBW2 action[13] have shown to suppress dislocations successfully, but at the expense of suppression of topology tunneling at smaller lattice spacings. Alternatively, additional fermion action can be used to suppress dislocations, as they are related to near zero modes of Wilson Dirac operators. This idea was first introduced in [17] and later used in dynamical overlap simulation by JLQCD collaboration[18], where a ratio of Wilson fermion determinants were used instead. It should be noted that while formally fermions are added, these are not related to physical quarks and this is effectively just changing gauge action after they are integrated out.

For the lattice spacings currently being investigated, determinants used in either [17] or [18] appear to suppresses the topology tunnelling too strongly. To circumvent this, a small imaginary
mass is also added to the numerator to ensure the zero eigenvalues are not completely suppressed.

Supression factor from these additional terms is given by

$$\Psi(M_5, \varepsilon_f, \varepsilon_b) = \frac{\det [D_W(-M_5 + i\varepsilon_f \gamma^5)^\dagger D_W(-M_5 + i\varepsilon_f \gamma^5)]}{\det [D_W(-M_5 + i\varepsilon_b \gamma^5)^\dagger D_W(-M_5 + i\varepsilon_b \gamma^5)]}$$

$$= \frac{\det [H_W(-M_5)^\dagger H_W(-M_5) + \varepsilon_b^2]}{\det [H_W(-M_5)^\dagger H_W(-M_5) + \varepsilon_b^2]} = \prod_i \frac{\lambda_i^2 + \varepsilon_f^2}{\lambda_i^2 + \varepsilon_b^2} \quad (2.1)$$

Where $\lambda_i$ are eigenvalues of Hermitian Wilson Dirac operator with mass $-M_5 = \gamma_5 D_W(-M_5)$. Eq. (2.1) gives $\sim 1$ for $\lambda_i \gg \varepsilon_b, \varepsilon_f$, while $\sim \varepsilon_f^2/\varepsilon_b^2$ for $\lambda_i \ll \varepsilon_f, \varepsilon_b$. $\varepsilon_f = 0$ corresponds to what is used in [18] and only numerator with $\varepsilon_f = 0$ was used in [17].

<p>| Aux. Det. $\beta = 1.75, a \sim 1.4$GeV, $\varepsilon_f/\varepsilon_b = 0.02/0.5, am_{res} \sim 0.0019$ |</p>
<table>
<thead>
<tr>
<th>L/a</th>
<th>$m_\pi$</th>
<th>$m_{1/2}$</th>
<th>L(fm)</th>
<th>$m_{PS}$ (Mev)</th>
<th>$\tau$ (MD)</th>
<th>Accept.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$32^3 \times 64 \times 32$</td>
<td>0.045</td>
<td>0.0042</td>
<td>$\sim 4.5$</td>
<td>$\sim 250$</td>
<td>$\sim 1200$</td>
<td>$\sim 70 %$</td>
</tr>
<tr>
<td>$32^3 \times 64 \times 32$</td>
<td>0.045</td>
<td>0.001</td>
<td>$\sim 4.5$</td>
<td>$\sim 200$</td>
<td>$\sim 250$</td>
<td>$\sim 70 %$</td>
</tr>
</tbody>
</table>

Table 1: List of ongoing ensemble generation using DWF with Eq. (2.1).

Table 1 shows the ongoing DWF simulations with the auxiliary determinant. Although it turned out it is necessary to use a rather large $\varepsilon_b$ to make enough suppression of the residual mass, which causes the shifts in gauge coupling $\beta$, a factor of 5-7 decrease in residual mass was observed after the scales are matched by locating transition temperature [19].

Figure 4 shows the topological charge evolution of these ensembles and Wilson Dirac operator eigenvalue flow diagram for an ensemble with same lattice spacing and smaller volume, which suggests the topology is being sampled well while dislocations which causes the small eigenvalues near $-M_5 = 1.8$ are suppressed as intended.

Preliminary results form $m_\pi \sim 250$Mev ensembles suggests that the scaling error between
Status of dynamical ensemble generation

Chulwoo Jung

\( a \sim 0.14 \text{fm} \) AuxDet ensembles and existing DWF ensembles is at a few percent level, and it is possible to fit DWF ensembles with and without the auxiliary determinant in a fashion in [16], by allowing different \( a^2 \) dependence to LO LEC’s in the ChPT fitting. Result of this analysis will be forthcoming.

2.3 \( a^2 \), tadpole improved staggered action (Asqtad)

\[ \frac{a^2}{\text{latt}} \geq 0.4 \] and \( m_{l}/m_{s} \geq 0.1 \), which gives \( m_{\pi} \geq 220 \text{ MeV} \) at multiple lattice spacings, \( a \sim 0.15 \) (usually referred as "extra-coarse"), 0.12 (coarse), 0.09 (superfine), 0.06 (ultrafine) fm. In addition to these, there are ensembles generated at \( m_{s} \) near 60% of physical strange quark mass, to aid SU(3) ChPT studies. Most recent continuum extrapolation of Sommer scale \( r_{1} \), set by \( f_{K} \), gives 0.3117 fm [21]. Typically staggered chiral perturbation theory [23, 24] is used for continuum extrapolations. An update of SU(3) ChPT fitting results are reported in [21].

It should also be noted that MILC ensembles have been extensively used for various mixed action studies, for example [25], where valence quarks with different discretizations are used, due to early availability of ensembles with multiple lattice spacings and quark masses. For these studies, the effects of taste symmetry breaking also has to be accounted for and this is typically done via mixed action chiral perturbation theories such as[26].

2.4 Highly Improved Staggered Quarks (HISQ)

HISQ action [27] is a staggered fermion action, improved further from Asqtad action by intro-
Status of dynamical ensemble generation

Chulwoo Jung

Figure 7: Taste splitting of Asqtad and HISQ action, measured by the difference of squared pseudoscalar meson masses with different tastes[28]. $M_I$ and $M_G$ is the mass of taste singlet and Goldstone pseudoscalar respectively.

Figure 8: Scaling plot of nucleon mass for Asqtad and HISQ actions[28].

Reducing additional Fat-7 smearing followed by projection to U(3) using Cayley-Hamilton theorem, used also in [29, 30], before combined in a similar fashion to Asqtad action.

Preliminary studies of dynamical ensembles generated with HISQ action indicates that while it is about 2 times more expensive per MD units compared to those of Asqtad actions at the similar lattice spacing, the mass splitting between pseudoscalar mesons with different tastes are reduced by a factor of $2.5 \sim 3$ (Fig. 7).

While the U(3) projection after additional smearing have shown to be effective in suppressing taste symmetry breakings, this also causes the fermion force for MD steps to be large when the smeared link happens to have a small eigenvalue, resulting in low acceptance. These difficulty is avoided by replacing $Q^{-1/2}$ with $(Q + \delta I)^{-1/2} (\delta \sim 5 \times 10^{-5})$ during molecular dynamics (MD) steps, where $Q$ is a smeared link to be projected to U(3). While this effectively changes the hamiltonian for MD step, Metropolis step after each trajectory corrects for the discrepancy. Also, dynamical charm quark can be included with $O((am_c)^4)$ errors removed, but the coefficient for a straight 3-link term (Naik term) should be set different for different quarks.

Preliminary studies with HISQ actions for both thermodynamic studies ($N_f = 2 + 1$) [31] and zero temperature studies with dynamical charm quark ($N_f = 2 + 1 + 1$)[28] are reported at the conference.

2.5 $O(a)$ improved Wilson (Clover) action

There are ensemble generation activities by several collaborations using Clover actions: Budapest-Marseille-Wuppertal(BMW) collaboration, PACS-CS collaboration, QCDSF collaboration and Coordinated Lattice Simulations(CLS) collaboration. All activities are aimed at generating en-
sembles with the pion mass close or at the physical value. The main features of simulations are summarized in table 2.

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>BMW</th>
<th>PACS-CS</th>
<th>QCDSF</th>
<th>CLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_f$</td>
<td>2+1</td>
<td>2+1</td>
<td>2</td>
<td>2+1(SLiNC)</td>
</tr>
<tr>
<td>Gauge action</td>
<td>Tree Symanzik</td>
<td>Iwasaki</td>
<td>Wilson</td>
<td>Tree Symanzik</td>
</tr>
<tr>
<td>$c_{sw}$</td>
<td>Tree level</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>Smearing</td>
<td>Stout</td>
<td>Stout</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm</td>
<td>(R)HMC</td>
<td>DDHMC +PHMC</td>
<td>(R)HMC</td>
<td>DDHMC +deflation</td>
</tr>
<tr>
<td>$a$(fm)</td>
<td>0.065-0.125</td>
<td>0.09</td>
<td>0.072-0.09</td>
<td>~0.08</td>
</tr>
<tr>
<td>$m_\pi$(Mev)</td>
<td>&gt;~ 190</td>
<td>156-702</td>
<td>140 - 1010</td>
<td>360 - 520</td>
</tr>
<tr>
<td>This conference</td>
<td>[32]</td>
<td>[33]</td>
<td>[34]</td>
<td>[35]</td>
</tr>
</tbody>
</table>

Table 2: Summary of dynamical Clover ensemble generation activities.

While nonperturbatively determined $c_{sw}$ is used for most of simulations, ensembles generated by BMW collaboration uses tree level value, but instead rely on up to six levels of stout smearing[29] to suppress lattice discretization errors. The main results from these ensembles were reported in[37, 38, 39] and update of $f_K/f_\pi$ is given in [32]. PACS-CS[40, 41] has applied mass reweighting technique(section 5.1) to tune both light and strange dynamical masses to reach physical point for figure 9. Currently available configurations have a physical size $L \sim 3$ fm, $m_\pi L \sim 2.2$. Simulation with $L \sim 6$ fm is under way.

QCDSF collaboration has generated extensive $N_f = 2$ ensembles and the results for hadron masses from those configurations are given in [34]. They are also working on $N_f = 2+1$ ensembles using stout smearing and nonperturbative $c_{sw}$, called SLiNC action. A preliminary study for tuning quark masses to physical point is described in [35]. $r_0 = 0.5$ fm is used to set the scale.

Figure 9: Deviation of hadron masses from physical values, normalized by $m_\Omega$, from [33].
Figure 10: Nucleon mass versus $m_\pi^2$ from QCDSF $N_f = 2$ ensembles. $M_N \sim m_\pi^2$ behavior is observed. $r_0 = 0.467$ fm is used.
CLS collaboration\cite{42} aims to generate Clover ensembles in a wide range of lattice spacings, quark masses and lattice volumes using deflation accelerated DD-HMC\cite{43}. Preliminary results are reported in \cite{36}.

### 2.6 Twisted mass Wilson (tmWilson)

European Twisted Mass Collaboration (ETMC) has been generating both $N_f = 2$ and $N_f = 2 + 1 + 1$ dynamical tmWilson ensembles \cite{44, 45} with maximal twist. $N_f = 2$ ensembles are generated with tree-level Symanzik gauge action, $0.053 \text{ fm} < a < 0.1 \text{ fm}$, $280 \text{ Mev} < m_\pi < 650 \text{ Mev}$ and $2.0 \text{ fm} < L < 2.5 \text{ fm}$.

![Figure 11: $(m^2_2) - (m^0_2)$ vs. $a^2$ for tmWilson $N_f = 2$ ensembles.](image1)

![Figure 12: $f_\pi$ vs $m_\pi$ for tmWilson $N_f = 2$ and 2+1+1 ensembles\cite{46}.](image2)

A more detailed scaling study on $N_f = 2$ ensembles shows a good scaling between ensembles with 4 different lattice spacings for many quantities. An exception to this is the mass splitting between charged and neutral pseudoscalar mesons from the breaking of isospin symmetry at nonzero lattice spacing. Figure 12 shows while the behavior is consistent with $O(a^2)$ effect, it still is significant in the lattice spacings studied. This may be understood via a Symanzik-type analysis\cite{47}. More details of $N_f = 2$ ensembles is reported in \cite{48}.

$N_f = 2 + 1 + 1$ ensembles are generated with Iwasaki gauge action, at $a \sim 0.078, 0.086 \text{ fm}$, $280 \text{ Mev}, < m_\pi < 500 \text{ Mev}, L \sim 2.8 \text{ fm}$. Polynomial HMC (PHMC)\cite{49} is used for the fermion simulation. Twist parameters for heavy quarks, $\mu_\sigma$ and $\mu_\delta$, are fixed by tuning to physical values of $K$ and $D$ mass while keeping $m_{PCAC}$, close to zero. This is done without calculating all the possible excited states such as $K + n \times \pi$ between $K$ and $D$ meson masses by assuming some properties about our trial states and the corresponding correlation matrices. it turned out $\mu_\delta$ was slightly underestimated for some of the ensembles. This is planned to be adjusted by reweighting\cite{50}. More details of $N_f = 2 + 1 + 1$ ensembles is reported in \cite{46}.
In addition to this, $N_f = 4$ ensemble is being generated for nonperturbative renormalization with RI-MOM scheme\[51\].

### 2.7 Anisotropic Clover

Hadron Spectrum Collaboration(HSC) has been generating $N_f = 2 + 1$ anisotropic clover configurations at $a_s = 0.125$ fm, $\xi = a_s/a_t = 3.5, a_t \sim 5.6$ Gev, $L = 3 \sim 4$ fm, $m_\pi = 230, 360$ Mev to study various quantities such as resonance spectroscopy and nuclear forces. Stout smearing is used on spatial links only to suppress discretization errors further while preserving locality of transfer matrix. Strange quark is tuned to keep $s_\Omega = 9(2m_K^2 - m_\pi^2)/(2m_\Omega)^2$ constant while approaching continuum, as shown in figure 13. More details of algorithm and mass tuning are given in [52, 53]. Results using these ensembles on excited nucleon spectroscopy [54], $\pi\pi$ states using distilled quark propagators [55], multi-hadron systems[56] and cascade baryons[57] are reported at the conference.

![Figure 13: $s_\Omega = 9(2m_K^2 - m_\pi^2)/(2m_\Omega)^2$ vs. $l_\Omega(3m_\pi/3m_\Omega)^2$ (Newport News plot).](image)

![Figure 14: Preliminary results for resonance spectrum from $N_f = 2 + 1, m_\pi = 380$ Mev, $16^3 \times 128$ anisotropic clover ensembles[54].](image)

### 2.8 Overlap and Optimal DWF

JLQCD and TWQCD collaboration has been generating $N_f = 2$ [18] and $N_f = 2 + 1$ overlap ensembles with $a \sim 0.1 - 0.12$fm, $L \leq 2$ fm. Recent results are covered separately in [58, 59].

Also, TWQCD collaboration has started generating $N_f = 2$ and $2+1$ ensembles using optimal domain wall fermion[60] with $a^{-1} > \sim 1.6$ Gev, $L = 2 \sim 3$ fm. These simulations are performed with CUDA-written codes on Nvidia GPGPU (Nvidia Tesla S1070, 46 Nvidia GTX285/280). More details of TWQCD activities can be found in [61].

### 2.9 Chirally Improved fermion(CI)

Bern-Graz-Regensburg(BGR) collaboration has started generating $N_f = 2$ configurations using CI action[62], a 4 dimensional action with links of length up to 4, tuned to satisfy Ginsparg-
Wilson relation approximately. Stout smearing is used to suppress discretization error further along with Lüscher-Weisz Gauge action. Figure 15 shows collection of links connected to the point at the center in CI action. Currently generated ensembles have \( N_f = 2, 16^3 \times 32, \sim 2.4 \text{ fm}, 318 < m_\pi < 526 \text{ MeV}, m_{AW} = 15 \sim 42 \text{ MeV}. \) This gives a significantly different action to be used for studies of excited states to anisotropic Clover in section 2.7. Preliminary results for excited hadron spectrum is reported in [63].

3. Simulation cost

As listed in the section above, there are many ongoing dynamical ensemble generation activities with different discretizations. This is not just because of increasing availability of computing resources, but rather the reflection of the fact that different actions have different lattice spacing error for different quantities at same lattice spacings, and the recent progress in ensemble generation algorithms makes ensemble generation relatively cheaper compared to the valence propagator generation, either with the same or different discretization, and relying on one discretization method to get every physical quantities of interest may not turn out to be the most cost-efficient approach. This also suggest a comparison of different actions at the same lattice spacing is not necessarily a fair one.

Having said that, it is still useful to have some idea on how expensive each action is at a given lattice spacing and quark mass. Here such a comparison is made by estimating the number of flops to generate the same number of MD units for each action from available performance data. Recent studies indicate it is necessary to keep physical volumes to be larger than previously deemed sufficient to control finite size effect, so for a given lattice spacing \( a \) and pseudoscalar meson mass \( m_\pi \) for each existing ensembles, the number of flops are scaled to a volume \( L^3 \times T \) which satisfies

\[
L = \text{Max}(4/m_\pi, \ 2.5 \text{ fm}), \quad T = 2L.
\]
Figure 16: Total number of TFlop-years needed for $10^4$ MD units versus $m_\pi$(Mev) at physical volume $L^3 \times T$ satisfying eq. (3.1), scaled from numbers for existing ensembles by (cost) $\sim V^{5/4}$. 2 separate lines with same colors for Asqtad and HISQ ensembles represent the lightest(Goldstone) and the heaviest(taste singlet) pseudoscalar masses. Numbers for Asqtad and HISQ simulations are calculated from eq. (3.3). tmWilson(0.08) is from ETMC[64]. Clover(0.09) is from PACS-CS[40]. Clover(0.00) is from CLS[68].

First, available cost formulas in TFlop-years to generate 10000 MD units or 100 statistically independent configurations are given below. here $L, T, a$ are in fm and $m_\pi, m_K, m_l, m_s$ are in Mev.

\[
\text{Cost}[DWF][65] \sim \left( \frac{L^3 \times T}{a^4} \right)^{5/4} \left( \frac{1}{m_\pi a} \right) \left[ C_0 + C_1 \frac{1}{m_K a} + C_2 \left( \frac{a}{m_\pi} \right)^2 \right] \tag{3.2}
\]

\[
\text{Cost}[Asqtad][66] \sim \frac{2.9}{64^3 \times 144} \left( \frac{L^3 \times T}{a^4} \right)^{5/4} \frac{0.1}{m_l/m_s} \frac{0.06}{a} \left( 0.5 + 0.5 \frac{0.1}{m_l/m_s} \frac{0.06}{a} \right) \tag{3.3}
\]
Status of dynamical ensemble generation

Chulwoo Jung

Cost[HISQ] \sim 2 \times Cost[Asqtad]

\begin{equation}
\text{Cost}[\text{Clover}(\text{DD} - \text{HMC})][67] \sim 0.05 \times \left( \frac{20}{\bar{m}_q} \right) \left( \frac{L}{3} \right)^5 \left( \frac{0.1}{a} \right)^6
\end{equation}

(3.4)

While these formulas show different dependencies in $a$ or $m_q(m_{\pi})$, they all have the same dependence in volume, (cost) $\sim V^{5/4}$, which is all we need for the comparison outlined above.

Figure 16 shows the number of total TFlop-years needed to generate 10000 MD units of gauge configurations with the lattice spacings and pseudoscalar masses of existing ensembles, with the volume scaled to satisfy eq.(3.1). While it should be noted that there are differences in precisely how flops are counted for each ensemble which results in some uncertainties in comparing different ensembles. However, it does appear that the difference in total cost between different actions tends to get smaller as the quark mass approaches the physical point. This is presumably because the low, physical eigenmodes of Dirac operators increasingly dominate the cost near the chiral limit, while simply the number of degrees of freedom per sites counts more for simulations at heavier masses.

Still, the cost is a rapidly changing function of the lattice spacing and any real cost comparison should be done not only with these numbers, but in combination with how small lattice spacing error is required for the quantities one aims to study with each action.

4. Autocorrelation

For obvious reasons, autocorrelations within ensembles is a significant factor in what the real cost of generating a number of independent configurations is. Unfortunately, in practice this is not necessarily something one can measure easily, even after ensemble generation is nominally finished, as it can and does vary greatly between different quantities.

While all the ensembles reported here have reasonable ($< 100$ MD units) autocorrelation time for quantities such as plaquette, meson propagators and number of CG iterations, many of them only a few MD units, a significant slowdown of change in global topological charge was observed in ensembles with relatively smaller lattice spacings. Time evolution of topological charge in both asqtad and DWF ensembles are shown in Figures 17 and 18. A similar behavior was also observed in clover action simulations with DD-HMC[69]. While the global topological charge itself may or may not be relevant in physics as long as the physical volume is large enough, it still is a sign of failures of evolution algorithms currently in use to ensure ergodicity of each ensemble.

While these results are still preliminary and requires more careful studies, if not longer ensembles, it would be crucial to check this for any ensembles, especially ensembles with smaller lattice spacings, since these ensembles are designed to have systematic errors small enough for precision studies, and long autocorrelations affect the cost of generating independent configurations to control the statistical errors significantly, or introduce systematic errors from failing to sample all the topological sectors.

This also presents a potential problem in relying on staying on the same lattice discretization and increasingly smaller lattice spacing, made possible by increasing computing resources,
to control the discretization error. This might make it preferable to make better use of increasing computing resources by keeping the lattice spacing moderate and try more improved actions with smaller lattice spacing error, instead of continuing more conventional actions at smaller lattice spacing. Unfortunately, currently this approach - comparing different discretizations - is hampered substantially by the lack of a quantity easily measurable with known and accurate physical value to fix the lattice spacing, as shown by different values of $r_{0.1}$'s used or measured for different discretizations in section 2.

This is not to say this problem cannot be circumvented by algorithmic development. The fact that the increase in autocorrelation is largely independent of the quark mass, choice of discretization and simulation algorithms[69] suggests the gauge action is the culprit, and techniques for more aggressive decorrelation while keeping the action the same, such as Noisy Monte Carlo[70], or even drastically different ideas such as [71], could turn out to be effective.

5. Algorithms and techniques

In the last several years, there have been much progress in fermion simulation algorithm.
Figure 18: Time history of topological charge for DWF ensembles. The ensembles are as follows from the top: \( (a \sim 0.11 \text{fm} : am_l = 0.005, 0.01), (a \sim 0.08 : am_l = 0.004, 0.006, 0.008) \).

Improvements such as Rational Hybrid Monte Carlo\cite{73}, higher order integrators, particularly Omelyan\cite{74, 75}, multiscale integrators and mass preconditioning\cite{76} achieved almost ubiquitous usage, similar to what \( \Phi \) and R algorithm\cite{77} used to be. While these advances made details of fermion action unique for each simulation and made extensive tuning necessary, they also made, in combination with advances in hardware, the dynamical simulations with pion masses at or close physical value and made the cost of ensemble generation much less dominating compared to valence propagator generation needed for various analyses. This is certainly not to say we have exhausted the possibility of improvement. Algorithmic development such as force gradient integrator and better tuning via shadow hamiltonian\cite{78, 79} suggests further reduction in cost of realistic QCD ensemble generation is possible.

Here some of more recent advances in algorithms are summarized, starting with mass reweighting technique, which has shown to be surprisingly useful.

5.1 Reweighting

The basic Idea of reweighting is well known. There are more than one Hamiltonians used in typical Monte Carlo simulations:

- Guiding Hamiltonian for Molecular Dynamics:
  \[ dU/dt_{MD} = -iHU, \quad d\mathcal{H}_l/dt_{MD} = 0 \] (As pointed out, for example in \cite{78}, MD integrators
are not exact. Here $H_1$ is the hamiltonian used to construct MD integrator, not the one actually being preserved by it.

- **Hamiltonian for Metropolis step:**
  
  $$\text{Acceptance} = \min\{1, \exp\left[-(H_2(U_{i+1}) - H_2(U_i))\right]\}$$

- **Hamiltonian for ensemble averaging**
  
  $$\langle O \rangle_{H} = \frac{\int [dU] O(U) W(U)}{\int [dU] W(U)}, \quad W(U) = \exp\left[-(H_3(U) - H_2(U))\right]$$

Typically it is called "reweighting" when $H_2$ is intentionally set different from $H_3$. In principle $H_1, H_2, H_3$ can be all different and there are working examples, ranging from subtle ones such as using less precision and/or relaxed stopping condition for Hybrid Monte Carlo, to more explicit ones used to avoid singularities in HISQ action (section 2.4) where $H_1 \neq H_2$, to many thermodynamic studies, to locate the phase transition temperature[80] or circumvent difficulties with finite density[81] where $H_2 \neq H_3$.

However, $H_{1,2,3}$ are extensive quantities ($H \propto V$) and conventional wisdom has been that it is hard to find $H$’s which the acceptance/reweighting factor $e^{-\Delta H}$ is close enough to 1 while it is different enough to give significant benefit for simulations with larger volumes. (For example, $32^3$ DWF simulations reported in section 2.1 has $H$ on the order of $10^8$.)

However, recently some more working examples have been reported, such as reweighting of light quark [82, 83] and strang quark [84, 33] toward the physical point. While more aggressive reweighting such as those for light quark have potential to be more beneficial in the long run, the reweighting of strange quark mass in particular has proved to be a cost effective way of eliminating one of the major sources of systematic error and provide an immediate benefit. A more detailed description of strange quark reweighting is given below.

**5.1.1 Reweighting of dynamical strange quark**

Due to the nonperturbative nature of QCD, lattice spacing of any lattice QCD simulation, with or without fermions, is not known a priori until it is measured on thermalized configurations.

Somewhat ironically, this has not posed problem for light quarks in practice as much, as typically multiple ensembles of different light quark masses are generated to do extrapolations to the chiral limit anyways, recent simulations near or at physical point notwithstanding. However, it is in principle possible and would be quite beneficial to simulate at the correct physical strange quark. Unfortunately, this is not the case and the dynamical strange quark is often different from physical value by up to 20%. Traditionally one of these approaches has been taken to address this problem:

- Generate multiple ensembles with different dynamical strange quark masses near physical value and use interpolation
- Using SU(3) ChPT to fit up to the strange quark.
- Do multiple parameter tuning runs in smaller volume before larger runs.
- Include the effect of discrepancy as a systematic error.
None of these approaches are particularly attractive and can be time consuming. It can save significant computing (and human) resources if this can be avoided.

It turned out the reweighting factor $W(U)$ (Eq. (5.1)) is often close to 1 even when strange quark mass is $10 \sim 20\%$ different from the dynamical value, which makes the tuning of strange quark to the physical value via reweighting possible. This technique can also be used to calculate derivative with respect to the dynamical quark mass [85].

Reweighting factor for 1 flavor can be calculated as follows:

$$W([U], m_s, m'_s) = \det \left( \frac{D^\dagger D'}{D D^\dagger} \right)^{1/2} = \det(\Omega)^{-1/2}, \Omega([U], m_s, m'_s) = D^{-1} DD^\dagger (D^\dagger)^{-1}$$

$$D = D([U], m_i, m_s), D' = ([U], m_i, m'_s)$$

$$W([U], m_s, m'_s) = \int d\xi d\xi'^\dagger \ e^{-\xi'^\dagger \sqrt{\Omega([U], m_s, m'_s)} \xi} = \left< e^{-\xi'^\dagger (\sqrt{\Omega([U], m_s, m'_s)} - 1) \xi} \right>$$  \hspace{1cm} (5.2)

where $\xi$ is a Gaussian random vector. $\sqrt{\Omega[U]}$ can be calculated using Rational approximation, similar to RHMC. Alternatively, the same reweighting factor can be calculated by

$$W([U], m_s, m'_s) = \sqrt{\left< e^{-\xi'^\dagger (\Omega[U] - 1)} \xi \right>}$$  \hspace{1cm} (5.3)

and it could be more efficient, as (5.3) can be calculated with just one inversion per random number. However, (5.3) is in principle a biased estimator and more accuracy than (5.2) is required for reweighting factors for each configurations to ensure the effect from inaccurate reweighting factor is under control, while (5.2) can be used without such requirement. Another scheme to use Taylor expansion to calculate square root is used in [33].

Due to the nature of (5.2) where each evaluation using random vector is exponentiated before average, it is essential to ensure $M = \sqrt{\Omega([U], m_s, m'_s)}$ has only eigenvalues close to 1. In fact, it can be shown that the error in (5.2) diverges if $M$ has eigenvalues smaller than 1/2. Even when this is not the case, eigenvalues significantly away from 1 can make the reweighting factor converge very slowly with finite statistics. For mass reweighting, determinant breakup by intermediate masses [82]

$$W([U], m_s, m'_s) = \Pi_{i=0}^{n-1} \left< e^{-\xi'^\dagger (\sqrt{\Omega([U], m_i, m_{i+1}) - 1}) \xi} \right>$$

$m_s = m_0 < m_1 < \cdots < m_n = m'_s$ or vice versa.

gives an additional benefit that the reweighting factors for the masses between $m_s$ and $m'_s$ is automatically available. For other reweighting where this is not available, for example [86], Breakup using rational approximation, similar to "root-N trick" in [73] is applicable.

Now, observables for $m'_s$ is calculated from ensemble $[U_i]$ generated $m_s$ by

$$\left< O \right> (m'_s) = \frac{\sum_i O[U_i] W([U_i], m_s, m'_s)}{\sum_i W[U_i]}$$  \hspace{1cm} (5.4)

Figure 19 shows reweighted pseudoscalar decay constant of DWF ensemble in section 2.1.
Figure 19: Reweighted $f_\pi$ and $f_K$ for $N_f = 2 + 1$ DWF ensembles, from[16]. Broken vertical lines represent the physical strange quark mass and the rightmost points shows the original sea strange quark mass for each ensemble.

Even though only 4 random vectors per mass step is used to calculate reweighting factor for $a^{-1} \sim 2.33$ Gev, $am_l = 0.004$ ensemble, the reweighted values at $\sim 20\%$ smaller than $m_s$ has only slightly larger error than the error at $m_s$. As shown in section 2.5, PACS-CS collaborations has done both strange quark and light quark reweighting to reach physical point[33] and tuning of twist for the strange/charm quarks via reweighting is also being carried out by ETMC collaboration[50].

Given the success of mass reweighting, we can also think of other cases where reweighting can be helpful.

- $\mathcal{H}_1 \neq \mathcal{H}_2$:
  - Using a relaxed approximation sign function during MD step of Overlap/Domain wall fermions, to avoid topology tunneling difficulties while preserving good chiral symmetry, this appears to be promising if the phase space where $\mathcal{H}_1 \neq \mathcal{H}_2 \geq 1$ is limited. One approach using DWF, namely reweighting to larger $L_s$ in DWF was explored in [86].

- $\mathcal{H}_2 \neq \mathcal{H}_3$:
  - Apply reweighing to part or all of differences in action in mixed action studies. This would effectively trade systematic error with a statistical one.
  - Adding new terms to the action via reweighting, such as adding charm quarks to $N_f = 2 + 1$ configurations or adding QED for studies of electromagnetic effects[87].

5.2 Nested/mixed precision solvers

Advent of many multicore chips, available currently or in the near future, present in principle significantly more computing power than currently available. However, this increase in computing power often comes without corresponding increase in bandwidth, both in memory and network. While it is not yet clearly shown that these architecture can achieve scalability suitable for large scale dynamical ensemble generations using conventional algorithms such as (R) HMC, or newer algorithms such as DD-HMC, it still is expected to provide competitive hardware platforms for
propagator generation, where “embarrassingly parallel” - the local volume is chosen to be optimal for performance and run multiple jobs concurrently - approach is applicable.

The relative lack of memory bandwidth also pose problems for propagator generation even when it is possible to fit it on one node to circumvent the network bottleneck. One way to lessen the pressure on memory bandwidth is to use less precision arithmetic. While the relative difference in speed between single and double precision is significant for many existing platforms, until recently most frequently used approaches to take advantage of this was somewhat simple-minded, such as first solving in single precision and finish with double(less effective if small residual is required), MD in lower precision, Metropolis in double precision (if $\Delta H$ from precision loss is small, expected to scale poorly for larger volume), or even using single precision MD and skip Metropolis when the step size error is deemed smaller than other systematic errors.

Recently there have been rediscovery of algorithms, such as defect correction(used in [38]) and Reliable updates[88] for effective use of lower precision solvers for double precision inversions. Reliable update technique in particular has shown to be very effective in using single or half(16bit) precision, useful in GPGPU. More detailed description can be found in [90, 91]. In addition to this, there are other nested inversion algorithms such as Generalized Conjugate Residual(GCR), used with Domain decomposition[89] or inexact deflation[92]. A more extensive studies of these algorithms could allow much more effective inversion algorithms on existing and emerging architectures.

6. Conclusions

Recent advances in dynamical fermion simulation algorithms and continuing progress in hardware has enabled multiple collaborations to generate dynamical ensembles with different lattice discretizations, lattice spacings and dynamical quark masses. This is a striking departure from only a few years ago, when Asqtad configurations generated by the MILC collaboration were the only choice for continuum extrapolation with multiple lattice spacings and quark masses. Some of the ensembles are even being generated at or close to the physical pion mass. $N_f = 2, 2 + 1$ and $2+1+1$ dynamical lattice QCD ensembles generation activities of various collaborations are summarized. It is shown that computing resources on the order of 10 to 100 TFlop-years, well within reach with technologies currently available, is enough to generate $N_f = 2 + 1$ ensembles with $> 100$ independent configurations at $m_\pi < 200$ Mev and $m_\pi L > 4$, barring the effect of autocorrelation observed by global topological charge.

A steep dependence of simulation cost on the lattice spacing and different scaling behavior of different actions, as well as the ambiguities in defining the lattice spacing itself, makes superficial comparisons between different actions not necessarily useful. Although it has become possible to use relatively inexpensive actions such as Wilson to generate ensembles very close to or at physical quark mass, the presence of significant autocorrelations observed by topological charge evolution at smaller lattice spacings ($a \leq 0.08$fm) should be better understood and circumvented to ensure usefulness of large scale ensembles. It is possible that multiple, more highly improved actions at a moderate lattice spacing is a better way to ensure that systematic errors are under control for all the quantities we are interested in studying. Presence of many ensemble generation activities is helpful in that respect.
Some of the recent progress in algorithms and techniques were also summarized. The mass reweighting technique, which turned out to be quite effective in eliminating one of the persistent source of systematic error is explained in more detail. Reweighting in general could alleviate the need to generate separate ensembles for slightly different parameters and/or actions, even for relatively large lattices.

Acknowledgments

I would like to thank the colleagues who sent previously unpublished materials before the conference for preparation of the talk, in particular A. Bazavov, T. Blum, T-W. Chiu, C. Detar, S. Gottlieb, G. Heroïda, R. Horsley, B. Joo, Y. Kuramashi, C. B. Lang, R. D. Mawhinney, D. Palao, S. Reker, G. Schierholz and H. Wittig. I am also indebted to many RBC/UKQCD colleagues for many helpful discussions. The author was supported by the U.S. DOE under contract DE-AC02-98CH10886.

References

Chulwoo Jung

Status of dynamical ensemble generation


[38] https://twiki.cern.ch/twiki/bin/view/CLS/WebHome


[57] N. Mathur, These proceedings, PoS(LAT2009)100.

[58] H. Fukaya, These proceedings, PoS(LAT2009)004.


C. Detar, private communications.


T. Ishikawa et al., These proceedings, PoS(LAT2009)035.


M. Clark, These proceedings, PoS(LAT2009)003.
