# Progress in Algorithms and Numerical Techniques

Chulwoo Jung
Brookhaven National Laboratory

August 2, 2013
Lattice 2013
Johannes Gutenberg Universität, Mainz, Germany

# Introduction

From hep-lat/0411006. 2f DWF

TABLE I: Small lattice comparison of HMC evolutions. All these evolutions use the Wilson gauge action with $\beta = 5.2$ and two flavours of Domain Wall Fermions with a bare mass of $m_{\rm sea} = 0.02$.

| Force Term | $\Delta t$ | Steps/Trajectory | Trajectories | Acceptance | CG-iterations/Trajectory | $C_{\Delta H}$ |
|---|---|---|---|---|---|---|
| Old | 1/64 | 33 | 1000-1880 | 87% | 8336 | 26.5 |
| Old | 1/32 | 17 | 1000-1929 | 59% | 4310 | 30.0 |
| New | 1/32 | 17 | 1000-1936 | 79% | 4179 | 12.9 |

Advances in integrator(multiscale,Omelyan,Force Gradient..), Mass preconditioning, Rational Hybrid Monte Carlo(RHMC), Domain decomposition, etc allowed exact dynamical evolution at or near physical $m_l$ with only relatively small ($\sim$10) number of light quark inversions per MD.
For review, S. Schaefer (Lat12), CJ(Lat09), K. Jansen(Lat08) M. Clark (Lat06), ....

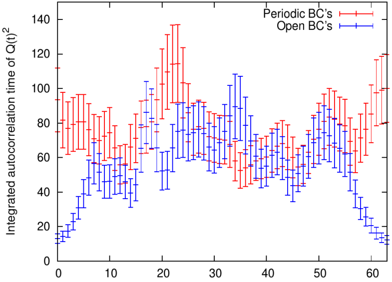
# What about measurements??

- Bigger step size for light quark during MD means often number of light quark inversions per configurations for measurements far exceeds those for evolution.
- While Rational approx. has made it possible to tune the approximation during MD and preserve acceptance, one is often forced to adopt conservative criteria (tighter stopping condition, etc) for propagators, which hampers aggressive use of progress in solver.
- Better ways of taking advantage of the characteristic of $\not{D}$ necessary, for a large number of light quark inversion per configuration.
- Improving solver performance : Deflation techniques
- Error reduction techniques: All Mode Averaging

# Evolution(autocorrelation)

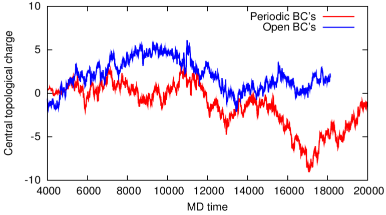
McGlynn, Mon. 1D

Iwasaki $a^{-1} \sim$ 3 Gev, $24^3x64$ $\longrightarrow$ $\downarrow$



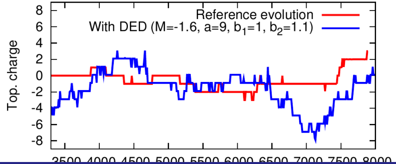

Wilson $a^{-1} \sim$ 4Gev, $32^4$ $\longrightarrow$

Topological charge on central half of lattice (excludes boundary regions)




Dislocation Enhancing Determinant (DED)

$$\det\left[f(D_W^\dagger D_W)\right] = \prod_i f(\lambda_W^2)$$



# Solvers

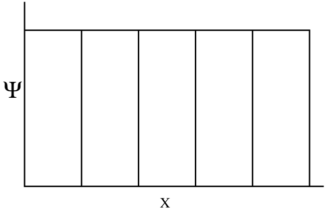
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - bD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ DC^{-1} & I \end{bmatrix}$$

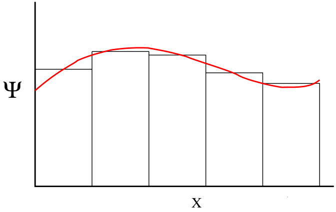
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}$$

Find $D$ which improves condition number of $A - BD^{-1}C$ efficiently. $A - BD^{-1}C$ is easy to apply if:

- $D = I_n$ (even-odd preconditioning)
- $D$ local in 4d (Clover, Mobius...)
- $D$ is approx. eigenspace (eigCG(Orginos et al.))
- None of the above? → Inexact deflation, Multigrid, ...

Local coherence(Luscher): Many low modes are locally similar $\rightarrow$ we can construct effective $D$ from small number of approx. low modes by domain decompsition. Crucial in controlling setup time for $D$.



However "$D$evil is in the $D$etails"... $D^{-1}$ can be very inefficient on parallel machines:

- $D$ built form block size $L^4(\times Ls)$ has

$$\left.\begin{array}{c}\text{flops reduced by } L^4(\times Ls) \\ \text{Communication size reduced by } L^3(\times Ls) \\ \text{latency not reduced}\end{array}\right\} \times \text{ by iter}(D^{-1})$$

  quickly becomes communication bounded. This will only get worse for future machines.
- For non-Wilson/Clover, even-odd precondited $\not{D}^\dagger \not{D}$ makes $D$ highly nontrivial.
- Have to find ways to invert $D$ efficiently, or ways to do less of it! Until recently, success varied (more successful for Wilson/Clover, less for other fermions)

# Deflated solvers: Wilson/Clover

- Domain Decomposition/aggregation based Adaptive Algebraic Multigrid (DD-$\alpha$AMG)
  (Rottmann, Mon. 1D )
  Multi-level DD-$\alpha$AMG(arXiv:1303.1377) implemented and tested for up to 4 levels. Found 3 level MG much better than 2 for larger Wilson lattices.

- Adaptive Multigrid on GPU (M. Clark, Tue. 4G )
  Multigrid being implemented in QUDA (USQCD SciDAC) framework.
  Staggered/Clover preconditioner.

# Hierarchically Deflated CG(HDCG) (P. Boyle, Mon. 1D)

First "practically successful" results on non-Wilson fermions.

- Preconditioned DWF, up to 4th-nearest neighbor interactions
- Use heavy ($M \sim 1$) inversion as high-frequency preconditioner.
- Move $D^{-1}$ to preconditioner, Exact deflation of $(128-256)$ low modes of $D$ + relaxed stoppint condition ($\sim 10^{-2}$) for $D^{-1}$ to decrease iteration number for $D^{-1}$ to $O(50)$
- Employed A-DEF2(Tang et al, J. Sci. Comput(2009) 340) preconditioned solver. Robust against relaxed precision of $D^{-1}$.

$48^3 \times 96 \times 24$, $a \sim 0.11$fm on 1024-node BG/Q(32 threads)

| | EigCG | HDCG |
|---|---|---|
| Vec # | 600 | 40 |
| Setup | 10h | 40min |
| Mixed($10^{-8}$) | 320s | 155s |
| Mixed($10^{-4}$) | 55s | 8s |

# Discussion:Evolution, Solvers

- Autocorrelation in HMC for $a^{-1} = 3 \sim 4$Gev may need a closer look. Further improvement with Dislocation Enhancing Determinant studied.
- Progress in deflated solvers, now in both Wilson and non-Wilson(DWF). Hierarchically Deflated CG achieves a significant speedup over CG, eigCG.
- There is still a significant variation in hardware specifications (GPU $\leftrightarrow$ IBM BG/Q). Optimal choice of details of deflation may depend on hardware environment.

# All Mode Averaging(AMA)

(Blum, Izubichi & Shintani arXiv:1208.4349)
Use the translation invariance of LatticeQCD Lagrangian and replace $\mathcal{O}^{(rest),g}$ with $\mathcal{O}^{(rest)}$ for a set of covariant shifts $\{g\}$.

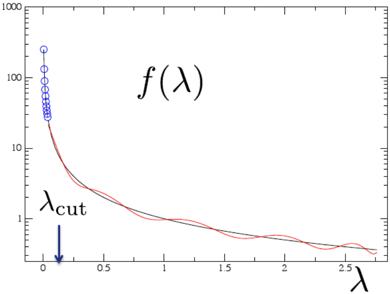
$$\mathcal{O}^{(imp)} = \frac{1}{N_G}\Sigma_g\left(\mathcal{O}^{(rest),g} + \mathcal{O}^{(approx),g}\right)$$
$$\sim \frac{1}{N_E}\mathcal{O}^{(rest)} + \frac{1}{N_G}\Sigma_g\mathcal{O}^{(approx),g}$$
$$\mathcal{O}^{(rest)} = \mathcal{O}^{(exact)} - \mathcal{O}^{(approx)}$$

This is cost effective when $\mathcal{O}^{(approx)}$ is such that

- $\langle(\Delta(\mathcal{O}^{(rest)})^2\rangle \ll \langle(\Delta(\mathcal{O}^{(approx)})^2\rangle$
- $\mathcal{O}^{(approx)}$ much less expensive than $\mathcal{O}^{(exact)}$
- Covariance: $\mathcal{O}^{(approx),g}[U] = \mathcal{O}^{(approx)}[U^g]$ (However, can be relaxed)

Low Mode Averaging(LMA):
$\mathcal{O}^{(approx)}$ constructed only from low modes. Efficacy depedent on observables. Improving $\mathcal{O}^{(approx)}$ by calculating more low modes challenging.

From T. Izubuchi

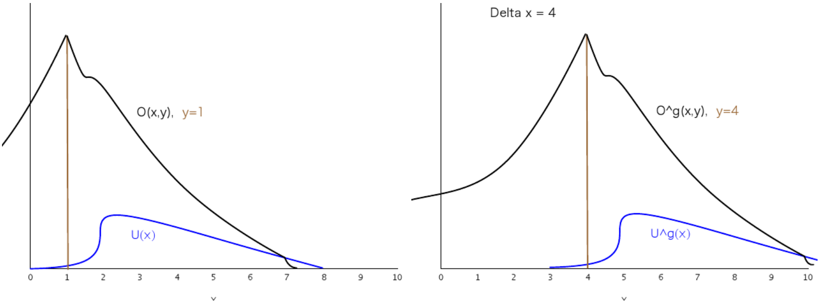



Not all observables are dominated by low modes, sensitive to fluctuations in high modes.
Combine low modes with the deflated inversion with relaxed ($\sim 10^{-4}$) stopping condition ("sloppy solve") $\rightarrow$ AMA.
Low modes crucial: $\mathcal{O}^{(approx)}$ with only sloppy solve fails for low energy quantities.

With (in)exact deflation, a significant saving is achieved by relaxing the stopping condition (typecially factor of $\sim 4$) compared deflated exact inversions.
Control fluctuation of $\mathcal{O}^{(rest)}$ by stopping condition of $\mathcal{O}^{(approx)}$ and/or $N_E$
Variety of deflation and approximation can be used

- (Almost) exact low modes(Lanczos, eigCG) + sloppy solve
- Approx. low modes (inexact deflation, HDCG, Multigrid,$\cdots$) + sloppy solve
- Deflation + Approx. action (eg. large $L_s$ DWF $\leftrightarrow$ small $L_s$ Mobius) + sloppy solve

# Stochastic choice of exact calcuation




Covariance requires that $\mathcal{O}^{(approx)}$ does not have any position dependence. Approx. solution often introduces translational invariance breaking. $\rightarrow$ Choose position for $\mathcal{O}^{(exact)}$ randomly among $\{g\}$. $\mathcal{O}^{(approx)}$ just has to be unique/deterministic.

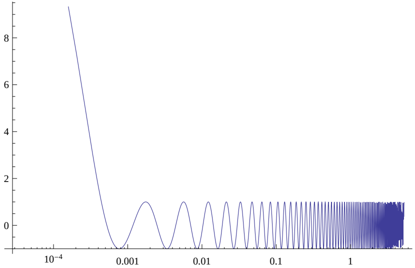
# Relation with other error reduction methods

- Low Mode Averaging(LMA) (Giusti et al., hep-lat/0402002, Degrand & Schaefer, hep-lat/0401011)
- Low Mode Substitution(LMS) (Li et al., arXiv:1005.5424): Random sources on grid points + exact low modes. Improved S/N for 2-pt functions.
- Truncated Solver Methods(TSM) (Bali et al., arXiv:0910.3970): Relaxed inversion (+ low mode deflation). Translational invariance $\leftrightarrow$ random source. Used for disconnected diagrams.
- All to all propagators (A2A)(Bali et al., hep-lat/0505012, Foley et al., hep-lat/0505023): Use random sources to allow stochastic evaluation of multiple operators.
- Distillation(Peardon et al., arXiv:0905.2106): Use low eigenmodes of 3d Laplace operator per each time slice. Can be combined with AMA in time direction.

# AMA applications

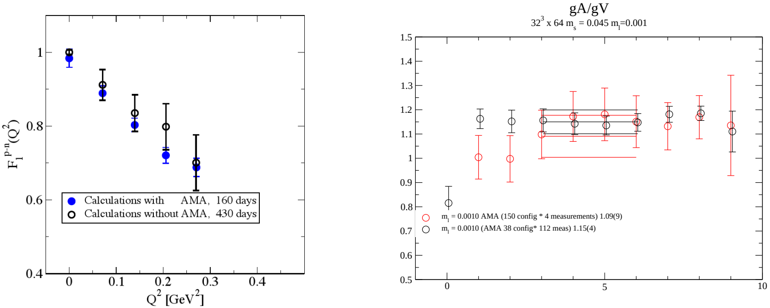
$g_A$ and nucleon form factors on DWF+DSDR $M_\pi \sim$ 170Mev,
S. Ohta, M. Lin, Thur. 7B

| | |
|---|---|
| $a$(fm) | 0.14 |
| $L$ | $32^3 \times 64 \times 32$ |
| # config | 39 |
| # LM(Lanczos) | 2000 |
| # Exact | 4 |
| # AMA | 112 |
| ($L_s = 16$ Mobius) | ‘ |
| Cost to 1 undeflated meas. | |
| Setup | 0.7 |
| Exact($10^{-8}$) | 1 |
| Sloppy($10^{-4}$) | 0.01 |
| gain | $\sim$19 |




Chevyshef acceleration(Neff et al. hep-lat/0106106) : Maps unwanted eigenvalues to {-1,1} to accelerate Lanczos.

w/o AMA : 150 configs $\times$ 4 sources $=$ 600 measurements
AMA: 39 configs$\times$ 112 source $=$ 4368 measurements

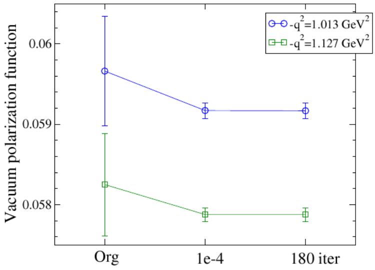



$g_A$ : 1.09(9) / 1.15(4)

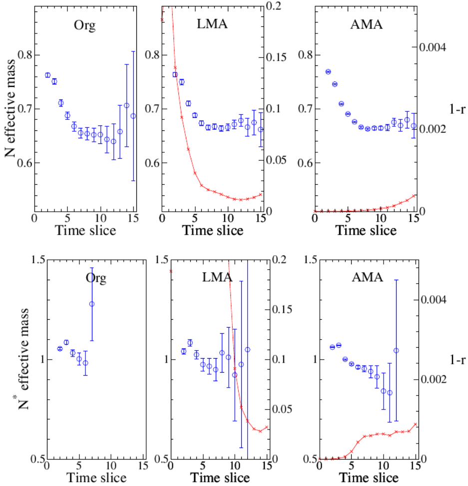
Error bar $\times$ 2 - 2.7 $\sim \sqrt{\frac{4400}{600}} \rightarrow$ all AMA measurements nearly independent.

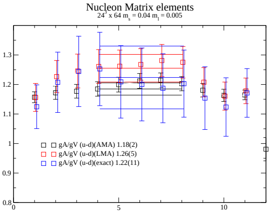
Hadronic Vacuum Polarization on DWF+I,

| | |
|---|---|
| $a$(fm) | 0.11 |
| $L$ | $24^3 \times 64 \times 16$ |
| # config | 391 |
| # LM | 400 |
| # Exact | 1 |
| # AMA | 32 |
| Cost compared to 1 undefl. meas | |
| Setup | 2 |
| Exact($10^{-8}$) | 0.17 |
| Sloppy($10^{-4}$) | 0.032 |
| gain | $\sim$10 |




$-\Pi(Q^2)$ (37 config.)
from Blum, Christ, Izubuchi & Shintani, in preparation.



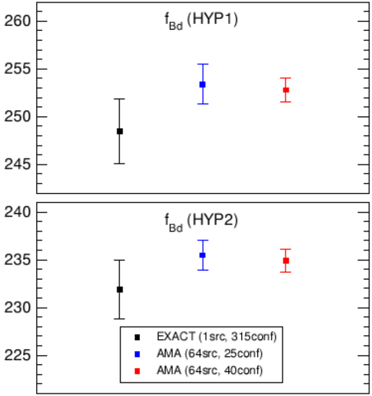




gA/gV(AMA/LMA/exact)=
1.18(2):1.26(5):1.22(11)

Error reduced by $\sim \sqrt{32}$ compared to exact, 2∼3 compared to LMA.

Static heavy-light w. AMA $f_{Bd}$ T. Ishikawa, Tues. 4C

| | |
|---|---|
| $a$(fm) | 0.08 |
| $L$ | $32^3 \times 64 \times 16$ |
| # config | 64 |
| # LM | 130 |
| # Exact | $1(10^{-8})$ |
| # AMA | $64\ (3 \times 10^{-3})$ |
| gain | $\sim$5 over defl. exact |




$$\left[\frac{\sigma(1\times 315 exact)}{\sigma(40\times 64 AMA)}\right]^2 = 2.2 \sim 3.8 \text{ Compared to } \frac{40\times 64}{1\times 315} = 8.13$$

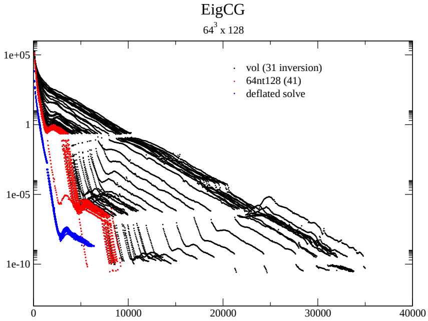
# DWF+I WME measurements @ Physical pion mass

Talks by Juettner (Thurs 7C) Janowski, Mawhinney (Thurs. 8C)

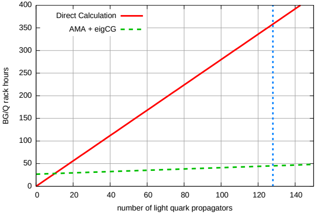
$B_K, f_\pi, f_K, kl_3, K \to \pi\pi \cdots$

Deflation with eigCG (Orginos et al., arXiv:0707.0131)

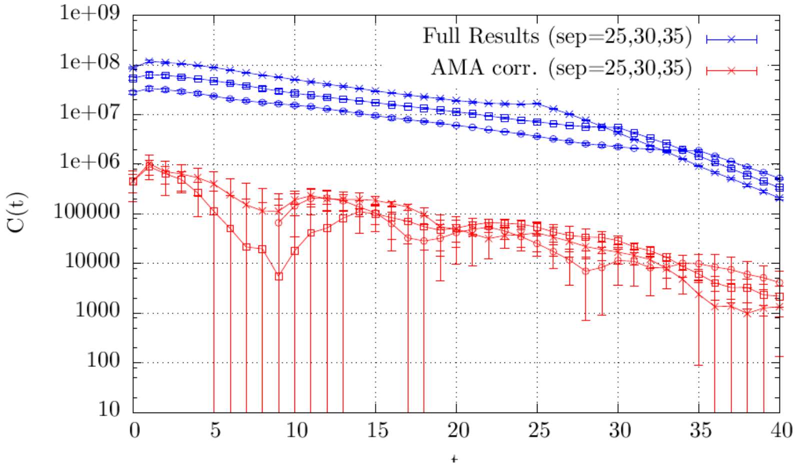
| $a$(fm) / $L$ | 0.08<br>$64^3 \times 128 \times 12$ | 0.11<br>$48^3 \times 96 \times 24$ |
|---|---|---|
| # config | 21 | 42 |
| # LM | 1500 | 600 |
| # Exact(Stochastic) | 8 | 7 |
| # AMA | 128 | 96 |
| Cost: BG/Q rack-hours | | |
| Setup | 26.2 | 9.8 |
| Exact prop($10^{-8}$) | 0.49 | 0.89 |
| Sloppy prop($10^{-4}$) | 0.12 | 0.22 |
| Orig(Undefl. prop) | 2.8 | 1.8 |
| gain(light prop.) | $\sim 7.9$ | $\sim 4.6$ |
| Total WME measurement per config | 170.2 | |
| 1 MD (8 rack) | 5.3 | |



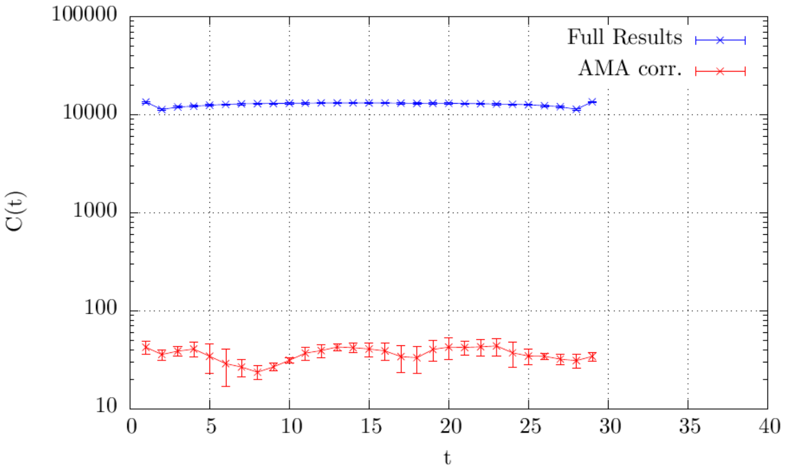

Residuals for EigCG setup(black), extra(Red), deflated exact solve(blue).




Cost scaling for $64^3 \times 128$ measurements.




$kl_3$ correlators on $64^3$ lattice, $\mathcal{O}^{(rest)} \sim \mathcal{O}^{(approx)}/50$



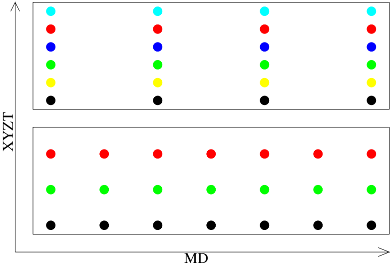

$B_k$ correlator for $64^3$ DWF lattices

Comparison of AMA and exact results for DWF+I endembles (from D. Murphy)
Exact/AMA $= 7/96(48^3), 8/128(64^3)$

| Ensemble | Observable | Exact | AMA | $\sigma_{\text{Exact}}/\sigma_{\text{AMA}}$ |
|---|---|---|---|---|
| $48^3 \times 96$ | $m_\pi$ | 0.08006(51) | 0.08065(18) | 2.81 |
| | $m_K$ | 0.28813(55) | 0.28840(23) | 2.39 |
| | $f_\pi$ | 0.07650(32) | 0.07601(13) | 2.38 |
| | $f_K$ | 0.09099(37) | 0.09063(13) | 2.94 |
| | $f_K/f_\pi$ | 1.1894(48) | 1.1924(18) | 2.65 |
| | $B_K$ | 0.58132(851) | 0.58363(85) | 10.0 |
| | $Z_A$ | 0.71374(153) | 0.71203(20) | 7.81 |
| $64^3 \times 128$ | $m_\pi$ | 0.05857(48) | 0.05891(26) | 1.89 |
| | $m_K$ | 0.21563(51) | 0.21510(21) | 2.50 |
| | $f_\pi$ | 0.05555(29) | 0.05545(11) | 2.71 |
| | $f_K$ | 0.06650(32) | 0.06643(13) | 2.40 |
| | $f_K/f_\pi$ | 1.1972(63) | 1.1980(26) | 2.44 |
| | $B_K$ | 0.5776(118) | 0.5623(12) | 10.2 |
| | $Z_A$ | 0.74302(147) | 0.74344(16) | 9.28 |

# Error analysis with AMA

As $\mathcal{O}^{(approx)}$ is factor of $4 \sim 5$ cheaper to calculate, maximizing $N_G$ is cost effective.
Appears to give statistically independent measurements for $N_G$ up to $\sim 100$.
As lattice spacing decreases, autocorrelation in MD is still a concern: AMA results shows we can mitigate this by measuring more per configuration.




Cost effective way to get maximum information per configuration
Relatively small number of configurations can make error analysis less straightforward (What is the error on error?)
This is nothing new - eg. check 1/N correction on jackknife.
What does it say about evolution algorithm in general?

# Summary: AMA

- All mode averaging is a class of covariant approximation which combines delfation with relaxed solver for high modes. Shown to be highly effective for a wide range of observable.
- Can be easily combined with existing deflation techniques(lanczos, eigCG, HDCG, inexact deflation, Multigrid...).
- Already in use for various quantities such as Nucleon stucture function, WME, HVP, Proton Decay, Heavy quark (Static)
- Explorations for other observables, approximations can improve this further.

Thank you!

Wilson $a^{-1} \sim 4$Gev