GRID and evolution algorithms for critical slowing down

Guido Cossu*

in collaboration with P. Boyle, N. Christ, C. Jung, A. Jüttner, F. Sanfilippo

*leaving to Ascent Robotics

Frontiers in Lattice quantum field theory 2018 – IFT Madrid May 31th

EXASCALE COMPUTING PROJECT

THE UNIVERSITY OF EDINBURGH

Computing resources

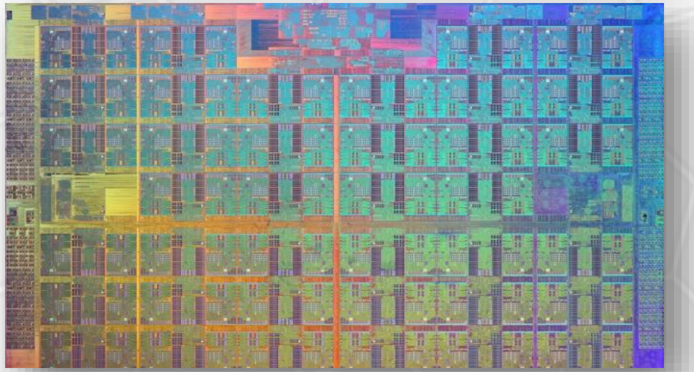
ECP
EXASCALE COMPUTING PROJECT

DiRAC Distributed Research utilizing Advanced Computing

Science & Technology Facilities Council

Argonne NATIONAL LABORATORY

NERSC
National Energy Research Scientific Computing Center



THE UNIVERSITY OF EDINBURGH

GRID library

Modern C++ library for Cartesian mesh problems
P. Boyle, G. Cossu, A. Portelli, A. Yamaguchi

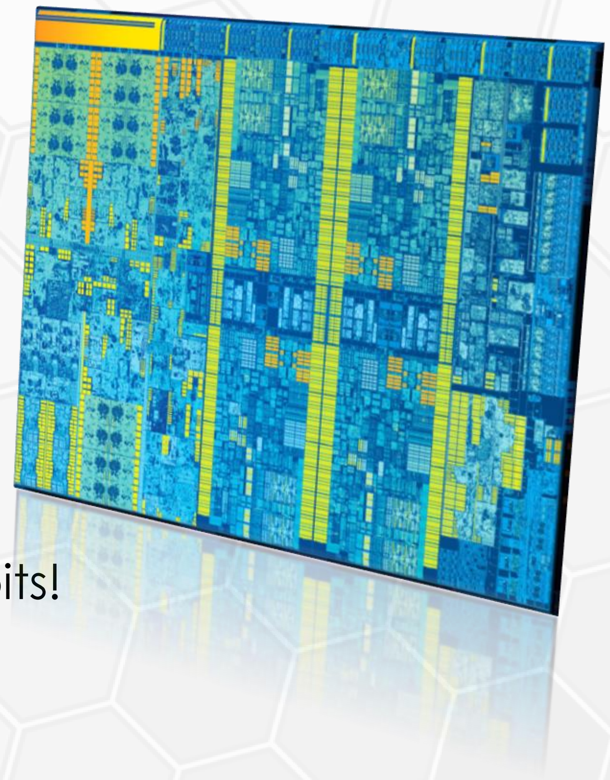
Grid github pages github.com/paboyle/Grid
(partial) documentation paboyle.github.io/Grid/
CI: Travis, TeamCity https://ci.cliath.ph.ed.ac.uk/

Design goals

- performance portability
- zero code replication

Source code: C++11, autotools
+ HADRONS physics measurement framework based on Grid (A. Portelli)
Intense work in progress, production stage on Tesseract

GRID architecture support



Current support

- SSE4.2 (128 bit)
- AVX, AVX2 (256 bit) (e.g. Intel Haswell, Broadwell, AMD Ryzen/EPYC)
- AVX512F (512 bit, Intel KNL, Intel Skylake)
- QPX (BlueGene/Q), experimental
- NEON ARMv8 (thanks to Nils Meyer from Regensburg University)
- Generic vector width support

Work in progress for

- CUDA threads (Nvidia GPUs) (GRID team & ECP collaboration)
- ARM SVE (Scalable Vector Extensions - Fujitsu post-K), from 128 up to 2048 bits!

Exploiting all levels of parallelism

- Vector units, Threading, MPI

Work on optimising communications triggered **3 major updates** for

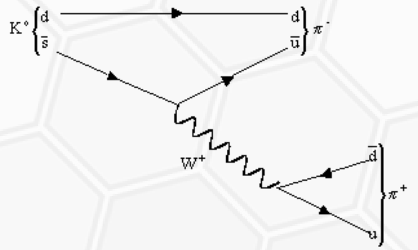
- Intel MPI stack (library and PSM2 driver)
- HPE-SGI Message Passing Toolkit (MPT)
- Mellanox HPC-X

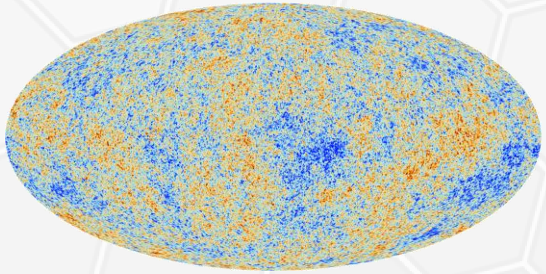
GRID features (selection)

- Actions
  - Gauge: Wilson, Symanzik, Iwasaki, RBC, DBW2, generic Plaquette + Rectangle
  - Fermion: Two Flavours, One Flavour (RHMC), Two Flavours Ratio, One Flavour Ratio, Exact one-flavour. All with the EO variant.
  - Kernels: Wilson, Wilson TM, Wilson Clover + anisotropy, generalised DWF (Shamir, Scaled Shamir, Mobius, Z-mobius, Overlap, …), Staggered
  - Scalar Fields
  - Integrators: Leapfrog, 2^{nd} order minimum-norm (Omelyan), force gradient, + implicit versions
- Fermion representations
  - Fundamental, Adjoint, Two-index symmetric, Two-index antisymmetric, and all possible mixing of these. Any number of colours. All fermionic actions are compatible.
- Stout smeared evolution with APE kernel (for SU(3) fields). Any action can be smeared.
- Serialisation: XML, JSON
- Algorithms: GHMC, RMHMC, LAHMC, density of states LLR (not public) easily implemented
- File Formats: Binary, NERSC, ILDG, SCIDAC (for confs). MPI-IO for efficient parallel IO
- Measurements:
  - Hadrons (2,3 point functions), Several sources, QED, Implicitly Restarted Lanczos, and many more…
- Split Grids: 3x speedup + deflation for extreme scalability

Some HMC features inherited from IroIro++

GRID current physics

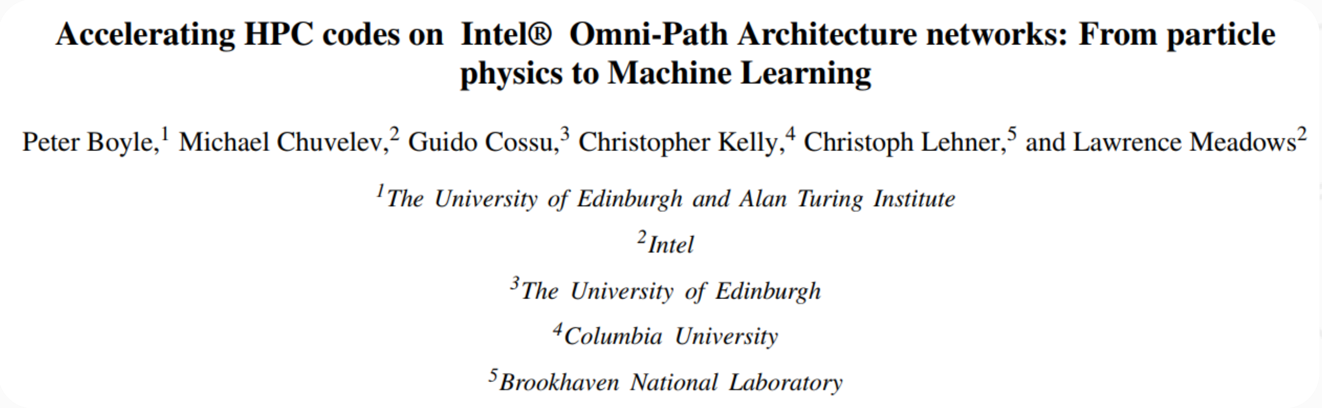




- RBC-UKQCD
  - This work on algorithms…
  - Kaon decay with G-parity $K \to \pi\pi$
  - QED corrections to Hadron Vacuum Polarization
  - Non Perturbative Renormalization
  - Holographic cosmology
  - BSM, composite Higgs with mixed representations
- Numerical Stochastic PT (Wilson Fermions) G. Filaci (UoE)
- Axial symmetry at finite temperature, Semi-leptonic B-decays (GC with JLQCD)
- Density of states (GC with A. Rago, Plymouth U.)
- …

GRID/Intel paper and on the tech news!

On the optimization of comms and how to drive the Intel Omni-Path
arXiv:1711.04883


Accelerating HPC codes on Intel® Omni-Path Architecture networks: From particle physics to Machine Learning

Peter Boyle,[1] Michael Chuvelev,[2] Guido Cossu,[3] Christopher Kelly,[4] Christoph Lehner,[5] and Lawrence Meadows[2]

[1]*The University of Edinburgh and Alan Turing Institute*

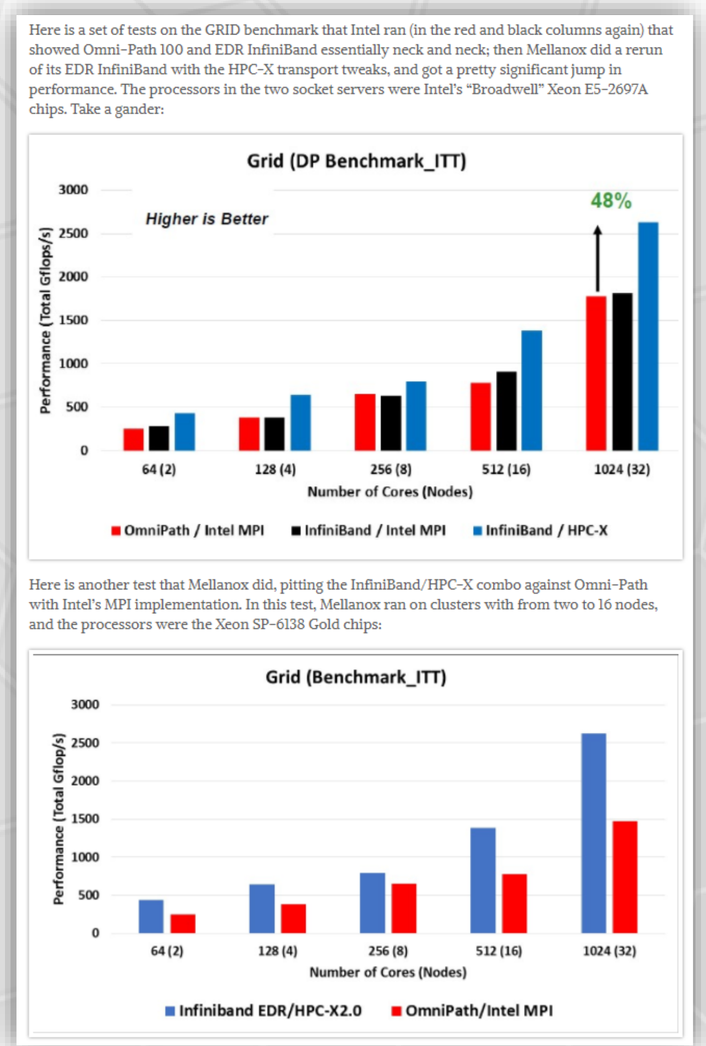
[2]*Intel*

[3]*The University of Edinburgh*

[4]*Columbia University*

[5]*Brookhaven National Laboratory*

HPC Tech news site reported GRID benchmarks in a *Battle of the InfiniBands* article (Nov 29) from the HPC Advisory Council slides



Critical slowing down

Continuum limit of lattice QFT: second order critical point

$$\tau_{\text{int}}(O) \propto a^{-\varepsilon}$$

The exponent depends on the:

- algorithm to generate the Markov Chain
- observable (if not universal)

Some observables couple more tightly to the slow modes of the transition matrix, e.g. the topological charge

Markov Chain Monte Carlo

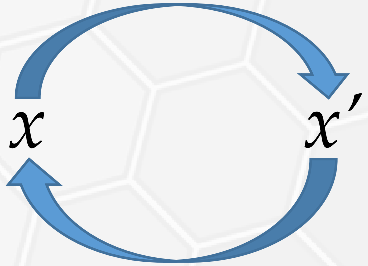
Problem: sampling of a target probability distribution $p(x) = \frac{1}{Z} \exp(-S(x))$

Generate a Markov chain $x_0 \to x_1 \to \ldots \to x_N$

such that $\int p(x) T(x'|x) dx = p(x')$

$p(x)$ is a fixed point of the transition matrix T, or eigenmode for a discrete set of states

$$\langle O \rangle = \frac{1}{Z} \int O \exp(-S(x)) dx \approx \frac{1}{N} \sum O(x_n)$$



Detailed balance $p(x) T(x'|x) = p(x') T(x|x')$
→ fixed point equation

Note: detailed balance is a sufficient but NOT a necessary condition

Generalised Hybrid Monte Carlo

Hamiltonian evolution $H(x) = \frac{1}{2}p^T p + S(x)$ $p(x) \propto \exp(-H(x))$

Duane et al. 1987

$$x' = FL(M,\epsilon)x^{(t,0)}$$

$x^{(t,1)} = x'$ Accept

$x^{(t,1)} = x^{(t,0)}$ Reject

$$\pi_{\text{accept}} = \min\left(1, \frac{p(x')}{p(x)}\right)$$

$$x^{(t,2)} = Fx^{(t,1)}$$

$$x^{(t+1,0)} = R(\alpha)x^{(t,2)}$$

Note: $(FL)^{-1} = L^{-1}F = FL$

simplifying the Metropolis-Hastings step $\pi_{\text{accept}} = \min\left(1, \frac{p(x')}{p(x)}\frac{T(x|x')}{T(x'|x)}\right)$

See e.g. Kennedy, Pendleton hep-lat/0008020

Markov transitions

Hamiltonian integration

M steps, ε step size

Symplectic, reversible

$$L(M,\epsilon)$$

Momenta flip F

Momenta randomization

α mixing angle

$$R(\alpha)$$

Generalised Hybrid Monte Carlo

Hamiltonian evolution $H(x) = \frac{1}{2}p^T p + S(x)$
Duane et al. 1987

$p(x) \propto \exp(-H(x))$

Some special cases
- complete randomization of momenta: Standard HMC
- no momenta randomization: microcanonical integration (non ergodic)
- single step integration and complete momenta randomization: Langevin Monte Carlo

Markov transitions

Hamiltonian integration
M steps, ε step size
Symplectic, reversible

$$L(M, \epsilon)$$

Momenta flip F

Momenta randomization
α mixing angle

$$R(\alpha)$$

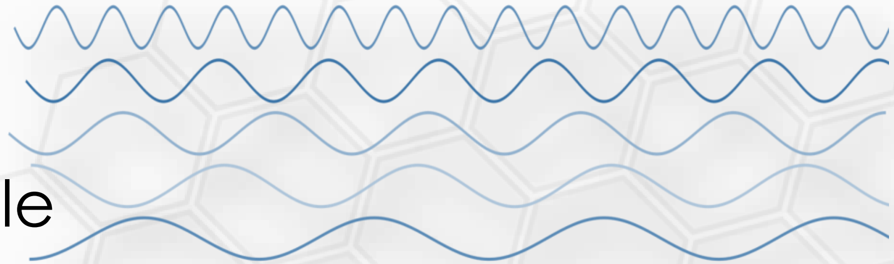
See e.g. Kennedy, Pendleton hep-lat/0008020

Riemannian Manifold Hybrid Monte Carlo
RMHMC

Accelerating slow modes

Transition matrix:

- Fastest modes limit the integration step size
- Slow modes take longer time to complete a cycle and decorrelate



Increasing ratio of frequencies determines the critical slowing down

Simple scalar free field example: characteristic frequency $\omega(p)^2 = m^2 + p^2$

Consider the Hamiltonian evolution of

$$H(\pi, \psi) = \frac{1}{2}\pi^T (M^2 - \partial^2)^{-1} \pi + S(\phi) \quad \Rightarrow \quad \omega(p)^2 = \frac{m^2 + p^2}{M^2 + p^2}$$

Extend this idea to gauge theories

Gauge invariant Fourier acceleration

Duane, Pendleton et al. 1986, 1988

Formulated in geometric terms recently by Girolami, Calderhead 2011

Covariant modifications of the kinetic term (metric in a Riemannian Manifold: RM-HMC)

$$\frac{1}{2}\pi^T M^{-1}\pi \qquad M\phi(x) = (1-\kappa)\phi(x) - \frac{\kappa}{4d}\nabla^2\phi(x)$$

Covariant laplacian

$\kappa \to 1$ max acceleration for free fields

The resulting Hamilton equations are non-separable

Leapfrog-like schemes do not work anymore as non reversible and non volume conserving.

Integration procedure by implicit integration (Leimkuhler, Reich 2004)

$$\pi^{n+\frac{1}{2}} = \pi^n - \frac{\epsilon}{2}\frac{\delta H}{\delta\phi}(\phi^n, \pi^{n+\frac{1}{2}})$$

$$\phi^{n+1} = \phi^n + \frac{\epsilon}{2}\left[\frac{\delta H}{\delta\pi}(\phi^n, \pi^{n+\frac{1}{2}}) + \frac{\delta H}{\delta\pi}(\phi^{n+1}, \pi^{n+\frac{1}{2}})\right]$$

$$\pi^{n+1} = \pi^{n+\frac{1}{2}} - \frac{\epsilon}{2}\frac{\delta H}{\delta\phi}(\phi^{n+1}, \pi^{n+\frac{1}{2}})$$

RM-HMC evolution

Another modification is necessary

- The new kinetic term introduces the inverse determinant of M in the distribution: it must be cancelled

Two solutions

- a set of auxiliary fields (scalars in the adjoint representation) with action $\frac{1}{2}p_\theta^T M^2 p_\theta + \theta^2$ and include this in the Hamiltonian integration
- Use pseudofermion-like fields and update only at the refresh step

New parameters: κ

Choice of operator M not unique.

We can imagine also operators with a bounded spectral content

Overhead: inversion of the Laplacian, implicit integration

Should be irrelevant once fermions are included

Tested originally in 2d SU(2) pure gauge (with non exact Runge-Kutta integration)

Now testing on 4d SU(3) pure gauge and CP^N models (Jüttner, Sanfilippo)

Look Ahead Hybrid Monte Carlo
LAHMC

Dropping detailed balance

$$p(x)T(x'|x) = p(x')T(x|x')$$

a sufficient but NOT a necessary condition
Can introduce a random walk behaviour to the evolution

Rejection, in general, leads to momentum reversal, wasteful

We would like to
- move further in the integration
- reduce the rejection steps optimizing the costs

Look Ahead HMC

Repeat the integration accepting with modified probabilities up to a maximum number K of times

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

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

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

Same as HMC (generalized) for K=1

Target distribution is a fixed point of the evolution

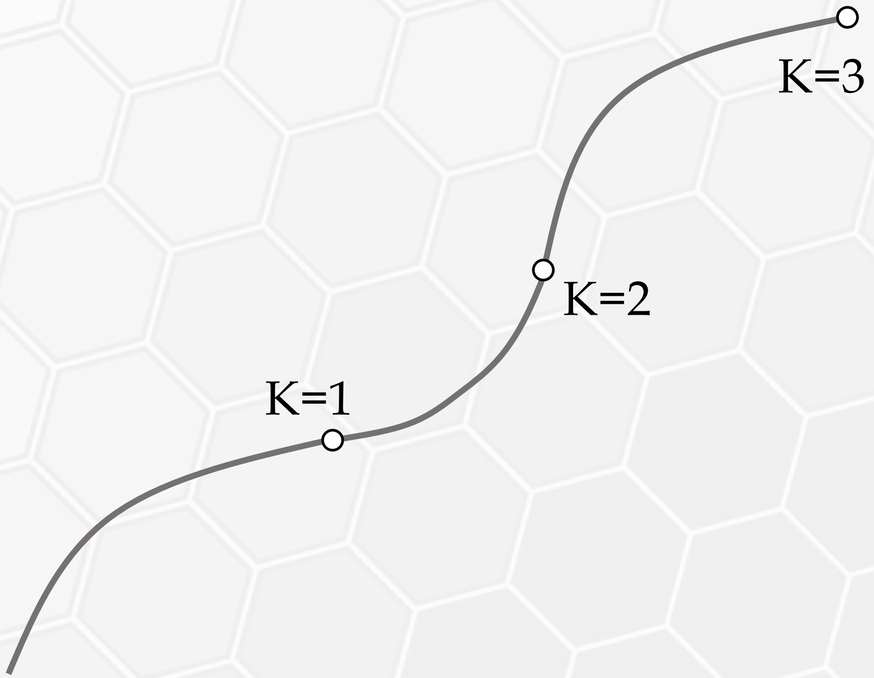
Sohl-Dickstein et al. (for machine learning) 2016

$$x' = FL(M,\epsilon)x^{(t,0)} \qquad \begin{array}{ll} x^{(t,1)} = x' & \text{Accept} \\ x^{(t,1)} = x^{(t,0)} & \text{Reject} \end{array}$$

$$\pi_{\text{accept}} = \min\left(1, \frac{p(x')}{p(x)}\right)$$

$$x^{(t,2)} = Fx^{(t,1)}$$

$$x^{(t+1,0)} = R(\alpha)x^{(t,2)}$$



Look Ahead HMC

Repeat the integration accepting with modified probabilities up to a maximum number K of times

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

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

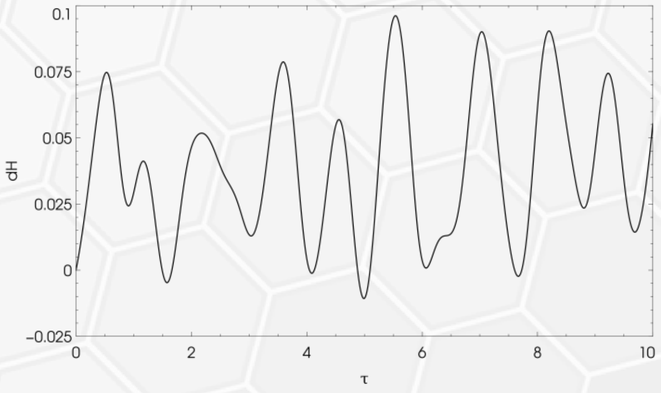
$$\pi_{L^a}(x) = \min\left[1 - \sum_{b<a} \pi_{L^b}(x), \frac{p(FL^a x)}{p(x)}\left(1 - \sum_{b<a} \pi_{L^b}(FL^a x)\right)\right]$$

Same as HMC (generalized) for K=1

Target distribution is a fixed point of the evolution

Sohl-Dickstein et al. (for machine learning) 2016

Symplectic integrators:
H oscillates around the original value during evolution, dH same order even after long trajectories



Hopefully more work per chain pays in terms of sample quality

Simulations: current status

Pure gauge SU(3)

Implemented both algorithm prototypes in Grid

- Few days of work for RMHMC (for any gauge theory w/o fermions)
- One day for LAHMC

Nice example of the higher level flexibility of Grid

Reference runs: HMC in Grid with 2nd order minimum norm integration.

Two volumes 16^4 and 32^4, Wilson Gauge action at $\beta = 6.2, 6.4$ respectively.

CP^N model (Jüttner, Sanfilippo)

Model has severe slowing down at small lattice spacing

Simulating N=10

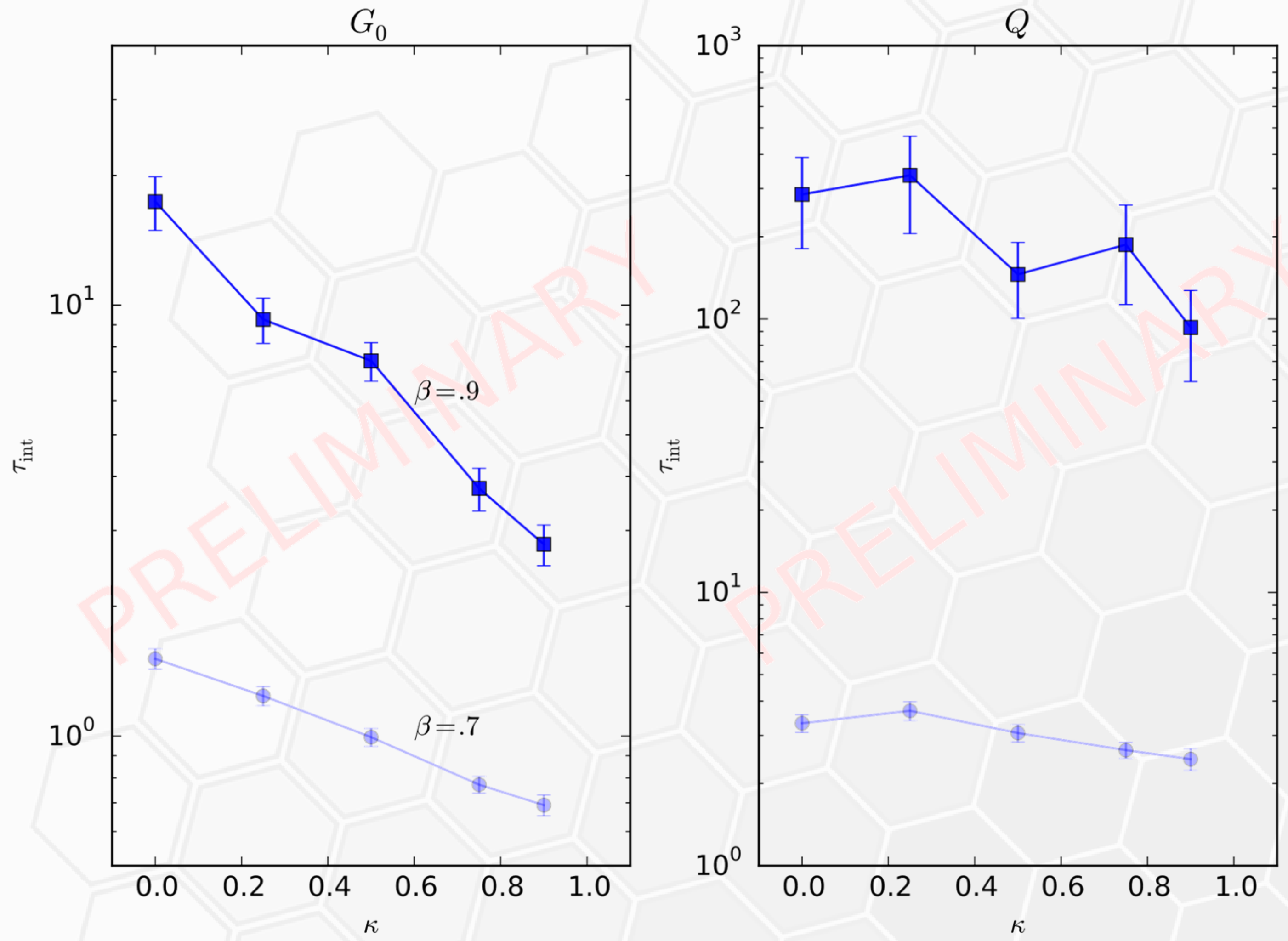
several βs and κ to check the acceleration efficiency in the continuum limit

Current status: RMHMC, CP^N

N=10
Showing 2 lattice spacings
$\beta = 0.7$ L=42
$\beta = 0.9$ L=90 statistics to be increased

Order of magnitude reduction of autocorrelation time observed in G_0

G_0 is the 2-point correlator (projected at zero momentum)
Q is the topological charge



Current status: SU(3)

RMHMC reminder: modification of the kinetic term in Hamiltonian $\frac{1}{2}\pi^T M^{-1}\pi$ $M\phi(x) = (1-\kappa)\phi(x) - \frac{\kappa}{4d}\nabla^2\phi(x)$

Pure gauge, Wilson action

$\beta = 6.2$ (a= 2.9 GeV^{-1} = 0.068 fm), L=16

$\beta = 6.4$ (a= 3.8 GeV^{-1} = 0.051 fm), L=32.

Reference: HMC 83-84% acceptance (Grid)

RMHMC overhead:

- Laplacian inversion ~20 iterations. Almost independent of *k*
- Implicit steps converge after 4-5 iterations

Observables:

- Topological charge and susceptibility
- $T_0 = 2\tau_{\text{flow}}^2 \frac{\langle E \rangle}{V}$

Current status: SU(3)

RMHMC reminder: modification of the kinetic term in Hamiltonian $\frac{1}{2}\pi^T M^{-1}\pi$ $M\phi(x) = (1-\kappa)\phi(x) - \frac{\kappa}{4d}\nabla^2\phi(x)$

Observables:

- Topological charge and susceptibility
- $T_0 = 2\tau_{\text{flow}}^2 \frac{\langle E \rangle}{V}$

Definition of cost related to the number of force computations (with an eye on simulations including fermions)

$$C(\mathcal{O}) = \frac{N_{\text{traj}} N_{\text{MD}}}{N_{\text{conf}}} \tau_{\text{acorr}}(\mathcal{O})$$

$\tau_{\text{acorr}}(\mathcal{O})$ is the integrated autocorrelation time for observable O

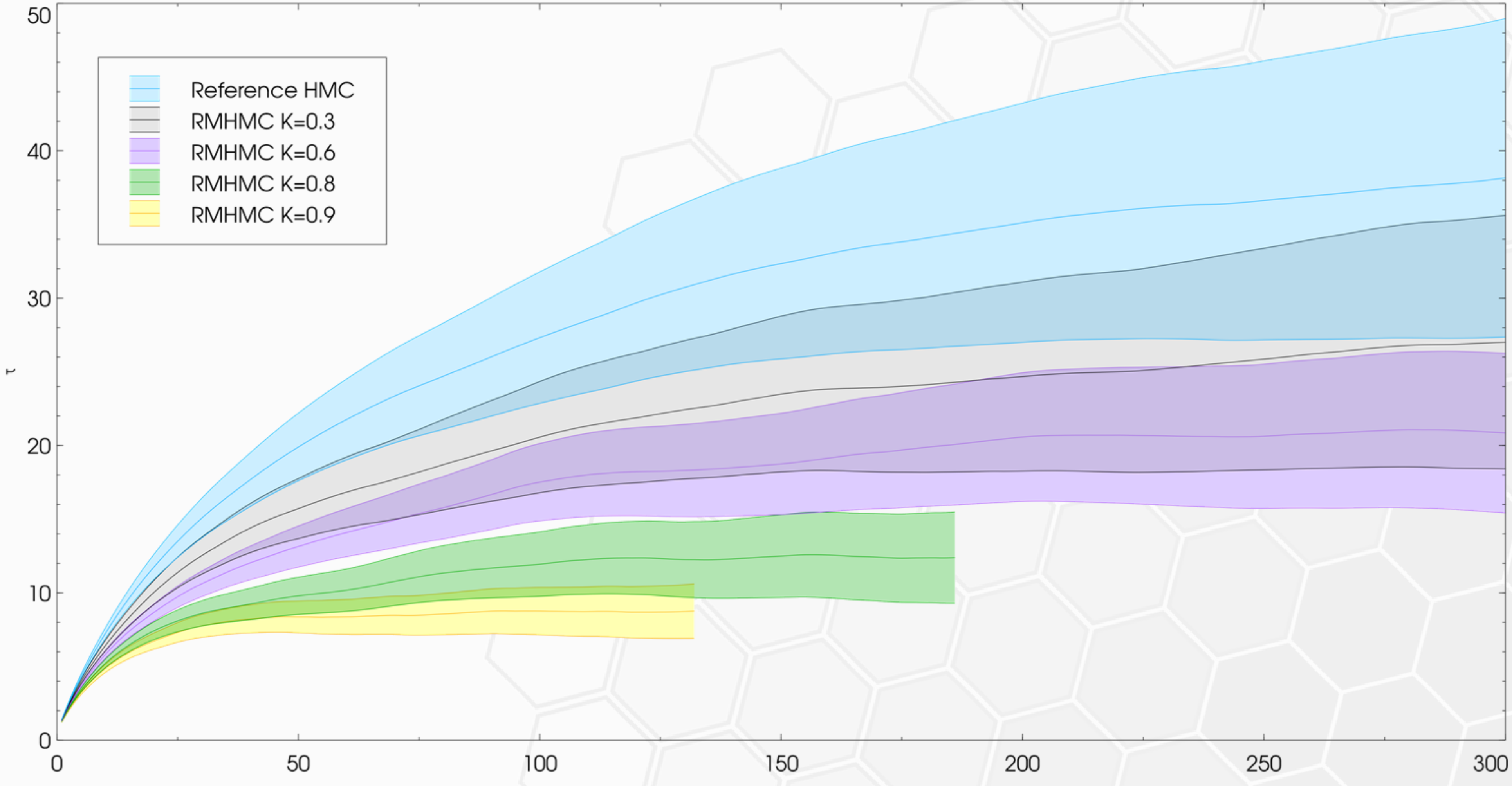
Trivial for HMC and RMHMC (N_{traj} = N_{conf}) but not for LAHMC.

Current status: RMHMC, SU(3)

Pure gauge, Wilson action, $\beta = 6.2$ L=16. Reference: HMC ~83% acceptance

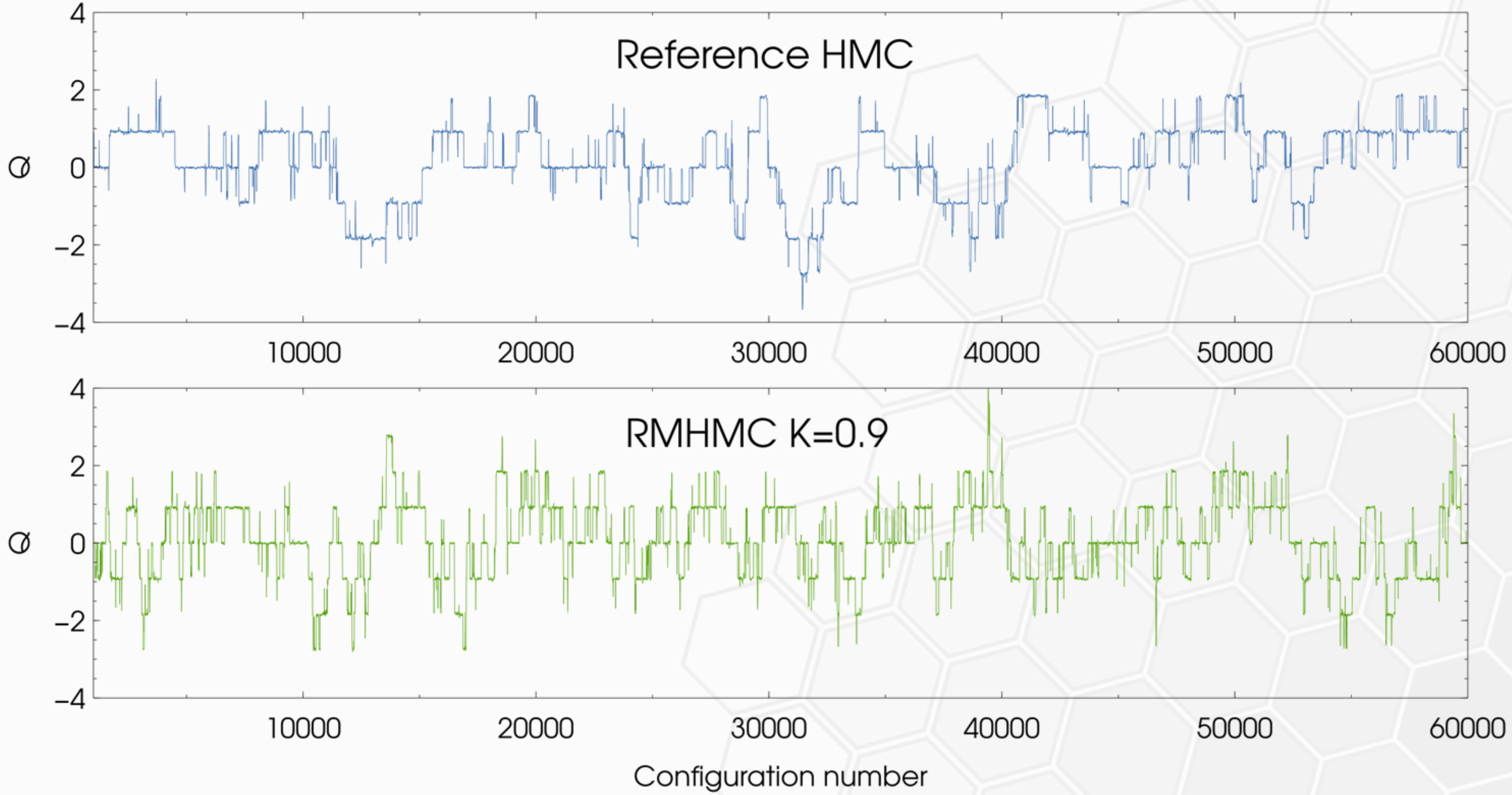
T_0

Integrated autocorrelation time



Current status: RMHMC, SU(3)

Pure gauge, Wilson action, $\beta = 6.2$ L=16. Reference: HMC ~83% acceptance

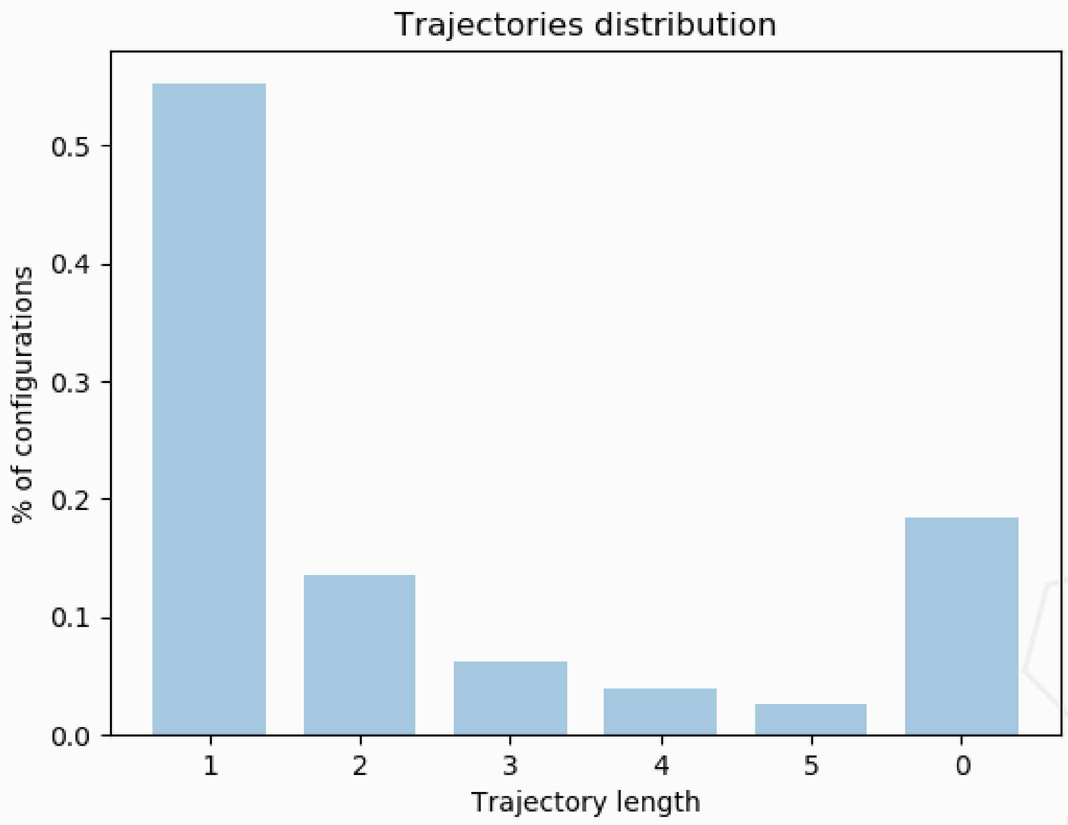


Current status: LAHMC

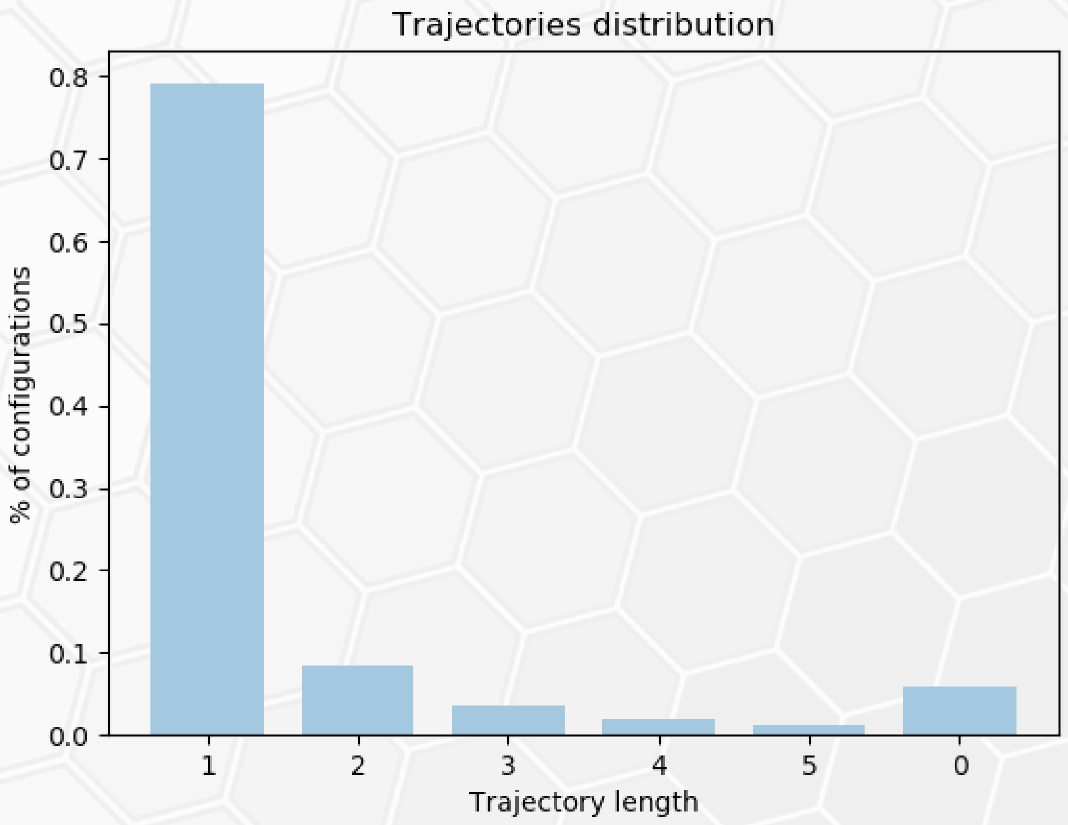
Pure gauge, Wilson action, $\beta = 6.2$ L=16, several pairs (K, α)

Effect on acceptance/rejection rate

MD steps = 10, K = 5



MD steps = 15, K = 5



Current status: LAHMC

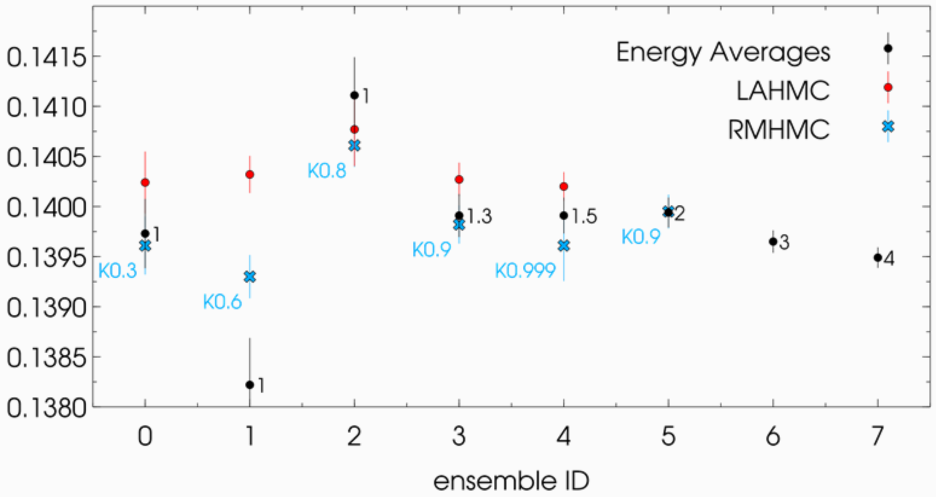
Pure gauge, Wilson action, $\beta = 6.2$ L=16, several pairs (K, α)

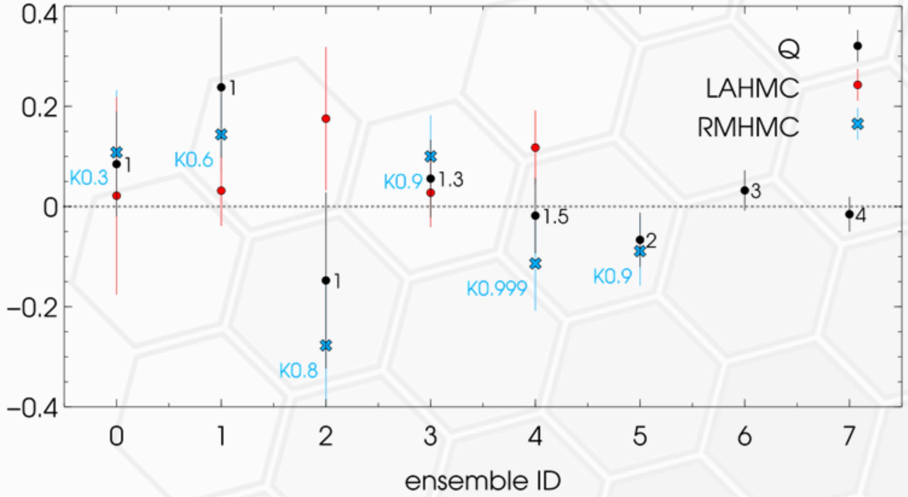
Effect on acceptance/rejection rate

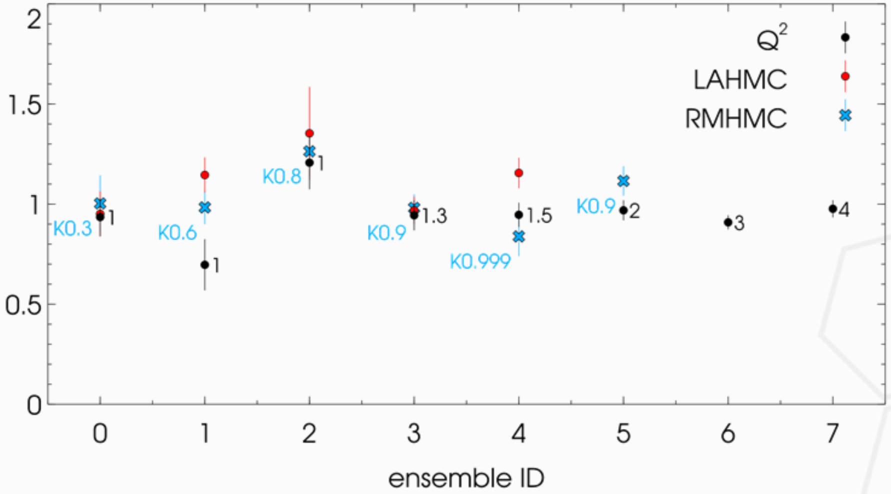
ID	L	β	N_{MD}	τ	K	α	dH 1	Acc.	Acc. 1	Avg τ	Avg N_{MD}
LA0	16	6.2	10	1.0	5	0.6	0.704 ± 0.0080	0.81	0.55	1.31	22.26
LA1	16	6.2	10	1.0	5	1.0	0.697 ± 0.0061	0.81	0.55	1.3	22.23
LA2	16	6.2	15	1.0	5	1.0	0.140 ± 0.0026	0.94	0.79	1.21	22.47
LA8	32	6.4	10	1.0	5	1.0	1.080 ± 0.0044	0.82	0.55	1.31	22.35
LA9	32	6.4	20	2.0	5	1.0	1.102 ± 0.0072	0.82	0.55	2.63	44.7

SU(3), coarse lattices summary

Pure gauge, Wilson action, $\beta = 6.2$ L=16. Reference: HMC ~83% acceptance





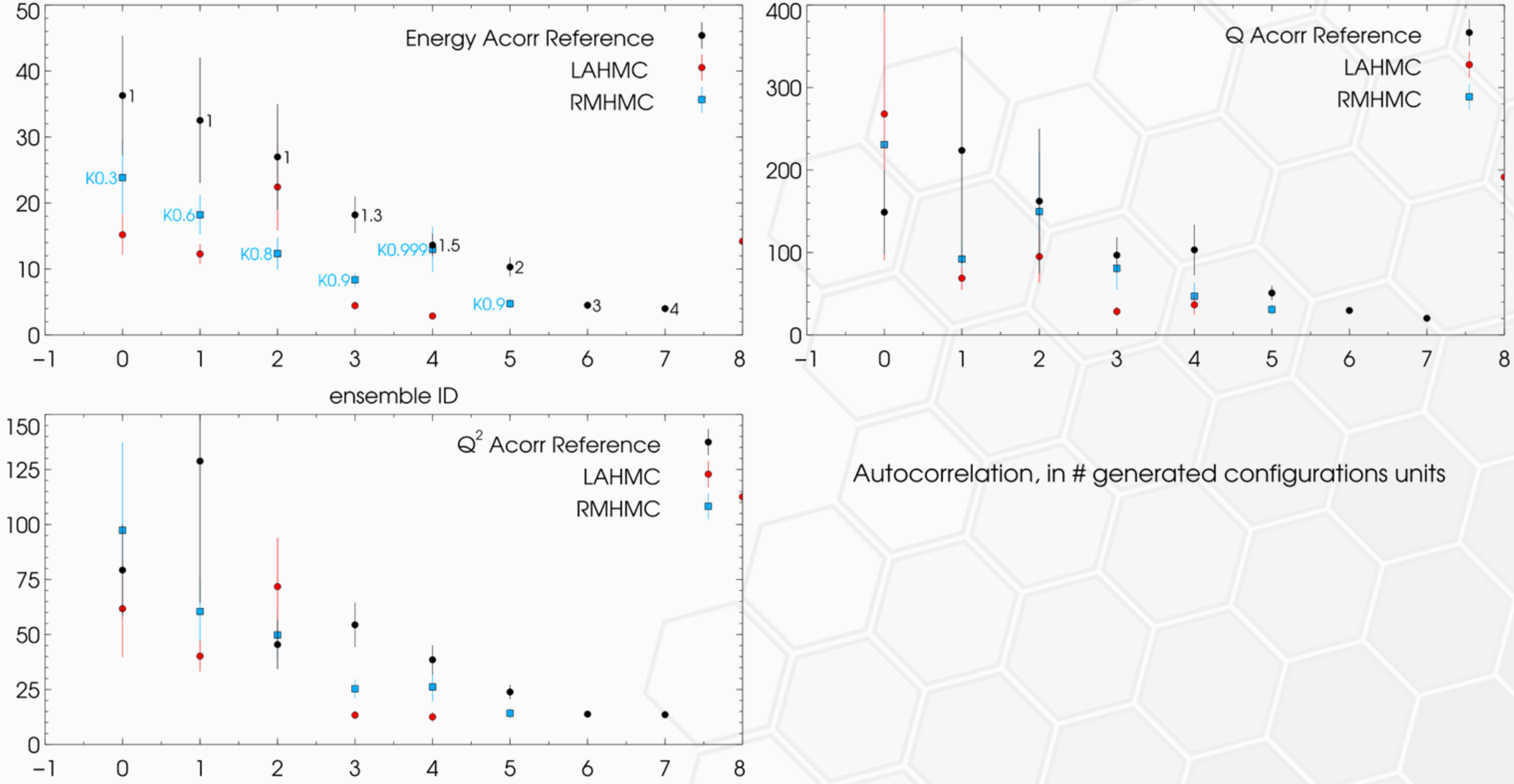


Averages: no biases

(ref runs 1,2 low stat. Not considered)

SU(3), coarse lattices summary

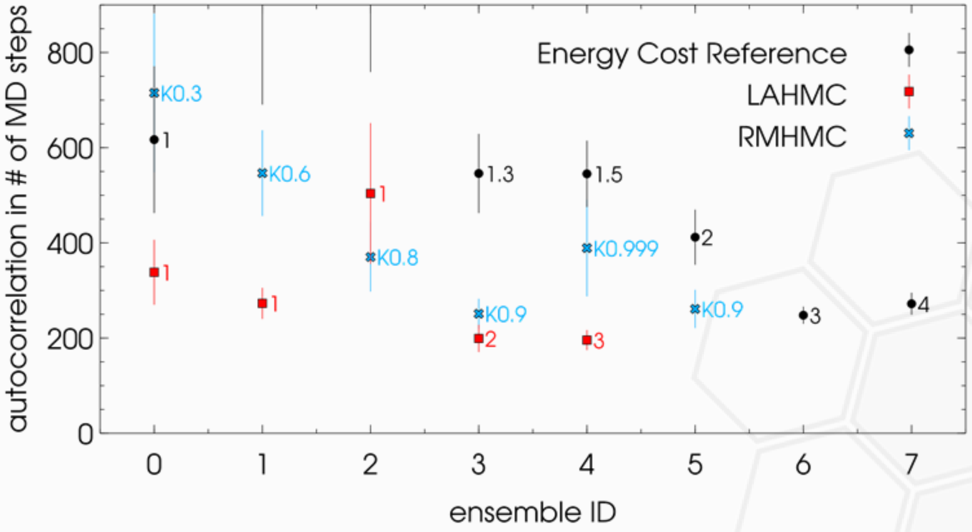
Pure gauge, Wilson action, $\beta = 6.2$ L=16. Reference: HMC ~83% acceptance

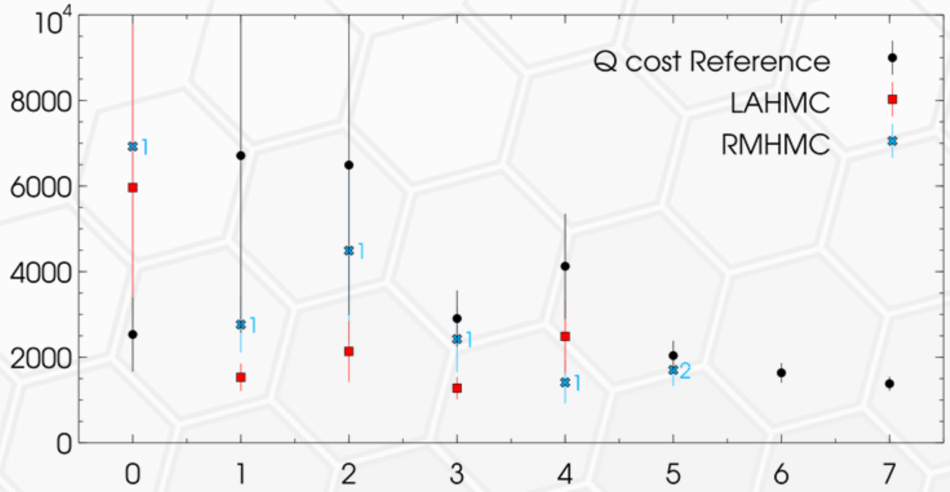


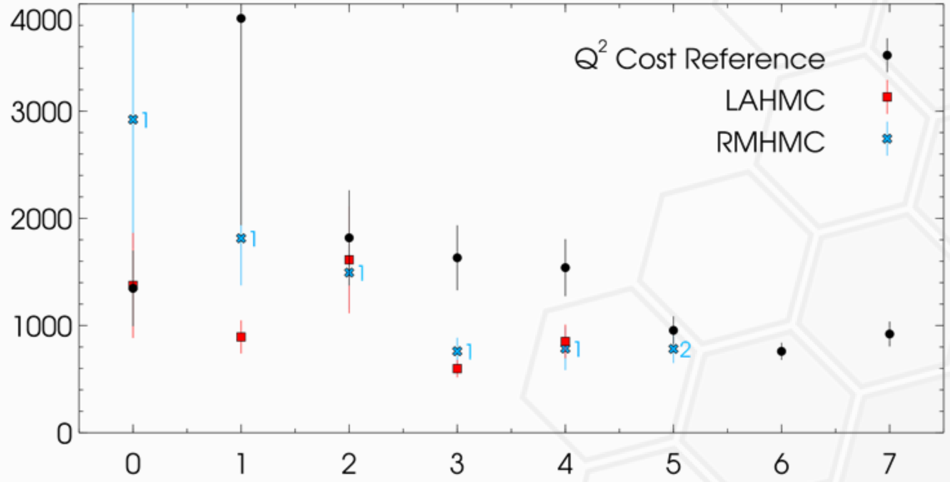
Autocorrelation, in # generated configurations units

SU(3), coarse lattices summary

Pure gauge,
Wilson action
$\beta = 6.2$ L=16





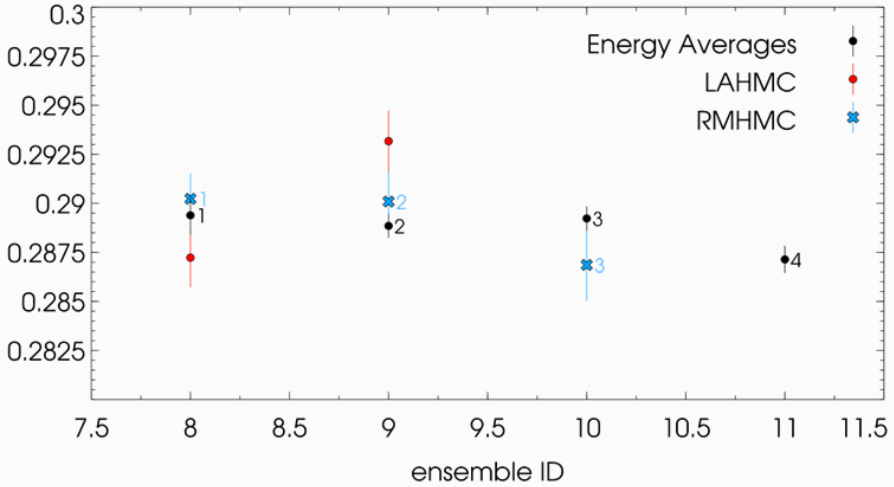


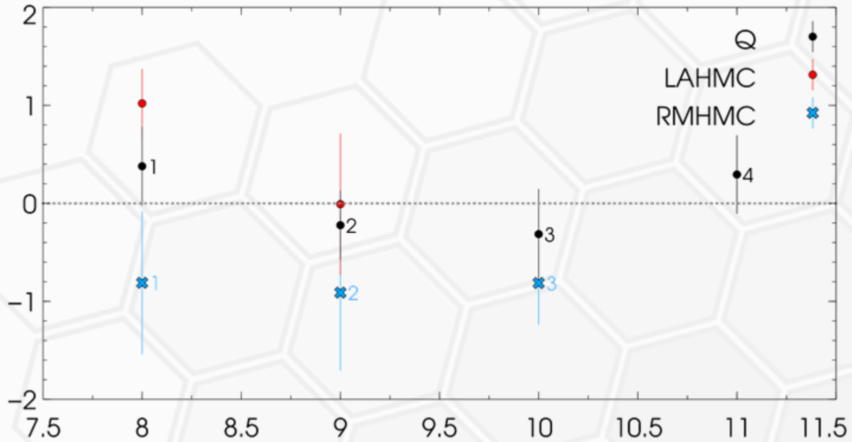
Cost in units of force computations (MD steps)

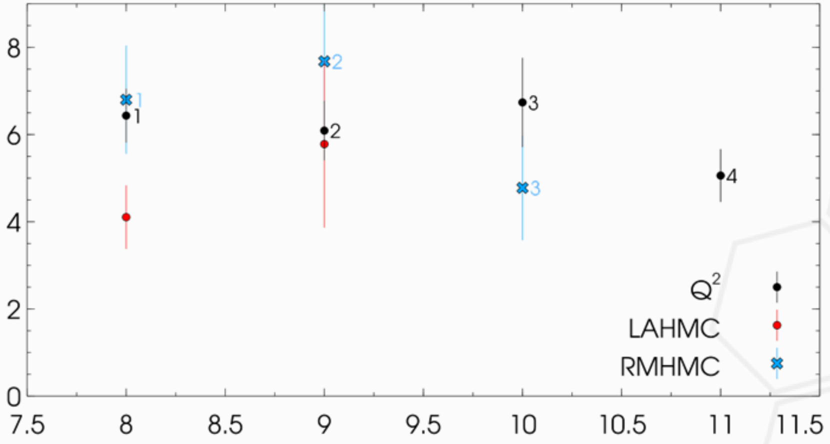
COST ESTIMATE

SU(3), fine lattices summary

Pure gauge, Wilson action, $\beta = 6.4$ L=32. Reference: HMC ~84% acceptance





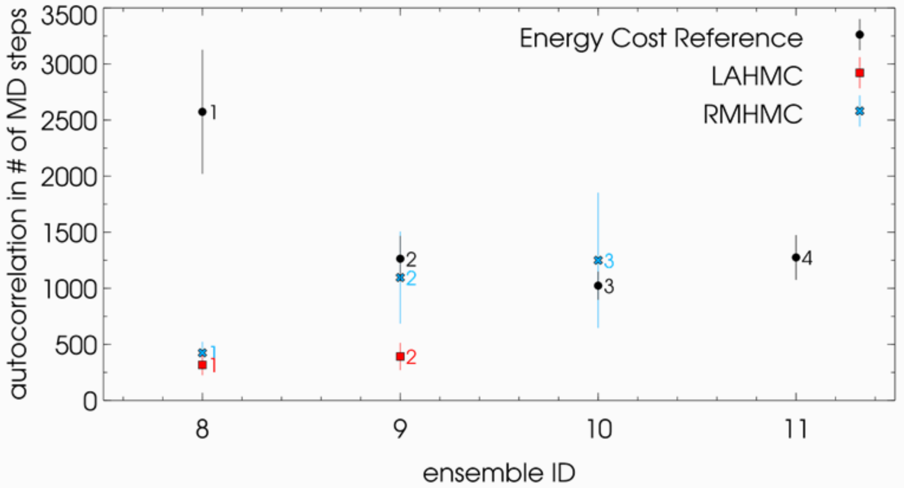


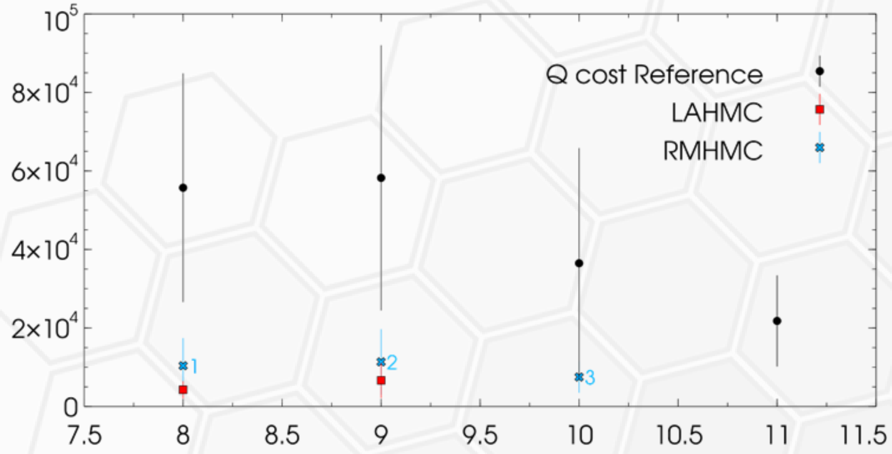
Averages L=32 β=6.4

WORK IN PROGRESS

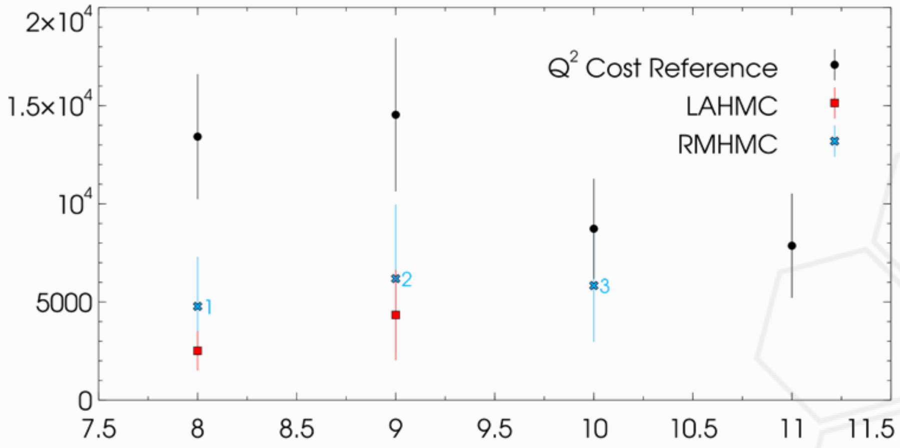
SU(3), fine lattices summary

Pure gauge, Wilson action, $\beta = 6.4$ L=32. Reference: HMC ~84% acceptance





Cost in units of force computations (MD steps) L=32 β=6.4



WORK IN PROGRESS

Outlook

Currently finalizing the scaling study towards the continuum

- Positive (not yet conclusive) results, quick to implement
- RMHMC: study other "acceleration" operators, e.g. Laplacian with a spectral bound (using Chebyshev polynomials for example)
- Methods can be combined, working in "orthogonal" directions
- Useful interaction between the ML and Statistical physics community

GRID and evolution algorithms for critical slowing down

Guido Cossu

in collaboration with P. Boyle, N. Christ, C. Jung, A. Jüttner, F. Sanfilippo

Frontiers in Lattice quantum field theory 2018 – IFT Madrid May 31th

EXASCALE COMPUTING PROJECT

THE UNIVERSITY OF EDINBURGH