Evolution algorithms for critical slowing down

Guido Cossu

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

Lattice 2017 - Granada

Critical slowing down

Continuum limit of lattice QFT: second order critical point

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

The exponent depends on the:
- observable
- algorithm to generate the Markov Chain

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 $p(x) = \frac{1}{Z} \exp\left(-S(x)\right)$

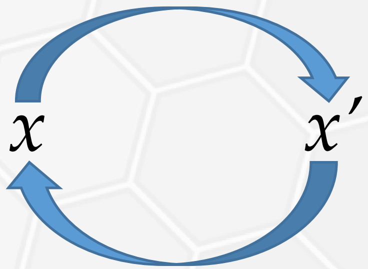
Generate a Markov chain $x_0 \to x_1 \to \ldots \to x_N$

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

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

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

Hybrid Monte Carlo

Hamiltonian evolution
Duane et al. 1987

$$H(x) = \frac{1}{2}p^T p + S(x)$$

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

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

$x^{(t,1)} = x'$ 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: $(FL)^{-1} = L^{-1}F = FL$

Markov transitions

Hamiltonian integration
M steps, ε step size
Symplectic, reversible

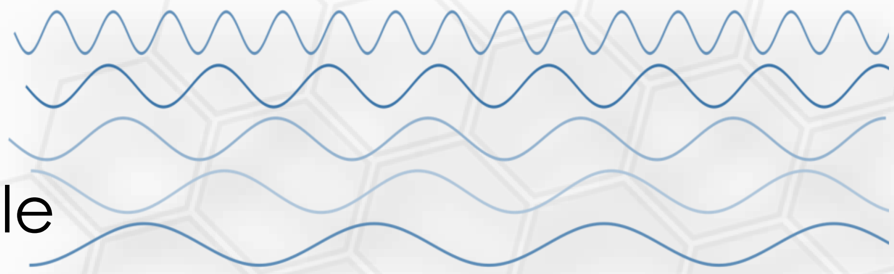
$$L(M,\epsilon)$$

Momenta flip F

Momenta randomization
α mixing factor

$$R(\alpha)$$

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 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 $\frac{1}{2}\theta^T M^2\theta + \theta^2$ 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 Runge-Kutta integration)

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

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

$$x' = FL(M,\epsilon)x^{(t,0)} \qquad x^{(t,1)} = x' \quad \text{Accept} \qquad x^{(t,1)} = x^{(t,0)} \quad \text{Reject}$$

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

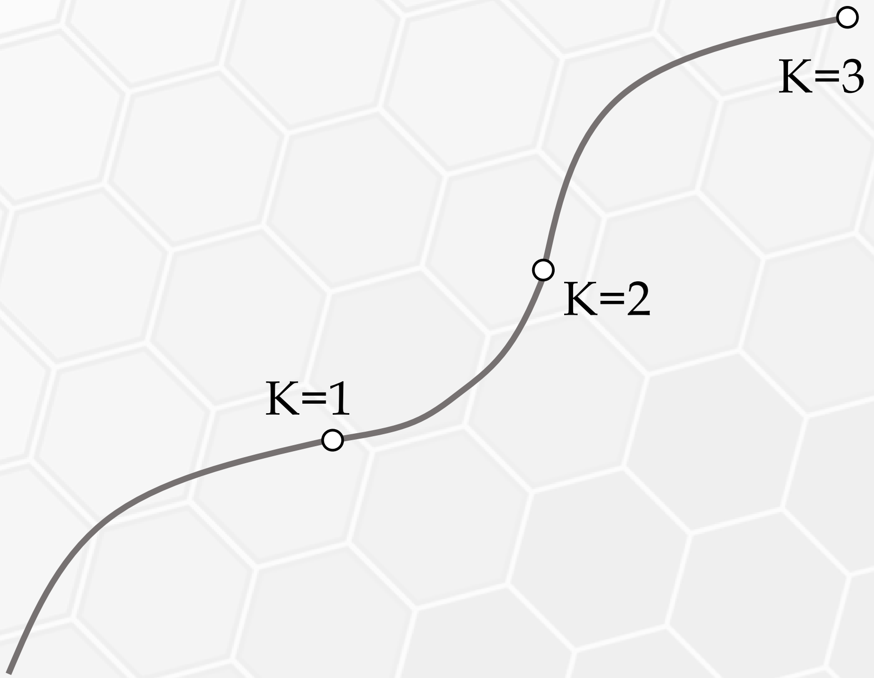
$$x^{(t,2)} = Fx^{(t,1)}$$

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

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

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

Look Ahead HMC

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

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

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

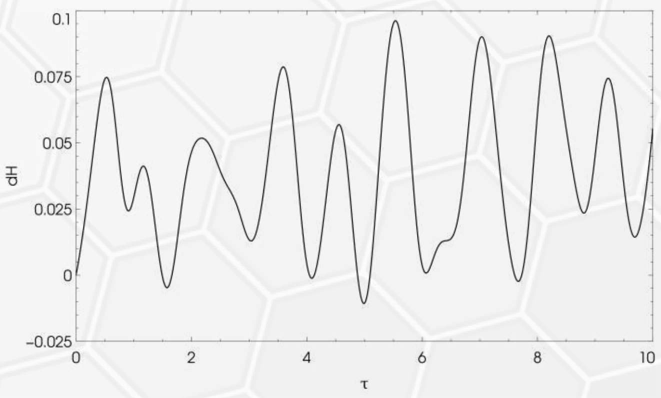
$$\pi_{L^a}(x) = \min\left[1 - \sum_{b<a} \pi_{L^b}(x), \frac{p(FL^a x)}{p(x)}\left(1 - \sum_{b<a} \pi_{L^b}(FL^a x)\right)\right]$$

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

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

Currently O(10K) trajectories or less

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

CP^N model (Jüttner, Sanfilippo)

Model has severe slowing down at small lattice spacing

Simulating N=10

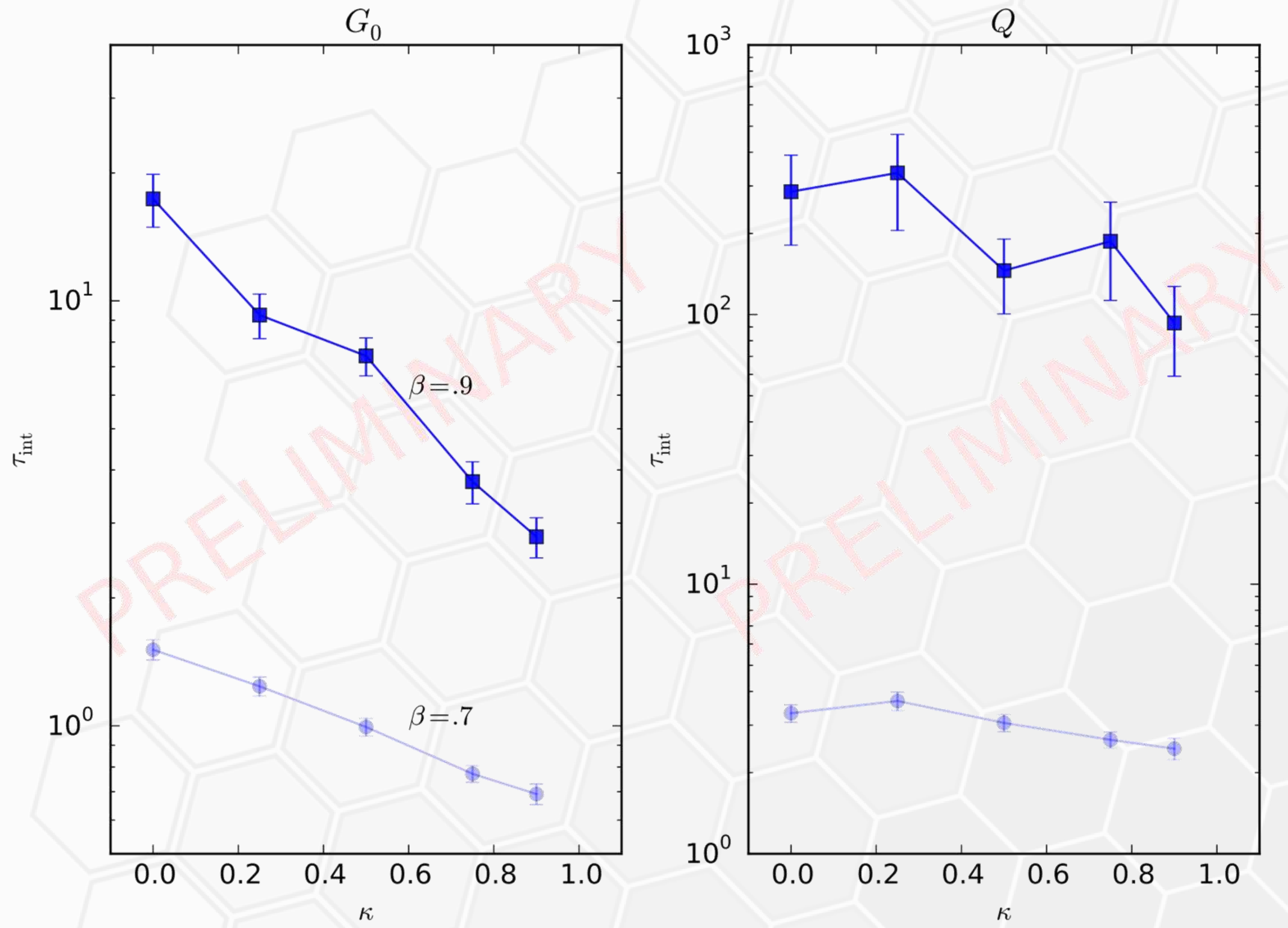
several βs and κ to check the acceleration efficiency in the continuum limit

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

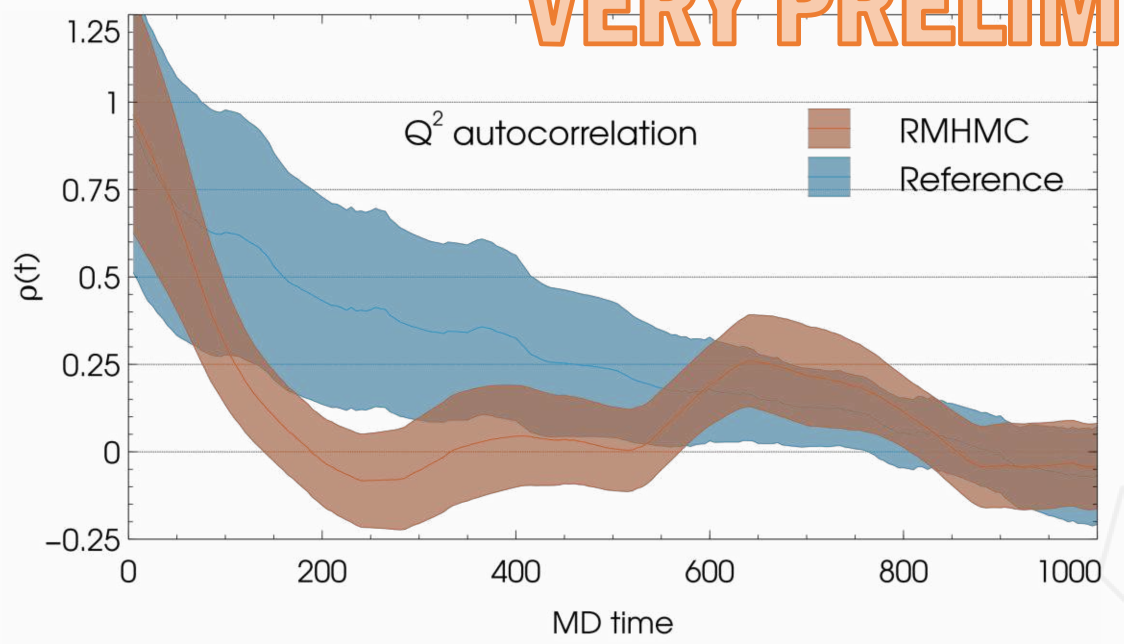


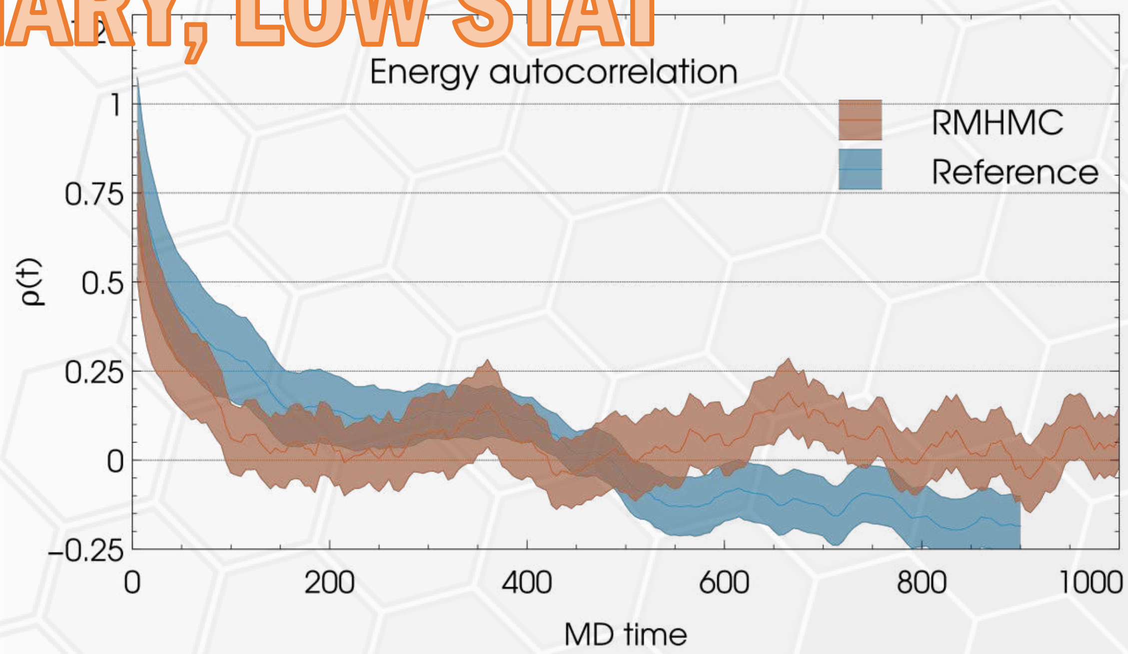
Current status: RMHMC, SU(3)

Pure gauge, Wilson action, $\beta = 6.2$ L=16, $\kappa = 0.999$. Reference: HMC ~95% acceptance (Grid)
RMHMC overhead:

- Laplacian inversion 25 iterations
- Implicit steps converge after 4-5 iterations

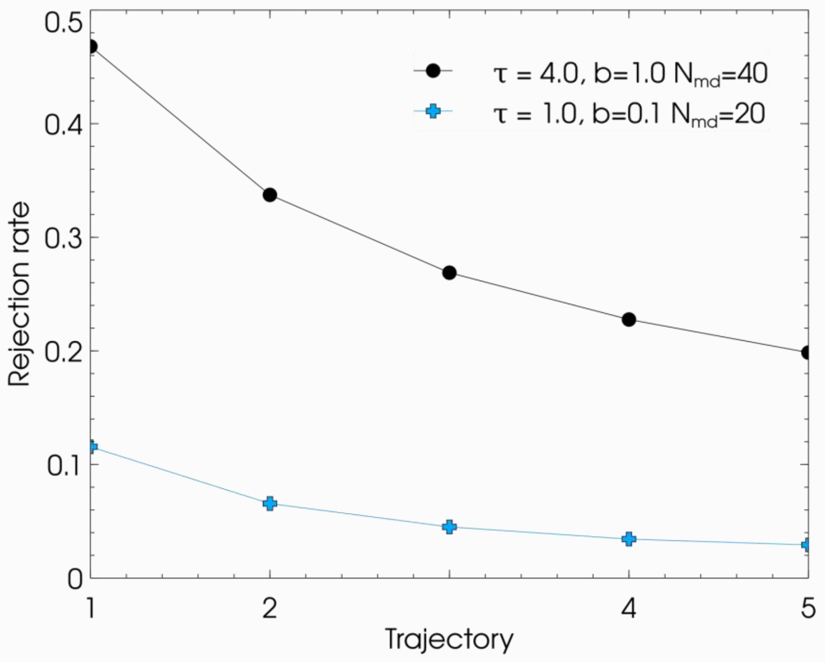
VERY PRELIMINARY, LOW STAT

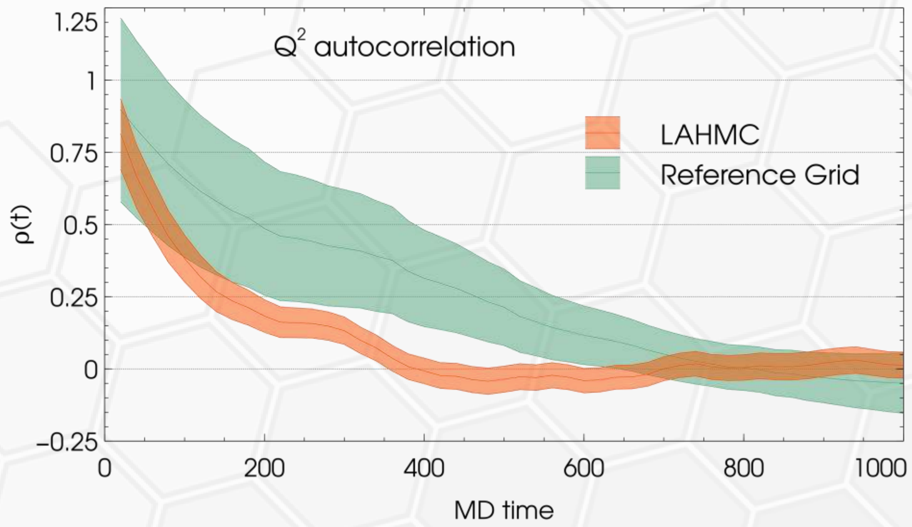


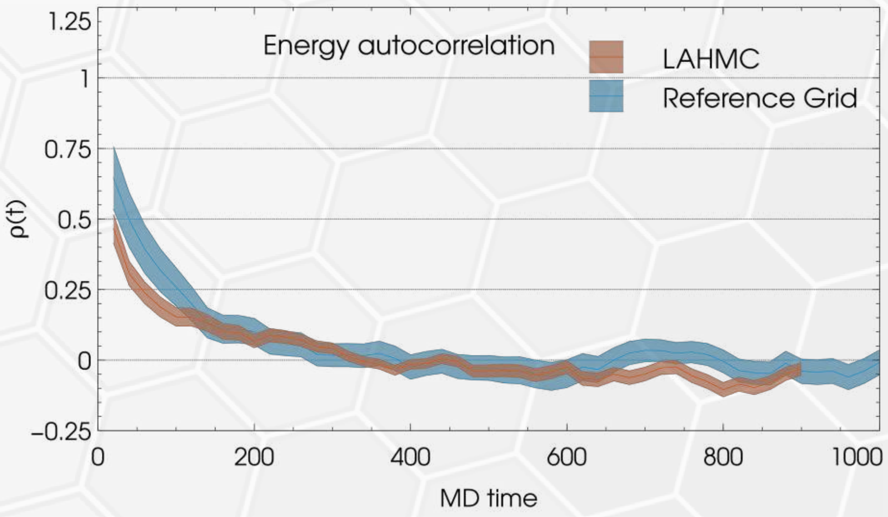


Current status: LAHMC

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







VERY PRELIMINARY, LOW STAT

Summary

Currently in the stage of investigation

Promising directions but lot of work to do

- Increase statistics
- Span a larger parameter space in both algorithms, optimize
- RMHMC: study other “acceleration” operators, e.g. Laplacian with a spectral bound (using Chebyshev polynomials for example)
- RMHMC, LAHMC: cost estimate

Evolution algorithms for critical slowing down

Guido Cossu

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

Lattice 2017 - Granada