Stabilised Wilson fermions for QCD

Antonio Rago

University of Plymouth

13^{th} January 2020

In collaboration with A. Francis, P. Fritzsch and M. Lüscher

Motivations

- A large part of the success of Lattice Gauge Theory is inherently tied with advances in Monte Carlo simulations.

- Monte Carlo methods used in Lattice Gauge Theory are importance sampling methods.

  - Generating an ensemble of configurations through a Markov process and estimating the expectation values on the ensemble averages.

  - On very large lattices translation averages in presence of a single gauge field (the master field) provide an alternative way of calculating the expectation values.

- However, when the gap of the lattice Dirac operator shrinks, algorithmic instabilities and precision issues hamper the stability of the configurations generation and affect the estimate of observables.

Ways to overcome these problems are described in this talk for the case of the O(a)-improved Wilson formulation of lattice QCD.

Identifying the critical aspects

- Algorithmic stability:
  - Update algorithm: Hybrid Monte-Carlo.
  - Integration schemes.
  - Global Metropolis accept-reject step.
- Fermion discretisation:
  - Spectral gap of Dirac operator.
  - Near zero-modes: MD evolution of smallest eigenvalue.
  - Solver stopping criteria.

All the above have a strong influence on the simulation cost and affects the reliability of the simulation.

Summary: Stabilized Wilson fermions

The proposed stabilizing measures include:

1. A modification of the standard O(a)-improved lattice Dirac operator.
2. The use of the Stochastic Molecular Dynamics (SMD) simulation algorithm.
3. Tuning of the numerical precision required to guarantee a sufficient level of accuracy on large lattices.

I will present results of some representative simulations of the theory with $2+1$ flavours of quarks, to demonstrate the viability of the framework.

$O(a)$-improvement revisited

- The traditional Wilson Dirac $O(a)$-improved operator is

$$D = \frac{1}{2}\{\gamma_\mu(\nabla^*_\mu + \nabla_\mu - a\nabla^*_\mu\nabla_\mu)\} + c_{\rm sw}\frac{i}{4}\sigma_{\mu\nu}\hat{F}_{\mu\nu} + m_0.$$

- If the lattice points are classified as even-odd

$$D = \begin{pmatrix} D_{ee} & D_{eo} \\ D_{oe} & D_{oo} \end{pmatrix}$$

with the diagonal part

$$D_{ee} + D_{oo} = 4 + m_0 + c_{\rm sw}\frac{i}{4}\sigma_{\mu\nu}\hat{F}_{\mu\nu}.$$

- The even-odd preconditioned form

$$\hat{D} = D_{ee} - D_{eo}\,(D_{oo})^{-1}\,D_{oe}.$$

$O(a)$-improvement revisited

- The coefficient $c_{\rm sw}$ is equal to 1 at tree-level PT and grows monotonically with the gauge coupling (~ 2 on coarse lattices).
- The Pauli term in these equations can be fairly large, particularly so on coarse lattices (saturating the bound).

$$\left|\left|\frac{i}{4}\sigma_{\mu\nu}\hat{F}_{\mu\nu}\right|\right|_2 \leq 3$$

- Positive and negative eigenvalues of the Pauli term are equally distributed.
  D_{oo} is not protected by small eigenvalues especially for small masses and rough gauge fields.
- EO preconditioning occasionally fails with probability growing with the lattice size.
- Impossible to use in master-field simulations.

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu} \right).$$

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu}\right).$$

- At leading order of perturbation theory this expression actually coincides with the traditional.

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu}\right).$$

- At leading order of perturbation theory this expression actually coincides with the traditional.

- The diagonal part of the Dirac operator is positive definite and safely invertible.

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu}\right).$$

- At leading order of perturbation theory this expression actually coincides with the traditional.
- The diagonal part of the Dirac operator is positive definite and safely invertible.
- Even-odd preconditioning is therefore guaranteed to be numerically unproblematic.

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu}\right).$$

- At leading order of perturbation theory this expression actually coincides with the traditional.
- The diagonal part of the Dirac operator is positive definite and safely invertible.
- Even-odd preconditioning is therefore guaranteed to be numerically unproblematic.
- Moreover, $\det \mathrm{D} = \det \hat{\mathrm{D}}$ up to a field-independent proportionality constant.

$O(a)$-improvement revisited

The improved Wilson-Dirac is not positive: is this why it tends to promote the instabilities?

- An alternative definition of the Wilson Dirac $O(a)$-improved operator is

$$D_{ee} + D_{oo} = (4 + m_0) \exp\left(\frac{\mathrm{c_{sw}}}{4 + \mathrm{m_0}} \frac{\mathrm{i}}{4} \sigma_{\mu\nu} \hat{\mathrm{F}}_{\mu\nu}\right).$$

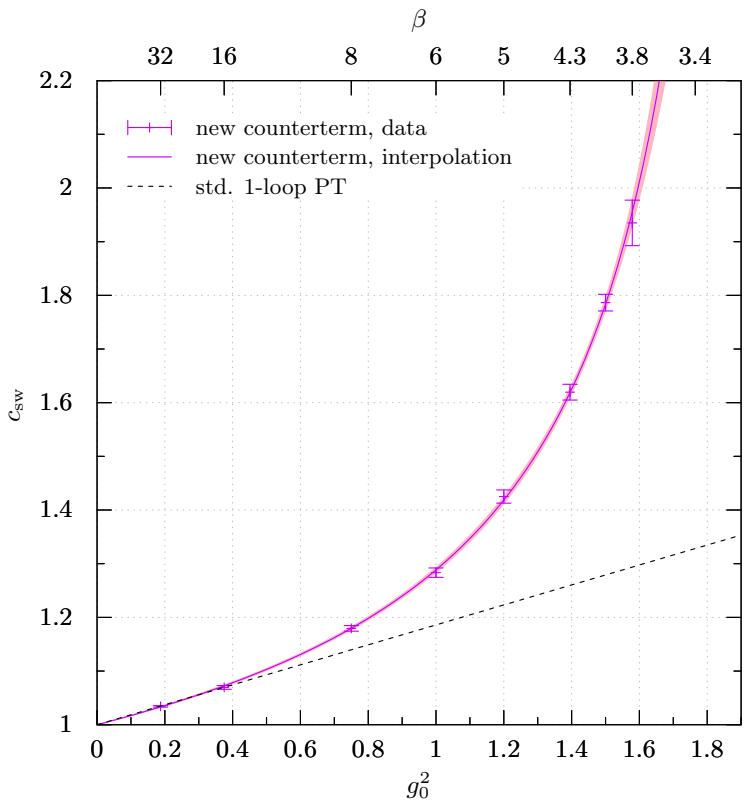
- At leading order of perturbation theory this expression actually coincides with the traditional.
- The diagonal part of the Dirac operator is positive definite and safely invertible.
- Even-odd preconditioning is therefore guaranteed to be numerically unproblematic.
- Moreover, $\det \mathrm{D} = \det \hat{\mathrm{D}}$ up to a field-independent proportionality constant.
- The exponential and the associated force can be evaluated with negligible computational effort.

Improved $c_{\rm SW}$ tuning

Is this a viable choice of Dirac-Wilson improvement?

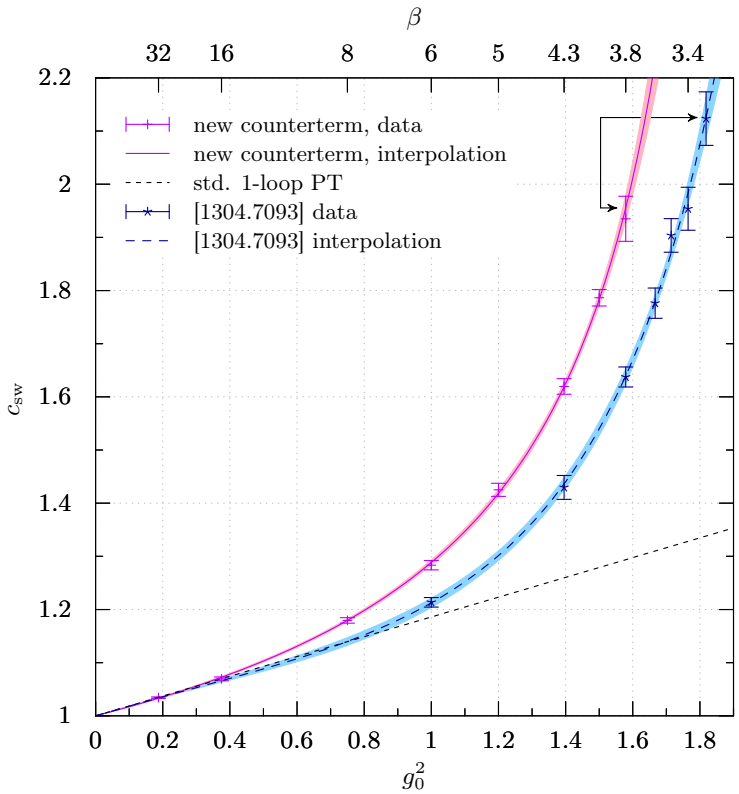
We need extensive simulations of the modified theory.

- $N_f = 2 + 1$ QCD simulations with a tree level improved Symanzik action.
- Tuning of the $c_{\rm SW}$ through the standard massless Schröedinger Functional scheme
- Scan up to very large β to make contact with PT.



Comparison with the traditional $c_{\rm SW}$ term

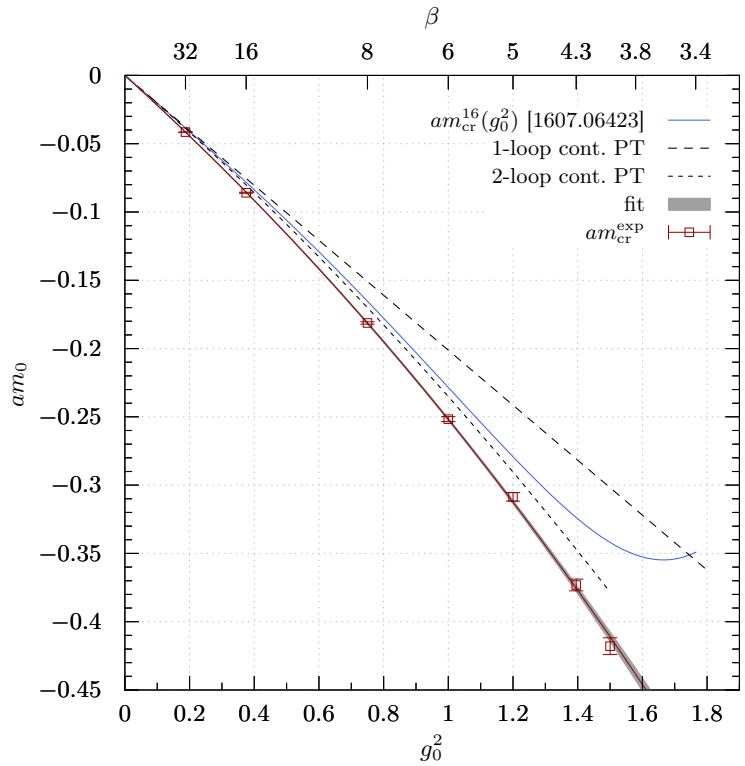
- How does it compare with the traditional $c_{\rm SW}$ tuned values?



- The scale setting is different. Arrows indicate $a \sim 0.095\text{fm}$
- For equal lattice spacing $c_{\rm SW}^{\rm new} < c_{\rm SW}^{\rm old}$
- A similar outcome is obtained also for the quenched theory

Comparison of the critical mass

- It is analogously possible to compare the critical mass: $am_{cr} = \frac{1}{2\kappa_{cr}} - 4$



- Also in this case the scale setting is different.
- A similar outcome is obtained also for the quenched theory

The SMD algorithm

The SMD algorithm is rather similar to the HMC algorithm.

Start with:
$U(x,\mu)$, the momentum $\pi(x,\mu)$ and the pseudo-fermion $\phi(x)$ with action $S_{\rm pf} = \phi(D^\dagger D)^{-1})\phi$

One cycle consists of:

- A random rotation of the momentum and the pseudo-fermion

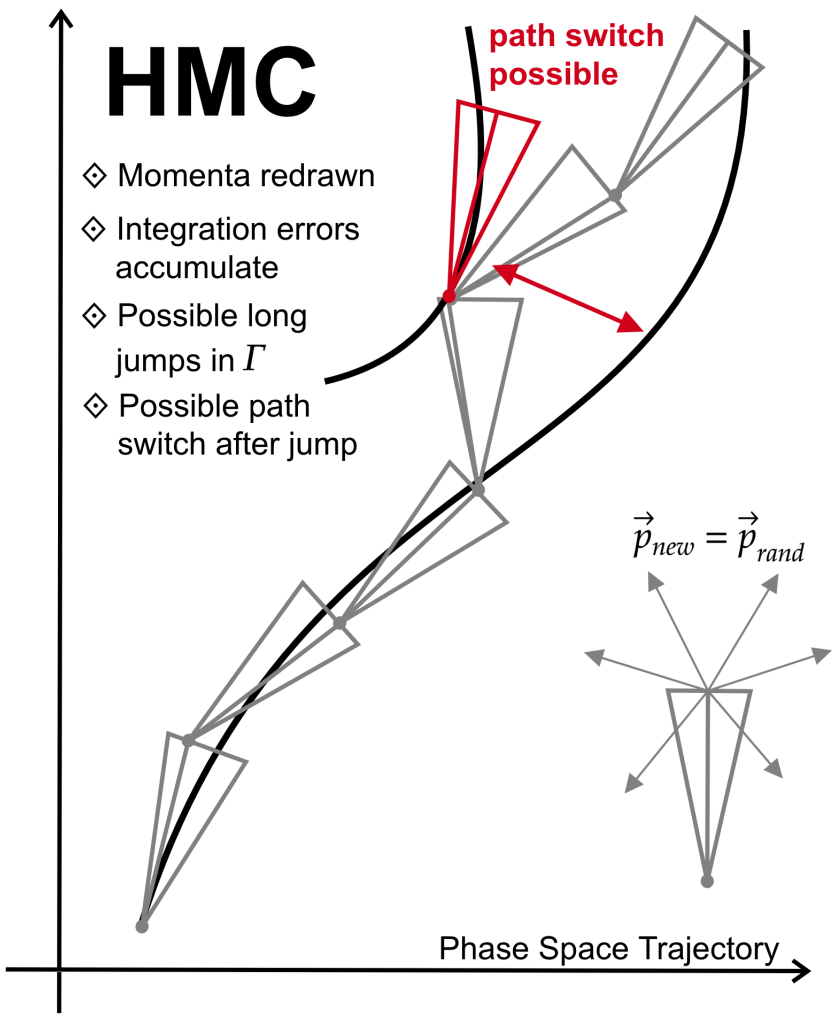
$$\pi \to c_1\pi + c_2 v \qquad\qquad \phi \to c_1\phi + c_2 D^\dagger \eta$$

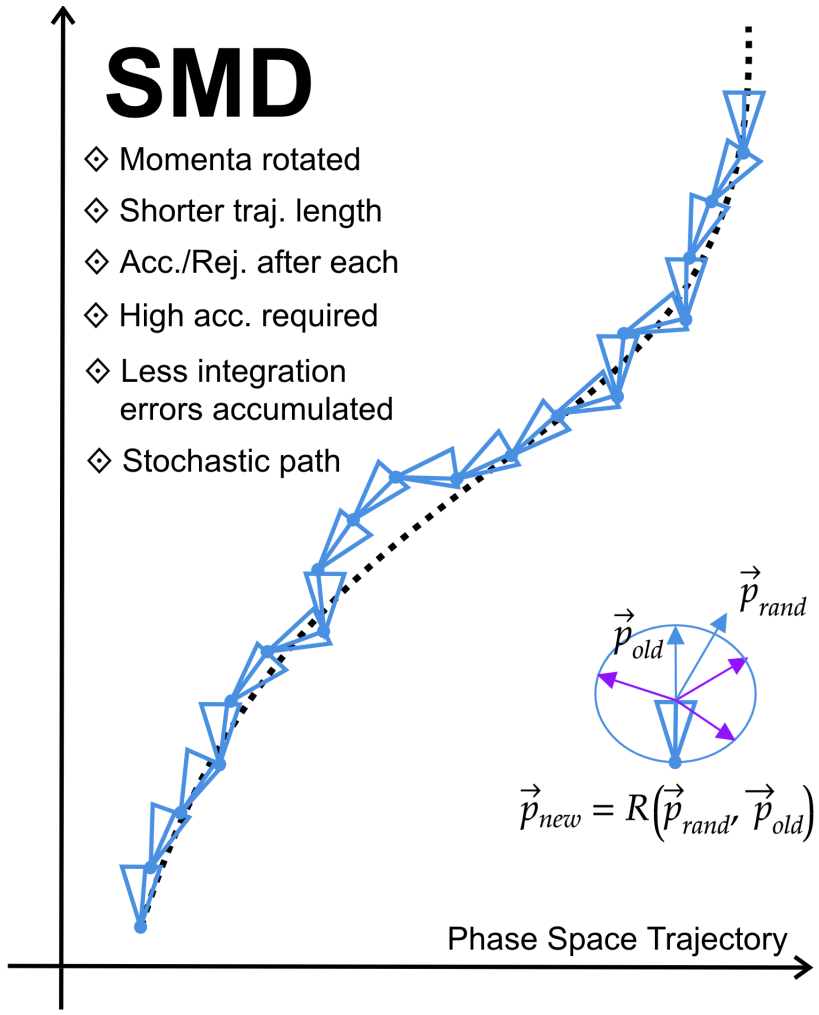
  with v and η random normally distributed, $c_1^2 + c_2^2 = 1$ and $c_1 = e^{-\epsilon\gamma}$.
  Where ϵ is the MD integration time and γ is the friction parameter.

- A short molecular-dynamics evolution

- an accept-reject step that makes the algorithm exact

The SMD algorithm





Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm

Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm
- If ϵ is sufficiently small, the SMD algorithm can be rigorously shown to be ergodic and to converge to a unique stationary state simulating the canonical distribution.

Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm
- If ϵ is sufficiently small, the SMD algorithm can be rigorously shown to be ergodic and to converge to a unique stationary state simulating the canonical distribution.
- When field configurations are rejected, the momentum is reversed and the algorithm tends to backtrack its trajectory in field space. Rejections must occur at large separations in simulation time.

Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm
- If ϵ is sufficiently small, the SMD algorithm can be rigorously shown to be ergodic and to converge to a unique stationary state simulating the canonical distribution.
- When field configurations are rejected, the momentum is reversed and the algorithm tends to backtrack its trajectory in field space. Rejections must occur at large separations in simulation time.
- As ΔH grows like $V^{1/2}$, to preserve acceptance the integration has to become more accurate with the volume. It is hence convenient to use integration rule of high order as the computational cost will scale with $V^{1/(2p)}$

Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm
- If ϵ is sufficiently small, the SMD algorithm can be rigorously shown to be ergodic and to converge to a unique stationary state simulating the canonical distribution.
- When field configurations are rejected, the momentum is reversed and the algorithm tends to backtrack its trajectory in field space. Rejections must occur at large separations in simulation time.
- As ΔH grows like $V^{1/2}$, to preserve acceptance the integration has to become more accurate with the volume. It is hence convenient to use integration rule of high order as the computational cost will scale with $V^{1/(2p)}$
- While the unit of MD the SMD tends to be slower than the HMC, the difference is (largely) compensated by a shorter autocorrelation time.

Considerations on the SMD algorithm

- At fixed ϵ and large γ, the SMD algorithm coincides with the HMC algorithm
- If ϵ is sufficiently small, the SMD algorithm can be rigorously shown to be ergodic and to converge to a unique stationary state simulating the canonical distribution.
- When field configurations are rejected, the momentum is reversed and the algorithm tends to backtrack its trajectory in field space. Rejections must occur at large separations in simulation time.
- As ΔH grows like $V^{1/2}$, to preserve acceptance the integration has to become more accurate with the volume. It is hence convenient to use integration rule of high order as the computational cost will scale with $V^{1/(2p)}$
- While the unit of MD the SMD tends to be slower than the HMC, the difference is (largely) compensated by a shorter autocorrelation time.

Significant reduction of the unbounded energy violations $|\Delta H| \gg 1$

Other algorithmic improvements

- Convergence criterion for the solver
  The solver uses an iterative procedure that is being stopped when the approssimate solution $\tilde{\psi}$ satisfies

$$||\eta - D\tilde{\psi}||_2 \leq w||\eta||_2 \qquad \text{with} \qquad ||\eta||_2 \propto V$$

- Global reductions
  Sum over all lattice points can cause accumulations errors.

$$\Delta H \propto \sqrt{V} \qquad \text{the numerical precision has to increase with V.}$$

Other algorithmic improvements

- Convergence criterion for the solver
  The solver uses an iterative procedure that is being stopped when the approssimate solution $\tilde{\psi}$ satisfies

$$||\eta - D\tilde{\psi}||_2 \leq w||\eta||_2 \qquad \text{with} \qquad ||\eta||_2 \propto V$$

  ✓ Replaced with the uniform norm: $||\eta||_\infty = \sup_x ||\eta(x)||_2$ (V-independent)

- Global reductions
  Sum over all lattice points can cause accumulations errors.

$$\Delta H \propto \sqrt{V} \qquad \text{the numerical precision has to increase with V.}$$

  ✓ Use quadruple precision in global sums

Ongoing investigations

Once $c_{\rm sw}$ is tuned and with all the algorithmic measures in place

We performed a set of (2+1)-flavour simulations

a/fm	β	$T \times L^3$	m_π/MeV	m_K/MeV	Lm_π	BC	status
0.095	3.8	96×32^3	410	410	6.3	P	✓
		96×32^3	294	458	4.5	P	✓
		96×32^3	220	478	3.4	P	✓
		144×64^3	135	494	4.2	P	planned
0.064	4.0	96×48^3	410	410	6.4	P	✓
0.055	4.1	96×48^3	410	410	5.5	O	thermalized

The runs are in direct comparison with ensambles from Coordinated Lattice Simulations (CLS) collaboration.

- The scale setting is on the symmetric point for the continuum limit:

$$\phi_4 \equiv 8t_0(\frac{1}{2}m_\pi^2 + m_K^2) = const \propto \mathrm{tr}(M_q)$$

- We have also planned a single β set of runs up to the physical point on the coarsest lattice.

Details of the SMD runs

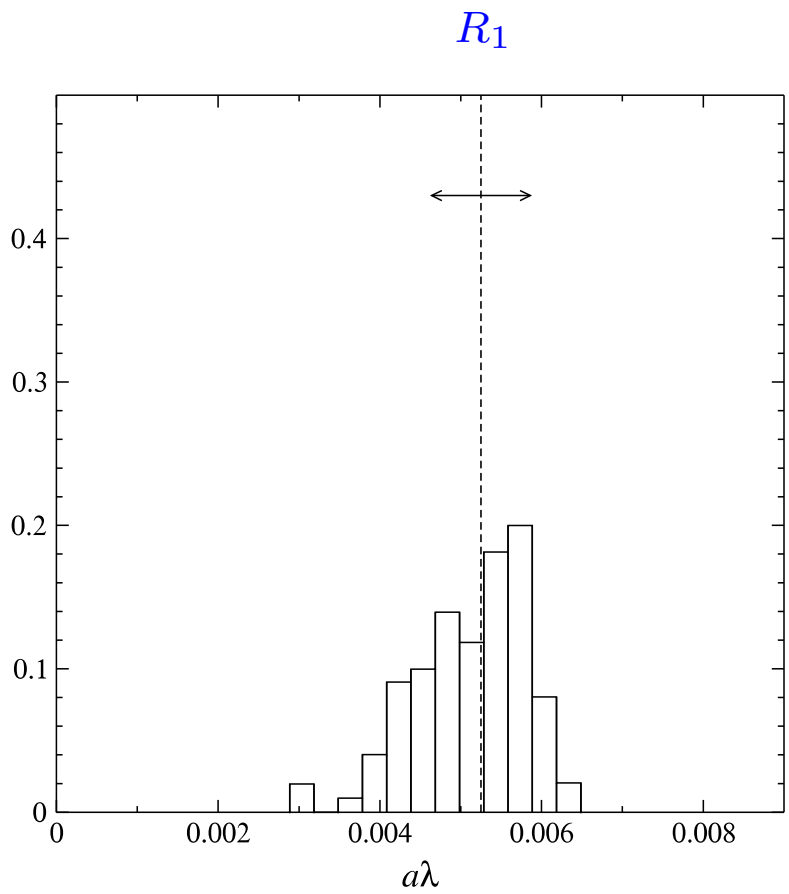
- The runs have shown no issues of instability
- Parameters setup: $\gamma = 0.3, \epsilon = 0.31$, 2-lvls of OMF-4, $N_{\rm pf} \leq 8$, $\deg(\mathrm{R}) \leq 10$

m_π/MeV	P_{acc}	$P(\lvert\Delta H\rvert \geq 2)$
410	97.5%	0.15%
294	98.6%	0.15%
220	98.2%	0.05%

- Physical m_π seems possible also at such coarse lattice spacing.
- The lowest eigenvalue of $\sqrt{D^\dagger D}$ was measured with a precision greater than 0.5%.

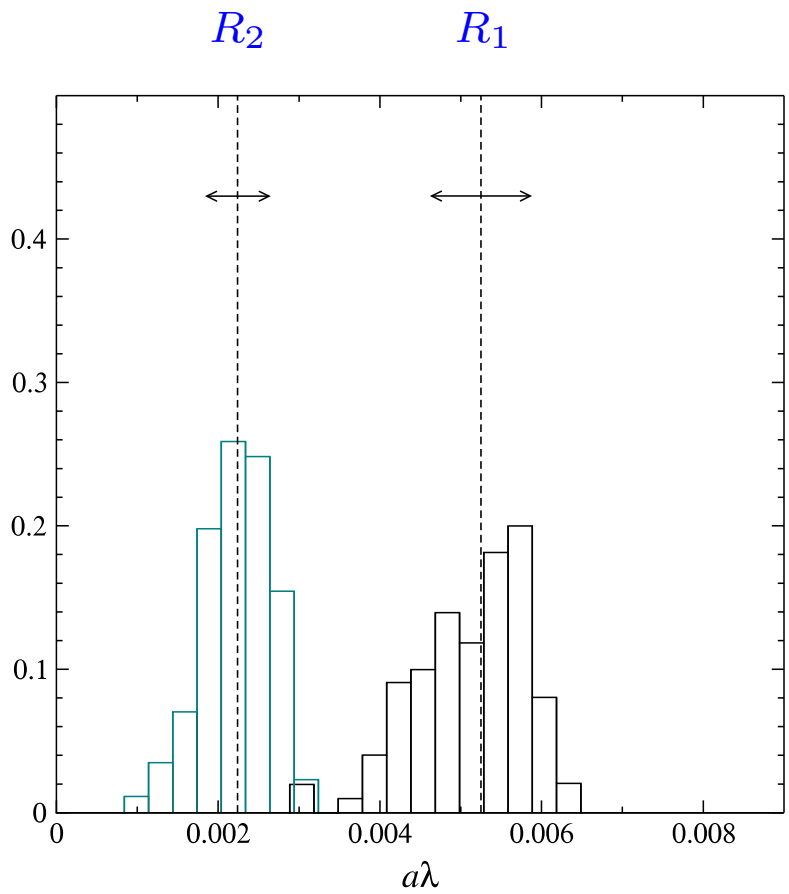
Spectral gap of the Dirac operator

- Width (σ) of the distribution is generally smaller than the traditional case.
- σ decreases for lighter pions (as observed with $N_f = 2$)
- Empirically $\sigma \propto 1/\sqrt{V}$
- No data for a direct comparison.
- For R_1 $m_\pi = 410$



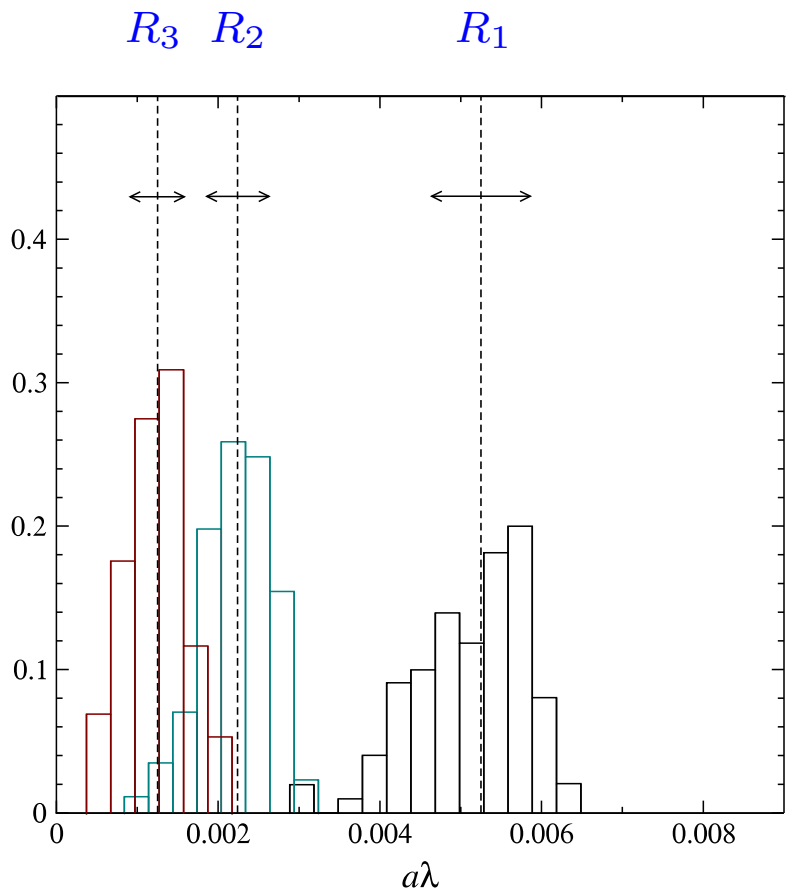
Spectral gap of the Dirac operator

- Width (σ) of the distribution is generally smaller than the traditional case.
- σ decreases for lighter pions (as observed with $N_f = 2$)
- Empirically $\sigma \propto 1/\sqrt{V}$
- No data for a direct comparison.
- For R_1 $m_\pi = 410$
- For R_2 $m_\pi = 294$



Spectral gap of the Dirac operator

- Width (σ) of the distribution is generally smaller than the traditional case.
- σ decreases for lighter pions (as observed with $N_f = 2$)
- Empirically $\sigma \propto 1/\sqrt{V}$
- No data for a direct comparison.
- For R_1 $m_\pi = 410$
- For R_2 $m_\pi = 294$
- For R_3 $m_\pi = 220$



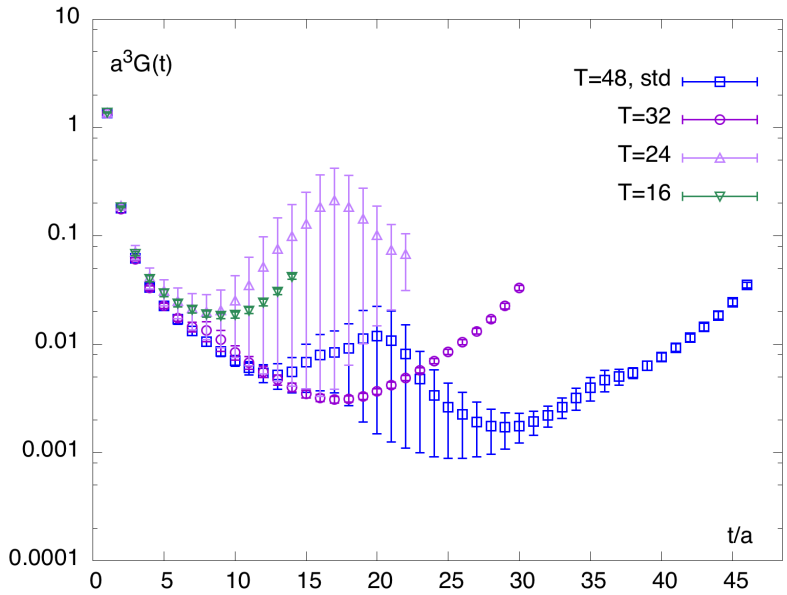
These simulations are all about equally expensive, as we lower m_u and raise m_s the cost balance stays constant.

Quenched observables

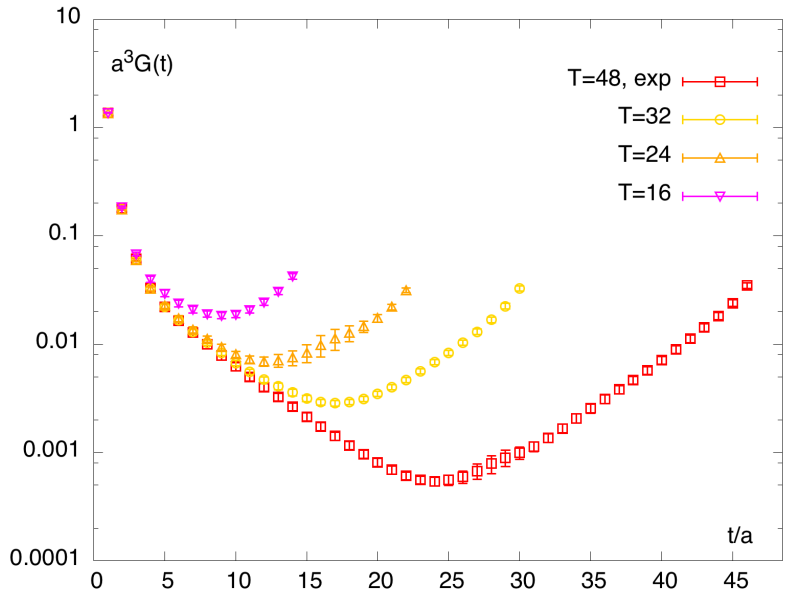
A direct comparison is easier in a quenched setup

- Quenched improved Wilson action, $\beta = 6.0$

Standard $c_{\rm sw}$



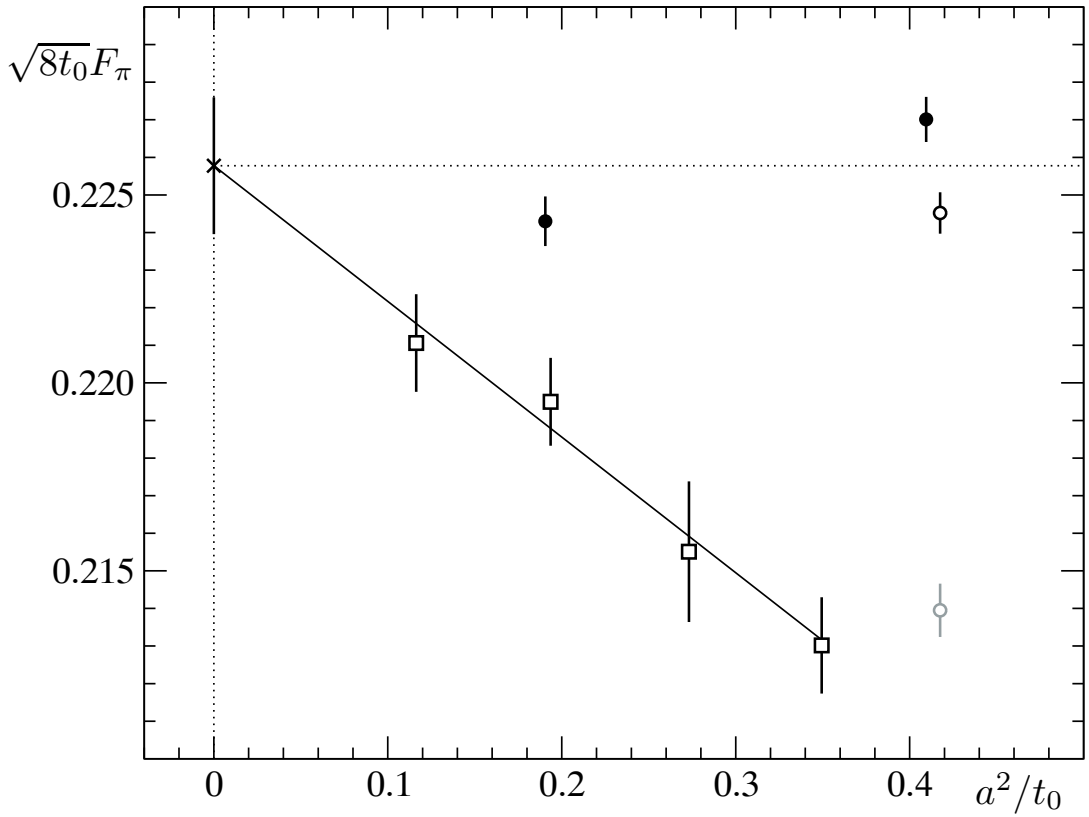
Modified $c_{\rm sw}$



Quantifying $\mathcal{O}(a^2)$ effects

- With the generation of the continuum limit trajectories still ongoing we can give estimates of the size of the cutoff effects.

CLS continuum extrapolation of symmetric point $f_{\pi K}$ data [M. Bruno et al. Phys. Rev. D 95, 074504]



NOTE: The CLS current has been renormalized with the Z_A from the chirally rotated SF [M. Dalla Brida et al. Eur. Phys. J. C79 (2019)]. We Use Z_A determined with the fermion flow [M. Lüscher,JHEP 1304 (2013) 123].

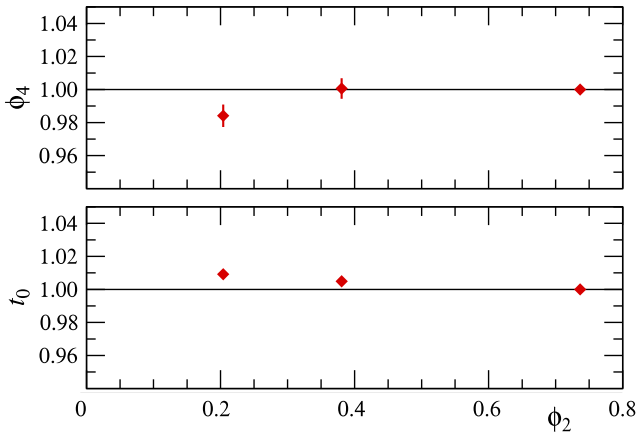
Lattice effects

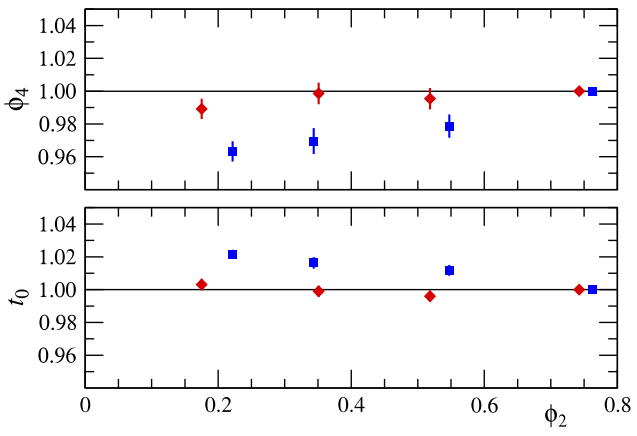
A fixed bare quark masses trajectory will show deviations in the observables of order $\mathcal{O}(a\,m)$
[M. Bruno et al. Phys. Rev. D 95, 074504]

Stabilized Wilson

CLS runs

a=0.095 fm





a=0.064 fm

a=0.086 fm

Where

$$\phi_4 \equiv 8t_0(\frac{1}{2}m_\pi^2 + m_K^2)$$

$$\phi_2 = 8t_0 m_\pi^2$$

H101,H102,H105,C101, blue squares.

N202,N203,N200,D200, red diamonds.

Summary and Perspectives

Our measures to stabilize the Wilson Dirac improved action comprise

- A modified O(a) improvement of the action.
- Stochastic Molecular Dynamics
- Uniform norm and quadruple precision

So far:

- ✓ Good behaviour of the proposal even on very coarse lattices
- ✓ Comparable runtime with respect to the traditional formulation
- ✓ No indication of unusually large lattice effects

Ongoing:

- Continuum limit scaling behaviour

Modified $c_{\rm SW}$ implementation

- The Cayley-Hamilton theorem can be used to express a $n \times n$ matrix as:

$$\exp_N(A) = \sum_{m=0}^{N} \frac{1}{m!} \sum_{i=0}^{n-1} c_i(m, A) A^i = \sum_{i=0}^{n-1} b_i(N, A) A^i.$$

- The Horner method states that for a sequence of polynomial $q_i(X)$ the iteration

$$\begin{aligned} q_N &= c_N \\ q_i &= X q_{i+1} + c_i \qquad\qquad i = N-1, \ldots, 0 \\ c_i &= 1/i! \end{aligned}$$

will converge and the last polynomial q_0 will coincide with the evaluation of the recursive sum up to grade N.

Modified $c_{\rm SW}$ implementation

- Specialising to the case of 6×6 matrices and the Cayley-Hamilton representation

$$
\begin{aligned}
q_{N,0} &= c_N = 1/N! \qquad & q_{N,1\ldots5} &= 0 \\
q_{n,0} &= -p_0 q_{n+1,5} + 1/n! & & \\
q_{n,i} &= -p_i q_{n+1,5} + q_{n+1,i-1} \qquad & i &= 1, \cdots, 5,
\end{aligned}
$$

  where the p_i are the coefficients of the characteristic polynomial.

- Rapid convergence: