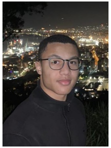
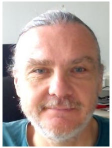
Parallel Tempered Metadynamics

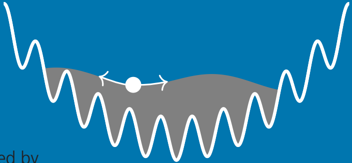
Slowing Down Critical Slowing Down

[PhysRevD.109.114504]/[2307.04742]

Timo Eichhorn, Gianluca Fuwa,
Christian Hoelbling, Lukas Varnhorst


NGT Algorithm Workshop - 10.11.2024



Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

BERGISCHE UNIVERSITÄT WUPPERTAL

Motivation - Cost of lattice QCD simulations

$$\text{Simulation cost} \propto \underbrace{V^{z_V}}_{\text{HMC}} \quad \underbrace{m_\pi^{-z_\pi}}_{\text{Solver}} \quad \underbrace{a^{-z_a}}_{\text{Autocorrelations}}$$

- $z_V \gtrsim 1$, depends on integrator, hard to imagine an algorithm where $z_V < 1$
  - HMC + second order integrator: $z_V = 1 + 1/4$
  - HMC + fourth order integrator: $z_V = 1 + 1/8$
- $z_\pi \sim 1$–2 after many algorithmic improvements, almost constant down to very small (below physical) quark masses
  - Hasenbusch mass preconditioning, domain decomposition, multiple pseudofermions
  - Multiple timescale integration
  - Multigrid solvers
  - Deflation
- $z_a \sim 2$ according to “naive” expectations
  $z_a \sim 5$–6 in the presence of topological sectors

Motivation - Cost of lattice QCD simulations

$$\text{Simulation cost} \propto \underbrace{V^{z_V}}_{\text{HMC}} \quad \underbrace{m_\pi^{-z_\pi}}_{\text{Solver}} \quad \underbrace{a^{-z_a}}_{\text{Autocorrelations}}$$

- $z_V \gtrsim 1$, depends on integrator, hard to imagine an algorithm where $z_V < 1$
  - HMC + second order integrator: $z_V = 1 + 1/4$
  - HMC + fourth order integrator: $z_V = 1 + 1/8$
- $z_\pi \sim 1$–2 after many algorithmic improvements, almost constant down to very small (below physical) quark masses
  - Hasenbusch mass preconditioning, domain decomposition, multiple pseudofermions
  - Multiple timescale integration
  - Multigrid solvers
  - Deflation
- $z_a \sim 2$ according to "naive" expectations
  $z_a \sim 5$–6 in the presence of topological sectors

Motivation - Cost of lattice QCD simulations

$$\text{Simulation cost} \propto \underbrace{V^{z_V}}_{\text{HMC}} \quad \underbrace{m_\pi^{-z_\pi}}_{\text{Solver}} \quad \underbrace{a^{-z_a}}_{\text{Autocorrelations}}$$

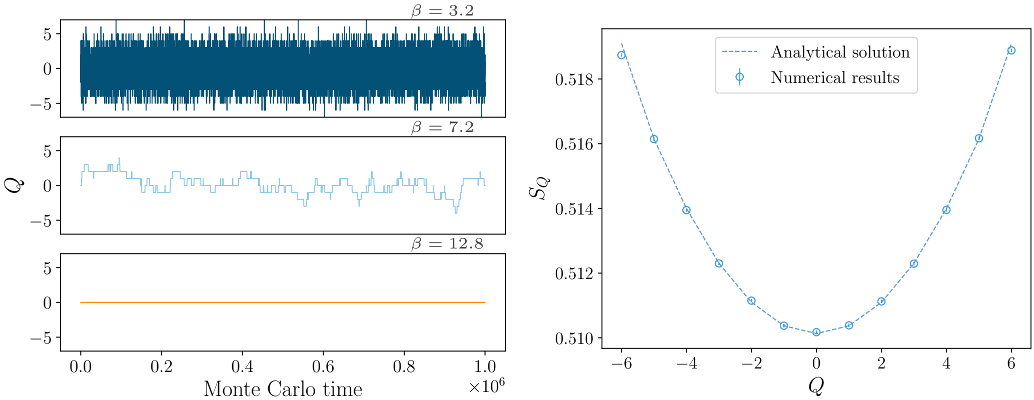
- $z_V \gtrsim 1$, depends on integrator, hard to imagine an algorithm where $z_V < 1$
  - HMC + second order integrator: $z_V = 1 + 1/4$
  - HMC + fourth order integrator: $z_V = 1 + 1/8$
- $z_\pi \sim 1$–2 after many algorithmic improvements, almost constant down to very small (below physical) quark masses
  - Hasenbusch mass preconditioning, domain decomposition, multiple pseudofermions
  - Multiple timescale integration
  - Multigrid solvers
  - Deflation
- $z_a \sim 2$ according to “naive” expectations
  $z_a \sim 5$–6 in the presence of topological sectors

Motivation - Cost of lattice QCD simulations

$$\text{Simulation cost} \propto \underbrace{V^{z_V}}_{\text{HMC}} \quad \underbrace{m_\pi^{-z_\pi}}_{\text{Solver}} \quad \underbrace{a^{-z_a}}_{\text{Autocorrelations}}$$

- $z_V \gtrsim 1$, depends on integrator, hard to imagine an algorithm where $z_V < 1$
  - HMC + second order integrator: $z_V = 1 + 1/4$
  - HMC + fourth order integrator: $z_V = 1 + 1/8$
- $z_\pi \sim 1$–2 after many algorithmic improvements, almost constant down to very small (below physical) quark masses
  - Hasenbusch mass preconditioning, domain decomposition, multiple pseudofermions
  - Multiple timescale integration
  - Multigrid solvers
  - Deflation
- $z_a \sim 2$ according to “naive” expectations
  $z_a \sim 5$–6 in the presence of topological sectors $\Rightarrow$ from now on z refers to z_a

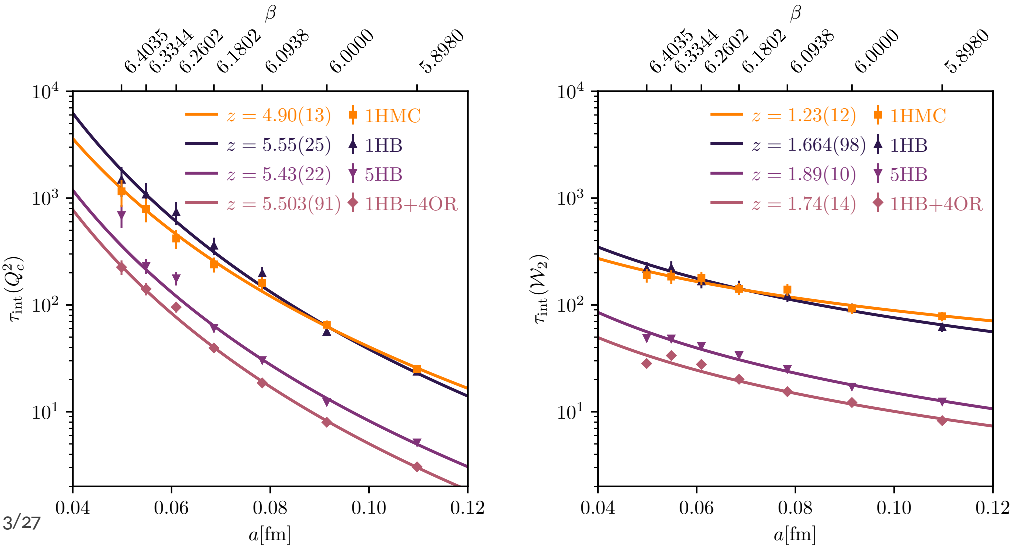
Topological freezing - 2D U(1)



Slow topological modes couple to other non-topological observables

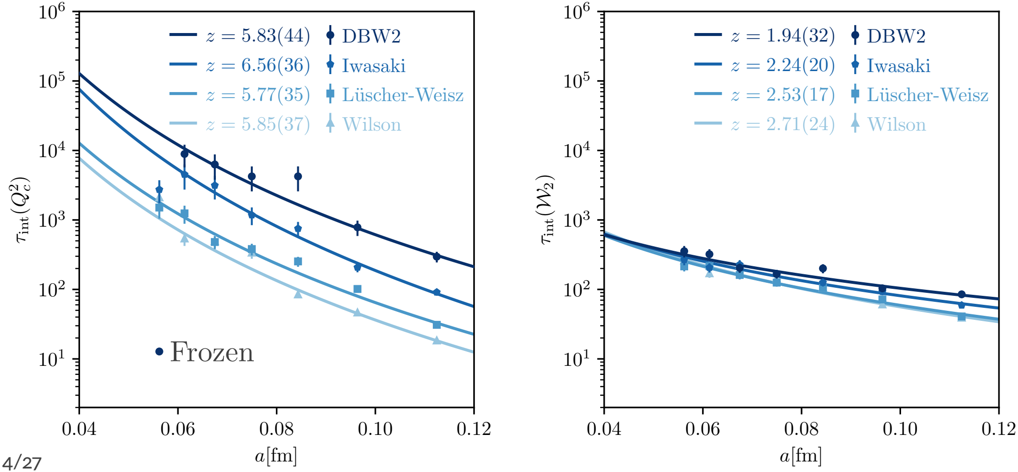
Correlation between topological charge and gauge action

Topological freezing - 4D SU(3) with Wilson action



Topological freezing - 4D SU(3) with different actions

Similar scaling for different actions/“surprisingly universal” behaviour



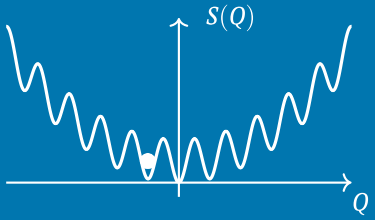
Possible approaches to the problem

- Fixed topology simulations
- Master field simulations
- Modified boundary conditions
  - Open boundaries, P-periodic
  - Parallel tempering in boundary conditions
- Nested sampling
- Trivializing maps
- Machine learning (normalizing flows, diffusion models...)
- Multiscale thermalization
- Instanton updates
- **Metadynamics**
- ...

Metadynamics: Remove barriers between sectors

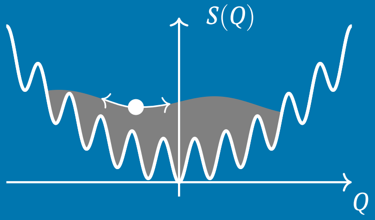
General idea

Original action landscape



Stuck in sector (here $Q = -1$)

Modified action landscape



Able to move between sectors

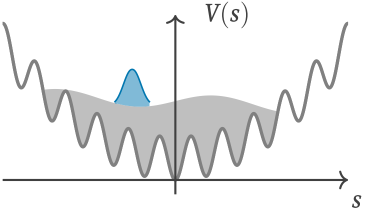
Metadynamics

- Well established method for molecular dynamics simulations [cond-mat/0208352]
- In context of lattice gauge theory first proposed for CP^{N-1} models in [1508.07270]
- Add (time-dependent) bias potential $V_t(s)$ to action, depending on set of collective variables (CVs) s
- Converges towards negative free energy $\pm$ const.
  $\Rightarrow$ Marginal distribution of CVs becomes flat
- **Reweighting necessary**

Construct from Gaussians

$$V_t(s) = \sum_{t' \leq t} \prod_i^N g(s_i - s_i(t'))$$

$$g(s) = w \exp\left(-s^2/(2\delta s^2)\right)$$



Metadynamics

- Single collective variable Q_{meta}
  - Clover-based topological charge
  - $n = 4$–10 stout smearing steps with $\rho = 0.12$
  - Trade-off between
    - Staying in relevant regions of phase space (less smearing)
    - Having sufficient resolution (more smearing)
- Resulting modified action $S_t^M = S + V_t(Q)$

$\Rightarrow$ Staple becomes less local, so use HMC

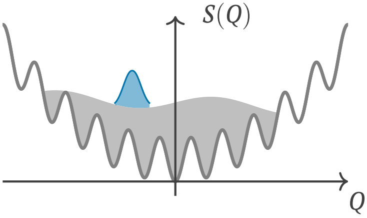
$\Rightarrow$ Stout force recursion (as with smeared fermions)

$$F_{\text{meta}} = \frac{\partial V}{\partial Q_{\text{meta}}} \frac{\partial Q_{\text{meta}}}{\partial U^{(n)}} \frac{\partial U^{(n)}}{\partial U}$$

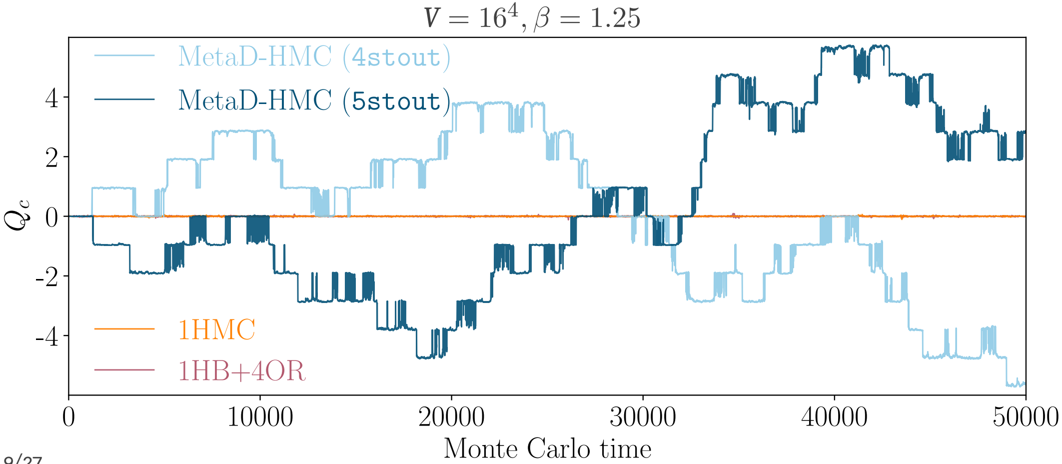
Construct from Gaussians

$$V_t(Q) = \sum_{t' \leq t} g(Q - Q(t'))$$

$$g(Q) = w \exp\left(-Q^2/(2\delta Q^2)\right)$$

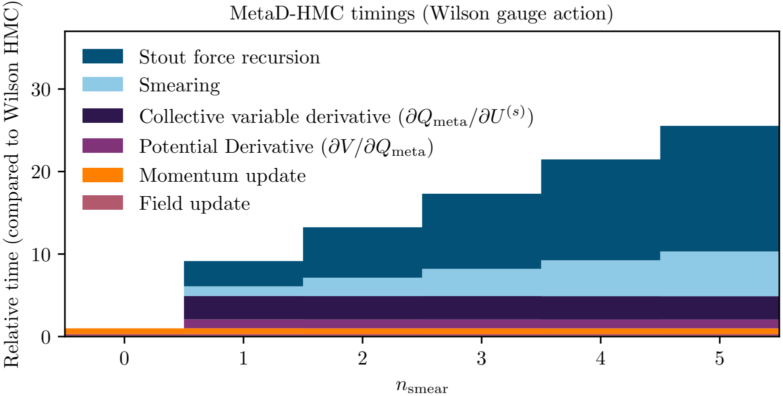


Metadynamics: 4D SU(3) with DBW2 action



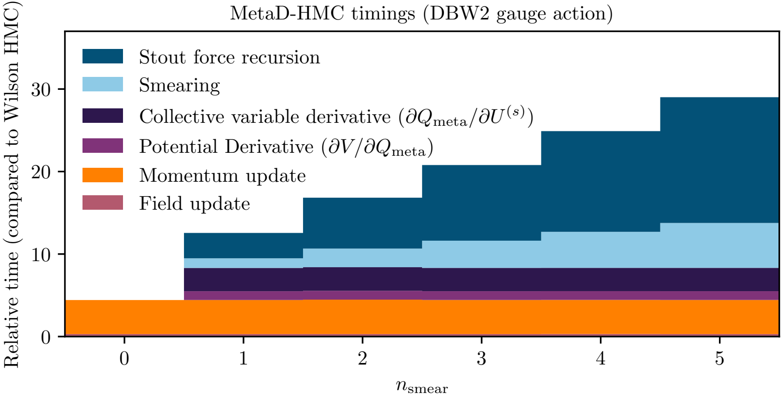
Metadynamics: 4D SU(3) with DBW2 action

- Improvement of autocorrelation times by at least two orders of magnitude:
  - HMC or 1HB+4OR: $\tau_{Q^2} > 400\,000$
  - MetaD: $\tau_{Q^2} \sim 2000$–3000
- Caveat: Large computational overhead (Caveat^2: much smaller for full QCD)



Metadynamics: 4D SU(3) with DBW2 action

- Improvement of autocorrelation times by at least two orders of magnitude:
  - HMC or 1HB+4OR: $\tau_{Q^2} > 400\,000$
  - MetaD: $\tau_{Q^2} \sim 2000$–3000
- Caveat: Large computational overhead (Caveat^2: much smaller for full QCD)



Metadynamics - Reweighting efficiency

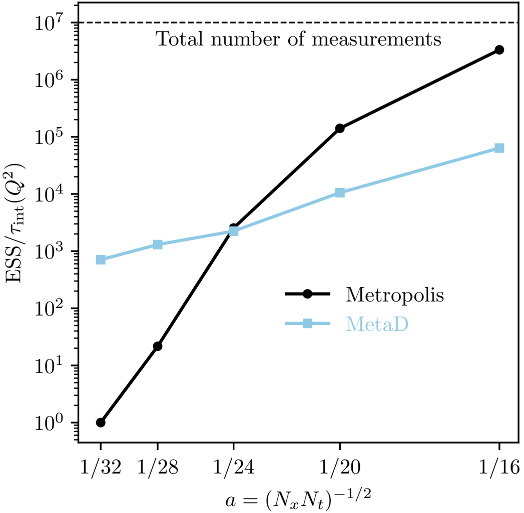
Efficiency of reweighting?

- Effective sample size

$$\mathrm{ESS} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}$$

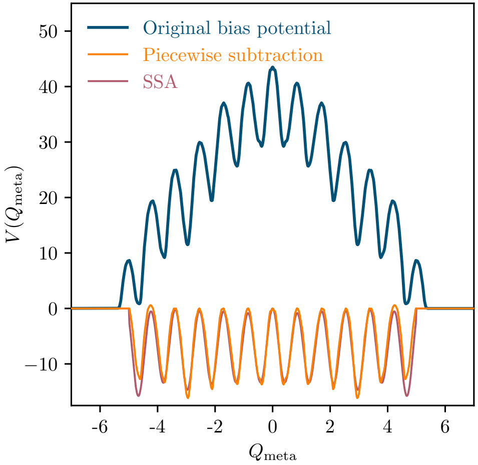
- Reweighting significantly reduces ESS down to $\mathcal{O}(10\,\%)$
- Two causes:
  - Oversampling of sectors with large $|Q|$
    $\Rightarrow$ Modify bias potential
  - Oversampling of configurations between sectors
    $\Rightarrow$ Combine with parallel tempering (PT-MetaD)

Scaling plot for 2D U(1)

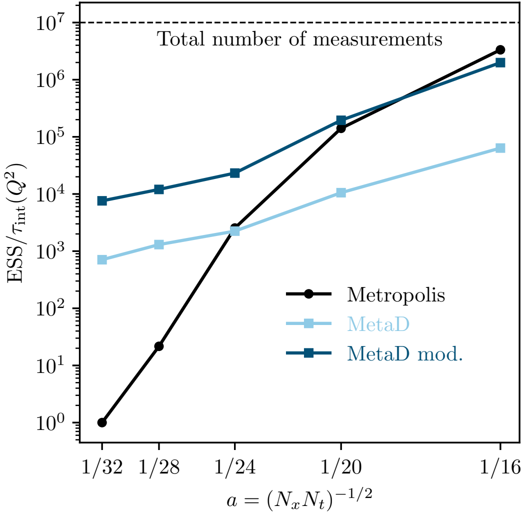


Metadynamics - bias potential modification

Oversampling of large $|Q|$



Scaling plot for 2D U(1)



Only remove barriers between sectors

Parallel Tempered Metadynamics (PT-MetaD)

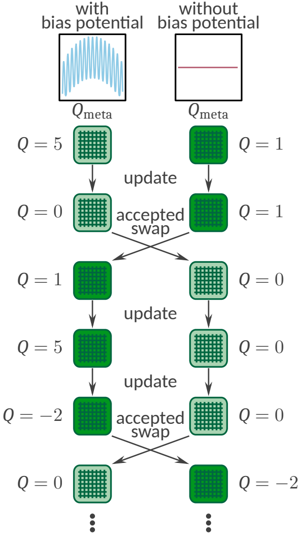
Address oversampling of configurations between sectors

- Two streams (inspired by [1706.04443]):
  - Stream 1 with bias potential ⇒ tunneling
  - Stream 2 without bias potential ⇒ measurements (no reweighting)
- Propose swaps like in standard parallel tempering
- Swaps **only depend on bias potential!**

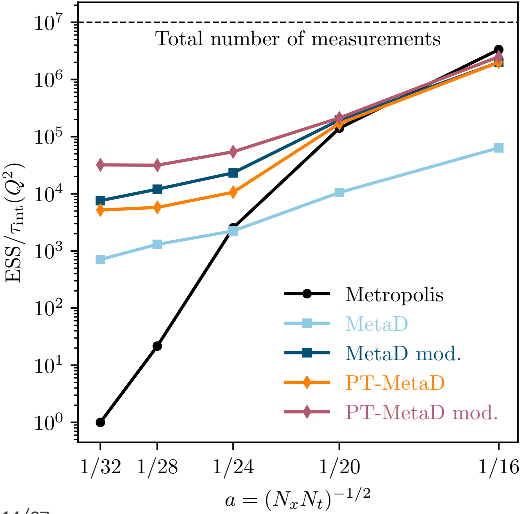
$$\Delta S_t^M = \left[S_t^M(U_1) + S(U_2)\right] - \left[S_t^M(U_2) + S(U_1)\right]$$
$$= V_t(Q_{\text{meta},1}) - V_t(Q_{\text{meta},2})$$

⇒ Fermions pose no additional difficulty

- Monitor τ_{int} of observables defined on product space



Scaling of PT-MetaD in 2D U(1)



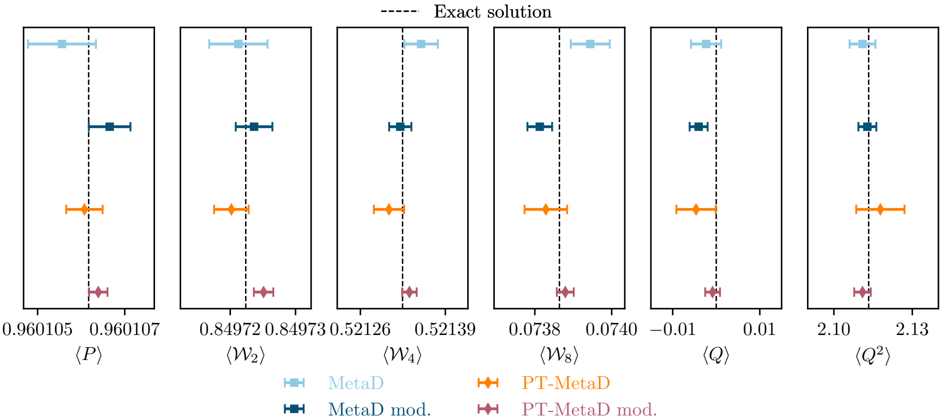
Putting everything together

- Further improvement compared to standard MetaD
- Best results when combined with modification of bias potential

⇒ Almost 5 orders of magnitude improvement for finest lattice spacing compared to conventional algorithm

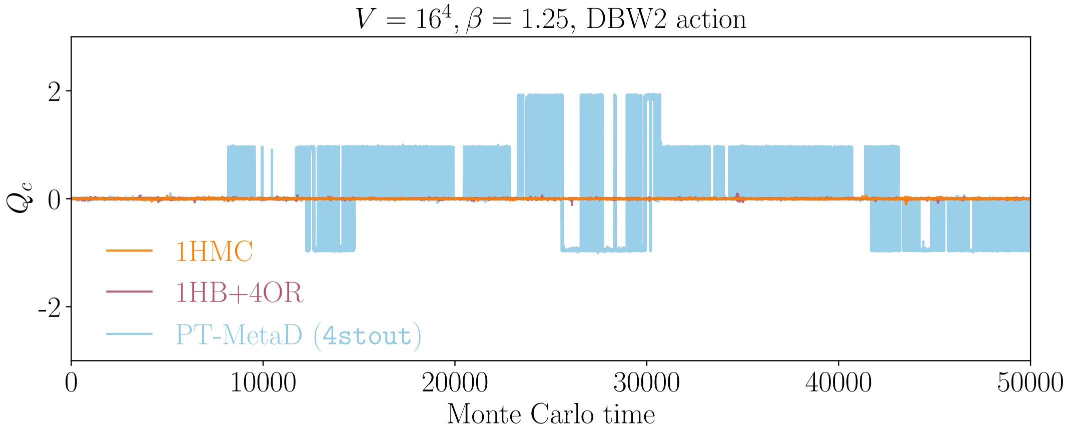
- Scaling of all (PT-)MetaD variations much milder compared to conventional algorithms

Scaling of PT-MetaD in 2D U(1)

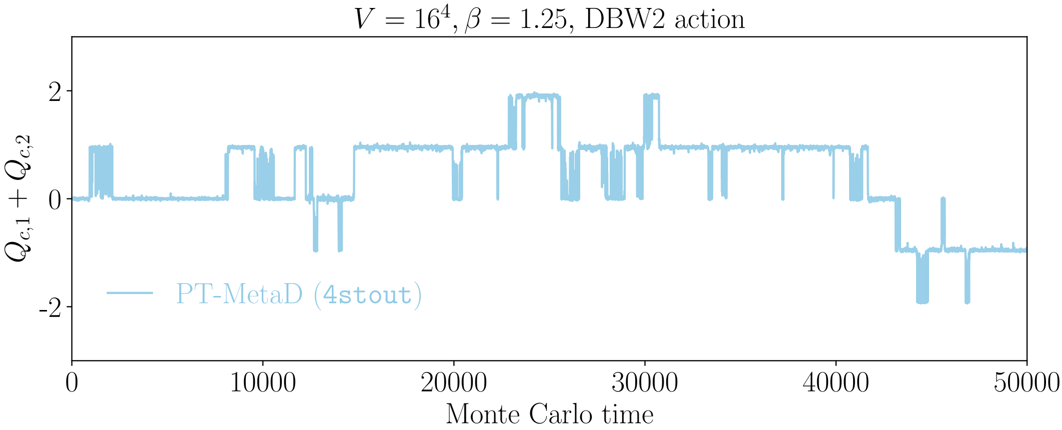


- Topological sampling as good as MetaD
- Non-topological sampling improved

PT-MetaD - 4D SU(3)



PT-MetaD - 4D SU(3)



Extension to QCD

- No conceptual difficulties
- In some ways better suited to QCD simulations than pure gauge
  - HMC already required
  - (Stout) smearing often used for fermions
  - $\Rightarrow$ Potentially no overhead from (stout) force recursion
- Buildup of potential may take too long (timescales of $\mathcal{O}(10^4)$ unfeasible)

Accelerating the bias potential thermalization

1. **Use prior knowledge/symmetries**
   - Charge conjugation symmetry of action ($Q \leftrightarrow -Q$)
     In practice: Whenever we update the bias potential at some Q_{meta} we also update it at $-Q_{\text{meta}}$
     $\Rightarrow$ Approximately $2\times$ speed-up
   - Expect a (near) Gaussian distribution of Q with variance $\langle Q^2 \rangle \sim V$
     $\Rightarrow$ Extrapolate from smaller volumes?
   - Barrier shapes seem to be periodic
     $\Rightarrow$ Extrapolate from inner sectors?

Accelerating the bias potential thermalization

2. Well-tempered Metadynamics [Barducci' 08]

- Standard Metadynamics

$$V_{t+1}(Q) = V_t(Q) + w \exp\left(-\frac{(Q_t - Q)^2}{2\sigma^2}\right)$$

- Well-tempered Metadynamics

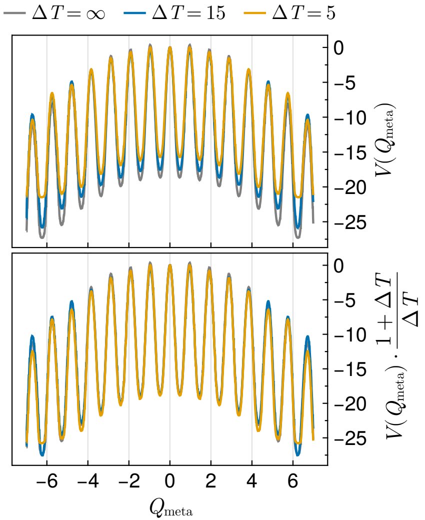
$$V_{t+1}(Q) = V_t(Q) + \exp\left(-\frac{V_t(Q)}{\Delta T}\right) w \exp\left(-\frac{(Q_t - Q)^2}{2\sigma^2}\right)$$

Tunable parameter ΔT:

- $\Delta T \to 0$: No Metadynamics
- $\Delta T \to \infty$: Standard Metadynamics

Accelerating the bias potential thermalization

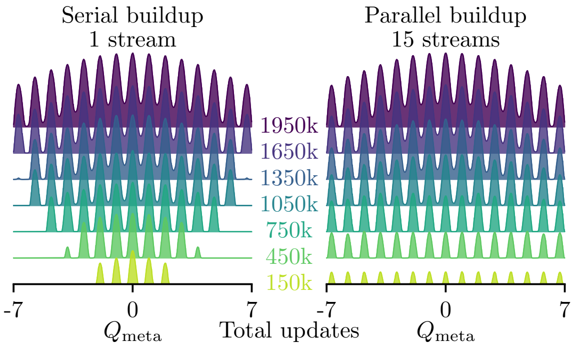
2. **Well-tempered Metadynamics** [Barducci' 08]
   - Able to choose larger w while maintaining smoothness in the end
   - Quirk: Bias does not converge to $-S(Q)$ but $-\frac{\Delta T}{1+\Delta T} S(Q)$
     $\Rightarrow$ Small ΔT may reintroduce ergodicity problems, as the barriers are not entirely canceled out



Accelerating the bias potential thermalization

3. Multiple walkers [Raiteri' 06]

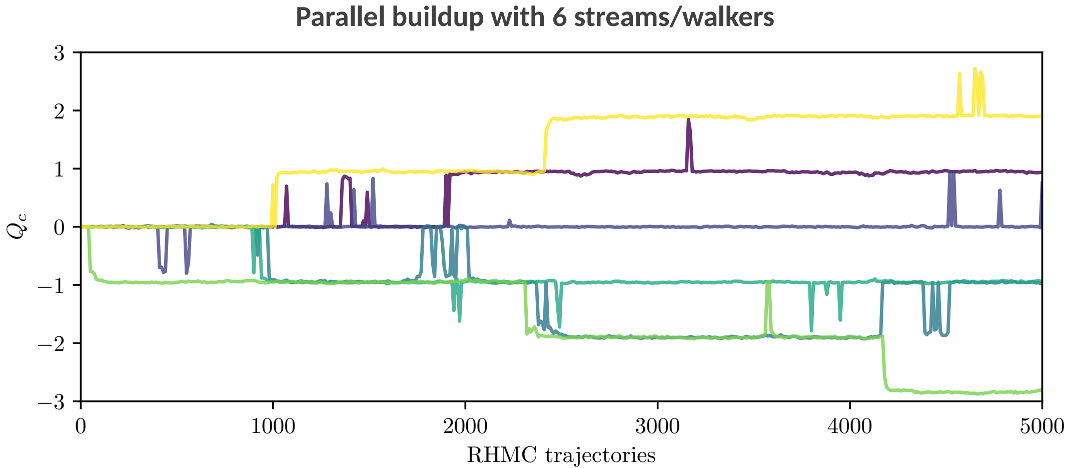
- Run N_{walkers} simulation streams in parallel, all working on the same potential
- Minimal communication between processes required (e.g., a single `MPI.Allgather` call per iteration)
  $\Rightarrow$ speed-up of factor $\sim N_{\text{walkers}}$
- **Possible variation:** Start each walker in a different topological sector to eliminate time before falling into unexplored sector



QCD results

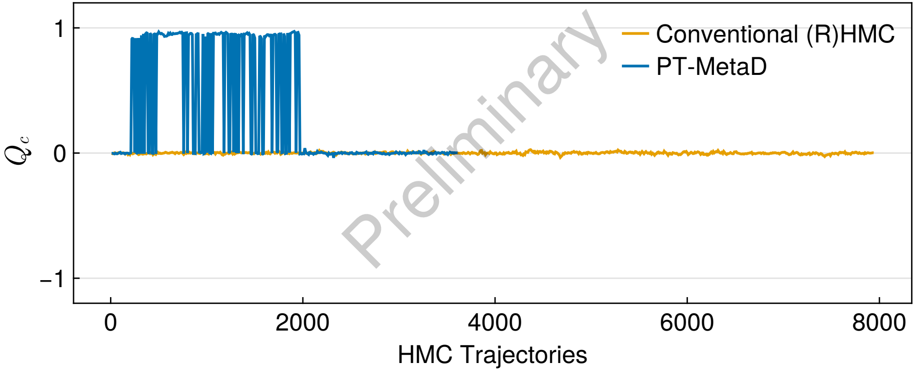
- Simulated action
  - **Fermion Action:**
    $N_f = 2$, $am = 0.02$
    4×0.125 Stout-smeared Staggered
  - **Gauge Action:**
    DBW2 $\beta = 1.05$
    $\Rightarrow a^{-1} \approx 0.05\,\mathrm{fm}$
  - **Bias potential:**
    Clover charge with 6×0.12 stout smearing
- Lattice Volume: $(16a)^4 \approx (0.8\,\mathrm{fm})^4$
- Bias parameters: $\delta Q = w = 0.02$, $\gamma = \infty$, $[Q_{\mathrm{min}}, Q_{\mathrm{max}}] = [-3, 3]$ and 6 walkers
- Relative overhead of only $\sim 20\,\%$ compared to conventional RHMC

QCD results - MetaD timeseries (potential buildup)



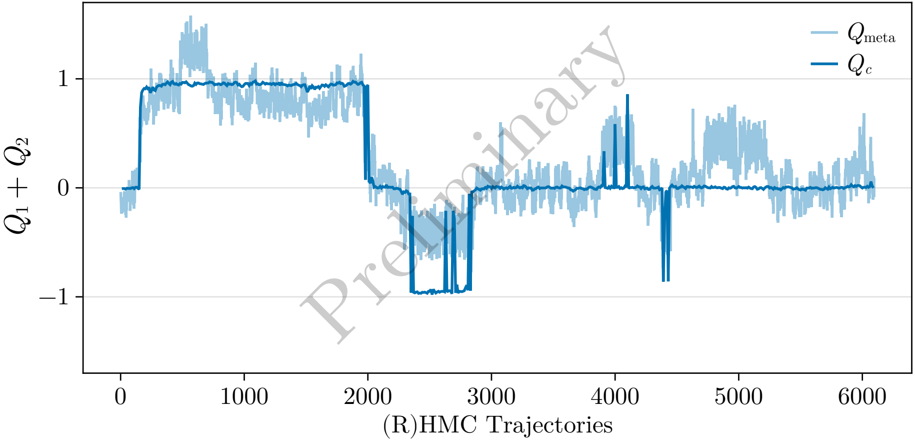
QCD results - PT-MetaD timeseries (static potential)

Better tunneling than conventional RHMC, but suboptimal potential?



QCD results - PT-MetaD timeseries (static potential)

Better tunneling than conventional RHMC, but suboptimal potential?

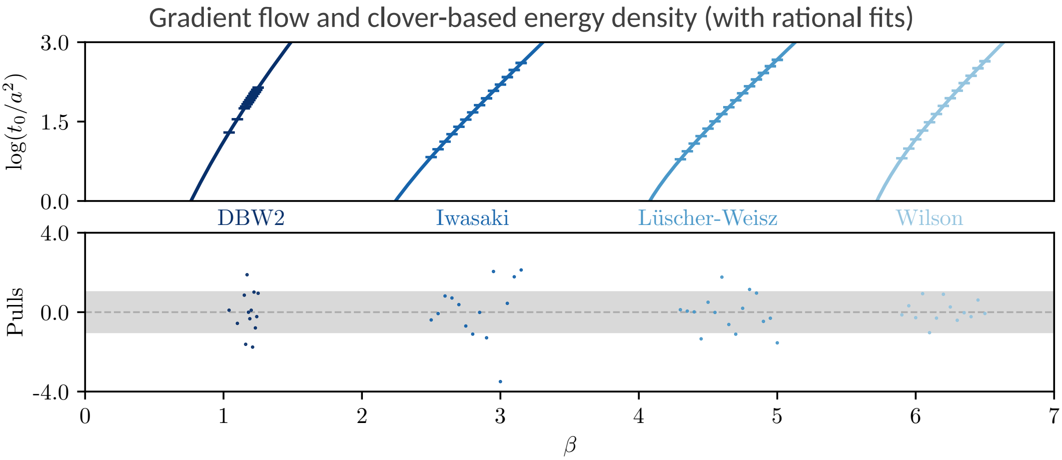


Summary

- PT-MetaD + modified bias potential
  - At least as efficient in reducing autocorrelation times as standard MetaD
  - No reweighting required $\Rightarrow$ No reduction of effective sample size
  - Improved scaling
    - With lattice spacing
    - With volume (compared to standard parallel tempering in β)
- Open questions:
  - Scaling behavior in dynamical simulations
  - Can the buildup of the potential be accelerated or entirely avoided?
  - Synergy with other approaches (only targets topological freezing, not critical slowing down in general)
- Pure gauge code: [Lettuce] (currently undergoing rewrite)
- Dynamical code: [MetaQCD.jl]

Backup

Scale setting



Scale setting - Fit parameters

- Rational fit ansatz with 3 and four parameters

$$\ln(*/a^2) = \frac{8\pi^2}{33}\beta\frac{1 + d_1/\beta + d_2/\beta^2}{1 + d_3/\beta}$$

$$\ln(*/a^2) = \frac{8\pi^2}{33}\beta\frac{1 + d_1/\beta + d_2/\beta^2}{1 + d_3/\beta + d_4/\beta^2}$$

  where $*$ is either t_0 or w_0^2
- Setup same as in [2307.04742], only difference between runs:
  - DBW2 action: Use RK3 integrator (from [1006.4518])
  - Other actions: Use RK3W7 integrator (from [2101.05320])
- Results should not be interpreted as an attempt at precise scale setting due to
  - Relatively small volumes (48×32^3)
  - Large autocorrelations for Iwasaki and DBW2 actions (only few tunneling events for finest lattice)

Scale setting - Fit parameters

Action	Definition	$\chi^2/\text{d.o.f.}$	AIC	BIC	d_1	d_2	d_3	d_4
Wilson	$t_{0,\text{plaq}}$	0.4040	-9.1934	-7.4985	-10.6943	28.4179	-5.4472	-
		0.3934	-8.9091	-6.6493	-10.7799	28.9095	-5.6475	0.9154
	$t_{0,\text{clov}}$	**0.3701**	**-10.3339**	**-8.6390**	**-10.5983**	**27.9044**	**-5.3532**	**-**
		0.4033	-8.5851	-6.3253	-10.6504	28.2014	-5.4607	0.4509
	$w_{0,\text{plaq}}$	0.5524	-5.1255	-3.4307	-10.5471	27.6060	-5.3367	-
		0.5857	-3.7355	-1.4757	-10.6810	28.3705	-5.6140	1.1800
	$w_{0,\text{clov}}$	0.5417	-5.3796	-3.6848	-10.5483	27.6182	-5.3341	-
		0.5633	-4.2417	-1.9819	-10.7023	28.4958	-5.6505	1.3363

Scale setting - Fit parameters

Action	Definition	$\chi^2/\text{d.o.f.}$	AIC	BIC	d_1	d_2	d_3	d_4
Lüscher-Weisz	$t_{0,\text{plaq}}$	1.0344	3.1601	5.2842	-7.6495	14.5316	-3.8449	-
		0.9789	3.0282	5.8604	-7.4807	13.8374	-3.5555	-0.7666
	$t_{0,\text{clov}}$	**0.9815**	**2.3723**	**4.4965**	**-7.5710**	**14.2438**	**-3.7571**	**-**
		0.9418	2.4476	5.2798	-7.2601	12.9838	-3.2940	-0.9920
	$w_{0,\text{plaq}}$	1.1786	5.1181	7.2423	-7.5313	14.0723	-3.7481	-
		1.2103	6.2105	9.0427	-7.2295	12.8438	-3.2836	-1.0750
	$w_{0,\text{clov}}$	1.1485	4.7296	6.8537	-7.5347	14.0911	-3.7475	-
		1.1911	5.9714	8.8036	-7.2632	12.9885	-3.3346	-0.9335

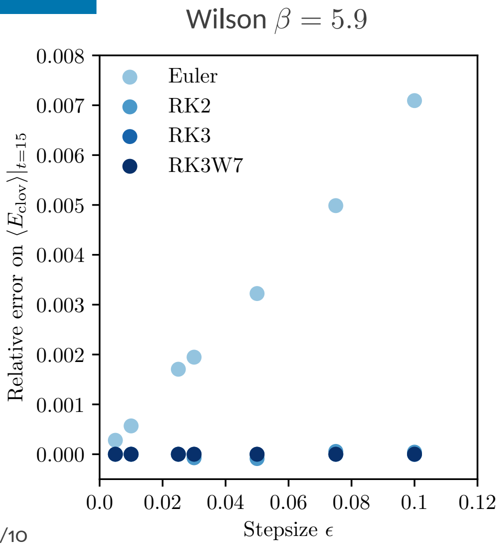
Scale setting - Fit parameters

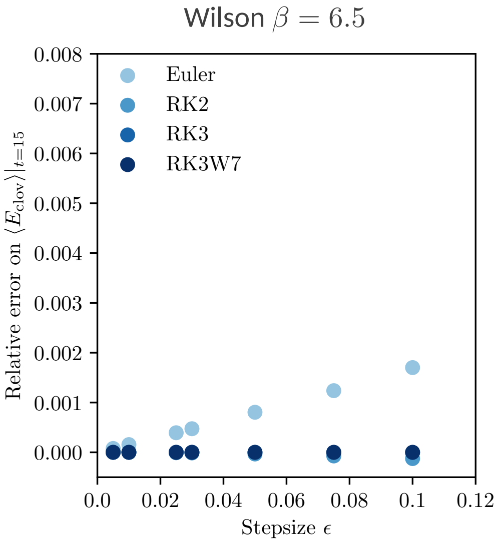
Action	Definition	χ^2/d.o.f.	AIC	BIC	d_1	d_2	d_3	d_4
Iwasaki	$t_{0,\text{plaq}}$	2.5849	15.9191	17.8363	-3.7990	3.4459	-1.8512	-
		2.3742	15.3946	17.9509	-4.2711	4.5271	-2.4362	0.6535
	$t_{0,\text{clov}}$	**2.6527**	**16.2816**	**18.1988**	**-3.5912**	**3.0251**	**-1.6435**	**-**
		2.4833	16.0236	18.5798	-4.3548	4.7300	-2.5136	0.7051
	$w_{0,\text{plaq}}$	2.3712	14.7111	16.6283	-3.6807	3.2020	-1.7553	-
		2.3945	15.5137	18.0699	-4.2616	4.5144	-2.4463	0.6869
	$w_{0,\text{clov}}$	2.4038	14.9027	16.8199	-3.6729	3.1919	-1.7434	-
		2.3323	15.1451	17.7013	-4.3325	4.6760	-2.5225	0.7476

Scale setting - Fit parameters

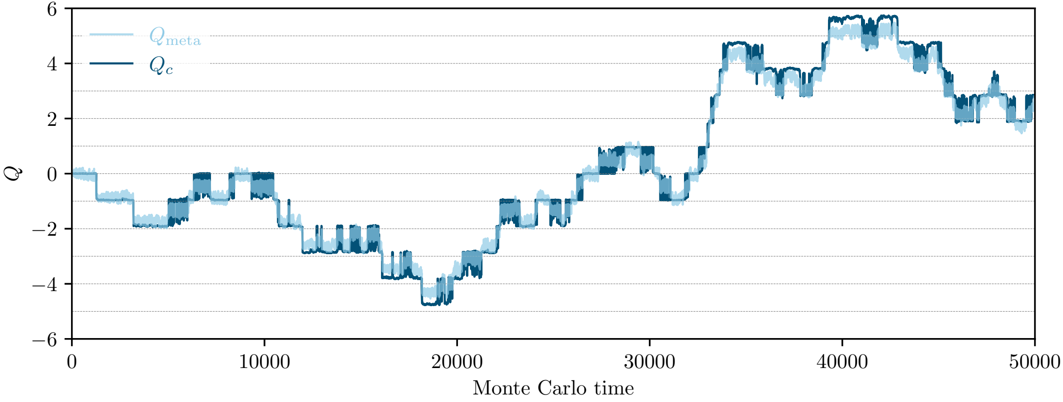
Action	Definition	$\chi^2/\mathrm{d.o.f.}$	AIC	BIC	d_1	d_2	d_3	d_4
DBW2	$t_{0,\mathrm{plaq}}$	1.2491	5.4809	7.1757	0.5953	-1.1083	0.1059	-
		1.3841	7.4450	9.7048	-0.3606	-0.3352	-0.6167	0.3067
	$t_{0,\mathrm{clov}}$	**1.3058**	**6.0581**	**7.7530**	**1.0351**	**-1.3763**	**0.4058**	**-**
		1.4430	7.9867	10.2465	-0.5080	-0.1750	-0.7426	0.4182
	$w_{0,\mathrm{plaq}}$	1.3569	6.5568	8.2516	0.8566	-1.2825	0.2488	-
		1.5076	8.5565	10.8163	0.6397	-1.1098	0.0882	0.0641
	$w_{0,\mathrm{clov}}$	1.3615	6.6006	8.2954	0.9223	-1.3198	0.2951	-
		1.5127	8.6000	10.8598	1.3160	-1.6310	0.5862	-0.1135

Scale setting - Integrator accuracy and scaling

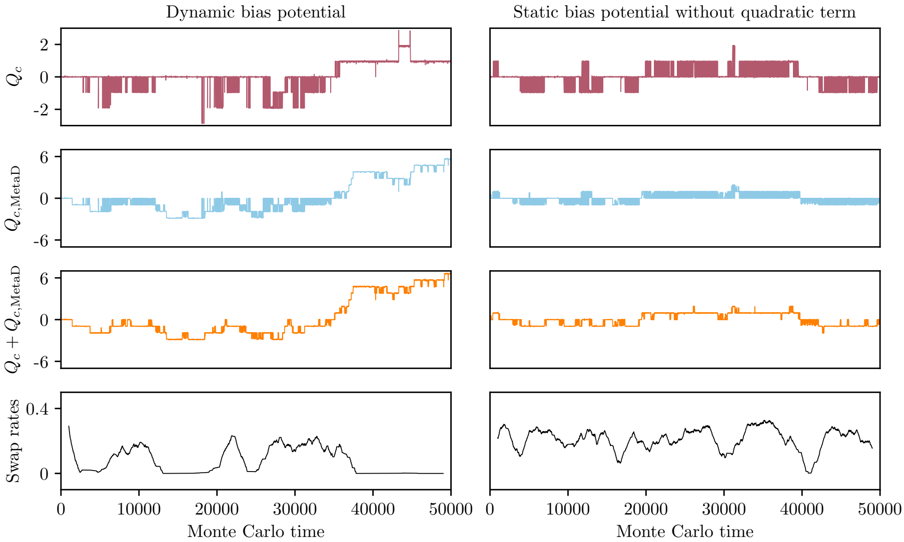




Metadynamics: 4D SU(3) with DBW2 action



Detailed timeseries of PT-MetaD in 4D SU(3)



QCD bias potential