Tackling critical slowing down using global correction steps with equivariant flows within the 2D Schwinger model

Jacob Finkenrath

Meinerzhagen
16.08.2022

The Cyprus Institute
RESEARCH·TECHNOLOGY·INNOVATION

CaSToRC

Table of Content

Simulations at the precision frontier

- critical slowing down via Dirac's index
- global corrections within Monte Carlo simulations

Generative models for U(1)

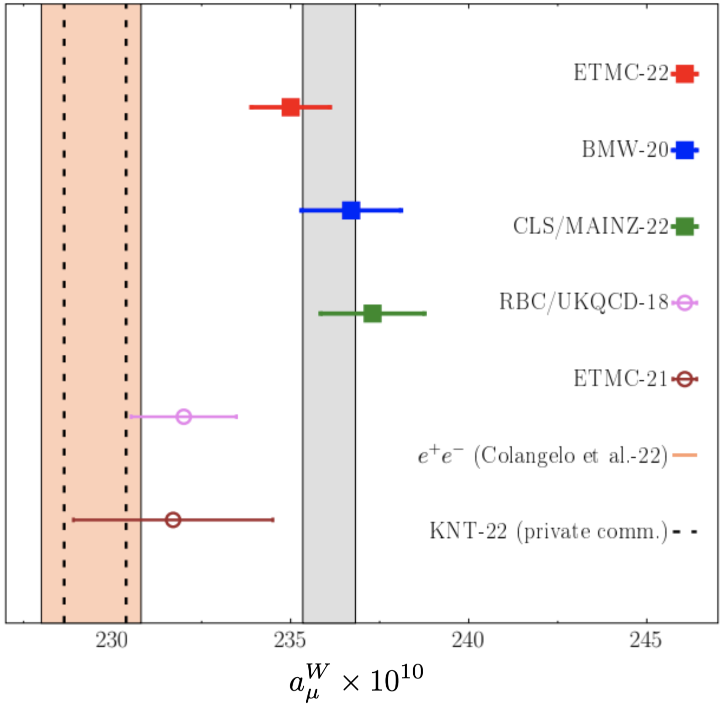
- insides into gauge invariant flows
- scalability via domain decomposition

Global corrections with the fermion determinant

- towards high acceptance rate
- towards low autocorrelation

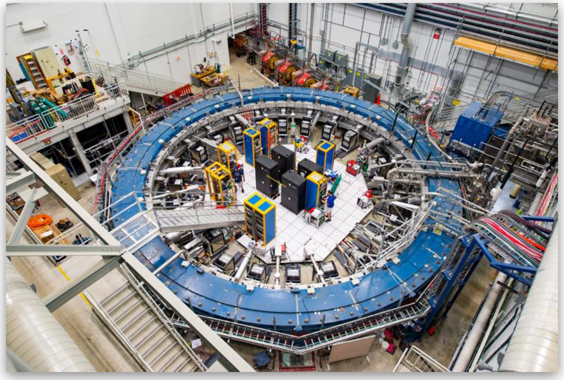
Comments on steps towards 4D-QCD …

The Cyprus Institute
RESEARCH•TECHNOLOGY•INNOVATION



Exciting times for
Lattice Quantum Chromodynamics

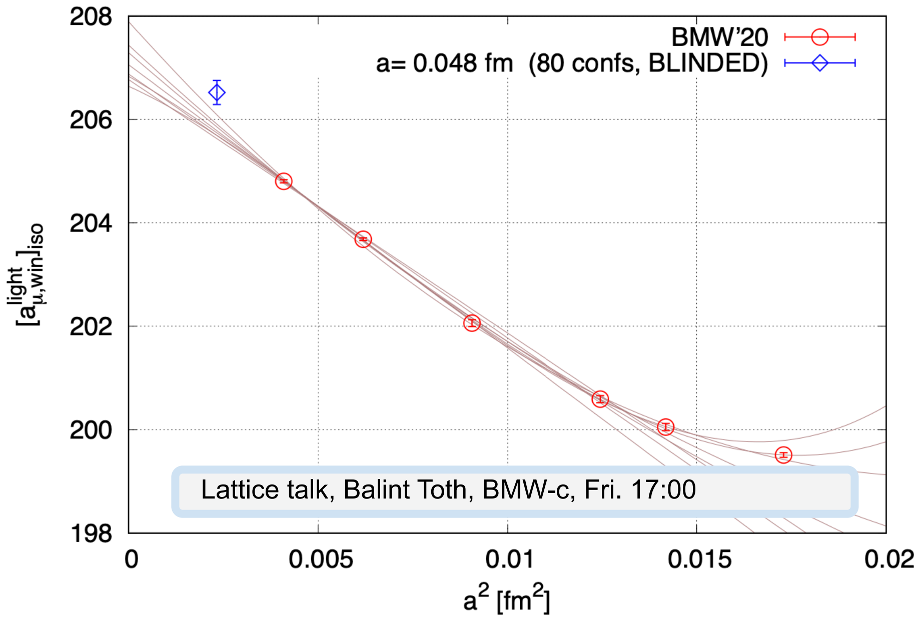
Muon and Flavor Physics
are indicating New Physics;
ab initio LQCD calculations are needed



Search for new physics in the precision frontier by

- high precision measurements
- theoretical prediction

deviations are signs for new physics



Anomalous magnetic moment of muon:
Muon g-2 Experiment at FermiLab confirmed results

- 4σ deviation between experiment and data-driven approach
- 4σ deviation between lattice and data-driven approach

To resolve this puzzle:

Precision Measurement of Lattice QCD are needed

aSTRC

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

Simulation at the Precision Frontier:

Very fine lattice spacing needed to match future experiments precision

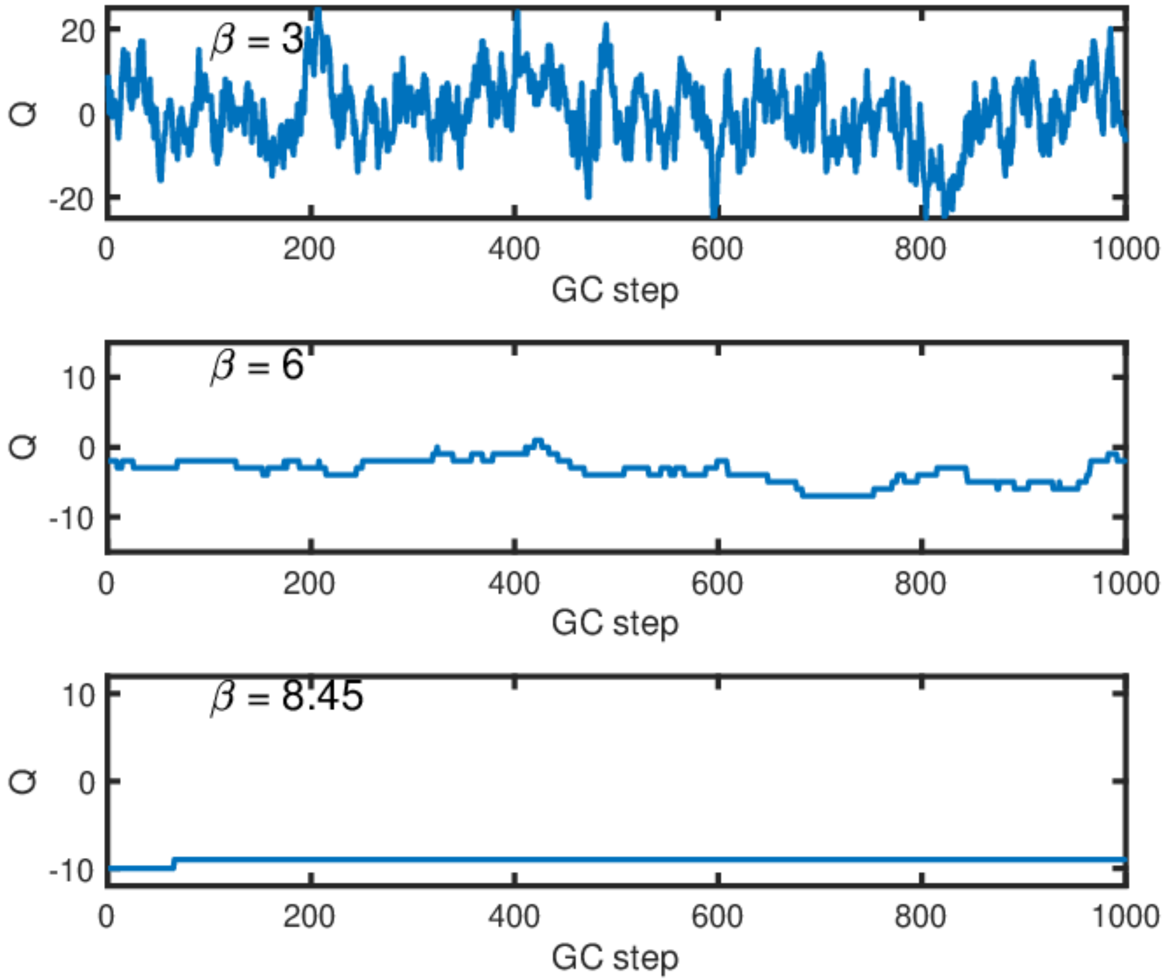
Standard large scale MCMC method:

- Hybrid Monte Carlo (HMC) algorithm
  - based on molecular dynamics

$$\dot{P} = -\frac{\partial H}{\partial U} \quad \text{and} \quad \dot{U} = \frac{\partial H}{\partial P}$$

for very fine lattice spacings a<0.05 fm
the HMC algorithm freezes out a topological sector

S. Schaefer et al., Null. Phys. B 845 (2011) 93-119



severe critical slowing down

- Efficient algorithm in QCD missing (openBC would be a possibility)

here, we will take a look to a simpler case: 2D Schwinger Model

aSTоRC

The Index theorem gives some illustrative insides:

$$N_R - N_L = Q^{geo}$$

Atiyah and Singer, 1963

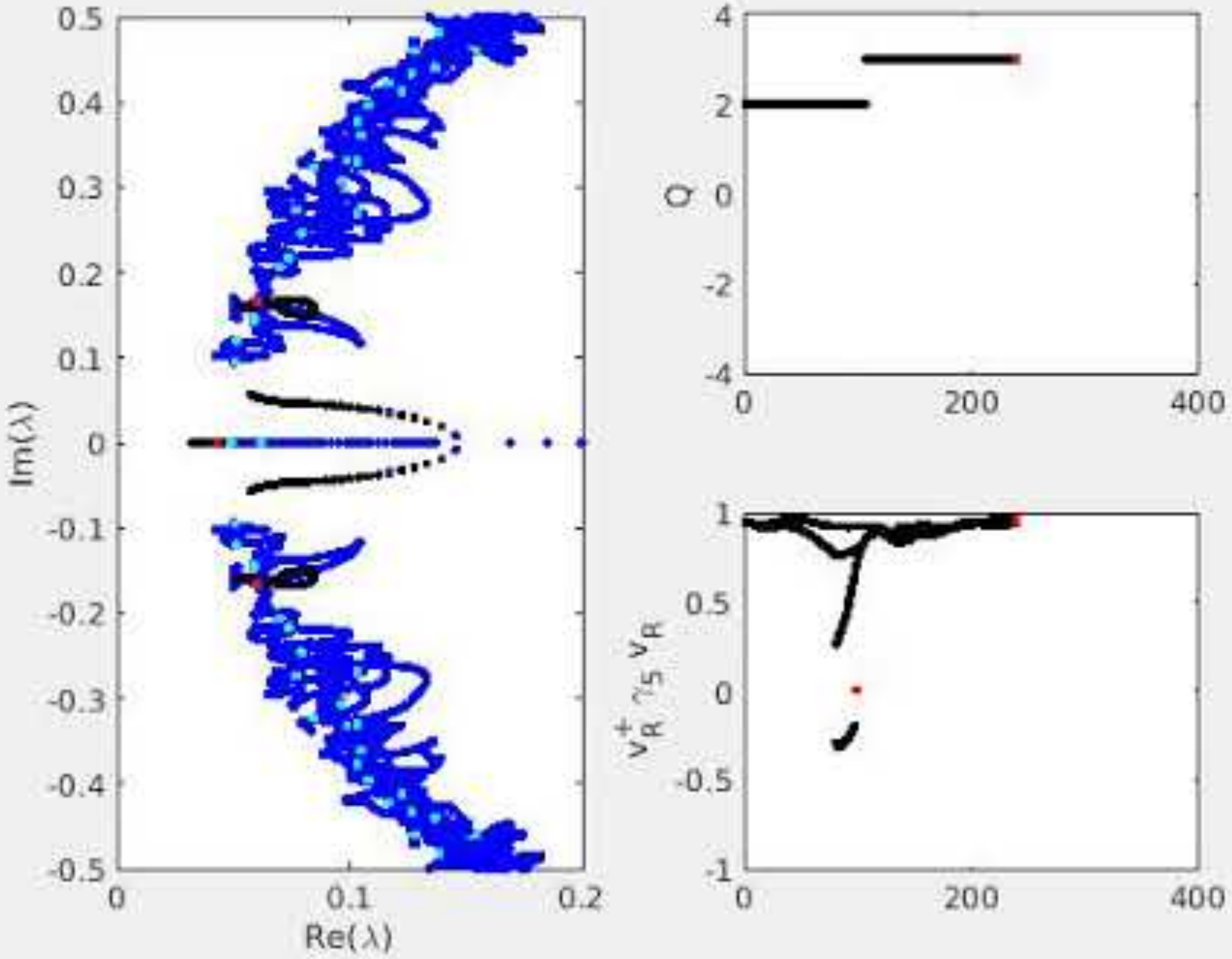
with the geometric definition: $Q^{geo} = \frac{1}{2\pi}\sum_x \theta_{12}(x)$

and $\mathrm{Index}(D) = N_R - N_L = \sum_i \chi_i|_{\lambda(D)=0}$

$$\chi_i \equiv \mathrm{sign}(v^\dagger_{i,R}\gamma_5 v_{i,R})$$

Microcanonical simulations suppresses Q transition

$$U \longrightarrow U'$$
$$H(U) \equiv H(U') = P^2 + \beta S(U') - 2\sum_i \ln \lambda_i$$



for 2D Schwinger see e.g. P. Hernandez *Nucl.Phys.B* 536 (1998)]

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

General structure:

1. Propose U' according to $T_0(U \to U')$
2. Correct with $P_{acc}(U \to U') = \min\left[1, \frac{\tilde{\rho}(U)\rho(U')}{\rho(U)\tilde{\rho}(U')}\right]$

In case ratio of distributions $(\tilde{\rho}(U)\rho(U'))/(\rho(U)\tilde{\rho}(U'))$ is log-normal distributed.

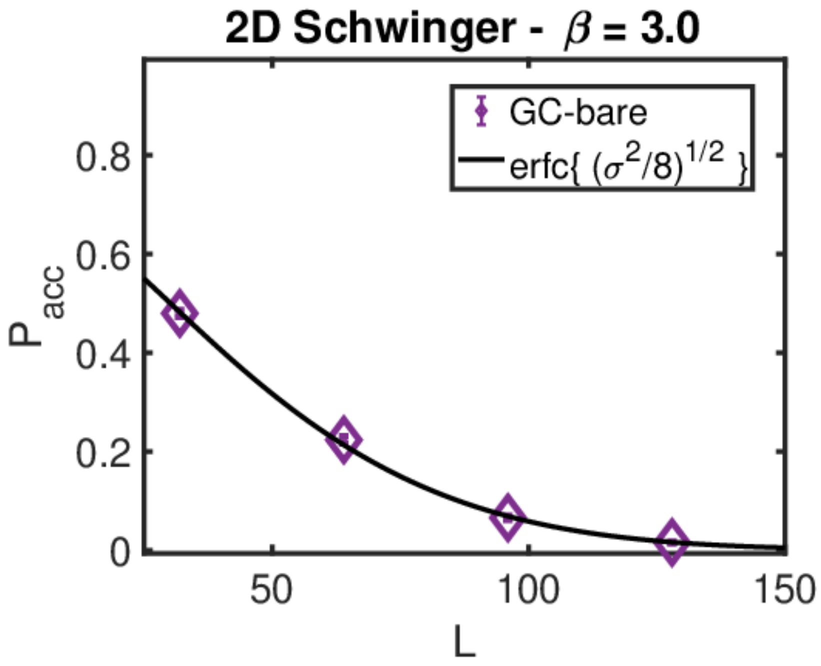
- for the acceptance rate follow

Creutz, Phys. Rev. D38 (1988) 1228–1238

$$P_{acc} = \mathrm{erfc}\{\sqrt{\sigma^2(\Delta S)/8}\}$$

with $\Delta S = \ln\{\rho(U')\} - \ln\{\rho(U)\} + \ln\{\tilde{\rho}(U)\} - \ln\{\tilde{\rho}(U')\}$

where $\ln(\rho(U))$ is an extensive quantity, thus $\sigma^2(\Delta S) \propto V$



- GC-step is very fast ineffective :

$$P_{acc} \to e^{-V}$$

aSTRC

How to control $\sigma^2(\Delta S)$

1. by using correlations between ρ and $\tilde{\rho}$
2. by reduction of degrees of freedom of ρ and $\tilde{\rho}$

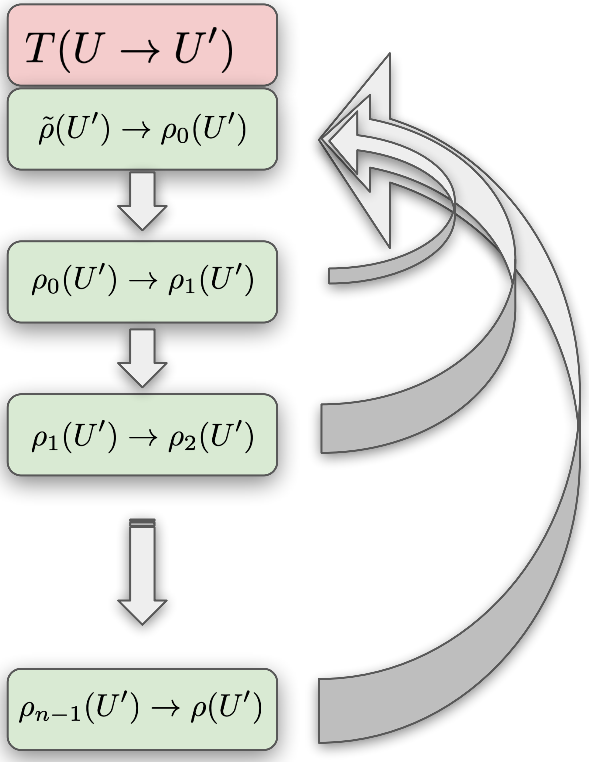
Generalization leads to factorization with parametrization of ρ via

$$\rho_n(U) = P_0(U,\alpha_i^{(0)})\, P_1(U,\alpha_i^{(1)})\, \ldots P_n(U,\alpha_i^{(n)})$$

and GC step is spliting up into *n* successive steps

$$P^i_{acc}(U \to U') = \min\left[1, \frac{\rho_{j-1}(U,\alpha_i^{(j-1)})\rho_j(U',,\alpha_i^{(j)})}{\rho_j(U,,\alpha_i^{(j)})\rho_{j-1}(U',\alpha_i^{(j-1)})}\right]$$

- Iterate each step to filter out local fluctuations

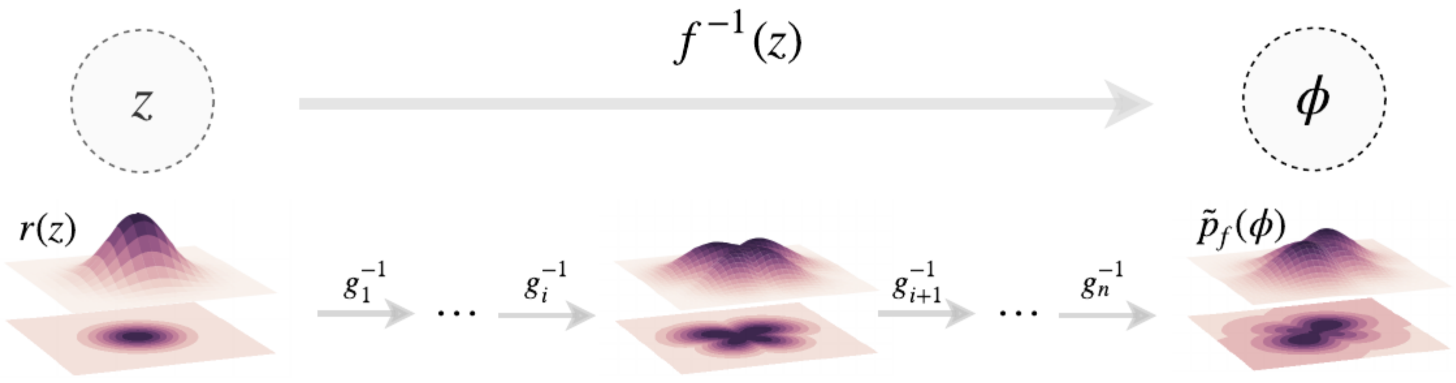


THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

An example: Generative model in U(1) with gauge invariant flow

Idee:
Use a flow map $f^{-1}(z)$ to propose new configurations with known distribution

$$\tilde{p}(\phi) = r(f(\phi)) \cdot \left| \det \frac{\partial f(\phi)}{\partial \phi} \right|$$



- introduce coupling layers with

$$g_i^{-1}(z) := \begin{cases} \phi_a = z_a \\ \phi_b = (z_b - t_i(z_a)) \odot e^{-s_i(z_a)}. \end{cases}$$

- train the coupling layers (s,t) by minimizing the loss-function

$$\begin{aligned} L(\tilde{P}) &:= D_{KL}(\tilde{P}||p) - \log Z \\ &= \int \prod_j d\phi_j \, \tilde{P}(\phi)(\log \tilde{P}(\phi) + S(\phi)). \end{aligned}$$

successfully applied to ultra local 2D discrete lattice models by
  - ϕ^4 Albergo et al., Phys.Rev.D 100 (2019) 3, 034515
  - U(1), Kanwar et al., Phys.Rev.Lett.125 (2020) 12, 121601
  - SU(2), SU(3) Boyda et al., Phys.Rev.D 103 (2021) 7, 074504
- can overcome critical slowing down

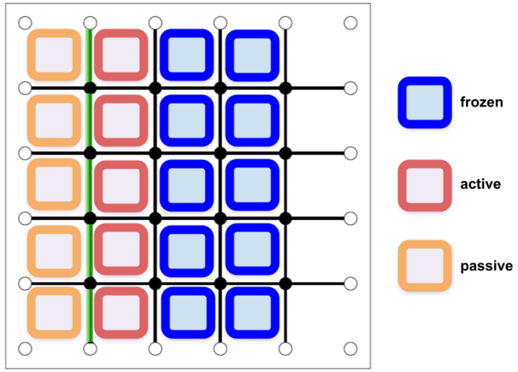
Albergo et al., arXiv:2101.08176

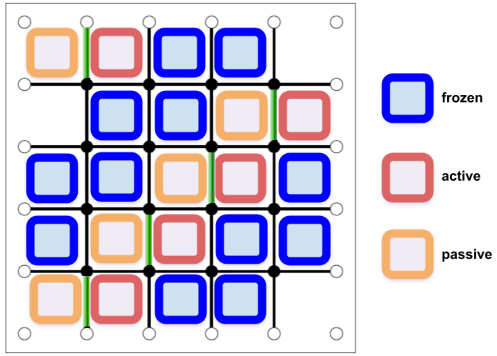
aSTORC

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

How to design coupling layers:

$$g_i^{-1}(z) := \begin{cases} \phi_a = z_a \\ \phi_b = (z_b - t_i(z_a)) \odot e^{-s_i(z_a)}. \end{cases}$$



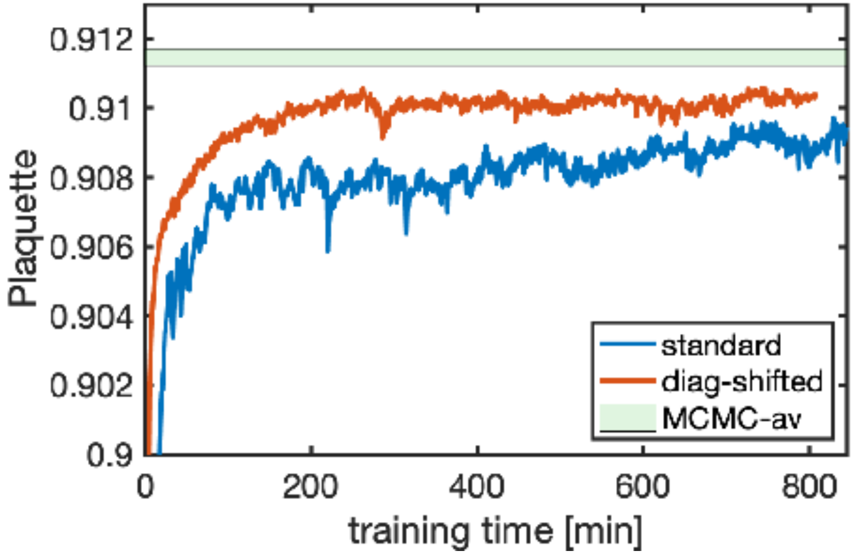


t_i and s_i consists of neural networks

- Can be design to contain symmetries
  - Gauge invariant by masks and proposing plaquettes
  - Partially translation invariant by convolutional networks

Structure of networks

- convolutional kernels with size 3
  - note that only frozen plaquettes are used as input values
- with hidden layers (here default 2 with 8 nodes)
- 8 coupling layers corresponds to a full update



Updating masks improve convergence rate and acceptance from 30% to 50%

aSTøRC

Let's defined our minimization condition:

The loss function:

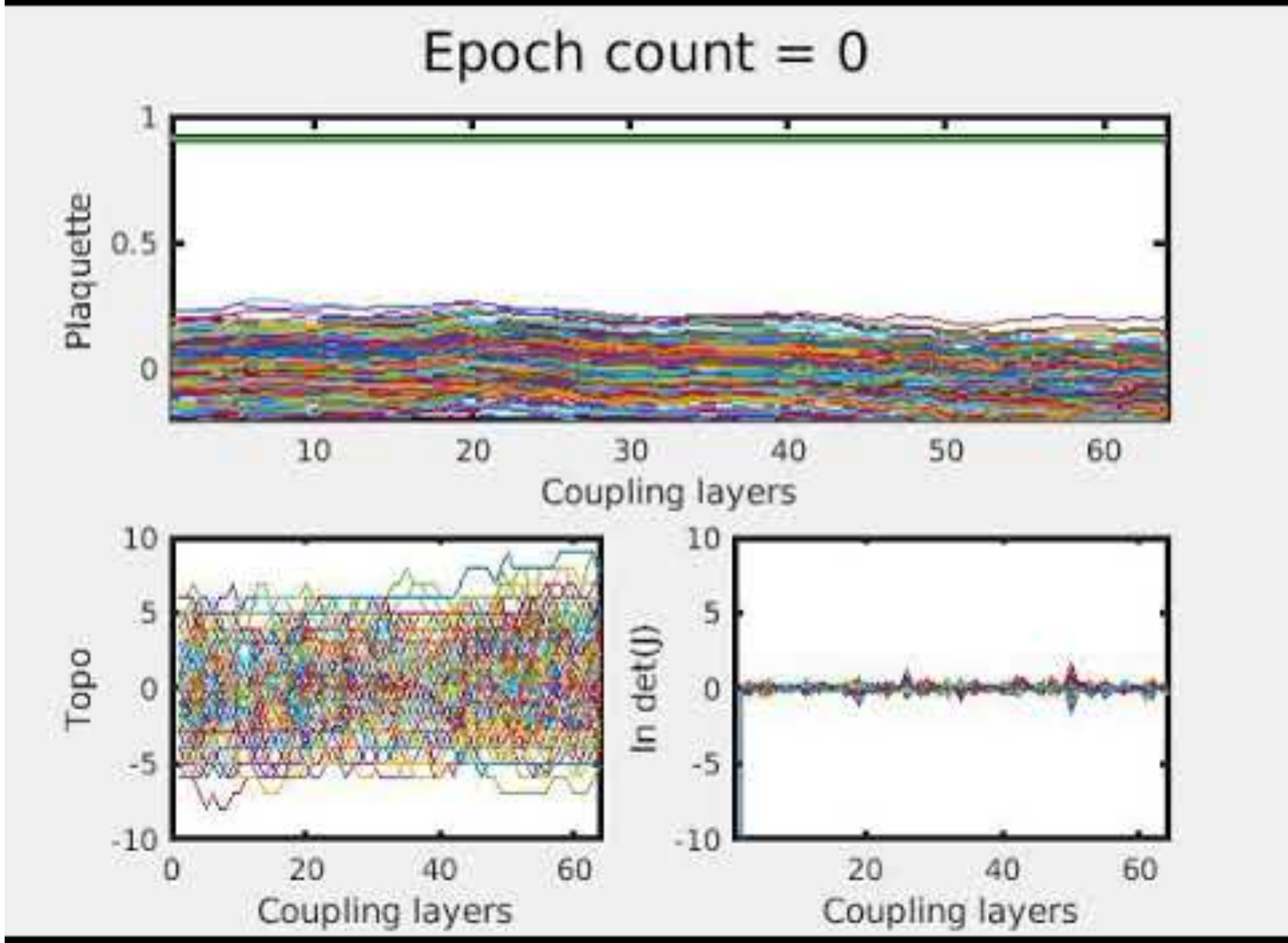
$$\begin{aligned} L(\tilde{P}) &:= D_{KL}(\tilde{P}||p) - \log Z \\ &= \int \prod_j d\phi_j \, \tilde{P}(\phi)(\log \tilde{P}(\phi) + S(\phi)). \end{aligned}$$

- with ultra-local plaquette action:

$$\ln(\rho(U)) = -\beta \sum_x P_{12}(U)$$

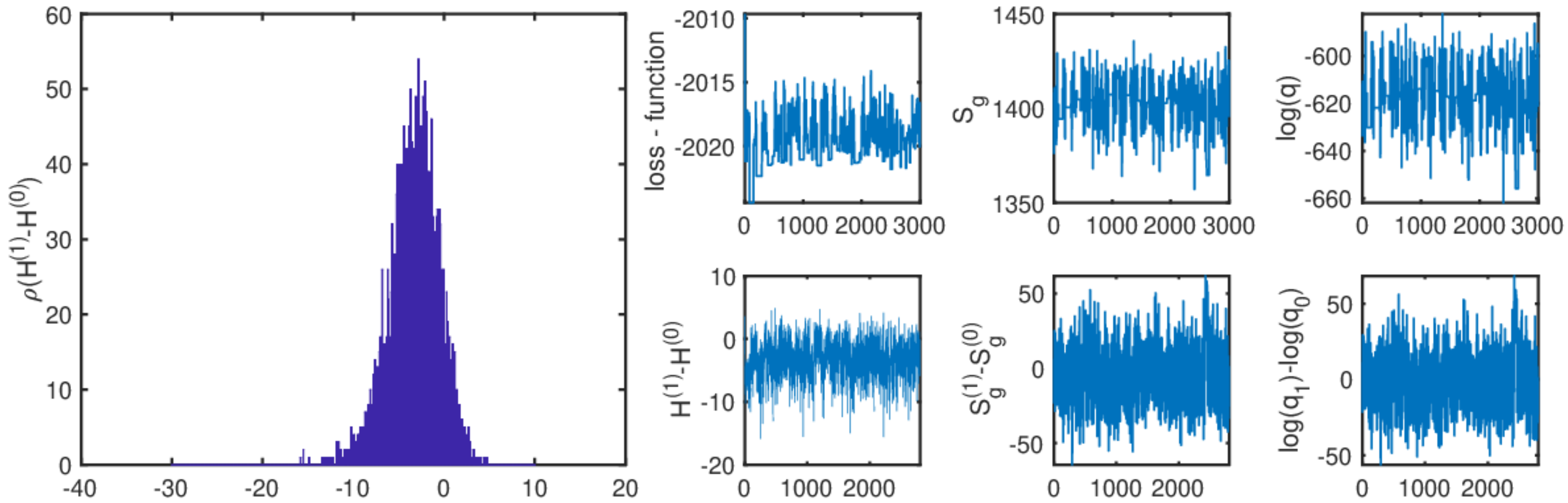
- and flow distribution:

$$\tilde{\rho}(U) = \rho_{trival}(m^{-1}(U)) \prod_j \det J(g_j^{-1}(\alpha_{i,j}^{(0)}))$$



Correlations of distribution $\tilde{\rho}$ and ρ

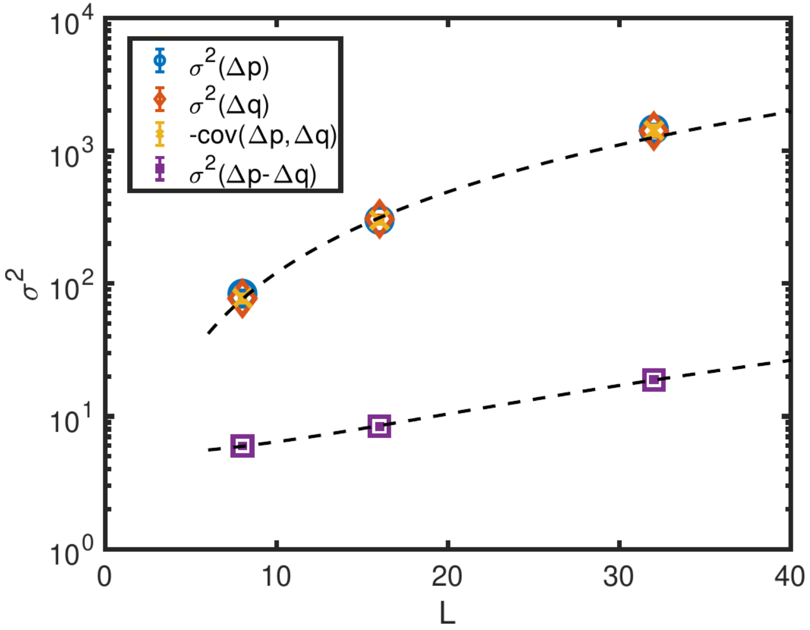
- covariance need to be of $\mathrm{cov}(\tilde{\rho}, \rho) \propto \mathcal{O}(V)$ to compensate extensive variances $\sigma^2 \propto \mathcal{O}(V)$

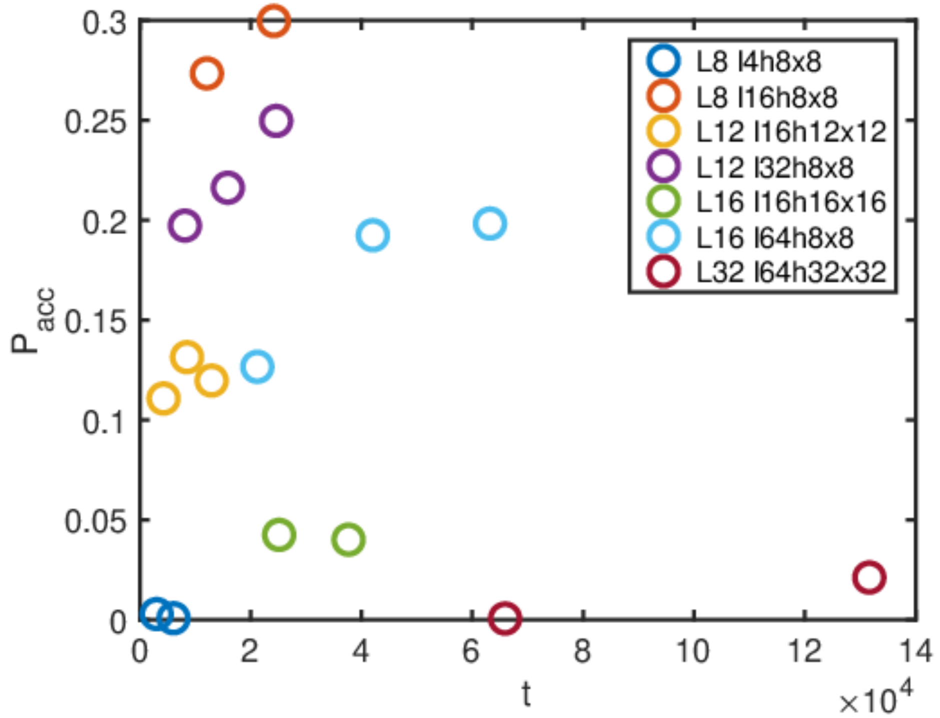


Works for L=8 → L=16

Volume scaling of gauge invariant flow:

- coupling layer dof are scaled with volume
  - *l*: coupling layers
  - *h*: hidden layers
- scaled *l* and *h* with V=L² while decreasing minimization rate





Fine tuning problem:

Covariances of distributions scales like variances

$$\mathrm{var}(\Delta p) + \mathrm{var}(\Delta q) \approx 2\mathrm{cov}(\Delta p, \Delta q)$$

But $\sigma^2 = \mathrm{var}(\Delta p) + \mathrm{var}(\Delta q) + 2 \cdot \mathrm{cov}(\Delta p, \Delta q)$
still grows with the volume

aSTRC

The Cyprus Institute
RESEARCH•TECHNOLOGY•INNOVATION

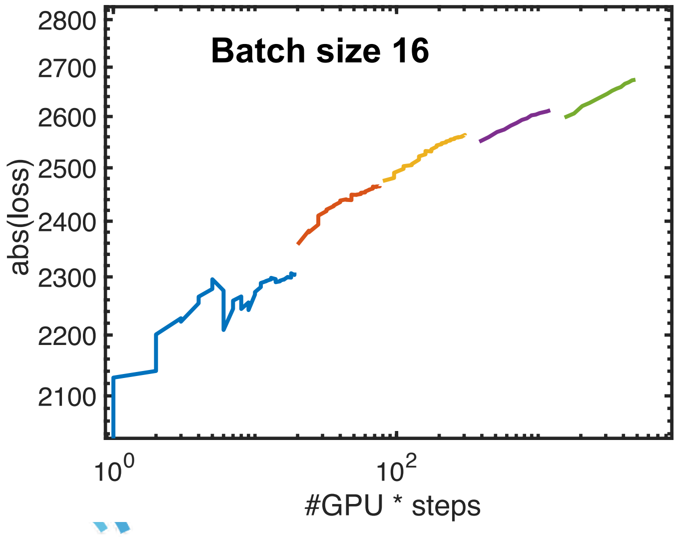
How to scale up:

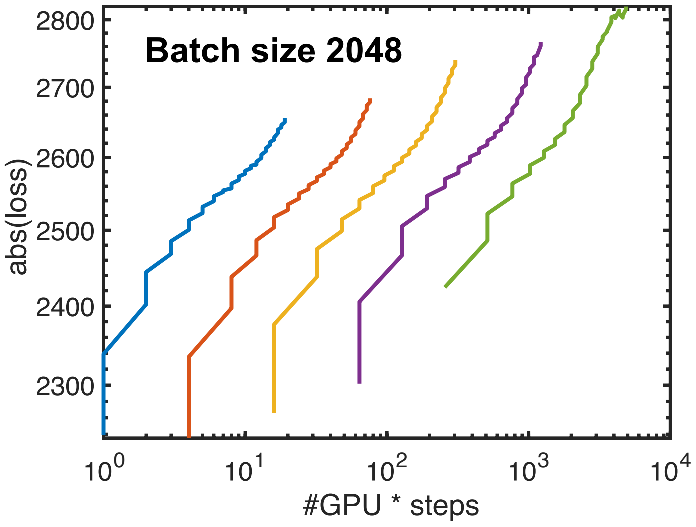
Exercise with horovod

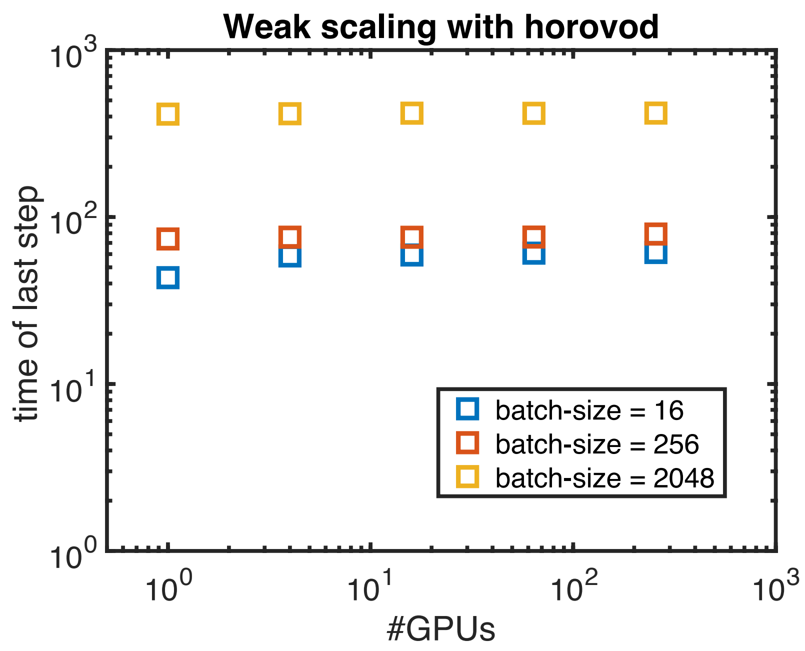
- Simple to implement but needs fine tuning
- adds new batch to each additional GPU
    - Total batch-size = #GPUs x local batch-size

Modifications:

- Switch to double precision
- Use Ada..
- Use stepsize decay







Benchmark runs on JUWELS-BOOSTER

- Loosely coupled scales weakly perfect
- For smaller batch-sizes works fine
- For larger batch-sizes convergence deteriorates

aSTORC

Scalability via Domain Decomposition

Lattice action are local

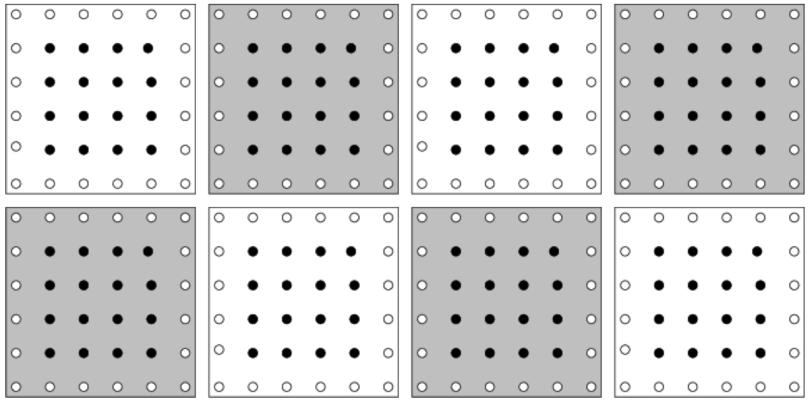
every highly optimized lattice algorithm are based on it

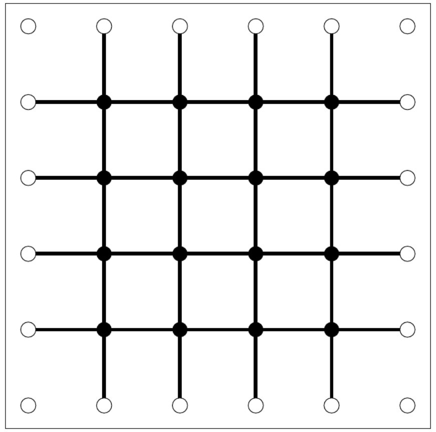
- multigrid, multi level, hierarchical probing, low-mode averaging, etc.

mainly based on Domain Decomposition of the lattice
then the ultra local plaquette action splits up into

$$S_g = -\beta \left\{ \sum_{b} \sum_{x \in b} P_{12}(x) + \sum_{\delta b} \sum_{x \in \delta b} P_{12}(x) \right\}$$

Plaquettes inside of the blocks   Plaquettes between blocks





Taken from: M. Luscher, CPC 165 (2005) 199-220

For the gauge invariant flow

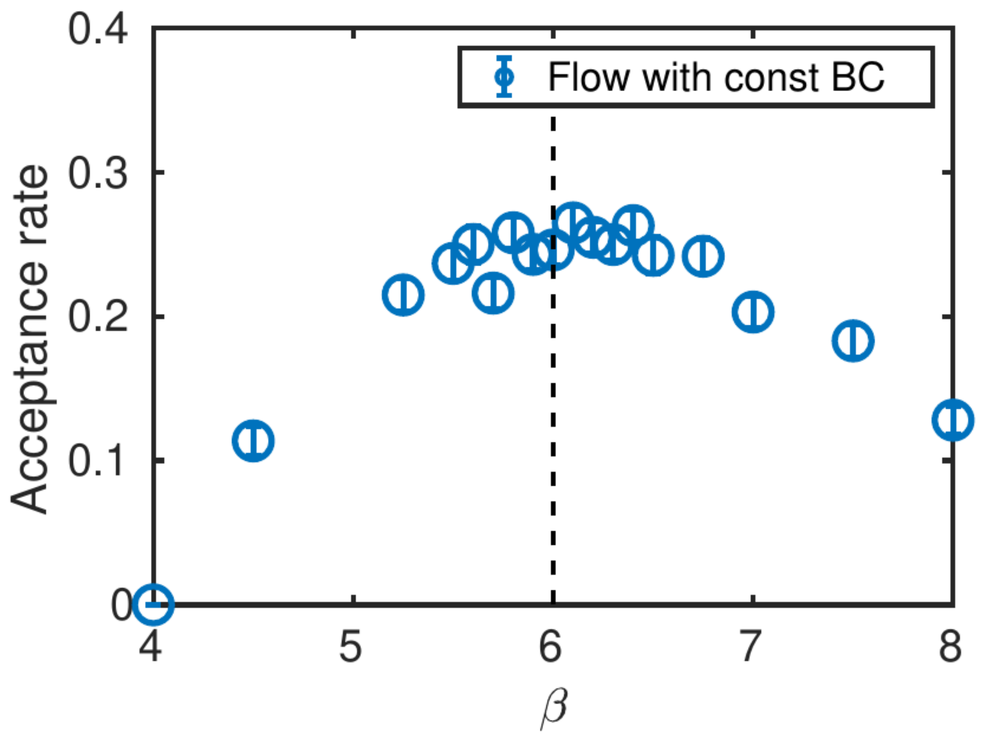
- update only links/plaquettes inside blocks
- create maps of active links within each block

Training within fixed domains

Adaptation of training procedure

By:

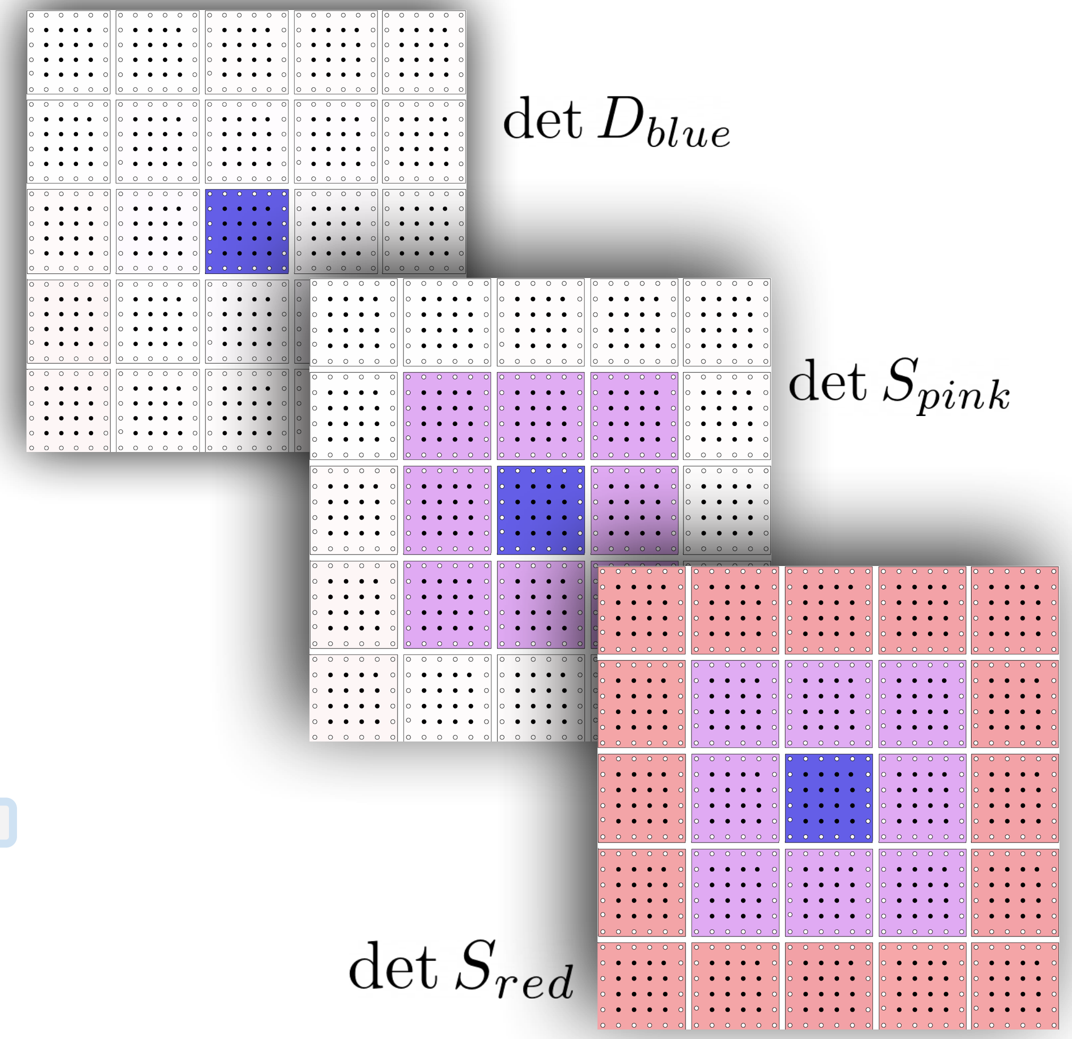
- Using the periodic trained model to generate boundaries or starting from random and shift lattice after each epoch
- Using different boundaries for each batch with total batch size 4096
- Increase iteration before boundaries updated to 1000
- Using diagonal masks to increase overlap with frozen plaquettes
  (faster convergence)



Acceptance rate of fixed boundaries drops down to ~25% with L = 8 (from 50% periodic case)

- due to the ultra locality of gauge action: larger volumes are trivial to generate

Recursive Domain Decomposition



Action with fermions:

$$P(U) = Z^{-1} \left(\prod_j^{N_f} \det D_j(U) \right) e^{-\beta S_g(U)}$$

with $\det D(U)$ is a *localised* action

- distance interaction decays with

$$cov(x, y) \propto \exp\{-m_{PS}|x-y|\}$$

Idea: using exact decomposition of fermion action:

$$\det D = \det S_{red} \cdot \det S_{pink} \cdot \det D_{blue}$$

effective long range decomposition of the fermion determinant

M. Luscher, CPC 165 (2005) 199-220

J. F. et al., CPC 184 (2013) 1522-1534

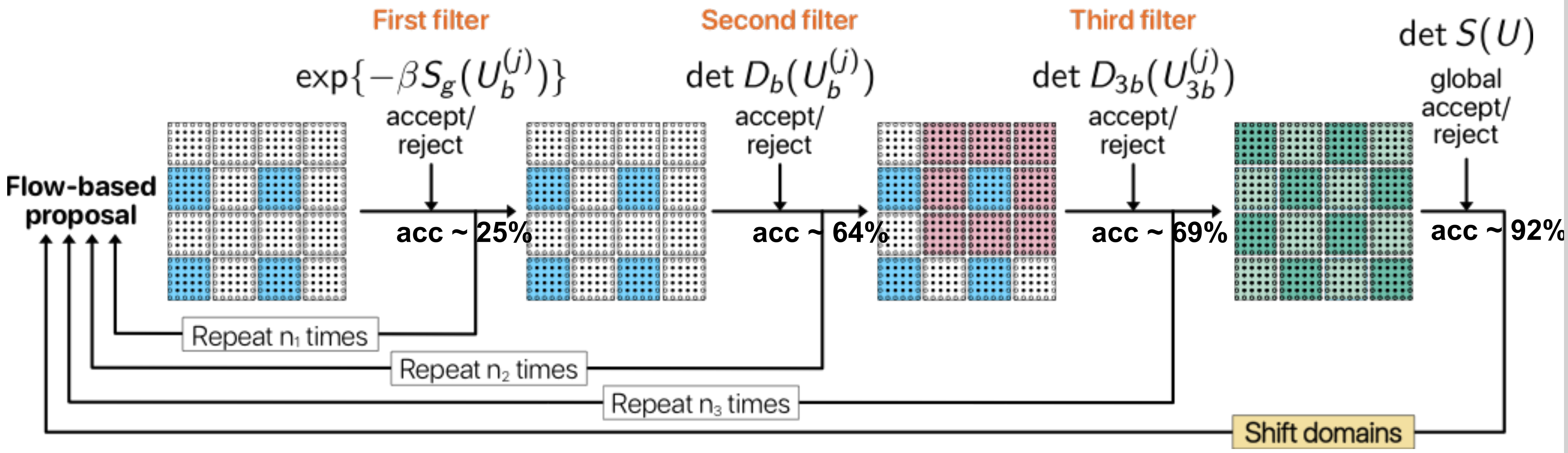
M. Cè et al., Phys.Rev.D 93 (2016) 9, 094507

M. Cè et al., Phys.Rev.D 95 (2017) 3, 034503

Towards high acceptance rate

Global Correction Monte Carlo algorithms with equivariant flows:

J. F., arxiv:2201.02216



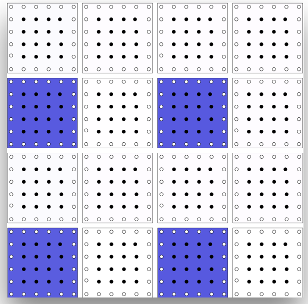
β	3.0	6.0	8.45
5 level flowGC with $d = 16$:			
Level 4			
with σ^2	0.0052	0.0369	0.0046
and P_{acc}	0.9713	0.9235	0.9727
$\delta\beta_4^{(3)}$	-2.0037	-2.0182	-2.0087
$\delta\beta_4^{(2)}$	1.0027	1.0061	1.0083
$\delta\beta_4^{(1)}$	-0.0003	0.0008	0.0004
Level 3	$n_1 = 2$		
with σ^2	0.6688	0.6190	0.1546
and P_{acc}	0.6826	0.6940	0.8441
$\delta\beta_3^{(2)}$	-1.1730	-1.3635	-1.3534
$\delta\beta_3^{(1)}$	-0.0006	0.0149	0.0125
Level 2	$n_2 = 4$		
with σ^2	1.4384	0.8325	0.1857
and P_{acc}	0.5487	0.6482	0.8294
$\delta\beta_2^{(1)}$	-0.2482	-0.3082	-0.2863
Level 1	$n_1 = 100$		
with P_{acc}	0.5669	0.2501	0.2794
2 level GC:			
with σ^2	12.3774	9.7119	3.7260
and P_{acc}	0.0786	0.1192	0.3345

Multilevel hierarchical filter steps with 4 levels

Enhancing acceptance rate by

- within level 1, 2, 3 each active block can be updated independently from each other
- use correlation between actions via parameterization,
  - e.g. for the gauge coupling

Towards high acceptance rate



Acceptance rate:

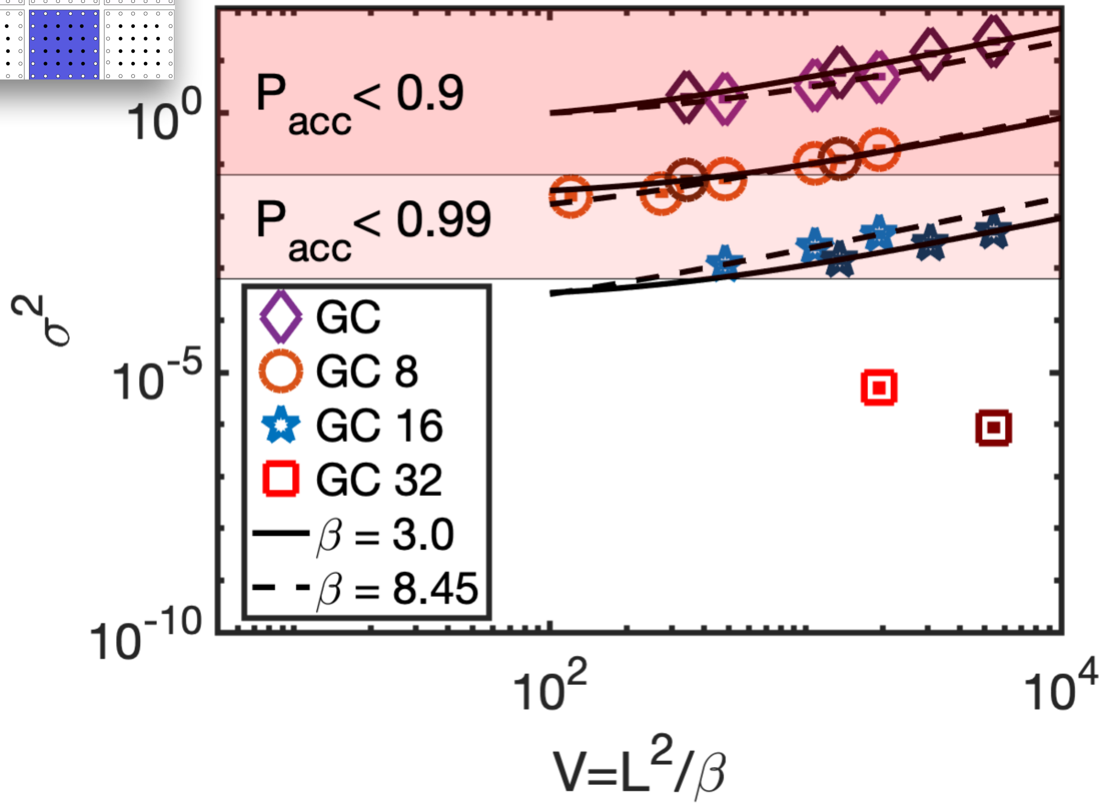
- select L=8 flow proposals
- updating every 4th block, which introduces a distance between active blocks by d = Lbs
  which results into 16% of links updated per step

(independent of global volume!)

runs for different Lbs = 8, 16, 32 with 4 lvl filter steps

- variance is very efficient reduced for larger Lbs
- volume scaling remains

How a change of 16% influence sampling rates ?



aSTORC

Topological charge:

Usually we are using the autocorrelation time for comparison, but HMC freezes and τ_{int} is not measurable

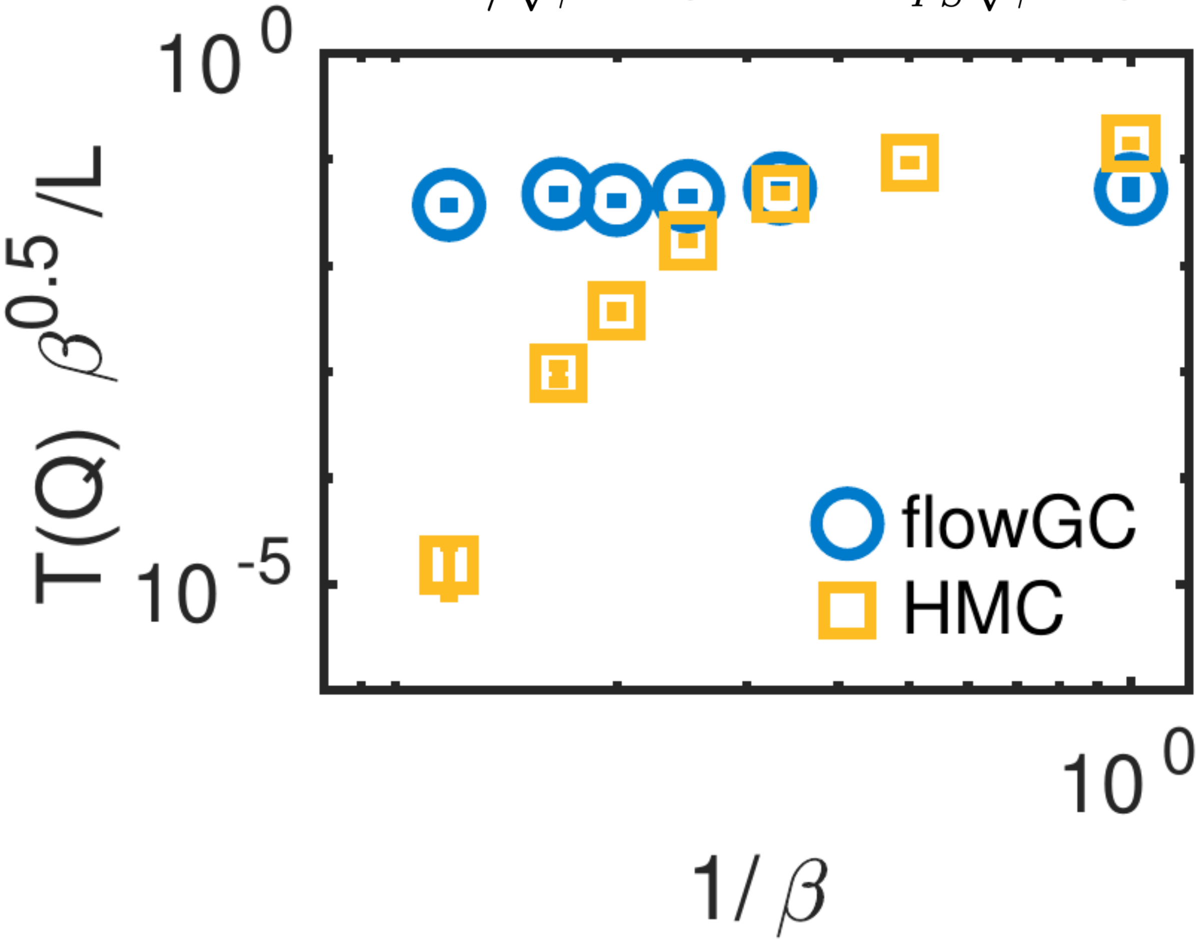
Instead one can define a tunneling rate:

$$T(Q) = \langle |Q_i - Q_{i+1}| \rangle$$

GC shows no critical slowing down and topological tunneling scales

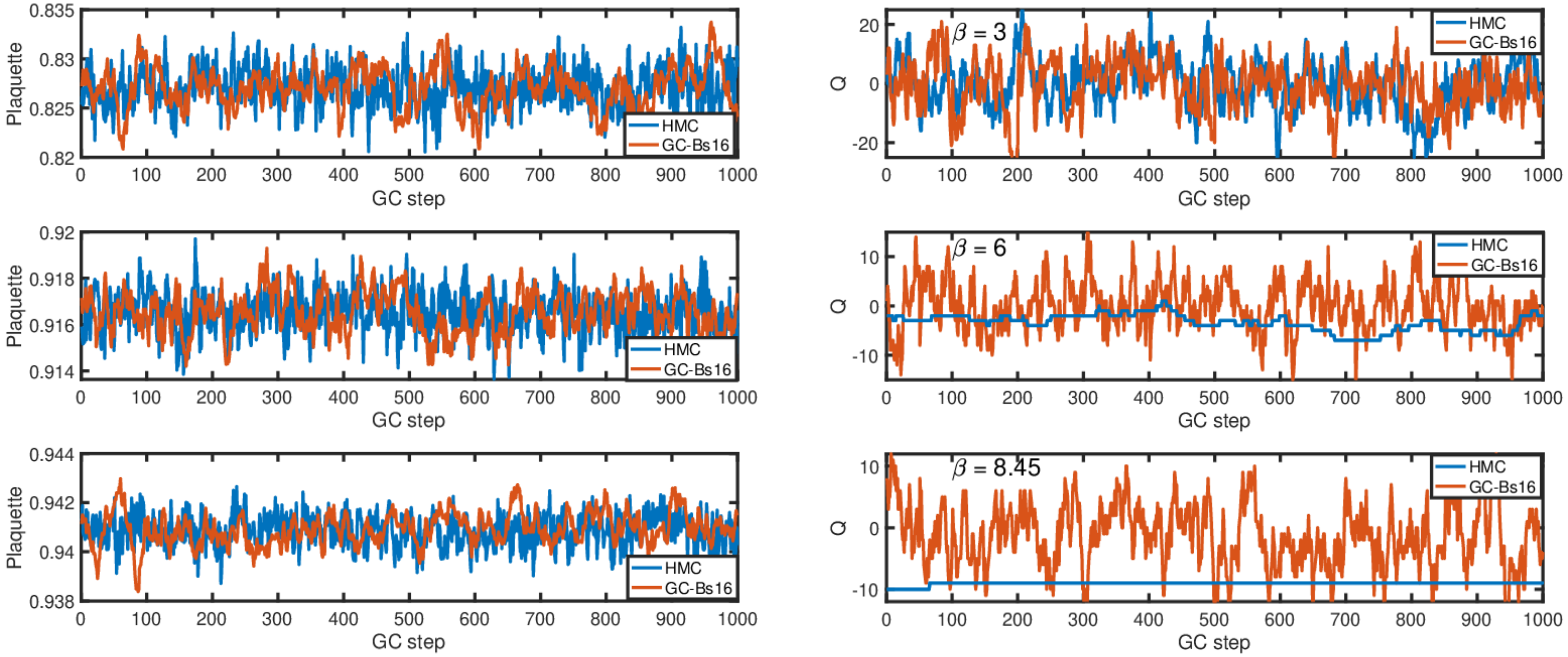
At constant line of physics:

$L/\sqrt{\beta} \sim 40$ and $m_{PS}\sqrt{\beta} \sim 0.4$



Plaquette and topological charge history

Runs at L=128



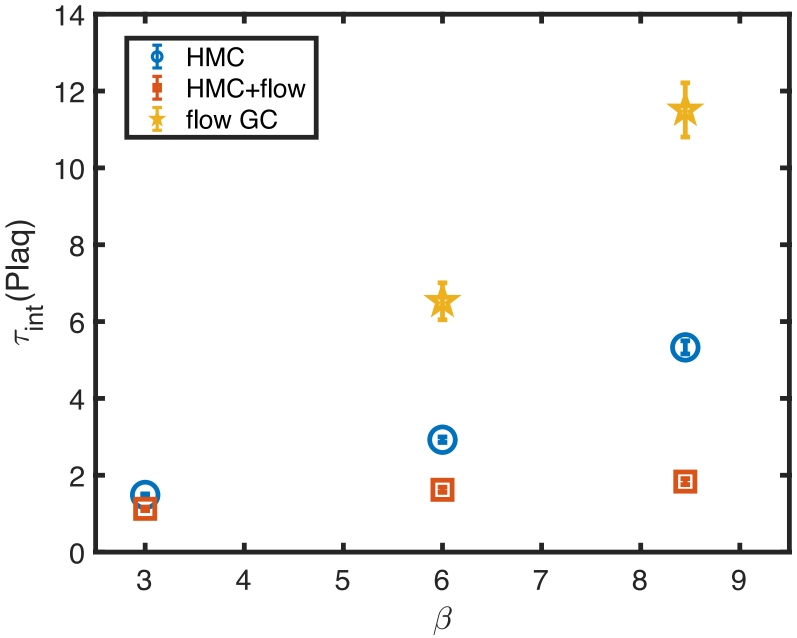
Combination with HMC

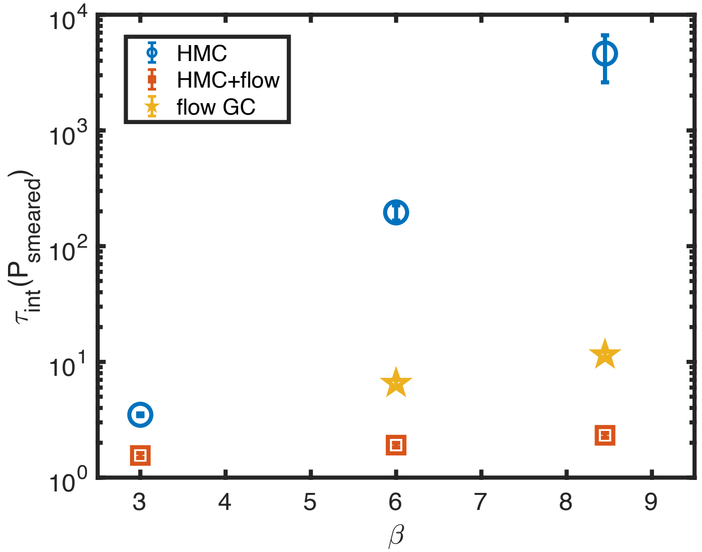
The Cyprus Institute
RESEARCH•TECHNOLOGY•INNOVATION

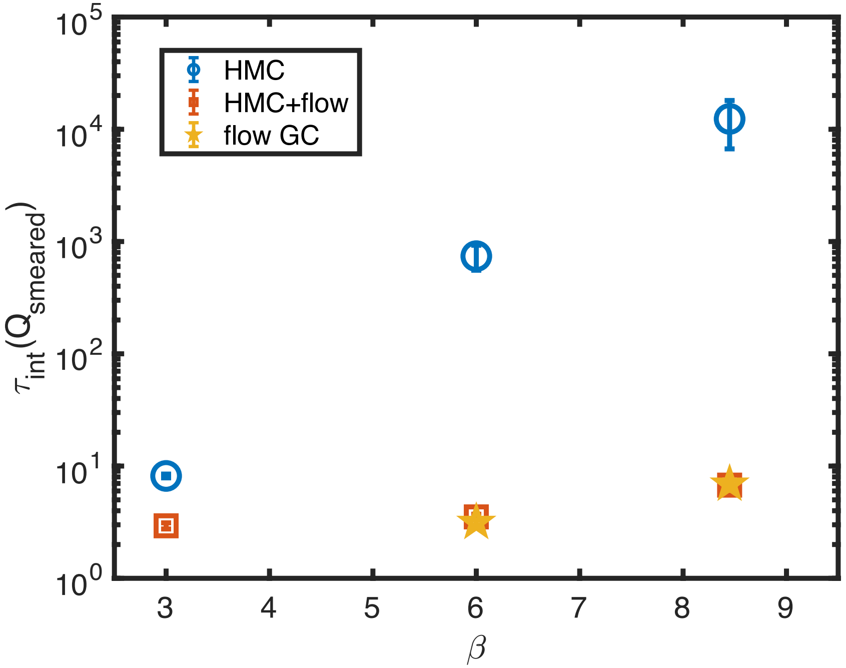
Idea: combination with HMC
and high statistic runs

Similar to D. Albandea et al., Eur.Phys.J.C 81 (2021) 10, 873

→ HMC step → flow GC step → HMC step →







with $\tau_{int}(P_0) \propto \beta^c$

- HMC : c = -0.5
- HMC+flow: c = -0.5
- flow GC : c = -1.5

HMC+flow:
Outperforming other Methods

Runs done on L=32
No constant line of physics

with $\tau_{int}(Q) \propto \beta^d$

- HMC : d = -7.0
- HMC+flow: d = -0.8
- flow GC : d = -2.0

aSTORC

GC+flow proposal can solve critical slowing down in the 2D Schwinger Model

Major challenges addressed

J. F., arXiv:2201.02216

- very high acceptance rate by keeping 16% of links active towards large volumes
- Tunneling rate of topological charge relative constant towards finer lattice spacings

Combination with HMC promising towards more complex and larger models
Which depends on:

- Flow proposals within 4D with SU(3)
- Block acceptance can break down (so far 6^4 are reached)
  flow proposals with fermions should help

J. F. et al., CPC 184 (2013) 1522-1534

M. Albergo et al., Phys.Rev.D 104 (2021) 11,11450

R. Abbott et al., arXiv:2207.08945

Normalising flows: Volume scaling needs to be addressed

- Parameter/function/method space is large
  - A lot of possibilities/potential : training procedure, mapping, factorizations …
  - … but there is the danger of the parameter/methods desert

P. Shanahan, Talk, 16.08, 10:40

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

GC - steps - Status

Techniques introduced in J. F. et al., CPC 184 (2013) 1522-1534
Factorisation of determinant and its computation

- Use LU until L=4
- Use Stochastic estimators for L>4
  - Only one source per ratio (need for rel. gauge fixing)

New developments (so far not implemented): M. Cè et al., Phys.Rev.D 95 (2017) 3, 034503

- increase distances between active domains L. Guisti et al., Phys. Let. B 829 (2022) 137103
- Use GC-steps as topological tunnelling steps and not as full MCMC method

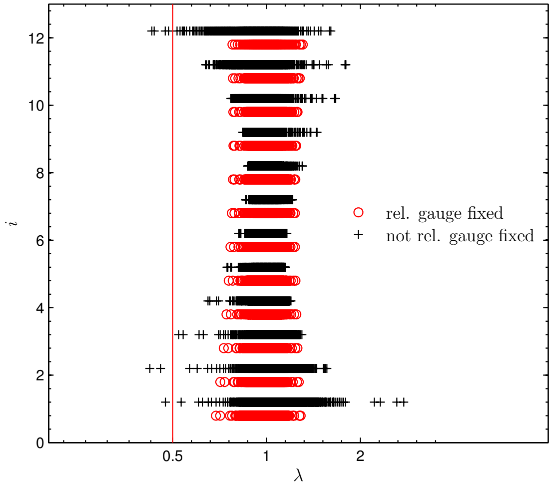
New implementation for an efficient steps
This should/could include:

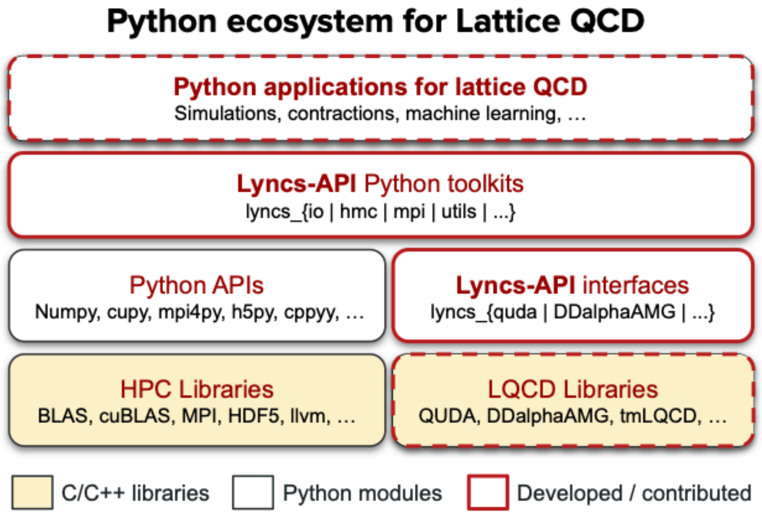
- Flexible parallelisation techniques
  - Decomposition is not equally distribute computing
    - Active domains are computational hot spots
- Modularity
  - LU-decomposition requires thick nodes
  - Sparse matrix inversions more efficient on GPUs

Included in design of lyncs
- requires lyncs-GC

?

Spectrum of Dirac Operator under relative gaugefixing





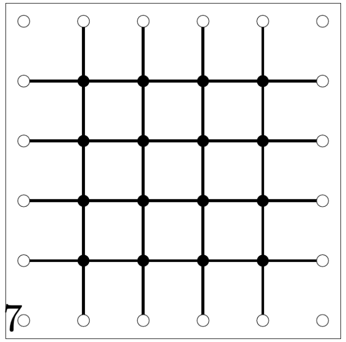
aSToRC

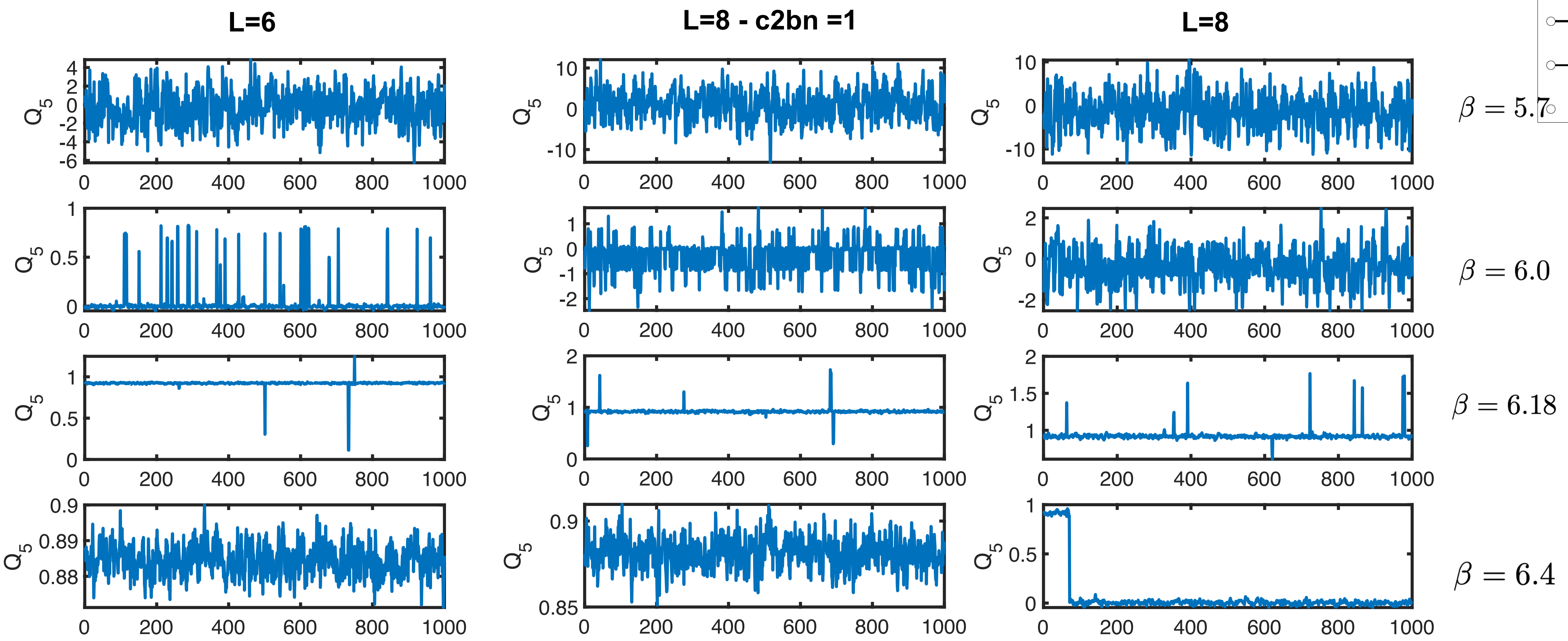
Discussions - Towards QCD

SU(3) - domain size

How large has to be the block ?Roughly L > 0.4 fm, which is ~10^4 at a=0.04 fm

- HB-Overrelaxation study seems to confirm that (here 8^4 within 16^4)





SU(3) - updates

Need for an update procedure which can (ideally guarantee) tunnelling of topology

- Generative models
- Continuous flows
- Instanton-updates (seems not to work)
  - maybe in combination with flows
- Re-thermalization (brute force)
- Local HMC (brute force)

- **at the physical point**
  - require at least 1 fm distance between active domains
    - Within a 5 fm box
      - 162 blocks of size 2.5 fm possible
      - should be okay (if acceptance rate is fine)

M. Dalla Brida et al., Phys.Lett.B 816 (2021) 136191

How large are the costs ?

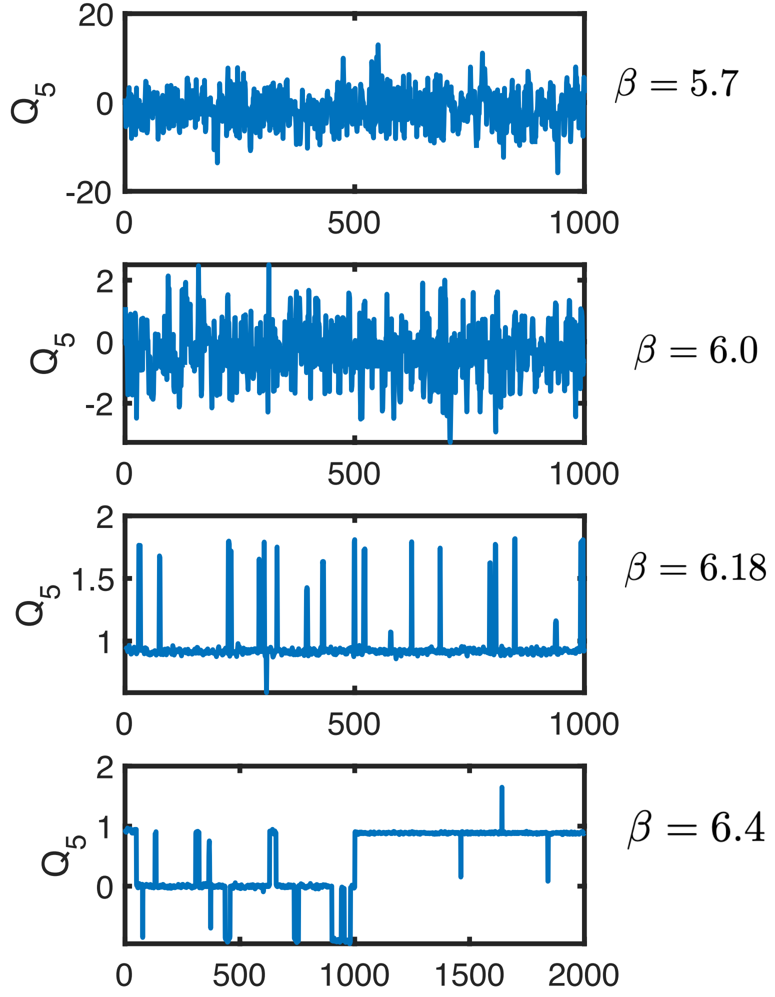
GC: nested accept-reject steps will scale with the most expensive step

- PG step or local determinant (potential V^2 scaling)

Multi-level/HMC-updates: of the larger domains at 0.04 fm

- Scales with 162*60^4 (~ 8x 128^4)

HB+OR re-thermalization on L=8



Not clear in the moment which method will work

Note that with computing at the exascale

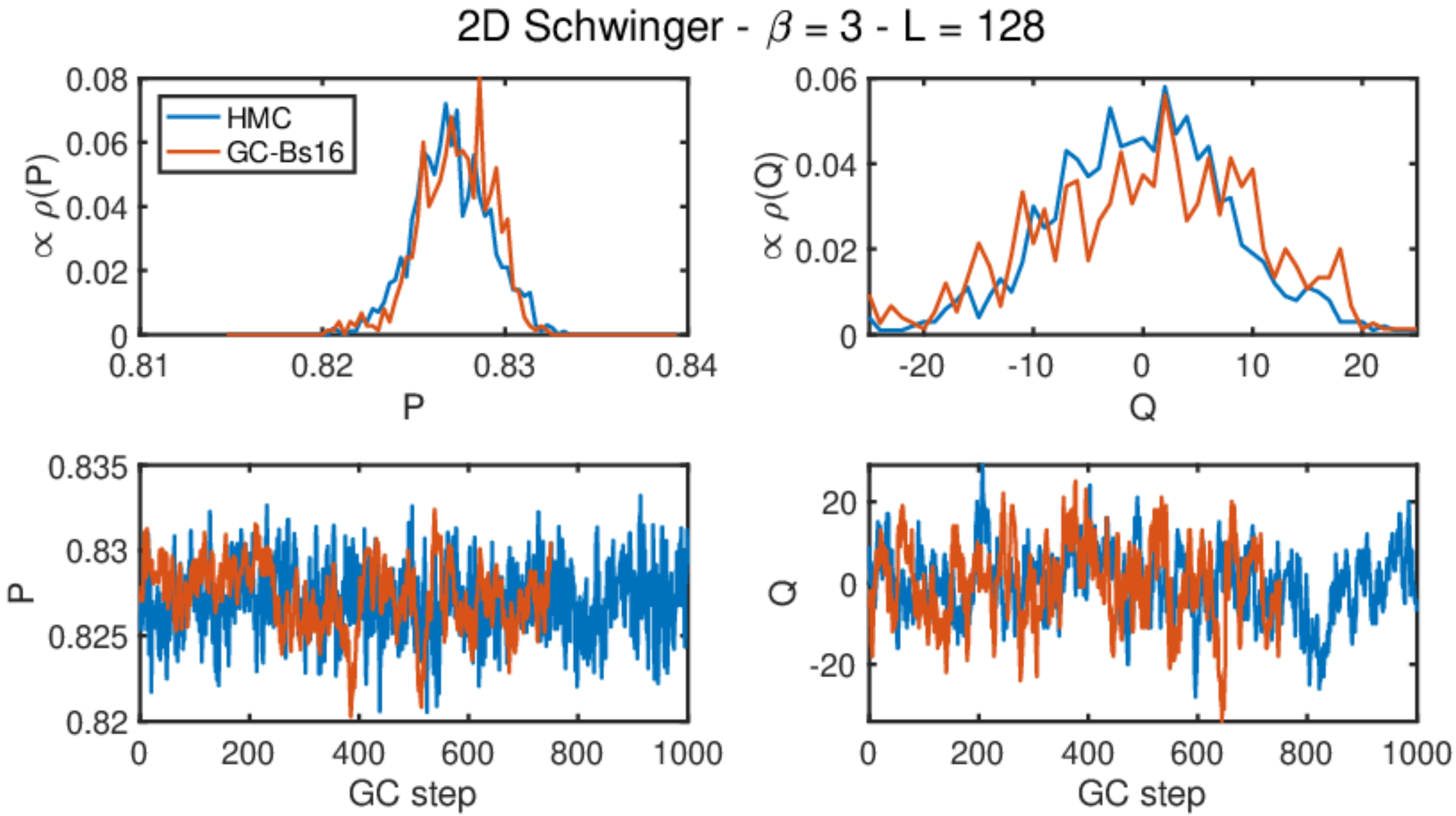
- Computational resources available to run 10k MDU for L=128 at physical pion masses reaching a < 0.05 fm in reach if we can mild down topological freezing

… novel idea’s and implementation are needed.

Thank you

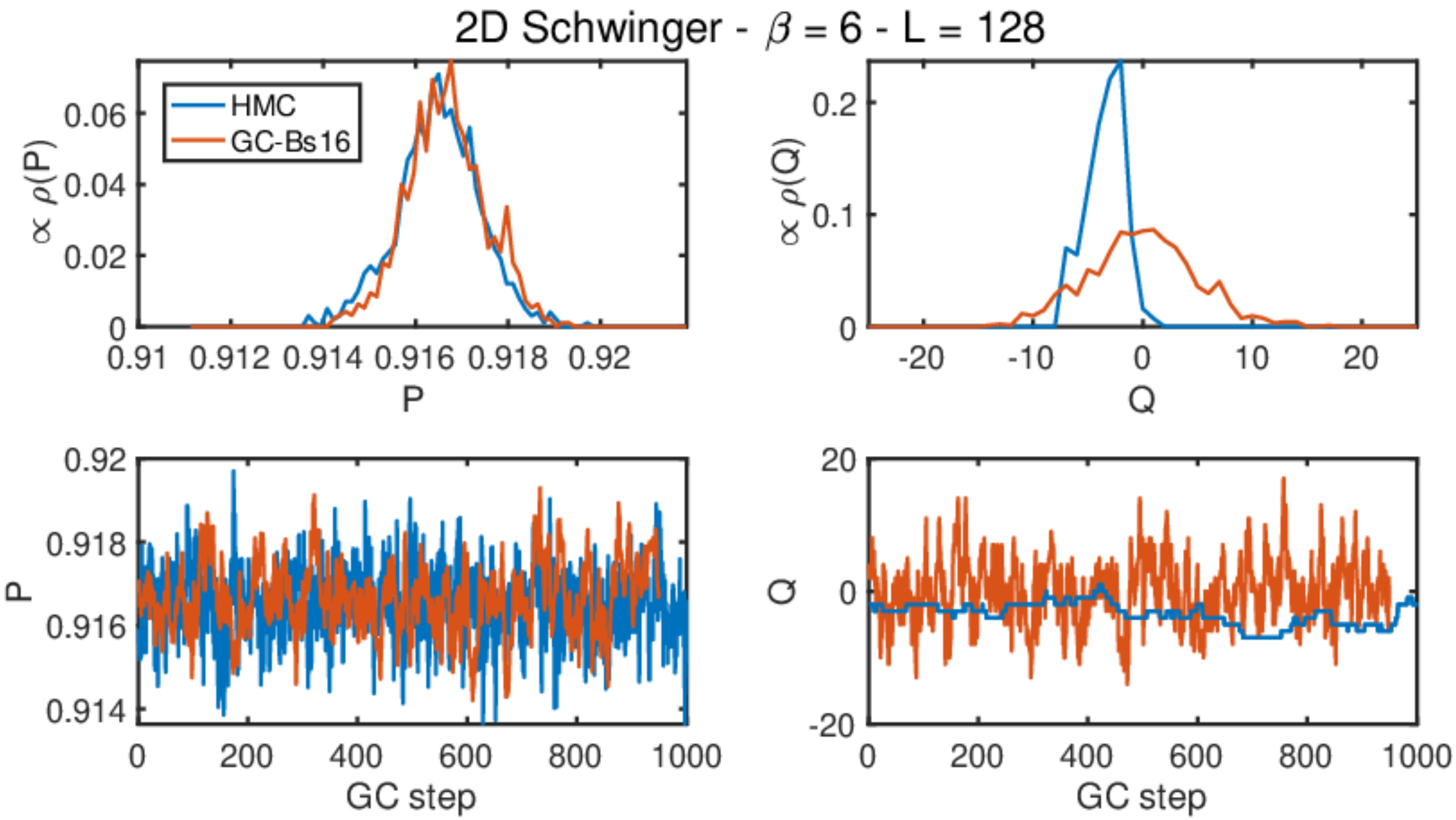
Appendix

- L = 128
- beta = 3.0
- m = -0.082626



Run Statistics

- L = 128
- beta = 6.0
- m = -0.0342



Run Statistics

- L = 128
- beta = 8.45
- m = 0.0