Review on Algorithms for Dynamical Fermions

Jacob Finkenrath

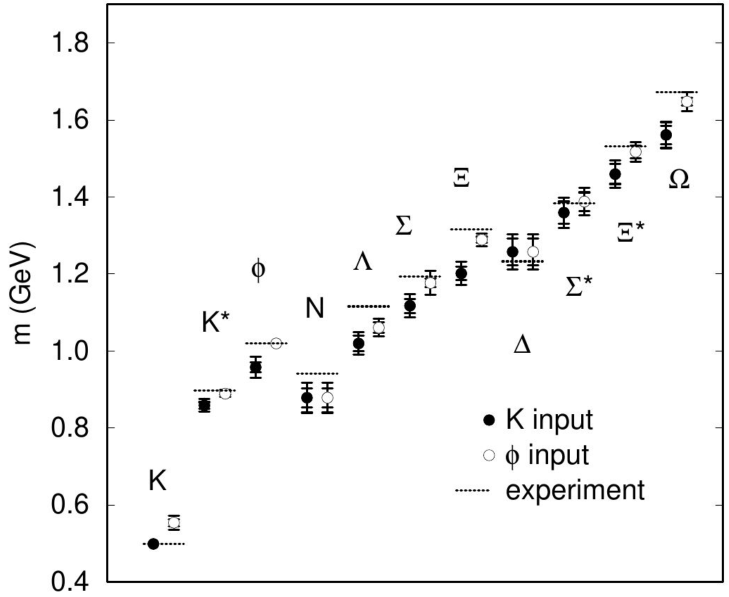
LATTICE 2022
AUGUST 8TH–13TH
BONN GERMANY

THE CYPRUS INSTITUTE
RESEARCH · TECHNOLOGY · INNOVATION

CaSToRC

New century - Era of dynamical fermions

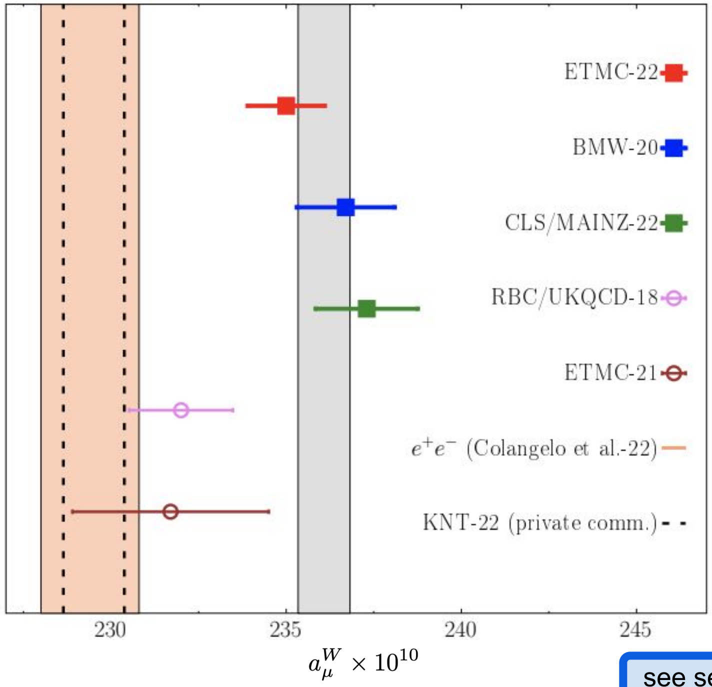
State of the art: ~ 2000
Hadron spectrum on large lattice
in quenched approximation



CP-PACS, *Phys.Rev.D* 67 (2003) 034503

❖ ~10% systematic effects due to neglecting fermion loops

State of the art: now
sub-percentage precision in intermediate
window of HVP with dynamical fermions



see session on Searches for BSM

❖ sub-percentage precision due to ensembles at physical pion mass
❖ remarkable consistency between independent lattice determinations

Physical point ensembles

Physical point ensembles:

- various lattice collaborations
- different fermion actions with Nf=2+1(+1)
- most > 5 fm
- between [0.05 - 0.2] fm

Lattice Data, Di. 14:00-16:00

Algorithm: Hybrid Monte Carlo

Duane et al., Phys. Lett. B195, 216 (1987)

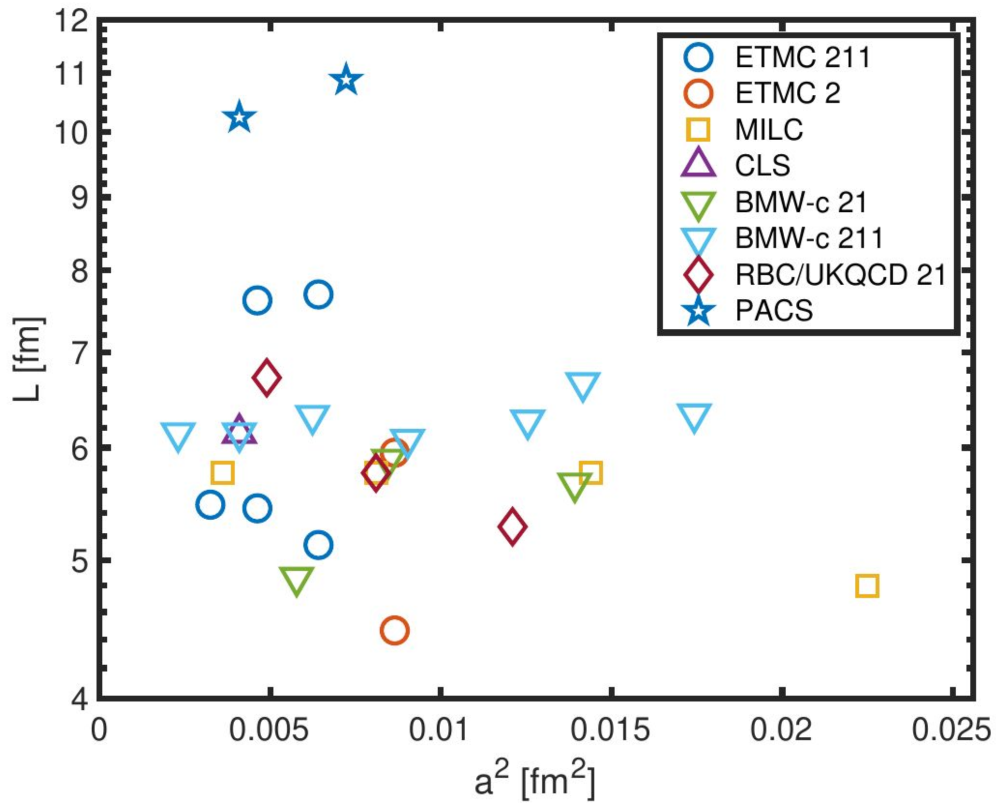
Gottlieb et al., Phys.Rev. D35 (1987) 2531

with variants:

- with IR/UV - preconditioning, even-odd-reduction
  - Hasenbusch-mass-preconding

    M. Hasenbusch, Phys.Lett.B 519 (2001) 177-182
  - Rational HMC

    Clark et al., Phys.Rev.Lett. 98 (2007) 051601



Challenges:

Two major systematics
- finite size effects
- finite discretization effects

required to minimize for new physics

N. Husung, Plenary, Saturday 8:50

N. Husung et al., Eur.Phys.J.C 80 (2020) 3, 200

M. Cè et al., JHEP 12 (2021) 215

- B- and Charm-physics
- multiple particle scattering
- g-2
- …

Minimize these effects by new ensembles
- generate large lattices > 8 fm
- generate finer lattices < 0.05 fm

Challenge
- large computational costs per MDU / realtime
- explosion of costs due to topological freezing

Snowmass target:
L=256, a = 0.04 fm

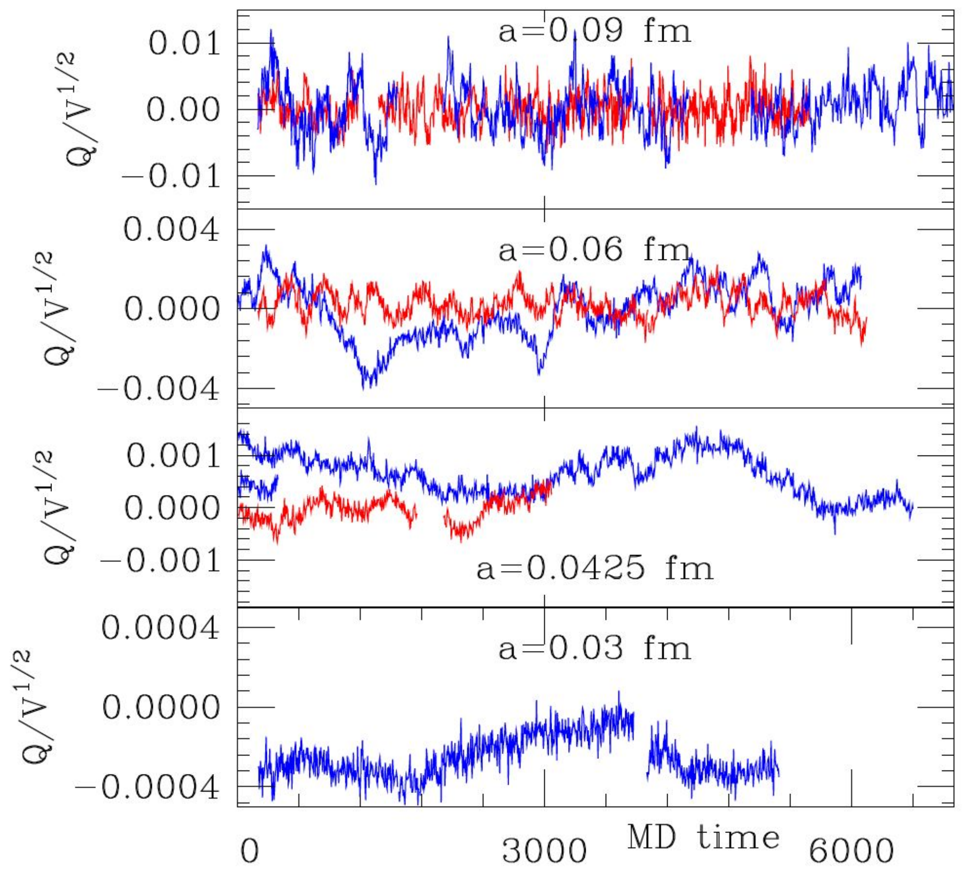
Snowmass 2022: arXiv:2205.15373

Snowmass 2022: arXiv:2204.00039

PACS (under production)
L=256, a = 0.043 fm

Y, Kuramashi, Lattice Data, Tuesday 14:50

Severe freezing towards fine lattice spacings



C. Bernard et al., Phys.Rev.D 97(2018) 7. 074502

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

F. Zimmermann, Algorithms, Thursday, 9:20

aSTRC

Table of Content

1. **HMC**
    - State of the art: Solvers
        - Conjugate gradient
        - Multigrid preconditioners
            - calculation on GPUs
    - Molecular dynamics
        - Symplectic integrators
            - Higher order integrators

1. **MCMC methods**
    - Methods without topological freezing
        - Modifications of molecular dynamics
        - Global corrections
            - Multi-level
            - Instanton update
        - Machine learning / Generative Models
            - normalizing flows

Proposal of a new set (U,P) via Hamilton's Equations:

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

Accept-reject step:

$$P_{acc} = \min\left[1, e^{-H(U_t)+H(U_0)}\right]$$

Challenge: Large lattices

How HMC scales towards larger lattices ?

- ❖ MD requires continues updates
  - ➢ speed up via strong scaling

limited by machines and algorithms

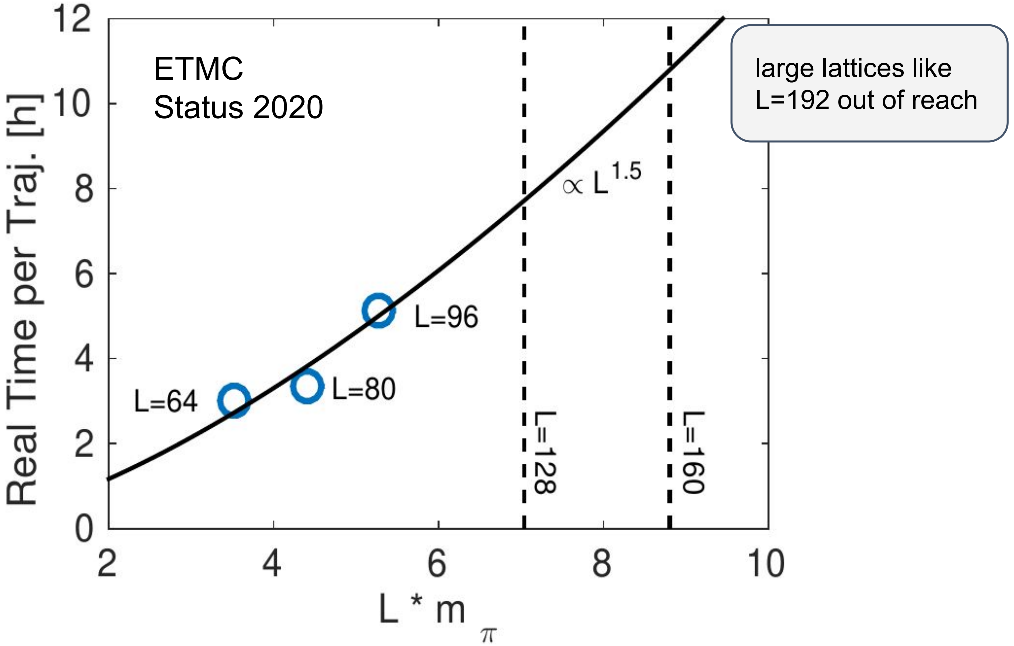
battle between computational costs and strong scalability

if strong scaling window does not scale as HMC costs

- ❖ resulting into increasing simulation time

Cost for HMC:

Solver	#steps	stat. error
$\propto V$	$V^{1/2n}$	N_{cnfg}



J.F. PRACE 6IP, Inter WP session 2020, LyNcs: Towards exascale ..

Overcome by:

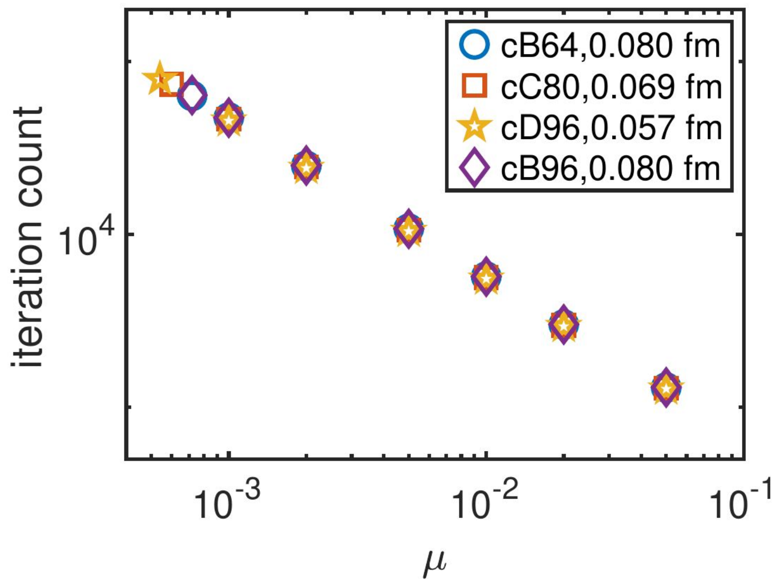
- ❖ increase scalability
- ❖ speed-up inversion time

Conjugate Gradient solver

basic Krylov subspace solver

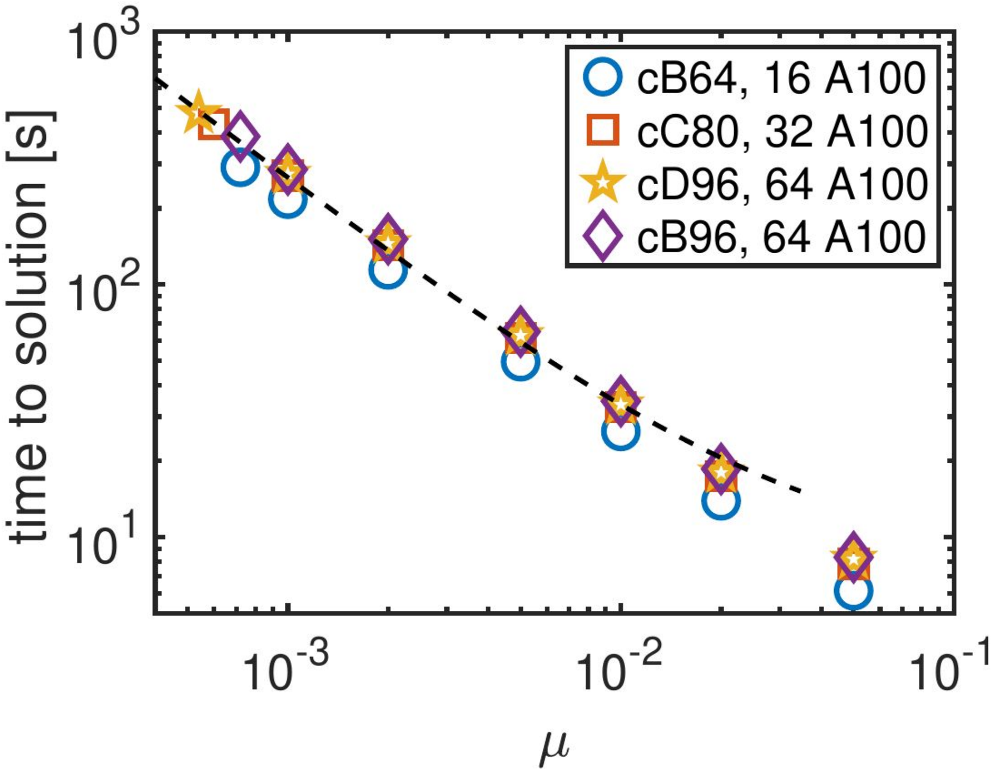
- used for large masses
- only depend on matrix-vector stencil
  - very good scalability
- sped up by using low-precision (40% single, 50% half)

K. Clark, Software, Thursday, 10:00



iteration proportional to condition number

- only depend on smallest mode (independent from density of modes)
- iteration count increases drastically at physical point



Cost for Conjugate gradient solver:

$$cost \approx V \cdot \left(\frac{b}{\mu} + a\right) \approx V \frac{b}{\mu}$$

with $b/a \sim 0.04$

Lowest level: Matrix vector product

$$D(U) \cdot x$$

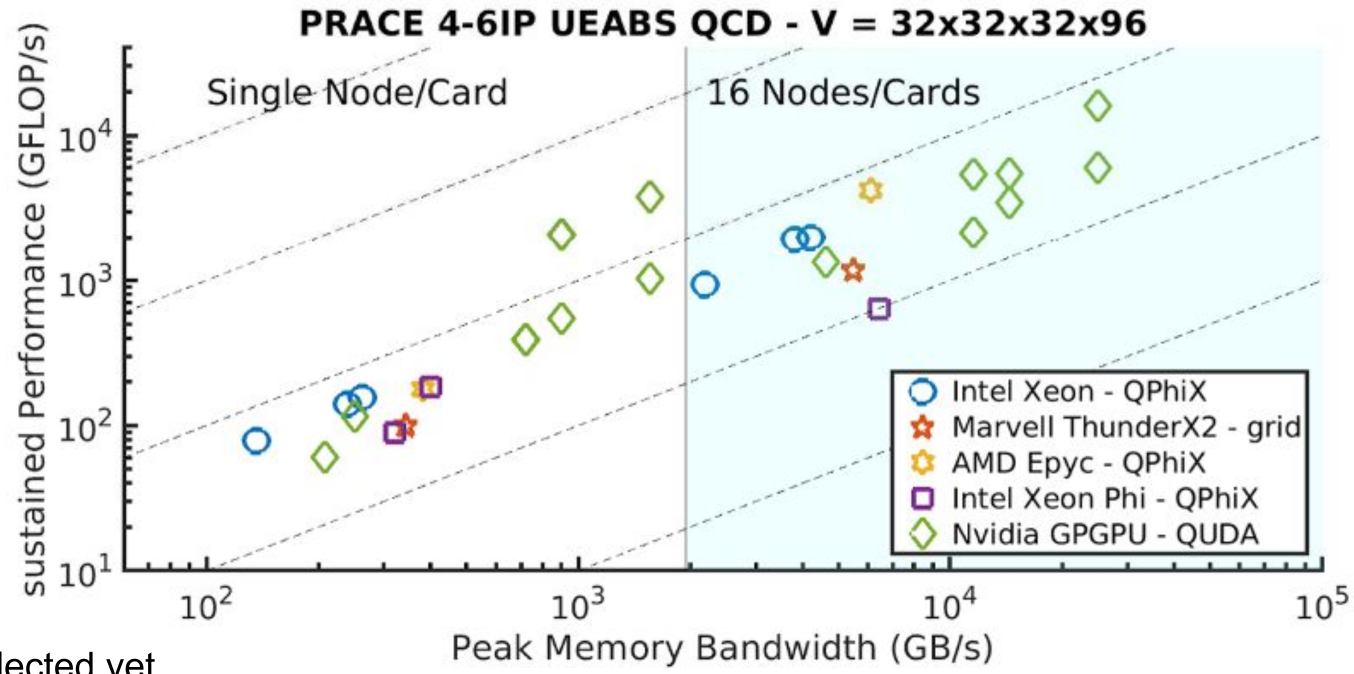
- ❖ arithmetic intensity ~ 1.0
- ❖ computational costs grows with **V**

on HPC-hardware:
- ❖ bandwidth bound
- ❖ latency bound

Benchmark kernels:
- ❖ QPhiX Intel Xeon (Phi)
- ❖ QUDA Nvidia
- ❖ grid Arm

performance results for novel HPC hardware not collected yet
- ❖ several lattice QCD packages offering optimized stencil, e.g. QUDA and grid
- ❖ scaling results from Fugaku I. Kanamori, Poster, Di. 20:00
- ❖ multiple RHS : decreasing arithmetic intensity, operators available within DDalphaAMG QUDA grid
N. Meyer, Poster, Di. 19:00

Performance on PRACE Tier 0 Machines
(Partnership for Advanced Computing in Europe)

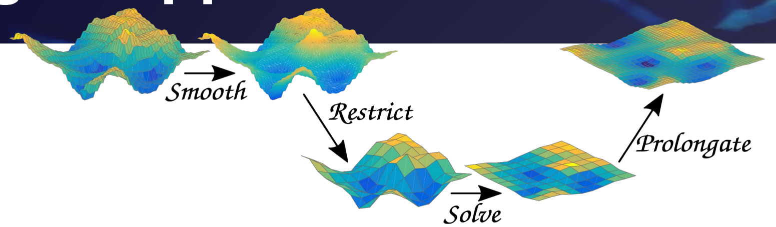


J.F. ,PRACE 4IP - 6IP, WP7 Task Benchmark

CG benchmarks on European HPC systems
- ❖ Tier 0 and Prototype systems with PRACE 4IP-6IP WP7

aSTоRC

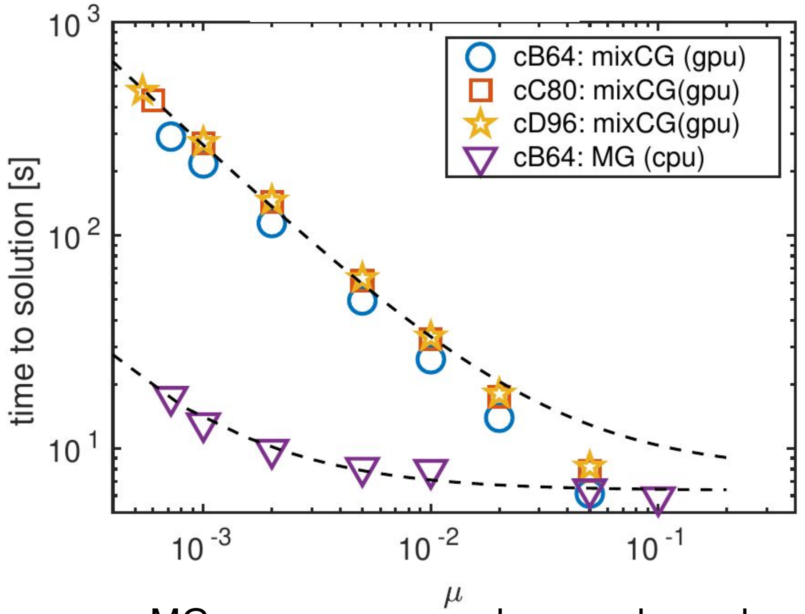
Multigrid approaches



using an very effective preconditioner

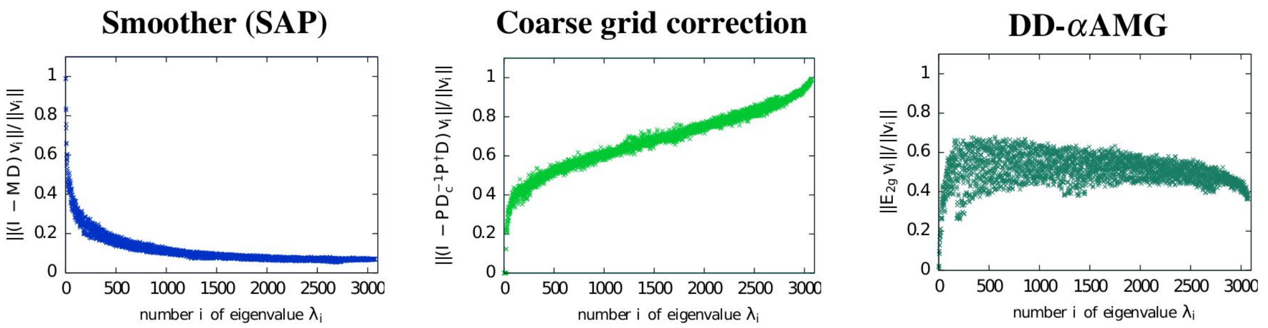
- ❖ tread UV modes by SAP smoother
- ❖ tread IR modes by an algebraic multigrid approach

$$D_c = RDP.$$



MG suppresses quark mass dependence

$$\propto V \cdot (0.001/\mu + 1)$$



(F)GMRES/GCR

- ❖ multi-grid preconditioning — M. Luscher, JHEP 12 (2007) 011, JHEP 07 (2007) 081
  - ➢ Wilson ~ 100x — R. Babich et al. Phys.Rev.Lett. 105 (2010) 201602 [QUDA] [DDalphaAMG]; A. Frommer, et. al., SIAM J.Sci.Comput. 36 (2014) [grid] [openQCD]
  - ➢ TM ~ 100x — C. Alexandrou, et al., PRD 94 (2016) 11, 114509 [QUDA] [DDalphaAMG]
  - ➢ staggered ~10x — Brower et al., Phys.Rev.D 97 (2018) 11, 114513 [QUDA]; V. Ayyar, Software, Thursday, 9:20
  - ➢ DW ~ 2-4 — Brower et al., Phys.Rev.D 102 (2020) 9, 094517 [QUDA] [grid]; P. Boyle et al., arXiv:2103.05034

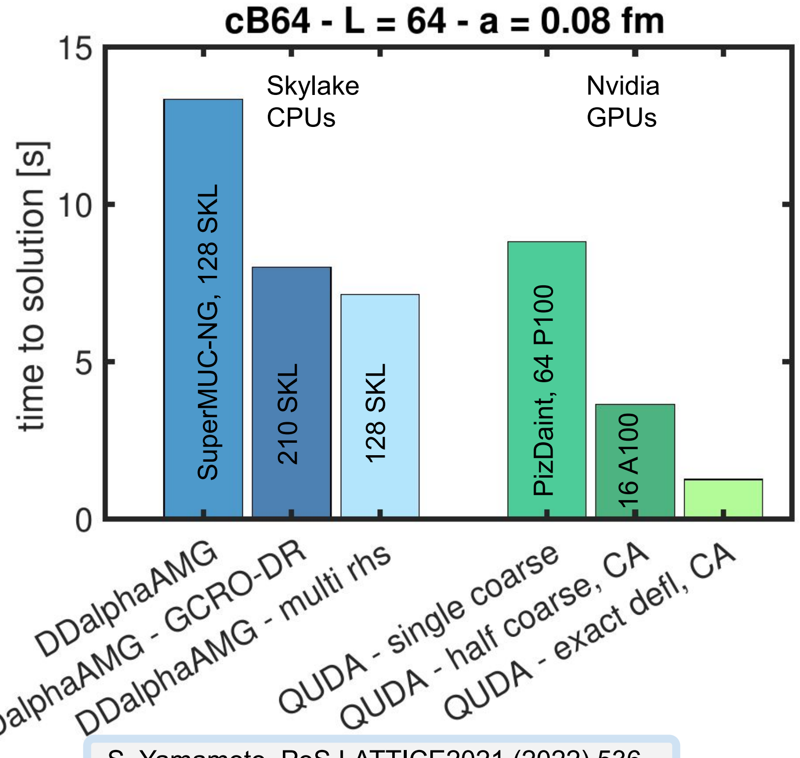
Multigrid solvers within the HMC

- ❖ limited scalability
  - ➢ strong scaling is limit to coarse grid size
  - ➢ lower limit bounded by memory
- ❖ additional overhead due to setup-update during MD

Multigrid improvements

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

Multigrid improvements in case of twisted mass



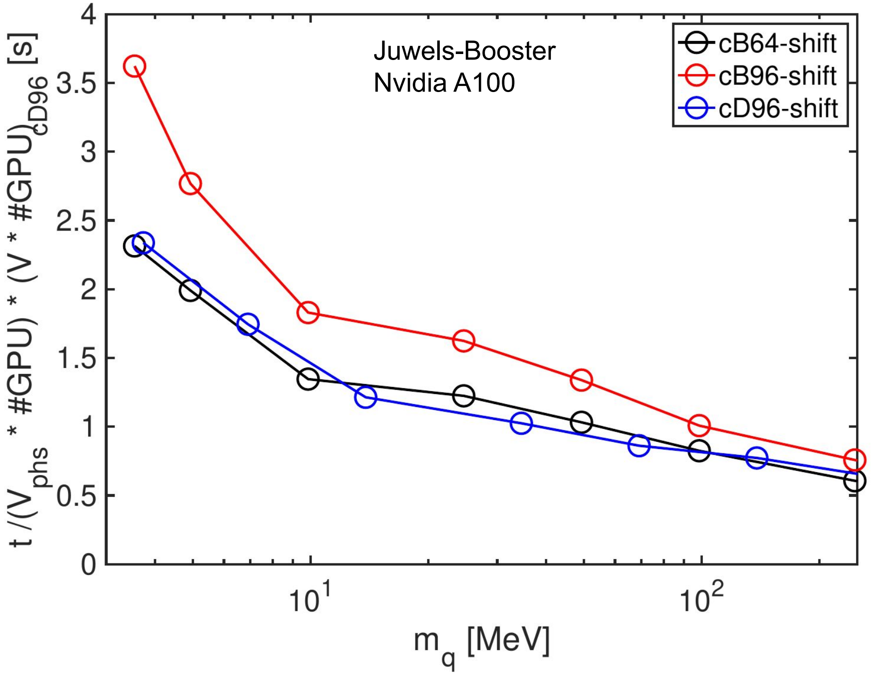
S. Yamamoto, PoS LATTICE2021 (2022) 536

J. Espinoza-Valverde et al., arXiv:2205.09104

coarse grid improvements
- ❖ communication avoiding
- ❖ coarse grid deflation

within the HMC ? GCRO-DR stable ?

Multigrid towards larger physical volumes



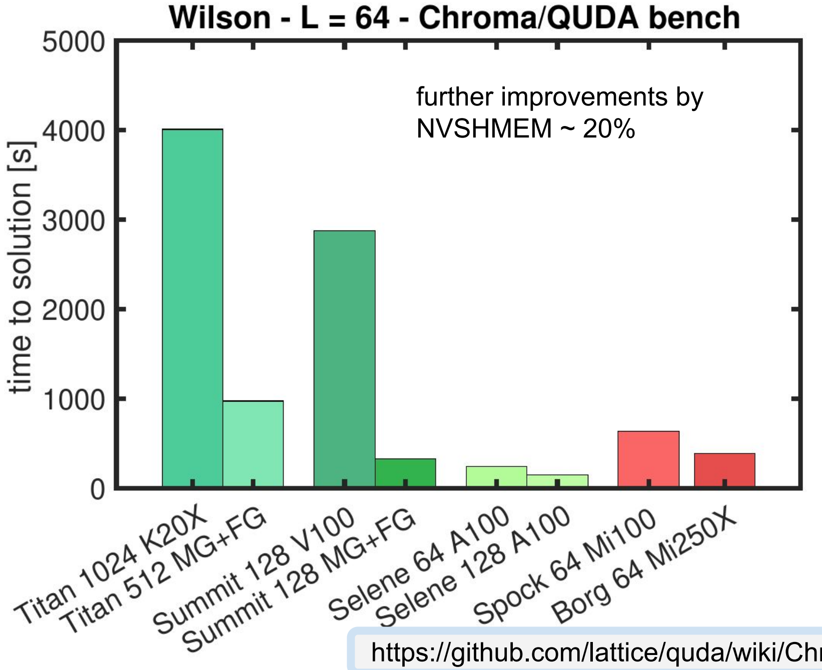
shows dependence on physical quark mass
- ❖ increases with physical volume

multiple rhs in grid on Juwels-Booster
- ❖ large improvements ~ 6x

N. Meyer, Poster, Di. 19:00

aSTORC

HMC on GPUs



https://github.com/lattice/quda/wiki/Chroma-ECP

M. Wagner, Software Session I , Monday,16:30

QUDA within HMC

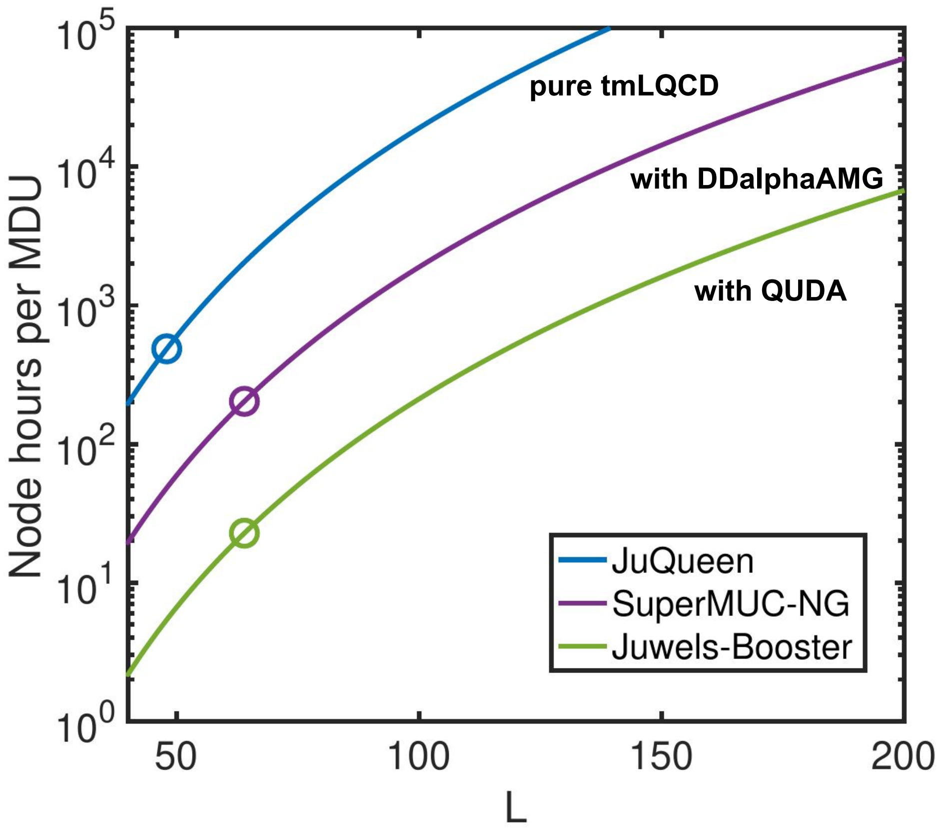
- ➢ Chroma | MILC | CPT
- ➢ tmLQCD — B. Kostrewza, Software Session I , Monday, 15:00
- ➢ lyncs-API — S. Bacchio, Software Session I , Monday, 14:40

S. Yamamoto, Software Session I , Thursday, 9

grid enables HMC on GPUs

- ❖ GPT — C. Lehner, Software Session I , Monday, 14:20
- ❖ with inexact deflation — F. Ziegler, Software Session I , Monday, 15:20

Cost for HMC: by 10x with MG, by 10x with GPUs



further minimization possible
using higher order integrators

Molecular dynamics - Overview

The Cyprus Institute
RESEARCH•TECHNOLOGY•INNOVATION

Proposal of a new set (U,P) via Hamilton's Equations:

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

HMC requires: symplectic, reversible integrators

- ❖ second order approaches: 2MN — Sexton et al., Nucl. Phys. B380, 665 (1992)
- ❖ collection of higher order schemes, now standard given by Omelyan, Mryglod and Folk — Omelyan et al., PRL. 86 (2001) 898

How to tune with various actions ?

- ❖ minimize costs at constant acceptance (dH normal distributed) — Creutz, Phys. Rev. D38 (1988) 1228–1238

$$P_{acc} = \mathrm{erfc}(\sqrt{\sigma^2/8}) \quad \text{with} \quad \sigma^2 = \mathrm{var}(\delta H)$$

- ❖ dependence of $\sigma^2(N_{steps}, \mu, L, \ldots)$ is necessary
  - ➢ calculate NLO terms of Shadow-Hamiltonian, given by Poisson Brackets
  - ➢ in case of second minimal norm scheme (2MN):

T. Takaishi et al., Phys.Rev.E 73 (2006) 036706

A. Kennedy et al., PRD. 87 (2013) 3.034511

$$\tilde{H} = T + S + \delta\tau^2 \left(\frac{6\lambda^2 - 6\lambda + 1}{12}\{S,\{S,T\}\} + \frac{1-6\lambda}{24}\{T,\{S,T\}\} \right) + \mathcal{O}(\delta\tau^4)$$

with $\lambda = 1/6$ leading term is given by force $\{S,\{S,T\}\} = tr(F^2)/a$

aSTRC

RHMC with Block Solvers and multiple Pseudofermions

NLO term can be also used to evaluate methods

Clark et al., Phys.Rev.Lett. 98 (2007) 051601

de Forcrand et. al, Phys.Rev.E 98 (2018) 4, 043306

RHMC with block solvers and multiple pseudofermions

- a variant of RHMC by splitting up

$$\det\left[M^\dagger M\right] = \det\left[\left(M^\dagger M\right)^{\frac{1}{n_{pf}}}\right]^{n_{pf}}$$

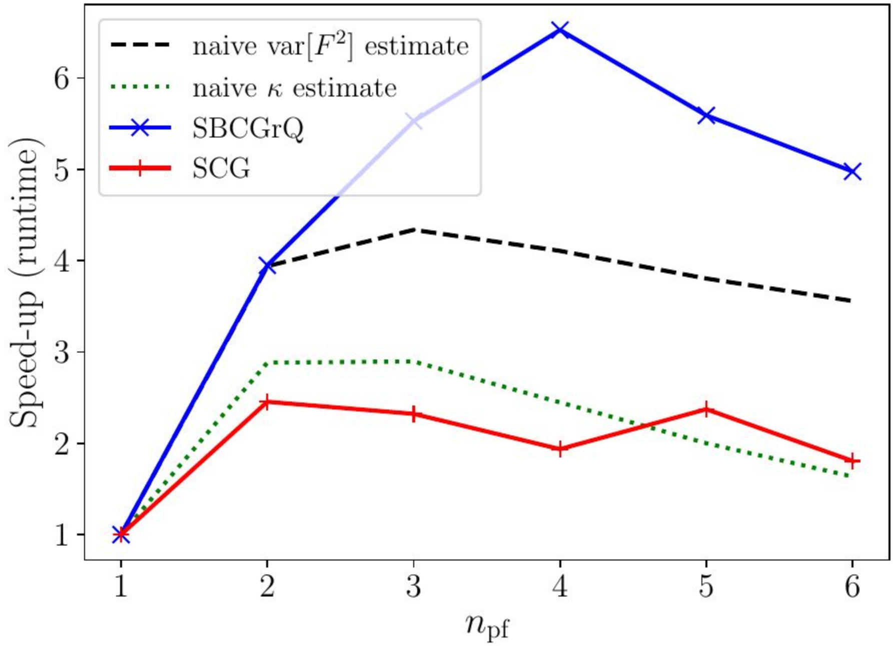
  - with the variance of the force

$$\mathrm{var}\left(F^2(n_{pf})\right) = c_2 n_{pf}^{-1} + c_3 n_{pf}^{-2} + \mathcal{O}(n_{pf}^{-3})$$

    which reduces the required steps at given acceptance

- ideally to combine with Block Krylov solvers
  - use SBCGrQ
  - faster convergence by increasing search space

lkeegan / blockCG



- in combination with Block solver speed ups of ~ 6 achieve
  - in case of Nf=4, L=8 and small mass

Hasenbusch mass preconditioning

Hasenbusch, Phys.Lett.B 519 (2001) 177-182

- terms are proportional to $\propto \left(\frac{\Delta^2 m}{\mu^2}\right)^k$
  - $\Delta m^2 = \mu_1^2 - \mu^2$ can be tuned,
    - while with RHMC terms are *static* due to Chebyshev approximation

aSTorC

Nested and higher order schemes

The Cyprus Institute
RESEARCH·TECHNOLOGY·INNOVATION

Nesting

$$\Delta(h) = e^{\frac{h}{6}\hat{B}_1} e^{\frac{h}{2}\hat{A}} e^{\frac{2h}{3}\hat{B}_1} e^{\frac{h}{2}\hat{A}} e^{\frac{h}{6}\hat{B}_1}$$

$\Delta(h/2, \hat{B}_0, \hat{A})$ $\Delta(h/2, \hat{B}_0, \hat{A})$

with $\Delta(h/2) = e^{\frac{h}{12}\hat{B}_0} e^{\frac{h}{4}\hat{A}} e^{\frac{h}{3}\hat{B}_0} e^{\frac{h}{4}\hat{A}} e^{\frac{h}{12}\hat{B}_1}$

J. Sexton et al., Nucl. Phys. B380, 665 (1992)

C. Urbach et al., Comput.Phys.Commun. 174 (2006) 87-98

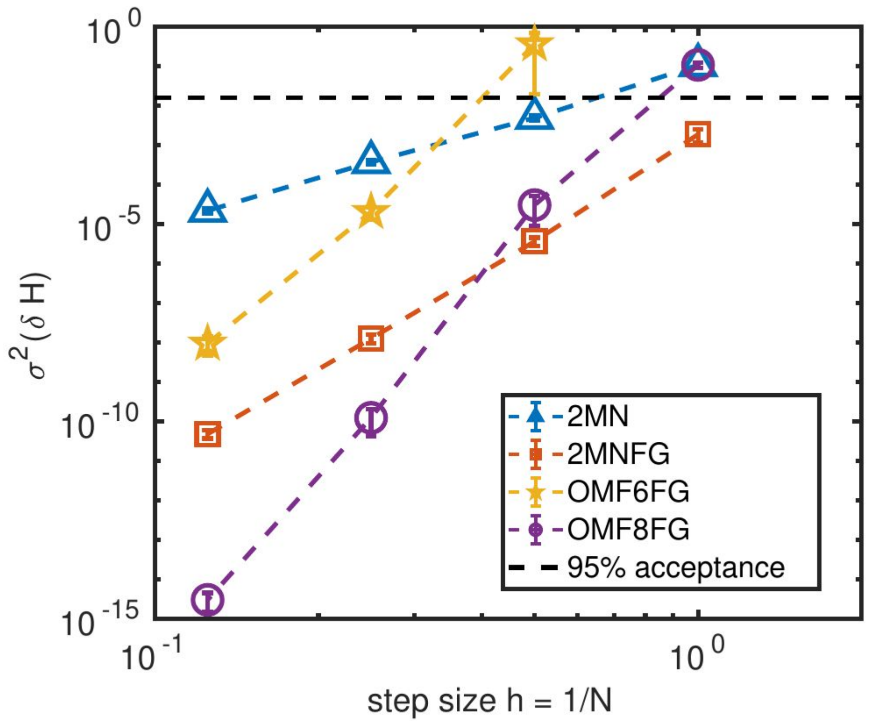
D. Shcherbakov et al., CCP. 21 (2017) 4, 1141-1153]

- Second order minimal norm scheme perfect for nesting
  - can be extended to fourth order scheme with force gradient term

with $$\Delta(h) = e^{h\frac{1}{6}\hat{B}} e^{\frac{1}{2}h\hat{A}} e^{\frac{2}{3}h\hat{B} - \frac{1}{72}h^3 C} e^{\frac{1}{2}h\hat{A}} e^{\frac{1}{6}h\hat{B}}$$

where $$C = 2\sum_{x=1,\nu=0}^{V,3} \frac{\partial S}{\delta U_\nu(x)} \frac{\partial^2 S}{\delta U_\nu(x)\delta U_\mu(x)}$$

- comes with second derivative and mixing actions



Trick by Lin and Mawhinney:

H. Lin et al., PoS LATTICE2011 (2011) 051

- approximate term numerically
  - requires additional memory for gaugefield
  - reduces inversions by 4/3
  - simple to implement, e.g. tmLQCD

aSTRC

HMC summary

Cost for HMC:

Solver	#steps	stat. error
$\propto V$	$V^{1/2n}$	N_{cnfg}

$$\propto \frac{(a^4 V)^{(1+1/2n)}}{a^{4(1+1/2n)}} \cdot a^{-5} \propto (a^4 V)^{(1+1/2n)} \cdot a^{-10}$$

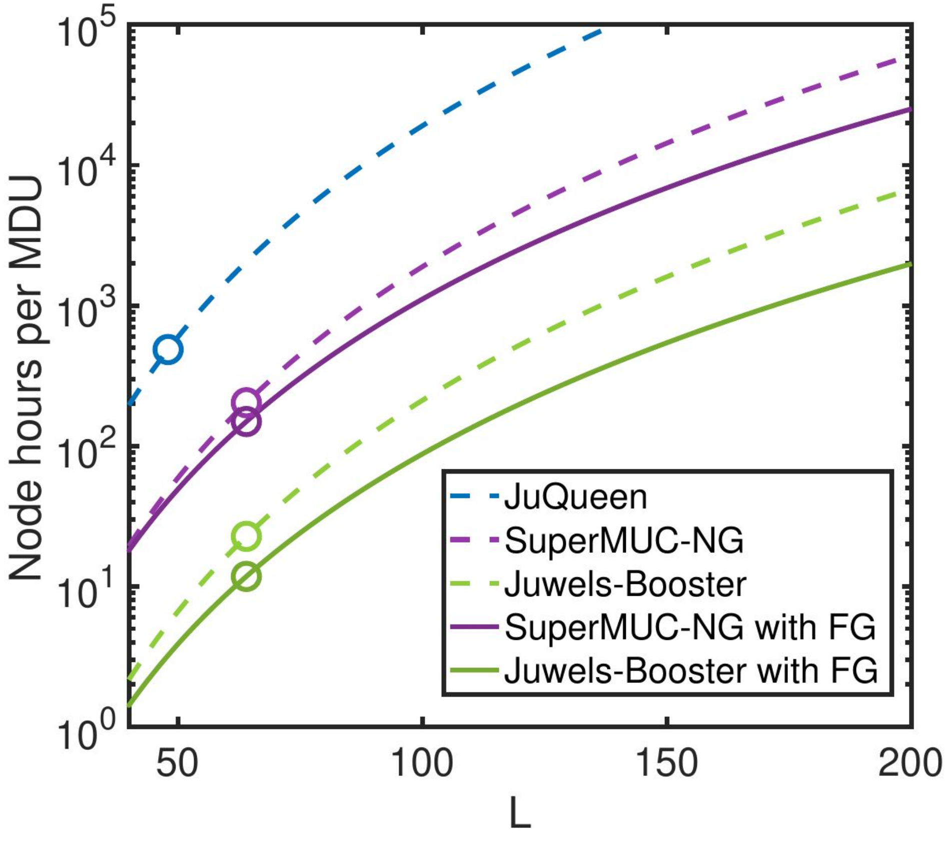
suppressed quark mass dependence

- ❖ usage of multigrid
- ❖ Hasenbusch mass-preconditioning

reducing volume scaling

- ❖ with higher order integrators

- ❖ how to reduce scaling with lattice spacing ?



for L=192:

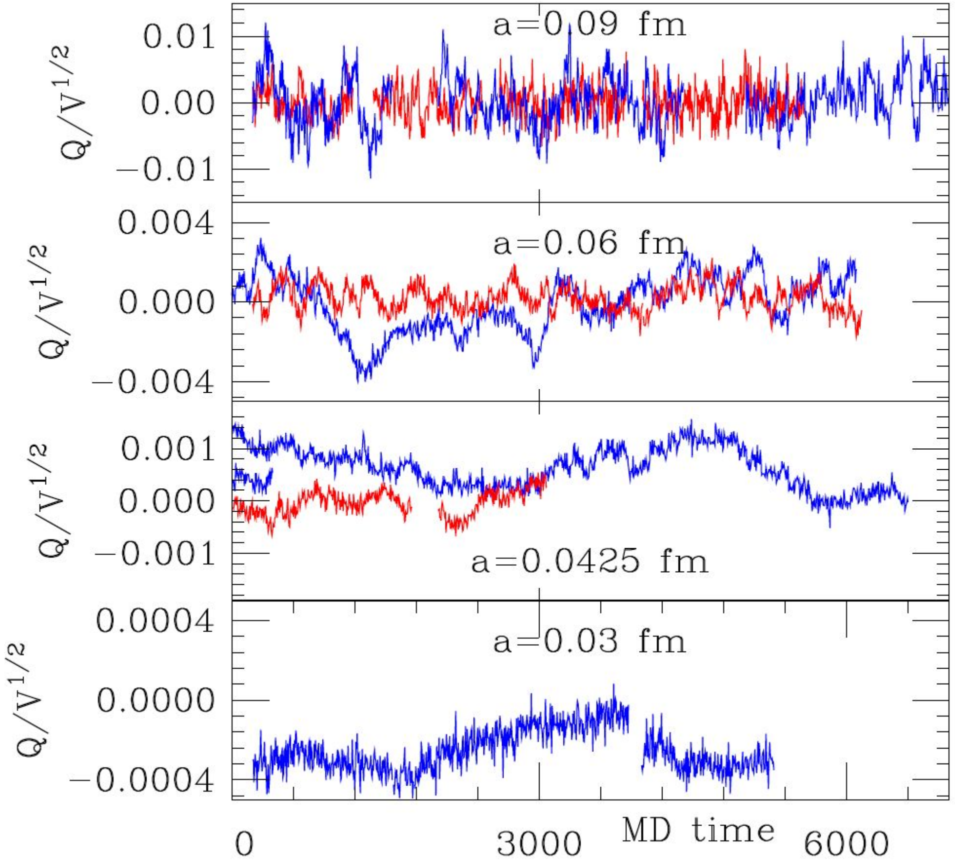
- ❖ ~ 1000 Node hours per MDU (4x A100)
  - ➢ in reach with exascale computing

THE CYPRUS INSTITUTE
RESEARCH•TECHNOLOGY•INNOVATION

1. HTML

1. HMC
   - ❖ State of the art: Solvers
     - ➢ Conjugate gradient
     - ➢ Multigrid preconditioners
       - ■ calculation on GPUs
   - ❖ Molecular dynamics
     - ➢ Symplectic integrators
       - ■ Higher order integrators

1. **MCMC methods**
   - ❖ Methods without topological freezing
     - ➢ Modifications of molecular dynamics
     - ➢ Global corrections
       - ■ Multi-level
       - ■ Instanton update
     - ➢ Machine learning / Generative models
       - ■ normalizing flows



C. Bernard, et al., Phys.Rev.D 97(2018) 7. 074502

aSToRC

Overview - Algorithms

General MCMC structure:

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

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

Distributions $(\tilde{\rho}(U)\rho(U'))/(\rho(U)\tilde{\rho}(U'))$ log-normal distributed

❖ for the acceptance rate follows

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

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

General approaches:

1. Change Update procedure
2. Modify MD integration
3. Change Hamiltonian/conditions

Smart modifications needed:
otherwise $P_{acc} \to e^{-V}$

❖ Change Hamiltonian/conditions
- ➢ open boundary conditions in time — Luscher et al., JHEP 07 (2011) 036
- ➢ topological freezed simulations — Czaban et al., Lat13, arXiv:1310.5258; Brower et al., Phys.Lett.B 560 (2003) 64-74
- ➢ Masterfield — Luescher, EPJ Web Conf. 175 (2018) 01002; Albandea et al., Eur.Phys.J.C 81 (2021) 10, 873
- ➢ Stochastic molecular dynamics — Patrick Fritzsch, Plenary talk , Saturday 9:20
- ➢ Multiscale equilibration/re-thermalization — Detmold et al., Phys.Rev.D 94 (2016) 11, 114502; Detmold et al., Phys.Rev.D 97 (2018) 7, 074507; Tu et al., EPJ Web of Conferences 175, 02006 (2018)

Modifications of MDs

Eliminating volume fluctuations via dynamics at fixed Hamiltonian
basically HMC, MD modifications

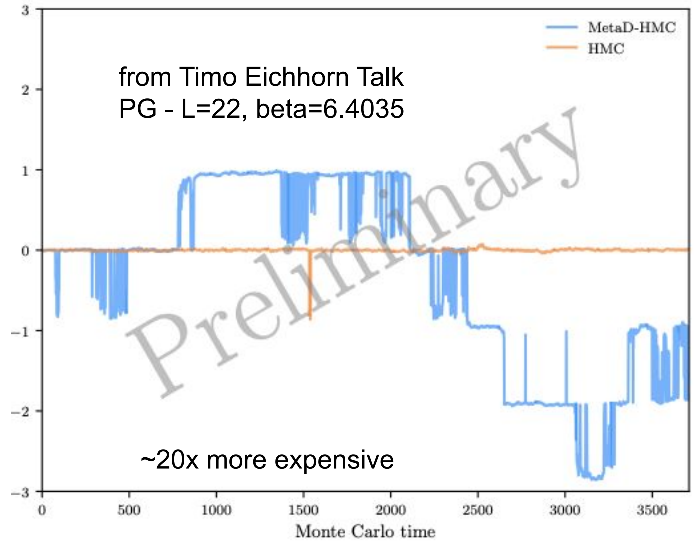
- Riemannian-manifolds/Fourier acceleration — T. Nguyen et al., LATTICE2021 (2022) 582
  - modify momentas to de/accelerate high/slow modes
- skewed detailed balance — J. Pinto Barros, Algorithms, Thursday, 10:40

- trivializing maps
  - integrate to a trivialized point, which mixes up frequencies
    - use Schwinger-Dyson equation to construct approximative map — N. Matsumoto, Plenary talk , Monday, 9:50
    - combination with normalizing flows — D. Albandea, Algorithms, Monday, 18:10; S. Foreman et al., PoS LATTICE2021 (2022) 073

- Metadynamics
  - add marginal terms
    - Pauli-Villars fields — A. Hasenfratz et al., Phys.Rev.D 104 (2021) 7, 074509
    - mild down large cut-off effects

$$\det(M_{PV}^\dagger M_{PV})^{-1} = \int d\phi \, \exp\{-\phi^\dagger M_{PV}^\dagger M_{PV} \phi\}$$

    - topology tunneling terms — T. Eichhorn et al., PoS LATTICE2021 (2022) 573; T. Eichhorn, Algorithms, Monday, 17:30
    - LeapFrogLayers — S. Foreman et al., PoS LATTICE2021 (2022) 508



Global corrections

Global Corrections via accept-reject step

- Exact algorithm with Metropolis-Hasting step

$$P_{acc}(U \to U') = \min\left[1, \frac{\tilde{\rho}(U)\rho(U')}{\rho(U)\tilde{\rho}(U')}\right]$$

- with proposal probability $\tilde{\rho}(U')$

Challenge:

1. **Find appropriated proposal procedure**
2. **Reduce the variance of**

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

- restrict dimension of the distributions $\rho(U')$
  - factorization of Boltzmann-factor via recursive Schur decomposition:
    $$\det D = \det S_{red} \cdot \det S_{pink} \cdot \det D_{blue}$$
  - complete time factorization of determinant
- use correlations between actions
  - via linear parameter, shift in $\delta\beta$
  - via machine learning

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

J. Finkenrath 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

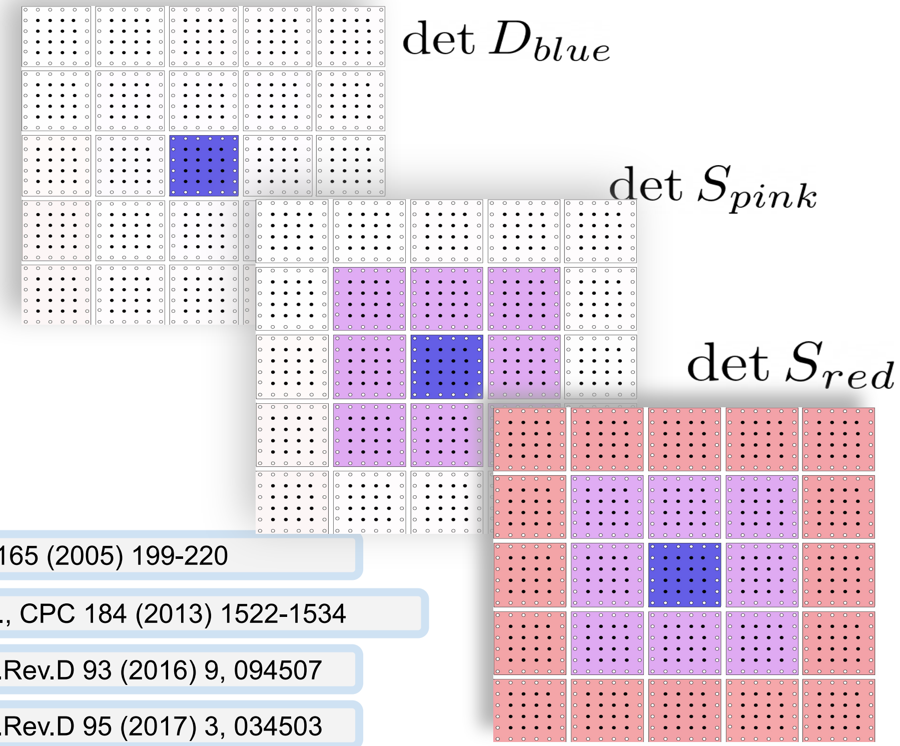
U. Wenger, Algorithms, Thursday, 9:40

A. Irving et al., Phys.Rev.D 55 (1997) 5456-5473

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

M. Albergo et al., Phys.Rev.D 100 (2019) 3, 034515

Recursive Domain Decomposition

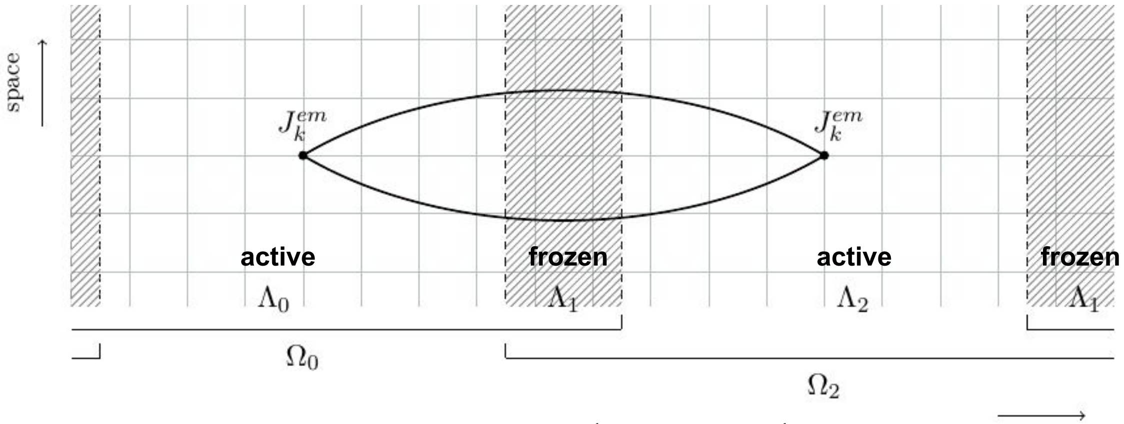


Multilevel - algorithms

Idea: factorize fermionic function (determinante and correlator)

$$\det Q = \frac{\det\,(1-w)}{\det Q_{\Lambda_1}\,\det Q_{\Omega_0}^{-1}\,\det Q_{\Omega_2}^{-1}}$$

- select active domains with separation Λ_1



global term $w = P_{\partial\Lambda_0} Q_{\Omega_0}^{-1} Q_{\Lambda_{1,2}} Q_{\Omega_2}^{-1} Q_{\Lambda_{1,0}}$

Algorithm:

- update via HMC each domain independent from each other n_1 times
- update global lattice and repeat n_0 times
  - (shifting possible but requires correction step)
- introduce global corrections via reweighting factors

gives in localist models a statistical error

$$\propto 1/(n_1\sqrt{n_0})$$

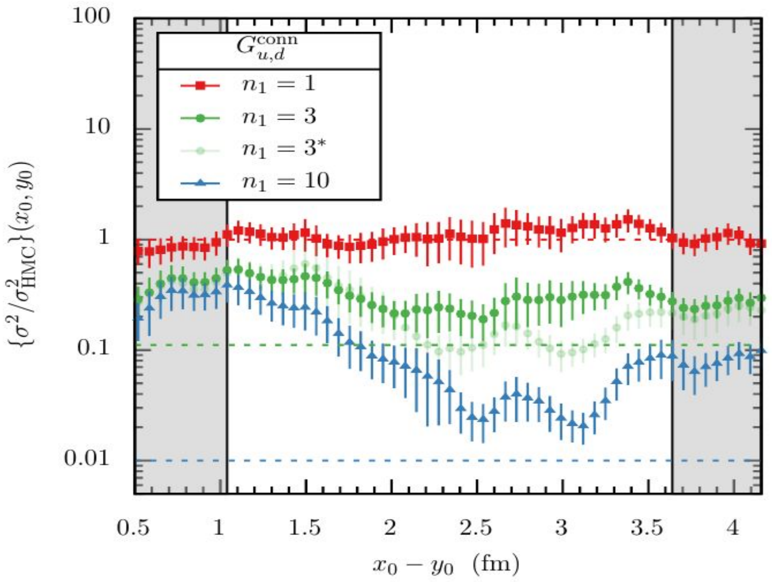
M. Cè et al., Phys.Rev.D 93 (2016) 9, 094507

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

Demonstrated in case of HVP of magnetic moment of the muon:

- L= 48, a= 0.065 fm, $\Lambda_1 = 8$, $\Lambda_{0/2} = 40$ and 300 MeV Pions
- error grows proportional to $\frac{\sigma^2_{G^{\rm conn}_{\rm u,d}}(x_0)}{[G^{\rm conn}_{\rm u,d}(x_0)]^2} \propto \frac{1}{n_0}\, e^{2\,(M_\rho - M_\pi)|x_0|}$,

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



- reduces the total statistical error in the full HVP to 1% accuracy with $n_0 = 25$ seperated with 48 MDUs and $n_1 = 10$

Idea: Minimize changed variables

D. Leinweber et al., Phys.Lett.B 585 (2004) 187-191

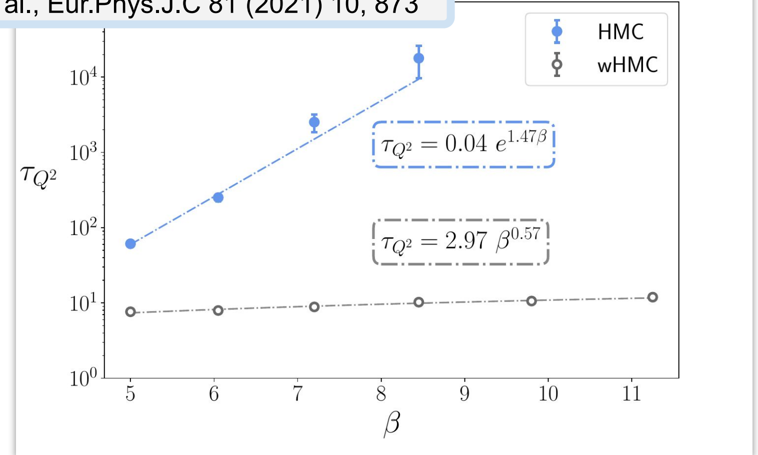
D. Albandea et al., Eur.Phys.J.C 81 (2021) 10, 873

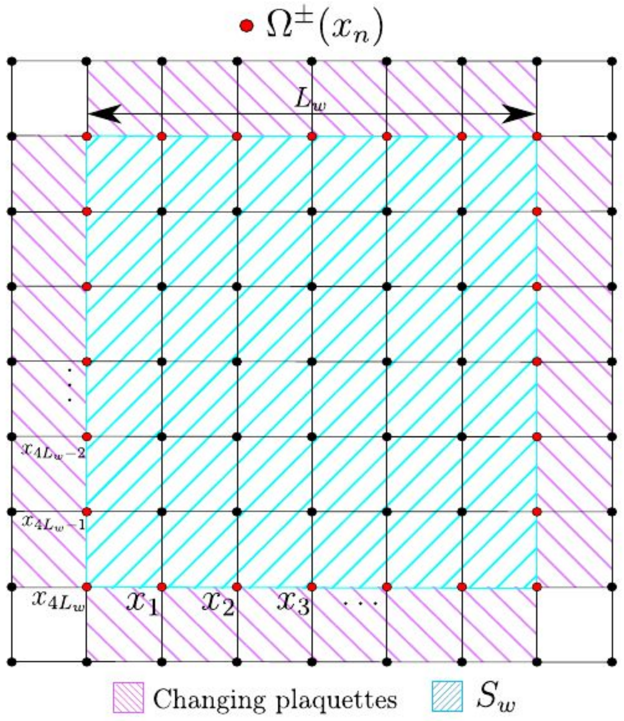
topological transitions by winding the field

$$U_\mu(x) \to U_\mu^\Omega(x) \equiv \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu})$$
$$\Omega^\pm(x_n) = e^{\pm i \frac{\pi}{2}\left(\frac{n}{L_w}+r\right)}$$

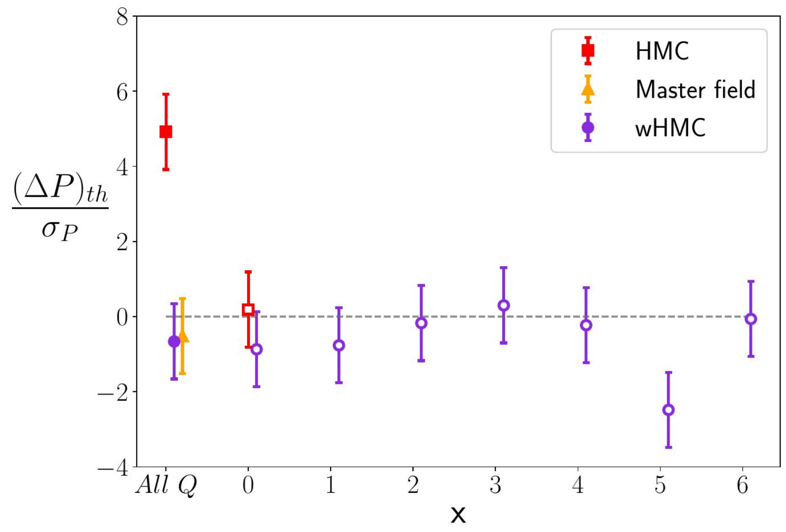
- accept reject step with
$$P_{acc}(U \to U') = \min\left[1, e^{-S[U']+S[U]}\right]$$
- **wHMC**: ergodic in combination with HMC:

→ HMC step → winding step → HMC step →





2D Schwinger Model
at beta = 11.25
- wHMC
- HMC at fix Q
- Masterfield L=8192



Breakdown of method (rough transformation)
- acceptance rate breaks down
  - ➢ with fermions and towards fine lattice spacings
  - ➢ trails in 4D SU(3) unsuccessful

T. Eichhorn, Algorithms, Monday, 17:30

Generative models for gauge theories

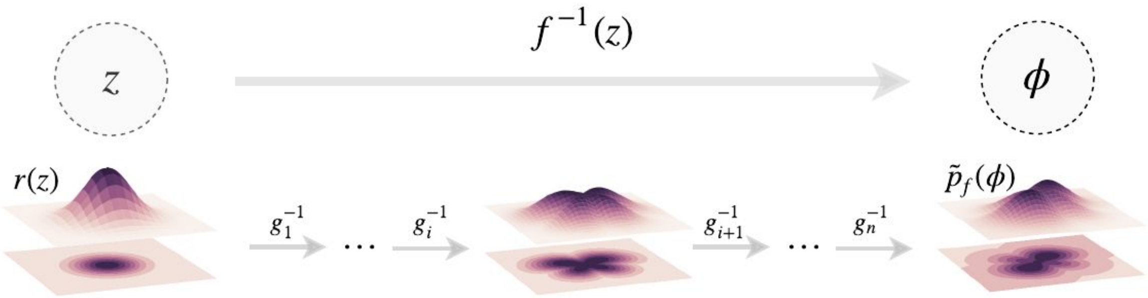
The Cyprus Institute
RESEARCH•TECHNOLOGY•INNOVATION

Idea: Train the correlations

Generative model in gauge theories with gauge invariant flow

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

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



- ❖ successfully applied to 2D discrete lattice models
- ❖ overcome critical slowing down by sampling from random distributions

- ❖ 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

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

M. Albergo et al., Phys.Rev.D 100 (2019) 3, 034515

G. Kanwar et al., Phys.Rev.Lett.125 (2020) 12, 121601

D. Boyda et al., Phys.Rev.D 103 (2021) 7, 074504

M. Albergo et al., arXiv:2101.08176 | Jupyter-notebook

P. Shanahan, Algorithms, Monday, 16:30

J. Marsh Rossney, Algorithms, Monday, 17:50

aSTRC

Generative models - normalizing flows

Why this approach is exciting ?

- new way to model physics distribution
  - huge potential to give new insides to QCD, sign problem
- exact: used as proposal within MCMC algorithm with:

$$P_{acc}(U \to U') = \min\left[1, \frac{q(U)p(U')}{p(U)q(U')}\right]$$

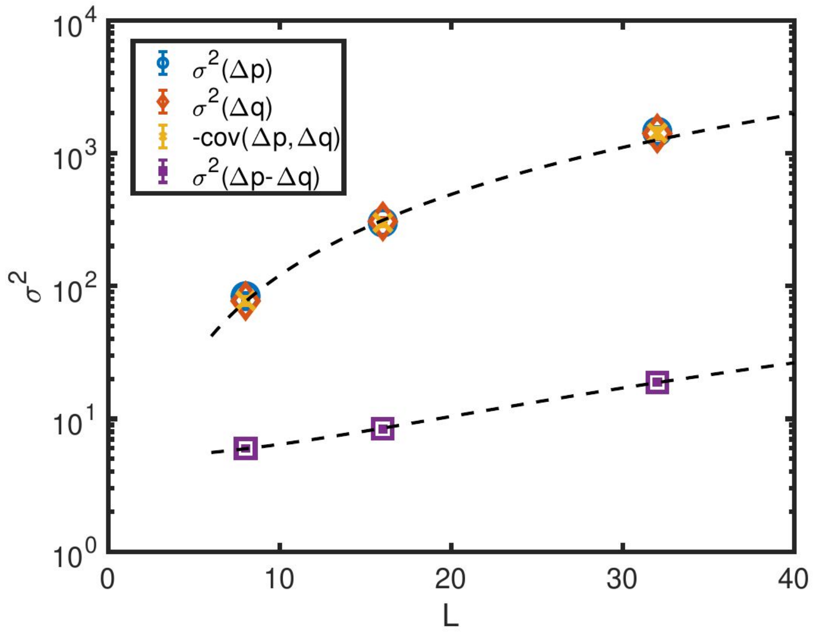
  - minimization of loss-function minimizes volume fluctuation
  - If log-normal distributed:

$$P_{acc} \approx 1 - \frac{\sigma}{\pi} \quad \text{for} \quad \sigma \ll 1$$

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

Fine tuning problem

- $\mathrm{var}(\Delta p) + \mathrm{var}(\Delta q) \approx -2 \cdot \mathrm{cov}(\Delta p, \Delta q)$
- Note that KL-divergence is minimizing distance between q and p



training at beta = 6.0,
two hidden layers with dimension L

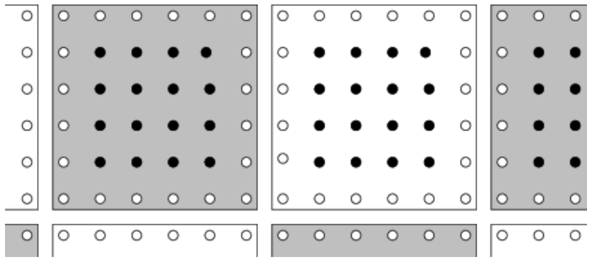
Challenge: How to scale ?
under active research

- Optimizing maps
  - modify neural networks
  - different flows, modification to normalizing flows, continues flows
- Use physical properties of the system
  - make use of the location of the action

Localization and factorization of action

Lattice action is local

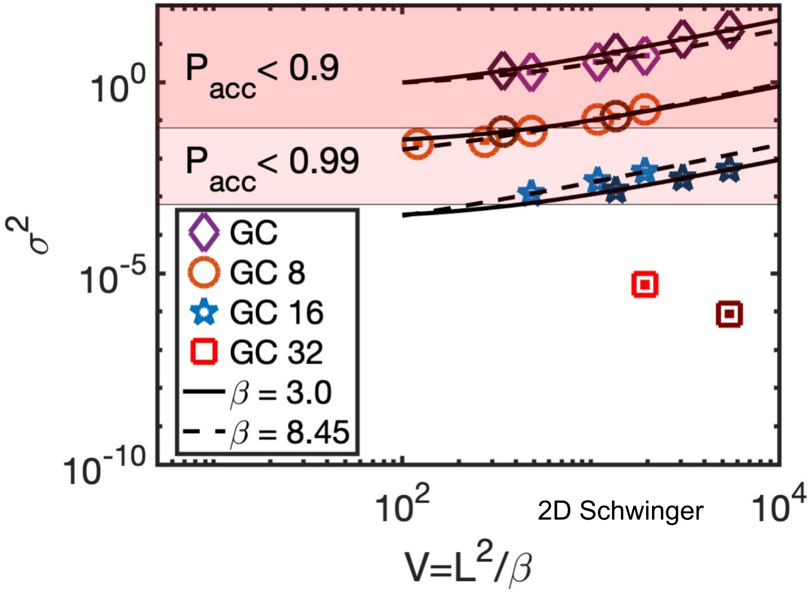
- without localization no proper continuum limit
- Fermions/quarks correlation decays with the lowest mode



Idea: use normalizing flows for local domain updates

- only update links within domain
- train the normalizing flows local
- decouple domains by freeze boundary terms

J. Finkenrath, arXiv:2201.02216



Combine this with accept-reject steps of the determinant

UV fluctuations can be filtered out by

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

- factorization and nested hierarchical filter (with correlations and cost ordered) by

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

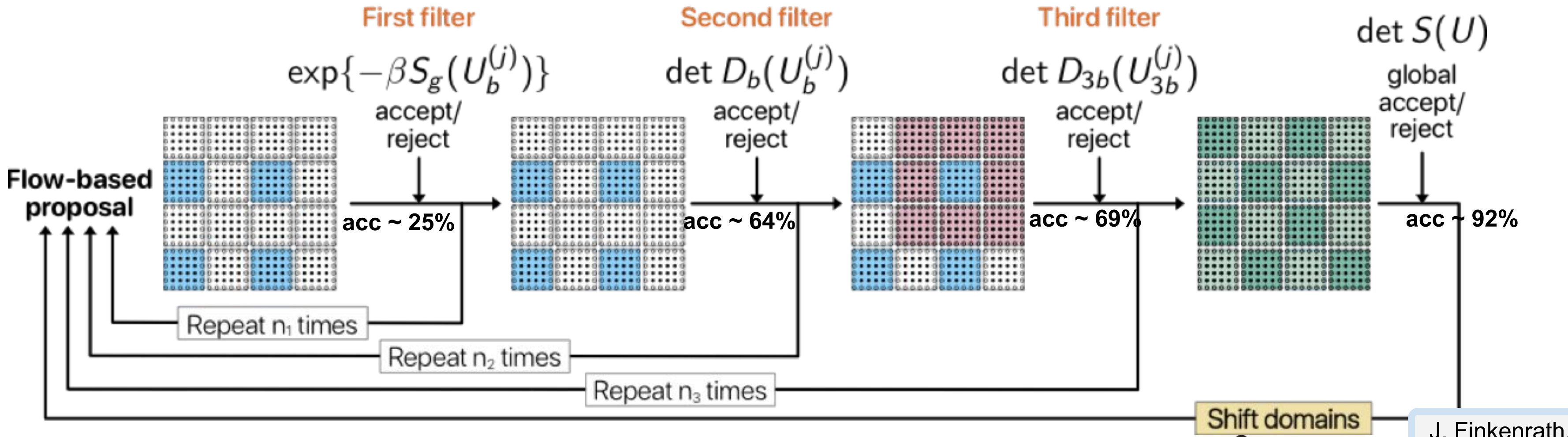
  - with the *i*th step

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

Long-range fluctuation highly suppressed

- in 2D-Schwinger model
  - acceptance rate ~ 100% with d=32

Domain decomposed normalizing flows



J. Finkenrath, arXiv:2201.02216

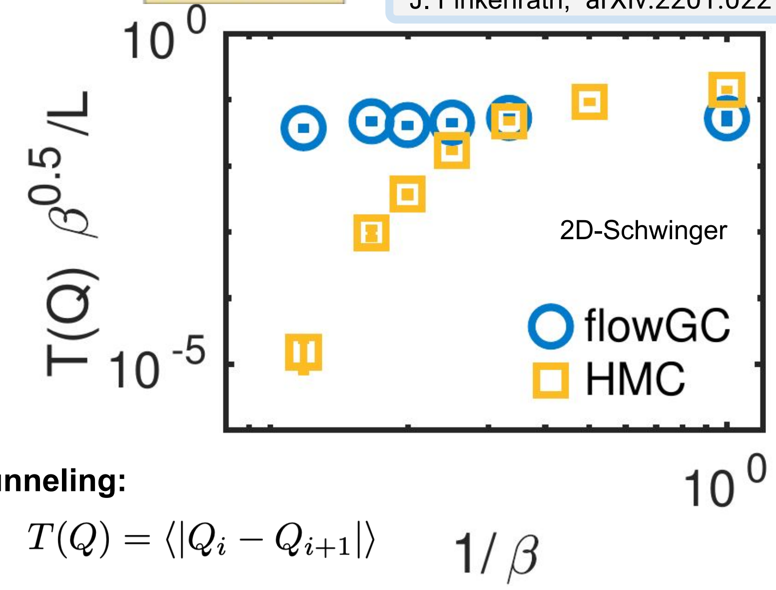
2D-Schwinger model:

- ❖ normalizing flow updates decorrelate topological charge
- ❖ update 16% of the links (combination with HMC possible)
- ❖ scaling at $m_{PS}\sqrt{\beta} \sim 0.4$ and $L/\sqrt{\beta} \sim 40$

for larger systems

- ❖ second filter potentially develop low acceptance rate
  - ➢ flow updates only take into account pure gauge weight

need to include fermions into the flow



Tunneling:

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

Flows with fermions

Fermion contributions can be included by sampling pseudofermions

- Idea: factorize distribution into

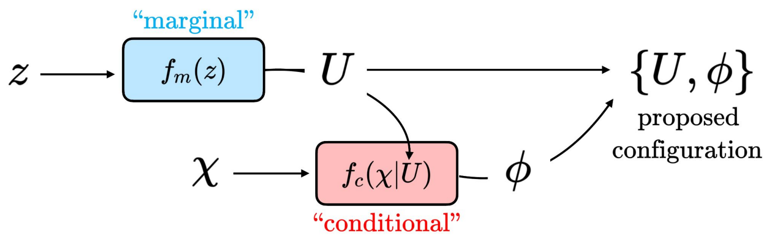
$$p(U,\phi) = p(U)\, p(\phi|U)$$

with *marginal* distribution $p(U) \propto \det DD^\dagger[U]\, e^{S_g(U)}$

and *conditional* distribution $p(\phi|U) \propto \dfrac{\det DD^\dagger[U]}{e^{-S_{pf}(U,\phi,\phi^\dagger)}}$

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

F. Romero-Lopez, Algorithms, Monday, 15:00



Split training accordingly with $q(U,\phi) = q(U)q(\phi|U)$

1. train the *marginal* to generate $\{U\}$
2. train on constant $\{U\}$ the pseudofermion map $f_c(\chi|U)$

Pseudofermions requires a new design of maps

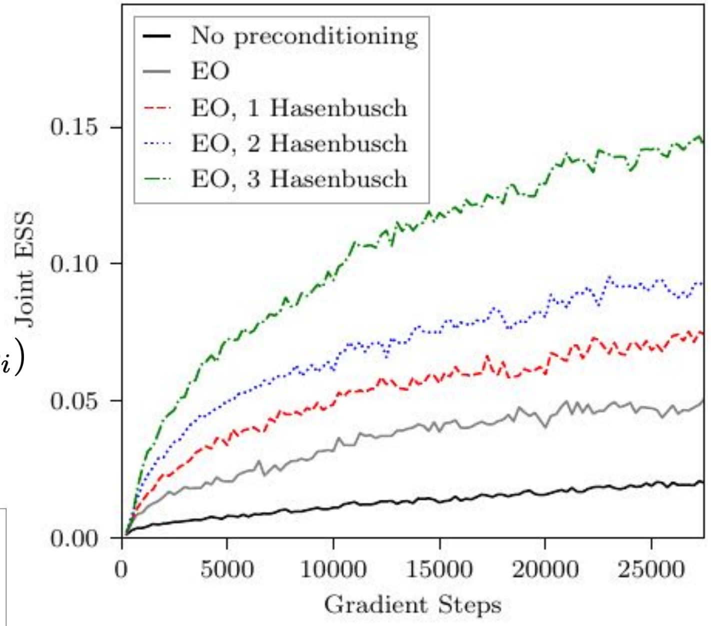
- using parallel transporter to approximate $D(U)$

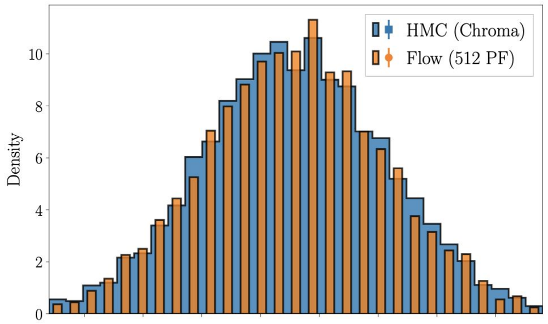
Effective Sample Size:

$$ESS = \frac{\frac{1}{N}(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2}$$

$$w_i = p(\phi_i)/q(\phi_i)$$

Combined with Hasenbusch-preconditioning





P. Shanahan, Algorithms, Monday, 16:30

R. Abbott et al., arXiv:2207.08945

First results on 4D-SU(3)+Nf(2) with normalizing flows:

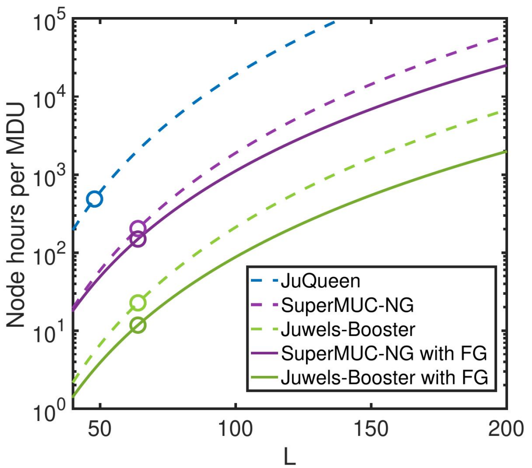
- using L=4, $\kappa = 0.1$ and $\beta = 1.0$

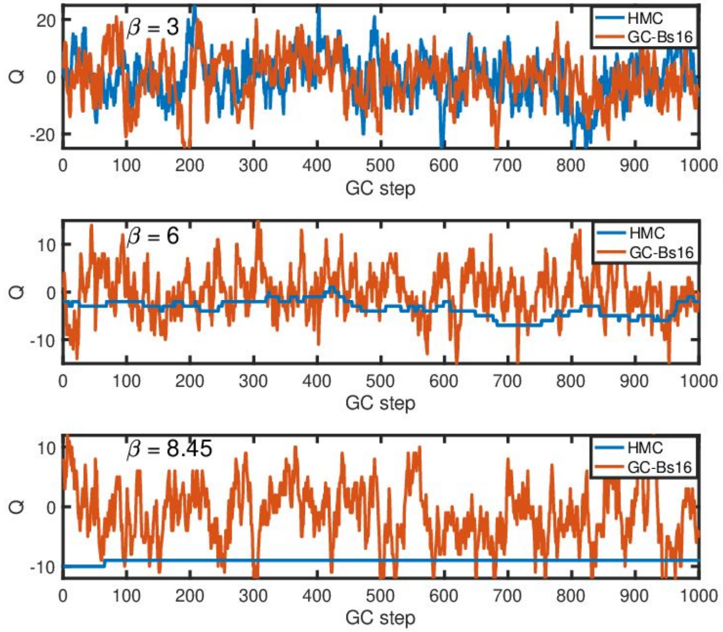
aSTcRC

Outlook

Status on HMC

- ❖ Software available to utilize pre-exa and exascale machine
  - ➢ scalability results still missing on novel architecture
  - ➢ L > 8 fm is feasible, next steps to further control finite volume effects in reach





Status on MCMC methods

- ❖ efficient algorithm to unfreeze topology are not available, expect for opening the boundaries
  - ➢ behind computational capability
    - ■ L = 128 for a = 0.04 fm would be in computational reach without freezing
    - ■ possible solution: combination of tunneling steps in combination with HMC updates
  - ➢ Need for flexible software for algorithms developments, to utilizing modular HPC hardware potential
    - ■ python APIs utilizing high performance packages QUDA and grid
  - ➢ community is very active, many ideas are under research
    - ■ I am looking forward to new ideas.

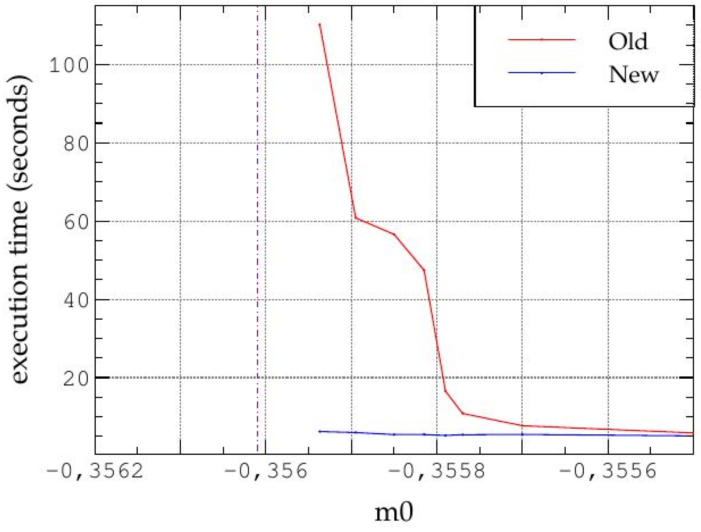
Thank you very much !!!

Appendix

Multigrid improvements: DDalpaAMG

DDalphaAMG J. Espinoza-Valverde et al., arXiv:2205.09104

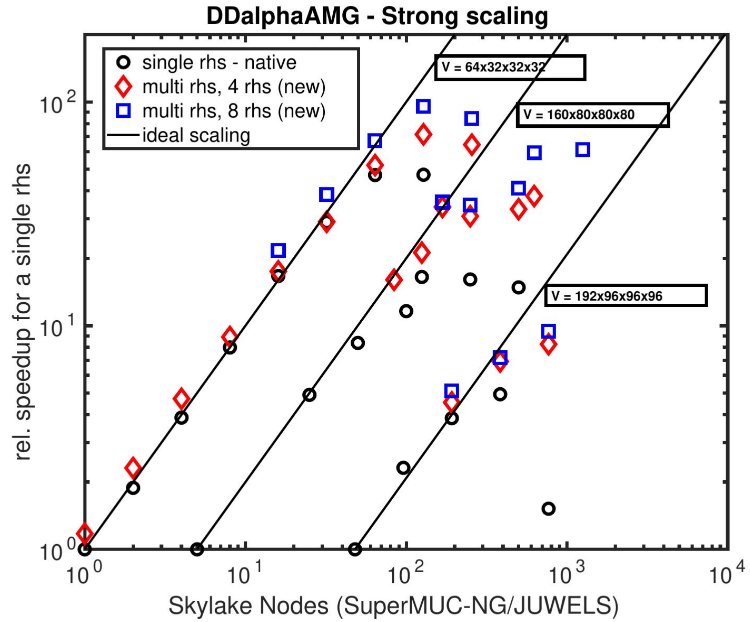
- pushing the strong scaling limit
  - pipeline versions
  - polynomial preconditioning
  - coarse grid Block-Jacobi
  - GCR-DRO
- improving mass scaling



- easily adapted within HMC (needs tests)

DDalphaAMG-multi rhs S. Yamamoto, PoS LATTICE2021 (2022) 536

- pushing the strong scaling limit
- testbed for Block-Krylov solver using fabulous



- requires high level redesign of HMC
  - need flexible software like lyncs-API

lyncs-API

S. Bacchio, Software, Monday, 14:40

S. Yamamoto, Software, Thursday, 9:40

  - combination with multiple pseudofermion variants

de Forcrand et al, Phys.Rev.E 98 (2018) 4, 043306

aSToRC