Status of the ETMC ensemble generation effort

C. Alexandrou[(a)(b)], **S. Bacchio**[(b)], **J. Finkenrath**[(c)], **R. Frezzotti**[(d)] **M. Garofalo**[(e)], **B. Kostrzewa**[(f)], **G. Koutsou**[(b)], **S. Romiti**[(g)], **A. Sen**[(e)], **C. Urbach**[(e)], **U. Wenger**[(g)]

[(a)] Department of Physics, University of Cyprus, 20537 Nicosia, Cyprus
[(b)] Computation-based Science and Technology Research Center, The Cyprus Institute, 2121 Nicosia, Cyprus
[(c)] Theoretical Physics Department, CERN 1211 Geneva 23, Switzerland
[(d)] Dipartimento di Fisica and INFN, Università di Roma "Tor Vergata", I-00133 Roma, Italy
[(e)] Helmholtz-Institut für Strahlen- und Kernphysik (Theorie), University of Bonn, 53115 Bonn, Germany
[(f)] High Performance Computing and Analytics Lab, University of Bonn, 53115 Bonn, Germany
[(g)] Institute for Theoretical Physics, A. Einstein Center for Fundamental Physics, University of Bern, CH-3012 Bern, Switzerland

Introduction

The $N_f = 2+1+1$ path integral for twisted mass Wilson clover fermions [4, 5, 8] is

$$Z = \int \mathcal{D}U \mathcal{D}\chi \mathcal{D}\bar{\chi}\, e^{-S_{\text{gauge}} - \bar{\chi} D_\ell \chi - \bar{\chi} D_h \chi}\,,$$

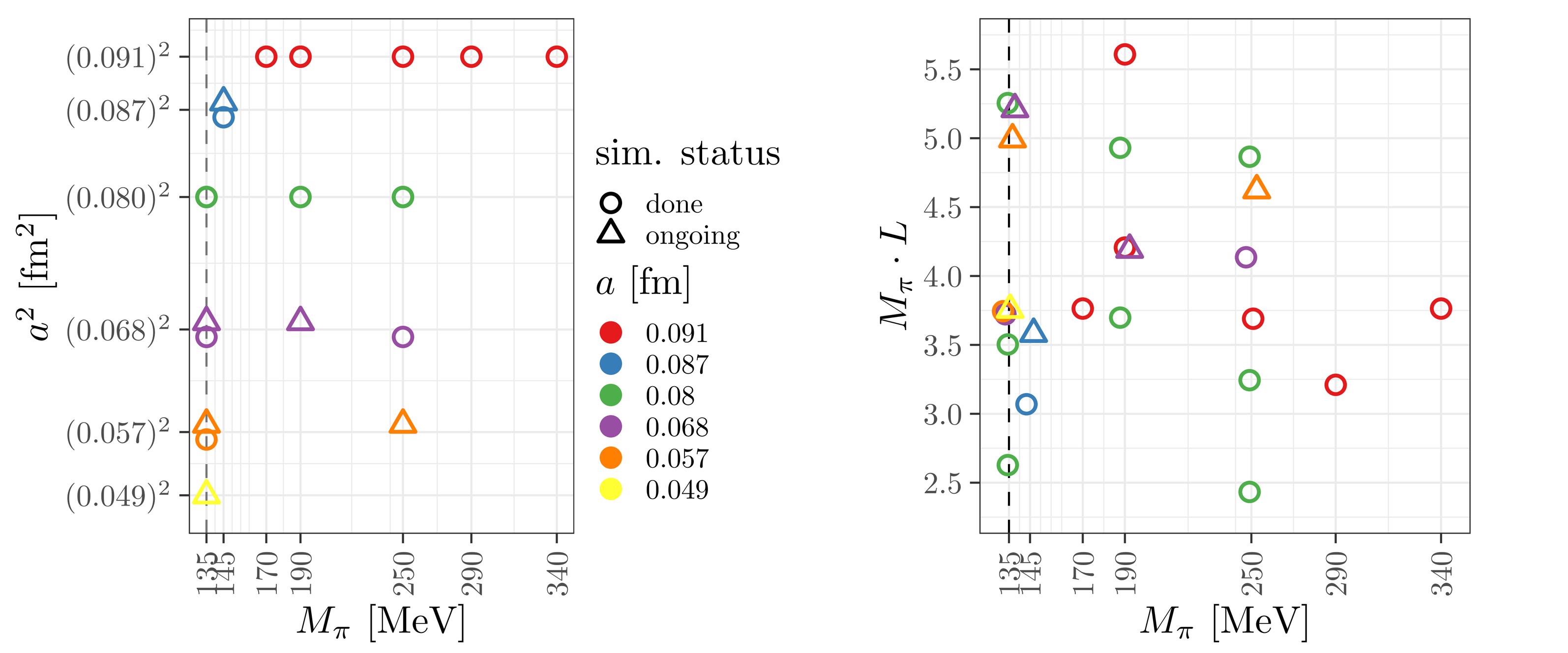
where D_ℓ is the Dirac operator for a doublet of light mass-degenerate quarks and D_h is the Dirac operator for a non-degenerate doublet of quarks corresponding to the strange and charm contributon:

$$D_\ell = (D_{\text{sw}}[U] + m_0)\, 1_f + i\mu_\ell \gamma_5 \tau_f^3\,, \qquad D_h = (D_{\text{sw}}[U] + m_0)\, 1_f + i\bar{\mu}\gamma_5\tau_f^3 - \bar{\epsilon}\tau_f^1$$

$$Q = \gamma_5 D \underset{\text{e/o precon}}{\rightarrow} \hat{Q} \underset{\text{Hasenbusch}}{\rightarrow} \hat{W}(\pm\rho) = \hat{Q} \pm i\rho$$

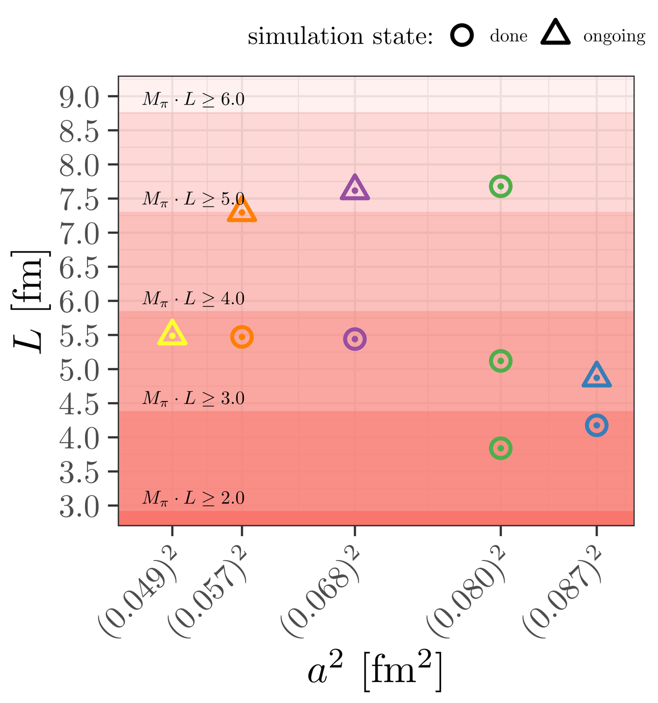
where D_{sw} is the Wilson clover operator while m_0, μ_ℓ, $\bar{\mu}$ and $\bar{\epsilon}$ are the various untwisted and twisted mass parameters. We employ Hasenbusch mass-preconditioning [6] to split the light quark determinant and rational HMC [3] with split partial fractions in the non-degenerate one. Even-odd preconditioning is used throughout.

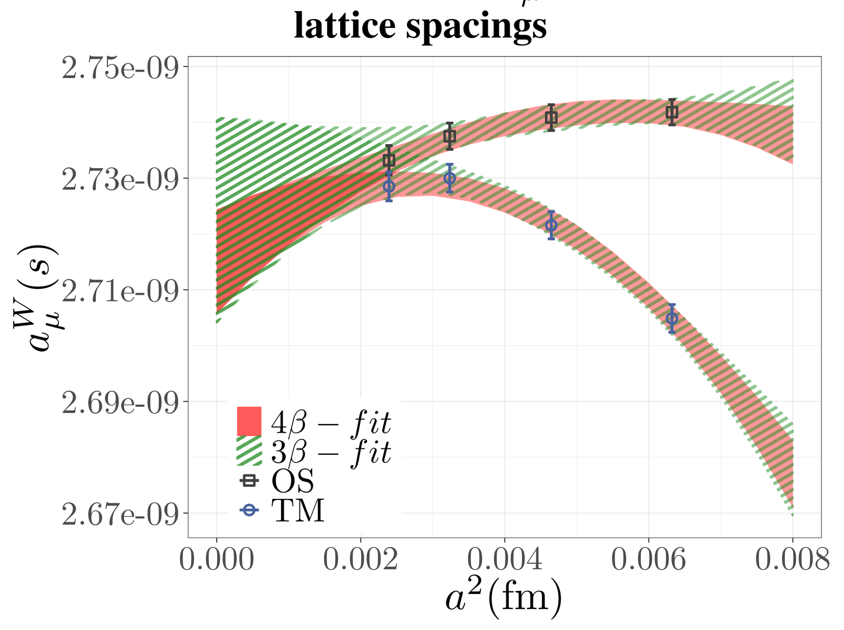
Overview of current ensembles



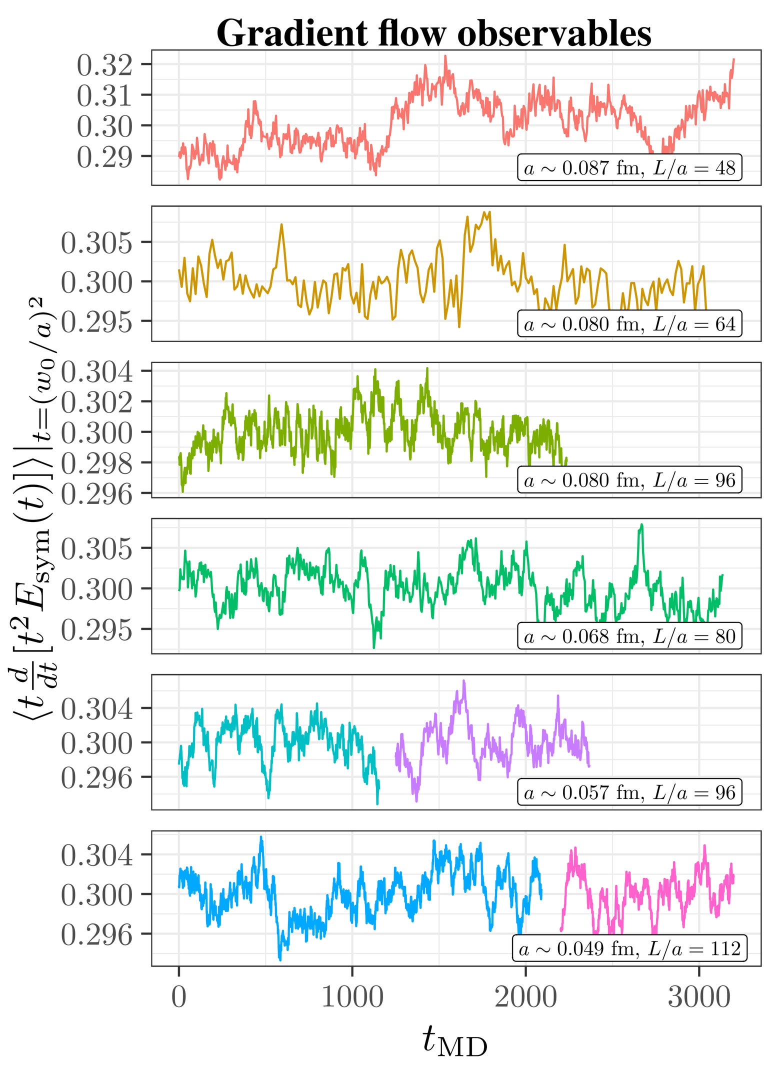
- 7 approximate pion mass values: 135 MeV $\leq M_\pi \leq$ 340 MeV
- 6 lattice spacings: 0.049 fm $\leq a \leq$ 0.091 fm, 5 at or close to the physical pion mass
- multiple volumes at many pion mass points: 2.0 fm $\leq L \leq$ 7.7 fm

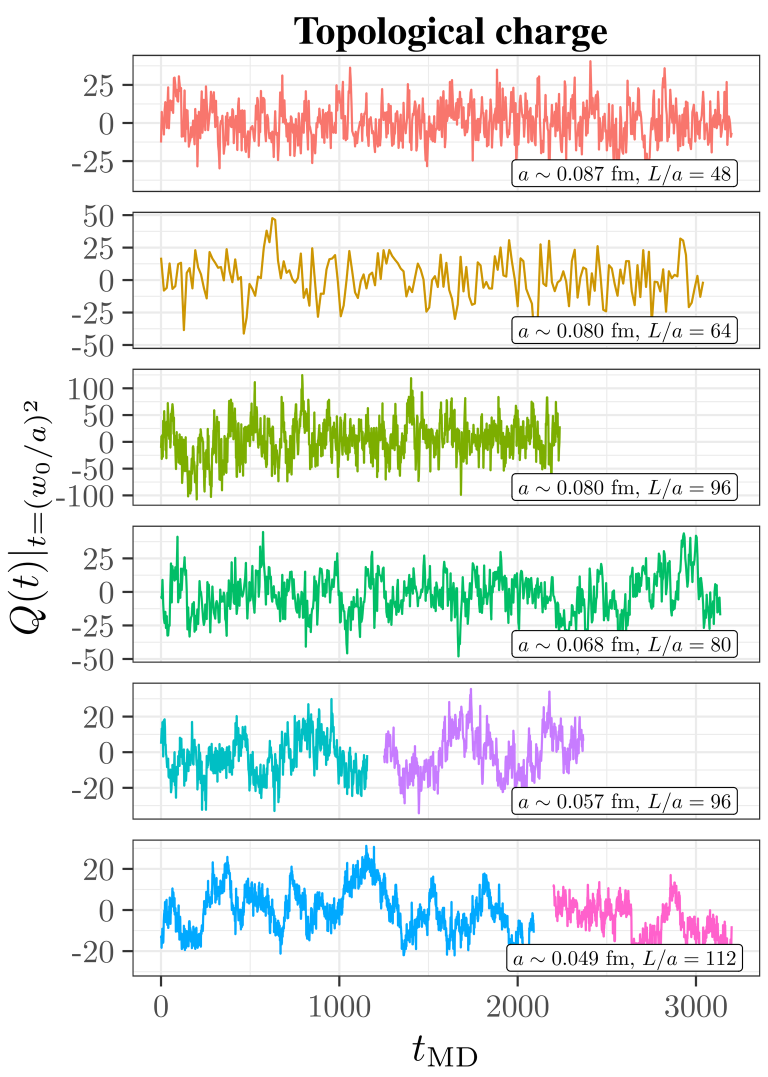
Ensembles close to the physical point





Statistical properties close to the physical pion mass





- Long τ_{int} in w_0/a at coarse lattice spacing, not negligible at fine lattice spacing.
- Topological charge moves well even at $a \approx 0.049$ fm

Algorithmic choices for the MD Hamiltonian

The molecular dynamics Hamiltonian is decomposed into monomials integrated on different time scales according to their contribution to the force using a second-order minimal norm integrator, including a force-gradient term for large volumes, 2MN(FG). The inversions are performed using the most appropriate solver for each monomial.

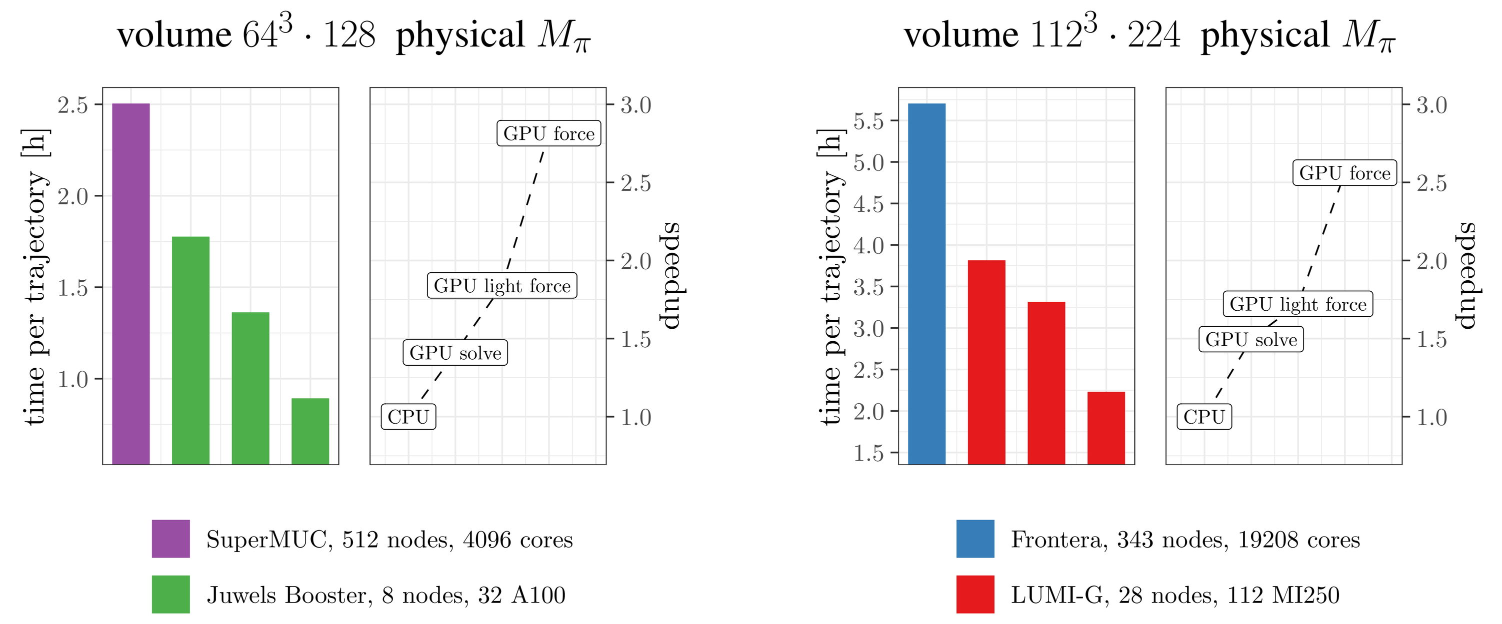
$\frac{1}{\hat{W}_+(\rho_\ell)\hat{W}_-(\rho_\ell)}$	$N_f = 2$ determinant $\rightarrow$ double-half mixed precision CG
$\hat{W}_-(\rho_\ell)\frac{1}{\hat{W}_+(\rho_h)\hat{W}_-(\rho_h)}\hat{W}_+(\rho_\ell)$	$N_f = 2$ determinant ratios $\rightarrow$ multigrid-preconditoned GCR or double-half mixed precision CG, depending on ρ
$\prod_{i=n_\ell}^{n_k} \frac{\hat{Q}_h^2 + a_{2i-1}}{\hat{Q}_h^2 + a_{2i}}$	$1+1$ non-degenerate determinant $\rightarrow$ RHMC, single precision multi-shift CG with double-half precision shift-by-shift refinement

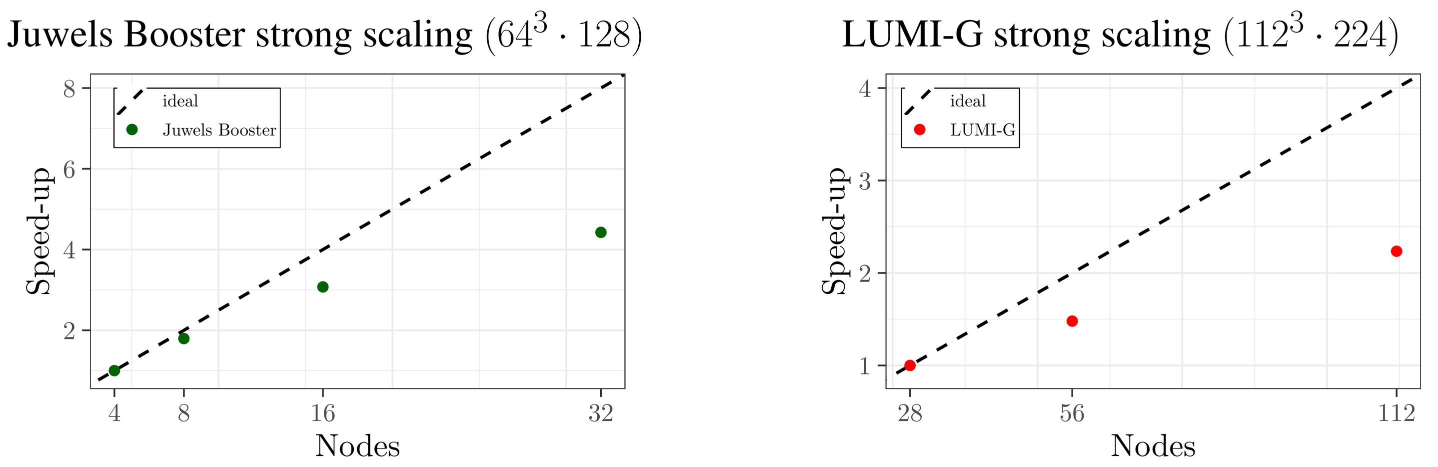
tmLQCD + QUDA

Efficient usage of GPUs is achieved in tmLQCD via the QUDA library [2, 1]. As a first step in interfacing tmLQCD to QUDA [7] we offloaded the most expensive part of the HMC, i.e. the inversion of the various Dirac operators and the gauge force computation. In the last year, we further offloaded all computations of the fermionic force. In this way, tmLQCD can reach GPU utilizations over 70% and even up to 90% depending on the number of available CPU cores per GPU.

real-time speedup through increased GPU offloading

(left to right) CPU only, GPU solves only, GPU $N_f = 2$ force, full GPU $N_f = 2+1+1$ force





Acknowledgments

We would like to thank the QUDA developers for their tremendous work as well as the many pleasant and productive interactions during this and previous efforts. We thank the ETMC for the most enjoyable collaboration. For part of this work. S.B. and J.F. were supported by the H2020 project PRACE 6-IP (grant agreement No. 82376) and the EuroCC project (grant agreement No. 951740). We acknowledge support by the European Joint Doctorate program STIMULATE grant agreement No. 765048. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and the NSFC through the funds provided to the Sino-German Collaborative Research Centre CRC 110 "Symmetries and the Emergence of Structure in QCD" (DFG Project-ID 196253076 - TRR 110, NSFC Grant No. 12070131001). We acknowledge support from projects NextQCD (EXCELLENCE/0918/0129) and "3D-Nucleon" (EXCELLENCE/0421/0043), co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation. We further gratefully acknowledge: The Gauss Centre for Supercomputing e.V. for funding this project through computing time on the GCS supercomputers SuperMUC and SuperMUC-NG at Leibniz Supercomputing Center as well as JUWELS Booster at the Jülich Supercomputing Centre. PRACE for awarding access to HAWK at HLRS within the project with Id Acid 4886. The Swiss National Supercomputing Centre (CSCS) and the EuroHPC Joint Undertaking for awarding access to the LUMI supercomputer, owned by the EuroHPC JU, hosted by CSC and the LUMI consortium through the Chronos programme under project IDs CH17-CSCS-CYP and CH21-CSCS-UNIBE as well as EuroHPC under project ID EHPC-REG-2021R0095. The Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources on Frontera (Project ID PHY21001). Some of the simulations were performed on the Luxembourg national supercomputer MeluXina.

References

[1] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower, and S. Gottlieb. Scaling lattice QCD beyond 100 GPUs. In *International Conference for High Performance Computing, Networking, Storage and Analysis*, 9 2011.

[2] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi. Solving Lattice QCD systems of equations using mixed precision solvers on GPUs. *Comput. Phys. Commun.*, 181:1517–1528, 2010.

[3] M. A. Clark and A. D. Kennedy. Accelerating dynamical fermion computations using the rational hybrid Monte Carlo (RHMC) algorithm with multiple pseudofermion fields. *Phys. Rev. Lett.*, 98:051601, 2007.

[4] R. Frezzotti and G. C. Rossi. Chirally improving Wilson fermions. 1. O(a) improvement. *JHEP*, 08:007, 2004.

[5] R. Frezzotti and G. C. Rossi. Chirally improving Wilson fermions. II. Four-quark operators. *JHEP*, 10:070, 2004.

[6] M. Hasenbusch. Speeding up the hybrid Monte Carlo algorithm for dynamical fermions. *Phys. Lett. B*, 519:177–182, 2001.

[7] B. Kostrzewa, S. Bacchio, J. Finkenrath, M. Garofalo, F. Pittler, S. Romiti, and C. Urbach. Twisted mass ensemble generation on GPU machines. *PoS*, LATTICE2022:340, 2023.

[8] B. Sheikholeslami and R. Wohlert. Improved Continuum Limit Lattice Action for QCD with Wilson Fermions. *Nucl. Phys. B*, 259:572, 1985.