PoS PROCEEDINGS OF SCIENCE

Lattice gauge ensembles and data management

Y. Aoki,[a,#] **E. Bennett,**[b,1,*] **R. Bignell,**[c,2,*] **K. U. Can,**[d,3,*] **T. Doi,**[e,4,*] **S. Gottlieb,**[f,5,*] **R. Gupta,**[g,6,*] **G. von Hippel,**[h,7,*] **I. Kanamori,**[a,8,*] **A. Kotov,**[i,9,*] **G. Koutsou,**[j,10,#,*] **R. Mawhinney,**[k,11,*] **A. Patella,**[l,12,*] **G. Pederiva,**[i,13,*] **C. Schmidt,**[m,14,*] **T. Yamazaki**[n,15,*] **and Y.-B. Yang**[o,16,*]

[a] *RIKEN Center for Computational Science (R-CCS), Kobe, 650-0047, Japan*
[b] *Swansea Academy of Advanced Computing, Swansea University, United Kingdom*
[c] *School of Mathematics, Trinity College, Dublin, Ireland*
[d] *CSSM, Department of Physics, The University of Adelaide, Australia*
[e] *Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Japan*
[f] *Department of Physics, Indiana University, IN, USA*
[g] *Theoretical Division, Los Alamos National Laboratory, NM, USA*
[h] *PRISMA+ Cluster of Excellence and Institut für Kernphysik*
[i] *Jülich Supercomputing Center, Forschungszentrum Jülich, Germany*
[j] *Computation-based Science and Technology Research Center, The Cyprus Institute, Cyprus*
[k] *Physics Department, Columbia University, USA*
[l] *Humboldt Universität zu Berlin, Institut für Physik & IRIS, Germany*
[m] *Fakultät für Physik, Universität Bielefeld, Germany*
[n] *Institute of Pure and Applied Sciences, University of Tsukuba, Japan*
[o] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China*

We summarize the status of lattice QCD ensemble generation efforts and their data management characteristics. Namely, this proceeding summarizes contributions to a dedicated parallel session during the 41[st] International Symposium on Lattice Field Theory (Lattice 2024), during which representatives of 16 lattice QCD collaborations provided details on their simulation program, with focus on plans for publication, data management, and storage requirements. The parallel session was organized by the International Lattice Data Grid (ILDG), following an open call to the lattice QCD for participation in the session.

The 41st International Symposium on Lattice Field Theory (LATTICE2024)
28 July - 3 August 2024
Liverpool, UK

[1] For the TELOS collaboration [2] For the FASTSUM collaboration [3] For the CSSM/QCDSF/UKQCD collaboration [4] For the HAL QCD collaboration [5] For the MILC collaboration [6] For the Jlab/W&M/LANL/MIT/Marseille effort [7] For the CLS [8] For the JLQCD collaboration [9] For the TWEXT collaboration [10] For the ETM collaboration (ETMC) [11] For the RBC-UKQCD collaboration [12] For the RC* collaboration [13] For the OPEN LAT initiative [14] For the HotQCD collaboration [15] For the PACS collaboration [16] For the CLQCD collaboration [#] Conveners [*] Speaker

Contents

1. Introduction

The simulation of Quantum Chromodynamics (QCD) via its Euclidean-time, discrete formulation on a lattice, has been one of the most compute-intensive applications in scientific computing, consuming substantial fractions of computer time at leadership HPC facilities internationally. In particular, the generation of ensembles of gauge configurations, for multiple values of the QCD parameters such as the QCD coupling, the quark masses, and the extent of the finite volume, requires multi-year simulation campaigns, coordinated by multi-member research collaborations. It is thus common that collaborations store and reuse the same gauge ensembles for multiple observables of interest, and in many cases also share the ensembles with researchers external to the collaboration that generated them.

The purpose of this proceeding is to summarize the available gauge ensembles generated by various lattice QCD collaborations internationally, with a focus on the data management practices each collaboration employs. It follows a parallel session at the 41$^{\text{st}}$ International Symposium on Lattice Field Theory (Lattice 2024), during which 16 collaborations provided status reports of their simulation efforts, responding to an open call for participation addressed to the lattice QCD community prior to the conference. The first such session was during Lattice 2022 and a report of the contributions presented during that session can be found in Ref. [1].

These sessions are organized by the International Lattice Data Grid (ILDG) with the intention of obtaining gathering and summarizing the evolving needs of the lattice community in terms of data storage and management. The ILDG was setup in the early 2000s [2–5] by the lattice community, which realized early on the value in standardizing data management practices across the field. ILDG is organized as a federation of autonomous *regional grids*, within a single Virtual Organization [6]. It standardizes interfaces for the services, which are to be operated by each regional grid, such as storage and a searchable metadata catalog, so that the regional services are interoperable. Within ILDG, working groups specify community-wide agreed metadata schemas (QCDml) [7] to concisely mark-up the gauge configurations and develop relevant middleware tools for facilitating the use of ILDG services. The middleware and metadata specifications developed by ILDG adhere to most of the FAIR (Findable, Accessible, Interoperable, Reusable) principles [8]. A summary of recent developments in ILDG, referred to as ILDG 2.0, was presented during the same session and can be found in a separate proceeding [9].

In the remainder of this proceeding, we present the status of ensemble generation of each of the 16 collaborations that contributed to the parallel session. We restrict to simulations of QCD, and at present these are carried out using N_f=2+1, N_f=2+1+1, and N_f=1+1+1+1 sea quark flavors with various fermion discretizations. The contributors were asked to specify whether their data are public or if they plan in making them public, their interest in using ILDG services and tools for that purpose, as well as some overall information regarding storage requirements. This information is collected in a table and summary section that follows the individual contributions.

2. Contributions

The contributions from each collaboration follow, in the order presented during the parallel session. The original presentations can be found on the conference website [10].

2.1 CLQCD

2.2 Jlab/W&M/LANL/MIT/Marseille

2.3 HotQCD

2.4 FASTSUM

The FASTSUM collaboration use N_{f}=2+1 flavour anisotropic gauge ensembles in the fixed-scale approach to study the behaviour of QCD as a function of temperature in hadronic and plasma phases. Specifically we have considered the behaviour of hadronic states including light, strange, charm and bottom quarks, the electrical conductivity of QCD matter, the interquark potential and properties of the chiral transition. FASTSUM gauge fields utilise an $O\left(a^2\right)$ improved Symanzik gauge action and an $O\left(a\right)$ improved spatially stout-smeared Wilson-Clover action following the parameter tuning and zero-temperature ensembles of the Hadron Spectrum collaboration [11, 12]. "Generation 2" ensembles were generated using the Chroma [13] software suite while the newer "Generation 2L" used a modfication [14] of the openQCD [15, 16] package which introduces stout-link smearing and anisotropic actions. Generation 2 and 2L have an anisotropy $\xi = a_s/a_\tau \sim 3.5$ with $a_s \sim 0.12$ fm, $N_s = 24$ or 32 and a wide range of N_τ corresponding to $T \in [44, 760]$ MeV. Full details of these ensembles may be found in Refs. [17, 18]. We are in the process of production for "Generation 3" - a parameter set similar to "Generation 2" but with twice the anistropy $\xi \sim 7$ - using openQCD-FASTSUM [14]. We maintain a centralised metadata repository detailing (among other information) who was responsible for each run, on which machine that run was produced and where copies may be found. The gauge fields are redundantly stored on two (well-separated) storage servers managed by Swansea University in the openQCD format. The "Generation 2" ensembles are publically available [19] while other ensembles will be available after an embargo period. We anticipate making ensembles available through the next incarnation of the ILDG with supplementary information also available on Zenodo [19, 20].

2.5 TELOS

2.6 HAL QCD

2.7 TWEXT

The TWEXT (Twisted Wilson @ Extreme conditions) collaboration studies the properties of QCD at high temperature using Wilson Twisted Mass fermions. Problems under investigation include chiral properties of QCD, in particular the behavior of QCD around the chiral phase transition and its scaling window [21], topological properties of QCD and QCD axion [22], hadron masses, symmetries of QCD and others. For this purpose, TWEXT generated a set of configurations for N_f=2+1+1 fermions at the physical pion mass and also uses older configurations with heavier pion mass. Configurations with the physical pion mass have three lattice spacings $a \in (0.057, 0.080)$ fm and cover a wide range of temperatures from $\sim$ 120 MeV to $\sim$ 900 MeV. It allows the TWEXT collaboration to perform the continuum extrapolation for quantities of interest in this temperature range. For the generation the tmLQCD software package [23–25] is used and the parameters of the ensembles were taken from the zero temperature simulations of the ETM collaboration [26]. Currently, the TWEXT collaboration has 80 ensembles (one ensemble corresponds to one point in the space temperature-pion mass-lattice spacing), which occupy $\sim$ 80 TB of disk space. Configurations are stored in the ILDG format. Possible collaborations are welcome and TWEXT plans to make configurations public/use ILDG in the future, after performing the ongoing analysis.

2.8 CSSM/QCDSF/UKQCD

2.9 RBC-UKQCD

2.10 OPEN LAT

2.11 RC*

2.12 ETMC

The ETM collaboration focuses on hadron spectroscopy, hadron structure, and flavor physics at zero temperature. Ensembles employ the twisted mass formulation, realizing $O(a)$-improvement by tuning to maximal twist, and include a clover term to further reduce the size of lattice artifacts. The Iwasaki gauge action is used. The main simulation effort is for the generation of ensembles with degenerate up- and down-, strange- and charm-quarks (N_f=2+1+1) with lattice spacing ranging between 0.049 and 0.091 fm. $M_\pi \cdot L$ varies from 2.5 up to ~5.5. At the time of writing, 24 ensembles are available or in the process of being generated, with 8 of these at approximately physical values of the quark masses. For a recent listing of the ensembles, see [27]. Simulations are performed using the Hybrid Monte Carlo (HMC) algorithm implemented in the tmLQCD software package [23–25]. See Ref. [26] for details on the simulation program, including the parameter tuning. The DD-αAMG [28, 29] multigrid iterative solver is employed for the most poorly conditioned monomials in the light sector while mixed-precision CG is used elsewhere. Multi-shift CG is used together with shift-by-shift refinement using DD-αAMG [30] for a number of small shifts for the heavy sector. tmLQCD has interfaces to QPhiX [31] and QUDA [32, 33]. tmLQCD automatically writes gauge configurations in the ILDG format, with meta-data including creation date, target simulation parameters, and the plaquette. ETMC policy is to make ensembles publicly available after a grace period. Older N_f=2 and N_f=2+1+1 ensembles [34–36] have made use of ILDG storage elements. The current ensembles are available upon request and the collaboration intends to use ILDG in the near future. For these ensembles, we expect storage requirements to reach 3 PB.

2.13 JLQCD

2.14 MILC

2.15 CLS

2.16 PACS

3. Summary

Table 1: Public: (2 = currently public, 1 = after an embargo period, 0 = no); ILDG: (N = no interest, I = interest, P = planned, U = already using); #end: Number of ensembles; #cfg: Total number of configurations; storage: Total storage needed in TBytes.

Collaboration	Public	ILDG	#ens	#cfg	Storage (TB)
CLQCD					
Jlab/W&M/LANL/MIT/Marseille					
HotQCD					
FASTSUM	Y	I	37	37,000	65
TELOS					
HAL QCD					
TWEXT	1	P	80	70,000	80
CSSM/QCDSF/UKQCD					
RBC-UKQCD					
OPEN LAT					
RC*					
ETMC	1	P,U	24	12,000	3,000
JLQCD					
MILC					
CLS					
PACS					

Acknowledgments

GK acknowledges support from EXCELLENCE/0421/0195, co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation and the AQTIVATE a Marie Skłodowska-Curie Doctoral Network GA No. 101072344. RB acknowledges support from a Science Foundation Ireland Frontiers for the Future Project award with grant number SFI-21/FFP-P/10186

The participating collaborations acknowledge the following HPC systems, for the generation of the gauge ensembles reported here,

LUMI-C and LUMI-G at CSC, Finland,

Irène Joliot-Curie at Très Grand Centre de Calcul (TGCC) in Bruyères-le-Châtel, France,

JUGENE, JUWELS and JUWELS-Booster at Jülich Supercomputing Centre (JSC),

HAWK at Höchstleistungsrechenzentrum Stuttgart (HLRS), Germany,

SuperMUC and SuperMUC-NG at Leibniz Rechenzentrum (LRZ) in Garching, Germany,

Kay and Stokes at the Irish Centre for High-End Computing (ICHEC) in Galway, Ireland,

Leonardo, Marconi 100 and Marconi A2 at CINECA in Bologna, Italy,

Tesseract at the DiRAC Extreme Scaling Service at the University of Edinburgh, United Kingdom,

DiRAC Data Intensive 2.5 & 3 at the University of Leicester, United Kingdom,
Sunbird of Supercomputing Wales at Swansea University, United Kingdom,
and Frontera at TACC, TX, US.

References

[1] G. Bali et al., *Lattice gauge ensembles and data management*, *PoS* **LATTICE2022** (2022) 203 [2212.10138].

[2] UKQCD collaboration, C. T. H. Davies, A. C. Irving, R. D. Kenway and C. M. Maynard, *International lattice data grid*, *Nucl. Phys. B Proc. Suppl.* **119** (2003) 225 [hep-lat/0209121].

[3] T. Yoshie, *Making use of the International Lattice Data Grid*, *PoS* **LATTICE2008** (2008) 019 [0812.0849].

[4] C. M. Maynard, *International Lattice Data Grid: Turn On, Plug In,and Download*, *PoS* **LAT2009** (2009) 020 [1001.5207].

[5] M. G. Beckett, B. Joo, C. M. Maynard, D. Pleiter, O. Tatebe et al., *Building the International Lattice Data Grid*, *Comput. Phys. Commun.* **182** (2011) 1208 [0910.1692].

[6] I. L. D. Grid, "Organization of ildg activities." https://hpc.desy.de/ildg/organization/. Accessed 2024-08-06.

[7] ILDG Metadata Working Group collaboration, P. Coddington, B. Joo, C. M. Maynard, D. Pleiter and T. Yoshie, *Marking up lattice QCD configurations and ensembles*, *PoS* **LATTICE2007** (2007) 048 [0710.0230].

[8] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton et al., *The fair guiding principles for scientific data management and stewardship*, *Scientific Data* **3** (2016) 160018.

[9] H. Matsufuru, H. Simma and C. Urbach, *ILDG 2.0*, *PoS* **LATTICE2024** (2024) .

[10] "Lattice data session." https://conference.ippp.dur.ac.uk/event/1265/sessions/1744/#20240802. Accessed 2024-08-02.

[11] R. G. Edwards, B. Joo and H.-W. Lin, *Tuning for Three-flavors of Anisotropic Clover Fermions with Stout-link Smearing*, *Phys. Rev. D* **78** (2008) 054501 [0803.3960].

[12] Hadron Spectrum collaboration, H.-W. Lin et al., *First results from 2+1 dynamical quark flavors on an anisotropic lattice: Light-hadron spectroscopy and setting the strange-quark mass*, *Phys. Rev. D* **79** (2009) 034502 [0810.3588].

[13] SciDAC, LHPC, UKQCD collaboration, R. G. Edwards and B. Joó, *The Chroma software system for lattice QCD*, *Nucl. Phys. Proc. Suppl.* **140** (2005) 832 [hep-lat/0409003].

[14] J. R. Glesaaen and B. Jäger, "openqcd-fastsum." https://gitlab.com/fastsum, Apr., 2018. 10.5281/zenodo.2216355.

[15] M. Lüscher and S. Schaefer, *Lattice QCD with open boundary conditions and twisted-mass reweighting*, *Comput. Phys. Commun.* **184** (2013) 519 [1206.2809].

[16] "openQCD: Simulation programs for lattice QCD." https://luscher.web.cern.ch/luscher/openQCD/.

[17] G. Aarts, C. Allton, A. Amato, P. Giudice, S. Hands et al., *Electrical conductivity and charge diffusion in thermal QCD from the lattice*, *JHEP* **02** (2015) 186 [1412.6411].

[18] G. Aarts et al., *Properties of the QCD thermal transition with Nf=2+1 flavors of Wilson quark*, *Phys. Rev. D* **105** (2022) 034504 [2007.04188].

[19] G. Aarts, C. Allton, A. Amato, R. Bignell, T. J. Burns et al., *FASTSUM Generation 2 Anisotropic Thermal Lattice QCD Gauge Ensembles*, July, 2024. 10.5281/zenodo.8403827.

[20] European Organization For Nuclear Research and OpenAIRE, *Zenodo*, 2013. 10.25495/7GXK-RD71.

[21] A. Y. Kotov, M. P. Lombardo and A. Trunin, *QCD transition at the physical point, and its scaling window from twisted mass Wilson fermions*, *Phys. Lett. B* **823** (2021) 136749 [2105.09842].

[22] A. Y. Kotov, A. Trunin and M. P. Lombardo, *QCD topology and axion's properties from Wilson twisted mass lattice simulations*, *PoS* **LATTICE2021** (2022) 032 [2111.15421].

[23] K. Jansen and C. Urbach, *tmLQCD: A Program suite to simulate Wilson Twisted mass Lattice QCD*, *Comput. Phys. Commun.* **180** (2009) 2717 [0905.3331].

[24] A. Deuzeman, K. Jansen, B. Kostrzewa and C. Urbach, *Experiences with OpenMP in tmLQCD*, *PoS* **LATTICE2013** (2014) 416 [1311.4521].

[25] A. Abdel-Rehim, F. Burger, A. Deuzeman, K. Jansen, B. Kostrzewa et al., *Recent developments in the tmLQCD software suite*, *PoS* **LATTICE2013** (2014) 414 [1311.5495].

[26] C. Alexandrou et al., *Simulating twisted mass fermions at physical light, strange and charm quark masses*, *Phys. Rev. D* **98** (2018) 054518 [1807.00495].

[27] B. Kostrzewa, "Status of the etmc ensemble generation effort." https://conference.ippp.dur.ac.uk/event/1265/contributions/7655/.

[28] A. Frommer, K. Kahl, S. Krieg, B. Leder and M. Rottmann, *Adaptive Aggregation-Based Domain Decomposition Multigrid for the Lattice Wilson–Dirac Operator*, *SIAM J. Sci. Comput.* **36** (2014) A1581 [1303.1377].

[29] C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl et al., *Adaptive Aggregation-based Domain Decomposition Multigrid for Twisted Mass Fermions*, *Phys. Rev. D* **94** (2016) 114509 [1610.02370].

[30] C. Alexandrou, S. Bacchio and J. Finkenrath, *Multigrid approach in shifted linear systems for the non-degenerated twisted mass operator*, Comput. Phys. Commun. **236** (2019) 51 [1805.09584].

[31] B. Joó, D. D. Kalamkar, K. Vaidyanathan, M. Smelyanskiy, K. Pamnany et al., *Lattice QCD on Intel® Xeon Phi Coprocessors*, Lect. Notes Comput. Sci. **7905** (2013) 40.

[32] QUDA collaboration, 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** (2010) 1517 [0911.3191].

[33] QUDA collaboration, R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower et al., *Scaling lattice QCD beyond 100 GPUs*, in *International Conference for High Performance Computing, Networking, Storage and Analysis*, 9, 2011, 1109.2935, DOI.

[34] R. Baron et al., *Light hadrons from lattice QCD with light (u,d), strange and charm dynamical quarks*, JHEP **06** (2010) 111 [1004.5284].

[35] European Twisted Mass collaboration, R. Baron et al., *Computing K and D meson masses with $N_f = 2{+}1{+}1$ twisted mass lattice QCD*, Comput. Phys. Commun. **182** (2011) 299 [1005.2042].

[36] ETM collaboration, R. Baron et al., *Light Meson Physics from Maximally Twisted Mass Lattice QCD*, JHEP **08** (2010) 097 [0911.5061].