

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,*}**

^aRIKEN Center for Computational Science (R-CCS), Kobe, 650-0047, Japan

^bSwansea Academy of Advanced Computing, Swansea University, United Kingdom

^cSchool of Mathematics, Trinity College, Dublin, Ireland

^dCSSM, Department of Physics, The University of Adelaide, Australia

^eInterdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, Japan

^fDepartment of Physics, Indiana University, IN, USA

^gTheoretical Division, Los Alamos National Laboratory, NM, USA

^hPRISMA+ Cluster of Excellence and Institut für Kernphysik

ⁱJülich Supercomputing Center, Forschungszentrum Jülich, Germany

^jComputation-based Science and Technology Research Center, The Cyprus Institute, Cyprus

^kPhysics Department, Columbia University, USA

^lHumboldt Universität zu Berlin, Institut für Physik & IRIS, Germany

^mFakultät für Physik, Universität Bielefeld, Germany

ⁿInstitute of Pure and Applied Sciences, University of Tsukuba, Japan

^oSchool 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 41st 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	2
2	Contributions	3
2.1	CLQCD	4
2.2	Jlab/W&M/LANL/MIT/Marseille	4
2.3	HotQCD	5
2.4	FASTSUM	5
2.5	TELOS	6
2.6	HAL QCD	6
2.7	TWEXT	7
2.8	CSSM/QCDSF/UKQCD	7
2.9	RBC-UKQCD	7
2.10	OPEN LAT	8
2.11	RC*	8
2.12	ETMC	9
2.13	JLQCD	9
2.14	MILC	10
2.15	CLS	10
2.16	PACS	11
3	Summary	11

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 41st 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.3 HotQCD

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.4 FASTSUM

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.5 TELOS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.6 HAL QCD

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.7 TWEXT

1 The TWEXT (Twisted Wilson @ Extreme conditions) collaboration studies the properties of
2 QCD at high temperature using Wilson Twisted Mass fermions. Problems under investigation in-
3 clude chiral properties of QCD, in particular the behavior of QCD around the chiral phase transition
4 and its scaling window [11], topological properties of QCD and QCD axion [12], hadron masses,
5 symmetries of QCD and others. For this purpose, TWEXT generated a set of configurations for
6 $N_f=2+1+1$ fermions at the physical pion mass and also uses older configurations with heavier pion
7 mass. Configurations with the physical pion mass have three lattice spacings $a \in (0.057, 0.080)$ fm
8 and cover a wide range of temperatures from ~ 120 MeV to ~ 900 MeV. It allows the TWEXT
9 collaboration to perform the continuum extrapolation for quantities of interest in this temperature
10 range. For the generation the tmLQCD software package [13–15] is used and the parameters of
11 the ensembles were taken from the zero temperature simulations of the ETM collaboration [16].
12 Currently, the TWEXT collaboration has 80 ensembles (one ensemble corresponds to one point in
13 the space temperature-pion mass-lattice spacing), which occupy ~ 80 TB of disk space. Configu-
14 rations are stored in the ILDG format. Possible collaborations are welcome and TWEXT plans to
15 make configurations public/use ILDG in the future, after performing the ongoing analysis.

2.8 CSSM/QCDSF/UKQCD

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.9 RBC-UKQCD

1
2
3

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.10 OPEN LAT

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.11 RC*

1
2
3
4

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.12 ETMC

1 The ETMC collaboration focuses on hadron spectroscopy, hadron structure, and flavor physics
2 at zero temperature. Ensembles employ the twisted mass formulation, realizing $O(a)$ -improvement
3 by tuning to maximal twist, and include a clover term to further reduce the size of lattice artifacts.
4 The Iwasaki gauge action is used. The main simulation effort is for the generation of ensembles
5 with degenerate up- and down-, strange- and charm-quarks ($N_f=2+1+1$) with lattice spacing ranging
6 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
7 are available or in the process of being generated, with 8 of these at approximately physical values of
8 the quark masses. For a recent listing of the ensembles, see [17]. Simulations are performed using
9 the Hybrid Monte Carlo (HMC) algorithm implemented in the tmLQCD software package [13–15].
10 See Ref. [16] for details on the simulation program, including the parameter tuning. The DD-
11 α AMG [18, 19] multigrid iterative solver is employed for the most poorly conditioned monomials
12 in the light sector while mixed-precision CG is used elsewhere. Multi-shift CG is used together
13 with shift-by-shift refinement using DD- α AMG [20] for a number of small shifts for the heavy
14 sector. tmLQCD has interfaces to QPhiX [21] and QUDA [22, 23]. tmLQCD automatically writes
15 gauge configurations in the ILDG format, with meta-data including creation date, target simulation
16 parameters, and the plaquette. ETMC policy is to make ensembles publicly available after a grace
17 period. Older $N_f=2$ and $N_f=2+1+1$ ensembles [24–26] have made use of ILDG storage elements.
18 The current ensembles are available upon request and the collaboration intends to use ILDG in the
19 near future. For these ensembles, we expect storage requirements to reach 3 PB.

2.13 JLQCD

1
2
3
4
5

6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.14 MILC

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

2.15 CLS

1
2
3
4
5
6

7
8
9
10
11
12
13
14
15
16
17
18
19

2.16 PACS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

3. Summary

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.

Table 1: Public: (Y = Yes, N = No); ILDG: (N = no interest, I = interest, P = planned, U = already using); #ens: 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					
TELOS					
HAL QCD					
TWEXT					
CSSM/QCDSF/UKQCD	Y	P	80	70,000	80
RBC-UKQCD					
OPEN LAT					
RC*					
ETMC	Y	P,U	24	12,000	3,000
JLQCD					
MILC					
CLS					
PACS					

The participating collaborations acknowledge the following HPC systems, for the generation of the gauge ensembles reported here, LUMI-C and LUMI-G at CSC in Finland, JUWELS and JUWELS-Booster at Jülich Supercomputing Centre (JSC), SuperMUC and SuperMUC-NG at Leibniz Rechenzentrum (LRZ) in Garching, Leonardo at CINECA in Bologna, Italy, 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] 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].
- [12] 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].
- [13] K. Jansen and C. Urbach, *tmLQCD: A Program suite to simulate Wilson Twisted mass Lattice QCD*, *Comput. Phys. Commun.* **180** (2009) 2717 [0905.3331].
- [14] A. Deuzeman, K. Jansen, B. Kostrzewa and C. Urbach, *Experiences with OpenMP in tmLQCD*, *PoS LATTICE2013* (2014) 416 [1311.4521].
- [15] 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].
- [16] C. Alexandrou et al., *Simulating twisted mass fermions at physical light, strange and charm quark masses*, *Phys. Rev. D* **98** (2018) 054518 [1807.00495].
- [17] B. Kostrzewa, “Status of the etmc ensemble generation effort.”
<https://conference.ippp.dur.ac.uk/event/1265/contributions/7655/>.
- [18] 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].
- [19] 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].

- [20] 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](#)].
- [21] 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.
- [22] 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](#)].
- [23] 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.
- [24] R. Baron et al., *Light hadrons from lattice QCD with light (u,d), strange and charm dynamical quarks*, *JHEP* **06** (2010) 111 [[1004.5284](#)].
- [25] 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](#)].
- [26] ETM collaboration, R. Baron et al., *Light Meson Physics from Maximally Twisted Mass Lattice QCD*, *JHEP* **08** (2010) 097 [[0911.5061](#)].