PROCEEDINGS OF SCIENCE

Lattice gauge ensembles and data management

Gunnar Bali,[a,#] **Ryan Bignell,**[b,1,*] **Anthony Francis,**[c,2,*] **Steven Gottlieb,**[d,3,*] **Rajan Gupta,**[e,4,*] **Issaku Kanamori,**[f,5,*] **Bartosz Kostrzewa,**[g,6,*] **Andrey Yu. Kotov,**[h,7,*] **Yoshinobu Kuramashi,**[i,8,*] **Robert Mawhinney,**[j,9,*] **Christian Schmidt,**[k,10,*] **Wolfgang Söldner**[a,11,*] **and Peng Sun**[l,m,12,*]

[a] *Institut für Theoretische Physik, Universität Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany*
[b] *Department of Physics, Swansea University, Swansea SA2 8PP, United Kingdom*
[c] *Institute of Physics, National Yang Ming Chiao Tung University, 30010 Hsinchu, Taiwan*
[d] *Department of Physics, Indiana University, Bloomington, IN 47405, USA*
[e] *Los Alamos National Laboratory, Theoretical Division T-2, Los Alamos, NM 87545, USA*
[f] *RIKEN Center for Computational Science (R-CCS), 7-1-26, Minatojima Minamimachi, Kobe 650-0047, Japan*
[g] *High Performance Computing and Analytics Lab, Rheinische Friedrich-Wilhelms-Universität Bonn, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany*
[h] *Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, 52428 Jülich, Germany*
[i] *Center for Computational Sciences, University of Tsukuba, 305-8577, Tsukuba, Japan*
[j] *Physics Department, Columbia University, New York, NY 10027, USA*
[k] *Fakultät für Physik, Universität Bielefeld, Universitätsstr. 25, 33615 Bielefeld, Germany*
[l] *Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China*
[m] *University of Chinese Academy of Sciences, Beijing 100049, China*

The generation of ensembles of gauge configurations is a considerable expense. The preservation and curation of these ensembles constitutes a valuable shared resource for the lattice field theory community. The organizers of Lattice 2022 dedicated a parallel session to the presentation of gauge ensembles and their generation, plans for ensemble publication and data management/storage activities of different collaborations. A summary of the twelve contributions is presented here.

The 39th International Symposium on Lattice Field Theory,
8th-13th August, 2022,
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

[1]For the FASTSUM Collaboration.
[2]For the Open Lattice Initiative.
[3]For the MILC Collaboration.
[4]For the JLab, College of William & Mary, LANL, MIT, OLCF, Marseille Collaboration.
[5]For the JLQCD Collaboration.
[6]For the Extended Twisted Mass Collaboration (ETMC).
[7]For the TWEXT Collaboration.
[8]For the PACS Collaboration.
[9]For the RBC-UKQCD Collaborations.
[10]For the HotQCD Collaboration.
[11]For the CLS Community Effort.
[12]For the CLQCD Collaboration.
[#]Convenor
[*]Speaker

PoS(LATTICE2022)203

Contents

1. Introduction

Lattice QCD calculations factorize into at least three parts, that we label "generation," "measurements" and "analysis." First of all, ensembles of gauge configurations are generated using the Markov chain Monte Carlo method. Second, a set of correlation functions that is relevant for the specific physics programme is computed on these ensembles. This process is often described as "making measurements" on gauge configurations. Finally, the correlation functions are statistically analysed and combined into the observables of interest. These data are then extrapolated to the infinite volume and continuum limits, as well as to the quark masses of interest, quantifying systematic and statistical errors.

The first two steps (generation and measurement) are compute intensive. During measurements, the building blocks are quark propagators with different types of sources. These are mostly computed on the fly and not stored for re-use, while perambulators [1] and eigenvectors, that have broader applications, are stored. The resulting correlation functions are observable specific. Perambulators, eigenvectors and correlation functions are mostly shared within one collaboration or project. This may change in the future. Gauge ensembles are much less specific, can easily be transferred between computer centres and are often shared far beyond the group that created them.

This parallel session is directed towards summarizing what gauge ensembles exist and the respective data sharing policies and data management strategies. We shall distinguish between "data providers" who generate gauge ensembles and manage the data and "data consumers" who use these (and can also be providers). Measurements with a specific physics goal typically require less computational resources, planning and parameter tuning than a campaign to generate a set of gauge ensembles. Therefore, the former sometimes can be carried out by smaller groups, without multiple computer time proposals and with less coordination and administrative overhead. Apart from these accessibility issues, gauge ensembles are highly valuable research products and, once these exist, the science output should be maximized to further our understanding of Nature.

Already twenty years ago, the lattice community initiated the International Lattice Data Grid (ILDG) [2–5] as a framework to enable and facilitate world-wide sharing of gauge configurations. This already anticipated most of the FAIR (Findable, Acessible, Interoperable, Reusable) principles [6]. A major achievement of ILDG was the specification of a community-wide agreed metadata schema (QCDml) [7] to concisely markup the gauge configurations. ILDG is organized as a federation of autonomous "regional grids" within a single Virtual Organization. It has a uniform user registration and defines a common format for the binary storage of the gauge configurations. Standardized interfaces for the services, which are to be operated by each regional grid, like storage and a searchable metadata catalogue, render these regional services interoperable. The status and plans of the modernization of ILDG have been presented in a plenary talk at this conference [8], as well as in a lunch meeting, organized by the working groups of ILDG.

There exist different levels of the sharing of gauge ensembles between the data providers and the data consumers. Sharing may be restricted to within a collaboration, e.g., between the members who generate the configurations and those who carry out measurements and the analysis. Often, gauge ensembles which initially are only shared internally are declared "public" after some embargo time, with guidelines how these and associated publications should be cited. Other levels of access and usage are possible, e.g., limited to specific projects or sharing between two or more collaborations. In an ideal world, one may envisage the following scenario.

- In a **collaboration-internal** context, the **data consumers** know everything about the configurations being produced (availability and provenance, simulation parameters, storage location, etc.). All data management within the collaboration follows a clear and transparent data management plan and the usage rules are well-defined and known to everyone within the collaboration. The actual management of all the data and metadata is taken care of by designated members of the collaboration.
- In a **collaboration-internal** context, the **data providers** follow a well-defined and smooth workflow for collecting and archiving all gauge configurations and the corresponding metadata. Internal and shared data can be handled in essentially the same way, so that data can be declared completely "public" or shared with specific external users with little extra effort (e.g., by requiring to switch an access permission flag for an ensemble at the end of an embargo time). Data that are declared "public" are easily citable, making their impact visible to funding agencies, computing centres and the wider research community.
- In a **community-wide** context, the **data consumers** know about the existence of all publicly available gauge ensembles that may be relevant or interesting for their own research projects. All scientists in the community should obtain these at essentially no cost in terms of human or computing resources and then be able to freely carry out their high quality research. Every

PoS(LATTICE2022)203

data consumer follows good scientific practice and properly acknowledges the source of the data and gives credit to the data providers.

- In a **community-wide** context, the **data providers** can make their valuable data available on some storage infrastructure at no extra cost in terms of human or hardware resources. Declaring data "public" will make these known to other researchers who will frequently use these in other projects, so that the data providers receive recognition and citations.

In the real world, where many of those responsible for generating, storing and managing the data are on temporary positions and where large, globally accessible, long-term storage is not for free, the situation is more challenging.

In this contribution, we collect the present status of ensemble generation to inform both data consumers and providers about the availability of gauge ensembles and present practices. We restrict ourselves to simulations of QCD. At present, these are mostly carried out using $N_f = 2+1$, $N_f = 2+1+1$ and also $N_f = 1+1+1+1$ sea quark flavours $q = u, d, s, c$, with various fermion discretizations. Naturally, we can only cover simulations by the groups who responded to the call. The next section provides the current status. This is followed by a brief summary.

2. Contributions

In this section, the different collaborations present their physics programmes, gauge ensembles and data policies. The contributions are ordered like the collaborations on the title page, i.e., alphabetically in terms of the respective author names. The original presentations can be found on the indico page of the conference [9].

2.1 FASTSUM Collaboration

The FASTSUM collaboration uses $N_f = 2+1$ flavour anisotropic gauge ensembles in the fixed-scale approach to study the behaviour of QCD in as a function of the temperature in the 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. The fixed-scale approach allows for a conceptually clear investigation of temperature effects while anisotropy allows more temporal points at each temperature.

FASTSUM gauge fields utilise an order a^2 improved Symanzik gauge action and an order a improved spatially stout-smeared Wilson-Clover action following the parameter tuning and zero-temperature ensembles of the Hadron Spectrum collaboration [10, 11]. "Generation 2" ensembles were generated using the Chroma [12] software suite while the newer "Generation 2L" used a modfication [13] of the openQCD [14, 15] package which introduces stout-link smearing and anisotropic actions. We use 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. "Generation 2" and "Generation 2L" differ mainly in their quark mass. Full details of these ensembles may be found in ref. [16, 17].

As we use a derivative of the openQCD package log, files include information such as the openQCD version, algorithmic parameters, plaquette values and the run time. Additionally 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 available upon request while other ensembles will be available after an embargo time. We anticipate making ensembles available through the next incarnation of ILDG with supplementary information also available on Zenodo [18].

2.2 Open Lattice Initiative

The increase of numerical cost and data complexity in precision era lattice QCD calculations poses new questions on data generation, transfer and access. The Open Lattice Initiative (OpenLat) was established as a response to these questions in 2019. The goal of this young initiative is to generate and make available QCD gauge ensembles for a broad range of physics applications. Emphasis is put on open science and user access.

The initiative uses the stabilised Wilson fermion (SWF) framework [19]. An important new component is the exponentiated clover term in the fermion action, which cures a large volume pathology in Wilson-Clover fermions and exhibits benefits in terms of discretization effects [20]. Furthermore SWFs include a number of algorithmic improvements, such as choosing the Stochastic Molecular Dynamics (SMD) algorithm, which shows favorable autocorrelation times and stability [21, 22]. Precision losses are further prevented by using a volume-independent solver stopping criterion and quadruple precision arithmetic in global sums. These measures are implemented in addition to the established techniques, such as the Schwarz-alternating-procedure (SAP), local deflation, mass-preconditioning, multiple time-scale integrators and others, for details see ref. [19, 20]. For generation we use the open source software packages `openQCD-2.0` and `2.4` [15].

All generated gauge ensembles are $N_f = 2+1$ full QCD. The scale is determined using the gradient flow time criterion [23] and results are converted to physical units via $\sqrt{8t_0} = 0.414(5)\text{fm}$ [24]. By default (anti-)periodic boundary conditions are set with the aim to preserve translation invariance. Open boundaries are chosen once a significant slowdown in topological tunneling is observed.

The production plan of ensembles proceeds in three stages [25]: In stage 1, after a high precision tuning, ensembles are generated at the $SU(3)$-flavor symmetric point ($m_\pi = m_K = 412$ MeV) at four lattice spacings, $a = 0.12, 0.094, 0.077$ and 0.064 fm. In stage 2, the light quark masses are reduced to $M_\pi = 300$ and 200 MeV, keeping the quark mass matrix trace fixed, and the lattice spacing $a = 0.055$ fm (open boundary conditions) is added. In stage 3, the light quark masses are lowered to the physical point. Throughout $L \gtrsim 3$fm with $m_\pi L \gtrsim 4$ and $T/L \geq 2$. The initiative is currently in the process of completing stage 1. Each completed stage is accompanied by a reference publication. With this publication all configurations and metadata are made openly available without further embargo time. The public data are: the metadata catalogue, the gauge configurations and the basic observables, see ref. [20] for a list. The metadata catalogue follows a detailed provenance policy and will be compliant with community standards, such as ILDG. All metadata is preserved on disk and in the main online repository [26]. Users may obtain access to the configurations of ongoing, i.e. unpublished, stages. This user-access is granted on a case-by-case basis.

The initiative maintains a repository of 100 TB total, which is projected to grow to 500+ TB with the completion of stage 2. At this point the repository includes ~ 20k saved configurations in the so-called openQCD format. The data is mirrored at two separate locations on disk and tape.

As per openQCD format standard the plaquette is given in the header of each configuration file, additionally `sha512` checksums are kept separately as means to monitor data integrity.

2.3 MILC Collaboration

The MILC Collaboration has been sharing code and lattice configurations for approximately 30 years. We started with $N_f = 2$ staggered quarks, transitioned to the asqtad action with $N_f = 2+1$, and are now using the HISQ action [27] with $N_f = 2+1+1$ or $1+1+1+1$. Our original physics interests included both zero and non-zero temperature QCD. Our work on non-zero temperature QCD culminated with our participation in HotQCD, which is covered later in this article. Our main interests now are leptonic and semileptonic decays of mesons containing charm or bottom quarks; properties of baryons using staggered quarks, all currently done together with the Fermilab Lattice Collaboration; and the muon anomalous magnetic moment, done with the Fermilab Lattice and HPQCD Collaborations. We are also interested in electromagnetic corrections.

We were initial users of the Gauge Connection at NERSC where many asqtad ensembles were made available in the ILDG format. Unfortunately, this service is in disarray, but there are plans for its revival. The asqtad physics program and the associated ensembles were extensively reviewed in ref. [28]. The more modern ensembles using the HISQ action are our current focus. We have approximately 25 publicly available ensembles with lattice spacings $a \approx 0.15$, 0.12, 0.09, 0.06, and 0.042 fm. At each lattice spacing, there is one ensemble where the light, strange, and charm quark masses are closely tuned to physical values. There are also several ensembles with the light quark mass higher than its physical value, enabling study of the chiral limit. At 0.12 fm, with $m_l = 0.1m_s$, we have three volumes, $L = 24$, 32, and 40. In addition, we have an ensemble with $a \approx 0.03$ fm with the light quark set to 0.2 times the strange quark mass. With the advent of exascale computers, we hope to generate a new ensemble with the light quark mass tuned so that the Goldstone pion mass is 135 MeV. There are also a number of ensembles in which the strange quark mass is reduced from its physical value to better enable chiral fits and determination of the low energy constants in the chiral lagrangian. It should also be mentioned that CalLat has been generating additional HISQ ensembles [29].

Configurations are generated using the rational hybrid molecular dynamics or rational hybrid Monte Carlo algorithms [30]. We also employ the Hasenbusch method [31]. Details may be found in refs. [32] and [33]. The performance of the MILC code, on various architectures, is enhanced by using QOP [34], QPhiX [35–38], or QUDA [39–42].

We have the log files from most of the gauge generation runs. These contain checksums for the configurations, the plaquette, the chiral condensate, and many Wilson loops. The configurations contain metadata including two checksums, the space-space and space-time plaquette, run parameters, and the time the configuration was generated. This information is also contained in an ASCII “info” file of about 600 bytes. Not all log files contain the name of the machine on which the code was run, but it is usually possible to tell from other details in the log file. We are trying to be more systematic about including such information, and the date on which the code was compiled. Both asqtad and HISQ ensembles are archived at Fermilab for the convenience of USQCD members. A second copy of most HISQ ensembles is kept on Ranch at the Texas Advanced Computing Center (TACC). A few years ago, we had to migrate our data from one tape system to a new one, and it was decided not to preserve the asqtad ensembles at TACC.

PoS(LATTICE2022)203

Our sharing policy is available on GitHub under the milc-qcd main page. The information is on the wiki of the repository called "sharing" [43]. It contains links to our sharing policy, a web page with a detailed list of the publicly available ensembles, and a document explaining which paper to cite when using each configuration and how we would like to be acknowledged.

2.4 JLab, College of William & Mary, LANL, MIT, OLCF, Marseille Collaboration

This US community effort is generating ensembles with $2+1$ flavors of Wilson-clover fermions with stout-link smearing of the gauge fields and a tree-level tadpole-improved Symanzik gauge action. One iteration of the four-dimensional stout smearing is used with the weight $\rho = 0.125$ for the staples in the rational hybrid Monte Carlo (RHMC) algorithm. After stout smearing, the tadpole-improved tree-level clover coefficient C_{SW} is very close to the nonperturbative value. This was confirmed using the Schrödinger functional method for determining the clover coefficient nonperturbatively. The tuning of the strange quark mass has been done in two ways. For the ensembles at $a = 0.127$ and 0.091 fm, we have required the ratio $(2M_{K^+}^2 - M_\pi^2)/M_{\Omega^-}$ to take on its physical value 0.1678. This tuning is done in the 3-flavor theory, and the resulting value of m_s is then kept fixed as the light-quark masses in the (2+1)-flavor theory are decreased towards their physical values. For the other ensembles, we have required $M_K = 495$ MeV, independent of the light quark masses. The lattice scale has, so far, been set using w_0.

In the HMC algorithm, the light quark generation is carried out utilizing the Hasenbusch method [31] of factorizing the 2-flavor determinant with chains of 3–4 ratios followed by a 2-flavor term. The Hasenbusch ratios are implemented by using different quark masses in the numerator and denominator of the determinant ratios, rather than by utilizing different twisted masses in these terms. The one flavor piece was carried out using a rational approximation for $\sqrt{M_s^\dagger M_s}$. The two flavor solves are carried out using the multi-grid preconditioned GCR solver implemented in QUDA [40, 44].

The integration is carried out with a nested Force Gradient integrator. We tuned the placement of terms in the Hamiltonian (monomials in the action) on various timescales, the number of steps on each timescale, and the Hasenbusch masses to minimize trajectory time while maintaining an acceptance rate of approximately 95% for our 4th order integrator. The procedure was somewhat ad-hoc, but primarily involved: a) tuning the quark masses so that the Hasenbusch ratio terms would have forces that were roughly equal in infinity-norm and less than 1 in value. Thereafter, we placed the rational approximation on the middle timescale and the gauge and 2-flavor fermion pieces on the finest timescale. Generally we set the step on the finest timescale to be very large and adjusted the middle time-scale to a level, where we would have no instability from the rational term having too big a time-step. Finally we reduced the number of steps on the finest time step.

Beyond making the multi-grid blocking sizes fit, we did not perform extensive tuning of the multigrid parameters from ensemble to ensemble. We refresh the multi-grid subspace through 'polishing' — iterating the existing subspace vectors with the current operator, once a given threshold is exceeded. If after the polish the solution does not converge within the iteration limit we regenerate the subspace. If ever the solution fails to converge even after regenerating the subspace we terminate the program.

Our computations are carried out with the Chroma code [12, 45] built over QDP-JIT [46, 47] and QUDA [40, 41] and the majority of the ensembles have been generated on the Summit system at OLCF in a variety of allocations. So far 13 ensembles have been generated at four values of the lattice spacing ($a \approx 0.127,\ 0.091,\ 0.071$ and 0.055 fm) and four values of the pion mass $M_\pi = 270,\ 220,\ 170$ and 130 MeV with between 10,000 and 20,000 thermalized trajectories each. At present 1000 lattices, each separated by 4 trajectories, are available on request to the USQCD collaboration for non-competing physics projects and the full set will be made available by the beginning of 2024 through a password protected repository at OLCF.

2.5 JLQCD Collaboration

The JLQCD collaboration puts emphasis on the chiral symmetry and uses Möbius domain-wall fermions [48] with scale factor 2 (Shamir type). Stout smearing with $\rho = 0.1$ and $n = 3$ is applied to the fermion action. The tree-level improved Symanzik gauge action is used. More details of the action are described in the supplemental material of ref. [49]. We have two different physics targets. One is B-physics with fine lattices at zero temperature. To this end, we have $2 + 1$ flavor ensembles at inverse lattice spacings ranging from 2.45 GeV to 4.5 GeV. The other target is finite temperature physics, to survey the phase structure and to reveal the nature of the $U(1)_A$ symmetry across the phase transition/cross over. We have 2, $2 + 1$, and 3 flavors ensembles for this. The same action is used for both zero and finite temperature targets. We have more than 200 ensembles in total and the size is around 20 TB, assuming we keep the configurations every 100 trajectories for the finite temperature ensembles.

The Hybrid Monte Carlo (HMC) [50] algorithm is used to generate configurations. The strange quark is treated with the rational HMC (RHMC) [51]. The one- or two-level Hasenbusch trick [31] is combined for light quarks. We employ the Omelyan integrator [52] for the molecular dynamics, and apply Metropolis accept/reject test and refresh the momenta at every unit of molecular dynamics time. We initially used the code set Iroiro++ [53] and now we use Grid [54] with local modifications.

Details of the up-to-date zero temperature ensembles can be found in ref. [49]. The lattice volume including the 5th extent is $32^3 \times 64 \times 12$ ($1/a = 2.45$ GeV) – $64^3 \times 128 \times 8$ (4.5 GeV). Except for the finest lattice spacing, we have several quark mass points covering a range from about 300 MeV to 500 MeV in terms of the pion mass and one ensemble with a 230 MeV pion on a $48^3 \times 96 \times 12$ ($1/a = 3.61$ GeV) lattice. Each ensemble has 50–100 configurations, separated by 50–100 molecular dynamics trajectories. Larger lattice volume and lighter quark mass ensembles are also planned.

Finite temperature ensembles are actively being generated and the status at the moment of Lattice 2022 is summarized in the slides by IK [9] (see also ref. [55, 56]). For 2-flavor ensembles, we have $N_T = 8$–14 with the aspect ratio 2–4, mainly for $T > T_c$. Each case has 1–6 different quark masses. We have in total more than 50 2-flavor ensembles and most of the ensembles have 20,000 trajectories or more and are stored every 100 trajectories with some exceptions. There are more than 70 ensembles for $2 + 1$ flavor finite temperature physics, each has 20,000 trajectories and is stored every 10 trajectories (the stored configurations will be reduced). The majority (60+) are tuned along the line of constant physics (LCP) where the quark mass is kept fixed in physical units and the temperature is varied in the range 120–205 MeV through changing the gauge coupling. The quark masses are carefully tuned on some ensembles by taking into account the effect of the residual mass of domain-wall fermions. We currently have only one LCP parameter series with $N_T = 16$ (10 ensembles) and the rest of the LCP ensembles use $N_T = 12$. The non-LCP $2 + 1$ flavor ensembles employ $N_T = 12$, 14 and 16, covering temperatures 153–204 MeV. There are 70 ensembles for 3-flavor finite temperature physics. The temperature is fixed to 180 MeV ($N_T = 8$) or 120 MeV ($N_T = 12$). Each ensemble has 300–1200 configurations, which are stored every 10 trajectories. We also have 3-flavor zero temperature ensembles with the volumes $12^3 \times 24 \times 16$ and $24^3 \times 48 \times 16$ for several quark masses.

The generated configurations are stored on the Japan Lattice Data Grid (JLDG) [57]. During the production, we also utilize the Gfarm storage system (the same as JLDG), provided by "High Performance Computing Infrastructure (HPCI)," which is accessible from major supercomputer sites in Japan.

The access policy is currently request based and we can provide configurations after the relevant publications. Every ensemble should be public in principle in the future. Details are to be determined.

2.6 ETMC (Extended Twisted Mass Collaboration)

The ETM collaboration addresses mainly zero-temperature QCD physics. This includes hadron spectroscopy, hadron structure and (heavy) flavour physics. As the collaboration name suggests, the so-called Wilson clover twisted mass formulation is employed. $O(a)$-improvement is realised by automatic improvement at maximal twist, while the clover term is used to further reduce the size of lattice artefacts. In the gauge sector, the Iwasaki gauge action is used.

The current simulation programme is two-fold: first, the collaboration produces gauge ensembles with $N_f = 2 + 1 + 1$ dynamical quark flavours, i.e. up/down, strange and charm dynamical quarks at five values of the lattice spacing ranging from 0.091 fm down to 0.049 fm. At all these lattice spacings ensembles with an (almost) physical pion mass value are available and additional ensembles with pion masses up to 350 MeV have been generated. The strange and charm quark masses are tuned to their physical values on all ensembles. In order to investigate finite volume effects, the collaboration has several parameter sets with only the volume varying, covering $M_\pi \cdot L$ values from 2.5 up to ~ 5.5. For details. see ref. [58] and for a listing of the parameter choices for most ensembles, see the appendix of ref. [59]. Second, $N_f = 4$ flavour ensembles are generated for each of the lattice spacing values mentioned above to be specifically used for the renormalisation effort of the collaboration. The overall goal of the simulation effort is to allow for a controlled continuum extrapolation at physical pion mass values. For more details, we refer to ref. [60].

Simulations are performed using the Hybrid Monte Carlo (HMC) algorithm implemented in the tmLQCD software package [61–63], which is publicly available under GPL. It implements multiple time scales and force gradient integration schemes, combined with Hasenbusch mass preconditioning for the light quarks [64]. For the strange and charm quarks a rational HMC is implemented with frequency splitting.

The DD-αAMG [65, 66] multigrid iterative solver is employed for the most poorly conditioned monomials in the light sector while mixed-precision CG is used elsewhere. In the heavy sector, multi-shift CG is used together with shift-by-shift refinement using DD-αAMG [67] for a number of smallest shifts (on machines where this is more efficient). To target SIMD architectures, tmLQCD interfaces with the QPhiX [35] library which was extended to support the necessary operators. Due to a more recent effort [68], the HMC is now also able to run on GPU machines by offloading the gauge force and iterative solves using QUDA [40, 41] (including employing QUDA's MG solver [44]). tmLQCD automatically writes gauge configurations in the ILDG format, including a header which provides some meta-data such as the creation date, the target simulation parameters, the trajectory number and the plaquette expectation value.

While the ETMC has uploaded previous sets of gauge ensembles (with $N_f = 2$ and $N_f = 2+1+1$ dynamical quark flavours, but without clover term) to ILDG storage elements — see for instance refs. [69–71], the current simulation campaign is not yet available via ILDG. However, the collaboration would immediately make use of the ILDG infrastructure as soon as a reliable service becomes available again.

The ETMC gauge ensembles are publicly available after a grace period. Thus, while all the previous gauge ensembles are available to everyone, the latest ones are currently available upon request. However, they will become public in the near future. Here, we expect to require in the order of 2.5 PB of disc space in the foreseeable future.

2.7 TWEXT Collaboration

The main goal of the Twisted Wilson @ EXTreme conditions (TWEXT) collaboration is to explore the properties of QCD at large temperatures, starting from below the chiral phase transition, up to the high temperature regime. We would like to address questions such as the QCD phase diagram, the scaling behaviour, chiral and topological properties of QCD at high temperatures and others.

We are using the Wilson twisted mass fermion discretization at maximal twist. The simulations are performed with $N_f = 2+1+1$ quarks, where we keep the parameters of the strange and charm sectors close to the physical values, and we have several ensembles for pion masses starting from the physical $m_\pi = 140$ MeV up to the heavy quark regime, $m_\pi = 370$ MeV. For the ensembles with the physical pion mass, we are using the Wilson clover-improved twisted mass action, while in the case of higher pion masses the action is a standard Wilson twisted mass action. In our simulations we are using the tuning of the parameters by the ETM collaboration [72]. The simulations are carried out in a fixed scale approach, i.e., to change the system temperature we change the lattice temporal extent N_t, keeping the lattice spacing a fixed.

Simulations are performed with the freely available tmLQCD code [73]. In total, currently we have 60 ensembles (one ensemble corresponds to one point in the temperature-pion mass-lattice spacing space), which have 50k configurations, occupying 26 TB of the disk space. A short summary of these ensembles can be found in ref. [74]. We plan to add new ensembles with physical pion mass and finer lattice spacings. Configurations are stored in the ILDG format. Ensembles are available upon request and we are open to collaboration. We plan to make TWEXT ensembles public or upload them to ILDG after some embargo time.

2.8 PACS Collaboration

There are two major problems in lattice QCD simulations. One is that it is still difficult to make a high precision measurement of the physical observables exclusively at the physical point. The other is that current lattice QCD simulations determine different physical observables by choosing different sets of gauge configurations. From a view point of predictability of lattice QCD it is highly desirable to make a high precision measurement of various physical observables from a unique set of gauge configurations. In order to overcome the above problems we perform very large scale simulations at the physical point towards master-field simulations named by Lüscher [22]. Lattice QCD simulations on very large lattices have inherent advantages: The statistical errors decrease thanks to the stochastic locality and the geometrical symmetries of the lattice, and the accessible minimum momentum is reduced in proportion to $1/L$ with the lattice extent L. We can give reliable predictions for the physical observables relevant for particle physics within and beyond the standard model.

In the past years the PACS Collaboration has been generating $2+1$ flavor QCD configurations on very large lattices at the physical point employing the stout-smeared $O(a)$-improved Wilson-clover quark action and Iwasaki gauge action. We use the stout smearing parameter $\rho = 0.1$ with six smearing iterations. The improvement coefficient for the clover term is nonperturbatively determined using the Schrödinger functional scheme. These gauge configurations, which keep the space-time volumes larger than (10 fm)^4, are called "PACS10" configurations. So far we have finished generating two gauge ensembles of (lattice spacing, lattice size)=(0.085 fm, 128^4) [75] and (0.064 fm, 160^4) [76]. We are now generating a third one of (lattice spacing, lattice size)=(0.041 fm, 256^4) at a finer lattice spacing. The degenerate up-down quarks are simulated with the domain-decomposed HMC algorithm [77] and the strange quark with the rational HMC algorithm [78]. The up-down quark determinant is separated into UV and IR parts after the even-odd preconditioning. We further apply the mass-preconditioning [31] to the IR part, which is divided into three forces of $F''_{\rm IR}$, $F'_{\rm IR}$ and $\tilde{F}_{\rm IR}$. In the end, the force terms consist of the gauge force $F_{\rm g}$, the up-down UV force $F_{\rm UV}$, the strange force $F_{\rm s}$ and the three up-down IR forces $F''_{\rm IR}$, $F'_{\rm IR}$, $\tilde{F}_{\rm IR}$. We adopt the multiple time scale integration scheme in the molecular dynamics steps according to the hierarchical structure of $||F_{\rm g}|| > ||F_{\rm UV}|| > ||F_{\rm s}|| \approx ||F''_{\rm IR}|| > ||F'_{\rm IR}|| > ||\tilde{F}_{\rm IR}||$. The trajectory length is chosen to be $\tau \geq 1.0$. We use the mixed precision nested BiCGStab [79] for the quark solver with the aid of the chronological inverter guess.

Many gauge ensembles generated by the PACS Collaboration and the predecessors, which were the CP-PACS and PACS-CS Collaborations, are publicly available. The current situation is available on a public webpage [57]. We plan to make the "PACS10" ensembles public through a new generation of ILDG after some embargo time. The details of the data policies are under discussion within the collaboration.

2.9 RBC-UKQCD Collaborations

The RIKEN-BNL Columbia (RBC) collaboration and the UKQCD collaboration, i.e., the RBC-UKQCD collaborations, began generating $2+1$ flavor ensembles with Domain Wall Fermions (DWF) in 2005, when the QCDOC machines became operational at the University of Edinburgh (a 10 Tflops computer) and Brookhaven National Lab (BNL) (2 and 10 Tflops computers). (This followed work by the RBC collaboration using DWF on quenched QCD configurations and preliminary investigations with 2 flavor DWF configurations that began in 1997.) These calculations were the first large-scale use of the recently invented Rational Hybrid Monte Carlo algorithm by Clark and Kennedy [30], which was used for the strange quark part of the evolution. Adding this to the standard HMC algorithm for two degenerate light quarks provided, and continues to provide, an exact evolution algorithm, since all of the RBC-UKQCD collaborations' ensembles include the accept/reject step at the end of each trajectory.

During the intervening years, 41 different ensembles have been produced with either DWF or Möbius DWF (MDWF), which achieves similar residual chiral symmetry breaking with a smaller value for L_s, the extent of the fifth dimension for (M)DWF. Most of the ensembles have been produced with the Iwasaki gauge action, which provides for smoother gauge fields at the lattice cutoff, for a given lattice spacing, than the Wilson action. The smoother gauge fields reduce the residual chiral symmetry breaking for a given value of L_s. The earliest ensembles have m_π in the 200 to 400 MeV range and volumes $\sim$ (2.5 fm)^3. With improvements in algorithms and computers, we have, over the last decade, produced 3 ensembles, with $1/a = 1.730$, 2.359 and 2.708 GeV, 2 + 1 flavors and physical quark masses and volumes of (5.5 fm)^3, (5.4 fm)^3 and (6.9 fm)^3, respectively [80]. (The last ensemble in this list is still being produced.) These 3 ensembles represent the only physical quark mass, large volume, 2 + 1 flavor ensembles generated with the chiral symmetries of the continuum essentially intact. They support much of the wide variety of physics observables of interest to the combined collaborations.

We also have a number of ensembles with the Iwasaki plus Dislocation Suppressing Determinant Ratio (DSDR) gauge action, refereed to as the ID gauge action, which suppress the copious topological tunneling at strong coupling. The tunneling produces large residual chiral symmetry breaking, which the ID action controls, allowing for 2 + 1 flavor simulations with physical quark masses on lattices as coarse as $1/a = 1$ GeV. We have ID+MDWF ensembles with $1/a = 1$ GeV and physical volumes of (4.8 fm)^3, (6.4 fm)^3 and (9.6 fm)^3 — an excellent place to investigate large volume physics with realistic quark masses. We have also generated ID+MDWF ensembles with G-parity boundary conditions and $1/a = 1.37$ GeV for our ongoing work to precisely determine $\mathrm{Re}(\epsilon'/\epsilon)$ [81].

The members of the RBC-UKQCD collaborations have interest in a broad range of physics topics including: precision electroweak physics in light hadrons, such as $K \to \pi\pi$ decays, ΔM_K, rare kaon decays, K_{l3}; heavy quark physics; the hadronic vacuum polarization contribution to $(g-2)_\mu$ and electromagnetic contributions to many of the above processes.

We are currently setting parameters and tuning our evolution algorithms for 2 + 1 + 1 flavor (M)DWF ensembles. We are targeting a series of ensembles with $1/a = 3$, 4 and 5 GeV and physical quark masses, using the Wilson action. The physical volumes are about (4.5 fm)^3. These will provide a set of ensembles to allow the continuum limit to be taken with physical quark masses and the good chiral symmetry properties of (M)DWF. This is a long-term goal which will require the resources of the Exascale machines that are now nearing completion.

The RBC-UKQCD collaborations make their ensembles publicly available after an initial publication of the first physics results from these ensembles. Earlier access can be possible for non-competing physics projects. Many of the ensembles are available through a Globus connection at Columbia.

2.10 HotQCD Collaboration

The HotQCD collaboration addresses physics questions that deal with the properties of strong interaction matter under extreme conditions. In particular, we are interested in universal critical behavior and the exploration of the QCD phase diagram, the QCD equation of state at zero and nonzero temperature, fluctuations of conserved charges and effective degrees of freedom and the in-medium properties of hadrons, respectively their melting.

Due to their computational advantages and the remainder of the chiral symmetry group, the HotQCD collaboration is predominantly using highly improved staggered quarks (HISQ) [27] with $2+1$ flavors of light and strange quarks at physical and lighter than physical masses. The algorithm used for the generation of the gauge ensembles is the rational hybrid Monte Carlo algorithm [30]. The integrator used for the integration of the molecular dynamical trajectories has three time scales, corresponding to the light, strange and gauge forces. For each time scale we usually use the standard leapfrog scheme, however, we have also implemented a 2^{nd} order minimum norm integrator (Omelyan) [52]. The acceptance is tuned to approximately 70% on a trajectory length in the range of 0.5-1.0 molecular dynamical time units. The code base for the calculations is now publicly available on GitHub and was named SIMULATeQCD [82, 83], which stands for *"a simple multi-GPU code for lattice QCD calculations"*. The code prints a log-file, containing the most important provenance information. The file format of the gauge configurations is by default NERSC, but the ILDG format is also supported.

Recently, the HotQCD collaboration has spent the majority of its awarded compute time on the generation of non-zero temperature lattices of size $32^3 \times 8$, $48^3 \times 12$ and $64^3 \times 16$, at physical quark masses. High statistics calculations exist at 9 values of the temperature in an interval $T \in [125, 175]$ MeV [84–86]. Aiming for the calculation of high order cumulants of conserved charges, the ensemble size increased to 1.4 mio. ($N_\tau = 8$), 400,000 ($N_\tau = 12$) and 20,000 ($N_\tau = 16$) gauge configurations per temperature value. This amounts to a total of 912 TB, 938 TB and 210 TB, respectively. Currently, the gauge configurations are stored in data projects at NERSC, ORNL, JLab and JSC. We are planning to make these gauge configurations available through ILDG 2.0 in the foreseeable future. In addition there exists $N_\tau = 8$ ensembles at lighter than physical quark masses, partly generated on lattices with large spatial extent ranging up to $N_\sigma = 56$ with a total size of 192 TB [87].

2.11 CLS Community Effort

The current Coordinated Lattice Simulations (CLS) effort started in 2013 as a continuation of an earlier $N_f = 2$ simulation programme carried out by some of the member groups. CLS generate gauge ensembles with $N_f = 2+1$ flavours of non-perturbatively order a improved Wilson fermions on a tree-level Symanzik improved gauge action. For more detail on the action, see ref. [88]. The lattice spacings at present range from 0.1 fm down to below 0.04 fm with volumes from $48 \cdot 24^3$ to $192 \cdot 96^3$ points. The main focus is to push lattice simulations further towards the continuum limit. This is particularly relevant for observables involving momenta and for heavy quark physics, however, a controlled continuum limit is also essential for a wide range of other applications. Ergodicity of the simulations for lattice spacings smaller than about 0.06 fm, where freezing of the topological charge sets in, is achieved by employing open boundary conditions in time [21].

All simulations are carried out using the GPL licensed package openQCD [14, 15], also in use by two of the other collaborations/initiatives who contributed to this article. The Hybrid Monte Carlo [50] (HMC) algorithm is employed and rational HMC (RHMC) [51] is used for the strange quark. This is combined with 2nd and 4th order Omelyan-Mryglod-Folk [52] integrators for different pseudo-fermions. Moreover, Hasenbusch frequency splitting [31] and chronological deflation acceleration of the domain decompositioned SAP-preconditioned GCR solver are employed. For details on the algorithmic choices and the implementation, see refs. [14, 88–90]. A small twisted mass term is added to stabilize the simulation. This, together with the error of the rational approximation of the strange quark determinant, is corrected for by including reweighting factors [14, 88, 91] into the expectation values. These factors are stored, together with the gauge ensembles.

An overview of the existing ensembles can be found on a public webpage [92], see also, e.g., ref. [90]. These are redundantly stored in the openQCD data format at DESY Zeuthen and at Regensburg. At present, runs exist for about 60 different combinations of the lattice spacing, quark masses and the volume, with a total of about 130000 gauge configurations, amounting to 1 PB of data. The log files include information on the algorithmic parameters, the openQCD [15] version in use, the target machine, run time and partition size, what person was responsible for the run, as well as various control measurements and check sums. The transfer of the configurations, log files and other run information to the storage elements is to a large extent handled automatically by regularly executed scripts. These also extract detailed information that is visualized on an internal overview/status web-page.

Many ensembles are available upon request. Further ensembles are available upon approval by CLS, either after some embargo time or for use in non-competing physics projects. The data policies are at present being reviewed within CLS, with the aim to make most ensembles available through a new generation of ILDG.

2.12 CLQCD Collaboration

The research interests of the CLQCD Collaboration lie in multiple areas of lattice QCD, and we are generating three different types of gauge configurations simultaneously.

The general-purpose ensembles are generated with the 2 + 1-flavor Wilson-Clover fermion action and the tadpole-improved Symanzik gauge action. Such ensembles are designed to address general topics in lattice QCD like hadron structure, the hadron spectrum and so on. The CLQCD collaboration has generated ensembles with the tree level tadpole improved clover action with a mild stout smearing on the gauge link with the smearing parameter $\rho = 0.125$. For both gauge and fermion fields, we also apply self consistent tuning for tadpole improvement factors. Based on the above setup, we generate ensembles with the lattice spacing ranging from about 0.1 fm to about 0.05 fm with volumes from $24^3 \times 72$ up to $48^3 \times 144$ points.

Ensembles on anisotropic lattices are generated for research on charmed hadrons, glueballs and exotic particles. The Wilson-Clover fermion action and the tadpole-improved Symanzik gauge action are adopted. The anisotropy ratios of fermion and gauge fields are set to 5 to have much finer temporal lattice spacings, while retaining a sufficiently large spatial box size. All the parameters are precisely tuned automatically with a set of sophisticated QScheme scripts. One feature of these ensembles is huge statistics; each ensemble contains 7000–10000 configurations. We have generated 2-flavor ensembles with $m_\pi^{min} = 349$ MeV, and 2-flavor and 2 + 1-flavor ensembles with $m_\pi \approx 200$ MeV are proposed.

The ensembles of $N_f = 2 + 1$ QCD with nonzero temperature and nonzero magnetic fields are generated using the HISQ/tree action with a modified version of the SIMULATeQCD [82, 83] code. Five different temperatures around the transition temperature and 9 different values of the magnetic

Table 1: Available gauge ensembles and data management. The numbers are only approximate. In some of the cases (e.g., MILC, CLS), these only account for the latest action used by the collaboration(s)/effort/initiative. The storage displayed for JLab/W&M/LANL/MIT/OLCF/Marseille is the total storage used, not only that for the gauge ensembles, while the storage given for ETMC includes future plans.
The second column indicates the public availability of the ensembles (0 = no, 1 = yes, after some embargo time, 2 = yes, already now) and the third column the interest in sharing these through ILDG (0 = no interest, 1 = interest, 2 = planned, 3 = already using).
#ens = number of ensembles, #cfg = total number of configurations.

collaboration	public	ILDG	#ens	#cfg	storage (TB)
FASTSUM	1	1	25	22k	40
OpenLat	1	2	8	10k	30
MILC	1	1	>25	>50k	650
JLab/W&M/LANL/MIT/OLCF/Marseille	0	0	13	105k	2000
JLQCD	1	2/3	>230	60k	20
ETMC	1	2/3	21	100k	2500
TWEXT	1	1	60	50k	26
PACS	1	2/3	3	100	60
RBC-UKQCD	1	0	41	20k	200
HotQCD	1	2	58	15M	2250
CLS	1	2	>60	130k	1000
CLQCD $T = 0$	1	1	10	5k	14
CLQCD $T > 0$	1	1	28	150k	120
HAL QCD	1	2	1	1.4k	70
QCDSF-UKQCD-CSSM	1	2/3	60	90k	300

field strength are realized. Additional ~5k configurations of $48^3 \times 12$ lattices are generated for each of the 25 parameter sets. The above-mentioned configurations occupy about 120 TB.

3. Summary

The fermionic actions used by the groups who responded to the call are staggered quarks, in particular using the HISQ action (MILC, HotQCD, CLQCD), improved Wilson quarks on isotropic (OpenLat, JLab/WM/LANL/MIT/OLCF/Marseille, PACS, CLS, CLQCD, HAL QCD, QCDSF-UKQCD-CSSM) and anisotropic lattices (FASTSUM, CLQCD), twisted mass fermions (ETMC, TWEXT) and Domain Wall fermions (JLQCD, RBC-UKQCD). Some collaborations have generated ensembles exclusively with temperatures $T > 0$ (TWEXT, HotQCD, CLQCD), others both for $T > 0$ and for $T = 0$ (FASTSUM, MILC) and the remaining ones only for $T = 0$. The total number of gauge ensembles/configurations and the storage requirements, public availability and ILDG plans are summarized in Table 1. For details on the action, simulation setup, data management strategy etc., we refer to the individual contributions in the previous section. After the session was filled, additional input was received from Takumi Doi on behalf of the HAL QCD Collaboration and from James Zanotti on behalf of QCDSF-UKQCD-CSSM. This is also included in the table.

PoS(LATTICE2022)203

The collected material is a snapshot of the activities of a large fraction of the "data providers" generating $N_f = 2 + 1$ and $N_f = 2 + 1 + 1$ QCD ensembles. We hope that this information on ongoing ensemble generation campaigns, the associated usage and access policies as well as the storage requirements is of use, both to "data providers" and to "data consumers." The intent is that this summary encourages communication, coordination and best practice within the community and beyond.

Most participants of the parallel session indicated an interest in sharing gauge configurations through ILDG, see the third column of Table 1. This may encourage and assist the efforts to modernize and extend ILDG as discussed at the ILDG lunch [9] and in a plenary contribution [8] at this conference. In helping to establish well-defined workflows and minimum quality standards for data management, ILDG can provide an important framework to support data sharing. However, acquiring storage space, where data can be held accessible (according to the envisaged access policies) and persistent for more than a decade, remains the responsibility of the regional grids and the contributing collaborations (irrespective of whether they join the ILDG effort or not). Many different regional funding agencies are involved. Therefore, this challenge — as well as supporting the necessary scientific personnel — cannot be addressed globally with one and the same strategy. However, local experiences and information can be shared and activities of the community as a whole can be coordinated.

Acknowledgments

Gunnar Bali thanks Hubert Simma for discussions and help in the organization of the session. For lack of space, we only acknowledge funding received by the co-authors and not by all the members of the respective collaborations. We also refrain from a detailed listing of the respective computer time grants, which are acknowledged in the publications cited in the text. Instead, we list the high performance computers that were used to generate the gauge ensembles. Our thanks go to all the institutions that run these machines, their user support and the funding agencies.

Gunnar Bali, Christian Schmidt and Wolfgang Söldner acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) under the grant "NFDI 39/1" (PUNCH4NFDI). Gunnar Bali and Wolfgang Söldner acknowledge support from the European Union's Horizon 2020 research and innovation programme under grant agreements 813942 (ITN EuroPLEx) and 824093 (STRONG-2020). Ryan Bignell is supported by Science and Technology Facilities Council (STFC) grant ST/000813/1. Anthony Francis is supported by the Japanese Ministry of Science and Technology Taiwan (MOST) under grant 111-2112-M-A49-018-MY2. Steven Gottlieb is supported by the U.S. Department of Energy (DOE) under Award No. DE-SC0010120. The work of Rajan Gupta is supported by the DOE High Energy Physics program (DOE HEP) under Award No. DE-AC52-06NA25396 and by the LANL LDRD program. Issaku Kanamori is partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in the "Program for Promoting Researches on the Supercomputer Fugaku" (Simulation for basic science: from fundamental laws of particles to creation of nuclei). He also acknowledges the Japanese Society for the Promotion of Science (JSPS) KAKENHI grant (JP20K03961). The "PACS10" configuration generation by the PACS Collaboration (presented here by Yoshinobu Kuramashi) is supported in part by Grants-in-Aid for Scientific Research from the MEXT (Grant No. 19H01892). Robert Mawhinney is supported in part by the DOE grant DE-SC0011941. Christian Schmidt is supported through the DFG Collaborative Research Centre SFB/TRR 211 as well as by the European Union grant 283286.

The participating collaborations have generated their gauge ensembles (list sorted alphabetically according to country, institute, computer)
on the HPC Cluster of the Institute of Theoretical Physics of the Chinese Academy of Sciences (CAS) in Beijing,
on computers at the Nuclear Science Computing Center (NSC3) at Central China Normal University (CCNU) in Wuhan,
on computers at the Southern Nuclear Science Computing Center (SNSC) at South China Normal University (SCNU) in Guangzhou,
on Siyuan Mark 1 at the Center for High Performance Computing at Shanghai Jiao Tong University (SJTU), China,
on Occigen at Centre Informatique National de l'Enseignement Supérieur (C.I.N.E.S.) in Montpellier,
on Jean Zay at Institut du Développement et des Ressources en Informatique Scientifique (IDRIS) in Orsay,
on Irène Joliot-Curie at Très Grand Centre de Calcul (TGCC) in Bruyères-le-Châtel, France,
on Goethe-HLR and Loewe-CSC at Goethe-Universität Frankfurt,
on Clover and HIMster-II at Helmholtz Institute Mainz,
on Hazel Hen and HAWK at Höchstleistungsrechenzentrum Stuttgart (HLRS),
on Mogon-II at Johannes-Gutenberg-Universität Mainz,
on JUGEEN, JUQUEEN, JURECA, JURECA-Booster, JUWELS and JUWELS-Booster at Jülich Supercomputing Centre (JSC),
on SuperMUC and SuperMUC-NG at Leibniz Rechenzentrum (LRZ) in Garching,
on Oculus, Nocutua and Nocutua2 at Paderborn Center for Parallel Computing (PC2),
on Fritz at Regionales Rechenzentrum Erlangen (RRZE),
on the Bielefeld GPU-Cluster of Universität Bielefeld,
on Bonna at Universität Bonn,
on Lise (HLRN-IV) at Zuse-Institut Berlin (ZIB),
on iDataCool and Athene 2 at Universität Regensburg, Germany,
on Stokes at the Irish Centre for High-End Computing (ICHEC) in Galway, Ireland,
on Fermi, Marconi A1, Marconi A2, Marconi A3 and Marconi 100 at CINECA in Bologna, Italy,
on Polaire and Grand Chariot at Hokkaido University in Sapporo,
on IBM Blue Gene/Q (BG/Q) computers at the High Energy Accelerator Research Organization (KEK) in Tsukuba,
on SQUID at Osaka University,
on Supercomputer Fugaku at the RIKEN Center for Computational Science (R-CCS) in Kobe,
on Oakforest-PACS at the Joint Center for Advanced High Performance Computing (JCAHPC) of the Universities of Tokyo and Tsukuba in Kashiwa,
on Wisteria at the Information Technology Center of the University of Tokyo in Kashiwa, Japan,
on Prometheus at Akademickie Centrum Komputerowe CYFRONET in Kraków,
on Okeanos at ICM Uniwersytet Warszawski, Poland,
on hpc-qcd at CERN in Geneva,

on Piz Daint at the Centro Svizzero di Calcolo Scientifico (CSCS) in Lugano, Switzerland,
on Tesseract and Tursa of the DiRAC Extreme Scaling service in Edinburgh,
on the BG/L, BG/Q and QCDOC at the University of Edinburgh,
on Sunbird of Supercomputing Wales at Swansea University, UK,
on Intrepid, Mira and Theta of the Argonne Leadership Computing Facility (ALCF) at the Argonne National Laboratory (ANL) in Lemont,
on the BG/Q, QCDOC and QCDSP of the RIKEN-BNL Research Center and the BG/L of the New York Center for Computational Sciences (NYCCS) at Brookhaven National Laboratory (BNL) in Upton,
on the QCDOC of USQCD at BNL,
on the QCDSP at Columbia University,
on the GPU and KNL Clusters of USQCD at Jefferson Lab (JLab) in Newport News and at the Fermi National Accelerator Laboratory (FNAL) in Batavia,
on Big Red 2+, Big Red 3, and Big Red 200 at Indiana University in Bloomington,
on Sequoia and Vulcan at Lawrence Livermore National Lab (LLNL) in Livermore,
on Badger, Chicoma and Grizzly at Los Alamos National Laboratory (LANL) Institutional Computing,
on Blue Waters of the National Center for Supercomputing Applications (NCSA) at the University of Illinois in Urbana-Champaign,
on Cori, Edison and Perlmutter at the National Energy Research Scientific Computing Center (NERSC) in Berkeley,
on computers of the National Center for Atmospheric Research (NCAR) in Boulder,
on Summit and Titan at Oak Ridge Leadership Class Facility (OLCF) at Oak Ridge National Laboratory (ORNL),
on computers of the National Institute for Computational Sciences (NICS) of ORNL and the University of Tennessee at ORNL,
on Frontera and Stampede 2 at Texas Advanced Computing Center (TACC) in Austin
and on RMACC Summit at the University of Colorado in Boulder, USA.

References

[1] Hadron Spectrum collaboration, M. Peardon et al., *A Novel quark-field creation operator construction for hadronic physics in lattice QCD*, *Phys. Rev. D* **80** (2009) 054506 [0905.2160].

[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 and T. Yoshie, *Building the International Lattice Data Grid*, *Comput. Phys. Commun.* **182** (2011) 1208 [0910.1692].

PoS(LATTICE2022)203

[6] M. Wilkinson, M. Dumontier, I. Aalbersberg et al., *The FAIR Guiding Principles for scientific data management and stewardship*, Sci Data **3** (2016) 160018.

[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] F. Karsch, H. Simma and T. Yoshie, *The International Lattice Data Grid – towards FAIR Data*, 2212.08392.

[9] S. Gottlieb et al., "Parallel session at Lattice 2021: Lattice Data." https://indico.hiskp.uni-bonn.de/event/40/sessions/98/#20220809.

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

[11] 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].

[12] 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].

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

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

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

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

[17] 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].

[18] European Organization for Nuclear Research and OpenAIRE, "Zenodo." https://www.zenodo.org/.

[19] A. Francis, P. Fritzsch, M. Lüscher and A. Rago, *Master-field simulations of O(a)-improved lattice QCD: Algorithms, stability and exactness*, Comput. Phys. Commun. **255** (2020) 107355 [1911.04533].

[20] A. S. Francis, F. Cuteri, P. Fritzsch, G. Pederiva, A. Rago, A. Shindler, A. Walker-Loud and S. Zafeiropoulos, *Properties, ensembles and hadron spectra with Stabilised Wilson Fermions*, PoS **LATTICE2021** (2022) 118 [2201.03874].

[21] M. Lüscher and S. Schaefer, *Lattice QCD without topology barriers*, JHEP **07** (2011) 036 [1105.4749].

[22] M. Lüscher, *Stochastic locality and master-field simulations of very large lattices*, EPJ Web Conf. **175** (2018) 01002 [1707.09758].

[23] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD*, JHEP **08** (2010) 071 [1006.4518], [Erratum: JHEP 03, 092 (2014)].

[24] M. Bruno, T. Korzec and S. Schaefer, *Setting the scale for the CLS* $2+1$ *flavor ensembles*, Phys. Rev. D **95** (2017) 074504 [1608.08900].

[25] F. Cuteri, A. Francis, P. Fritzsch, G. Pederiva, A. Rago, A. Shindler, A. Walker-Loud and S. Zafeiropoulos, *Gauge generation and dissemination in OpenLat*, 2212.07314.

[26] "Openlat online repository and point of reference." https://openlat1.gitlab.io.

[27] HPQCD, UKQCD collaboration, E. Follana et al., *Highly improved staggered quarks on the lattice, with applications to charm physics*, Phys. Rev. D **75** (2007) 054502 [hep-lat/0610092].

[28] MILC collaboration, A. Bazavov et al., *Nonperturbative QCD simulations with* $2 + 1$ *flavors of improved staggered quarks*, Rev. Mod. Phys. **82** (2010) 1349 [0903.3598].

[29] N. Miller et al., *Scale setting the Möbius domain wall fermion on gradient-flowed HISQ action using the omega baryon mass and the gradient-flow scales* t_0 *and* w_0, Phys. Rev. D **103** (2021) 054511 [2011.12166].

[30] M. A. Clark and A. D. Kennedy, *The RHMC algorithm for two flavors of dynamical staggered fermions*, Nucl. Phys. B Proc. Suppl. **129** (2004) 850 [hep-lat/0309084].

[31] M. Hasenbusch, *Speeding up the hybrid Monte Carlo algorithm for dynamical fermions*, Phys. Lett. B **519** (2001) 177 [hep-lat/0107019].

[32] MILC collaboration, A. Bazavov et al., *Scaling studies of QCD with the dynamical HISQ action*, Phys. Rev. D **82** (2010) 074501 [1004.0342].

[33] MILC collaboration, A. Bazavov et al., *Lattice QCD Ensembles with four flavors of highly improved staggered quarks*, Phys. Rev. D **87** (2013) 054505 [1212.4768].

[34] "USQCD software page." https://www.usqcd.org/software.html.

[35] B. Joó, D. D. Kalamkar, K. Vaidyanathan, M. Smelyanskiy, K. Pamnany, V. W. Lee, P. Dubey and W. Watson, *Lattice QCD on Intel® Xeon Phi coprocessors*, Lect. Notes Comput. Sci. **7905** (2013) 40.

[36] R. Li and S. Gottlieb, *Staggered Dslash performance on Intel Xeon Phi architecture*, PoS **LATTICE2014** (2015) 034 [1411.2087].

[37] C. DeTar, D. Doerfler, S. Gottlieb, A. Jha, D. Kalamkar, R. Li and D. Toussaint, *MILC staggered conjugate gradient performance on Intel KNL*, PoS **LATTICE2016** (2016) 270 [1611.00728].

[38] C. DeTar, S. Gottlieb, R. Li and D. Toussaint, *MILC code performance on high end CPU and GPU supercomputer clusters*, EPJ Web Conf. **175** (2018) 02009.

[39] K. Barros, R. Babich, R. Brower, M. A. Clark and C. Rebbi, *Blasting through lattice calculations using CUDA*, PoS **LATTICE2008** (2008) 045 [0810.5365].

[40] 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].

[41] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower and S. Gottlieb, *Scaling Lattice QCD beyond 100 GPUs*, in *SC11: International Conference for High Performance Computing, Networking, Storage and Analysis*, 9, 2011, 1109.2935, DOI.

[42] M. A. Clark and R. Babich, "QUDA: A library for QCD on GPUs." http://lattice.github.io/quda/.

[43] "MILC wiki page." https://github.com/milc-qcd/sharing/wiki/LatticeSharing.

[44] M. A. Clark, B. Joó, A. Strelchenko, M. Cheng, A. Gambhir and R. Brower, *Accelerating lattice QCD multigrid on GPUs using fine-grained parallelization*, in *SC16: Proceedings of*

the International Conference for High Performance Computing, Networking, Storage and Analysis, p. 795, 12, 2016, 1612.07873, DOI.
[45] R. G. Edwards and B. Joó, "The Chroma Software System for LatticeQCD." http://github.com/jeffersonlab/chroma.
[46] F. T. Winter, M. A. Clark, R. G. Edwards and B. Joó, *A framework for lattice QCD calculations on GPUs*, in *28th IEEE International Parallel and Distributed Processing Symposium*, 8, 2014, 1408.5925, DOI.
[47] F. Winter, "QDP-JIT Download." http://github.com/jeffersonlab/qdp-jit.
[48] R. C. Brower, H. Neff and K. Orginos, *The Möbius domain wall fermion algorithm*, *Comput. Phys. Commun.* **220** (2017) 1 [1206.5214].
[49] JLQCD collaboration, B. Colquhoun, S. Hashimoto, T. Kaneko and J. Koponen, *Form factors of $B \to \pi \ell \nu$ and a determination of $|V_{ub}|$ with Möbius domain-wall fermions*, *Phys. Rev. D* **106** (2022) 054502 [2203.04938].
[50] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.
[51] A. D. Kennedy, I. Horvath and S. Sint, *A new exact method for dynamical fermion computations with nonlocal actions*, *Nucl. Phys. B Proc. Suppl.* **73** (1999) 834 [hep-lat/9809092].
[52] I. Omelyan, I. Mryglod and R. Folk, *Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations*, *Comput. Phys. Commun.* **151** (2003) 272.
[53] G. Cossu, J. Noaki, S. Hashimoto, T. Kaneko, H. Fukaya, P. A. Boyle and J. Doi, *JLQCD IroIro++ lattice code on BG/Q*, *PoS* **Lattice 2013** (2014) [1311.0084].
[54] P. Boyle, A. Yamaguchi, G. Cossu and A. Portelli, *Grid: A next generation data parallel C++ QCD library*, 1512.03487.
[55] JLQCD collaboration, A. Tomiya, G. Cossu, S. Aoki, . Fukaya, S. Hashimoto, T. Kaneko and J. Noaki, *Evidence of effective axial $U(1)$ symmetry restoration at high temperature QCD*, *Phys. Rev. D* **96** (2017) 034509 [1612.01908], [Addendum: Phys.Rev.D 96, 079902 (2017)].
[56] JLQCD collaboration, S. Aoki, Y. Aoki, G. Cossu, H. Fukaya, S. Hashimoto, T. Kaneko, C. Rohrhofer and K. Suzuki, *Study of the axial $U(1)$ anomaly at high temperature with lattice chiral fermions*, *Phys. Rev. D* **103** (2021) 074506 [2011.01499].
[57] "Japan Lattice Data Grid." https://www.jldg.org/.
[58] C. Alexandrou et al., *Simulating twisted mass fermions at physical light, strange and charm quark masses*, *Phys. Rev. D* **98** (2018) 054518 [1807.00495].
[59] ETM collaboration, C. Alexandrou et al., *Ratio of kaon and pion leptonic decay constants with $N_f = 2 + 1 + 1$ Wilson-clover twisted-mass fermions*, *Phys. Rev. D* **104** (2021) 074520 [2104.06747].
[60] ETM collaboration, C. Alexandrou et al., *Quark masses using twisted-mass fermion gauge ensembles*, *Phys. Rev. D* **104** (2021) 074515 [2104.13408].
[61] K. Jansen and C. Urbach, *tmLQCD: A Program suite to simulate Wilson twisted mass lattice QCD*, *Comput. Phys. Commun.* **180** (2009) 2717 [0905.3331].
[62] A. Deuzeman, K. Jansen, B. Kostrzewa and C. Urbach, *Experiences with OpenMP in tmLQCD*, *PoS* **LATTICE2013** (2014) 416 [1311.4521].

PoS(LATTICE2022)203

[63] A. Abdel-Rehim, F. Burger, A. Deuzeman, K. Jansen, B. Kostrzewa, L. Scorzato and C. Urbach, *Recent developments in the tmLQCD software suite*, *PoS* **LATTICE2013** (2014) 414 [1311.5495].

[64] C. Urbach, K. Jansen, A. Shindler and U. Wenger, *HMC algorithm with multiple time scale integration and mass preconditioning*, *Comput. Phys. Commun.* **174** (2006) 87 [hep-lat/0506011].

[65] 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].

[66] C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl and M. Rottmann, *Adaptive aggregation-based domain decomposition multigrid for twisted mass fermions*, *Phys. Rev. D* **94** (2016) 114509 [1610.02370].

[67] 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].

[68] B. Kostrzewa, S. Bacchio, J. Finkenrath, M. Garofalo, F. Pittler, S. Romiti and C. Urbach, *Twisted mass ensemble generation on GPU machines*, *PoS* **LATTICE2022** (2022) 340 [2212.06635].

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

[70] ETM 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].

[71] ETM collaboration, R. Baron et al., *Light meson physics from maximally twisted mass lattice QCD*, *JHEP* **08** (2010) 097 [0911.5061].

[72] C. Alexandrou et al., *Nucleon axial and pseudoscalar form factors from lattice QCD at the physical point*, *Phys. Rev. D* **103** (2021) 034509 [2011.13342].

[73] "tmLQCD code." https://github.com/etmc/tmLQCD.

[74] 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].

[75] PACS collaboration, K.-I. Ishikawa, N. Ishizuka, Y. Kuramashi, Y. Nakamura, Y. Namekawa, Y. Taniguchi, N. Ukita, T. Yamazaki and T. Yoshie, *Finite size effect on pseudoscalar meson sector in $2 + 1$ flavor QCD at the physical point*, *Phys. Rev. D* **99** (2019) 014504 [1807.06237].

[76] PACS collaboration, E. Shintani and Y. Kuramashi, *Hadronic vacuum polarization contribution to the muon $g - 2$ with $2 + 1$ flavor lattice QCD on a larger than (10 fm)^4 lattice at the physical point*, *Phys. Rev. D* **100** (2019) 034517 [1902.00885].

[77] M. Lüscher, *Lattice QCD and the Schwarz alternating procedure*, *JHEP* **05** (2003) 052 [hep-lat/0304007].

[78] 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** (2007) 051601 [hep-lat/0608015].

[79] PACS-CS collaboration, S. Aoki et al., $2+1$ *Flavor lattice QCD toward the physical point*, *Phys. Rev. D* **79** (2009) 034503 [0807.1661].

[80] RBC-UKQCD collaboration, T. Blum et al., *Domain wall QCD with physical quark masses*, *Phys. Rev. D* **93** (2016) 074505 [1411.7017].

[81] RBC-UKQCD collaboration, R. Abbott et al., *Direct CP violation and the $\Delta I = 1/2$ rule in $K \to \pi\pi$ decay from the standard model*, *Phys. Rev. D* **102** (2020) 054509 [2004.09440].

[82] "SIMULATeQCD public code repository." https://github.com/LatticeQCD/SIMULATeQCD.

[83] D. Bollweg, L. Altenkort, D. A. Clarke, O. Kaczmarek, L. Mazur, C. Schmidt, P. Scior and H.-T. Shu, *HotQCD on multi-GPU Systems*, *PoS* **LATTICE2021** (2022) 196 [2111.10354].

[84] HotQCD collaboration, A. Bazavov et al., *Skewness, kurtosis, and the fifth and sixth order cumulants of net baryon-number distributions from lattice QCD confront high-statistics STAR data*, *Phys. Rev. D* **101** (2020) 074502 [2001.08530].

[85] HotQCD collaboration, D. Bollweg et al., *Second order cumulants of conserved charge fluctuations revisited: Vanishing chemical potentials*, *Phys. Rev. D* **104** (2021) [2107.10011].

[86] HotQCD collaboration, D. Bollweg et al., *Taylor expansions and Padé approximants for cumulants of conserved charge fluctuations at nonvanishing chemical potentials*, *Phys. Rev. D* **105** (2022) 074511 [2202.09184].

[87] HotQCD collaboration, H. T. Ding et al., *Chiral Phase Transition Temperature in* $2+1$*-Flavor QCD*, *Phys. Rev. Lett.* **123** (2019) 062002 [1903.04801].

[88] CLS collaboration, M. Bruno et al., *Simulation of QCD with $N_f = 2 + 1$ flavors of non-perturbatively improved Wilson fermions*, *JHEP* **02** (2015) 043 [1411.3982].

[89] D. Mohler, S. Schaefer and J. Simeth, *CLS* $2+1$ *flavor simulations at physical light- and strange-quark masses*, *EPJ Web Conf.* **175** (2018) 02010 [1712.04884].

[90] RQCD collaboration, G. S. Bali et al., *Scale setting and the light baryon spectrum in $N_f = 2+1$ QCD with Wilson fermions*, 2211.03744.

[91] D. Mohler and S. Schaefer, *Remarks on strange-quark simulations with Wilson fermions*, *Phys. Rev. D* **102** (2020) 074506 [2003.13359].

[92] "Status of CLS configurations for $N_f = 2 + 1$ flavours." https://www-zeuthen.desy.de/alpha/public-cls-nf21/.