Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 286 (2026) 24–33

Procedia Computer Science

www.elsevier.com/locate/procedia

Proceedings of the Fourth EuroHPC user day

Future Requirements of Lattice Field Theory Calculations on European High-Performance Computing Facilities

Gert Aarts[a], Gunnar Bali[b], Jacob Finkenrath[c], Stefan Krieg[d,e], Antonio Rago[f,*], Carsten Urbach[e], Hartmut Wittig[g]

[a] *Centre for Quantum Fields and Gravity, Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom*
[b] *Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany*
[c] *Department of Physics, Bergische Universität Wuppertal, 42119 Wuppertal, Germany*
[d] *Jülich Supercomputing Centre (JSC), Center for Advanced Simulation and Analytics (CASA), Forschungszentrum Jülich, 54245 Jülich, Germany*
[e] *Helmholtz-Institut für Strahlen- und Kernphysik, Rheinische Friedrich-Wilhelms-Universität Bonn, Nussallee 14-16, 53115 Bonn, Germany*
[f] *Quantum Theory Center (QTC), IMADA, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*
[g] *PRISMA++ Cluster of Excellence, Institut für Kernphysik and Helmholtz-Institut Mainz, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*

Abstract

Lattice field theory provides a first-principles framework for studying properties of strongly interacting quantum field theories in elementary particle physics. Researchers in lattice field theory are also among the largest and most efficient users of high-performance computing resources in fundamental science. In this contribution, we outline the computational profile of lattice QCD, from gauge-field generation to large-scale measurements, and discuss the main hardware, software, and human resource requirements needed to sustain progress on current and future European HPC infrastructures.

© 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0)
Peer-review under responsibility of the scientific committee of the Proceedings of the fourth EuroHPC User Days 2026

Keywords: Lattice field theory, QCD, Computational cost, Efficiency, HPC, Exascale

1. Introduction

The Standard Model of particle physics represents one of the most successful theoretical frameworks in modern science. Major international experiments carried out over several decades have established with remarkable precision the particle content and interactions of the electroweak and strong sectors. Measurements at facilities such as LEP, the Tevatron, and the Large Hadron Collider have confirmed the gauge structure of the Standard Model and determined many of its parameters with high accuracy.

Within this framework, the strong interaction described by Quantum Chromodynamics (QCD) plays a central role. QCD governs the dynamics of quarks and gluons and is responsible for the formation of hadrons, including baryons, the particles that make up most of the visible matter in the universe. A key feature of this theory is that the interaction becomes stronger at low energies. At the scale relevant for hadrons, the theory is therefore strongly coupled, meaning that standard perturbative techniques are no longer applicable. Understanding the properties of strongly interacting matter consequently requires genuinely non-perturbative approaches.

Lattice field theory provides such a framework. In this approach space and time are discretised on a finite grid, allowing the quantum field theory to be formulated in a way that can be evaluated numerically. Physical observables are

*Corresponding author. Tel.: +4565509258
Email address: rago@qtc.sdu.dk (Antonio Rago)

1877-0509 © 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0)
Peer-review under responsibility of the scientific committee of the Proceedings of the fourth EuroHPC User Days 2026
10.1016/j.procs.2026.08.016

computed by sampling gauge field configurations using Monte Carlo methods. This formulation enables calculations directly from the fundamental theory with systematically improvable uncertainties. Although originally developed for QCD, the method can be applied to any quantum field theory that admits a Euclidean path integral formulation and therefore has become a powerful tool for studying any strongly interacting field theory.

Lattice Quantum Chromodynamics represents the most mature realisation of this approach. Over the past decades lattice QCD has produced a number of major achievements. One of the earliest successes was the quantitative demonstration of colour confinement through the computation of the potential between a heavy quark and an antiquark, showing the emergence of a constant attractive force at large distances. Lattice calculations have also enabled precise determinations of the hadron spectrum, reproducing the masses of light hadrons with percent-level agreement with experimental measurements.

In recent years, lattice QCD has become an essential tool for determining fundamental parameters of the Standard Model, including the quark masses and the strong coupling constant. Its scope has expanded to increasingly complex observables, such as nucleon structure, hadron scattering amplitudes, and electroweak decay matrix elements. Lattice calculations also provide crucial theoretical input for precision tests of the Standard Model. A notable example is the determination of the hadronic contributions to the anomalous magnetic moment of the muon $(g-2)$, a quantity that showed a tension between experimental measurements and theoretical predictions [1]. Improving the precision of hadronic contributions is therefore essential for interpreting this discrepancy and assessing possible indications of physics beyond the Standard Model (BSM) [2]. Including temperature in simulations is straightforward and yields results relevant for the early universe and for relativistic heavy-ion collisions, e.g., at the Large Hadron Collider [3]. The community has embraced this challenge in line with the priorities identified by the European Strategy for Particle Physics[1] and the Nuclear Physics Collaboration Committee (NuPECC) Long Range Plan[2].

The challenges encountered are not limited to QCD. Many extensions of the Standard Model predict additional strongly interacting sectors, often motivated by attempts to explain phenomena such as electroweak symmetry breaking, dark matter, or the hierarchy problem. These BSM theories frequently involve new gauge dynamics whose low-energy behaviour is inherently non-perturbative. In summary, lattice field theory has become a mature computational discipline for obtaining quantitative predictions for strongly interacting systems directly from the underlying quantum field theory, while also providing a versatile first-principles framework for interpreting present experimental data and exploring possible new physics scenarios beyond the Standard Model.

1.1. Lattice formulation

The lattice formulation replaces continuous space–time with a discrete grid characterised by a lattice spacing a and a finite physical extent L. The lattice spacing acts as an ultraviolet regulator of the theory, while the finite volume provides an infrared cutoff.

Physical observables are computed as expectation values of operators defined on the lattice. In practice, these expectation values are estimated by averaging over ensembles of gauge-field configurations. Each configuration can be interpreted as a snapshot of the quantum fields over the entire space–time lattice. Since quantum fluctuations are intrinsic to the theory, a single configuration does not carry direct physical meaning. Instead, reliable predictions emerge only after sampling a sufficiently large number of configurations distributed according to the probability measure defined by the lattice action.

Markov-chain Monte Carlo algorithms produce a sequence of such field configurations. Observables are then evaluated on each configuration ("measurements") and averaged over the ensemble. The statistical uncertainty associated with this procedure decreases as more configurations are included, allowing one to assign predictive errors to the final results. Controlling these statistical uncertainties requires careful treatment of correlations between configurations and different observables on a given ensemble.

In addition to statistical uncertainties, lattice calculations must address systematic effects introduced by the discretisation of space–time. The physical theory corresponds to the continuum limit, which is obtained by extrapolating results computed at several lattice spacings towards $a \to 0$, a procedure called "continuum limit". Achieving control over this extrapolation is challenging and requires simulations at multiple lattice spacings together with careful analysis of discretisation effects. At the same time, the lattice volume must be large enough to suppress finite-volume

[1] https://europeanstrategy.cern/
[2] https://www.nupecc.org/

artefacts, and the parameters of the theory must be tuned to reproduce some observables of reference.

Reducing the lattice spacing, increasing the lattice volume, and approaching physical quark masses all lead to rapidly increasing computational costs. As a result, lattice QCD simulations require extremely large computational resources and have historically been among the most demanding applications in high-performance computing.

1.2. Lattice field theory community

A key feature of lattice field theory is that physical results, once extrapolated to the continuum limit, do not depend on the specific lattice formulation used. Different discretisations of the theory may differ at finite lattice spacing, but they all converge to the same continuum physics. This property allows for independent calculations of the same observables using different lattice formulations. In practice, this provides an important and, in fact, scientifically necessary form of cross-validation of the method: results obtained by different groups using different discretisations can be compared and tested against each other. In this respect, lattice QCD resembles large experimental programmes such as those at CERN, where multiple experiments independently measure the same quantities using different detectors to ensure robustness and reliability of the results.

The choice of a particular lattice formulation often shapes the organisation of the research collaborations that carry out the simulations. Over time, communities have formed around specific discretisations of the theory, together with the software frameworks and algorithmic strategies required to implement them efficiently. These collaborations typically involve multiple research institutions and national computing infrastructures, pooling expertise in physics, numerical algorithms, and high-performance computing.

The lattice field theory community, therefore, operates through large collaborative projects that generate and analyse ensembles of gauge-field configurations. Producing such ensembles requires extensive and sustained access to high-performance computing resources and is frequently coordinated across several computing centres. Once generated, these configurations become long-lived scientific assets within an open-science framework, shared across collaborations, and reused by multiple groups for diverse physics analyses, retaining and even increasing their value over more than a decade.

Because the same ensembles support many independent studies, the community has historically developed a strongly interconnected structure. This model maximises the scientific return of large computational investments and has fostered a culture of data sharing, common standards, and joint development of software tools that underpin modern lattice field theory research. It also enables the community to scrutinise results obtained with different lattice formulations and to provide validated input to the wider community, in particular flavour physics, where lattice QCD supplies the non-perturbative hadronic quantities needed to interpret precision measurements from LHCb, Belle II, BESIII, kaon experiments, and related programmes. Decay constants, form factors, bag parameters, and mixing matrix elements enter directly into CKM phenomenology, unitarity tests, and searches for physics beyond the Standard Model. The Flavour Lattice Averaging Group (FLAG) provides a central example of this community-wide validation process: it reviews lattice results, applies agreed quality criteria, and provides reference averages or estimates for the wider particle-physics community [4, 5, 6].

1.3. EuroLFT

The European lattice field theory community has recently established the EuroLFT initiative (https://eurolft.github.io/) as a platform to represent the interests and needs of the European lattice community towards European institutions. EuroLFT acts as a point of reference for institutions seeking to engage with the lattice field theory community and facilitates dialogue on topics relevant to research, computing infrastructure, and long-term scientific strategy. Similar initiatives exist in other regions of the world; for example, in the United States the lattice community is organised through the USQCD collaboration, which coordinates community activities and interactions with national laboratories and funding agencies (https://www.usqcd.org).

EuroLFT aims to strengthen connections within the European lattice community and to improve interactions between research groups, computing centres, and policy bodies. By serving as a representative forum for the community, it articulates the computational requirements of the field, provides a collective perspective on future developments in computing architectures, and ensures that the specific needs of lattice field theory are visible in broader strategic discussions concerning high-performance computing and scientific data infrastructures in Europe.

1.4. International Lattice Data Grid

The lattice field theory community has historically embraced the principles of open science. A central element of this approach is the sharing of gauge-field configurations, which represent a major fraction of the consumed computer time of large-scale lattice simulations. Once generated, these configurations can be reused by many research groups to study a wide range of physical observables. This practice greatly amplifies the scientific return of large computational investments and is consistent with the culture of collaboration and data sharing across the field.

To support this model, the community established the International Lattice Data Grid (ILDG) [7, 8, 9, 10], a distributed service to share gauge-field ensembles produced by lattice collaborations worldwide. The ILDG connects multiple regional storage nodes through a common metadata catalogue, enabling researchers to discover, access, and reuse configurations generated on major high-performance computing systems. The ILDG represented, and still represents, an implementation of FAIR principles [11] for the precious lattice gauge-field ensembles.

More recently, the European part of the community has launched the ILDG 2.0 initiative [12], aimed at revitalising this infrastructure by adapting it to modern data management practices and evolving computing environments. This effort seeks to improve interoperability with contemporary storage technologies and data services while preserving the core mission of enabling open access to lattice simulation data. In addition, ILDG 2.0 strives to extend its service to additional data types.

2. Computational profile

The computational workflow of lattice QCD spans several stages, each characterised by different computational patterns. At its core lies the generation of gauge-field configurations using large-scale Monte Carlo simulations. These configurations form ensembles that are subsequently used for a wide range of physics studies.

Once generated, ensembles often remain scientifically valuable for many years. New analysis techniques are continuously applied to existing configurations and novel observables measured, meaning that the computational investment in ensemble generation continues to produce results over long timescales.

2.1. Typical workflows

Lattice workflows can involve a wide variety of computational tasks and analysis steps, often combining several software frameworks and numerical techniques. In practice, the full scientific pipeline may include configuration generation, “measurements”, and increasingly sophisticated analysis procedures. In this section we provide a simplified overview of these workflows, focusing on the main components that dominate the overall computational cost.

Configuration generation. Gauge configurations are generated using the Hybrid Monte Carlo (HMC) algorithm [13] and related methods that sample the probability distribution defined by the lattice action as a Markov chain. In practice, the simulation evolves a gauge field configuration through a sequence of molecular-dynamics trajectories, with a Metropolis accept/reject step after each individual trajectory. The size of the lattice is determined by two main physical requirements. The lattice spacing a must be small enough to control discretisation effects, while the physical volume $\propto L^4$ must be large enough to accommodate the relevant hadronic states without significant finite-volume distortions. These constraints are closely related to the characteristic physical scales that must be resolved in the simulation. At long distances, the box size must be large enough to accommodate the lightest hadronic state in the theory, the pion. At the same time, the lattice spacing must be small enough to resolve the shortest physical scales relevant for the observables under investigation, which are typically set by the heaviest states or largest momenta entering the problem.

For present-day calculations with close to physical quark masses, this typically leads to lattice spacings in the range $a \sim 0.04$–0.1 fm and spatial volumes with linear sizes of several fermis (1 fm = 10^{-15} m). As a result, typical lattice grids used in modern simulations contain between 10^7 and 10^9 lattice sites, corresponding, for example, to lattices of size $64^3 \times 128$, $96^3 \times 192$, or $128^3 \times 256$.

The computational cost of this procedure is dominated by repeatedly solving large sparse linear systems involving the lattice Dirac operator, which describes the propagation of quarks in the background gauge field. These linear systems typically involve matrices with dimensions proportional to the number of lattice sites multiplied by internal spin and colour degrees of freedom. For the lattice volumes quoted above this amounts to order 10^{10} degrees of freedom in the linear system, or more.

In addition, as the lattice spacing is decreased, autocorrelation times grow rapidly [14, 15] and increasingly long Markov chains need to be generated in order to obtain a sufficiently large number of statistically independent configu-

rations. This effect, combined with large lattice volumes required to control finite-volume effects, makes configuration generation one of the most computationally demanding components of lattice QCD simulations and of workloads in high-performance computing in general.

While the discussion above provides a useful estimate of the computational cost of lattice QCD simulations, it represents only a simplified picture. Over the past decades the lattice community has continuously worked to overcome these limitations through the development of improved numerical algorithms alongside advances in computing hardware.

Key developments include refinements of the Hybrid Monte Carlo algorithm [13], such as mass-preconditioning [16], domain-decomposition techniques [17], deflation methods [18], and, more recently, multigrid solvers for the lattice Dirac operator [19]. These innovations have significantly reduced the cost of the most demanding parts of the calculation and have enabled simulations at or near the physical quark masses, at small lattice spacings.

Measurements of observables. Once gauge-field configurations have been generated, physical observables are extracted through "measurement calculations" performed on these ensembles. Most observables in lattice QCD are obtained from correlation functions constructed from quark and gluon fields. In practice, computing these correlation functions requires solving additional linear systems involving the Dirac operator, similar to those encountered during configuration generation. These measurements typically require many independent inversions of the Dirac operator for each configuration. To improve statistical precision, the calculations are repeated for multiple source locations on the lattice and often for several operator structures.

Observables of interest such as hadron masses, matrix elements, or form factors are extracted from correlation functions at large Euclidean times. A central challenge in these calculations is the signal-to-noise problem. For many observables, particularly those involving baryons or multi-hadron systems, the signal contained in correlation functions decreases exponentially with the Euclidean time separation, while the statistical noise decreases much more slowly. As a result, the ratio between signal and noise deteriorates exponentially at large time separations, making it increasingly difficult to extract the desired physical information.

Controlling this effect requires ensembles of large numbers of configurations and a significant number of measurements per configuration. Various strategies are used to mitigate the problem, including improved operator constructions, variance-reduction techniques, and stochastic estimation methods. Nevertheless, signal-to-noise degradation remains a fundamental limitation that often drives the overall computational cost of many lattice QCD studies.

Machine-learning assisted workflows. Recent developments have introduced machine-learning techniques into several aspects of lattice QCD workflows. These approaches are being explored for tasks along the entire pipeline of a lattice QCD simulation: from the tuning of the action parameters, via the generation of ensembles and the computation of observables, to the data analysis [20]. As an example, it is now possible to machine-learn a renormalisation–group-improved lattice gauge action ("perfect action") [21] via the use of *lattice gauge equivariant convolutional neural networks* [22]. Variance or noise reduction in correlation functions can be achieved via the use of (automated) differentiation of lower n-point functions [23, 24].

A particularly active area of research concerns the use of generative models to improve sampling in Monte Carlo simulations. In this context, methods in generative AI are being adopted as a way to learn transformations that map simple probability distributions to the highly structured distributions describing lattice field configurations. A distinction has to be drawn between approaches that learn from the action, such as normalising flow [25], and those that learn from data, such as diffusion models [26], and hence rely on large-scale ensembles generated using standard (HMC) approaches. Once trained, such models can potentially generate candidate configurations with reduced autocorrelations, or provide improved proposal distributions for Monte Carlo algorithms. These approaches may help mitigate some of the algorithmic challenges associated with critical slowing down, which increasingly affect simulations at fine lattice spacings. While there are some attempts to go to four dimensions [27], so far the emphasis is on lower-dimensional models. Importantly, there are deep links between methods proposed in the machine-learning literature and those developed in lattice field theory, which provides a fruitful path for future interaction [28, 29]. A strategic roadmap, based on a survey across particle and nuclear physics, is now available [30].

Machine-learning techniques are also being explored as surrogate models capable of approximating expensive components of lattice calculations, thereby reducing the number of costly solver applications required in certain analysis tasks. While these approaches remain at an exploratory stage and do not yet replace the established numerical algorithms used in lattice simulations, they offer promising avenues for improving efficiency and scalability. As modern HPC architectures increasingly integrate AI-oriented hardware and software ecosystems, machine-learning-

assisted methods may become an important complement to traditional lattice QCD workflows.

2.2. Main hardware bottlenecks

Modern lattice QCD codes are dominated by repeated applications of sparse matrices, such as the Dirac operator, within iterative linear solvers. While the computational effort is perfectly balanced over the space-time grid, the corresponding kernels have relatively low arithmetic intensity and are, therefore, often limited by memory bandwidth and communication rather than peak floating-point performance. Modern multigrid and deflated solvers further amplify this effect, as key components of these algorithms reduce the local arithmetic workload and make performance increasingly dependent on low-latency, high-bandwidth network communication. Thus, high-bandwidth, low-latency interconnects and large local CPU/GPU memory are key requirements, especially for measurement applications where solutions of the Dirac equation are reused extensively.

At the same time, increasing floating-point precision is becoming an important requirement. Although mixed-precision algorithms are often used in inner iterations of solvers to maximise performance, many observables require high numerical accuracy and strict control of rounding errors. This requirement becomes more pronounced as lattice volumes grow, where maintaining numerical stability over large configuration volumes becomes essential. In this context, current trends in GPU hardware, which increasingly prioritise lower-precision arithmetic to maximise throughput, could be suboptimal for lattice workloads, where sustained and reliable double-precision performance remains critical. A potential solution may be the use of emulated high-precision arithmetic. This would, however, require a reliable and verified implementation of the emulation for simulation results to be trustworthy.

2.3. Evolution of computational effort

Over the past decades the computational scale of lattice QCD simulations has grown dramatically. Early simulations were limited to comparatively small lattice volumes, using computing resources that were state of the art at the time but modest by present-day standards. Modern simulations routinely employ volumes exceeding $64^3 \times 192$ with near-physical quark masses [31, 32] and require large statistics [33]. A non-exhaustive list of large collaborations based in Europe, or with a significant European component, includes CLS, CSSM/QCDSF/UKQCD, ETMC, FASTSUM, HotQCD, openLAT, RBC/UKQCD, RC*, TELOS, and TWEXT. These collaborations coordinate ensemble production campaigns across multiple HPC systems and the subsequent sharing of the resulting ensembles.

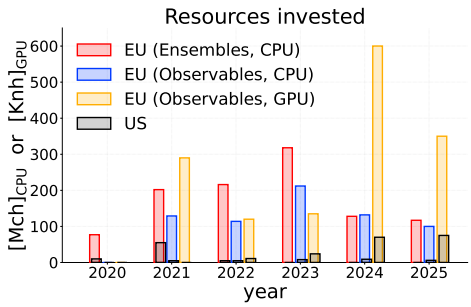

Figure 1: Evolution of computational resources invested in the OpenLAT collaboration, separated into configuration generation and observable measurements.

As an illustrative example, fig. 1 shows the evolution of computational resources within the OpenLAT collaboration, one of the youngest large-scale initiatives. Although this is a single case, it reflects a general pattern: the upfront cost of configuration generation supports a sustained physics programme through community reuse. In addition, a constant effort to develop novel algorithms and new software kernels is required to utilise state-of-the-art machines [34, 35]. An example for this ongoing effort is the development of `tmLQCD`, see fig. 2a, which could substantially speed up simulations using multi-grid solvers [36], as well as lately offloading solver and force computations to the `QUDA` library [37, 32]. Another example is the recent porting of `HiRep` to GPU architectures [38], which demonstrates excellent strong-scaling performance, see fig. 2b.

3. Main requirements for the future

3.1. Core hardware requirements for lattice investigations

The performance of lattice applications is primarily determined by a combination of memory bandwidth, communication efficiency, and the ability to sustain performance across large parallel systems. The dominant computational kernel, namely the application of the Dirac operator within iterative solvers, is inherently bandwidth-bound. This situation has become even more pronounced with the adoption of novel algorithms, which reduce the overall floating-point workload while increasing pressure on memory access and data movement.
As a consequence:

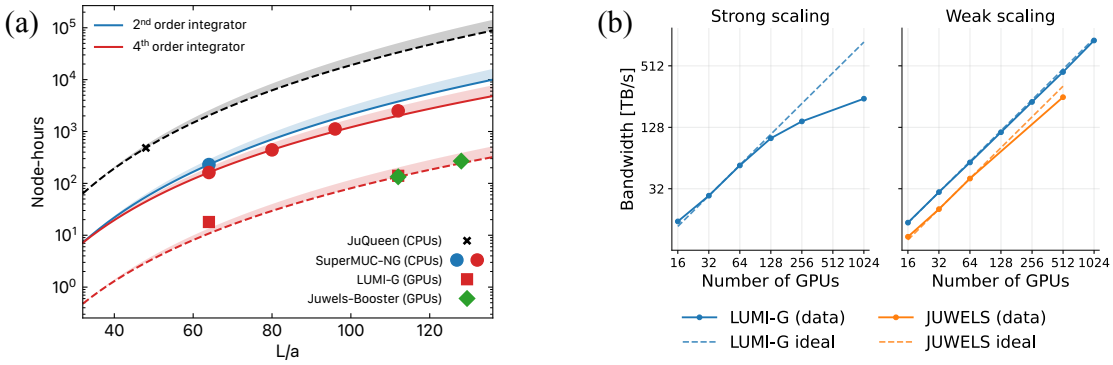

Figure 2: (a): `tmLQCD`: Cost per molecular dynamics unit for a simulation with light, strange and charm quarks tuned to physical masses. Utilising multi-grid solvers accelerated the simulations by a factor 10. Using GPUs decreases the required node-hours by another factor 10. (b): `HiRep`: Strong and weak scaling of the Dirac operator on LUMI-G and JUWELS Booster with all dimensions parallelised up to 1024 GPUs.

- **High memory bandwidth** is a critical requirement for future architectures. Equally important is the availability of low-latency, high-bandwidth interconnects, as lattice simulations rely heavily on nearest-neighbour communication patterns combined with frequent global reductions. Achieving efficient strong scaling to large fractions of modern supercomputers therefore depends both on network performance and the ability of the runtime system to overlap communication and computation effectively.
- **Large on-device memory**, such as high-bandwidth memory on accelerators, is also of central importance. Many applications require storing multiple fields and propagators simultaneously, particularly during measurement phases where solutions of the Dirac equation are reused extensively.
- In this context, a **balanced ratio between compute capability and memory bandwidth** is essential. Peak floating-point performance alone is not a reliable indicator of application performance; instead, sustained performance for stencil-like kernels with regular memory access patterns is the relevant metric for lattice workloads.
- **High and stable double precision performance** remains essential to ensure numerical accuracy, in particular for global reductions and long molecular-dynamics trajectories where rounding errors may accumulate. In selected cases, support for extended precision may also be required, for example in the presence of ill-conditioned systems or long summations. In the past this was usually realised in hardware, however, emulated high-precision arithmetic may be feasible, if performant and sufficiently verified.
- Support for **mixed-precision computations** is increasingly important. Modern solvers often exploit lower precision (such as single or half precision) in inner iterations to improve performance, while maintaining double precision for outer corrections and for the computation of physical observables.
- Lattice QCD simulations require efficient exploitation of **massive parallelism**. Architectures based on GPUs or many-core processors are well suited to this task, provided that lattice domains can be mapped efficiently onto the available parallel resources. Strong scaling to very large system sizes remains a key requirement, especially for large-volume simulations and long Markov chains. This, in turn, depends critically on both hardware capabilities and software support, including efficient MPI implementations and runtime systems. Given the regular structure of lattice discretisations, topology-aware network designs—such as **hyper-torus or locality-aware networks** offer clear advantages by minimising communication overhead and preserving locality in nearest-neighbour exchanges.

3.2. Software

Given the high demand in computing resources, the lattice QCD community has a long and successful history in self-developing highly optimised software, often including hand-optimised code targeting specific HPC architectures. For instance, the lattice community had already realised and applied the power of GPUs for scientific applications at a time when GPUs were mostly used for gaming [39]. This has led to the development of a handful of simulation software suites, maintained by small groups of physicists often connected to one of the collaborations. Examples are `Chroma`, `Grid`, `HiRep`, `MILC-QCD`, `openQCD`, `QUDA`, `SIMULATeQCD`, and `tmLQCD`.

With the increasing complexity of modern computer architectures, performance portability has become a major challenge, as writing hand-optimised code for each platform is increasingly demanding. This has also sparked the

development of novel approaches, for example implemented in `Kokkos`, a performance portability library that provides abstractions for parallel execution and data management. By expressing parallelism and memory access patterns at a high level, these approaches aim to enable compilers and backends to map the same code efficiently onto different architectures, such as multi-core CPUs and GPU accelerators, from a single source code base.

Maintaining such code suites requires continuous effort: continuous integration systems and automated performance regression tests are becoming increasingly important for sustaining reliability, reproducibility, and numerical efficiency. Moreover, these software packages require not only a stable and well maintained software stack, but also well-trained and sufficient personnel at the relevant computing centres: lattice QCD software often tests the boundaries of the hardware, and relies on the most recent features of the development toolchain. Therefore, small updates in the software stack regularly break the build chain of lattice QCD software, or lead to crashes. To address this, lattice QCD codes should be included into continuous testing pipelines of the high-performance computing centres. This is also useful for detecting hardware and compiler/module stack issues.

3.3. Human resources

Efficient utilisation of modern HPC systems requires close and sustained collaboration between research groups and computing centres. Performance analysis, architecture optimisation, and management of increasingly complex simulation workflows demand specialised expertise that goes beyond the traditional skills of individual research groups.

Dedicated support personnel with expertise in parallel programming, GPU optimisation, and performance engineering are crucial for translating algorithmic requirements into efficient implementations on evolving HPC architectures. Such roles are particularly critical given that they require a combination of in-depth knowledge of numerical algorithms, hardware architectures, and software engineering practices, while at the same time being highly attractive for the broader job market. As a result, maintaining and developing this expertise within academia remains a structural challenge for many scientific communities, including lattice field theory, where software development is still largely distributed across individual groups and lacks a coherent, long-term European framework.

A successful model to address these challenges is provided by the Swiss Platform for Advanced Scientific Computing (PASC, https://pasc-ch.org/about/index.html). PASC fosters close collaboration between universities and the Swiss National Supercomputing Centre, with a strong focus on application-driven software development and co-design between domain scientists and HPC experts. Extending similar frameworks at the European level would significantly strengthen the ability of lattice field theory and related disciplines to exploit current and future HPC systems effectively. In this context, the EuroHPC Centres of Excellence could represent a decisive step forward in establishing such coordinated efforts at the European scale.

3.4. Critical view of present trends in EuroHPC

European initiatives such as EuroHPC provide unprecedented computational capabilities and represent a major strategic investment for European science and technology. The deployment of large-scale systems across Europe has significantly expanded access to leadership-class computing resources. At the same time, several emerging trends offer new opportunities to further strengthen the impact of these infrastructures across a broad range of scientific domains.

- An important opportunity lies in broadening the criteria used to guide hardware procurement decisions. While current approaches capture peak floating-point performance and accelerator-driven workloads, complementing them with community-specific benchmarks would ensure a more balanced representation of scientific needs. This is not only relevant for lattice QCD, but also for other numerical approaches to problems in science and engineering, including hydrodynamics and numerical gravity [40]. These applications share demanding requirements in terms of communication costs, memory bandwidth, sustained strong scaling, network topology, latency, floating-point precision, and sustained bandwidth. Incorporating representative workloads from these communities into benchmarking suites would therefore help ensure that future systems suit a broader range of scientific applications.
- A further strategic opportunity is the development of a clear long-term European roadmap towards post-exascale computing, covering both an intermediate five-year horizon and a longer ten-year perspective. Building on the strong foundation established by current EuroHPC systems, increased visibility on future architectures and infrastructure evolution would provide valuable guidance to scientific communities engaged in long-term code development and optimisation. Establishing such a roadmap, in alignment with other international efforts, would strengthen Europe's position in the global HPC landscape and support sustained scientific leadership.
- The rapid growth of AI-driven workloads represents a major and welcome evolution in the HPC landscape, opening

new scientific and technological opportunities. At the same time, this trend should be accompanied by a balanced allocation of resources that continues to support other computational science domains.

- Established fields such as lattice field theory have historically been key drivers of innovation, producing advances in algorithms, software, and hardware whose broader impact is often difficult to anticipate but frequently transformative. Maintaining strong support for these research directions alongside emerging AI workloads will help preserve a diverse and resilient scientific ecosystem, maximising both immediate and long-term societal benefits.

A balanced approach to procurement, access policies, and long-term planning is essential for communication-bound applications such as lattice field theory.

4. Conclusions

Researchers in lattice field theory are among the largest and most efficient users of high-performance computing resources in fundamental science. The computational workloads place sustained demands on memory bandwidth, communication, and scalability, providing stringent benchmarks for current and future HPC architectures. At the same time, lattice field theory illustrates the broader impact of curiosity-driven research. Its first-principles description of strongly interacting systems has driven methodological advances beyond its original domain, including the development of algorithms such as Hybrid Monte Carlo (Hamiltonian Monte Carlo) [13] and contributions to the co-design of HPC systems, exemplified, e.g., by IBM's BlueGene architectures. These developments have had a lasting impact on computational science and data-driven methodologies.

The field remains one of the most computationally demanding areas of computational science, with progress tightly coupled to advances in algorithms, computing infrastructures and, potentially, machine learning. Looking beyond the exascale era, continued progress will require balanced HPC architectures capable of sustaining memory bandwidth and communication performance, together with robust and sustainable software ecosystems.

A clear vision for post-exascale computing is therefore essential. This must ensure balanced support for fundamental research domains, such as lattice field theory, that drive conceptual advances and technological innovation, along with emerging trends.

Sustained investment in human expertise is equally critical. The increasing complexity of hardware and software demands dedicated support at the interface of physics, numerical methods, and HPC. Establishing stable career paths for research software engineers and strengthening collaboration between scientific communities and computing centres are essential components of an effective long-term strategy.

Acknowledgements

The authors thank L. Del Debbio, A. Francis, B. Kostrzewa, C. Pica, S. Ryan, and S. Schaefer for valuable discussions and input. We also thank the OpenLAT collaboration for sharing information on their computational resource allocations. GA is supported by STFC grant ST/X000648/1 and a Royal Society Leverhulme Trust Senior Research Fellowship. JF received funding from the European Research Council via the project "LEEX" (GA 101170304). CU and SK were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the CRC 1639 NuMeriQS – project no. 511713970 and under Germany's Excellence Strategy – EXC 3107 – Project-ID 533766364 as part of the Cluster of Excellence Color-meets-Flavor. GB, SK and CU are supported in part by DFG project 460248186 (PUNCH4NFDI). We acknowledge EuroHPC JU for awarding access under the projects EHPC-EXT-2025E01-079 (222 knh on LUMI-G), EHPC-EXT-2023E01-010 (111 Mch on LUMI-C), EHPC-EXT-2025E01-091 (JUPITER), EHPC-EXT-2022E01-096 (120 Mch on LUMI-C), and EHPC-EXT-2023E01-018 (462 knh on Leonardo and 120 Mch on LUMI-C), which contributed to the studies presented.

References

[1] R. Aliberti, et al., The anomalous magnetic moment of the muon in the Standard Model: an update, Phys. Rept. 1143 (2025) 1–158. doi:10.1016/j.physrep.2025.08.002.

[2] J. de Blas, M. Dunford, E. Bagnaschi, A. Freitas, P. P. Giardino, et al., Physics Briefing Book: Input for the 2026 update of the European Strategy for Particle Physics doi:10.17181/CERN.35CH.2O2P.

[3] G. Aarts, et al., Phase Transitions in Particle Physics: Results and Perspectives from Lattice Quantum Chromo-Dynamics, Prog. Part. Nucl. Phys. 133 (2023) 104070. doi:10.1016/j.ppnp.2023.104070.

[4] S. Aoki, et al., FLAG Review 2019: Flavour Lattice Averaging Group (FLAG), Eur. Phys. J. C 80 (2) (2020) 113. doi:10.1140/epjc/s10052-019-7354-7.

[5] Y. Aoki, et al., FLAG Review 2021, Eur. Phys. J. C 82 (10) (2022) 869. doi:10.1140/epjc/s10052-022-10536-1.

[6] Y. Aoki, et al., FLAG review 2024, Phys. Rev. D 113 (1) (2026) 014508. doi:10.1103/nfzp-p5dn.

[7] C. T. H. Davies, A. C. Irving, R. D. Kenway, C. M. Maynard, International lattice data grid, Nucl. Phys. B Proc. Suppl. 119 (2003) 225–226. doi:10.1016/S0920-5632(03)01509-3.

[8] A. C. Irving, R. D. Kenway, C. M. Maynard, T. Yoshie, Progress in building an International Lattice Data Grid, Nucl. Phys. B Proc. Suppl. 129 (2004) 159–163. doi:10.1016/S0920-5632(03)02517-9.

[9] A. Ukawa, Status of international lattice data grid: An Overview, Nucl. Phys. B Proc. Suppl. 140 (2005) 207–208. doi:10.1016/j.nuclphysbps.2004.11.251.

[10] M. G. Beckett, B. Joo, C. M. Maynard, D. Pleiter, O. Tatebe, T. Yoshie, Building the International Lattice Data Grid, Comput. Phys. Commun. 182 (2011) 1208–1214. doi:10.1016/j.cpc.2011.01.027.

[11] M. D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 160018. doi:10.1038/sdata.2016.18.

[12] F. Karsch, H. Simma, T. Yoshie, The International Lattice Data Grid – towards FAIR data, PoS LATTICE2022 (2023) 244. doi:10.22323/1.430.0244.

[13] S. Duane, A. D. Kennedy, B. J. Pendleton, D. Roweth, Hybrid Monte Carlo, Phys. Lett. B 195 (1987) 216–222. doi:10.1016/0370-2693(87)91197-X.

[14] L. Del Debbio, G. M. Manca, E. Vicari, Critical slowing down of topological modes, Phys. Lett. B 594 (2004) 315–323. doi:10.1016/j.physletb.2004.05.038.

[15] S. Schaefer, R. Sommer, F. Virotta, Critical slowing down and error analysis in lattice QCD simulations, Nucl. Phys. B 845 (2011) 93–119. doi:10.1016/j.nuclphysb.2010.11.020.

[16] M. Hasenbusch, Speeding up the hybrid Monte Carlo algorithm for dynamical fermions, Phys. Lett. B 519 (2001) 177–182. doi:10.1016/S0370-2693(01)01102-9.

[17] M. Lüscher, Solution of the Dirac equation in lattice QCD using a domain decomposition method, Comput. Phys. Commun. 156 (2004) 209–220. doi:10.1016/S0010-4655(03)00486-7.

[18] M. Lüscher, Local coherence and deflation of the low quark modes in lattice QCD, JHEP 07 (2007) 081. doi:10.1088/1126-6708/2007/07/081.

[19] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, T. A. Manteuffel, S. F. McCormick, J. C. Osborn, C. Rebbi, Adaptive multigrid algorithm for the lattice Wilson-Dirac operator, Phys. Rev. Lett. 105 (2010) 201602. doi:10.1103/PhysRevLett.105.201602.

[20] D. Boyda, et al., Applications of Machine Learning to Lattice Quantum Field Theory, in: Snowmass 2021, 2022.

[21] K. Holland, A. Ipp, D. I. Müller, U. Wenger, Machine-Learned Renormalization-Group-Improved Gauge Actions and Classically Perfect Gradient Flows, Phys. Rev. Lett. 136 (3) (2026) 031901. doi:10.1103/k41k-2pnc.

[22] M. Favoni, A. Ipp, D. I. Müller, D. Schuh, Lattice Gauge Equivariant Convolutional Neural Networks, Phys. Rev. Lett. 128 (3) (2022) 032003. doi:10.1103/PhysRevLett.128.032003.

[23] G. Catumba, A. Ramos, Stochastic automatic differentiation and the signal to noise problem, Eur. Phys. J. C 85 (9) (2025) 1037. doi:10.1140/epjc/s10052-025-14690-0.

[24] R. Abbott, D. Boyda, Y. Fu, D. C. Hackett, G. Kanwar, F. Romero-López, P. E. Shanahan, J. M. Urban, Variance reduction in lattice QCD observables via normalizing flows (2026). arXiv:2603.02984.

[25] K. Cranmer, G. Kanwar, S. Racanière, D. J. Rezende, P. E. Shanahan, Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics, Nature Rev. Phys. 5 (9) (2023) 526–535. doi:10.1038/s42254-023-00616-w.

[26] G. Aarts, D. E. Habibi, A. Ipp, D. I. Müller, T. R. Ranner, L. Wang, W. Wang, Q. Zhu, Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory (2026). arXiv:2601.19552.

[27] R. Abbott, D. Boyda, G. Kanwar, F. Romero-López, D. C. Hackett, P. E. Shanahan, J. M. Urban, Progress in Normalizing Flows for 4d Gauge Theories, PoS LATTICE2024 (2025) 066. doi:10.22323/1.466.0066.

[28] M. Caselle, E. Cellini, A. Nada, M. Panero, Stochastic normalizing flows as non-equilibrium transformations, JHEP 07 (2022) 015. doi:10.1007/JHEP07(2022)015.

[29] L. Wang, G. Aarts, K. Zhou, Diffusion models as stochastic quantization in lattice field theory, JHEP 05 (2024) 060. doi:10.1007/JHEP05(2024)060.

[30] S. Caron, et al., Strategic white paper on AI infrastructure for particle, nuclear, and astroparticle physics: insights from JENA and EuCAIF, Mach. Learn. Sci. Tech. 7 (1) (2026) 013002. doi:10.1088/2632-2153/ae35cd.

[31] K. K. Szabo, L. Lellouch, Z. Fodor, F. Stokes, B. C. Toth, G. Wang, New physics in the muon magnetic moment?, Procedia Comput. Sci. 240 (2024) 91–98. doi:10.1016/j.procs.2024.07.012.

[32] C. Alexandrou, et al., Large-scale simulations of lattice QCD for nucleon structure using Nf=2+1+1 flavors of twisted mass fermions, Procedia Comput. Sci. 267 (2025) 92–101. doi:10.1016/j.procs.2025.08.236.

[33] S. Borsányi, Z. Fodor, J. N. Guenther, P. Parotto, A. Pásztor, L. Pirelli, K. K. Szabó, C. H. Wong, The QCD crossover line in a finite volume, Procedia Comput. Sci. 267 (2025) 4–13. doi:10.1016/j.procs.2025.08.229.

[34] J. Finkenrath, Review on Algorithms for dynamical fermions, PoS LATTICE2022 (2023) 227. doi:10.22323/1.430.0227.

[35] J. Finkenrath, Future trends in lattice QCD simulations, PoS EuroPLEx2023 (2024) 009. doi:10.22323/1.451.0009.

[36] C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, M. Rottmann, Adaptive Aggregation-based Domain Decomposition Multigrid for Twisted Mass Fermions, Phys. Rev. D 94 (11) (2016) 114509. doi:10.1103/PhysRevD.94.114509.

[37] B. Kostrzewa, S. Bacchio, J. Finkenrath, M. Garofalo, F. Pittler, S. Romiti, C. Urbach, Twisted mass ensemble generation on GPU machines, PoS LATTICE2022 (2023) 340. doi:10.22323/1.430.0340.

[38] V. Drach, S. Martins, C. Pica, A. Rago, High-performance simulations of higher representations of wilson fermions, Computer Physics Communications 322 (2026) 110061. doi:https://doi.org/10.1016/j.cpc.2026.110061.

[39] G. I. Egri, Z. Fodor, C. Hoelbling, S. D. Katz, D. Nogradi, K. K. Szabo, Lattice QCD as a video game, Comput. Phys. Commun. 177 (2007) 631–639. doi:10.1016/j.cpc.2007.06.005.

[40] The scientific and innovation case for computing in europe 2026–2034, iSBN: 9789464597738 (Apr. 2026). URL https://prace-ri.eu/scientific-case/the-scientific-and-innovation-case-for-computing-in-europe-2026-2034/