Brookhaven National Laboratory

USQCD US Lattice Quantum Chromodynamics

CompF2: Lattice QCD

Seattle Snowmass Summer Meeting 2022

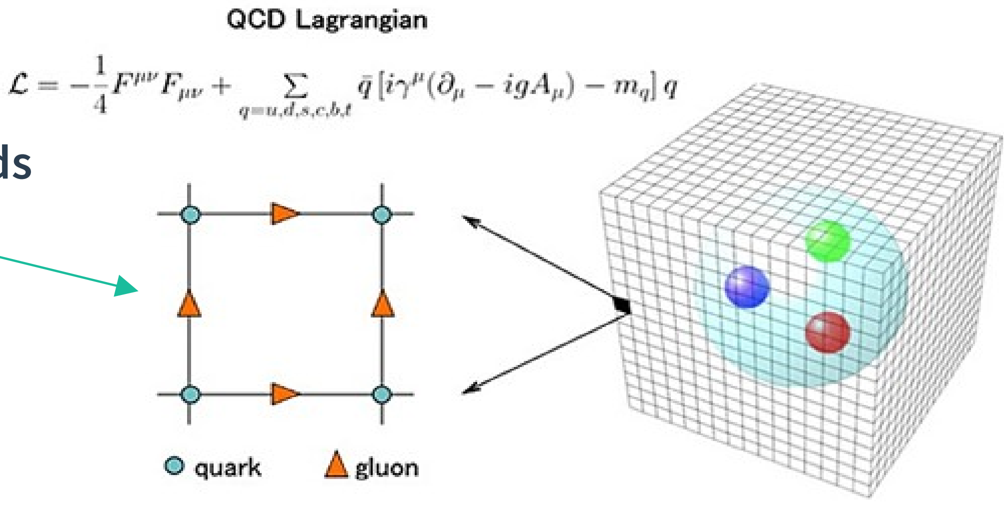
Christopher Kelly

07/18/22

Whitepaper : arXiv:2204.00039

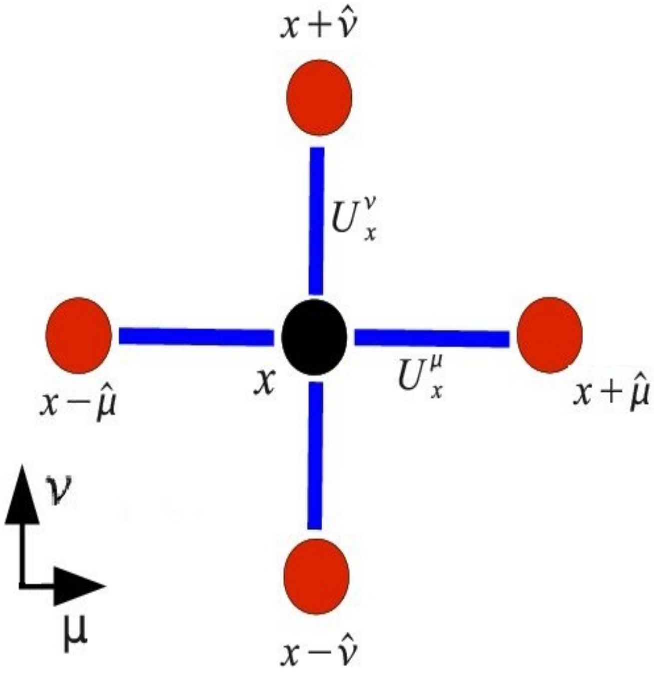
What is lattice QCD?

- Systematically improvable technique for studying non-perturbative Quantum Chromodynamics (QCD)
- Discretization of QCD on a finite volume represented by "gauge configurations": 4D grids of 3x3 complex matrices on the links between sites
- Naturally decomposes onto nodes arranged in 4D mesh
- USQCD primarily focuses on the *staggered* and *domain wall* formulations



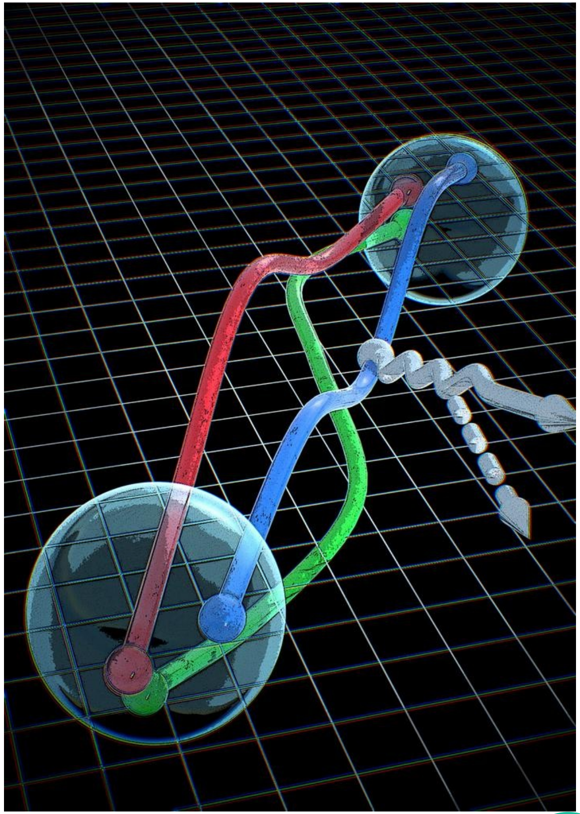
Computational challenge

- Core computation is inversion of Dirac operator typically via iterative (Krylov) methods
- Dirac op. is a (next-to)nearest neighbor 4D stencil operation
  - 1-2 Flops/Byte arithmetic intensity (single prec) means it is typically memory bandwidth bound
  - Neighbor gather requires large off-node comms bandwidth ~1/16 of local *cache* bandwidth for optimal balance
- Algorithms to reduce network bandwidth requirements are an active research topic
- Condition number diverges as a→0. Algorithmic challenge as we generate finer lattices.



Simulation phases

1) **Markov chain Monte Carlo methods used to generate an ensemble of gauge configurations**
   - Serially dependent
   - Requires a computer with good strong scaling
   - Critical slowing down issues appear at fine lattice spacings
2) **Measurements of hadronic observables are performed on each configuration**
   - Opportunity to amortize setup costs of more sophisticated inversion algorithms such as deflation and multi-grid methods
   - A high degree of trivial parallelism can be exploited



Physics focus of US community

Lattice calculations of Standard Model processes play a vital role in supporting US HEP experimental efforts:

[arXiv:2203.15810]

- Calculation of muon anomalous magnetic moment critical to interpreting results of FermiLab Muon g-2 experiment

[arXiv:2203.09030]

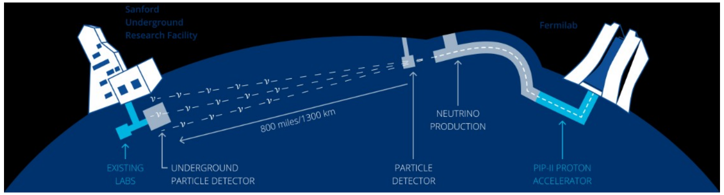
- Nucleon structure and parton physics calculations directly impact analysis of neutrino experiments (e.g. DUNE)

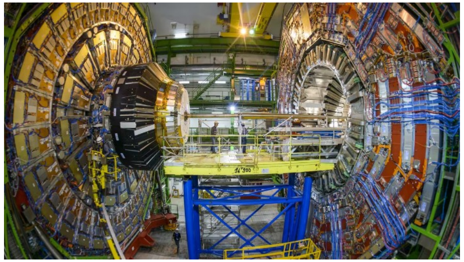
[arXiv:2205.15373]

- Beauty quark QCD physics calculations play important role in understanding b-physics anomalies in LHC and Belle II

[arXiv:2203.10998]

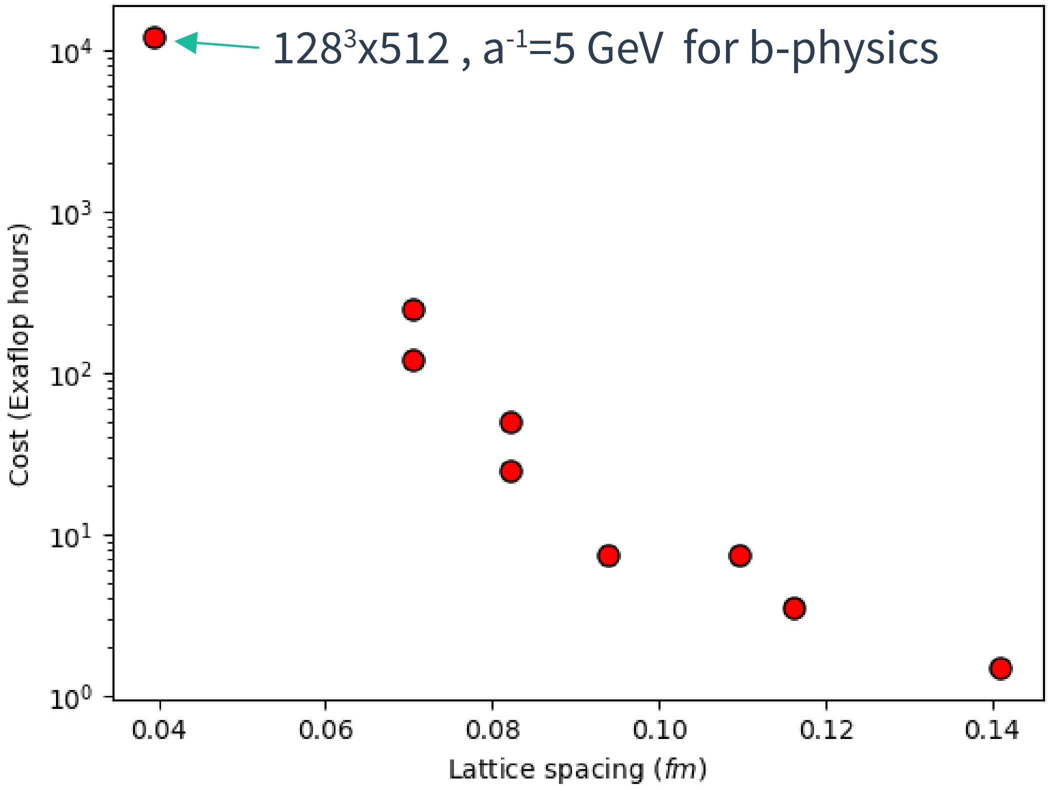
- Kaon QCD physics helps understand Standard Model CP violation and aids in the search for new physics in rare kaon decay processes.





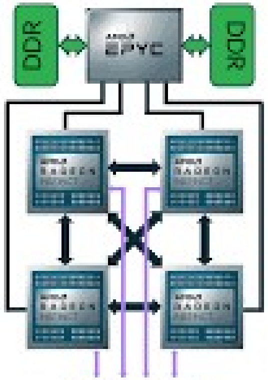
Computational requirements to support physics program

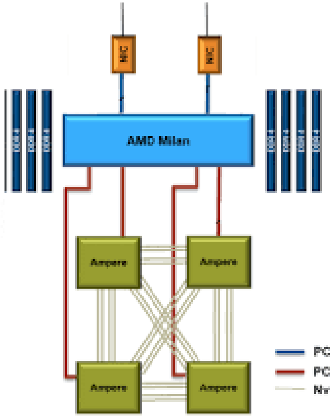
- **Precision to meet physics goals requires**
  - high statistics (# of gauge configs.)
  - large lattices to control finite-volume errs
  - fine lattice spacings to describe wide range of scales up to and including b-quark mass
- **Require computing resources at least 10x more capable than upcoming Exascale machines over Snowmass period**
- **Significant investment in algorithm development also necessary to complement hardware improvements and combat critical slowing down**



Software challenges

- Increasing heterogeneity in hardware:
  - complex memory hierarchies requiring manual control of data movement
  - accelerator variety; GPU, FGPA and spatial architecture, each with distinct programming models differing in both syntax and semantics
- Offers significant opportunity but places great burden on domain scientists to create performance portable codes
- Community is heavily reliant on a few custom codebases offering highly tuned and portable APIs for performing lattice simulations
- Software developed largely at National Labs through DOE programs e.g. ECP, SciDAC
- Long-term funding for support and development is minimal and insecure, increasing risk





Recommendations:
Hardware and programming model solutions

- At least a 10x increase in HPC compute capability over Exascale systems is required
- To support calculations that do not map easily to accelerator architectures, suggest provisioning of general purpose CPU-based clusters with GPU-like fast memory solutions
- A standard, unified interface for accelerator programming capable of generating high performance output is needed to reduce the burden of porting to new architectures
  - e.g. Sycl , OpenMP 5.0 and Parallel STL but all need better support and to demonstrate performance portability
- Investment in developing a generalized strategy for automating data movement is suggested
  - e.g. using a virtual memory paging system (à la Unix virtual memory or GPU managed memory) amortizing costs with larger page sizes if necessary

Recommendations:
Supporting algorithm/software development

- Continued investment in algorithm development is vital to achieving physics goals
- Continued software development funding is required at sufficient levels to provide complete products supporting the diverse array of computers and incorporating latest portability solutions
- Recommend DOE support dedicated software development career paths at Labs to retain expertise, enabling continued adapation to changing hardware environments
- Joint DOE funded Lab-university 5-year tenure track positions to likewise foster and develop talent at leading universities
- Support for training in new architectures and AI/ML techniques aimed at domain scientists (e.g. Hackathons, code examples) is necessary to keep abreast of new developments
- Early-access to new computing hardware is necessary to provide time to adapt to new environments