PROJECT NARRATIVE

1. SIGNIFICANCE OF RESEARCH

We request a three year allocation on the Argonne Leadership Computing Facility's Blue Gene/Q, Mira, and Theta and on the Oak Ridge Leadership Computing Facility's Cray XK7, Titan, in order to compute non-perturbatively aspects of hadron structure. The calculations we propose are directly supportive of the experimental program of the Thomas Jefferson National Accelerator Facility (TJNAF) facility, the Mainz Microtron (MAMI) in Germany, as well as aspects of the experimental program at the Large Hadron Collider (LHC). This proposal is submitted with the support of the USQCD Collaboration, and implements the research program of a subgroup of USQCD that focuses its efforts on understanding aspects of hadron structure. This research program is of high importance to the Collaboration, and utilizes gauge configuration ensembles that were generated with past USQCD INCITE proposals.

A key goal of the nuclear physics program is a first-principles description of nucleon and hadron structure from Quantum Chromodynamics (QCD), the theory of the strong interactions. Our knowledge of hadron structure is encapsulated in a variety of measures. From the earliest observations of Bjorken scaling in Deep Inelastic Scattering, a one-dimensional longitudinal description of the nucleon has been provided through the unpolarized and polarized *Parton Distribution Functions* (PDFs). In contrast, the transverse distribution of charge and currents was probed in elastic scattering, encapsulated in the electric and magnetic form factors. More recently, new measures have been discovered correlating both the longitudinal and transverse structures: the *Generalized Parton Distributions* (GPDs) describing hadrons in longitudinal fractional momentum x and transverse impact-parameter space, and the *Transverse-Momentum-Dependent Distributions* (TMDs) providing a description in x and transverse momentum space. These new descriptions have opened new vistas on the nucleon, enabling us to understand its spin decomposition and allowing the orbital angular momentum to be discerned.

From the inception of these new measures, the lattice community has attempted to calculate them from first principles. However, the formulation of lattice QCD in Euclidean space provided a formidable restriction: the x dependence could not be computed directly, but only the x moments of the distributions. Furthermore, the breaking of rotational symmetry on the lattice in practice restricted such calculations to only the first few moments. Recently, new ideas have been proposed that aim to circumvent one or more of these restrictions [29, 30, 36] known as the "Large Momentum Effective Field Theory" (LaMET).

The work proposed here will explore these new methods to investigate the properties of the pion, the lightest hadron as well as the nucleon. In particular, we will calculate the one- and three-dimensional images of the pion and the nucleon by computing, PDFs, GPDs and TMDs, and hence reveal vital information about the quark and gluon degrees of freedom in hadrons. This program is particularly timely given the approved experimental program at Jefferson Lab (12 GeV), which has the same physics goals as our proposed studies. Furthermore, a future electron-ion collider (EIC) will have the mission of uncovering the role of quarks and gluons in hadrons. Indeed, this was recognized as a Theory Initiative in the 2015 Nuclear Physics Long-Range Plan:

> *To meet the challenges and realize the full scientific potential of current and future experiments, we require new investments in theoretical and computational nuclear physics...*
>
> *...We recommend new investments in computational nuclear theory that exploit the US leadership in high-performance computing...*

Petascale computers, such as Mira and Titan, have made a transformational impact on our field. Without them projects such those described here are not possible. The USQCD collaboration has achieved important results with unprecedented precision in both high energy and nuclear physics by performing challenging numerical calculations that were previously not possible. Members of USQCD also continue to play leading roles in the development of algorithms and community codes. In this proposal we seek to build on these achievements to make possible further major advances in hadronic physics.

Computations of Parton Distribution functions

In the last few years, a major achievement in hadron structure has been the development of new methods that allow for direct computations [29, 36, 32, 20, 12]. This is a ground breaking development as it allows for the

first time to determine the full longitudinal momentum fraction dependence of the PDFs from lattice QCD, and thus opens up a new window for the theoretical study of the structure of fundamental building blocks of matter such as the pion and the nucleon. Experimentally hadron structure studies are a central part of DOE's nuclear physics programs both with current experimental facilities such as the JLab 12, GeV, RHIC, as well as at the future electron-ion collider (EIC). Furthermore, in the LHC era parton distribution functions (PDFs) of the nucleon are essential input in the search of new physics. Detailed studies of the Higgs boson as well as searches for heavy new particles (such as gluinos) require precise knowledge of PDFs at all ranges of the momentum fraction x, and in many cases PDF uncertainties at medium to high x dominate the total error budget. Furthermore, determining parton distribution functions in the *small x* region is particularly challenging. The functional forms of the PDFs are not well-known. In particular, the unknown glue helicity distribution at small x results in a large (more than 200%) error in fitting the total glue helicity from current experiment, and will need to rely on a future electron-ion collider to have small x data. In the meantime, *ab initiao* lattice calculation has the promise of being able to give PDF in certain x, and Q^2 ranges and can provide guidance to the global theoretical analysis from experiments.

The pion is of inherent importance as it is the Goldstone boson resulting from spontaneous chiral symmetry breaking and understanding its properties is essential to QCD theory and phenomenology. Furthermore, the pion is the simplest hadron that provides a good theatre for developing new ideas such as those discussed in this proposal. At JLab there are approved experiments for measuring the pion PDFs through the Sullivan process. In this case theoretical computations are particularly important as the experimental uncertainties are not well under control. Furthermore, the JAM collaboration at the JLab theory center is currently pursuing global fit analysis of the pion PDFs and having input from lattice QCD calculations will be essential in better controlling their global fit analysis.

The proton spin

Determining the contributions of the quarks and gluons to the nucleon spin is one of the most challenging issues in QCD both experimentally and theoretically. It was a big surprise 28 years ago when the deep inelastic scattering experiments revealed that the quark spin contributes only $\sim$25% to the total proton spin [3]. This is contrary to the conventional wisdom at that time which was based on the quark model predicting that 100% of the proton spin is due to the quark spin. This experimental finding was then called the 'proton spin crisis.' Recent analysis [19] of the experimental search for the remaining spin components due to the gluon helicity based on high-statistics 2009 STAR [21] and PHENIX [1] data showed evidence of non-zero glue helicity in the proton. For $Q^2 = 10\,\mathrm{GeV}^2$, they found gluon helicity distribution $\Delta g(x, Q^2)$ positive and away from zero in the momentum fraction region $0.05 \leq x \leq 0.2$. However, the result presented in [19] has very large uncertainty in the small x-region. There are also experimental attempts to measure the quark orbital angular momentum from the generalized parton distribution (GPD) in deeply virtual Compton scattering experiments to search for the missing proton spin components. This is an active research area both experimentally and theoretically and is an ideal subject for modern lattice QCD calculations to address. The lattice calculation we propose to carry out is aimed resolving the 'proton spin crisis' and understanding the origin of the small quark spin contribution to the proton spin. It also will address the 'proton spin puzzle' in finding out how the proton spin is apportioned among the quark and glue spins and their orbital angular momentum. These calculations should serve as predictions to be checked and verified by planned experiments to be conducted on the upcoming electron-ion collider.

Transverse Momentum Distributions

The three-dimensional distribution of the momenta of quarks in the nucleon is one of the pillars on which our understanding of nucleon structure rests. It is captured formally in transverse momentum-dependent parton distribution functions (TMDs). Experimentally, TMDs manifest themselves in angular asymmetries observed in processes such as semi-inclusive deep inelastic scattering (SIDIS) and the Drell-Yan (DY) process. Corresponding signatures have emerged at COMPASS, HERMES and JLab, and that has motivated targeting a significant part of the physics program at the upgraded JLab 12 GeV facility to the study of TMDs. Moreover, TMDs provide a central motivation for the physics program at the proposed electron-ion collider (EIC).

Concomitantly, the theoretical study of TMDs is being pursued vigorously; efforts within the U.S. theoretical nuclear physics community have been integrated through the recent launching of a DOE Topical Collabora-

tion on TMDs. The Lattice QCD studies of TMDs proposed here constitute a central element of the Topical Collaboration's program. We have developed and explored a scheme to calculate TMD observables using Lattice QCD in the past several years [38, 25, 24], and propose to build on that work in two important ways. Firstly, the studies to date have utilized pion masses considerably higher than the physical pion mass, 135 MeV; we currently only have reliable data down to 300 MeV. Notwithstanding the increasing computational expense as the pion mass is lowered, it is necessary to face the challenge of calculation at the physical pion mass.

Secondly, having gathered the expertise developed to access the quark transverse momentum structure of the proton puts a particularly attractive prize within reach: by combining transverse momentum information with transverse position information, as encoded in generalized TMDs (GTMDs), one can gain insight into (the longitudinal component of) orbital angular momentum. Generalizing the developed Lattice TMD methods, a first Lattice QCD calculation of quark orbital angular momentum in the proton according to the definition of Jaffe and Manohar has recently been presented [24], a significant advance in the investigation of proton spin structure. We propose to build on this exploration and advance the calculation of GTMDs to yield accurate, quantitative information on the proton spin puzzle and related aspects of proton structure, such as quark spin-orbit coupling.

2. RESEARCH OBJECTIVES AND MILESTONES

Parton Distribution Functions and Distribution amplitudes

The Pion: It has been shown that parton distribution functions can be probed via the computation of the so-called quasi parton distributions (quasi-PDFs) that are not defined on the light cone, but rather in terms of currents or quark fields at space-like separations; in the original proposal by Ji, these are defined by

$$\tilde{q}(x,\mu^2,P_z) = \int \frac{dz}{4\pi} e^{izk_z} \langle P \mid \bar{\psi}(0)\gamma_z \exp\left(-ig\int_0^z dz' A^z(z')\right)\psi(z) \mid P\rangle. \quad (1)$$

In contrast to the light-cone distributions, these matrix elements are at equal time, and thereby admit calculation on a Euclidean lattice. For states at high three-momentum, these quasi distributions can be related to the familiar standard distributions, and several recent works have computed these quasi distributions, and attempted to relate them to the well-known standard distributions in the nucleon [31, 2]. The matching of quasi distributions to light-cone PDFs requires considerable study, and an important aim of this proposal is to explore this matching. PI's and collaborators of this proposal have already considerable body of work on this topic [27, 37, 13, 39]. Note, however, that the quasi parton distribution functions [42], and the "quasi" parton distribution amplitude of the pion [43], are themselves of inherent physical interest. As shown in recent works [42] the quasi-PDF apart from containing information about the collinear PDFs, also contains information about the transverse structure of the hadron. A goal of this project is to extract the transverse content of the quasi-PDF along side with the collinear PDF following the approach discussed in our recent publication [39].

The pion is a particularly attractive theatre in which to explore these new ideas in hadron structure. The statistical uncertainty in lattice calculations involving the pion is considerably smaller than that for the nucleon, and the computation of correlation functions much simpler than in the baryon case. Most notably, the pion allows the computation both of the quasi parton-distribution functions, and of the quark distribution amplitude (DAs) that control exclusive process such as form factors at high-momentum transfers. An important question here, which we will be able to address, is whether the pion wave function is flat, asymptotic or has the highly peaked form suggested by sum-rule calculations[14]; this is a question that calculations of the moments of the distribution amplitude cannot reliably address.

The goals of this program are therefore as follows. Firstly, to compute the x-dependent quasi quark-distribution amplitude of the pion, and thereby to extract the leading Fock-space contribution to the pion's electromagnetic form factor encapsulated in the Distribution Amplitudes. By performing this matching at increasing pion momentum, we will investigate the approach of the quasi-DA to the light-cone DA that is related to experiment as well as extract information about transverse structure of the pion. Secondly, we will compute the quasi-PDF's of the pion, and thereby obtain a calculation of the collinear PDFs and once again explore the matching as the pion momentum increases. Finally, we will perform a direct calculation of

Project	Year 1	Year 2	Year 3	Total
Pion distribution amplitude				
$64^3 \times 128$, $a \approx 0.091$ fm	36	0	0	36
$72^3 \times 192$, $a \approx 0.075$ fm	0	24	32	56
Pion colinear PDFs				
$64^3 \times 128$, $a \approx 0.091$ fm	20	24	0	44
$72^3 \times 192$, $a \approx 0.075$ fm	0	15	55	70
Total Time (10^6 SUs)	56	63	87	206

Table 1: Resources requested for calculations of the pion DA and the pion PDFs using distillation expressed in Titan service units (SUs).

the pion form factor at values of Q^2 commensurate with the approved experiment E12-06-101 of Hall C at Jefferson Lab. This will allow us to compare the direct computation of the the form factor at high Q^2 values to the contribution that comes from the distribution amplitude.

All our calculations will be performed with a method developed at JLab, called distillation [44, 41], that allows for the isolation of the ground state matrix element with moderate Euclidean time separations resulting in an exponential reduction of the statistical noise. Statistical noise, as well as excited state contamination are major challenges for the calculation of quasi-PDFs and quasi-DAs at high momentum of the hadron and therefore using state of the art methods such as distillation will be of ground breaking nature. Leadership class computing is essential in applying such computationally demanding methods, however, the return on this investment will be an unprecedented precision to these calculations.

At the end of the project, we anticipate a comprehensive understanding of quasi-PDFs and quasi-DA of the pion, and their matching to collinear PDFs and DAs. More generally, the matching between the quasi- and standard distributions is independent of the hadronic state, and therefore the insight gained will be invaluable for general studies of hadron structure and in particular that of the nucleon.

The resource request for this project is listed in Tab. 1. It is computed using actual timings from Titan. The basic computational kernel is large sparse matrix inversion which is performed using multi-grid and takes 2.5s per right-hand side on 128 Titan nodes. For the pion DA calculations we use 32 correlation functions per configuration while for the pion PDFs we compute 8. For the pion DA each correlator requires one distillation propagator which corresponds to 4×256 inversions. For the pion PDF we need 5 distillation propagators per correlator hence we only compute 8 correlators per configuration. The cost for computing the generalized propagator beyond the cost of inversions is very small for this project (about 1M SUs per year) and for that reason we do not include it in the computation of costs. This because of the optimized software we have that performs the entire computation on the GPUs. In all cases we use 400 configurations.

Timetable of Activities

The timeline for major activities and deliverables is outlined as follows.

1. First year (Jan. 1, 2018 – Dec. 31, 2018):
   - Computation of the pion distribution amplitude using the distillation method. For this task will employ the $64^3 \times 128$, $N_f = 2 \oplus 1$ isotropic clover lattices, at a pion mass $M_\pi \simeq 170$ MeV, and with a lattice spacing $a \simeq 0.091$fm.
   - Begin computations of the pion collinear PDFs using distillation.
2. Second year (Jan. 1, 2019 – Dec. 31, 2019):
   - Complete computations of pion collinear PDFs using distillation.
   - Begin computations of the pion distribution amplitude using distillation at lattice spacing $a \simeq 0.075$fm with physical pion mass on a $72^3 \times 256$ lattice.

- Begin computations of the pion collinear PDFs using distillation at lattice spacing $a \simeq 0.075$fm with physical pion mass on a $72^3 \times 256$ lattice.

3. Third year (Jan. 1, 2020 – Dec. 31, 2020):

- Complete computations of the pion distribution amplitude using distillation at lattice spacing $a \simeq 0.075$fm with physical pion mass on a $72^3 \times 256$ lattice.
- Complete computations of the pion collinear PDFs using distillation at lattice spacing $a \simeq 0.075$fm with physical pion mass on a $72^3 \times 256$ lattice.

Project Team

This part of the project will be carried out by Kostas Orginos at College of William and Mary, Raul Briceno, Bipasha Chakraborty, Robert Edwards, Balint Joo, David Richards, and Frank Winter at JLab.

The Proton PDFs:

We will carry out precision calculations of the nucleon PDFs using the LaMET method, for which the motivation, methodology, and numerical procedure for the isovector case was detailed in earlier sections. Over the span of the award we will focus on performing calculations of the PDFs at finer lattice spacing to improve the systematics over our exploratory studies, and provide constraints for the global fits. Calculations at finer lattice spacings will allow us to to control discretization errors as well as have reliable matching of the quasi-PDF to the light cone PDF. In addition, small lattice spacings enable us to improve the reliability of the PDF at smaller values of the momentum fraction x, providing better insight to the anti-quark contributions to the PDF. Finally, finer resolution in the temporal direction, will also help to resolve the excited-state contamination which is expected to become more severe as the energy gaps between the excited boosted states become smaller. With all the above considerations, we propose to work with a mixed-action approach; that is, using HYP smeared clover fermion action as the valence quark on gauge configurations obtained using highly improved staggered quark (HISQ) action. Very fine HISQ gauge configurations (generated by MILC collaboration) already exist ready for use in this research: $a = 0.045$fm and multiple coarser lattice spacing with $M_\pi L \simeq 4$ at $M_\pi \leq 310$ MeV , including physical pion mass. The generation of gauge configurations for even finer lattice, namely for $a = 0.030$ fm and $M_\pi \leq 310$ MeV is also in progress. The HISQ gauge ensembles, generated with past USQCD INCITE awards provide the most economical approach for performing nucleon PDF calculations at fine lattice spacings. We have performed in the past single nucleon matrix element calculations with such ensembles with the same mixed action scheme proposed here. Our results indicate that in this set up one can achieve excellent control on all systematics associated with such calculations [6], and in particular the systematic errors associated with the lattice spacing extrapolation which is one of the main goals of this project.

Following the success of our pioneering studies, using the LaMET method for computing PDFs, we propose to improve our original lattice QCD calculation by using a much finer lattice spacing, $a = 0.045$ fm, with pion mass around 310 MeV and a $64^3 \times 192$ volume. This will be the first step towards a significant reduction of the systematics as discussed above. Subsequently we will perform the calculation on even finer lattices with $a = 0.03$fm and lattice volume $96^3 \times 288$ at the same pion mass and perform a detailed study of the continuum extrapolation. Our team will continue to investigating further improvements of the nonperturbative renormalization that was recently applied to the lattice quasi-PDFs [13]. Finally, in the third year we will work with global PDF community to incorporate lattice-QCD results as experimental data inputs to take advantage on both experimental and theoretical inputs. Together, these will significantly improve our understanding of PDFs, make predictions for the EIC era and assist new-physics searches at LHC.

With INCITE resources, our team will maintain international leadership of this fast developing research area.

The resources requested for this project are listed in Tab. 2. The basic computational kernel is large sparse matrix inversion using multigrid method and the timings have been obtained from the actual runs on Titan. We intend to use 500 gauge configurations for $64^3 \times 192$ lattice and 225 gauge configurations for $96^3 \times 288$ lattice. The propagators will be calculated using 256 sources to reduce the statistical noise.

Timetable of activities

Lattice	Year 1	Year 2	Year 3	Total
$64^3 \times 192$, $a \simeq 0.045$ fm	43	0	0	43
$96^3 \times 288$, $a \simeq 0.030$ fm	23	69	74	166
Total Time (10^6 SUs)				209

Table 2: Resources requested for the calculations of the nucleon quasi-PDF expressed in Titan service units (SUs).

1. First year (Jan. 1, 2018 - Dec. 31, 2018)

   Perform the calculation of quasi-parton distributions of the nucleon using $64^3 \times 192$ HISQ lattices at $a \simeq 0.045$ fm and pion mass of 310 MeV. In addition, begin production on the $a \simeq 0.045$ fm ensemble

2. Second year (Jan. 1, 2019 - Dec. 31, 2019) Continue the calculation of quasi-parton distributions of the nucleon using $96^3 \times 288$ HISQ lattices at $a \simeq 0.03$ fm and pion mass of 310 MeV. At the end of this year the calculations will have been performed on 125 configurations

3. Third year (Jan. 1, 2020 - Dec. 31, 2020) Complete the calculation of quasi-parton distributions of the nucleon using $96^3 \times 288$ HISQ lattices at $a \simeq 0.03$ fm and pion mass of 310 MeV by adding 100 gauge configurations into the analysis.

Project team

This part of the project will be carried out by Huey-Wen Lin at Michigan State University, Jiunn-Wei Chen at National Taiwan University, Jianhui Zhang at University of Regensburg, Yibo Yang at Michigan State University, Tomomi Ishikawa at Shanghai Jiao Tong University, Yong Zhao at Massachusetts Institute of Technology, Luchang Jin at BNL and University of Conneticut, Taku Izubuchi, Swagato Mukherjee, Peter Petreczky at BNL, and Sergey Syritsyn at Stony Brook University and RIKEN-BNL Research Center.

Small-x structure of the proton

The hadronic tensor has been formulated in the Euclidean path-integral formalism [35, 32]. Numerical methods have been proposed to address the inverse Laplace problem in order to convert the the hadronic tensor to the Minkowski space [33]. The main advantage of this approach is the fact that no renomalization is needed for the hadronic tensor which involves vector currents and it is frame-independent so that it can be calculated in any momentum frame of the nucleon. Furthermore, the valence+connected-sea (VCS) partons, the connected sea (CS) anti-partons, and the disconnected-sea (DS) partons and anti-partons are separately calculated in different path-integral diagrams [35, 32] and, thus, $\bar{u}$ and $\bar{d}$ in the CS which are responsible for the Gottfried sum rule violation can be revealed explicitly [34]. The challenging task of converting the Euclidean hadronic tensor calculated on the lattice to the Minkowski space is tackled with several methods. We propose to adopt the recently developed improved Maximum Entropy method [11], the Backus-Gilbert method [26] and model spectral function fitting to perform the inverse Laplace transform to see if the results are stable and if there are other algorithmic developments can be done. We will consider the hadronic tensor for the pion and both the unpolarized and polarized nucleon hadronic tensors and study the volume dependence and separate out different above-mentioned parton degrees of freedom via calculation of three different path-integral diagrams. We shall work on several anisotropic gauge field configurations produced by the JLab.

Timetable of Activities

The timeline for major activities and deliverables is outlined as follows.

1. First year (Jan. 1, 2018 – Dec. 31, 2018):

   - We will start by developing an efficient code to calculate the 4-point function for the nucleon hadronic tensor and test the inverse Laplace transform algorithms. We will use the JLab anisotropic lattices – $16^3 \times 128, 20^3 \times 128$ and $24^3 \times 128$ with the spatial lattice spacing $a = 0.123$ fam and $m_\pi = 389$ MeV for this test. The quark matrix inversion of the 16×128 lattice takes 0.06 hours

on one K40 GPU. This will be 3.6 service units on Titan. We will need to invert the quark matrix 600 times for the 300 configurations each for different nucleon momenta $|\vec{p}|$ and the injected photon momentum $|\vec{q}|$ to cover a range of the quark momentum fraction x, and different source, sink and current positions. This will take $3.6 \times 800 \times 300 = 1.6$ M service units for the $16^3 \times 128$ lattice, 1.5 M for the $20^3 \times 128$ lattice, and 2.4 M for the $24^3 \times 128$ lattice, for a total of 10.1 M service units.

2. Second year (Jan. 1, 2019 – Dec. 31, 2019):

   - We will extend our calculation to the $32^3 \times 256$ lattice to study the volume dependence of the inverse Laplace inverse transform and calculate the unpolarized nucleon and pion hadronic tensors and the polarized nucleon hadronic tensor. To scale the computer time from the $16^3 \times 128$ and taking into account the improved efficiency of Summit lattice for 250 configurations, this will require 10 M service units.

3. Third year (Jan. 1, 2020 – Dec. 31, 2020):

   - We will continue the calculation on a lattice with a pion mass at 236 MeV which is closer to the physical pion mass at 139 MeV on the $32^3 \times 256$ lattice. We will calculate the unpolarized nucleon and pion hadronic tensors and the polarized nucleon hadronic tensor at 4 Q^2 with 6 values of x each to compare with experiments. We will also calculate the disconnected sea partons. We estimate that this will require 20 M service units on Summit.

Project Team

This part of the project will be carried out by Jain Liang and Keh-Fei Liu at University of Kentucky, Kostas Orginos at College of William and Mary, and Yi-Bo Yang at Michigan State University.

Proton Spin:

To solve the proton spin decomposition with lattice QCD with all the systematic errors under control, we propose to use the overlap fermion as the valence quarks on the following 2+1 flavor domain wall fermion (DWF) configurations from the RBC and UKQCD collaborations.

Table 3: The parameters for the RBC/UKQCD configurations[8]: spatial/temporal lattice sizes, lattice spacings, physical spatial sizes, pion masses with degenerate light sea quark and number of configurations.

Lattice	$L^3 \times T$	a (fm)	L(fm)	m_π (MeV)	# Configurations
24I	$24^3 \times 64$	0.1105(3)	2.65	339	203
32I	$32^3 \times 64$	0.0828(3)	2.65	302	309
48I	$48^3 \times 96$	0.1141(2)	5.48	139	81
32ID	$32^3 \times 64$	0.1432(7)	4.58	172	200
24MID	$24^3 \times 64$	0.193(8)	4.8	137	200
32MID	$32^3 \times 64$	0.193(8)	6.4	137	200

These 6 ensembles of configurations have light sea quark masses corresponding to the pion mass range of 300 MeV down to the physical pion mass at 137 MeV, a range of volumes with the largest spatial size of 6.4 fm, and span a range of lattice spacings between 0.083 fm to 0.193 fm. They are sufficient for carrying out global fittings in chiral extrapolation including 3 lattices at physical pion mass, infinite volume and continuum extrapolations as have been done for the strangeness content, the glue spin, and the strange quark magnetic moment.

In view of the large normalization factor found in evaluating the quark spin from the local flavor-singlet axial axial-vector current through the anomalous Ward identity (AWI), we propose to use the chiral axial current derived for the overlap fermion which satisfies AWI configuration by configuration and will remove the systematic error due to the local current. We shall use both the overlap operator and gradient flow to

define the glue energy-momentum tensor to evaluate the glue momentum and angular momentum and its spin.

Timetable of Activities

The timeline for major activities and deliverables is outlined as follows.

1. First year (Jan. 1, 2018 – Dec. 31, 2018):

   - We shall produce 1000 pairs of overlap eigenmodes on each of the 200 configurations of the 32MID and 24MID lattices in Table 3. Based on our previous experience with Titan through the 2015 ALCC award which we used to generate eigenmodes for the 48I lattice, we estimate that it will take 23.8 M service units (30 service units for each node hour) for 32MID and 9.5 M core hours for 24MID for a total of 32 M service units. We plan to finish this production in 5 months.
   - Once the eigenmodes are calculated, we will proceed to invert the quark matrix of these lattices with eigenvector deflation and construct two-point and three-point correlators for calculating the exact chiral axial-vector current of the overlap fermion for the nucleon matrix element of the quark spin for both the connected and disconnected insertions. Matrix inversion for each quark propagator takes 1379/292 core hours for the 32MID/24MID configuration. To resolve the excited state contamination, we will need 4 source-sink separation for a total of 115 inversions for both the connected insertion (CI) and disconnected insertion (DI) calculations. With 200 configurations each for the two lattices, this part of the project will need $(1379 + 292) \times 115 \times 200$ = 38.4 M service units. The glue spin part takes only 1 M hours once the nucleon propagators are calculated.

2. Second year (Jan. 1, 2019 – Dec. 31, 2019):

   - We will start calculating the quark and glue spins for the 48I lattice (Table 3). Eigenmodes have been calculated before and a copy of them will be transferred to ALCF from NERSC. It takes 5,100 hours to do an inversion, we will need 2 source-sink separations (the other two were calculated elsewhere) and 115 inversions for 81 configurations for a total of $5100 \times 60 \times 81$ = 24.8 M service units. For the combined rest of the 32ID, 32I and 24I, we will need 15 M hours. Totally, this part will take 30M service units on Summit. We plan to finish this part of the calculation in 6 months so that we can have a manuscript on the quark spin for all the ensembles in Table 3 and produce a result with all the systematics under control.
   - Calculation of the quark and glue momentum and angular momentum will begin with the 48I lattice. This requires momentum transfers and will be more costly. We estimate it will need 120 inversions for a total of 30 M service units on Summit. We estimate that his part can be finished in 6 months.

3. Third year (Jan. 1, 2020 – Dec. 31, 2020):

   - We will continue the calculation of quark and glue momenta and angular momenta for the 32MID and 24MID lattices. Since the nucleon propagators are calculated in the first year, this will save time in both the CI and DI calculations. We will need 105 inversions. This gives an estimated time of 30 M service units on Summit. We will finish this part in 7 months.
   - Finally, we will finish the calculation of the quark and glue momentum and angular momentum for the remainder 32ID, 32I and 24I lattices. This will cost 25 M service units on Summit and should be finished in 5 months.

Project Team

This part of the project will be carried out by Jain Liang and Keh-Fei Liu at University of Kentucky and Yi-Bo Yang at Michigan State University.

Transverse Momentum dependent Distributions:

	Samples, #/cfg	(type)	Cost/sample Props, [c*h]	Cost/sample Contracts, [c*h]	Cost/cfg, [10^3 c*h]	Cost [10^6 c*h]
Transv. pol.	256	(approx)	280	1200	379	
	8	(exact)	2200	1200	27	
		(SUM)			406	81
Long. pol.	256	(approx)	420	655	275	
	8	(exact)	3300	655	32	
		(SUM)			307	62
Eigenvectors					115	23
Total						166

Table 4: Costs of computing quark propagators and contractions over 3 project years. Effort can be distributed flexibly, since calculations on each gauge configuration are independent. Transverse polarization: For each sample, 1 forward and, effectively, 1 sequential propagator is needed. The latter results from computing sequential propagators for 2 flavors, 2 source-sink separations, and 2 sink momenta, but combining propagators for 4 source times into one coherent propagator, and needing to propagate only the "upper" six components from the proton sink. Subsequently, 330,000 contractions per sample are evaluated. Longitudinal polarization: For each sample, 1 additional forward and, effectively, 2 sequential propagators are needed. The doubling compared to the transverse case results from the fact that one requires propagators containing a vector current insertion in order to implement a scheme to evaluate derivatives w.r.t. momentum transfer [18], needed for GTMD observables. Subsequently, 180,000 contractions per sample are evaluated.

TMD and GTMD observables are derived from proton matrix elements of the general form $\langle p'|\bar{q}(x)\Gamma U(x,y)q(y)|p\rangle$, where Γ denotes a Dirac structure, and the separation $(x-y)$ of the quark operators is the variable Fourier conjugate to the quark momentum. By using a gauge connection $U(x,y)$ between the quark operators that takes the form of a staple going to infinity, one sums the final state interactions between a high-energy quark ejected from the nucleon and the nucleon remnant. This limit is evaluated by extrapolation of calculations at finite staple lengths. The computational tasks involved in evaluating the aforementioned matrix elements are two-fold: One has to produce quark propagators from proton source and sink, and one has to calculate a very large number of contractions between them to survey all gauge connection geometries and Dirac structures needed for analysis. To estimate excited-state effects, two temporal source-sink separations will be used. To eliminate operator mixing effects induced when the quark discretization breaks chiral symmetry, the chirally symmetric domain-wall fermion (DWF) discretization will be employed. Concretely, a RBC/UKQCD DWF ensemble on $48^3 \times 96$ lattices at the physical pion mass will be used. The efficiency of propagator evaluation will be improved by employing Möbius acceleration, low-mode deflation, all-mode averaging (AMA), and a coherent sequential propagator scheme. This computational approach has been previously used successfully on the BlueGene/Q in an ALCC project [7]. We will use 200 thermalized gauge configurations and evaluate 256 AMA approximate samples on each, achieving a total of 51.2k samples (plus bias correction) for each individual matrix element calculated. The computational time needs are summarized in Table 4.

The computational effort can be divided into two main campaigns: Firstly, observables obtained with transverse proton polarization, which chiefly yield standard TMD observables related to SIDIS and Drell-Yan experiments, such as the Sivers and Boer-Mulders transverse momentum shifts; and, secondly, observables obtained with longitudinal proton polarization, which are most relevant for quark orbital angular momentum, spin-orbit coupling, and related effects, but also include the standard helicity TMD. In the latter case, a larger set of propagators is needed, since the additional transverse position information needed for GTMDs is extracted by using off-forward kinematics; standard TMDs require only forward matrix elements.

For our propagator computations, we will use the highly-optimized BFM implementation of the Möbius domain wall fermion operator (BFM) [9]. The actual inversion timings obtained in the project [7] are shown in Table 5. Contractions are implemented via code built using the Chroma library [22]. The contraction cost is estimated from the performance in previous projects carried out on Linux clusters, and rescaled according to the USQCD benchmark factor. cf. Table 5. The I/O performance of our software has been timed on a 1024 BlueGene/Q partition and found to be very efficient (cf. Table 5).

As far as storage needs are concerned, we plan to keep eigenvector data and quark propagators of active

	Note	time, s	BGQ c*h
Save Eigenvectors(2,000)	9.27 TiB	≈ 500	2280
Load Eigenvectors(2,000)	9.27 TiB	≈ 600	2730
Quark propagator(12 comp)	approx	30	140
Quark propagator(12 comp)	exact	240	1100
100,000 contractions			364

Table 5: Summary of I/O and inverter benchmarks using 1k (1024 node) partition on BlueGene/Q [7], as well as estimate of contraction costs.

gauge configurations on disk, to prevent data loss and enable recovery from system failures. We plan to process configurations in blocks of 5 to keep disk usage reasonable, but guarantee uninterrupted work flow. This will require approximately 200 TB of temporary online storage.

Project Team

This part of the project will be carried out by M.Engelhardt in collaboration with members of the Lattice TMD collaboration, including J.R.Green, Taku Izubuchi, J.Negele, A.Pochinsky and S.Syritsyn.

3. COMPUTATIONAL READINESS

3.1 Leadership Classification:

The work we propose requires the capability and capacity of leadership class computers. This project will exploit the largest gauge field ensembles generated on leadership class super-computers by USQCD. As such, computation of the correlation functions we need for extracting the physical observables, represents a capability and a capacity challenge. Firstly, we have a large number of configurations we need to analyze, and secondly some of the ensembles used in this work consist of very large lattices. The analysis on different configurations of an ensemble can be carried out in parallel, therefore leadership class computers can be used to ensure the timely completion of our program. We plan to run most of our projects on 20% or more of the system.

3.2 Computational Approach:

3.2.1 Underlying mathematical formulation: In lattice QCD, physical observables can be expressed in terms of Feynman path integrals on regular four-dimensional, hyper-cubic lattices,

$$\langle O \rangle = \frac{\int \mathcal{D}U \, O(U) \exp\left[-S(U)\right]}{\int \mathcal{D}U \exp\left[-S(U)\right]} \, .$$

Here, U represents the gauge field, which consists of one matrix assigned to each link of the lattice. Typical state of the art lattices now have $O(10^7 - 10^9)$ matrices. For QCD, each of the components of U is a 3×3 complex matrices that is a member of the group SU(3). $\mathcal{D}U$ represents an integral over all of the elements of U with the Haar measure, and $S(U)$ is the Euclidean action of the lattice theory after the quark fields have been integrated out. O is the physical observable being studied, and $O(U)$ is the value of O in the gauge configuration U. The first step is to use importance sampling techniques to generate an ensemble of gauge configurations U_i, $i = 1, \ldots N$, with the probability distribution

$$P(U_i) \sim \exp\left[-S(U_i)\right] .$$

Once an ensemble of representative gauge configurations is available, an unbiased estimator for the observable O is given by

$$\langle O \rangle = \frac{1}{N} \sum_i^N O(U_i) \, .$$

The projects proposed here have only this last step of the computation as the gauge fields are already available. The basic computational kernel is a the solution of a linear system with a large sparse coefficient matrix (of dimension $O(10^7 - 10^9)$). This task can only be performed on leadership class computers.

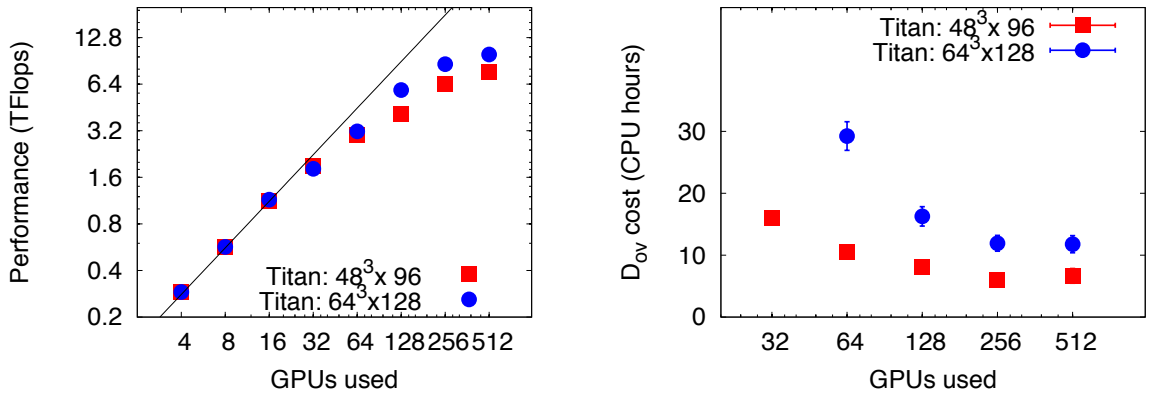

Figure 1: (a) Strong scaling performance for the double precision Wilson kernel. The fit line corresponds to 70 GFlops/gpu, the single gpu performance for our implementation on K20X GPUs with ECC turned on. (b) Service unit cost for a single overlap matrix-vector multiplication. The approximation error for the sign function we use for this test is $\varepsilon = 10^{-11}$, the approximation we use to run the eigensolvers.

3.2.2 Algorithms and numerical techniques:

Petascale computers, such as Mira and Titan, are enabling us to carry out calculations with smaller lattice spacings than previously possible using, for the first time, physical values of all quark masses. This development is greatly improving the precision of our calculations, but it requires major improvements in our solvers. The combination of increased condition number resulting from simulating at physical quark mass and more complex measurements with many operator inversions favors the use of deflation techniques in which one computes and exploits the first 500-2000 eigenvectors with smallest eigenvalues. The computation of the eigenvectors is done using the Lanczos algorithm or the implicitly restarted Arnoldi [45] algorithm in the case of non-hermitian matrices (which occur in our application). These methods are now providing a $5\times$ speed-up, or more, for our more difficult measurements. We have developed an adaptive multigrid (AMG) algorithm for clover quarks formulation, and are working to extend this approach to other formulations. We have also developed domain decomposition preconditioners for solvers running on multiple GPUs.

The inverter used for the overlap Dirac operator is a deflated conjugate gradient inverter which can be adapted to compute multiple quark masses without extra matrix-vector multiplication [28]. This is possible because the change in mass of the operators amounts to a simple shift and the Krylov spaces are identical. To further increase the efficiency of our codes we use an adaptive CG method [17]: the precision of the $\psi \rightarrow D_{\rm ov}\psi$ is decreased as we near convergence. The savings come from using a smaller order polynomial to approximate the sign function and, when the target precision is lower than 10^{-7}, we employ the single precision version of D_w, which speeds the multiplication by a factor of two. For a quark mass corresponding to $m_\pi \approx 200$ MeV the adaptive method is 60% faster than the regular CG.

The domain wall fermion propagators computations required by this project we will be computed with the highly-optimized BFM implementation of the Möbius domain wall fermion operator (BFM) [9] in the Conjugate Gradient and Lanczos algorithms. Its strong and weak scalings are shown in Fig. 2. Because of the memory required to store 2,000 eigenvectors, we will have to use 1024 nodes (one rack) of the BlueGene/Q. Our work flow will involve substantial I/O, and saving and loading eigenvectors has been optimized using an "ad-hoc" binary format to take advantage of the parallel file system and performs control check sums on the fly to ensure data integrity. The I/O performance in the Columbia Physics System lattice software package using the same approach has been timed on a 1024 BlueGene/Q partition and found to be very efficient (cf. Table 5).

For the clover quark propagators we will use the Chroma software system [23] in conjunction with the QUDA library [16, 5, 4] to calculate the quark solution vectors on Titan's GPUs. For this purpose, we will utilize the recently developed implementation of Adaptive-Smoothed Aggregation multi-grid technique

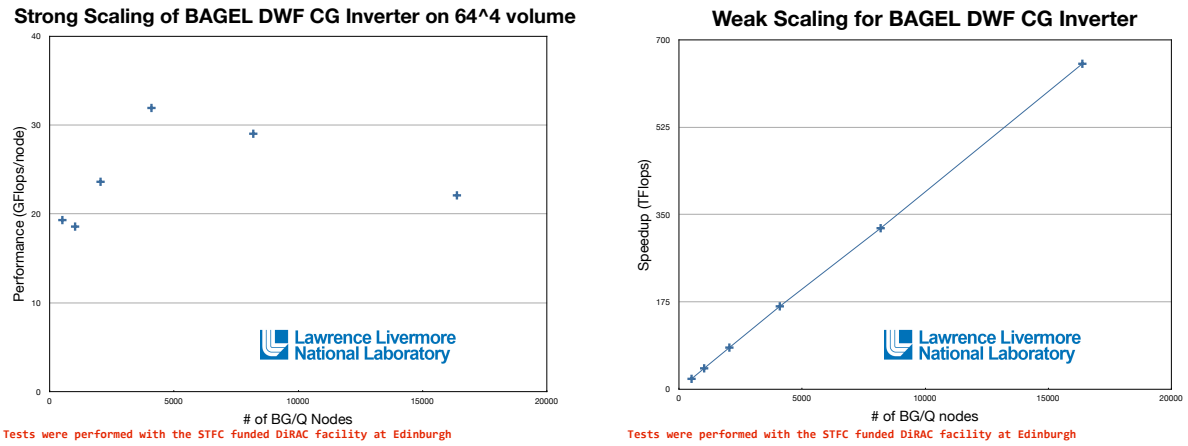

Figure 2: Performance of the optimized BFM implementation of the fermion operator [9] on the Blue-Gene/Q.

implemented in the QUDA library for lattice QCD on GPUs (QUDA-MG) [15]. Previously, we have used the CPU implementation in the QOPMG library [10, 40]

The GPU implementation of QUDA-MG been developed by Kate Clark of NVIDIA in collaboration with members of our team. Currently the library implements a multi-grid preconditioned GCR algorithm. The outer GCR solver features the usual QUDA performance features, including mixed precision allowing the use of single precision for subspace generation and gauge field compression, which has been discussed previously. The multigrid preconditioner is currently in single precision. The subspace creation is done, using the BiCGStab algorithm to solve the homogeneous system Mx = 0 to a fixed residuum to generate vectors rich in the low modes of M. Smoothing is performed using iterations of the MR algorithm. The multi-grid preconditioner can be either a classic V-cycle (V-cycle) in which the coarsest level solve uses GCR; or it can be a recursive tower of multi-grid preconditioned GCR solvers. Details of the optimizations employed in the multigrid algorithm are available in our contribution to SC'16 [15]. The development of this solver also served as an OLCF highlight.

Since the initial publication, the implementation has been further enhanced. The computation of the coarse operators can now be carried out entirely on the GPU, which substantially accelerates the setup phase. Further initial steps to add the solver into the HMC Gauge generation process have been carried out during an OLCF GPU Hackathon in Knoxville, TN in the fall of 2016. Finally, we note that the code-base has been ported to SumitDev at OLCF, and that the code base should be ready to run on the Summit system when that arrives. Current development focuses on relaxing scaling limitations to improve strong scaling (such as the minimum coarse lattice volume), improving the setup algorithm, and on completing the work on folding the multi-grid solver into the gauge configuration generation process for production use.

In figure 3 we re-plot the data from [15] comparing the performance of the multi-grid solver with the BiCGStab implementation available in QUDA. The timings were carried out on one of our target lattices with $64^3 \times 128$ sites with lattice spacing $a = 0.091$ fm, and quark masses corresponding to a pion with mass $m_\pi \approx 170$ MeV[1] For this lattice volume, using two levels of coarsening using blocking sizes of 4^4 and 2^4 sites, respectively, our strong scaling was limited to 512 GPUs, at which the local, per GPU lattice volume reached its minimum currently allowed value of 2^4 sites. In terms of production running in this proposal, we we would carry out ensemble style calculations utilizing between 64-128 Titan nodes per configuration. In this regime we should achieve an acceleration of between 5×-10× compared to the BiCGStab algorithm,

[1] In the SC'16 paper the lattice spacing was not yet fully determined, and lead the authors to estimate and quote that the quark mass corresponded to $m_\pi \approx 200$ MeV rather than the current value of 170 MeV.

as can be seen in 3. By working concurrently on up to 128 solver jobs with 128 nodes each, we could easily utilize up to 16,384 nodes of the 18,688 nodes available on Titan.

With the inversion algorithms above we are also using the all-mode-averaging (AMA) variance reduction technique. Here the same measurement is performed many times on each gauge configuration at low precision, with the sources of each of these measurements being a space-time translation of the others. High-precision measurements, performed less frequently, provide a correction, giving an exact result. For DWF measurements on Mira, speed-ups of $20\times$ are seen, but require this leadership class machine.

3.2.3 Programming languages, libraries and other software: Under grants from the DOE's SciDAC Programs, the USQCD Collaboration developed the QCD Applications Program Interface (QCD API), which will be used in this project. At the core of all computations is QMP (a component of the lowest level of QCD API) which is a communication library containing routines with the functionality of that portion of MPI relevant to lattice QCD calculations. It also has extensions that 1) partition the QCD space–time lattice, and map it onto the geometry of the hardware network; and 2) contain specialized routines designed to aid in the use of low level protocols on parallel computer networks, such as SPI on the Blue Gene/Q. A version of QMP that runs over MPI has been developed to ensure portability of codes.

The Chroma software package which is used in this work contains a high level interface of the basic QCD operations to low level efficient implementation. This one can achieve portability and good performance for the complex algorithms needed to solve our problems. Chroma is based on C++.

Finally, the codes used in this project use an I/O library, QIO, that enables users to read and write the different types of files that arise in our work in standard formats, which enables sharing of the large data sets that are created in the generation of gauge configurations and in their analysis. QIO supports a logical partitioning of the computer into I/O partitions with one core per partition handling I/O for the data in just that partition. Thus, our codes can read and write data in multiple files. The files thus created can be flattened into one large file offline on a single processor machine. There are no unusual memory requirements for this process. Likewise, for input one of our large files can be fragmented on a single-processor machine, copied to local disks, and read by the individual partitions. By tuning the size of the I/O partitions, we can maximize the I/O bandwidth and avoid contention. With a robust file system, such as Lustre or GPFS, we have written a 48.9 GB file from 4096 cores in 9 seconds and reread it in 6 seconds.

We have made a major effort to develop lattice gauge theory code for GPUs. The code has been incorporated into a new library, QUDA (QCD in CUDA). The first step was to develop code for all Dirac solvers commonly used in QCD (Clover-Wilson, Overalp, and DWF). The GPU solvers have already had a major impact on our physics analysis work. More recently, we have developed GPU codes for the full evolution of Clover-Wilson gauge configurations. The major QUDA codes run on multiple GPUs with support for partitioning of the lattice in all four space-time dimensions. The QUDA library incorporates a number of sophisticated optimizations. For example, core GPU kernels are subject to an automated tuning procedure which optimizes the assignment of thread blocks to the streaming multi-processors in a GPU. GPUDirect is also utilized, enabling peer-to-peer communication between GPUs sharing a PCIe bus.

Our team members Joo, Winter and .Pochinsky are active participants in the USQCD software development efforts. They have close relation with nVIDIA and Intel developing and porting our software to take advantage of the new machines (Summit and Aurora). Therefore we expect that our codes will be ready when the new systems become available.

3.2.3 Parallel programming model: Our lattices are divided into identical four-dimensional sub-lattices, each of which is assigned to a node. Communications between nodes are handled by message passing, MPI in the case of Crays, and SPI in the case of the Blue Gene/Q. Threaded codes, OpenMP, CUDA threads, are used on nodes and on GPUs.

3.2.4 Project workflow: The physics analysis on a configuration generally requires more floating point operations than were needed to generate it, configurations can be analyzed in parallel within a single job or multiple ones. In addition, the same configurations can be used to study a wide range of phenomena, again in parallel. Thus, physics analysis usually presents both a capacity and a capability problem for the most challenging problems such as those proposed here. Therefore, this analysis work must be carried out on leadership class machines, if results are to be obtained in a timely fashion. The analysis jobs are dominated by sparse matrix inversions using the BiCGStab or CG and AMG algorithms, so they are also suitable for

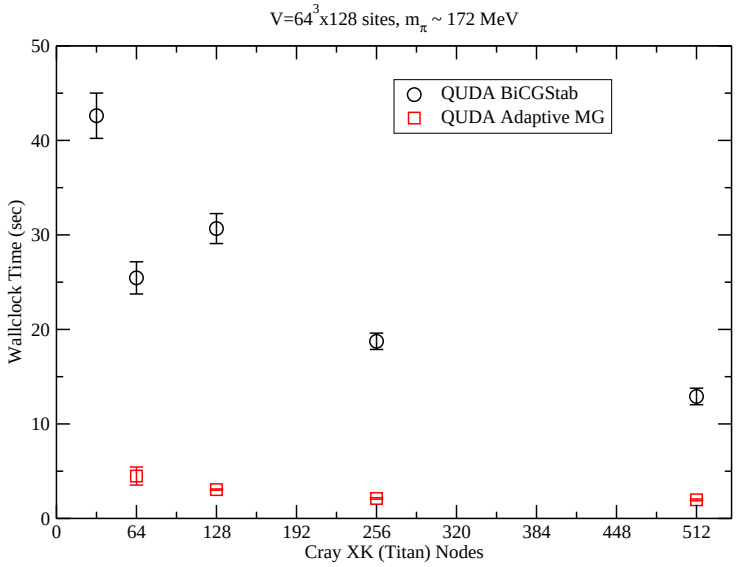

Figure 3: Performance of our algebraic multigrid solver (AMG) on Titan. The AMG solver achieves more than a factor of 10 increase in performance compared to a more traditional BiCGstab solver.

running on large partitions.

3.2.5 Software workflow: We have written a number of scripts which help to manage workflow. These scripts provide input parameters for jobs and in many cases submit new jobs when previous ones finish. In our analysis work, we use scripts to combine calculations of different observables in order to reuse expensive quark propagators whenever possible. Our calculations that are based on distillation, use a workflow engine called *Redstar* which is implemented in C++. *Redstar* is driven by a high level description of the correlators that need to be calculated. This program, which adds a minimal overhead, runs on the CPUs and generates the more complex input files that are needed to drive the computation on large GPU partitions. After all building blocks (quark propagators) have been generated *Redstar* proceeds to compute correlation functions. This last stage of the computation use optimized implementations of LAPACK and runs on the CPUs. Overall, for the application discussed here the automation achieved by Redstar comes at negligible computational overhead.

3.2.6 I/O requirements:

The largest files we will deal with in the proposed work will contain the $96^3 \times 192$ HISQ configurations, which are approximately 48.9 GB in size. As indicated in Section 3.2.3, we will use the QIO library to read and write files in parallel. This can be done in well under a minute per file, so I/O consumes a very small fraction of the time used for configuration generation. Our analysis work will also not strain the I/O systems of Mira or Titan. For an analysis job on the $96^3 \times 192$ HISQ gauge configurations, one would read in a single 48.9 GB configuration file, and write a few Megabytes of data in correlation functions. In the case of distillation calculations the output requirements higher and consist of $O(100)$ Gb per configuration. Using our IO software we can generally write such files in one to two minutes.

3.2.7 Data Storage:

The gauge configurations used in this project have been generated at ALCF or ORNL and have been used already in tape storage in these centers. No additional storage will be required for these configurations. When using deflated solvers eigenvectors are generated and those eigenvectors have to be stored on tape to be used in the future. In addition distillation propagators need to be archived for future use as they are

basic building blocks of correlation functions and can be used in a variety of future projects. In this case we need 500Gb of tape storage per configuration for long term archiving of distillation propagators for an aggregate of 500Tb of tape storage for the whole project. Temporary storage is also needed for our projects. The maximum request for the most data demanding application discussed here will require 200Tb of online temporary storage of intermediate steps in the calculation.

3.3 Development Work: We have an ongoing program to continue optimization of our codes for the Blue Gene and Cray architectures, which is funded by our DOE SciDAC and ECP grants. Members of USQCD at Jefferson Laboratory have a long term project to develop high performance code for QCD with Clover-Wilson quarks. This work has included the development of new algorithms to extend the scaling and efficiency of the Clover-Wilson solver, on Titan's XK nodes. Their success in the former is illustrated in Fig. 3.

Members of USQCD are also collaborating with Intel engineers to prepare our codes for the coming Aurora KNL architecture.

3.4 Use of Resources Requested:

3.4.1 Job Description: In Section 2 we provide tables showing the specific calculations we plan in each of the areas of research we propose. The text in that section describes each of the projects, and indicates how these calculations are related to the goals set out in Section 1.

3.4.2 Processor/core Utilization: The part of our project based on DWF calculations require a 1024 node BlueGene/Q partition primarily due to memory utilization. The rest of our project is requesting time on Titan and as our scaling plots indicate can run efficiently on a few as 128 nodes to as many as 512. Furthermore, these jobs can be bundled to processes more than one configuration at a time utilizing this way as many nodes as are available. Minimum running times range from one to two hours, but we would prefer to run jobs as long as allowed by the queue limits. We would also prefer to run long jobs for the analysis projects.

3.4.3 Calculation of Resource Request: The resources needed to analyze a single configuration are known from production runs already in progress for projects that have similar algorithmic components to the components needed by the projects proposed here. In some cases benchmark calculations of the exact production proposed here have already been performed as members of our team have access to ALCF and ORNL systems.

3.4.4 Anticipated Annual Burn Rate: Production codes for all of the calculations proposed, are ready and are already in use at ORNL and ALCF in ongoing projects. For that reason our teams are ready to start production right away and can have a very high burn rate if the queues allow it.

3.5 Parallel Performance:

In performing lattice QCD calculations one divides the lattice into a set of identical sub-lattices, and assigns one sub-lattice to each core or node. The number of lattice points on a core or node is often referred to as the local volume. In most cases the calculations performed on the various cores are identical, and only require data from a few neighboring cores. The major exceptions are the global sums that enter the conjugate gradient algorithm used for sparse matrix inversion. Thus, solvers for the Dirac operator provides the most stringent test of the scaling of our codes. It also gives a good indication of their overall performance, so we focus on it here. We have performed scaling tests on the Blue Gene/Q and Titan using the codes that we plan to employ on them. In Fig. 2, Fig. 3 and Fig. 1(b) we present results from scalings test for the DWF solver, Clover-Wilson solver and the overlap operator. The Clover-Wilson test was run on Titan, but the DWF test was run on the Blue Gene/Q Sequoia at Lawrence Livermore Laboratory because it has many racks than Mira, enabling a more extensive test.

References

[1] A. Adare et al. Inclusive double-helicity asymmetries in neutral-pion and eta-meson production in $\vec{p}+\vec{p}$ collisions at $\sqrt{s} = 200$ GeV. *Phys. Rev.*, D90(1):012007, 2014.

[2] Constantia Alexandrou, Krzysztof Cichy, Vincent Drach, Elena Garcia-Ramos, Kyriakos Hadjiyiannakou, Karl Jansen, Fernanda Steffens, and Christian Wiese. Lattice calculation of parton distributions. *Phys. Rev.*, D92:014502, 2015.

[3] J. Ashman et al. A Measurement of the Spin Asymmetry and Determination of the Structure Function g(1) in Deep Inelastic Muon-Proton Scattering. *Phys. Lett.*, B206:364, 1988.

[4] 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 Computi ng, Networking, Storage and Analysis Seattle, Washington, November 12-18, 2011*, 2011.

[5] Babich, Ronald and Clark, Michael A. and Joó, Bálint. Parallelizing the QUDA library for multi-GPU calculations in lattice Quantum Chromodynamics. *ACM/IEEE Int. Conf. High Performance Computing, Networking, Storage and Analysis, New Orleans*, 2010.

[6] Tanmoy Bhattacharya, Saul D. Cohen, Rajan Gupta, Anosh Joseph, Huey-Wen Lin, and Boram Yoon. Nucleon Charges and Electromagnetic Form Factors from 2+1+1-Flavor Lattice QCD. *Phys. Rev.*, D89(9):094502, 2014.

[7] T. Blum. Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment from Lattice QCD with chiral fermions. ALCC, 2016.

[8] T. Blum et al. Domain wall QCD with physical quark masses. *Phys. Rev.*, D93(7):074505, 2016.

[9] P. Boyle, M. Buchoff, N. Christ, T. Izubuchi, C. Jung, and T. Luu. The origin of mass, in: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. *SC'13*, 2013.

[10] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn, and C. Rebbi. Adaptive Multigrid Algorithm for Lattice QCD. *Phys. Rev. Lett.*, 100:041601, 2008.

[11] Yannis Burnier and Alexander Rothkopf. Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories. *Phys. Rev. Lett.*, 111:182003, 2013.

[12] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti. Nucleon structure functions from lattice operator product expansion. 2017.

[13] Jiunn-Wei Chen, Tomomi Ishikawa, Luchang Jin, Huey-Wen Lin, Yi-Bo Yang, Jian-Hui Zhang, and Yong Zhao. Parton Distribution Function with Non-perturbative Renormalization from Lattice QCD. 2017.

[14] V. L. Chernyak and A. R. Zhitnitsky. Exclusive Decays of Heavy Mesons. *Nucl. Phys.*, B201:492, 1982. [Erratum: Nucl. Phys.B214,547(1983)].

[15] M. A. Clark, Bálint Joó, Alexei Strelchenko, Michael Cheng, Arjun Gambhir, and Richard. C. Brower. Accelerating lattice qcd multigrid on gpus using fine-grained parallelization. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '16, pages 68:1–68:12, Piscataway, NJ, USA, 2016. IEEE Press.

[16] 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.*, 181:1517–1528, 2010.

[17] Nigel Cundy, Jasper van den Eshof, Andreas Frommer, Stephan Krieg, Th. Lippert, and Katrin Schafer. Numerical methods for the QCD overlap operator. 3. Nested iterations. *Comput. Phys. Commun.*, 165:221–242, 2005.

[18] G. M. de Divitiis, R. Petronzio, and N. Tantalo. On the extraction of zero momentum form factors on the lattice. *Phys. Lett.*, B718:589–596, 2012.

[19] Daniel de Florian, Rodolfo Sassot, Marco Stratmann, and Werner Vogelsang. Evidence for polarization of gluons in the proton. *Phys. Rev. Lett.*, 113(1):012001, 2014.

[20] William Detmold and C. J. David Lin. Deep-inelastic scattering and the operator product expansion in lattice QCD. *Phys. Rev.*, D73:014501, 2006.

[21] Pibero Djawotho. Gluon polarization and jet production at STAR. *Nuovo Cim.*, C036(05):35–38, 2013.

[22] Edwards, Robert G. and Joó, Bálint. The Chroma software system for lattice QCD. *Nucl.Phys.Proc.Suppl.*, 140:832, 2005.

[23] Edwards, Robert G. and Joó, Bálint. The Chroma software system for lattice QCD. *Nucl.Phys.Proc.Suppl.*, 140:832, 2005.

[24] M. Engelhardt. Quark orbital dynamics in the proton from Lattice QCD – from Ji to Jaffe-Manohar orbital angular momentum. *Phys. Rev.*, D95(9):094505, 2017.

[25] M. Engelhardt, P. Hgler, B. Musch, J. Negele, and A. Schfer. Lattice QCD study of the Boer-Mulders effect in a pion. *Phys. Rev.*, D93(5):054501, 2016.

[26] Backus George and Gilbert Freeman. The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, 16:169–205, 1968.

[27] Tomomi Ishikawa, Yan-Qing Ma, Jian-Wei Qiu, and Shinsuke Yoshida. Practical quasi parton distribution functions. 2016.

[28] Beat Jegerlehner. Krylov space solvers for shifted linear systems. 1996.

[29] Xiangdong Ji. Parton Physics on a Euclidean Lattice. *Phys. Rev. Lett.*, 110:262002, 2013.

[30] Xiangdong Ji. Parton Physics from Large-Momentum Effective Field Theory. *Sci. China Phys. Mech. Astron.*, 57:1407–1412, 2014.

[31] Huey-Wen Lin, Jiunn-Wei Chen, Saul D. Cohen, and Xiangdong Ji. Flavor Structure of the Nucleon Sea from Lattice QCD. *Phys. Rev.*, D91:054510, 2015.

[32] Keh-Fei Liu. Parton degrees of freedom from the path integral formalism. *Phys. Rev.*, D62:074501, 2000.

[33] Keh-Fei Liu. Parton Distribution Function from the Hadronic Tensor on the Lattice. *PoS*, LATTICE2015:115, 2016.

[34] Keh-Fei Liu, Wen-Chen Chang, Hai-Yang Cheng, and Jen-Chieh Peng. Connected-Sea Partons. *Phys. Rev. Lett.*, 109:252002, 2012.

[35] Keh-Fei Liu and Shao-Jing Dong. Origin of difference between anti-d and anti-u partons in the nucleon. *Phys. Rev. Lett.*, 72:1790–1793, 1994.

[36] Yan-Qing Ma and Jian-Wei Qiu. Extracting Parton Distribution Functions from Lattice QCD Calculations. 2014.

[37] Chirstopher Monahan and Kostas Orginos. Quasi parton distributions and the gradient flow. 2016.

[38] B U Musch, Ph Hägler, M Engelhardt, J W Negele, and A Schäfer. Sivers and Boer-Mulders observables from lattice QCD. *Physical Review D*, 85(9):094510, May 2012.

[39] Kostas Orginos, Anatoly Radyushkin, Joseph Karpie, and Savvas Zafeiropoulos. Lattice QCD exploration of pseudo-PDFs. 2017.

[40] J.C. Osborn, R. Babich, J. Brannick, R.C. Brower, M.A. Clark, et al. Multigrid solver for clover fermions. *PoS*, LATTICE2010:037, 2010.

[41] Michael Peardon, John Bulava, Justin Foley, Colin Morningstar, Jozef Dudek, Robert G. Edwards, Balint Joo, Huey-Wen Lin, David G. Richards, and Keisuke Jimmy Juge. A Novel quark-field creation operator construction for hadronic physics in lattice QCD. *Phys. Rev.*, D80:054506, 2009.

[42] Anatoly Radyushkin. Nonperturbative Evolution of Parton Quasi-Distributions. 2016.

[43] Anatoly V. Radyushkin. Pion Distribution Amplitude and Quasi-Distributions. 2017.

[44] Christian J. Shultz, Jozef J. Dudek, and Robert G. Edwards. Excited meson radiative transitions from lattice QCD using variationally optimized operators. *Phys. Rev.*, D91(11):114501, 2015.

[45] D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. on Matrix Analysis and Applications*, 13(1):357–385, 1992.