PROJECT NARRATIVE

1 SIGNIFICANCE OF RESEARCH

We request a three-year allocation on the Oak-Ridge Leadership Computing Facility's IBM/NVidia Summit computer to advance the DOE's high-priority search for new fundamental physics by testing the Standard Model at unprecedented precision. This proposal is submitted by two pairs of collaborations engaged in numerical simulations of quantum chromodynamics (QCD), namely, the joint Fermilab Lattice and MILC collaboration and the joint RBC and UKQCD collaboration. It has the support of the USQCD collaboration, a federation of nearly all US scientists engaged in numerical QCD calculations. This proposal describes two projects requiring the capabilities of Summit, one is in heavy-quark physics and the second is for synergistic calculations that use chiral quarks.

Since its inception, the INCITE program, through the availability of petascale resources, such as Mira and Titan, has had a transformational impact on our field. INCITE awards have been instrumental in allowing us to perform challenging numerical calculations that were previously not possible and to obtain many important results with unprecedented precision. In this proposal we seek to utilize the unique capabilities of Summit to reach long-sought precision goals for even more challenging quantities, central to the B-physics experiments at CERN and the KEKB accelerator in Japan, to the rare-kaon-decay experiments at CERN and the J-PARC facility in Japan, and the new $g_\mu - 2$ experiment at FNAL. The specific calculations we propose here are distinct from USQCD's current INCITE project on the ALCF Mira computer.

The search for new particles and interactions lies at the heart of high-energy-physics research. At the Intensity Frontier one looks for indirect signs of new particles through their quantum-mechanical effects that would result in deviations from Standard-Model expectations. However, many observables receive important contributions from strong-interaction effects that are not well known, which in turn limits the precision of these tests. Lattice Quantum Chromodynamics (QCD) is a general purpose tool for calculating the needed parameters with controlled uncertainties. The objective of this proposal is to bring the lattice-QCD errors down to, or below, the experimental ones for the quantities considered here, in order to maximize the new-physics discovery potential of current and future experiments at the Intensity Frontier.

Recent years have brought enormous progress, establishing lattice QCD as an essential tool in high energy phenomenology, especially for quark-flavor physics. In particular, decays of kaons and B mesons offer a unique and powerful way to search for physics beyond the Standard Model. Neutral quark-flavor-changing processes are especially interesting because they are suppressed in the Standard Model, offering the prospect of revealing properties of the underlying new physics. Consequently, high-precision flavor-physics studies probe physics at energies many orders of magnitude above those which are accessible to present or planned accelerators at the Energy Frontier. Thus, a strong quark-flavor program is an essential complement to high-energy collider experiments at the LHC and future facilities. Given the importance of quark flavor physics to the U.S. HEP mission, it has been consistently highlighted as a strategic goal of USQCD, including in the 2013 white paper *"Lattice QCD at the Intensity Frontier."*

1.1 High-precision semileptonic B-meson decay form factors

For the past decade, lattice QCD flavor-physics calculations, including those by the Fermilab Lattice and MILC Collaborations, have made critical contributions to determinations of CKM matrix elements (fundamental parameters of the Standard Model that control transitions between quark flavors) and to searches for beyond-the-Standard-Model effects in flavor physics. In particular, our recent results for the leptonic B, B_s, D, and D_s decay constants [1, 2] with crucial support from USQCD INCITE awards have sub-percent errors
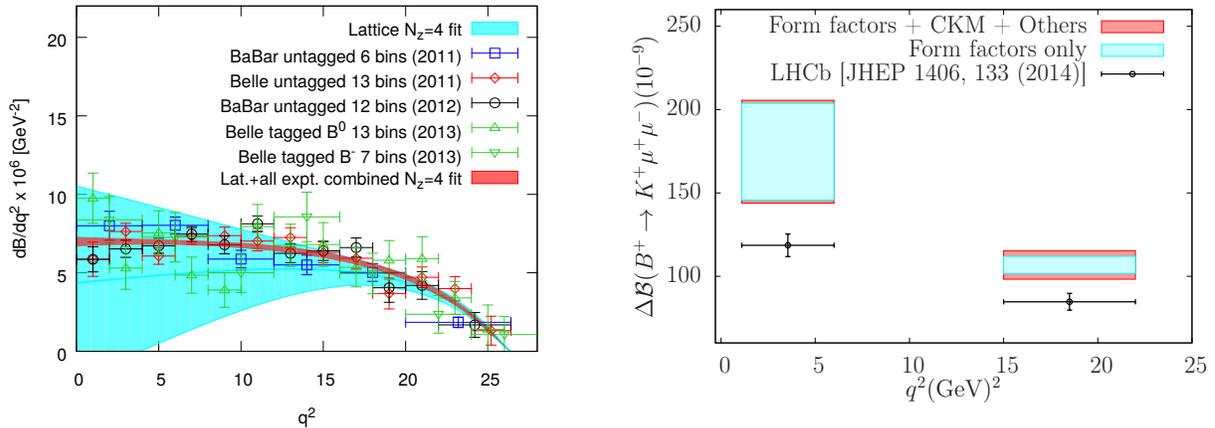
Figure 1. *(left)* **Partial branching fraction for the decay $B \to \pi \ell \nu$. The lattice-only fit result (cyan band) comes from our prior calculation of $f_+(q^2)$ [5] and the combined-fit result (red band) comes from fitting the experimental data together with the lattice form factor, leaving $|V_{ub}|$ as an unknown parameter to match the two results.** *(right)* **Standard-Model partially-integrated branching ratio for $B^+ \to K^+ \mu^+ \mu^-$ decay using our prior calculation of the form factors [12, 5, 11], compared with experimental measurements from LHCb [15, 16] for the wide q^2 bins above and below the charmonium resonances.**

and are 3 to 10 times more precise than any previous lattice-QCD calculation. Indeed, these parameters are now so well known that QCD theory is a subdominant source of error in the corresponding CKM matrix element determinations. Additionally, further reductions of the QCD theory errors will not improve the Standard-Model predictions of the corresponding leptonic decay rates without also including electromagnetic effects. In this sense, our results for the decay constants are the definitive QCD-only calculations.

The goal of the calculations proposed here is to bring the QCD-theory uncertainties on the semileptonic B-meson decay form factors to a similar precision level. Our previous calculations of these parameters, when combined with experimental measurements of the corresponding differential decay rates, yield the most precise determinations of the associated CKM matrix elements, $|V_{cb}|$ [3, 4] and $|V_{ub}|$ [5], which control transitions between charm and bottom quarks and between up and bottom quarks, respectively. For example, the total error on $|V_{ub}|$ is 3.8%, of which the lattice QCD form factors contribute about 3% and the experimental average [6, 7, 8, 9, 10] contributes 2.5% respectively.

Other recent highlights from our collaboration are the complete set of semileptonic decay form factors for tree-level and rare B-meson decays to pions, and kaons [11, 12], yielding new, interesting constraints on models of new physics [13] and the complete set of the neutral B and B_s meson mixing matrix elements [14], yielding the best-to-date constraints on $|V_{td}|$, $|V_{ts}|$, and their ratio. These difficult-to-measure CKM matrix elements control transitions involving the top quark.

These results form the basis for stringent tests of CKM unitarity. Neutral B-meson mixing and rare semileptonic B-meson decays (such as $B \to \pi \ell^+ \ell^-$, $B \to K \ell^+ \ell^-$) are especially promising channels for new-physics searches because their rates are suppressed in the Standard Model. Indeed, the Standard-Model expectations for these processes, obtained with our lattice results for the corresponding form factors and mixing matrix elements, differ from experimental measurements at the 2σ level, and point to an emerging tension between weak processes that are mediated at the loop and tree levels. Tests of lepton-flavor universality using semi-

leptonic B-meson decays have recently become very interesting, with the observation of several significant tensions involving both rare $b \to s$ and tree-level $b \to c$ transitions [17, 18, 19, 20, 21, 22, 23]. Here, our form factor results have been used to sharpen the Standard-Model expectations [3, 13, 24].

Figure 1 shows two examples from our prior calculations that indicate the need for additional precision. On the left, we plot the partial branching fraction for the decay $B \to \pi \ell \nu$. The cyan band shows that the errors in our lattice-QCD calculation are large at small q^2, but small at large q^2. The experimental results have the opposite characteristic. We determine $|V_{ub}|$ by matching the lattice result to experiment. A combined fit to both theory and experiment is shown in red and has smaller errors at all q^2. It is important to reduce the theory errors to keep pace with experimental improvements. On the right, we show results for the rare decay $B \to K \ell^+ \ell^-$. There is a tension between our Standard-Model prediction from lattice QCD and the experimental measurements. Here it is crucial to reduce the theory errors to see if this difference is statistically significant, which would be evidence of beyond-the-Standard-Model physics.

A wealth of experimental data already exists and more will become available in the coming years yielding improved measurements. The Belle II experiment has started running and expects to collect data sets that are 50–100 times larger than what was obtained by the Belle experiment [25]. For example, Belle II measurements of the semileptonic $B \to \pi \ell \nu$ differential distributions could yield a determination of $|V_{ub}|$ at close to 1% uncertainty [26], but only if there are commensurate improvements in the accuracy of the corresponding lattice QCD form factors. Similarly, it is expected that the LHC luminosity and detector upgrades will yield improved measurements at LHCb [27] with, again, significant reductions in the experimental uncertainties.

1.2 Second-order weak processes

Rare phenomena which can occur only at second order in the weak interactions as described by the Standard Model provide important opportunities to discover new phenomena that lie outside the Standard Model. Current mysteries such as dark matter, the absence of strong CP violation, the relatively large size of the baryon density in the Universe and the small value of the mass of the Higgs boson all point to limitations of the Standard Model and provide strong motivation to search for physics beyond it. However, for these rare processes to be used in such a search, the predictions of the Standard Model for these processes must be known. Here we propose to study two such phenomena: the mass difference, ΔM_K between the long- and short-lived neutral K mesons and the decay rate for the rare $K^+ \to \pi^+ \nu \bar{\nu}$ decay.

The kaon mass difference arises from a process which changes strangeness by two units and hence in the Standard Model is a very small second-order effect. It was first measured in the 1950's and is now accurately known: $\Delta M_K = 3.483(6) \times 10^{12}$ MeV. This is the smallest particle mass difference ever measured and because of its small size is exquisitely sensitive to physics at very high energies. At 10% precision, ΔM_K is sensitive to physics at an energy scale as high as 10^4 TeV. The major Standard Model contribution to ΔM_K comes from the charm energy scale, somewhat too low an energy to be computed reliably using QCD perturbation theory. In fact a NNLO calculation by Brod and Gorbahn [28] found that the NNLO contributions are 36% of the leading and next-leading result, explicitly indicating the difficulty of using perturbation theory.

There is another important quantity, closely related to ΔM_K, which is not included in this proposal because its calculation is not yet sufficiently mature. This is the long-distance contribution to the indirect CP violation parameter ϵ_K. Related to the imaginary part of the $K_L - K_S$ mixing matrix, this long-distance contribution is needed if the Standard Model prediction is to be compared with the measured value of ϵ_K at a precision of $\lesssim 1\%$. We hope to be able to include a calculation of this quantity in the later years of this proposal.

A second quantity that provides a highly sensitive test of the Standard Model is the decay rate for the rare process $K^+ \to \pi^+ \nu \bar{\nu}$. This is a forbidden "strangeness-changing neutral current" process and can occur in the Standard Model only in second order. This decay is suppressed by at least a factor of 10^{-10}. A 10% accurate measurement of this process is now the goal of the NA62 experiment at CERN. This rare kaon process was selected as the major goal of this experiment because this decay is dominated by the top quark and other short-distance effects and is expected to be computed accurately using QCD perturbation theory. However, there is a $\approx$ 5% contribution that comes from energies near the scale of the charm-quark mass where QCD perturbation theory may not be reliable.

Just as in the case of ΔM_K, lattice QCD can be used to compute the contributions coming from the energy scale of the charm-quark mass to this process, here identified as "long-distance" effects because of the uncertainty in using QCD perturbation theory to determine them. A reliable lattice QCD calculation of these long-distance contributions will both provide important backup for the 5% phenomenological estimate and explicit motivation for a follow-on version of the NA62 experiment to measure this decay to greater precision.

The method to compute such second-order effects using lattice QCD was invented by RBC/UKQCD [29]. While this approach is beginning to be used for other second-order processes by other groups, ours is the only group at present which is studying ΔM_K and $K^+ \to \pi^+ \nu \bar{\nu}$ [30, 31, 32, 33]. Because the presence of the charm quark requires a small lattice spacing while working at physical light quark masses requires a large physical volume, these are difficult calculations which are now being undertaken under Incite using Mira. The first, physical mass calculations of ΔM_K were begun in the final year of the 2014-2016 USQCD Incite proposal LatticeQCD_2 and have continued during the first two years of the follow-on 2017-2019 Incite proposal LatticeQCD_3. So far 126 configurations have been measured. During 2018 this calculation was extended to include $K^+ \to \pi^+ \nu \bar{\nu}$ and four configurations have been measured. These first calculations are being performed on a $64^3 \times 128$ ensemble with an inverse lattice spacing $1/a = 2.38$ GeV. We expect both of these calculations to be completed in the final year (2019) of this project on Mira.

The current Summit proposal builds on this work by increasing $1/a$ to 2.774 GeV. Comparing the results at these two values of the lattice spacing will provide an estimate of the size of the finite lattice spacing errors and allow an extrapolation to the continuum limit. With this information a complete estimate of systematic errors for both of these calculations will be possible.

1.3 The anomalous moment of the muon

The creation of this new ensemble with dynamical chiral quarks and small lattice spacing will make possible the calculation of the contributions of the strongly interacting quarks and gluons to the anomalous magnetic moment of the muon (referred to below as $g_\mu - 2$), reducing the systematic error in the result to below 1%. The accurate determination of this quantity within the Standard Model has been a critical goal of particle physics since the E821 experiment [34] reported an experimental measurement differing by more than 3 standard deviations from the Standard Model prediction. A more accurate Standard-Model prediction for this quantity is now even more urgent because a new high-profile experiment is being performed at Fermilab, E989, which expects to reduce the experimental error by a factor of four. A very different experiment E34, expected to have different systematic errors, is being undertaken at J-PARC.

In this proposal we focus on the contribution of the order α_{EM}^2 hadronic vacuum polarization (HVP) contribution to $g_\mu - 2$. A recent lattice QCD calculation of this strong interaction contribution [35] using chiral quarks has a total error of 3%, an error which will be reduced to the 1% level by the end of the year. In order to reduce the error in this calculation below 1% a more accurate continuum extrapolation is required, an

objective of this proposal. (On this scale the current experimental error is 0.9%.) Because of the importance of the Standard Model prediction of $g_\mu - 2$, many calculations using experimental data or lattice QCD calculations (or both in the case of Ref. [35]) are underway. However, the RBC/UKQCD collaboration is the only group exploiting a chiral fermion formulation which allows efficient calculation of the long-distance contribution, a major source of systematic error. This long-distance contribution can be accurately determined by a direct calculation multi-pion intermediate states [35]. The contribution of these light physical pions can be easily identified and accurately determined using chiral fermions.

The calculation of $g_\mu - 2$ within lattice QCD has been a long term goal of RBC/UKQCD, and we have received four ALCC proposal supporting this work: HadronicLight_1 (starting in 2015) through HadronicLight_4 (starting in 2018). These ALCC projects have focused on calculating the contribution of hadronic light-by-light scattering to $g_\mu - 2$, although the more recent proposals have included an HVP component. The current proposal builds upon those, moving to a smaller lattice spacing and larger lattice volume.

1.4   Other opportunities

While not the physics focus of this proposal, it is important to recognize that the ensemble of $96^3 \times 192$ gauge configuration generated as part of this proposal will have great value for several calculations carried out by our collaboration and by many others for at least the next ten years. This will be an ensemble generated with chiral quarks at the smallest lattice spacing yet studied. With earlier $48^3 \times 96$, $1/a = 1.73$ GeV and $64^3 \times 128$, $1/a = 2.38$ GeV ensembles this will be the third ensemble with the same action and physical quark masses but smaller lattice spacing. In fact, part of this proposal is the measurement of standard quantities involving the light, strange and charm quark. These calculations are needed to characterize this new ensemble properly. When combined with earlier results at larger lattice spacing, these results will further improve the accuracy of many important quantities such as f_π, f_K, f_D, f_{D_s} and B_K, to name a few.

2   RESEARCH OBJECTIVES AND MILESTONES

2.1   Objectives and milestones of the heavy-quark-physics program

Here we propose to calculate in lattice QCD the complete set of decay form factors for tree-level and rare $B \to \pi$, $B_s \to K$, and $B \to K$, $B \to D$, $B_s \to D_s$, and $B \to D^*$ transitions, as well as for the tree-level $D \to K$, $D \to \pi$, and $D_s \to K$ transitions. Our uncertainties in the form factors for B-meson decays will be at the 1% level and the sub-percent level for D-meson decays.

The form factors encode the nonperturbative QCD contributions to weak semileptonic decays, and are obtained from hadronic matrix elements of the relevant charged or neutral weak current. Using the $B \to \pi$ process as an example, we have:

$$\langle \pi | V^\mu | B \rangle = f_+(q^2) \left[p_B^\mu + p_\pi^\mu - \frac{M_B^2 - M_\pi^2}{q^2} q^\mu \right] + f_0(q^2) \frac{M_B^2 - M_\pi^2}{q^2} q^\mu \tag{1}$$

$$\langle \pi | S | B \rangle = \frac{M_B^2 - M_\pi^2}{m_b - m_u} f_0(q^2)$$

$$\langle \pi | T^{\mu\nu} | B \rangle = 2 \frac{p_B^\mu p_\pi^\nu - p_B^\nu p_\pi^\mu}{M_B + M_\pi} f_T(q^2) .$$

where V^μ, S, and $T^{\mu\nu}$ are the vector, scalar, and tensor $b \to u(d)$ currents and $q \equiv p_B - p_\pi$. There are three independent form factors: the vector (f_+), scalar (f_0), and tensor (f_T).

In the Standard Model, tree-level semileptonic decays such as $B \to \pi \ell \nu_\ell$ ($\ell = e, \mu, \tau$) are mediated by

Table 1. HISQ gauge-configuration ensembles with strange and charm quark masses set at close to their physical values. The first column gives the lattice spacing for which we were aiming, which could only be determined after the lattices were created. The second column gives the ratio of the simulation mass of the light quark to the physical mass of the strange quark, the third, the lattice dimensions, the fourth, the product of the Goldstone pion mass and the spatial extent of the lattice, and the fifth, the Goldstone pion mass in MeV. The sixth column gives the number of equilibrated gauge configurations as of June 2018. Generation of the $a \approx 0.042$ fm physical-mass and $a \approx 0.03$ fm ensembles is still in progress.

$\approx a$ (fm)	m_l/m_s	$N_s^3 \times N_t$	$M_\pi L$	M_π (MeV)	N_{lattices}
0.15	1/5	$16^3 \times 48$	3.78	306.9(5)	9947
0.15	1/10	$24^3 \times 48$	3.99	214.5(2)	1000
0.15	1/27	$32^3 \times 48$	3.30	131.0(1)	1902
0.12	1/5	$24^3 \times 64$	4.54	305.3(4)	1040
0.12	1/10	$24^3 \times 64$	3.22	218.1(4)	1020
0.12	1/10	$32^3 \times 64$	4.29	216.9(2)	1000
0.12	1/10	$40^3 \times 64$	5.36	217.0(2)	1028
0.12	1/27	$48^3 \times 64$	3.88	131.7(1)	1000
0.09	1/5	$32^3 \times 96$	4.50	312.7(6)	1011
0.09	1/10	$48^3 \times 96$	4.71	220.3(2)	1000
0.09	1/27	$64^3 \times 96$	3.66	128.2(1)	1556
0.06	1/5	$48^3 \times 144$	4.51	319.3(5)	1016
0.06	1/10	$64^3 \times 144$	4.30	229.2(4)	1246
0.06	1/27	$96^3 \times 192$	3.69	135.5(2)	1003
0.042	1/5	$64^3 \times 192$	4.35	309.3(9)	1167
0.042	1/27	$144^3 \times 288$	4.17	134.2(2)	455
0.03	1/5	$96^3 \times 288$	4.84	308.7(1.2)	837

the charged current interaction, and the resulting differential decay rate depends upon both the vector and scalar form factors. For decays to light charged leptons ($\ell = e, \mu$), the contribution from the scalar form factor is suppressed by m_ℓ^2 and hence negligibly small. In contrast, the scalar form-factor contribution to decays into τ leptons is numerically significant. Lattice-QCD calculations of of the form factors for tree-level decays $B \to \pi \ell \nu_\ell$, $B_s \to K \ell \nu_\ell$, $B_{(s)} \to D_{(s)} \ell \nu_\ell$, $D \to \pi \ell \nu_\ell$, and $D \to K \ell \nu_\ell$ enable determinations of the CKM elements $|V_{ub}|$, $|V_{cb}|$, $|V_{cd}|$, and $|V_{cs}|$, from experimental measurements of the corresponding differential decay rates.

Semileptonic flavor-changing neutral-current decays such as $B \to \pi \ell \ell$ and $B \to K \ell \ell$ also receive contributions from the tensor form factor, f_T. Beyond the Standard Model, the same three form factors f_+, f_0, and f_T still suffice to describe the hadronic contribution to the semileptonic transition, but enter the new-physics predictions multiplied by Wilson coefficients which depend on the underlying theory. In the Standard Model, flavor-changing-neutral current decays are mediated by loop interactions, so their rates are much smaller than those of related tree-level decays. Because of this, contributions to these "rare" processes from new particles may be easier to observe above the small Standard-Model rates. For rare decays, lattice-QCD calculated form factors enable Standard-Model predictions for the decay rates, which can then be compared to experimental measurements to look for discrepancies.

Table 2. Estimate of the computational cost for the heavy-quark part of this project. The columns show the approximate lattice spacing a in fm, the ratio of the light sea-quark mass to the strange-quark mass m_l/m_s, the lattice dimensions, the calculation proposed (either generating 190 new gauge-field configurations or calculating form factors from the stated number of configurations times source times (cfg-src)), and the cost in thousands of Summit node-hours.

$\approx a$ (fm)	m_l/m_s	$N_s^3 \times N_t$	calculation	cost (K node-h)
0.042	1/27	$144^3 \times 288$	generate 190 cfg	290
0.042	1/27	$144^3 \times 288$	analyze 2000 cfg-src	920
0.03	1/5	$96^3 \times 288$	analyze 8000 cfg-src	340
Total				1550

Our precision goals for this project are informed by the high precision achieved in our previous work on the kaon, D, and B-meson decay constants [1, 2], and we plan to employ the same methods here. The first key element of Refs. [1, 2] is the set of gluon-field ensembles (listed in Table 1) that include configurations with very fine lattice spacings and with sea quarks at their physical masses. With roughly 1000 configurations (or more) per ensemble, this set has a large statistical power. The second key element is the highly improved fermion formulation, called the HISQ action, used for all sea and valence quarks, including the heavy bottom quark. Based on experience from our previous work on the semileptonic form factors for $B \to \pi$, $B \to K$, $D \to \pi(K)$, and $K \to \pi$ [5, 11, 12, 36, 37, 38], we expect that these elements together will yield exquisite control over all the important sources of systematic error in the lattice calculation of B-meson form factors.

In summary, we expect to obtain these parameters with percent-level errors for pion recoil energies below 1-2 GeV. For rare B-meson decays the high recoil range (with 1 GeV^2 $\lesssim q^2 \lesssim$ 6 GeV^2) is also of high phenomenological interest. There we expect $2 - 3$ times larger errors on the form factors, due to the need to extrapolate the lattice results at low recoil to this kinematic region [5, 12, 11, 13]. Since, as in Refs. [1, 2], we plan to perform our calculation over a range of heavy quark masses, from the charm-quark mass all the way up to the physical b-quark mass on the finest ensembles, the form factors for the corresponding D-meson transitions, $D \to \pi$, $D \to K$, and $D_s \to K$ will be a by-product of this project, and we expect to obtain them with sub-percent precision over entire kinematic range. In order to achieve these goals on the time scale of the experimental program the Summit resources listed in Table 2 will be crucial.

The hadronic matrix elements (and the corresponding semileptonic form factors) are extracted from two- and three-point hadronic correlation functions calculated on the gauge-field ensembles listed above. In order to reach the desired statistical precision, we must calculate the needed correlation functions on each configuration at eight different time sources on each of 1000 configurations per ensemble for each of the ensembles in Table 1. The smallest ensembles will be analyzed on USQCD cluster resources, while, for the medium-sized ensembles, we plan to use a one-year ALCC allocation. The most challenging ensembles are the last two in the table. They are incomplete. We are adding to them with other resources, but need Summit to complete the 0.042 fm physical-mass ensemble. Based on our Summit benchmarks and on calculations on Mira, we estimate that it costs 1500 Summit node-h to add one new gauge configuration to the 0.042 physical-mass ensemble. We have budgeted for 190 such configurations.

As shown in the accompanying Milestone Table document, we plan to start with the form-factor calculation on the smaller of the two ensembles, $a \approx 0.03$ fm with $N_s^3 \times N_t = 96^3 \times 288$ in year 1. At the same time we will generate approximately 110, 0.042 fm $144^3 \times 288$ gauge configurations. In the second year we will

begin the analysis of the 0.042 fm gauge configurations and generate 80 new ones. In the final year, we will complete the analysis of the proposed total of 2000 configuration/source-time for the 0.042 fm ensemble. We hope to reach the goal of 8000 with Exascale resources in approximately three years.

We provide, here, more detail for the cost estimate for calculating the needed correlation functions on a single source time t_{source} and a single gauge configuration. That cost is given by

$$T = T_{2pt} + T_{3pt} \tag{2}$$

where T_{2pt} is the cost of calculating the needed hadronic two-point functions, based on the light and heavy quark propagators and T_{3pt} is the cost of calculating the hadronic three-point functions from extended heavy-quark propagators. These times break down further, as

$$T_{2pt} = 4T_{\text{full}} \quad T_{3pt} = 32T_{\text{heavy}}, \tag{3}$$

where T_{full} is the cost of a single refined multishift solve including all light and heavy quarks and T_{heavy} is the cost of a single refined multishift solve with only heavy quarks. These costs are obtained from our Summit benchmarks and from known Flop counts. For example, on the 0.03 fm ensemble with $m_l/m_s = 1/27$ and $N_s^3 \times N_t = 96^3 \times 192$ we estimate $T_{\text{full}} = 6$ Summit node-h and $T_{\text{heavy}} = 0.53$ Summit node-h, giving a total of $T = 42.5$ node-h per configuration/source-time. When this figure is multiplied by the required number of configuration/source-times, we get the cost estimate shown in Table 2. A similar calculation applies to the 0.042 fm ensemble.

2.2 Objectives and milestones of the chiral fermion calculations

The proposed chiral fermion calculations rely on the existence of the new ensemble of $96^3 \times 192$ gauge configurations with the increased inverse lattice spacing of $1/a = 2.744$. This is an improved version of an existing ensemble of $48^3 \times 96$ gauge configurations with the same small lattice spacing but a heavier-than-physical light-quark mass: ensemble F1 in Ref. [39]. The generation of this ensemble is part of our Summit Early Science proposal, and the current proposal assumes that this Summit ESP is successful. While the 1000 molecular dynamics time units planned in that ESP should be sufficient to begin our proposed physics calculations, this is an important ensemble with many other uses, and there is also some risk that the Early Science proposal may not be as successful as we expect. Therefore, we propose to continue to extend this ensemble, adding 200 time units each year, a goal which appears in the table of milestones. If the first 500 time units of the initial ESP ensemble are discarded for thermalization, these additional 600 time units will more than double the value of this ensemble.

The next component of this proposal is the calculation of two-point Green's functions which will determine the masses of the lowest-energy two- and three-quark states as well as physical quantities of interest such as f_π, f_D, etc. These will be used to determine more accurately the scale and required physical quark masses for this ensemble and are listed as "standard observables" in years one and two in the table of milestones. The first year will provide the accuracy needed to support the calculations of ΔM_K and $K^+ \rightarrow \pi^+ \nu \bar{\nu}$, while the second year will allow publication quality results for the basic quantities being studied. The milestones associated with these calculations are the preliminary results after year one and final results at the end of the second year.

Our goal for the calculation of ΔM_K is a 20% statistical error at this smaller lattice spacing and the resulting control of the finite lattice spacing error that will come from a comparison between this new calculation and the Mira INCITE calculation that is currently underway. This is a difficult calculation because disconnected graphs with large statistical noise make a substantial contribution. This is not the case for calculation of

Table 3. Computational cost in thousands of Summit node hours for the chiral-quark part of this project.

quantity	number of gauge configurations	cost (K node-h)
ensemble generation	600	250
standard observables	36	237
ΔM_K	78	460
$K^+ \to \pi^+ \nu \bar{\nu}$	20	207
$g_\mu - 2$ HVP	66	396
Total		1550

$K^+ \to \pi^+ \nu \bar{\nu}$. Here the disconnected diagrams are relatively unimportant and we hope to achieve a statistical error of 2-3% of the total branching ratio from less than 1/2 the number of samples. Both calculations are very demanding and require the power of Summit applied consistently over three years to complete the proposed matrix element evaluation.

The goal of the calculation of the HVP contribution to $g_\mu - 2$ is a reduction of the current total error from the $\approx 1\%$ that we expect to achieve by the end of 2018, to less than 0.5%. This can be accomplished by performing the proposed calculations on about 20 gauge configurations for each of the three years of this proposal, as listed in the table of milestones.

We list the estimated computational cost of the five components of the chiral quark portion of this proposal in Tab. 3. For the case of top four lines, these estimates were obtain from the execution times and measured code performance on Mira scaled up by the increased lattice size. For the fifth line we use an explicit count for the required operations. In all cases we assume a sustained performance on Summit of 4 Tflops per node. This assumption is supported by the 1.6 Tflops/node obtained on 1024 summit nodes and described later in this section. For both the ensemble generation and the calculation of observables we expect to increase this network-limited, 1.6 Tflops/node number at least 4.0 Tflops/node. For the measurements we will use the communication avoiding split grids algorithm (described below) which increases the execution speed on Cori by 3×. In fact, by using the blocked CG algorithm and performing inversions for multiple right hand sides together, we expect an additional 2× acceleration. These methods do not work for the ensemble generation task because of an insufficient number of inversions that can be performed at the same time. However, the MSPCG algorithm, also described below as well as the availability of RDMA on Summit should provide the assumed performance increase.

3 COMPUTATIONAL READINESS

For the heavy-quark project we use the MILC code [40] in conjunction with the QUDA library [41] for GPUs. For the chiral-quark project we use the CPS code [42], also in conjunction with QUDA. All of these codes are mature and have been in production use for many years for similar projects.

Most of the heavy-quark project is an "ensemble" project in INCITE-RFP terminology. However, its sheer size easily makes it a capability project. We have divided the needed gauge-field ensembles listed in Table 1 into four resource categories according to lattice size and computational difficulty: cluster, petascale (Mira, Theta), multipetascale (Summit), and exascale. As indicated in the Milestone Table and the resource request in Sec. 2, for Summit, we propose to analyze completely the two most computationally demanding ensem-

bles, namely part of the 0.042 fm physical-sea-quark mass $144^3 \times 288$ ensemble and the 0.03 fm $96^3 \times 288$ ensemble with light sea-quark mass equal to 1/5 the strange quark mass. We are also requesting time to generate 190 new 0.042 fm lattices, a capability project that may require full use of 1024 nodes. A complete analysis of the 0.042 fm ensemble will probably require Exascale resources.

The proposed calculations involving chiral quarks are demanding, capability jobs that easily require the resources of Summit. At the heart of this set of calculations is the generation of an ensemble incorporating chiral dynamical quarks with the largest space-time volume created to date. This job is best run as a single stream in which each configuration in the ensemble is generated from the one before as a traditional Markov chain. Generating a new configuration can be done at the rate of 2.5 per hour on a 1024-node partition. If the LookAhead algorithm is successful and allows up to five molecular dynamics (MD) time units to be combined to reduce Metropolis rejection, such a 5 MD time unit trajectory would require 2 hours. We propose to generate 200 MD time units per year. The ΔM_K and $K^+ \to \pi^+ \nu \bar{\nu}$ calculations might be run at a slower rate on smaller partitions, since they need not be run sequentially. However, when performed on a single configuration these calculations are expected to require up to 18 and 25 hours respectively on a 1024-node partition, so a reasonable time to solution will require a Summit partition of at least that size. The HVP calculation is more flexible but can be run efficiently on a large Summit partition.

3.1 Use of Resources Requested

3.1.1 Heavy-quark project

Our benchmarks suggest that the optimum heavy-quark form factor job will need full use of six GPUs on 128 to 512 nodes for up to twelve hours, depending on the lattice size. There will be several thousand such jobs with a steady burn rate. Since run times are uniform, jobs can be bundled efficiently to suit queue requirements. We will also use the burst buffer for intermediate storage of eigenvectors and propagators – approximately 20 TB to process a single lattice. The heavy-flavor gauge-configuration-generation component will require 1024 nodes and take approximately 1.5 hours to produce a new gauge configuration.

Requirements for I/O external to a job are very modest. We will be reading and possibly rereading 49 to 248 GByte gauge-configuration files in each job. These files will be transferred with Globus from archives at Fermilab and staged for analysis at Summit. Correlator output will be transferred for post-processing elsewhere. The only longer-term storage requirements could be for eigenvectors used for deflation. How much storage will be decided based on the economics of tape storage *vs.* the expected future savings in computer time. The total storage cost could amount to 7 PB over the three years of this project.

3.1.2 Calculations with chiral fermions

The run plan needed to execute the portion of this proposal using chiral fermions is straight forward. The first calculation must focus on generating the gauge field configurations. This task executes as a sequence of 1024-node jobs, each taking approximately two hours per 5 MD time unit trajectory, assuming that the LookAhead method is used. Each such configuration produced would be stored as a 5.4 Gbyte file. Since the ΔM_K, $K^+ \to \pi^+ \nu \bar{\nu}$ and HVP calculations require the same set of 5.5K eigenvectors, these would be computed as part of the ΔM_K or HVP jobs and stored on disk as 50 Tbyte files to be used by a subsequent $K^+ \to \pi^+ \nu \bar{\nu}$ job. Note, for chiral fermions these 5.5K eigenvectors would conventionally require 0.5 Pbytes of storage each. However, we can exploit local coherence to achieve a 10× compression [43]. These eigenvector files represent valuable data and will be archived if possible. Both the ΔM_K and $K^+ \to \pi^+ \nu \bar{\nu}$ jobs are performed in two varieties. The ΔM_K jobs are evenly divided between low precision jobs requiring 4 hours and 19 hour high precision jobs, if both are run on 1024 Summit nodes. The proportions are different for the $K^+ \to \pi^+ \nu \bar{\nu}$ job. If run on 1024 Summit nodes four fifths of the jobs will be 19 hour low-precision

jobs and one fifth will be 25 hour, high-precision jobs. All four types of jobs will produce modest data files that can be easily transferred elsewhere for analysis. The g-2 HVP calculation will generate 4 TB of data per configuration and time-slice that will be stored to disk for further processing. Having 100 TB of disk-space buffer would make the processing of the generated data comfortable. If tape-storage space allows, we could store data long-term for further processing by other projects, however, this is not needed for the proposed g-2 HVP calculation. The jobs for the g-2 HVP calculations will run on 1024 nodes for four hours each.

3.2   Computational Approach

3.2.1   *Underlying mathematical formulation:*

In lattice QCD, physical observables are expressed in terms of Feynman path integrals on regular four-dimensional, hyper-cubic lattices,

$$\langle O \rangle = \frac{\int \mathcal{D}U \, O(U) \, \exp\left[-S(U)\right]}{\int \mathcal{D}U \, \exp\left[-S(U)\right]} \; . \tag{4}$$

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 matrix 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

$$P(U_i) \sim \exp\left[-S(U_i)\right] \; . \tag{5}$$

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

$$\langle O \rangle = \frac{1}{N} \sum_{i}^{N} O(U_i) \; . \tag{6}$$

3.2.2   *Algorithms and numerical techniques:*

The act of integrating out the quark fields makes $S(U)$ highly nonlocal since it involves the determinants of the Dirac operators for the "sea" quarks. Integration over the gauge fields uses sophisticated importance sampling. The preferred method uses a molecular dynamics algorithm to sample the dominant part of the integration space. One must integrate numerically a set of nonlinear, coupled first order differential equations. To treat the sea quarks, at each step of the integration one must solve a set of linear equations of the form $(M + \sigma_j I)x_j = y_j$, where I is the unit matrix, the σ_j are real positive numbers, and M is either the Dirac operator for a quark, a large sparse matrix of dimension of order the volume of the lattice, or the product of the Dirac operator with its Hermitian conjugate. Solving this system, along with the calculation of the force that enters into the integration of the molecular dynamics equations, consumes the bulk of the floating point operations in configuration generation.

Our gauge evolution codes incorporate all the state-of-the-art methods which are successful in accelerating the generation of independent chiral fermion ensembles. We use the fourth-order, force-gradient integrator and exploit well-developed tools to determine a pattern of Hasenbusch masses which will lead to fermion forces of approximately equal size from each Hasenbusch ratio. A particular concern when generating a

gauge ensemble with small lattice spacing is the slow evolution of topological charge. As can be seen in Fig. 15 of Ref. [39], a finite auto-correction time of 60 ± 2 was found for the topological charge with our proposed action and lattice spacing. We are currently tuning the planned chiral-quark evolution code with the LookAhead algorithm [44], which promises to shorten auto-correlation times by a factor of two or more.

The generation of the 0.042 fm HISQ gauge configurations follows similar computational methods. The calculation of the heavy-quark form factors involves generating propagators for the quarks involved and tying them together with appropriate operators. We use a variety of state-of-the-art techniques for reducing the variance, including the use of stochastic sources to improve volume averaging, low-mode deflation to speed the calculation of the light-quark propagators, and the truncated solver method that exploits correlations between low and high-precision solutions.

In the calculation of all of the chiral fermion observables low-mode deflation is used to accelerate the Dirac matrix inversions with the 5500 low modes shared between the four sets of calculations. These eigenvectors are used both to deflate the CG inversions and to construct the all-to-all propagators needed to evaluate diagrams with a closed fermion loop, allowing the vertex where the quark is emitted and absorbed to be summed over the entire lattice volume.

For both ΔM_K and $K^+ \to \pi^+ \nu \bar{\nu}$ significant noise reduction strategies developed previously will be used. These include the "sample AMA" procedure in which inexact propagators are computed on most gauge configuration while an accurate correlated correction is determine on a few. In addition the product form of the noisy disconnected amplitudes is exploited to substantially increase the statistic at limited cost.

Critical to the HVP calculation is obtaining accurate results when the two electromagnetic (EM) currents which enter the HVP calculation are separated by a large distance. Here we will apply GEVP (generalized eigenvalue problem) and distillation methods using as many as 200 3-dimensional Laplace eigenvectors as sources on each time slice to perform a direct calculation of the matrix elements of an EM-current operator between the vacuum and a multipion state with a well-resolved energy. With this detailed spectral information we can explicitly determine the long-distance HVP contribution as a series of terms decreasing exponentially in the current-current separation with an increasing exponential fall-off.

Finally an important improvement over our previous large-volume studies will be achieved by averaging over the temporal locations of the sources as part of the final analysis. While the accuracy of the final average will not be affected, separating the results according to location will allow us to identify a much larger statistical sample [45] and obtain a significantly more reliable error estimate.

3.2.3 *Programming languages, libraries and other software*

As mentioned above, two major code bases and several supporting libraries will be used for calculations on Summit, namely, the MILC code [40] for the heavy-quark project and the CPS code [42] for lattice generation and analysis. The most time-consuming parts of the calculation, namely the sparse-matrix solvers and the other components of gauge-field molecular dynamics (gauge force, fermion force) are carried out on multiple GPUs using the QUDA (QCD in CUDA) library [41] developed by Kate Clark and collaborators for NVIDIA GPUs [46, 47, 48, 49]. This library is written with detailed attention to the capabilities of the architecture. It uses many sophisticated optimization strategies, including autotuning, data compression, use of texture tools, and a variety of communication-avoidance techniques. QUDA supports GPUDirect, enabling peer-to-peer communication between GPUs sharing a PCIe bus. Autotuning optimizes the assignment of thread blocks to the streaming multi-processors. QUDA is under continual improvement. For example, at present we are developing adaptive multigrid algorithms for the solvers used in this project.

Table 4. Weak-scaling performance for (left) the MILC/QUDA multimass HISQ (heavy-quark) solver with a local volume of 32^4 per GPU and (right) the DWF (chiral) inverter with a local volume of 16×12^3 per GPU. Columns give the number of nodes, the performance in Gflops (HISQ) or Tflops (DWF) per node for the full node and for the CPUs alone, the time to solution in seconds divided by the number of nodes for both cases. For the DWF benchmarks in the CPU-only case we have run only a single node benchmark. The other figures, identified with *, are upper/lower bounds assuming perfect weak scaling.

nodes	Gflops/node		Time/(# nodes)	
	GPU	CPU	GPU	CPU
16	432	50	0.75	6.4
32	426	47	0.38	3.4
64	306	44	0.31	1.8
128	300	44	0.133	0.9
256	282		0.070	
512	258	43	0.039	0.23
1024	228	43	0.021	0.12

(a) HISQ (heavy-quark

nodes	Tflops/node		Time/(# nodes)	
	GPU	CPU	GPU	CPU
1	8.85	0.21	11.9	427.
8	2.38	0.21^*	5.35	53.4^*
64	1.99	0.21^*	0.80	6.68^*
512	1.59	0.21^*	0.12	0.834^*
920	1.63	0.21^*	0.07	0.464^*
1024	1.55	0.21^*	0.06	0.417^*

(b) DWF (chiral)

### 3.2.4 *Parallel programming model, workflow, I/O requirements*

Our lattices are divided into identical sub-lattices, each of which is assigned to a node. Communication between nodes is handled by MPI, either directly or indirectly via the SciDAC-developed QMP message-passing system. The codes also support threading with OpenMP and parallel I/O.

For both projects, the generation of gauge configurations proceeds in a continuous sequence of calculations, each depending on the previous one. For the heavy-quark analysis, most of the needed gauge-field configurations have already been generated, so we are requesting time for generating only a few. For the form-factor calculation, the workflow consists in reading the gauge-field configurations in any order and carrying out a calculation that produces several thousand small hadron correlators for each gauge configuration. In the course of the analysis, a couple hundred intermediate propagator files are saved temporarily and reread. Depending on the ensemble and our workflow scheduling, they could amount to up to 20 TB. We will also be generating eigenvectors for deflated solves. Typically, one thousand such eigenvectors are needed per gauge configuration. The eigenvector storage requirement for the lattice sizes in this project is a few Terabytes per gauge configuration. We have found that a burst buffer is especially effective for both intermediate-propagator and eigenvector storage. We estimate that intermediate I/O will account for approximately 20% of the project cost.

For the calculation of chiral quark observables, all of the needed gauge configurations will be available at the OLCF. These are multi-hour jobs with parallel I/O needed to write or read a 50 Tbyte compressed eigenvector file which should be a modest overhead on Summit.

3.3 Parallel Performance

All of our proposed heavy-quark physics calculations use our MILC code [40] with the QUDA GPU package[41]. We have used this combination extensively in production running on the BNL Institutional Cluster, which features NVIDIA Tesla K80s and Pascal P100s. QUDA is also optimized for Volta V100s.
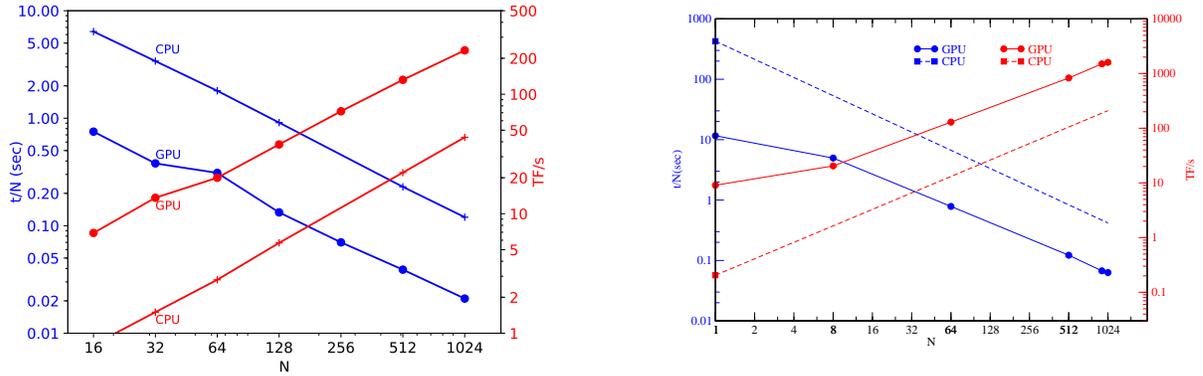
Figure 2. Left: Weak scaling performance of the MILC/QUDA multimass HISQ solver showing (left scale, blue) time to solution divided by node count and (right scale, red) total teraflops/s *vs.* the node count with local volume 32^4. Circles show GPU performance and plusses, CPU. Right: A similar plot for the CPS/QUDA DWF solver for the $16 \times 12^3 \times 12$ local volumes per GPU required to study the $96^3 \times 192$ lattice on 1024 Summit nodes.

3.3.1 *Heavy-quark code performance: MILC/QUDA*

During early access time on Summit we were able to run a few benchmarks with our code. We present results for the most expensive part of the calculation, namely, the repeated solution of the Dirac equation for the light-quark propagator. In production running, this operation accounts for more than 2/3 of the run time. More specifically, we present results for the QUDA solver for our preferred "highly improved staggered quark" (HISQ) formulation of lattice quarks. We have found that it delivers approximately 300 GF/s per node in double-precision when running on a 128-node problem with a local lattice volume of 32^4 per GPU, a size that is optimal for the proposed calculation. Considering the inherently low computational intensity (0.7 flops per byte) of our fermion formulation, this is good performance. We expect substantial improvements when RDMA is enabled on Summit and we optimize the intranode subdivision of our lattices. Nonetheless, even now, our results show convincingly that we can make efficient use of Summit for the proposed projects.

In Fig. 2 and Table 4, we show weak scaling performance of the MILC/QUDA code using our standard MILC-code benchmark problem with the multi-mass QUDA HISQ solver. The time to solution is defined as the time to complete 600 conjugate-gradient iterations. (The typical production run requires hundreds of thousands of such solutions.) We use this benchmark because it is amenable to varying the problem size and number of nodes. For the CPU benchmark we used six MPI ranks per node with seven threads per rank so as to balance the workload among the 42 SMCs. It is clear from Fig. 2 that the QUDA multimass HISQ solver scales well. Furthermore, the GPU boosts performance by a factor of 4 to 6.

3.3.2 *Chiral fermion code: CPS/QUDA*

The code which will initially be used for this project is a combination of the Columbia Physics System (described in Ref. [42] and on GitHub (CPS)) and the QUDA code developed and maintained by Kate Clark (NVIDIA) and collaborators [46, 47, 48, 49]. The CPS code provides a flexible, high-performance environment in which a broad range of lattice QCD code is available including state-of-the-art sparse matrix solvers and algorithms needed to generate the ensemble of gauge configurations proposed here. The CPS code, built upon Peter Boyle's Bagel code generator, has delivered 6 Pflops performance on Sequoia and been run in production to perform lattice QCD calculations on 32 racks of Mira sustaining in excess of 1

Pflops.

In order to run on Summit or other GPU-based machines, we have replaced the BG/Q optimized code by QUDA code, a project carried out four years ago. The combination is a flexible and effective system to allow efficient use of the heterogeneous Summit resource. While not essential for our use of Summit, we are also devoting substantial effort to evolve the Grid framework by Peter Boyle and collaborators [50] so that it will execute with high efficiency on GPU-based machines. When completed in perhaps 6 months this will provide a modern C++11 environment and significantly expand the code that is available to run with high efficiency on Summit. With support from NVIDIA (Kate Clark) and the OLCF we have already had considerable success in tuning our code to run efficiently on the six-GPU Summit node achieving a sustained 1.6 Pflops for our target local volume on 1024 Summit nodes using a double-precision conjugate gradient solver that exploits half-precision code internally. The weak scaling plot is shown in Fig. 2 and numerical results in Tab. 4. Also shown is the performance and time-to-solution of optimized CPU code which was run on a single node and then weak scaling used to set the rigorous upper and lower bounds given.

Since more than 90% of the computer cycles required by both the ensemble generation and the proposed Green's function calculations are consumed by the Dirac matrix solver we have limited our benchmarks to this critical component. However, we have also compiled the elaborate gauge evolution code and run it with adequate efficiency on Summit nodes so we expect no difficulties in running our full application.

3.4 Developmental Work

Early access to Summit has allowed us to ready our Summit code for all heavy-quark milestones. However, we are continually improving the MILC and QUDA codes. Among the planned improvements is a revision to the communications layer, QMP, that will support optimum intranode partitioning of the lattice. We are also improving our analysis framework to support sparse-matrix solves that combine multiple right-hand sides. This strategy should improve on-node performance. Since RDMA has not yet been enabled on Summit, we have not been able to include it in our tuning and benchmarks. However, that may become possible over summer 2018. Finally, we are involved in an Exascale Computing Project effort to develop a multigrid solver for the HISQ fermion formulation. It promises significant performance improvement. We hope it will become available in the later years of this project.

The six-GPU, 1.6 Tflops/node performance represents substantial development effort carried out this Spring. In addition to incorporating RDMA when available and continuing to tune of our QUDA code, there are three further algorithmic improvements which offer substantial speedups. For ensemble generation we will employ the Multisplitting Preconditioned Conjugate Gradient (MSPCG) [51] which promises a greater than 2× acceleration of the conjugate gradient code by using an on-node preconditioner capable of exploiting Summit's very large, local floating point performance. The porting of this code from Mira to Summit is presently underway and will be tested as soon as Summit becomes available. For the calculation of both $g_\mu - 2$, ΔM_K and $K^+ \rightarrow \pi^+\nu\bar{\nu}$ we will use the "split grids" algorithm which remaps the large Summit partition on which the first deflation steps of the Dirac matrix solver must be performed, so that later stages in the conjugate gradient algorithm can be performed for multiple right-hand sides on smaller sub-partitions. Finally, we will employ the new LookAhead Hybrid Monte Carlo algorithm (LAHMC) [44] to reduce critical slowing down and allow useful results to be obtained from Monte Carlo samples separated by fewer MD time units. LAHMC gauge evolution using the parameters for our Summit job is currently running on Mira to determine the optimal parameter choice.

4   REFERENCES

[1] **Fermilab Lattice, MILC** Collaboration, A. Bazavov *et. al. Phys. Rev.* **D90** (2014), no. 7 074509 [`1407.3772`].

[2] **Fermilab Lattice, MILC** Collaboration, A. Bazavov *et. al.*, *B- and D-meson leptonic decay constants from four-flavor lattice QCD, submitted to Phys. Rev. D* (2017) [`1712.09262`].

[3] **MILC** Collaboration, J. A. Bailey *et. al.*, *B → Dℓν form factors at nonzero recoil and |V_cb| from 2+1-flavor lattice QCD*, *Phys. Rev.* **D92** (2015), no. 3 034506 [`1503.07237`].

[4] **Fermilab Lattice, MILC** Collaboration, J. A. Bailey *et. al.*, *Update of |V_cb| from the B̄ → D* ℓν̄ form factor at zero recoil with three-flavor lattice QCD*, *Phys. Rev.* **D89** (2014), no. 11 114504 [`1403.0635`].

[5] **Fermilab Lattice, MILC** Collaboration, J. A. Bailey *et. al.*, *|V_ub| from B → πℓν decays and (2+1)-flavor lattice QCD*, *Phys. Rev.* **D92** (2015), no. 1 014024 [`1503.07839`].

[6] **HFLAV** Collaboration, Y. Amhis *et. al. Eur. Phys. J.* **C77** (2017), no. 12 895 [`1612.07233`].

[7] **BaBar** Collaboration, P. del Amo Sanchez *et. al. Phys. Rev.* **D83** (2011) 032007 [`1005.3288`].

[8] **BaBar** Collaboration, J. P. Lees *et. al. Phys. Rev.* **D86** (2012) 092004 [`1208.1253`].

[9] **Belle** Collaboration, H. Ha *et. al. Phys. Rev.* **D83** (2011) 071101 [`1012.0090`].

[10] **Belle** Collaboration, A. Sibidanov *et. al. Phys. Rev.* **D88** (2013), no. 3 032005 [`1306.2781`].

[11] **Fermilab Lattice, MILC** Collaboration, J. A. Bailey *et. al.*, *B → πℓℓ form factors for new-physics searches from lattice QCD*, *Phys. Rev. Lett.* **115** (2015), no. 15 152002 [`1507.01618`].

[12] J. A. Bailey *et. al.*, *B → Kl⁺l⁻ decay form factors from three-flavor lattice QCD*, *Phys. Rev.* **D93** (2016), no. 2 025026 [`1509.06235`].

[13] D. Du, A. X. El-Khadra, S. Gottlieb, A. S. Kronfeld, J. Laiho, E. Lunghi, R. S. Van de Water and R. Zhou, *Phenomenology of semileptonic B-meson decays with form factors from lattice QCD*, *Phys. Rev.* **D93** (2016), no. 3 034005 [`1510.02349`].

[14] **Fermilab Lattice, MILC** Collaboration, A. Bazavov *et. al.*, *B⁰_(s)-mixing matrix elements from lattice QCD for the Standard Model and beyond*, *Phys. Rev.* **D93** (2016), no. 11 113016 [`1602.03560`].

[15] **LHCb** Collaboration, R. Aaij *et. al. JHEP* **06** (2014) 133 [`1403.8044`].

[16] **LHCb** Collaboration, R. Aaij *et. al. JHEP* **10** (2015) 034 [`1509.00414`].

[17] **BaBar** Collaboration, J. P. Lees *et. al. Phys. Rev. Lett.* **109** (2012) 101802 [`1205.5442`].

[18] **LHCb** Collaboration, R. Aaij *et. al. Phys. Rev. Lett.* **113** (2014) 151601 [`1406.6482`].

[19] **LHCb** Collaboration, R. Aaij *et. al. Phys. Rev. Lett.* **115** (2015), no. 11 111803 [`1506.08614`]. 159901(E) (2015).

[20] **LHCb** Collaboration, R. Aaij *et. al. JHEP* **02** (2016) 104 [`1512.04442`].

[21] **Belle** Collaboration, Y. Sato *et. al. Phys. Rev.* **D94** (2016), no. 7 072007 [`1607.07923`].

[22] **Belle** Collaboration, S. Hirose *et. al. Phys. Rev. Lett.* **118** (2017), no. 21 211801 [`1612.00529`].

[23] **Belle** Collaboration, S. Wehle *et. al. Phys. Rev. Lett.* **118** (2017), no. 11 111801 [`1612.05014`].

[24] J. A. Bailey *et. al. Phys. Rev. Lett.* **109** (2012) 071802 [`1206.4992`].

[25] J. Bennett *J. Phys. Conf. Ser.* **770** (2016), no. 1 012044.

[26] **Belle II** Collaboration, M. Lubej in *52nd Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 2017. `1705.05289`.

[27] **LHCb** Collaboration, G. Wilkinson, *LHCb—selected highlights and future prospects*, *Subnucl. Ser.* **53** (2017) 413–428.

[28] J. Brod and M. Gorbahn, *Next-to-Next-to-Leading-Order Charm-Quark Contribution to the CP Violation Parameter ϵ_K and ΔM_K*, *Phys. Rev. Lett.* **108** (2012) 121801 [`1108.2036`].

[29] **RBC, UKQCD** Collaboration, N. H. Christ, T. Izubuchi, C. T. Sachrajda, A. Soni and J. Yu, *Long distance contribution to the KL-KS mass difference*, *Phys. Rev.* **D88** (2013) 014508 [`1212.5931`].

[30] Z. Bai, N. H. Christ, T. Izubuchi, C. T. Sachrajda, A. Soni and J. Yu, *$K_L - K_S$ Mass Difference from Lattice QCD*, *Phys. Rev. Lett.* **113** (2014) 112003 [`1406.0916`].

[31] **RBC, UKQCD** Collaboration, N. H. Christ, X. Feng, A. Portelli and C. T. Sachrajda, *Prospects for a lattice computation of rare kaon decay amplitudes II $K \to \pi\nu\bar{\nu}$ decays*, *Phys. Rev.* **D93** (2016), no. 11 114517 [`1605.04442`].

[32] N. H. Christ, X. Feng, A. Jüttner, A. Lawson, A. Portelli and C. T. Sachrajda, *First exploratory calculation of the long-distance contributions to the rare kaon decays $K \to \pi\ell^+\ell^-$*, *Phys. Rev.* **D94** (2016), no. 11 114516 [`1608.07585`].

[33] Z. Bai, N. H. Christ, X. Feng, A. Lawson, A. Portelli and C. T. Sachrajda, *Exploratory Lattice QCD Study of the Rare Kaon Decay $K^+ \to \pi^+\nu\bar{\nu}$*, *Phys. Rev. Lett.* **118** (2017), no. 25 252001 [`1701.02858`].

[34] **Muon g-2** Collaboration, G. W. Bennett *et. al.*, *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys. Rev.* **D73** (2006) 072003 [`hep-ex/0602035`].

[35] **RBC, UKQCD** Collaboration, T. Blum, P. A. Boyle, V. Gãijlpers, T. Izubuchi, L. Jin, C. Jung, A. Jüttner, C. Lehner, A. Portelli and J. T. Tsang, *Calculation of the hadronic vacuum polarization contribution to the muon anomalous magnetic moment*, `1801.07224`.

[36] **Fermilab Lattice, MILC** Collaboration, E. Gamiz *et. al.*, *Kaon semileptonic decays with $N_f = 2 + 1 + 1$ HISQ fermions and physical light-quark masses*, *PoS* **LATTICE2016** (2016) 286 [`1611.04118`].

[37] **Fermilab Lattice, MILC** Collaboration, T. Primer *et. al.*, *D meson semileptonic form factors with HISQ valence and sea quarks*, *PoS* **LATTICE2016** (2017) 305.

[38] A. Bazavov *et. al. Phys. Rev. Lett.* **112** (2014), no. 11 112001 [`1312.1228`].

[39] P. A. Boyle, L. Del Debbio, A. Jüttner, A. Khamseh, F. Sanfilippo and J. T. Tsang, *The decay constants* f_D *and* f_{D_s} *in the continuum limit of* $N_f = 2 + 1$ *domain wall lattice QCD*, *JHEP* **12** (2017) 008 [`1701.02644`].

[40] **MILC** Collaboration, C. DeTar *et. al.*, "MILC Collaboration code for lattice QCD calculations." `https://github.com/milc-qcd/milc_qcd`.

[41] K. Clark *et. al.*, "QUDA library for performing calculations in lattice QCD on GPUs." `https://github.com/lattice/quda`.

[42] **RBC, UKQCD** Collaboration, C. Jung, *Overview of Columbia Physics System*, *PoS* **LATTICE2013** (2014) 417.

[43] M. A. Clark, C. Jung and C. Lehner, *Multi-Grid Lanczos*, *EPJ Web Conf.* **175** (2018) 14023 [`1710.06884`].

[44] G. Cossu, P. Boyle, N. Christ, C. Jung, A. Jüttner and F. Sanfilippo, *Testing algorithms for critical slowing down*, *EPJ Web Conf.* **175** (2018) 02008 [`1710.07036`].

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

[46] G. Shi, S. Gottlieb, A. Torok and V. Kindratenko, *Design of MILC lattice QCD application for GPU clusters* , in *Proceedings of the 2011 IEEE Parallel and Distributed Processing Symposium, IPDPS '11. IEEE, 2011* , 2011.

[47] G. Shi, R. Babich, M. Clark, B. Joó, S. Gottlieb and V. Kindratenko, *The Fat-Link Computation On Large GPU Clusters for Lattice QCD*, in *Proceedings of the 2012 IEEE Parallel and Distributed Processing Symposium, IPDPS '12. IEEE, 2012* , 2012.

[48] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower and S. Gottlieb, *Scaling Lattice QCD beyond 100 GPUs*, in *SC11 International Conference for High Performance Computing, Networking, Storage and Analysis Seattle, Washington, November 12-18, 2011*, 2011. `1109.2935`.

[49] R. Babich, R. Brower, M. Clark, S. Gottlieb, B. Joo and G. Shi, *Progress on the QUDA code suite*, *PoS* **LATTICE2011** (2011) 033.

[50] P. Boyle *et. al.*, "Data parallel C++ mathematical object library." `https://github.com/paboyle/Grid`.

[51] D. Guo, R. Mawhinney and J. Tu, *Solving DWF Dirac Equation Using Multisplitting Preconditioned Conjugate Gradient*, `1804.08593`.