B_K and related matrix elements with unquenched, improved staggered fermions

Chulwoo Jung (Brookhaven National Lab)
Weonjong Lee (Seoul National University)
Stephen Sharpe (University of Washington)

March 11, 2011

Abstract

This is a class B proposal, requesting time on Infiniband-connected GPU clusters, to continue our calculation of kaon matrix elements using HYP-smeared improved staggered quarks. (The project web site is
`http://www.phys.washington.edu/users/sharpe/qcdoc/index.html`.)

We have made major progress during the current year using USQCD resources, adding a fourth lattice spacing and improving statistics. Our total error in $\widehat{B}_K$ has been reduced from 7.2% to 4.9% in the last year. The dominant error is from the truncation of the matching factor, and is being addressed in another USQCD proposal (Lytle and Sharpe) and in work at Seoul National University (SNU). Here we propose to use GPUs to continue reducing the statistical errors. These are presently at the $\sim 1\%$ level, but we want to lower them to the sub-percent level, which will allow us to better test our estimates of systematic errors due to fitting and chiral extrapolations.

A second focus of this year's proposed work will be to extend our analysis to matrix elements of four-fermion operators other than that needed to calculate B_K. These matrix elements can be used to constrain new physics contributions to kaon mixing. This extension needs perturbative matching factors (recently completed) and staggered chiral perturbation theory calculations.

Graduate students and postdoctoral associates at both the University of Washington and SNU will contribute to this project.

We are requesting 100,000 GPU-hours on a GPU clusters connected with Infiniband, and 170,000 Jpsi-equivalent core hours for storage.

1 Scientific Background and Overview

A major goal of lattice QCD is to calculate electroweak matrix elements. This would allow precision tests of the standard model (SM), and is one of the major objectives highlighted in the USQCD white papers and proposals. Here we propose to continue our study of B_K and related hadronic matrix elements using improved staggered fermions.

For the last two years we have submitted "type A" proposals so that we could bring our calculation of B_K on the MILC asqtad lattices to completion. This we have nearly done, with the addition of a fourth lattice spacing during the current year. The 2011-12 allocation year will be a period of transition for our calculation—in which we shift the focus to new quantities, improve our codes on the GPUs, and tool-up for using HISQ lattices (with HISQ valence quarks). We also think it worthwhile to reduce the statistical error in B_K on some of the ensembles. Thus we think it appropriate that we scale back to a "type B" proposal for this year.

Lattice calculations of B_K have been highly successful. Results using different fermion discretizations are in good agreement, with the present average (taken from Laiho, Lunghi and Van de Water [1]) being

$$\widehat{B}_K = 0.737 \pm 0.020\,. \tag{1}$$

One input into this average is our published result [2]

$$\widehat{B}_K = 0.724 \pm 0.012 \text{ (stat)} \pm 0.043 \text{ (syst)}\,, \tag{2}$$

which uses HYP-smeared [3] improved staggered fermions on coarse, fine and superfine MILC ensembles. Simulations during the last year, largely carried out with USQCD resources, allow us to now include the the ultrafine ensemble with $a \approx 0.045$ fm, and to improve statistical errors. Our updated (preliminary) result is

$$\widehat{B}_K = 0.719 \pm 0.009 \text{ (stat)} \pm 0.034 \text{ (syst)}\,; \tag{3}$$

more details will be presented below. Our calculation has the advantage over others of being based on four lattice spacings. Although the dominant error is from the truncation of the matching factor (4.4%), statistical and fitting errors are each $\sim 1\%$.

The lattice result for B_K provides an important constraint on the CKM matrix. At present, while all such constraints are consistent, there are some tensions. These provide motivation for further reducing the errors in B_K and other lattice inputs. In addition, if, as is hoped, the LHC finds evidence for new physics, it is important to have lattice results for matrix elements of $\Delta S = 2$ operators having all Dirac structures, so as to determine whether models of new physics are consistent with the constraints coming from $\bar{K} - K$ mixing. Although the required accuracy is not as great as for B_K, results with $5 - 10\%$ errors are desirable.

For both of these reasons we propose to continue our present project, although with a change of focus. Clearly, reducing the matching error is a priority, and this is addressed in a separate proposal by Lytle and Sharpe, and by an ongoing work by the group at Seoul National University (SNU). At the same time, we think it important to improve the statistics of our B_K calculation so that they are the same level on all ensembles, and to do further tests by adding new ensembles (e.g. a second superfine ensemble allows a further test of sea-quark mass dependence). In addition, throughout our calculation of matrix elements, we have, in tandem with B_K, obtained the tree-level matrix elements for all other Dirac structures. We have only recently, however, completed the one-loop matching calculation for these other operators [4], and are now working out the fitting forms predicted by staggered chiral perturbation theory. We plan to focus a significant part of our effort this coming year on obtaining results for these other matrix elements .

In previous years we have worked primarily on the QCDOC, and indeed we continue adding measurements on the ultrafine lattices using this resource. Last year we rewrote our core codes in CUDA, and did some running on GPUs. This proposal is entirely for time on GPUs, and is specifically for increasing the statistics on the superfine ensembles. We will do parallel work on the fine and coarse ensembles on a GPU cluster at SNU. We are gradually optimizing our codes. We are also beginning development of code for HISQ fermions, which will open up the possibility of calculating matrix elements involving charm quarks.

Since we use the MILC asqtad configurations, we must assume that the use of a rooted fermion determinant leads to the correct continuum limit. This is plausible, based on the extensive numerical results and theoretical work summarized in Ref. [5, 6, 7].

2 Methods

We use valence HYP-smeared quarks on MILC lattices.[1] Staggered fermions have the advantage of a remnant $U(1)$ partially conserved axial symmetry, which guarantees that the lattice matrix elements satisfy Ward identities analogous to those in the continuum [8]. We use HYP-smearing because it substantially reduces the taste-breaking in the spectrum (by the same amount as the HISQ action) [9], it is computationally simpler than both asqtad and HISQ fermions, and perturbative corrections to matching factors for bilinears and four-fermion operators are greatly reduced [10, 4]. We have also found that HYP-smeared fermions reduce discretization errors in general [11]. We note that we are using a mixed action, with different versions of staggered fermions for the valence and sea quarks. The impact of this can be taken into account using chiral perturbation theory.

On each (suitably chiral) MILC configuration we use 10 different valence quark masses, running approximately from $m_s^{\rm phys}$ down to $m_s^{\rm phys}/10$, and calculate B_K and the Goldstone-taste "kaon" mass for all 55 mass combinations. B_K is defined by

$$\frac{8}{3} m_K^2 f_K^2 B_K = \langle \bar{K} | \bar{s}\gamma_\mu(1-\gamma_5)d\bar{s}\gamma_\mu(1-\gamma_5)d|K\rangle \,. \tag{4}$$

The methodology for calculating B_K is well established [11, 2] and we do not repeat it here.

We use gauge-invariant operators having spin-taste structure $(\Gamma\otimes\xi_5)(\Gamma\otimes\xi_5)$, with all 16 choices of Dirac matrices Γ. An important feature is that the links in the operators are HYP-smeared, and mean-field improved.[2] This reduces perturbative corrections and may also reduce the size of discretization errors [12].

Suitable subsets of the resulting data are then fit to forms predicted by staggered chiral perturbation theory (SChPT) [13]. We have found fits based on $SU(2)$ SChPT to give smaller errors than those using $SU(3)$ fits. Details of our methodology are given in Refs. [2], [14] and [15].

---

[1] For much of the following, we use B_K as a shorthand for "B_K and the related four-fermion kaon matrix elements discussed above".

[2] Mean-field improvement is implemented here by measuring the average plaquette composed of HYP-smeared links, and then dividing all HYP links by the fourth root of the plaquette. For HYP links, the resulting $u_0 \approx 0.97$, so the impact of mean-field improvement is relatively small.

3 Code and production details

Our GPU code is written in CUDA, and developed on the SNU clusters of NVidia GTX 295 GPUs and Fermi Tesla GPUs (with nodes connected by Infiniband). So far we only have converted the inverter (conjugate gradient) to CUDA. The B_K calculation is run on the CPUs and is a relatively small overhead (see below).

Our present inverter attains 12.5 Gflop/s/GPU in double-precision when running $28^3 \times 96$ fine lattices on a 4 node partition with 1 Fermi Tesla GPU/node. This is at the upper end of the range we predicted in last year's proposal. On the superfine $48^3 \times 144$ lattices, however, running on 4 nodes with 4 Fermi Tesla GPUs/node, the performance drops to about 6 Gflop/s/GPU. This is related to our use of two-dimensional parallelization. We think that we can use GPU-MPI direct technology to regain the lost factor of 2.

With our present code, the inversions (20 in total) on a single superfine lattice take 140 minutes on 16 GPUs, while the calculation of B_K takes about 40 minutes (running on the CPUs). Thus one measurement costs 48 GPU-hours.

We have fairly frequent repeat failure error messages, and so, at present, have been running a repeat test for each configuration, doubling the effective cost.

We have been running solely on the JLAB 32-node cluster equipped with 4 NVidia C2050 (Fermi Tesla) GPUs/node, in order to make use of the fast QDR Infiniband (since our code is communications limited). Our progress has been slowed by long waits in the queues. Our jobs run on 16 GPUs, and, queueing 6 jobs in parallel. we have only been able to obtain a throughput of 400 GPU-hours/day. We began production in mid-November 2010, and to date have used about 28 K GPU-hours. At the present rate of running, we will use most, but not all, of our allocation. (We have about 70 K GPU-hours left, so in 110 days, we expect to use $110 \times 400\text{hrs} \times 1.2 = 55\text{Khrs}$, where 1.2 is the new charge factor for the Fermi Tesla GPUs.)

4 Progress in 2010-11

Table 1 summarizes the present state of our B_K analysis, as well as the additional lattices that will be available by the end of this allocation year.

We were allocated 20 M QCDOC-node hours, and approximately 100K GPU-hrs (the latter in two stages). The QCDOC allocation was for continuing our calculation of B_K on the ultrafine lattices. This running has been successful. So far in this allocation year, we have used about 16.2 M QCDOC node-hrs, and added 474 measurements of B_K (somewhat more than the 386 we proposed). We are still running, and hope to reach ≈ 1000 measurements in total (since we have about 450 from the previous year). The results presented below are based on 705 measurements.

On the GPU's, we have so far added 221 measurements on the S1 ensemble, and, at our present rate of progress, will add an additional ≈ 450 with the remainder of our allocation (listed as "rest of alloc" in the table). These have not yet been included in our B_K analysis.

We have also significantly increased the number of measurements on several other ensembles on the SNU GPU-clusters, leading to the rows marked "update" in the table.

The addition of the fourth lattice spacing, and, to a lesser extent, the improvement in statistics, has led to a significant reduction in the error. As an example of the quality of the data on the ultrafine lattices, we show the "X-fit" and "Y-fit" in Figs. 1 and 2, respectively.

ID	a (fm)	am_l/am_s	Size	Configs.	status
C1	0.12	0.03/0.05	$20^3 \times 64$	564×1	old
C2	0.12	0.02/0.05	$20^3 \times 64$	486×9	update
C3	0.12	0.01/0.05	$20^3 \times 64$	671×9	old
C3-2	0.12	0.01/0.05	$28^3 \times 64$	275×8	old
C4	0.12	0.007/0.05	$20^3 \times 64$	651×10	old
C5	0.12	0.005/0.05	$24^3 \times 64$	509×9	update
F1	0.09	0.0062/0.031	$28^3 \times 96$	995×9	update
F2	0.09	0.0031/0.031	$40^3 \times 96$	853×1	update
S1	0.06	0.0036/0.018	$48^3 \times 144$	744×2	old
S1	0.06	0.0036/0.018	$48^3 \times 144$	744×1.05	rest of alloc
S1	0.06	0.0036/0.018	$48^3 \times 144$	744×2.6	**Proposed**
U1	0.045	0.0030/0.015	$64^3 \times 192$	705×1	new
U1	0.045	0.0030/0.015	$64^3 \times 192$	300×1	rest of alloc

Table 1: Status of B_K analysis of MILC ensembles. "Old" indicates used in analysis presented in Ref. [2], which was based on data mainly generated in previous allocation years. "Update" indicates an increase in statistics during the present allocation year (typically a 9-fold increase in the number of measurements), which has been included in the analysis. "New" indicates a new ensemble included in the analysis. "Rest of alloc" indicates additional lattices that will be available by the end of the present allocation year. Finally, "proposed" indicates the lattices we propose to calculate in 2011-12.

Here X_P is the squared mass of the pion composed of valence $d\bar{d}$, and, in the X-fit, we are extrapolating m_d^{val} to its physical value (for kaons in which the strange valence quark mass is fixed at our largest value). In the Y-fit, we extrapolate (linearly) the valence strange quark mass to its physical value.

In Fig. 3 we give an update on the dependence on the light sea-quark mass, once chiral loops have been taken into account. This dependence, expected to be linear, is very mild.

Finally, in Fig. 4, we show the continuum extrapolation using four lattice spacings. Clearly the extra lattice spacing reduces the error in the extrapolated value. Even more important, however, is that it reduces the impact of the truncation error, since this is estimated as $\alpha_s^2(1/a)$ for the smallest lattice spacing.

We have quoted the result from this analysis in the Introduction. Here we present, in table 2, the updated error budget.

5 Proposed running in 2011-12

Our goal is to reduce the statistical error on the superfine ensemble S1. We have doing multiple measurements on each configuration, differing by the source positions, leads to results that are statistically independent.

As noted above, our present code loses a factor of 2 in performance when run on the superfine lattices, but we think that we can regain this factor. Thus we will make our estimates assuming a speed of 50 GPU-hours per measurement. We are also in process of developing a mixed-precision

cause	error (%)	description
matching factor (*)	4.4	α_s^2 on (U1)
statistics (*)	1.9	4X3Y-NNLO fit
discretization (*)	0.94	diff. of (U1) and $a = 0$
fitting (1)	0.92	X-fit
fitting (2)	0.08	Y-fit
am_ℓ extrap (*)	0.06	(C3) versus linear extrap
am_s extrap	0.5	constant vs. linear extrap
finite volume (*)	0.59	finite vol. vs. $V = \infty$ fit
r_1	0.14	r_1 error budget
f_π	0.38	132 MeV vs. 124.4 MeV

Table 2: Present status of error budget for B_K (preliminary). Estimates marked with an asterisk are updates from our published paper [2].

inverter. This should lead to substantial speed-ups, but we will not include this in our estimate.

We propose to do 2000 measurements on the S1 ensemble, leading to an estimate of **100 K GPU-hours**. (If the cost of using Fermi Tesla GPU's is increased by 1.2, then we will reduce the number of measurements accordingly.) Any further speed-ups will allow us to move closer to our goal of 9 measurements per configuration.

Our major need for storage is gauge configurations. The simplest mode of running is to keep all HYP smeared configurations available on disk for multiple calculations while running a given ensemble. For the S1 ensemble, the 743 lattices require 7 Tbytes of storage. Adding in temporary space for propagators (which will not be stored long term unless there is a demand), we request 10 Tbytes of disk space.

According to the conversion factors in the call for proposals, this converts to 1.7×10^5 **Jpsi-equivalent core-hours for disk storage**.

We are presently running at Jlab, and it would be simplest to continue running there, but we are also open to moving to the new cluster at FNAL, as long as the nodes are connected by fast Infiniband.

References

[1] J. Laiho, E. Lunghi and R. S. Van de Water, Phys. Rev. D **81**, 034503 (2010) [arXiv:0910.2928 [hep-ph]], with updated values from `http://krone.physik.unizh.ch/ lunghi/webpage/LatAves/`

[2] T. Bae, *et al.*, Phys. Rev. **D82**, 114509 (2010). [arXiv:1008.5179 [hep-lat]].

[3] A. Hasenfratz and F. Knechtli, Phys. Rev. D **64**, 034504 (2001) [arXiv:hep-lat/0103029].

[4] J. Kim, W. Lee, S. R. Sharpe, [arXiv:1102.1774 [hep-lat]].

[5] S. R. Sharpe, PoS **LAT2006**, 022 (2006) [arXiv:hep-lat/0610094].

[6] A. S. Kronfeld, PoS **LAT2007**, 016 (2007) [arXiv:0711.0699 [hep-lat]].

[7] M. Golterman, PoS C **ONFINEMENT8**, 014 (2008) [arXiv:0812.3110 [hep-ph]].

[8] G. W. Kilcup and S. R. Sharpe, Nucl. Phys. B **283**, 493 (1987).

[9] Taegil Bae, David H. Adams, Chulwoo Jung, Hyung-Jin Kim, Jongjeong Kim, Kwang-woo Kim, Weonjong Lee and Stephen R. Sharpe, Phys. Rev. D **77**, 094508 (2008) [arXiv:0801.3000 [hep-lat]].

[10] W. Lee and S. R. Sharpe, Phys. Rev. D **68**, 054510 (2003) [arXiv:hep-lat/0306016].

[11] W. Lee, T. Bhattacharya, G. T. Fleming, R. Gupta, G. Kilcup and S. R. Sharpe, Phys. Rev. D **71**, 094501 (2005) [arXiv:hep-lat/0409047].

[12] W. Lee, PoS **LAT2006**, 015 (2006) [arXiv:hep-lat/0610058].

[13] W. J. Lee and S. R. Sharpe, Phys. Rev. D **60**, 114503 (1999) [arXiv:hep-lat/9905023];
C. Bernard [MILC Collaboration], Phys. Rev. D **65**, 054031 (2002) [arXiv:hep-lat/0111051];
C. Aubin and C. Bernard, Phys. Rev. D **68**, 034014 (2003) [arXiv:hep-lat/0304014]; *ibid* **68**, 074011 (2003) [arXiv:hep-lat/0306026].

[14] J. Kim, C. Jung, H.J. Kim, W. Lee and S.R. Sharpe [arXiv:1101.2685 [hep-lat]].

[15] B. Yoon, Y.C. Jang, C. Jung, W. Lee, [arXiv:1101.2248 [hep-lat]].

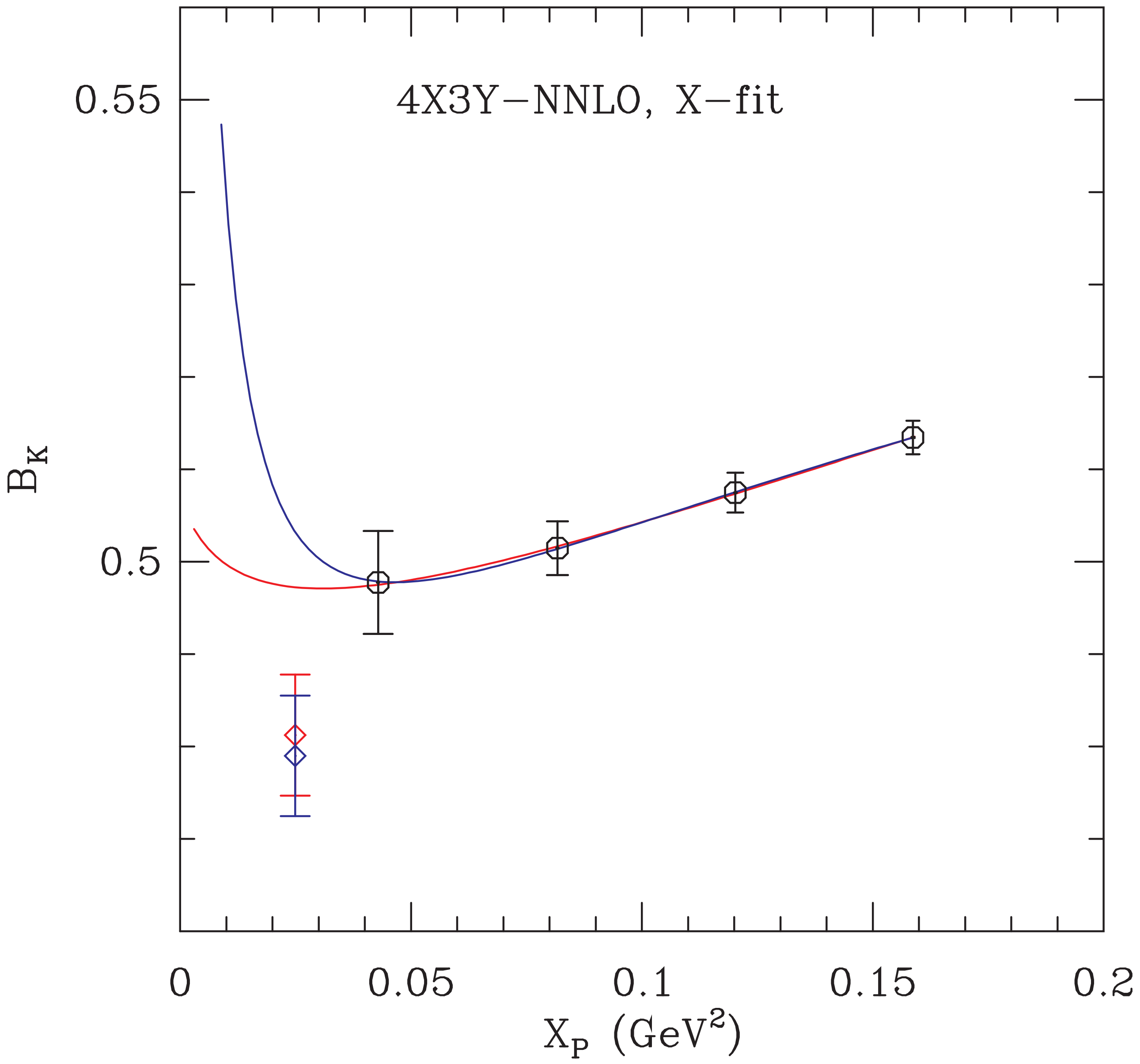

Figure 1: $B_K(1/a)$ versus X_P on superfine ensemble U1, for $am_s^{\rm val} = 0.014$. A fit to next-to-leading order SChPT with one analytic next-to-next-to-leading order term is shown. The blue curve shows the fit to the finite volume SChPT form, while the red curve is to the infinite volume form. The two lower points are determined by the fit once taste-breaking is removed, and the quark masses are set to physical values. We use the finite volume fit (corrected for finite volume effects) to give our central value.

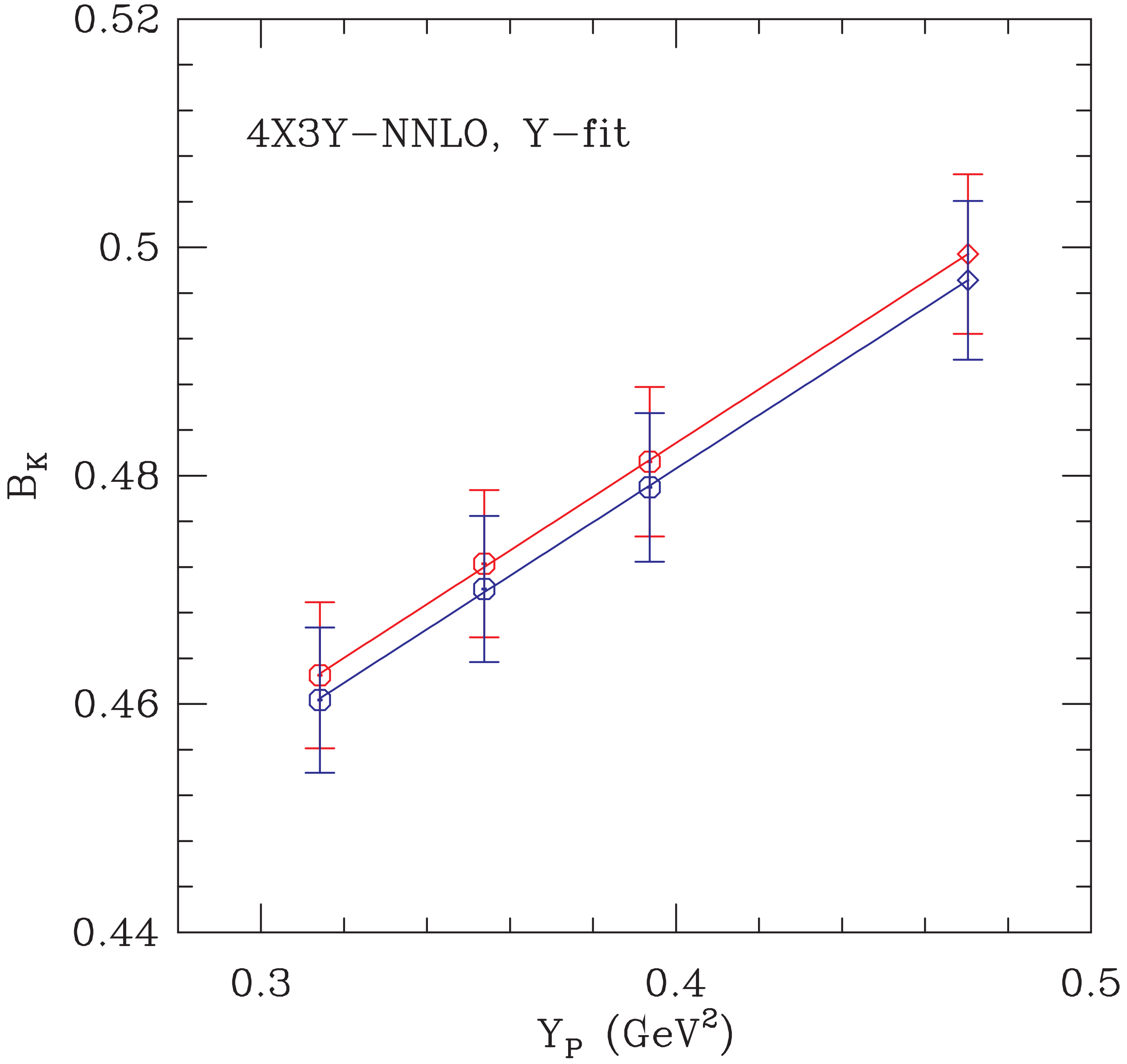

Figure 2: $B_K(1/a)$ from the fit shown in the previous figure, extrapolated in the squared mass of the valence $s\bar{s}$ meson, Y_P. A linear extrapolation to $m_s^{\rm val} = m_s^{\rm phys}$ (the rightmost points, shown as diamonds) is shown. Blue and red points and curves correspond, respectively, to those from the finite volume and infinite volume fits.

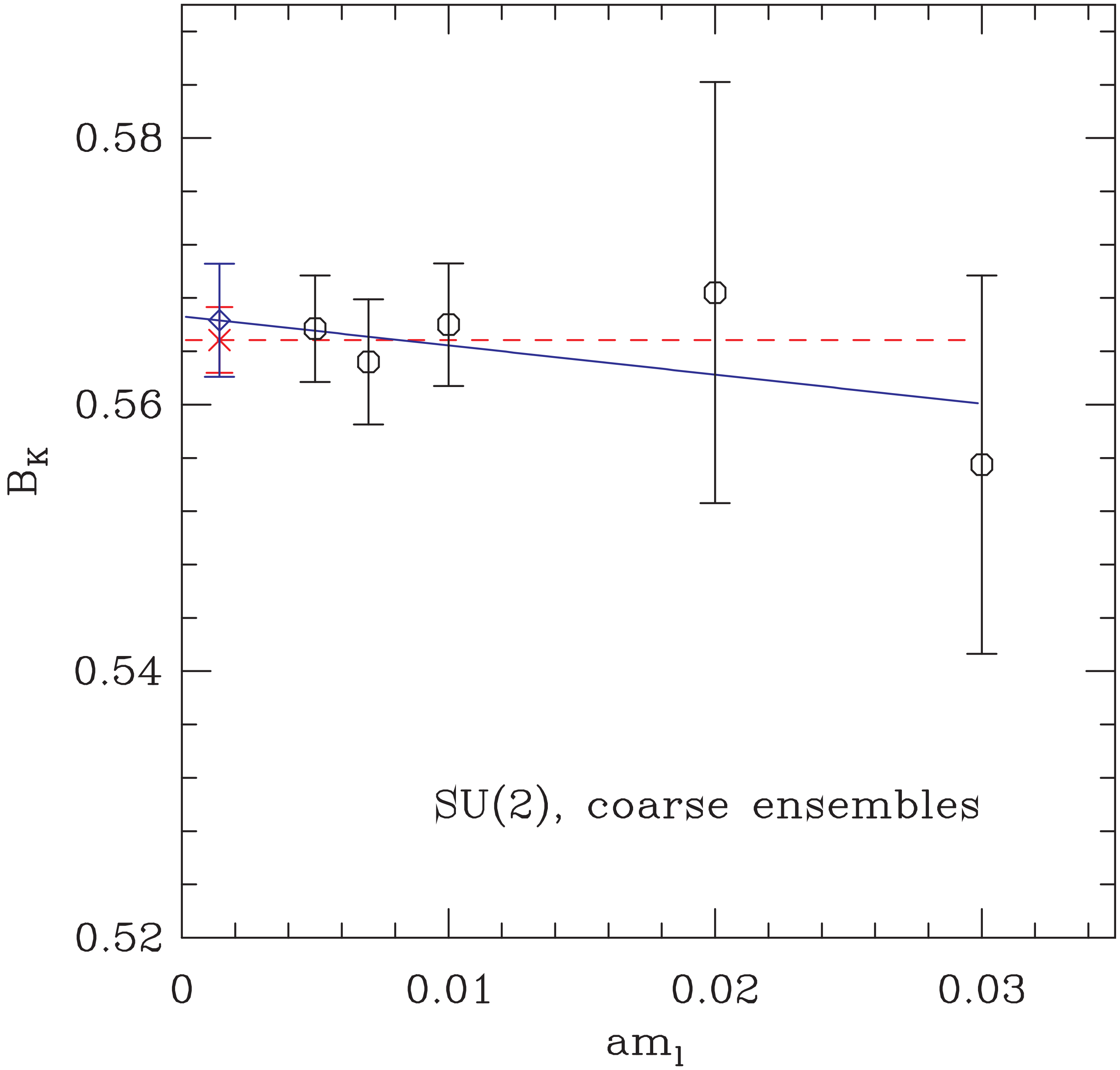

Figure 3: $B_K(\mu = 2\,\mathrm{GeV})$ from 4X3Y-NNLO SU(2) fits versus am_l (light sea-quark mass) on MILC coarse ensembles, using both a constant (red line) and linear (blue line) fit to extrapolate to the physical light quark mass.

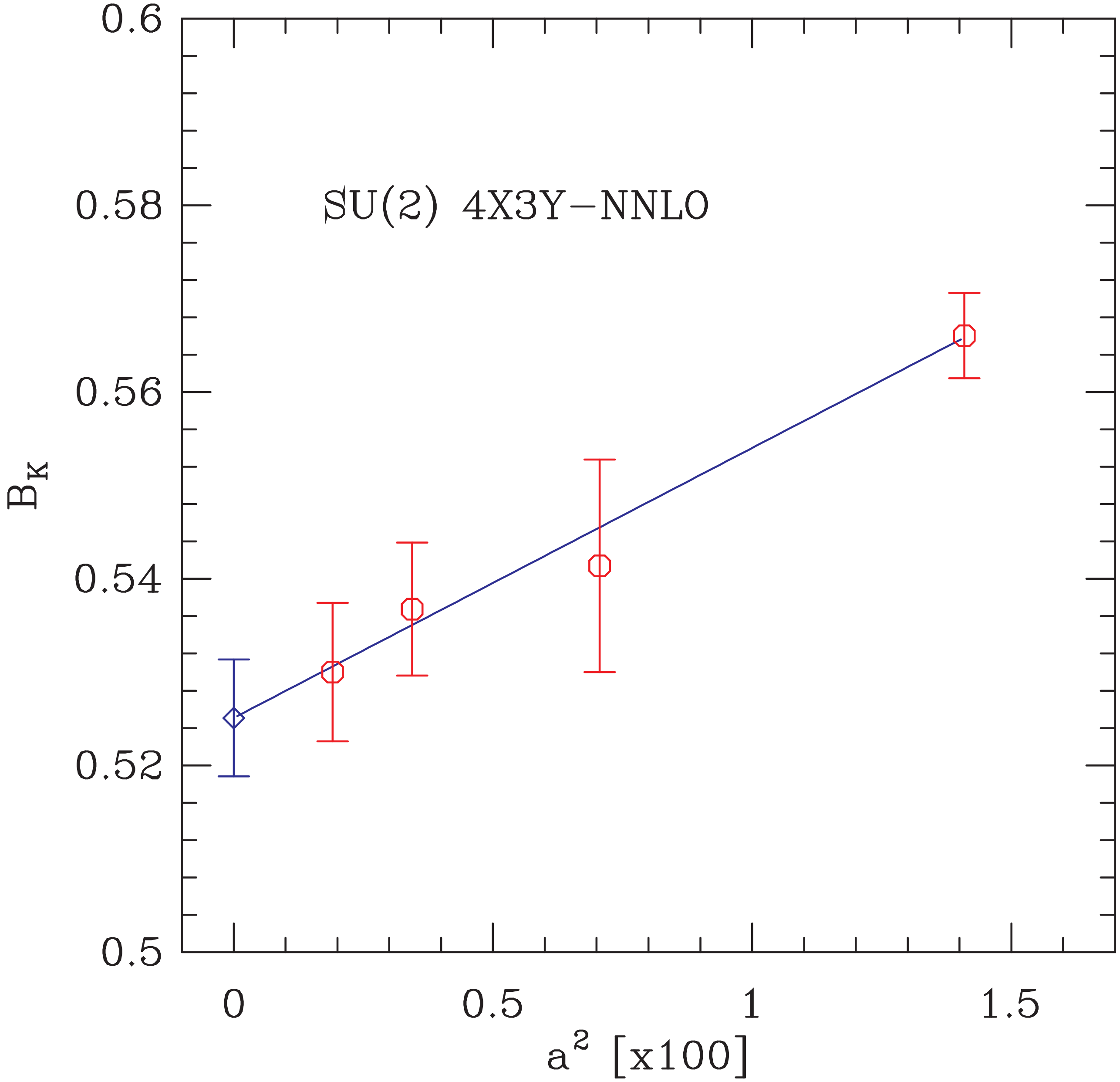

Figure 4: $B_K(2\,\mathrm{GeV})$ versus a^2, together with a linear fit to the continuum limit. The data is obtained using SU(2) 4X3Y-NNLO fits.