



2-15-2018

# Variance Reduction and Cluster Decomposition

Keh-Fei Liu

*University of Kentucky*, [kehfei.liu@uky.edu](mailto:kehfei.liu@uky.edu)

Jian Liang

*University of Kentucky*, [jli364@uky.edu](mailto:jli364@uky.edu)

Yi-Bo Yang

*Michigan State University*

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

Follow this and additional works at: [https://uknowledge.uky.edu/physastron\\_facpub](https://uknowledge.uky.edu/physastron_facpub)



Part of the [Elementary Particles and Fields and String Theory Commons](#)

## Repository Citation

Liu, Keh-Fei; Liang, Jian; and Yang, Yi-Bo, "Variance Reduction and Cluster Decomposition" (2018). *Physics and Astronomy Faculty Publications*. 567.

[https://uknowledge.uky.edu/physastron\\_facpub/567](https://uknowledge.uky.edu/physastron_facpub/567)

This Article is brought to you for free and open access by the Physics and Astronomy at UKnowledge. It has been accepted for inclusion in Physics and Astronomy Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

**Variance Reduction and Cluster Decomposition****Notes/Citation Information**

Published in *Physical Review D*, v. 97, issue 3, 034507, p. 1-7.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

**Digital Object Identifier (DOI)**

<https://doi.org/10.1103/PhysRevD.97.034507>

## Variance reduction and cluster decomposition

Keh-Fei Liu,<sup>1</sup> Jian Liang,<sup>1</sup> and Yi-Bo Yang<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506, USA*

<sup>2</sup>*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*



(Received 3 June 2017; published 15 February 2018)

It is a common problem in lattice QCD calculation of the mass of the hadron with an annihilation channel that the signal falls off in time while the noise remains constant. In addition, the disconnected insertion calculation of the three-point function and the calculation of the neutron electric dipole moment with the  $\theta$  term suffer from a noise problem due to the  $\sqrt{V}$  fluctuation. We identify these problems to have the same origin and the  $\sqrt{V}$  problem can be overcome by utilizing the cluster decomposition principle. We demonstrate this by considering the calculations of the glueball mass, the strangeness content in the nucleon, and the  $CP$  violation angle in the nucleon due to the  $\theta$  term. It is found that for lattices with physical sizes of 4.5–5.5 fm, the statistical errors of these quantities can be reduced by a factor of 3 to 4. The systematic errors can be estimated from the Akaike information criterion. For the strangeness content, we find that the systematic error is of the same size as that of the statistical one when the cluster decomposition principle is utilized. This results in a 2 to 3 times reduction in the overall error.

DOI: [10.1103/PhysRevD.97.034507](https://doi.org/10.1103/PhysRevD.97.034507)

### I. INTRODUCTION

As the physical pion mass is accessible in lattice QCD simulations nowadays with larger physical volumes and several lattice spacings with different lattice actions to estimate the associated systematic errors, lattice QCD calculation is getting mature, particularly for flavor physics where the quark masses, heavy-light decay constants, Cabibbo-Kobayashi-Maskawa (CKM) matrices, and strong coupling constant are reviewed and averaged by FLAG [1]. On the other hand, the baryon physics is not as settled as that of mesons. Part of the reasons is illustrated in the Paris-Lepage consideration of the signal-to-noise ratio of the nucleon propagator has three-pions as the lowest state in the correlator, the signal-to-noise (S/N) ratio is proportional to  $e^{-(m_N-3/2m_\pi)t}$  [2,3] and noticeably grows exponentially with  $t$  when the pion mass is close to the physical one in lattice calculations. This is why baryon physics is more noisy than that of mesons.

One special problem associated with the correlators of mesons involving annihilation channels or glueballs is that the signal falls off exponential with time, but the noise remains constant. Thus, after certain time separation, the signal falls below the noise and succumbs to the sign

problem. Another aspect of the disconnected insertions (DI) is observed in the DI three-point function involving a quark loop or the topological charge in the neutron electric dipole moment (nEDM) calculation from the  $\theta$  term, the fluctuations of the quark loop and the topological charge are proportion to  $\sqrt{V}$  which pose a challenge for calculations large volumes lattices. In this work, we shall show that the constant error and  $\sqrt{V}$  fluctuation in the DI have the same origin and they can be ameliorated with the help of the property of the cluster decomposition principle so that the S/N ratio can be improved by a factor of  $\sqrt{V/V_{R_s}}$  where  $V_{R_s}$  is the volume with radius  $R_s$  which is the effective correlation length between the operators.

### II. CLUSTER DECOMPOSITION PRINCIPLE AND VARIANCE REDUCTION

One often invokes the locality argument to justify that experiments conducted on Earth is not affected by events on the Moon. This is a consequence of the cluster decomposition principle (CDP) in that if color-singlet operators in a correlator are separated by a large enough space-like distance, the correlator will be zero. In other words, the operators are not correlated in this circumstance. To be specific, it is shown [4] that under the assumptions of translation invariance, stability of the vacuum, existence of a lowest nonzero mass and local commutativity, one has

$$|\langle 0 | \mathcal{B}_1(x_1) \mathcal{B}_2(x_2) | 0 \rangle_s| \leq A r^{-\frac{2}{3}} e^{-Mr} \quad (1)$$

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

for a large enough spacelike distance  $r = |x_1 - x_2|$ , where  $\langle 0 | \mathcal{B}_1(x_1) \mathcal{B}_2(x_2) | 0 \rangle_s \equiv \langle 0 | \mathcal{B}_1(x_1) \mathcal{B}_2(x_2) | 0 \rangle - \langle 0 | \mathcal{B}_1(x_1) | 0 \rangle \times \langle 0 | \mathcal{B}_2(x_2) | 0 \rangle$  is the vacuum-subtracted correlation function.  $\mathcal{B}_1(x_1)$  and  $\mathcal{B}_2(x_2)$  are two color-singlet operator clusters whose centers are at  $x_1$  and  $x_2$  respectively,  $M$  is the smallest nonzero mass for the correlator, and  $A$  is a constant. This is the asymptotic behavior of a boson propagator  $K_1(r)/r$ . This means the correlation between two operator clusters far apart with large enough space-like distance  $r$  tends to be zero at least as fast as  $r^{-\frac{3}{2}} e^{-Mr}$ . Given that the longest correlation length in QCD is  $1/m_\pi$ , one has  $M \geq m_\pi$ . Since the Euclidean separation is always “space-like,” the cluster decomposition principle (CDP) is applicable to the Euclidean correlators. Some of the recent attempts to reduce variances in the calculations of strangeness in the nucleon [5], the  $\rho$  meson mass [6], the light-by-light contribution in the muonic  $g - 2$  [7], the factorization of fermion determinant [8], and reweighting of nEDM calculation with topological charge density [9] have applied concepts similar or related to that of CDP. In this work, we prove that, applying the CDP explicitly, the error of an DI correlator can be improved by a factor of  $\sqrt{V/V_{R_s}}$ .

In evaluating the correlators, one often takes a volume sum over the three-dimensional coordinates. To estimate at what distance the large distance behavior saturates, we integrate the fall-off to a cut off distance  $R$ ,

$$\int_0^R d^3 r r^{-\frac{3}{2}} e^{-Mr} = 4\pi \left( \frac{\sqrt{\pi} \text{erf}(\sqrt{MR})}{2M^{\frac{3}{2}}} - \frac{\sqrt{R} e^{-MR}}{M} \right), \quad (2)$$

where erf is the error function. Since the kernel of the integral decays very quickly, the integral has already gained more than 99.5% of its total value for  $R = 8/M$ . Assuming the fall-off behavior dominates the volume-integrated correlator, we consider  $R_s \sim \frac{8}{M}$  as an effective cutoff and the correlation with separation  $r > R_s$  has negligible contribution.

To test the principle of cluster decomposition with lattice data, we consider the two-point correlator for a fixed  $t$  with a cutoff of  $R$  in the relative coordinate between the two color-singlet operators  $\mathcal{O}_1$  and  $\mathcal{O}_2$

$$C(R, t) = \frac{1}{V} \left\langle \sum_{\vec{x}} \sum_{r < R} \mathcal{O}_1(\vec{x} + \vec{r}, t) \mathcal{O}_2(\vec{x}, 0) \right\rangle, \quad (3)$$

where  $r = \sqrt{|\vec{r}|^2 + t^2}$ . The correlation functions in the present work are calculated using valence overlap fermions on the RBC-UKQCD 2 + 1 flavor domain-wall configurations. More detailed definitions and numerical implementations can be found in previous works [10–13].

We examine the nucleon two-point function first on the  $48^3 \times 96$  lattice (48I) with the physical sea quark mass [14]. We use 3 valence quark masses corresponding to pion masses 70 MeV, 149 MeV and 260 MeV respectively.

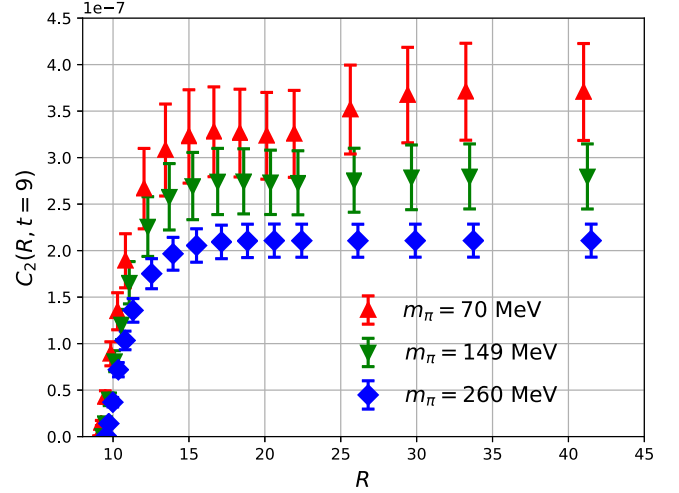


FIG. 1. Nucleon two-point functions at  $t = 9$  for three different valence quark masses as a function of the cutoff  $R$ .

The results for the nucleon correlators at  $t = 9$  for three different valence quark masses are plotted in Fig. 1 as a function of  $R$  which is the cutoff of the Euclidean distance  $r$  between the point source and the sink. We see that the nucleon correlator basically saturates after  $R \sim 15 = 1.71$  fm with  $a = 0.114$  fm for the three cases. This agrees well with our earlier estimate of a saturation radius  $R_s = 8/M$  which corresponds to  $\sim 1.66$  fm. This shows that the CDP works and Eq. (2) gives a good estimate of  $R_s$ .

Since the signal of the correlators falls off exponentially with  $r$ , summing over  $r$  beyond the saturating radius  $R_s$  does not change the signal and will only gather noise. Let us consider the disconnected insertion next and see how the S/N ratio can be improved with this observation. In the case of the DI, the variance of the correlator in Eq. (3)  $\frac{1}{V^2} \langle |\sum_{\vec{x}} \sum_{r < R} \mathcal{O}_1(\vec{x} + \vec{r}, t) \mathcal{O}_2(\vec{x}, 0)|^2 \rangle$  can have a vacuum insertion in addition to the exponential fall off in  $t$  due to the  $\mathcal{O}^\dagger \mathcal{O}$  operator.

$$\text{Var}(R, t) = \frac{1}{V^2} \sum_{\vec{x}, \vec{y}} \left( \left\langle \left( \sum_{r_1 < R} \mathcal{O}_1(\vec{x} + \vec{r}_1, t) \sum_{r_2 < R} \mathcal{O}_1^\dagger(\vec{y} + \vec{r}_2, t) \right) \cdot \langle \mathcal{O}_2(\vec{x}, 0) \mathcal{O}_2^\dagger(\vec{y}, 0) \rangle \right\rangle + \dots \right), \quad (4)$$

where  $r_1 = \sqrt{|\vec{r}_1|^2 + t^2}$  and  $r_2 = \sqrt{|\vec{r}_2|^2 + t^2}$  respectively. For the case where  $\vec{r}_1$  and  $\vec{r}_2$  are integrated over the whole lattice volume, the sum over the positions  $\vec{x}, \vec{y}, \vec{x} + \vec{r}_1$  and  $\vec{y} + \vec{r}_2$  can be carried out independently. Consequently,  $\mathcal{O}_1$  and  $\mathcal{O}_2$  in the DI fluctuate independently which leads to a variance which is the product of their respective variances. In this case, the leading vacuum insertion is a constant, independent of  $t$ . This is the reason why the noise remains constant over  $t$  in DI. On the other hand, the constant

variance is reduced to  $V_{R_s}/V$  when  $r$  is integrated to  $R_s$ , while the signal is not compromised. The subleading contribution (denoted by ...) in Eq. (4) has an exponential decay in  $t$  with a mass in the scalar channel. It is clear that to leading order, the ratio of the cutoff S/N at  $R_s$  to that without cutoff is

$$\frac{S/N(R_s)}{S/N(L)} \sim \sqrt{\frac{V}{V_{R_s}}}. \quad (5)$$

We shall consider several DI examples involving volume summations over two or more coordinates. Since the convoluted sum with a relative coordinate in Eq. (3) can be expensive, we shall invoke the standard convolution theorem by calculating the product of two functions  $\tilde{K}(\vec{p}, t) = \tilde{\mathcal{O}}_1(-\vec{p})\tilde{\mathcal{O}}_2(\vec{p})$ , where  $\tilde{\mathcal{O}}_1(-\vec{p})/\tilde{\mathcal{O}}_2(\vec{p})$  is the Fourier transforms of  $\mathcal{O}_1(\vec{x})/\mathcal{O}_2(\vec{x})$  in each configuration on their respective time slices. Then

$$C(R, t) = \left\langle \int_{r < R} d\vec{r} K(\vec{r}, t) \right\rangle. \quad (6)$$

where  $K(\vec{r}, t)$  is the Fourier transform of  $\tilde{K}(\vec{p}, t)$ . In this way, the cost of the double-summation, which is order  $V^2$ , is reduced to that of the fast Fourier transform (FFT) which is in the order of  $V \log V$ .

### A. Scalar matrix element of the strange quark

The first example is the disconnected insertion for the nucleon matrix element with a scalar loop which involves a three-point function and can be expressed as

$$C_3(R, \tau, t) = \left\langle \sum_{\vec{x}} \sum_{r < R} \mathcal{O}_N(\vec{x}, t) S(\vec{x} + \vec{r}, \tau) \bar{\mathcal{O}}_N(\mathcal{G}, 0) \right\rangle, \quad (7)$$

where  $S$  is the vacuum-subtracted scalar loop,  $\mathcal{G}$  denotes the source grid for increasing statistics [10]. Note here  $r = \sqrt{|\vec{r}|^2 + (t - \tau)^2}$  is the 4-D distance and  $r_x$  is the spatial separation between the loop and the sink. Since the low-modes dominate the strangeness in the nucleon [15], we calculate the strange quark loop with low-modes only to illustrate the CDP effect. The sum over the spatial relative coordinate between the scalar quark loop  $S$  and the sink interpolation operator  $\mathcal{O}_N(\vec{x}, t)$  is carried out through the convolution in Eq. (6). This calculation is done on the domain-wall  $32^3 \times 64$  (32ID) lattice [14] with pion mass  $\sim 170$  MeV and the lattice size is 4.6 fm.

The upper panel of Fig. 2 gives the value and the error of the ratio of three-to-two point functions  $C_3/C_2(R, \tau, t)$  in Eq. (7) as a function of  $R$  at  $\tau = 4$  and  $t = 9$ . The nucleon source-sink separation is 1.29 fm in this case. We see the

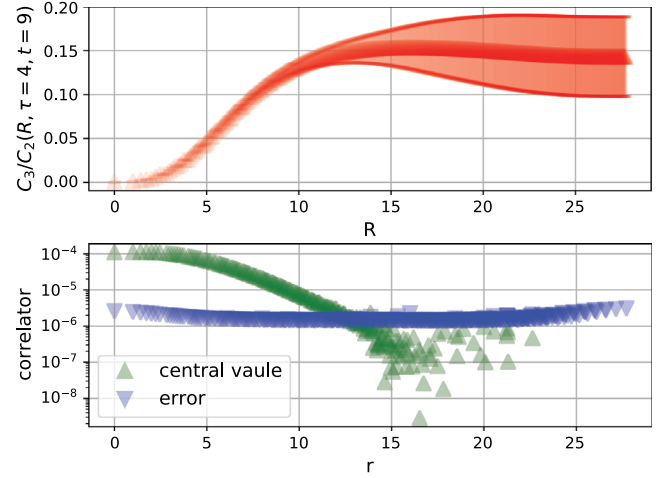


FIG. 2. The value and error of  $C_3/C_2(R, \tau = 5, t = 10)$  are plotted in the upper panel as a function of  $R$ . The lower panel displays the three-point function in Eq. (7) as a function of  $r$  without summing over it. The green band shows the signal and the blue band the error.

error grows after  $R$  is greater than  $\sim 12$  ( $\sim 1.7$  fm) while the central value remains constant within errors. This behavior reflects the fact that the signal without summing over  $|\vec{r}'|$  falls off exponentially with  $r$ , while the error remains constant as shown in the lower panel.

Fig. 3 shows  $DI(\text{scalar})$ —the disconnected three-point to two-point function ratio to obtain the scalar matrix element for the strange quark in the nucleon as a function of  $\tau - t/2$  for two source-sink separations at 1.00 fm (upper panel) and 1.57 fm (lower panel). Two results with cutoffs of  $R_s = 27$  and  $R_s = 12$  are plotted.  $R_s$  is the cutoff radius for the relative coordinate between the sink and the quark

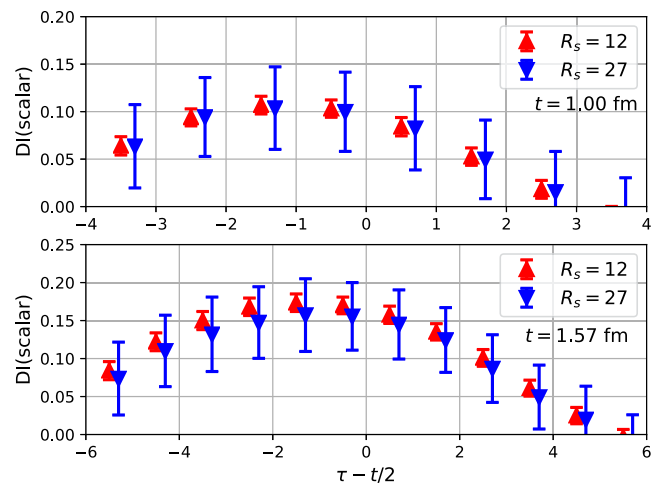


FIG. 3. DI calculation for the strange scalar matrix element in the nucleon as a function of  $\tau - t/2$ . For each of the source-sink separations at 1.00 fm (upper panel) and 1.57 fm (lower panel), two results with a cutoff of  $R_s = 27$  and  $R_s = 12$  are plotted.

loop in the spatial sum. It can be seen that the central values of the two cutoffs are all consistent within errors, while the errors with cutoff  $R_s = 12$  are smaller than the ones with cutoff 27, which includes the whole spatial volume, by a factor of 4 or so. Thus, cutting off the spatial sum at the saturation distance is equivalent to gaining  $\sim 16$  times more statistics.

### B. Glueball mass

Next, we consider the glueball correlators in Eq. (3) on the 48I lattice with  $La = 5.5$  fm. The correlators from the scalar  $E^2$  and  $B^2$  operators and the pseudoscalar  $E \cdot B$  operator are presented in Fig. 4, where they are plotted as a function of  $R$  in Eq. (3) at  $t = 4$ . We note the scalar correlators saturate after  $R \sim 9$  ( $\sim 1.0$  fm) and the pseudoscalar one saturates after  $R \sim 12$  ( $\sim 1.2$  fm), which can be understood in terms of the different ground state masses in these two channels. Again, comparing the error at  $R = 9$  to that at  $R = 24$ , the latter includes the whole spatial volume, for the scalar case, the error is reduced by a factor of  $\sim 4$  which is in reasonable agreement with the prediction of  $\sim (\frac{24}{9})^3 = 4.4$  from Eq. (5). For the pseudoscalar case, the improvement is around 3 times and is consistent with the estimate of  $\sim (\frac{24}{12})^3 = 2.8$ .

### C. Neutron electric dipole moment:

Finally, we examine the  $CP$ -violation phase  $\alpha^1$  on the same 48I lattice which is needed for calculating the neutron electric dipole moment (nEDM). The phase is defined as

$$\alpha^1 = \frac{\text{Tr}[C_{3Q}(t)\gamma_5]}{\text{Tr}[C_2(t)\Gamma_e]} \quad (8)$$

for large enough  $t$ , where  $C_2(t)$  is the common nucleon two-point function,  $\Gamma_e = \frac{1+\gamma_4}{2}$  is the parity projector,  $C_{3Q}(t)$

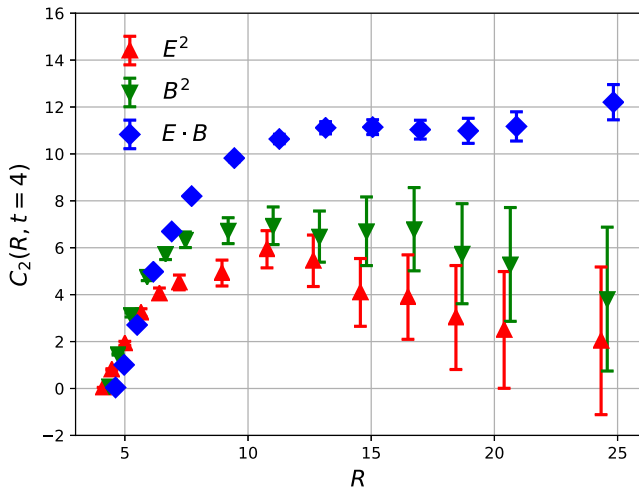


FIG. 4. Scalar (operator  $E^2$  and  $B^2$ ) and pseudoscalar ( $E \cdot B$ ) glueball correlators at  $t = 4$  as a function of cutoff  $R$ .

is the nucleon propagator weighted with the total topological charge  $Q$

$$C_{3Q}(t) = \left\langle \sum_{\vec{x}} \mathcal{O}_N(\vec{x}, t) \bar{\mathcal{O}}_N(\mathcal{G}, 0) Q \right\rangle. \quad (9)$$

We can turn the total topological charge into the summation of its density, i.e.,  $Q = \sum_x q(x)$  where we use the plaquette definition for  $q(x)$ . Then the expression of  $C_{3Q}(t)$  with a cutoff  $R$  can be cast in the same form as in Eq. (7), except the scalar quark loop  $S$  is replaced with the local topological charge  $q(x)$  and the sum of the topological charge density is over the four sphere with a radius  $R$ .

The result of  $\alpha^1$  as a function of  $R$  in Fig. 5 shows that the signal saturates after  $R \sim 16$ . Cutting off the sum of  $q(x)$  at this  $R$  leads to a factor of  $\sim 3.6$  times reduction in error compared to the case of reweighting with the total topological charge as in Eq. (9). This example indicates that for four-dimensional sums, our new method employing the CDP can also improve the S/N. As we illustrated in the introduction, the nEDM from the  $\theta$  term suffers from a  $\sqrt{V}$  problem. It is shown here it is related to the vacuum insertion in the variance. This problem is resolved by turning the topological charge into a 4-D sum of the local charge density and applying the CDP by cutting off the relative 4-D distance in the sum.

### III. SYSTEMATIC AND STATISTICAL ERRORS

So far, we have taken a simple cutoff  $R_s = 8/M$  to illustrate the efficacy of the variance reduction. This *ad hoc* choice inevitably incurs a systematic error. Since the asymptotic behavior of the integral of the correlator as a function of the separation  $R$  is similar to that of the effective mass, we fit it as such and apply the Akaike information criterion (AIC) [16] to obtain the statistical error and the

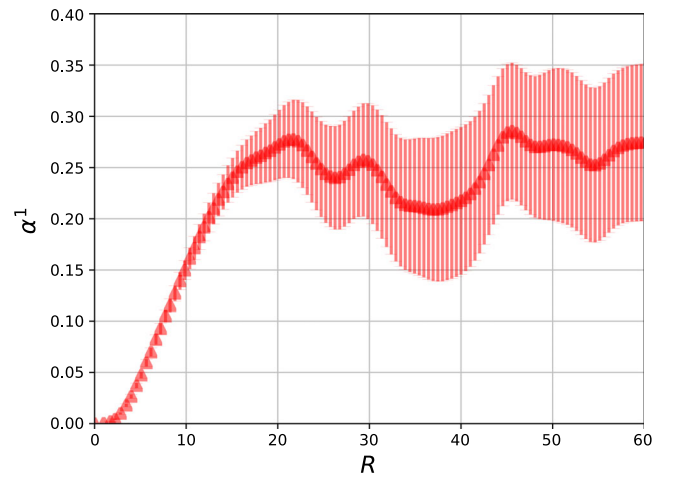


FIG. 5. The  $CP$ -violation phase  $\alpha^1$  calculated on the 48I lattice as a function of cutoff  $R$ . For each  $R$ , the value is averaged from  $t = 6$  to 13.

estimated systematic error by using analysis with different fitting windows and models. The details of the application of AIC to the strange matrix element is provided in the Appendix as an example. It turns out that both the systematic and statistical errors are stable against multiple choices of windows and two fitting models. The statistical error is close to that at the cutoff distance when the plateau emerges (i.e.  $R_s = 8/M$ ). Using a representative fit with 2 fitting formulas and 80 combinations of 8 data points each for a total of 160 fits and 100 bootstrap samplings, we obtain the value of the strange matrix element that we considered earlier to be 0.160(15) (15). The statistical error (first one) and the systematic error (second one) are practically the same. This is to be compared with the original value of 0.143(45) without taking the CDP into account. The systematic error is to be added to the total systematic error of the calculation.

#### IV. DISCUSSION AND SUMMARY

Regarding the nucleon correlator in Fig. 1, we notice that there is no conspicuous increase of the error as a function of  $R$ . This is because, unlike the DI, the variance does not have a vacuum insertion for the connected insertion (CI). The leading contribution to the variance is expected to be  $e^{-3m_\pi r}$ . This has a longer range than that of the signal which falls off with the nucleon mass. Therefore, in principle, one would expect some gain in the S/N when  $R$  is cut off at  $R_s$ . Therefore, the corresponding ratio of S/N in Eq. (5) is

$$\frac{S/N(R_s)}{S/N(L)} \sim \sqrt{\frac{A(L/2, 3m_\pi)}{A(R_s, 3m_\pi)}}. \quad (10)$$

For the 48I lattice in Fig. 1, this ratio is 1.13 for the physical pion mass with the cutoff  $R_s = 8/M$ . This is not nearly as much a gain as in the DI where the variance is dominated by the vacuum insertion. In the CI case, the noise saturates at  $\sim 8/(3m_\pi) = 3.5$  fm. There is no gain for a lattice with a size larger than this.

In summary, we have shown that the exponential fall off of the cluster decomposition principle (CDP) seems to hold numerically for the several correlators that we examined. For the disconnected insertions (DI), we find that the vacuum insertion dominates the variance so that the relevant operators fluctuate independently and is independent of the time separation. This explains why the signal fall off exponentially, while the error remains constant in the DI. To demonstrate the efficacy of employing the CDP to reduce the variance, we have restricted the volume sum of the relative coordinate between the operators to the saturation radius  $R_s$  to show that there is an effective gain of  $V/V_s$  in statistics without compromising the signal. This applies to all DI cases. For the cases we have considered, namely the glueball mass, the strangeness content in the

nucleon, and the  $CP$  violation angle in the nucleon due to the  $\theta$  term, we found that for lattices with a physical sizes of 4.5–5.5 fm, the errors of these quantities can be reduced a factor of 3 to 4. We have applied the Akaike information criterion (AIC) [16] in the Appendix to estimate the systematic and statistical errors incurred by applying the CDP. For the strangeness content, we find that the systematic error is practically of the same size as that of the statistical one when the cluster decomposition is taken into account. This results in a 2 to 3 times reduction in the overall error. For the connected insertions, there is no vacuum insertion in the variance, the gain in statistics is limited.

#### ACKNOWLEDGMENTS

This work is supported in part by the U.S. DOE Grant No. DE-SC0013065. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Stampede at TACC through allocation MCA01S007, which is supported by National Science Foundation Grant No. ACI-1053575 [17]. We also thank National Energy Research Scientific Computing Center (NERSC) for providing HPC resources that have contributed to the research results reported within this paper. We acknowledge the facilities of the USQCD Collaboration used for this research in part, which are funded by the Office of Science of the U.S. Department of Energy.

#### APPENDIX: ERROR ESTIMATE USING AKAIKE INFORMATION CRITERION (AIC)

AIC is founded in information theory. It is an estimator of the relative relevance of the statistical models that are used to describe a given set of data [16,18].

The usual task is to fit the data generated by some unknown process (function)  $f$  with different trial models. However, we cannot tell which model is a better representation of  $f$  with certainty in practice, because we do not know  $f$ . Akaike (1974) showed that we can estimate, via AIC, how much more (or less) information is lost when comparing one model to another. The estimate, though, is only valid asymptotically. If the number of data points is small, then some correction is often necessary (e.g., AICc) [18].

##### 1. Definition of the AIC value

Considering a model  $M$  for some data  $x$  with  $k$  parameters, the maximum value of the likelihood function for the model  $L_{\max}$  is represented by the following probability

$$L_{\max} = P(x|\vec{\theta}, M), \quad (A1)$$

where  $\vec{\theta}$  is the parameter vector that maximizes the likelihood function. The maximum likelihood function is related to the minimum  $\chi^2$  values in the standard fitting via

$$L_{\max} = e^{-\frac{\chi_{\min}^2}{2}}. \quad (\text{A2})$$

The AIC value of the model is defined as (see Ref. [18] and the references therein for details)

$$\text{AIC} = 2k - 2 \ln(L_{\max}) = 2k + \chi_{\min}^2, \quad (\text{A3})$$

This favors models with the minimum AIC value and penalizes those with many fitting parameters.

## 2. Practical application

In practice, we take a weighted average of all the models as is carried out in Refs. [19,20]. The normalized weight for each model is

$$w_i = \frac{e^{-\frac{\text{AIC}_i}{2}}}{\sum_i e^{-\frac{\text{AIC}_i}{2}}}, \quad (\text{A4})$$

where  $i$  is the index of the models.

When handling the systematic errors of lattice calculations, we usually need to take consideration of various fitting formulas with different combinations of data points used for the fit. The AIC method can be helpful in these cases. Let us consider a case where we have  $N$  configurations and, for each configuration, there are  $M$  data points (e.g., different time separation  $t$  of a hadron two-point correlator). We plan to use  $P$  models to fit the data (the number  $P$  includes different fit ranges and different combinations of data points) in order to obtain the mean value, the systematic error and the statistical error of some model parameters (e.g., the mass of the ground state). The detailed procedure are given as follows:

- (1) The mean value is the weighted average of all the  $P$  fit models (formulas and combinations of data points).

$$\bar{x} = \sum_{i=1}^P w_i x_i, \quad (\text{A5})$$

$x_i$  is the fitting result from each model and  $w_i$  is the normalized weight in Eq. (A4).

- (2) The systematic error is taken to be the standard error of the weighted mean,

$$\sigma_{\text{sys}} = \sqrt{\sum_{i=1}^P w_i \sigma_i^2}, \quad \sigma_i = (x_i - \bar{x}). \quad (\text{A6})$$

- (3) The statistical error can be obtained from bootstrap resampling. We first do  $N_b$  times bootstrap operation, in each bootstrap sample, we fit these

$P$  models and save the weighted mean value. After that, we will have  $N_b$  AIC weighted mean values. The bootstrap error of these weighted mean values gives the final statistical error.

In summary, we need to do  $(1 + N_b) \times P$  times of correlated fits to obtain all the relevant results.

## 3. The cluster decomposition case

In the cluster decomposition case, we shall consider the strange matrix element as a function of the cutoff radius  $R$  as illustrated in the following figure. We can use the AIC method to estimate the mean value, the systematic error and the statistical error of the ratio.

To apply the AIC method to this particular case, we need first to determine our fitting formulas. In view of the fact that the ratio between the three-point function and the two-point function falls off exponentially as a function of the relative separation between the quark loop and the sink of the nucleon propagator, the accumulated sum of the separation with a cutoff  $R$  is expected to be a constant after certain  $R$ , such as  $R_s = 8/M$ , as illustrated in the upper panel of Fig. 2. This is much like fitting the effective mass plot to isolate the ground state. The two formulas (models) we will use are

$$f_1(R) = C_0 \quad (\text{A7})$$

and

$$f_2(R) = C_0 + C_1 \frac{\sqrt{R} e^{-mR}}{m}. \quad (\text{A8})$$

The first one is the asymptotic form for  $R \rightarrow \infty$  and the second one is the form commensurate with that from the cluster decomposition principle to cover more range of  $R$  in the fitting.

Then, we need to choose the combinations of data points. To make sure that every combination has the same weight, we set the number of data points (marked as  $N_d$ ) contained

TABLE I. The mean values, and the systematic and statistical errors are given for various fits.  $N_b$  is the number of bootstrap samples,  $N_d$  the number is data points in each fit, and  $N_c$  is the number of different combinations of the  $N_d$  data points.

$N_b$	$N_d$	$N_c$	mean	$E_{\text{sys}}$	$E_{\text{sta}}$
100	6	80	0.161	0.013	0017
100	8	80	0.160	0.015	0.015
100	10	80	0.163	0.019	0.012
100	8	60	0.161	0.015	0.015
100	8	80	0.160	0.015	0.015
100	8	100	0.161	0.015	0.015
50	8	80	0.160	0.015	0.015
100	8	80	0.160	0.015	0.015
200	8	80	0.160	0.015	0.015



in each combination to be equal. And to enlarge the number of different combinations (marked as  $N_c$ ) we can have, we do not force the points in each combination to be contiguous. (e.g., if the total range of data points is  $R \in [8, 27]$ , and we set  $N_d = 4$ ,  $N_c = 5$ , the possible combinations can be  $[8, 9, 12, 15]$ ,  $[15, 18, 21, 22]$ ,  $[15, 16, 17, 18]$ ,  $[10, 12, 20, 22]$  and  $[20, 22, 24, 27]$ .) If  $N_c$  is large enough, combinations will include both contiguous data points and noncontiguous data point. So in this sense, this is a more general way to choose data points.

Having the formulas and combinations, we can then proceed to do the fittings. The final results might be affected by 3 factors:  $N_b$  (number of bootstraps),  $N_d$  (number of data points) and  $N_c$  (number of combinations). We vary these 3 factors to check if the results are stable in these fits.

The data range used is  $R \in [10, 27]$ . After several hundreds of thousands of correlated fits, the results are shown in Table I. The values and errors are pretty stable no matter how we vary  $N_b$ ,  $N_d$ , or  $N_c$ . We decide to take the representative results from  $N_b = 100$ ,  $N_d = 8$ , and  $N_c = 80$  which gives the value of  $0.160(15)(15)$  as our final estimation. The systematic error (the second parenthesis) is comparable to the statistical one (the first parenthesis) from 100 bootstrap samples. This is to be compared to the original value of  $0.143(45)$  when the sum over the relative coordinate between the quark loop and the sink of the nucleon propagator is carried out to cover the whole spatial volume. The total number of analysis  $P$  is the product of the number of models (2 in this case) and  $N_c$ , the number of different combinations of  $N_d$  data points. In this case,  $P = 2 \times 80 = 160$ .

- 
- [1] S. Aoki *et al.*, *Eur. Phys. J. C* **77**, 112 (2017).  
 [2] G. Parisi, *Phys. Rep.* **103**, 203 (1984).  
 [3] G. P. Lepage, in Boulder ASI **1989**, 97 (1989).  
 [4] H. Araki, K. Hepp, and D. Ruelle, *Helv. Phys. Acta* **35**, 164 (1962).  
 [5] G. Bali, S. Collins, and A. Schafer (QCDSF Collaboration), *Proc. Sci.*, LAT2009 (2009) 149 [arXiv:0911.2407].  
 [6] Y. Chen, A. Alexandru, T. Draper, K.-F. Liu, Z. Liu, and Y.-B. Yang, arXiv:1507.02541.  
 [7] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, and C. Lehner, *Phys. Rev. D* **93**, 014503 (2016).  
 [8] M. Cè, L. Giusti, and S. Schaefer, *Phys. Rev. D* **95**, 034503 (2017).  
 [9] E. Shintani, T. Blum, T. Izubuchi, and A. Soni, *Phys. Rev. D* **93**, 094503 (2016).  
 [10] A. Li *et al.* ( $\chi$ QCD Collaboration), *Phys. Rev. D* **82**, 114501 (2010).  
 [11] M. Gong *et al.* ( $\chi$ QCD Collaboration), *Phys. Rev. D* **88**, 014503 (2013).  
 [12] R. S. Sufian, Y.-B. Yang, A. Alexandru, T. Draper, J. Liang, and K.-F. Liu, *Phys. Rev. Lett.* **118**, 042001 (2017).  
 [13] Y.-B. Yang, A. Alexandru, T. Draper, M. Gong, and K.-F. Liu, *Phys. Rev. D* **93**, 034503 (2016).  
 [14] T. Blum *et al.* (RBC, UKQCD Collaborations), *Phys. Rev. D* **93**, 074505 (2016).  
 [15] Y.-B. Yang, A. Alexandru, T. Draper, J. Liang, and K.-F. Liu ( $\chi$ QCD Collaboration), *Phys. Rev. D* **94**, 054503 (2016).  
 [16] H. Akaike, *IEEE Trans. Autom. Control* **19**, 716 (1974).  
 [17] J. Towns *et al.*, *Comput. Sci. Eng.* **16**, 62 (2014).  
 [18] [https://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](https://en.wikipedia.org/wiki/Akaike_information_criterion).  
 [19] S. Borsanyi *et al.*, *Science* **347**, 1452 (2015).  
 [20] S. Durr *et al.*, *Phys. Rev. Lett.* **116**, 172001 (2016).