Parton Distribution Functions and Amplitudes of the Pseudoscalar Mesons and the Nucleon from Lattice QCD

May 10, 2023

Abstract

A long-standing challenge in lattice QCD is the direct computation of key measures of hadron structure, including parton distribution functions, quark distribution amplitudes, and three-dimensional measures such as the transverse-momentum-dependent distributions, and generalized parton distributions. Recently, new approaches have been proposed that enable their direct computation, and these are characterized by a requirement that the hadron of interest be increasingly relativistic. This **Class A Continuation** proposal aims to capitalize on recent developments, and our previous USQCD allocation, to compute the structure functions of the pion and nucleon using the pseudo-PDF formulation and the distillation framework, to extend the calculation to that of the gluon and sea-quark contributions to hadron structure, and to embark on calculations of three-dimensional imaging through the calculation of GPDs in the pseudo-PDF approach. This project is relevant to the hadron structure experimental programs at JLab, at RHIC-spin, and at the future EIC. The emphasis in this proposal on flavor-singlet structure is particularly timely with the increasing efforts directed towards the physics program at the EIC. We request **108K MI100 GPU-hours, 19.2M KNL core-hours, and 523K RTX GPU-hours**. In addition, we request an additional 200 TByte tape storage, and a total of 400 TByte of disk storage.

1 Goals and Milestones

This is a *continuation* proposal aimed at understanding the internal structure of pion, kaon and nucleon encapsulated in the one-dimensional Parton Distribution Functions (PDFs) and through the three-dimensional Generalized Parton Distribution Functions (GPDs). The overall goals, methods and software were described in detail in our proposal last year, and in that of 2019, and we repeat them here only as necessary to provide a self-contained document for review. We emphasise, however, that the goals of our project, and the specific aims this year, are directly related to the experimental programs at the 12 GeV upgrade of Jefferson Lab, and the work on the gluon distributions in particular is key to the program of the future Electron-Ion Collider.

The most urgent task in our program of understanding the structure of hadrons within the pseudo-PDF framework is the control over the systematic uncertainties, and in particular the need to have as great a reach in Ioffe time, and hence spatial momentum, as possible whilst ensuring that we remain in a regime where short-distance factorization is applicable. Thus the emphasis is on calculations at a fixed, somewhat larger-than-physical pion mass, but at a variety of lattice spacings. Specifically, we have two goals for this year. The first is the study of the flavor-singlet contributions to hadron structure, and in particular those arising from the disconnected quark diagrams thereby completing our program of the contribution of the gluons.

With the importance of a high-precision study in mind, we now have a suite of configurations with 20,000 trajectories, enabling us to perform studies on ensembles containing around 2,000 independent trajectories. The ensembles employed in our program of hadron-structure calculations are listed in Table 1. For this year of our program, we will primarily exploit the ensembles E2 and E3, the work presented in the progress report exploits E1.

Ensemble	β	L/a	T/a	m_π (MeV)	a (fm)	L (fm)	Traj.
E1	6.3	32	64	360	0.093	2.98	20,000
E2	6.3	32	64	270	0.093	2.98	20,000
E3	6.5	48	128	270	0.072	3.36	20,000
E4	6.7	64	192	270	0.055	3.52	20,000

Table 1: Parameters for the ensembles used in our hadron-structure computations. Column two gives the bare coupling β, columns three and four the spatial and temporal extents, respectively, in lattice units. We give the approximate pion mass in MeV in column five, and the approximate lattice spacing, a and spatial lattice extent, L, in columns six and seven, in physical units (1 fm = 10^{-15} m). The final column lists the number of available HMC trajectories.

The layout of the rest of the proposal is as follows. We will begin by describing progress since submission last year. We will give a brief description of changes to our code base before presenting our plan of work for the coming year, together with the associated resource request.

Progress since last USQCD Submission

The focus of the work published during the current allocation period has been on understanding the internal collinear structure of the nucleon within the pseudo-PDF framework.

The key to our program is the use of a smearing method known as *distillation*[1] which allows us to perform calculations at a significantly higher precision than other competing methods. This year we computed the quark and gluon helicity distributions in the proton [2, 3] and in addition finalized the computation of the transversity distribution in the proton [4]. Furthermore, we explored how

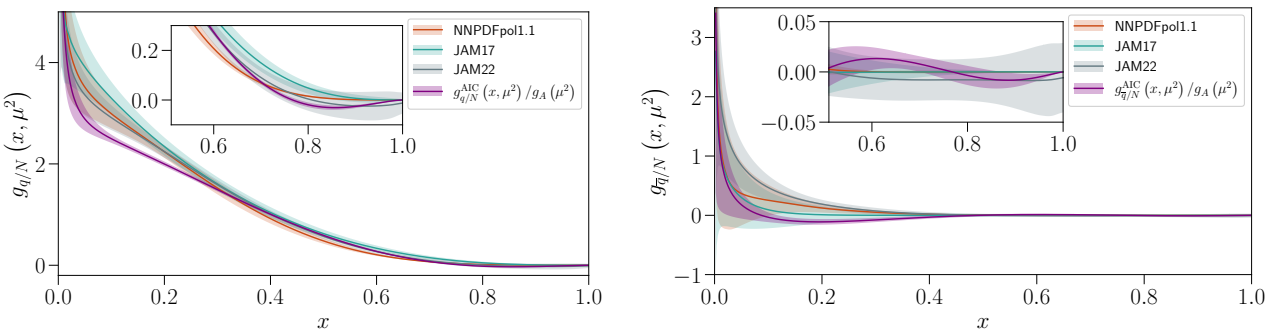

Figure 1: Derived $g_{q/N}\left(x,\mu^2\right)/g_A\left(\mu^2\right)$ (left) and $g_{\overline{q}/N}\left(x,\mu^2\right)/g_A\left(\mu^2\right)$ (right) PDFs obtained from the AIC prescription at an input scale $\mu^2 = 4$ GeV^2 compared with $g_{q/N}\left(x,\mu^2\right)$ and $g_{\overline{q}/N}\left(x,\mu^2\right)$ at the same scale isolated in the global analyses `NNPDFpol1.1` [6], `JAM17` [7], and `JAM22` [8].

our lattice results may affect the phenomenological extractions of parton distribution functions from experiment [5].

Non-singlet quark helicity PDFs of the nucleon from pseudo-distributions The non-singlet helicity quark parton distribution functions (PDFs) of the nucleon are determined from lattice QCD, by jointly leveraging pseudo-distributions and the distillation spatial smearing paradigm. A Lorentz decomposition of appropriately isolated space-like matrix elements reveals pseudo-distributions that contain information on the leading-twist helicity PDFs, as well as an invariant amplitude that induces an additional z^2 contamination of the leading-twist signal. An analysis of the short-distance behavior of the space-like matrix elements using matching coefficients computed to next-to-leading order (NLO) exposes the desired PDF up to this additional z^2 contamination. Due to the non-conservation of the axial current, we elect to isolate the helicity PDFs normalized by the nucleon axial charge at the same scale μ^2. The leading-twist helicity PDFs as well as several sources of systematic error, including higher-twist effects, discretization errors, and the aforementioned z^2 contaminating amplitude are jointly determined by characterizing the computed pseudo-distribution on a basis of Jacobi polynomials. The Akaike Information Criterion (AIC) is exploited to effectively average over distinct model parameterizations and cuts on the pseudo-distribution. Encouraging agreement is observed with recent global analyses of each non-singlet quark helicity PDF, notably a rather small non-singlet anti-quark helicity PDF for all quark momentum fractions. The final results for the non-singlet quark helicity distribution in the nucleon can be seen in Fig. 1. Our results are compared with state-of-the art phenomenological extractions and are found to be in good agreement with them.

Toward the determination of the gluon helicity distribution in the nucleon from lattice quantum chromodynamics In this work we present the first exploratory lattice quantum chromodynamics (QCD) calculation of the polarized gluon Ioffe-time pseudo-distribution (ITD) in the nucleon. The Ioffe-time pseudo-distribution provides a frame-independent and gauge-invariant framework to determine the gluon helicity in the nucleon from first principles. We employ a high-statistics computation using a $32^3 \times 64$ lattice ensemble characterized by a 358 MeV pion mass and a 0.094 fm lattice spacing. We establish the pseudo-distribution approach as a feasible method to address the proton spin puzzle with successive improvements in statistical and systematic uncertainties anticipated in the future. Within the statistical precision of our data, we find a good comparison

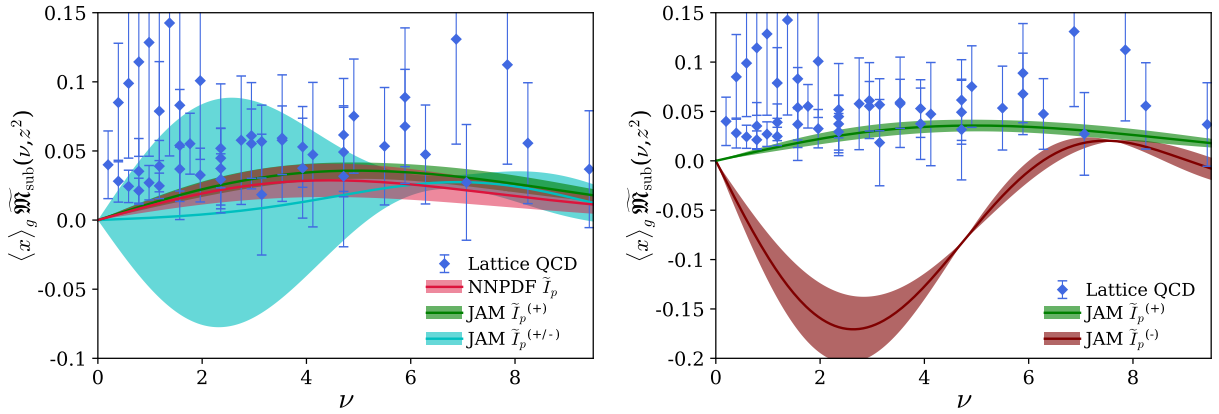

Figure 2: A comparison between the lattice reduced Ioffe-time pseudo-distribution $\mathfrak{M}(\nu, z^2)$ in the zero flow-time limit obtained through the subtraction method using the $p = 0$ matrix elements, and the gluon helicity ITD constructed from global fits of PDFs. In the left plot, the red band denotes the ITD constructed from the gluon helicity distribution by the NNPDF collaboration. The green band labeled by $\mathcal{I}_p^{(+)}$ and the cyan band labeled by $\mathcal{I}_p^{(+/-)}$ represent the gluon helicity ITD determined by the JAM collaboration with and without the positivity constraint on the gluon helicity PDF, respectively. On the right plot, the gluon helicity ITDs for positive and negative helicity PDFs are compared with the lattice data. The green band labeled by $\mathcal{I}_p^{(+)}$ and the maroon band labeled by $\mathcal{I}_p^{(-)}$ represent the gluon helicity ITD determined by the JAM collaboration associated with the positive and negative gluon helicity PDF solutions, respectively.

between the lattice determined polarized gluon Ioffe-time distribution and the corresponding expectations from the state-of-the-art global analyses. We find a hint for a nonzero gluon spin contribution to the proton spin from the model-independent extraction of the gluon helicity pseudo-distribution over a range of Ioffe-time, $\nu \lesssim 9$.

In the most recent global fit of the gluon helicity distribution [9], the Jefferson Lab Angular Momentum (JAM) collaboration showed that without the assumption of positivity constraints on the quark and gluon helicity PDFs, the magnitude or sign of the gluon polarization in the nucleon cannot be properly constrained. In other words, the ITD extracted from the JAM global fit (labeled by JAM $\mathcal{I}_p^{(+/-)}$ in Fig. 2) may have a similar or even larger magnitude of uncertainty than our lattice QCD calculation. We show a comparison of the polarized gluon ITDs obtained from global fits and our lattice calculation in Fig. 2. Most importantly, Fig. 2 shows that the ITD data in the $\nu \lesssim 6$ region is primarily controlled by whether the gluon polarization in the nucleon is positive or negative, according to the JAM analysis.

The positivity constraint on the gluon distributions, namely helicity-aligned and helicity-antialigned both being non-negative, in the analysis of experimental data in [9] leads to a substantial reduction of the variance of $x\Delta g(x)$ in the large-x region, as seen in Fig. 6 of [9]. Specifically, the PDFs without the positivity assumption were organized into a band of solutions with a negative PDF and a band of solutions with a positive PDF. We compare the ITDs resulting from the two bands with positive and negative $x\Delta g(x)$ to our results in the right panel of Fig. 2. The current matrix elements, albeit with an unphysical pion mass and finite lattice spacing, are inconsistent within statistical uncertainties with the negative PDF branch. We note, however, that the large ν behavior of the lattice-calculated ITD, beyond the range currently accessible, could turn negative and compensate for the positive trend observed in Fig. 2.

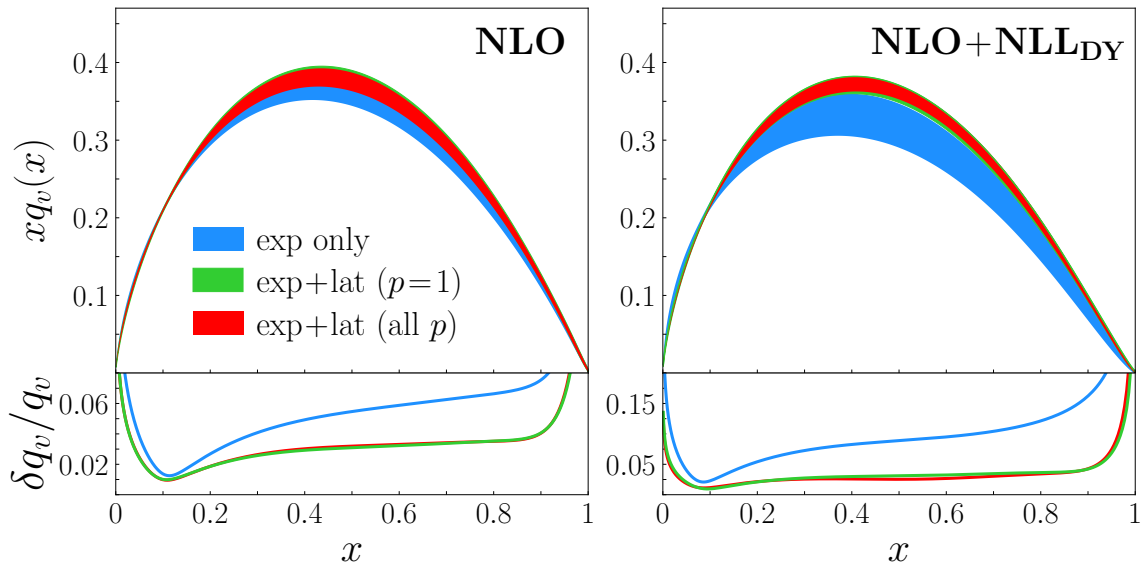

Figure 3: Valence quark distributions (**top**) when extracted from experimental data alone (blue), combined with the $p = 1$ lattice data (green), and combined with all the lattice data (red) for the NLO (**left**) and NLO+NLL$_{\rm DY}$ (**right**) cases, along with the relative uncertainties (**bottom**). The bands represent a 1σ uncertainty level.

Complementarity of Experiment and Lattice QCD on Pion PDF We have extracted the pion PDF from a combined global QCD analysis [5] of experimental data, together with Ioffe-time pseudo distributions [10] and matrix elements of current-current correlators [11, 12], in collaboration with the Jefferson Angular Momentum (JAM) Collaboration. The impact of the lattice calculation using the more precise, reduced pseudo distribution [10] to significantly reduce the uncertainty on the extracted PDF with respect to the experimental analysis alone is clear from Figure 3.

Software

We will use the *Chroma* software framework, which supports multigrid on the Nvidia and AMD GPUs, using *QUDA*, and on KNLs and Skylake nodes, using *QPhiX/mgproto*. For the computation of the elementals, we have a CPU and GPU implementation within *Chroma* that supports the phased distillation that is needed to reach high momenta. For the computation of the disconnected diagram, we employ the recently developed [13] methodologies that rely on multigrid deflation as well state-of-the-art probing methods [14] to reduce by orders of magnitude the variance of the stochastic estimator. This has been implemented on the KNL architecture, and cannot be performed on GPU-type architectures.

Finally, we use *redstar* for the contraction of the perambulators, generalized perambulators and elementals to form the resulting correlation functions.

We construct the genprops needed for our correlation functions with a newly developed software that computes quark propagators and contract them in one task, avoiding the large I/O of storing the quark propagators. The calculation of the sink and source tensors,

$$T^{(i)}_{\rm sink} = v^{(i)\dagger}(t_f)M^{-1}(t_f,t) \qquad T^{(i)}_{\rm src} = M^{-1}(t,t_i)V^{(i)}(t_i),$$

involves the inversion of M, the lattice Dirac operator, and uses program code for high-performance and advanced algorithmic inversion of the Dirac operator, which relies on a lattice grid that uses a

Ens.	$C_{\rm inv}$ (c-s)	$N_{\rm inv,d}$	$N_{\rm cfg}$	Total (c-hrs)
E2	303	3,072	2,000	516K
E3	2048	3,072	2,000	3,496K
Total:				**4,012K**

Table 2: Resources for computing the u/d-quark disconnected diagrams on our E2 and E3 ensembles on the KNL nodes. Here $C_{\rm inv}$ is the cost of a single inversion of the Dirac Matrix in core-seconds, $N_{\rm inv,d}$ is the number of inversions for estimating the disconnected components, and $N_{\rm cfg}$ is the number of gauge configurations. We will employ $N_p = 256$ probing vectors, corresponding to 256×12 inversions, corresponding to $N_{\rm inv,d} = 3,072$. Note that the cost of inserting the matrix element increases the number of inversions by a factor of 2.

four-dimensional space-time layout. This part of the calculation runs in capacity mode and utilizes all compute nodes within the job, where each node processes a four-dimensional sub-grid of the lattice. The correlation length t varies only over a sub-domain of the full time-extent of the lattice, so we are interested in computing the contraction of $T_{\rm sink}$ and $\Gamma(t)\mathcal{D}\mathcal{P}T_{\rm src}$ on this sub-domain only. However, because of the way the data is arranged, only a small portion of the computing nodes will participate in the computation, dramatically impacting the overall performance of generating genprops. The new code reduces by an order of magnitude the time expended in contractions by redistributing the data to balance the computation among the nodes, and rearranging the contractions as dense matrix-matrix multiplications, which are done with an optimized BLAS.

Work Plan for 2023-2024

Flavor-singlet hadron structure Our work plan for the coming year will focus on the contribution of the disconnected diagrams, both for those of the u/d quarks and of the s quarks, to hadron structure. Our preliminary analysis on the $32^3 \times 64$ lattice reveals that the number of configurations required to compute the disconnected diagrams is at least comparable to those needed to determine the gluonic contributions to hadron structure, described in the progress report. Thus our proposal is to compute the disconnected diagrams for both the u/d and s quarks on a total of 2,000 configurations for the ensembles E2 and E3. The cost of this computation for the u/d quarks is shown in Table 2. We assume that the cost is the same for the s quarks, yielding a total request for this part of the computation of **8,014K KNL-core-hours**.

In addition to the calculation of the quark disconnected diagrams, it is necessary to calculate the perambulators and the corresponding two-point functions for the target hadrons to high precision. Thus our request is both to compute the disconnected diagrams on these ensembles, and to compute the additional perambulators needed for the necessary statistics. Thus we propose to compute these on both the E2 and E3 ensembles for an additional 1500 configurations. This element of the computation can exploit the RTX nodes of Jefferson Lab, *and can also be run on the KNL and MI100 nodes*; the requested resources are shown in Table 3.

Parton Distribution Functions and GPDs

The formalism for the PDF calculations within the pseudo-PDF framework was introduced in ref. [15], and extended to GPDs within in ref. [16]. GPDs are three-dimensional measures of hadron structure described in terms of the momentum fraction x, the so-called skewness ξ, and the momentum

Ens	Soln (g-s)	N_ν	N_{t_i}	N_b	$N_{\rm inv}$	$N_{\rm cfg}$	Tot (GPU-h)
E2	10	64	8	5	10,240	1500	43K
E3	50	96	12	5	23,240	1500	480K
TOTAL							**523K**

Table 3: The cost of computing the light-quark perambulators needed for the pion, and nucleon two-point functions. Here N_ν, N_b are the number of eigenvectors and boosts, respectively, and N_t is the number of source time slices.

Ens	Task	Soln (n-s)	N_v	$N_{\rm src}$	N_b	$N_{\rm sep}$	$N_{\rm inv}$	$N_{\rm cfg}$	TOTAL (GPU-h)
E2	u/d	0.26	64	2	3	4	12,288	200	1.4K
E3	u/d	13	96	2	3	4	18,432	200	106K
TOTAL									**108K**
Ens	Task	Soln (c-s)	N_v	$N_{\rm src}$	N_b	$N_{\rm sep}$	$N_{\rm inv}$	$N_{\rm cfg}$	TOTAL (c-h)
E4	u/d	7282	96	2	3	4	18,432	300	11.2M
TOTAL									**11.2M**

Table 4: Resources for the calculation of the PDFs and GPDs, as discussed in the text, using MI100 resouces (upper table) and KNL resources (lower table). The timings in the upper table are obtained from runs on the "21g" cluster. The timings on the ensemble E2 are obtained from the measured timings of 0.26 node-seconds for a single solve; those on the ensembles E3 and E4 are based on the measured time of 13 node-seconds for a single solve on 3 nodes on the E3 ensemble. The timings for the lower Table is obtained from the measured time on the KNL cluster of 24 node-seconds for an inversion on a $48^3 \times 96$ lattice.

transfer t. An important feature of the pseudo-PDF framework is that it will enable the GPD to be charted both as a function of x, and for a discrete set of ξ and t. This task requires the generation of perambulators [1], quark propagators projected onto a "distillation" basis, and generalized propagators (genprops), quark propagators projected on the distillation basis that contain an insertion of an appropriate operator at a particular time slice. These genprops are the primary data products of our work, and can be re-used in multiple lattice calculations for a variety of physics applications. In particular, the same set of perambulators and genprops can be employed for the computation of the PDFs and GPDs of both the nucleon and of the pion, and a further set, corresponding to a strange quark, can be used for the calculation of the kaon.

Our request this year is to continue the generation of the perambulators and generalized perambulators, as tabulated in Table 4. Here the total number of inversion per configuration is given by $N_{inv} = 2 \times 4 \times N_b \times N_{\rm src} \times N_v \times N_{\rm sep}$, where 4 counts the number of quark spin states, the factor of 2 accounts for the source and sink solution vectors, and N_b the number of boosts corresponding to the different phasing factors (see Ref. [17]) needed to improve data quality at high-momenta. Also, $N_{\rm src}$ is the number of time-sources, N_v is the number of distillation vectors, and $N_{\rm sep}$ is the number of time-source separations. For the largest ensemble E4, we propose to use the KNL resource. Note that we can use the KNL resource to compute the generalized perambulators but *cannot use the RTX resource.*

In summary, we request **108K MI100 GPU-hours, 19.2M KNL core-hours, and 523K RTX GPU-hours**. There is some flexibility in terms of the resources required to implement each stage of our program, as noted in the text.

Readiness and Run Schedule

The software used in this proposal is that of USQCD, and in particular *Chroma*. Note that we had some impediments related to health issues that are now largely resolved.

Data sharing and exclusive rights

We note that most of the data generated as part of this proposal comprises the correlators, the "perambulators", "generalized perambulators", and elementals. These are in general very specific to our projects rather than of general use.

References

[1] Peardon M, Bulava J, Foley J, Morningstar C, Dudek J, Edwards R G, Joo B, Lin H W, Richards D G and Juge K J (Hadron Spectrum) 2009 *Phys. Rev. D* **80** 054506 (*Preprint* 0905.2160)

[2] Edwards R G *et al.* 2022 *arXiv:2211.04434* (*Preprint* 2211.04434)

[3] Egerer C *et al.* (HadStruc) 2022 *Phys. Rev. D* **106** 094511 (*Preprint* 2207.08733)

[4] Egerer C *et al.* (HadStruc) 2022 *Phys. Rev. D* **105** 034507 (*Preprint* 2111.01808)

[5] Barry P C *et al.* (Jefferson Lab Angular Momentum (JAM), HadStruc) 2022 *Phys. Rev. D* **105** 114051 (*Preprint* 2204.00543)

[6] Nocera E R, Ball R D, Forte S, Ridolfi G and Rojo J (NNPDF) 2014 *Nucl. Phys. B* **887** 276–308 (*Preprint* 1406.5539)

[7] Ethier J J, Sato N and Melnitchouk W 2017 *Phys. Rev. Lett.* **119** 132001 (*Preprint* 1705.05889)

[8] Cocuzza C, Melnitchouk W, Metz A and Sato N (Jefferson Lab Angular Momentum (JAM)) 2022 *Phys. Rev. D* **106** L031502 (*Preprint* 2202.03372)

[9] Zhou Y, Sato N and Melnitchouk W (Jefferson Lab Angular Momentum (JAM)) 2022 *Phys. Rev. D* **105** 074022 (*Preprint* 2201.02075)

[10] Joó B, Karpie J, Orginos K, Radyushkin A V, Richards D G, Sufian R S and Zafeiropoulos S 2019 *Phys. Rev.* **D100** 114512 (*Preprint* 1909.08517)

[11] Sufian R S, Karpie J, Egerer C, Orginos K, Qiu J W and Richards D G 2019 *Phys. Rev.* **D99** 074507 (*Preprint* 1901.03921)

[12] Sufian R S, Egerer C, Karpie J, Edwards R G, Joó B, Ma Y Q, Orginos K, Qiu J W and Richards D G 2020 *Phys. Rev. D* **102** 054508 (*Preprint* 2001.04960)

[13] Romero E, Stathopoulos A and Orginos K 2020 *Journal of Computational Physics* **409** 109356 ISSN 0021-9991 URL http://www.sciencedirect.com/science/article/pii/S0021999120301303

[14] Gambhir A S, Stathopoulos A and Orginos K 2017 *SIAM Journal on Scientific Computing* **39** A532–A558

[15] Radyushkin A 2017 *Phys. Rev. D* **96** 034025 (*Preprint* 1705.01488)

[16] Radyushkin A V 2019 *Phys. Rev.* **D100** 116011 (*Preprint* 1909.08474)

[17] Egerer C, Edwards R G, Orginos K and Richards D G 2021 *Phys. Rev. D* **103** 034502 (*Preprint* 2009.10691)