

RECEIVED: October 16, 2020

REVISED: December 21, 2020

ACCEPTED: January 4, 2021

PUBLISHED: February 16, 2021

# Neural-network analysis of Parton Distribution Functions from Ioffe-time pseudodistributions

**Luigi Del Debbio,<sup>a</sup> Tommaso Giani,<sup>a</sup> Joseph Karpie,<sup>b</sup> Kostas Orginos,<sup>c,d</sup>  
Anatoly Radyushkin<sup>c,e</sup> and Savvas Zafeiropoulos<sup>f</sup>**

<sup>a</sup>*The Higgs Centre for Theoretical Physics, The University of Edinburgh,  
Peter Guthrie Tait Road, Edinburgh EH9 3FD, U.K.*

<sup>b</sup>*Physics Department, Columbia University,  
538 West 120th Street, New York City, NY 10027, U.S.A.*

<sup>c</sup>*Thomas Jefferson National Accelerator Facility,  
12000 Jefferson Ave, Newport News, VA 23606, U.S.A.*

<sup>d</sup>*Physics Department, William & Mary,  
300 Ukrop Way, Williamsburg, VA 23187, U.S.A.*

<sup>e</sup>*Physics Department, Old Dominion University,  
4600 Elkhorn Ave, Norfolk, VA 23529, U.S.A.*

<sup>f</sup>*Aix Marseille Univ., Université de Toulon, CNRS, CPT,  
Marseille, France*

*E-mail:* [luigi.del.debbio@ed.ac.uk](mailto:luigi.del.debbio@ed.ac.uk), [tommaso.giani@ed.ac.uk](mailto:tommaso.giani@ed.ac.uk),  
[jmk2289@columbia.edu](mailto:jmk2289@columbia.edu), [kostas@wm.edu](mailto:kostas@wm.edu), [radyush@jlab.org](mailto:radyush@jlab.org),  
[savvas.zafeiropoulos@cpt.univ-mrs.fr](mailto:savvas.zafeiropoulos@cpt.univ-mrs.fr)

**ABSTRACT:** We extract two nonsinglet nucleon Parton Distribution Functions from lattice QCD data for reduced Ioffe-time pseudodistributions. We perform such analysis within the NNPDF framework, considering data coming from different lattice ensembles and discussing in detail the treatment of the different source of systematics involved in the fit. We introduce a recipe for taking care of systematics and use it to perform our extraction of light-cone PDFs.

**KEYWORDS:** Lattice field theory simulation, QCD Phenomenology

**ARXIV EPRINT:** [2010.03996](https://arxiv.org/abs/2010.03996)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Lattice data and observables</b>	<b>4</b>
<b>3</b>	<b>Fits over lattice data: statistical uncertainties only</b>	<b>7</b>
<b>4</b>	<b>Systematic effects</b>	<b>10</b>
4.1	Discussion	10
4.2	Results	13
<b>5</b>	<b>Conclusions</b>	<b>15</b>
<b>A</b>	<b>Pion mass dependence for 170 ensemble</b>	<b>17</b>
<b>B</b>	<b>Comparison with results from quasi-PDFs matrix elements</b>	<b>17</b>

---

## 1 Introduction

Parton Distribution Functions (PDFs), encoding the structure of the proton in terms of quarks and gluons, are one of the main ingredients required to do precise high-energy phenomenology. The available PDFs sets are extracted through global fits over experimental data [1–5]. Their non-perturbative nature makes them a natural candidate for a lattice QCD investigation, however it has been known for a long time that it is not possible to obtain them directly from first principle computations, due to the Euclidean metric of the lattice. In the last few years, several methods have been formulated [6, 7], which would allow us to compute on the lattice specific quantities that, in turn, can be related to PDFs through a factorization theorem. For a detailed discussion of the theoretical background, we refer the reader to recent reviews such as refs. [8–11].

Examples of such quantities are the equal time correlators underlying the definition of quasi- and pseudo-PDFs [12, 13], given by

$$M_{\mu}^{(0)}(z, P) = \langle P | \bar{\psi}^{(0)}(z) \gamma_{\mu} U^{(0)}(z, 0) \psi^{(0)}(0) | P \rangle, \quad (1.1)$$

with  $P$  denoting the momentum of the external proton states, while the suffix (0) reminds us that these are bare quantities. The matrix element of the vector bilocal operator of eq. (1.1) can be decomposed in terms of two form factors which only depend on the Lorentz invariants  $z^2$  and  $\nu \equiv -z \cdot P$  as

$$M_{\mu}^{(0)}(z, P) = 2P_{\mu} \mathcal{M}^{(0)}(\nu, z^2) + z_{\mu} \mathcal{N}^{(0)}(\nu, z^2). \quad (1.2)$$

As pointed out in [14], only the first form factor,  $\mathcal{M}^{(0)}$ , contains leading twist information. This can be seen by choosing a light-cone separation  $z = (0, z^-, 0_\perp)$  together with  $\gamma^\mu = \gamma^+$  and  $P = (P^+, 0, 0_\perp)$ , then we get

$$M_+^{(0)}(z, P) = 2P_+ \mathcal{M}^{(0)}(\nu, 0) = 2P_+ \int_{-1}^1 dx e^{ix\nu} f^{(0)}(x) \quad (1.3)$$

with  $f^{(0)}(x)$  being the bare collinear nonsinglet parton distribution. Because of the light-cone separation  $z$  involved in its definition,  $M_+^{(0)}$  is not directly computable on a Euclidean lattice. We can define a different quantity that is amenable to lattice simulations by choosing a purely spatial separation,  $z = (0, 0, 0, z_3)$ , together with  $\gamma^\mu = \gamma^0$  and  $P = (E, 0, 0, P_3)$ . Then taking the time component of eq. (1.2) we get

$$M_0^{(0)}(z, P) = 2E \mathcal{M}^{(0)}(\nu, -z_3^2). \quad (1.4)$$

The correlators defined in eqs. (1.3) and (1.4) are known in the literature as (bare) *Ioffe-time* distribution (ITD) and pseudodistribution (pseudo-ITD) respectively [13, 15]. For  $z_3^2 \neq 0$ , in addition to usual ultraviolet (UV) divergences (leading to coupling renormalization), they have specific link-related UV divergences, which are regularized by a finite lattice spacing  $a$ . Thus,  $\mathcal{M}^{(0)}(\nu, -z_3^2)$  is in fact  $\mathcal{M}^{(0)}(\nu, -z_3^2; a^2)$ .

The  $a \rightarrow 0$  UV divergences are multiplicatively renormalizable [16, 17]. The relevant renormalization factor  $Z(z_3^2, a^2)$  does not depend on  $\nu$  and, for small  $z_3^2$ , is known at one loop. Its explicit form is inessential if one introduces the so-called reduced Ioffe-time pseudo-distributions first defined in ref. [13] as

$$\mathfrak{M}(\nu, z_3^2) = \frac{\mathcal{M}(\nu, -z_3^2; a^2)}{\mathcal{M}(0, -z_3^2; a^2)}. \quad (1.5)$$

The  $Z$ -factors of the numerator and denominator are the same and cancel in the ratio leaving the reduced distribution on the left-hand side without any residual dependence on the lattice spacing.

Working in the small- $z_3^2$  limit, the pseudo-ITD can be matched at one-loop level to the corresponding ITD through a finite perturbative kernel, expressing the pseudo-ITD in terms of the collinear PDFs through a factorization formula based on the operator product expansion (OPE). The computation of the relevant QCD diagrams has been performed in a number of independent papers. The original QCD computation is reported, for example, in refs. [18–21]. A simple discussion of the basic features of the derivation of the factorization formula in non-gauge theories can be found in ref. [22].

The QCD result reads

$$\mathfrak{M}(\nu, z_3^2) = \int_{-1}^1 dx C(x\nu, \mu^2 z_3^2) f(x, \mu^2) + \mathcal{O}(z_3^2 \Lambda^2), \quad (1.6)$$

with

$$C(\xi, \mu^2 z_3^2) = e^{i\xi} - \frac{\alpha_s}{2\pi} C_F \int_0^1 dw \left[ \frac{1+w^2}{1-w} \log \left( z_3^2 \mu^2 \frac{e^{2\gamma_E+1}}{4} \right) + 4 \frac{\log(1-w)}{1-w} - 2(1-w) \right]_+ e^{i\xi w} + \mathcal{O}(\alpha_s^2). \quad (1.7)$$

Eqs. (1.6), (1.7) allow to relate collinear PDFs to quantities which are computable in lattice QCD simulations, through a factorized expression similar to those relating collinear PDFs to physical cross sections. In the spirit of the “good lattice cross sections” proposed in refs. [23, 24], this formula can be used in a fitting framework, to extract PDFs from lattice data, performing the same kind of analysis which is usually done when considering experimental data. This approach was first studied in ref. [25], and subsequently in refs. [26–28]. In ref. [27], it has been implemented within the `NNPDF` fitting environment. Considering data for quasi-PDFs matrix elements produced in refs. [29, 30] and starting from the momentum space factorization formula connecting quasi-PDFs to collinear PDFs, upon numerical implementation of the Fourier transform an expression analogous to eq. (1.6) was obtained, relating parton distributions directly to position space quasi-PDFs matrix elements. A similar analysis was very recently performed by the JAM collaboration in ref. [31] for the spin-averaged and spin-dependent PDFs employing quasi-PDF lattice data.

In the present work we perform an analogous exercise considering this time the reduced pseudo-ITD approach, implementing within the `NNPDF` environment the lattice data presented in refs. [32, 33], and using the position space factorization formula of eqs. (1.6), (1.7). This, besides being a complementary exercise to the one performed in ref. [27], has also some practical advantages. First, when working in the pseudo-ITD approach, the factorization is realized in the limit of small- $z^2$ . Unlike in the quasi-PDFs approach, where the factorization is realized for high values of  $P$ , here we are allowed to keep in the analysis data coming from a wide range of momentum values, without having to remove those with lower  $P$ . This advantage is particularly important, because in lattice QCD, the low momentum data are significantly more precise for a fixed computational cost. Second, we can directly use the position space factorization formula of eq. (1.6), relying on the analytical expression for the perturbative coefficient of eq. (1.7) and without having to perform the numerical Fourier transform described in the appendix A of ref. [27].

In this article we extend the general strategy that has been developed within the `NNPDF` framework and which allows us to systematically extract parton distributions from the available lattice data. In the implementation of this idea once the lattice data have reached some level of maturity in terms of precision and systematic effects, one could combine data from all pertinent lattice formalisms such as quasi-distributions [26, 29, 30, 34–50] and pseudo-distributions [28, 32, 33, 51–55]. One can also include results from the so-called “Good Lattice Cross-Sections” (LCS) approach, which is described in [23] and represents a general framework, where one computes matrix elements that can be factorized into PDFs at short distances. Papers [56–60] describe implementations of the latter formalism. Clearly a global analysis only makes sense after having scrutinised each set of data individually, and having understood the systematics that affect them. The structure of the paper is as follows. In section 2 we define the lattice observable considered in the fit, describe the corresponding data and briefly recall the main features of the NLO terms entering the factorization formulas. In section 3 we present the first set of results: we consider the fits where only the statistical uncertainties of the lattice data are taken into account. Analyzing data from different lattice ensembles we show that, in general, without accounting for systematic effects it is not possible to obtain a good fit. In section 4 we discuss and

quantify some of the systematic uncertainties affecting the lattice data. We include them into the analysis and study their impact on the final PDFs and on the fit quality. Section 5 summarizes our conclusions.

## 2 Lattice data and observables

In this section we describe the lattice observables we will consider in the present work, together with the corresponding data. By lattice observable, we mean a quantity which can be computed on the lattice on one hand, and related to some collinear PDFs through some kind of factorization theorem on the other. We will consider two different observables corresponding to the real and imaginary part of the reduced pseudo-ITD defined in eq. (1.5).

Considering the case of the unpolarized isovector parton distribution and recalling the definition of the two nonsinglet PDFs  $V_3$  and  $T_3$

$$V_3(x) = u(x) - \bar{u}(x) - [d(x) - \bar{d}(x)], \quad (2.1)$$

$$T_3(x) = u(x) - \bar{u}(x) + [d(x) - \bar{d}(x)], \quad (2.2)$$

taking the real and complex parts of eq. (1.6) and using eq. (1.7), we can define the two lattice observables

$$\text{Re}[\mathfrak{M}](\nu, -z_3^2) = \int_0^1 dx C^{\text{Re}}(x\nu, \mu^2 z_3^2) V_3(x, \mu^2), \quad (2.3)$$

$$\text{Im}[\mathfrak{M}](\nu, -z_3^2) = \int_0^1 dx C^{\text{Im}}(x\nu, \mu^2 z_3^2) T_3(x, \mu^2), \quad (2.4)$$

with

$$C^{\text{Re}}(\xi, \mu^2 z_3^2) = \cos(\xi) - \frac{\alpha_s}{2\pi} C_F \int_0^1 dw \left[ B(w) \log\left(z_3^2 \mu^2 \frac{e^{2\gamma_E+1}}{4}\right) + L(w) \right] \cos(\xi w), \quad (2.5)$$

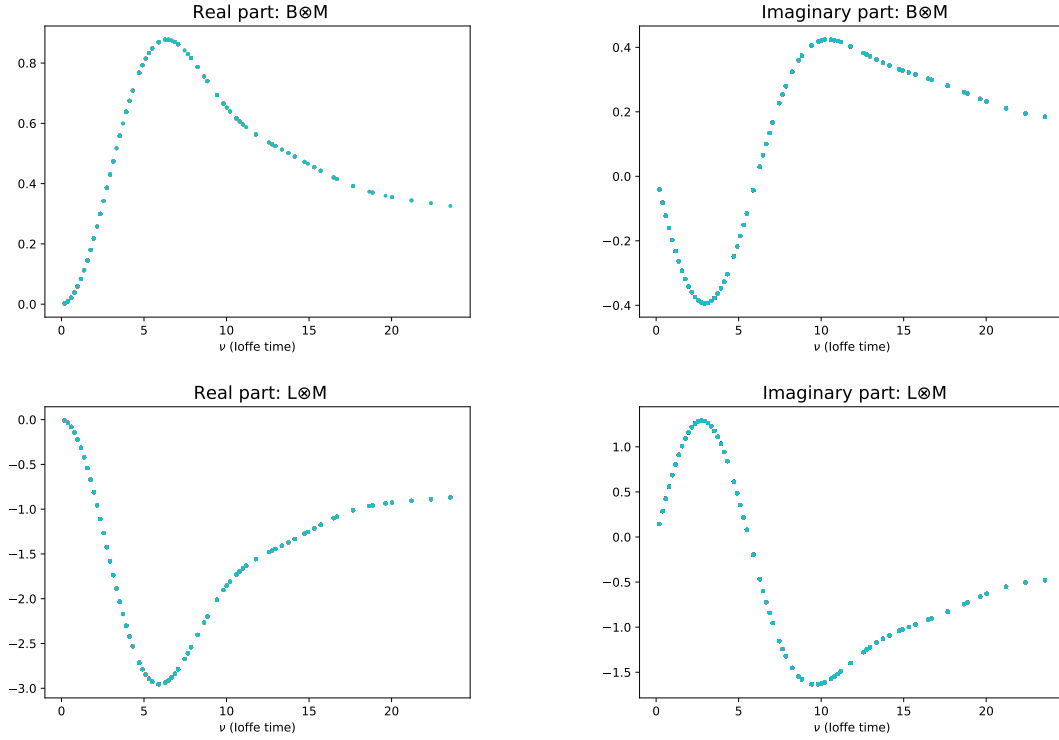
$$C^{\text{Im}}(\xi, \mu^2 z_3^2) = \sin(\xi) - \frac{\alpha_s}{2\pi} C_F \int_0^1 dw \left[ B(w) \log\left(z_3^2 \mu^2 \frac{e^{2\gamma_E+1}}{4}\right) + L(w) \right] \sin(\xi w), \quad (2.6)$$

where the kernels  $B(w)$  and  $L(w)$ , according to eq. (1.7), are given by

$$B(w) = \left[ \frac{1+w^2}{1-w} \right]_+, \quad (2.7)$$

$$L(w) = \left[ 4 \frac{\log(1-w)}{1-w} - 2(1-w) \right]_+. \quad (2.8)$$

It is worth recalling some important features of the NLO coefficients given in eqs. (2.5), (2.6). The contributions proportional to the two kernels  $B(w)$  and  $L(w)$  of eqs. (2.7), (2.8) can be seen as an evolution and a scheme change term respectively [32, 61]: while the former is responsible for the evolution from the PDF scale  $\hat{z}^{-2} = \mu^2 \frac{e^{2\gamma_E+1}}{4}$  to the pseudo-ITD scale  $z^2$ , the latter takes into account the finite terms characterizing the specific choice of the renormalization scheme. They are plotted in figure 1 for both the real and imaginary part, using the PDFs set NNPDF31\_nlo\_as\_0118 as input. The evolution



**Figure 1.** Upper plot: the NLO evolution term for the real (left) and imaginary part (right). Lower plot: the NLO scheme change term for the real (left) and imaginary part (right).

term  $B(w)$  also connects pseudo-ITD points having different values of  $z^2$ : considering for example the real part, from eqs. (2.3), (2.5) it follows

$$\begin{aligned} \text{Re}[\mathcal{M}](\nu, z_0^2) &= \text{Re}[\mathcal{M}](\nu, z^2) \\ &\quad - C_F \frac{\alpha_s}{2\pi} \log \frac{z_0^2}{z^2} \int_0^1 dx \left[ \int_0^1 dw B(w) \cos(x\nu w) \right] V_3(x, \mu^2), \end{aligned} \quad (2.9)$$

which relates the real part of the pseudo-ITD point at the scale  $z^2$  with the one having the same Ioffe time at the scale  $z_0^2$  [51, 62].

In the present work, we will consider the data for reduced pseudo-ITD from refs. [32, 33]: the datasets presented in ref. [32] have been produced starting from three different lattice ensembles, denoted as *fine*, *big* and *coarse* and which differ for the volume and lattice spacing used in the simulations. They have been produced using values of the pion mass ranging from 358 MeV (*fine*) to 415 MeV (*coarse* and *big*). In the present work we will focus on the datapoints produced from the *fine* ensemble, while those from the *coarse* and *big* ones will be used to estimate systematic effects due to continuum limit and finite lattice volume. We will also consider pseudo-ITD points presented in ref. [33], produced using pion mass equal to 172 MeV. Following the original convention of ref. [33] we will denote the corresponding lattice ensemble as *170*. Points from the ensemble *280*, presented in the same paper and produced using similar lattice spacing and pion mass 278 MeV, will be used to estimate the pion mass effects in the analyses for the ensembles *fine* and *170*. These

Lattice ensemble	a(fm)	$M_\pi$ (MeV)	$L^3 \times T$	$n_{\text{dat}}$	Reference
fine	0.094(1)	358(3)	$32^3 \times 64$	48	
big	0.127(2)	415(23)	$32^3 \times 96$	48	[32, 33]
coarse	0.127(2)	415(23)	$24^3 \times 64$	36	
280	0.094(1)	278(3)	$32^3 \times 64$	64	[33]
170	0.091(1)	172(6)	$64^3 \times 128$	80	

**Table 1.** Lattice data details.

five ensembles of  $2+1$  flavor lattice QCD were generated by the JLab/W&M collaboration using clover Wilson fermions and a tree level tadpole-improved Symanzik gauge action. One iteration of stout smearing with the weight  $\rho = 0.125$  for the staples is used in the fermion action. A direct consequence of the stout smearing is that the value of the tadpole corrected tree-level clover coefficient  $c_{\text{sw}}$  used is very close to the non-perturbative value determined, a posteriori, using the Schrödinger functional method. The detailed features of these ensembles are reported in table 1, together with the number of reduced pseudo-ITD datapoints  $n_{\text{dat}}$  computed from each of them.

Given a set of lattice data for the real and imaginary part of the reduced pseudo-ITD, the distributions  $T_3$  and  $V_3$  can be extracted from them through a standard minimum- $\chi^2$  fit, following the approach described in refs. [22, 27]: the unknown  $x$ -dependence of the PDFs is parameterized at the chosen scale  $\mu^2$ , using a suitable parametric form, whose best parameters are determined minimizing the  $\chi^2$  built using eqs. (2.3), (2.4) and the corresponding lattice results. In this work we will perform this exercise using the NNPDF fitting framework, running the same machinery commonly used to extract PDFs from experimental data, already applied to lattice results in ref. [27]. In the following we briefly recall its main relevant features, referring to ref. [27] for more details.

The  $x$ -dependence of the distributions  $f_q(x)$  ( $V_3$  and  $T_3$  in our case) is parameterized through a neural network  $\text{NN}_q$  multiplied by a preprocessing polynomial factor, as

$$f_q(x) = x^{\alpha_q} (1-x)^{\beta_q} \text{NN}_q(x), \quad (2.10)$$

$\alpha_q, \beta_q$  being additional free parameters to be determined during the fit, alongside the weights and biases defining the neural network. Denoting the free parameters of the model as  $\theta$ , the best fit is determined minimizing the  $\chi^2$  function, defined as

$$\chi^2(\theta) = \frac{1}{N_{\text{dat}}} \sum_{i,j} \left( \mathcal{O}(z_i) - \mathcal{O}^{\text{th}}(z_i, \theta) \right) \left[ \text{Cov}^{-1} \right]_{ij} \left( \mathcal{O}(z_j) - \mathcal{O}^{\text{th}}(z_j, \theta) \right), \quad (2.11)$$

where  $\mathcal{O}(z_i)$  denotes the measured lattice observable and  $\mathcal{O}^{\text{th}}(z_i, \theta)$  is the corresponding theoretical prediction, expressed using the matching coefficients of eqs. (2.5), (2.6) and the parameterized parton distribution of eq. (2.10). The implementation of the convolution entering eqs. (2.3), (2.4) is performed by means of FastKernel tables, introduced and validated in refs. [63, 64] in the context of global QCD fits, and currently used within the NNPDF code to obtain all the required theoretical predictions in a global fit. The covariance

matrix entering eq. (2.11) describes the distribution of the data, and takes into account the statistical and systematic uncertainties and their correlations. Considering  $N_c$  independent sources of correlated systematics, its explicit expression is given by

$$[\text{Cov}]_{ij} = \delta_{ij} \left( \sigma_{i,s}^{\text{stat}} \right)^2 + \sum_{l=1}^{N_c} \sigma_{i,l}^{\text{sys}} \sigma_{j,l}^{\text{sys}}, \quad (2.12)$$

where  $\sigma_i^{\text{stat}}$  and  $\sigma_{i,l}^{\text{sys}}$  represent the statistical and the  $l$ -th correlated systematic uncertainty of the  $i$ -th point.

The covariance matrix defined in eq. (2.12) enters both the definition of the  $\chi^2$  and the generation of Monte Carlo replicas [65], being therefore important for both the central value of the fit and the final PDFs error. A solid knowledge of the covariance matrix is therefore an essential ingredient to get reliable results. The minimization of the  $\chi^2$  is performed numerically: different algorithms can be implemented, here the CMA algorithm [66] is used, employing a cross-validation technique to avoid overfitting. The specific code used in the present work is the one employed for the production of the PDFs set NNP31 [1], together with the **ReportEngine** software [67].

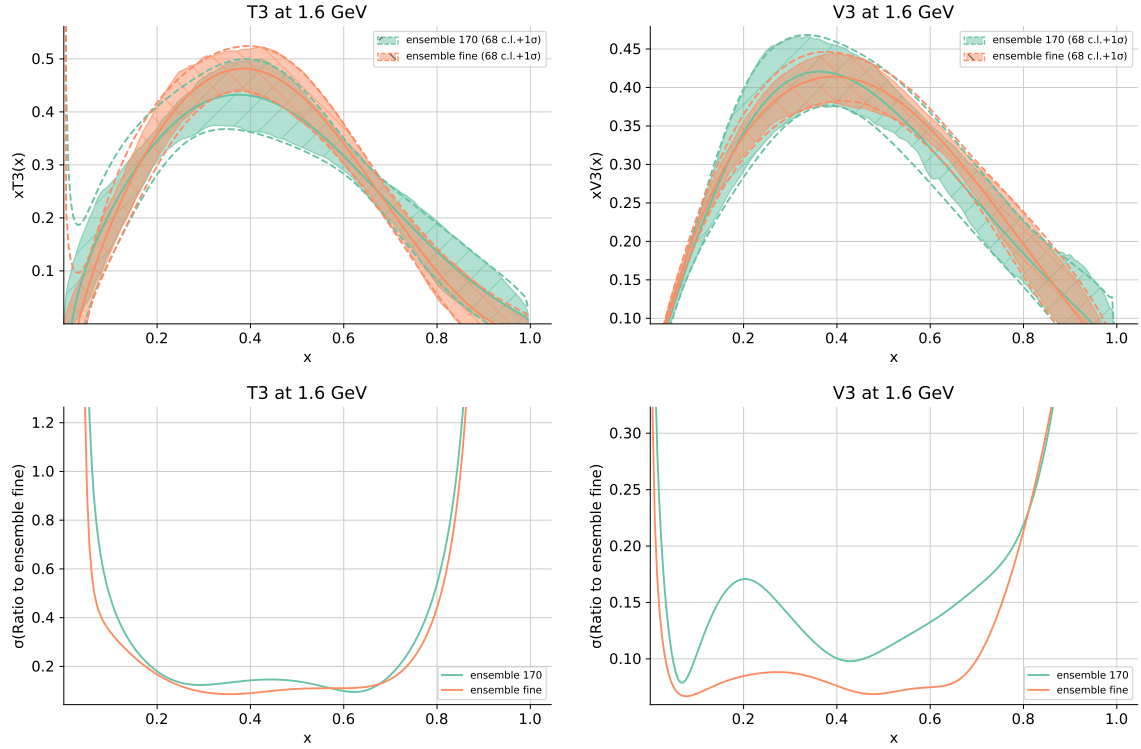
The NNP31 methodology has been used to produce PDF sets for many years now, and provides a flexible environment within which it has been possible to fit more than 4000 experimental points, coming from a variety of different high energy processes in different kinematic ranges [1, 65]. Therefore it represents a reliable framework which can be used to study and analyze the available lattice data, to assess how well these are able to constrain the PDFs and to compare lattice results with those coming from standard PDF sets. It is important to emphasise once again that in this analysis, once the FastKernel tables have been generated, the lattice data are treated exactly on the same footing as any other data, viz. the exact same methodology and code are used for fitting experimental and lattice data.

### 3 Fits over lattice data: statistical uncertainties only

In this section, we will present results for fits performed over the lattice data computed from the ensembles *fine* and *170*, denoted as *fine-stat* and *170-stat* respectively. Such fits have been produced considering statistical uncertainties only. We will show how, in general, without having the complete information regarding the lattice systematic uncertainties it is not always possible to obtain a good fit. In the next section, taking as example the case of the *fine* ensemble, we will discuss and estimate some of the possible systematic effects, studying their impact on the fit quality and on the resulting PDFs.

Parton distributions resulting from fits *fine-stat* and *170-stat*, together with the corresponding error bands, are plotted in the upper and lower plots of figure 2, and the  $\chi^2$  values are reported in table 2: despite the PDFs extracted from the two datasets are compatible within one  $\sigma$ , the error band of the fit *fine-stat* appears to be slightly smaller than the other, with an average  $\chi^2$  value per datapoint equal to 8.36, pointing out a possible underestimation of the error and a bad fit quality. This could be caused by inconsistencies between different datapoints, due to unknown systematic uncertainties affecting them. On





**Figure 2.** Upper plots: PDFs from datapoints computed from the ensembles fine and 170. The shaded bands represent the PDFs error computed as the 68 c.l. of the fit replicas, while the dashed line is obtained by computing the standard deviation point by point in  $x$ . Lower plots: corresponding PDFs errors, computed as standard deviation over fit replicas and displayed in function of  $x$ .

the other hand, the fit *170-stat* shows better  $\chi^2$  values, with an average value per datapoint equal to 1.38.

Focusing on the more problematic case of the fine ensemble results, in order to assess which points are more likely to be affected by large systematic errors, we will study the contribution to the  $\chi^2$  coming from each datapoint

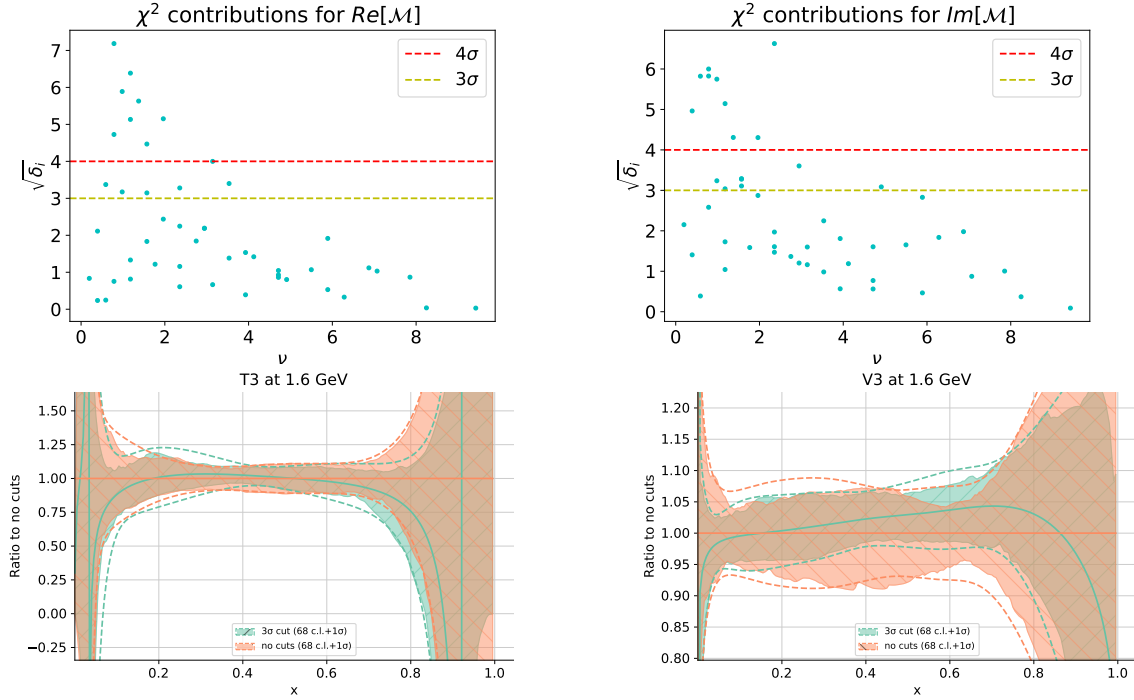
$$\delta_i = \frac{(D_i - T_i)^2}{\sigma_i^2}, \quad (3.1)$$

$D_i$  and  $T_i$  being the  $i$ -th lattice point and the corresponding prediction from the fit respectively, and find out which points  $D_i$  are more than  $4\sigma$  (or  $3\sigma$ ) off from the fitted distribution  $T_i$ . These are the points that, most likely, do not belong to the fitted distribution and which therefore might be affected by larger systematic effects.

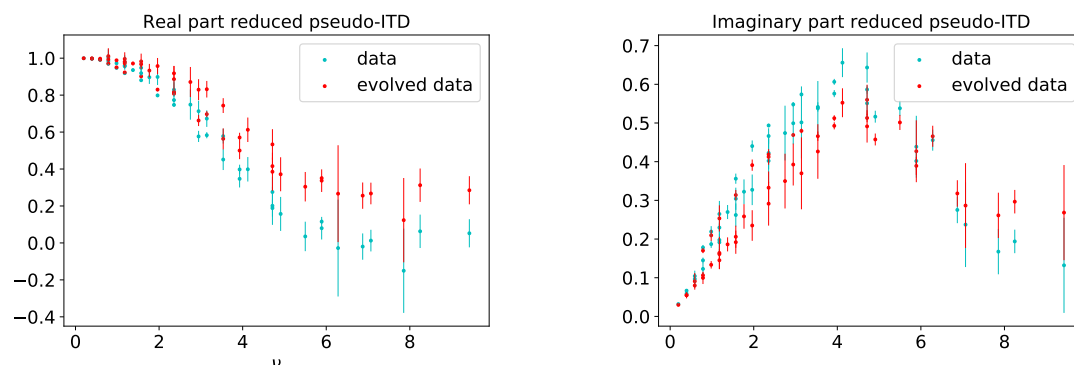
The contributions  $\sqrt{\delta_i}$  are plotted in the upper plot of figure 3 as a function of the Ioffe-time  $\nu$ , with the red and yellow lines highlighting the  $4\sigma$  and  $3\sigma$  cut respectively: it is clear that a bunch of points having small Ioffe-time values are those giving the highest contribution to the total  $\chi^2$ , being more than  $3\sigma$  or  $4\sigma$  off. We can implement  $4\sigma$  and  $3\sigma$  cuts, removing the problematic points from the dataset and producing new fits, denoted as *fine-stat-3σ* and *fine-stat-4σ*: the new fits show more reasonable  $\chi^2$  values, reported in

Ensemble	fit	Obs	$n_{\text{dat}}$	$\chi^2$	$\chi^2_{\text{tot}}$
fine	fine-stat	Re $[\mathfrak{M}]$	48	7.94	8.36
		Im $[\mathfrak{M}]$	48	8.77	
	fine-stat-4 $\sigma$	Re $[\mathfrak{M}]$	39	2.68	3.28
		Im $[\mathfrak{M}]$	39	3.89	
	fine-stat-3 $\sigma$	Re $[\mathfrak{M}]$	34	1.45	1.86
		Im $[\mathfrak{M}]$	32	2.27	
170	170-stat	Re $[\mathfrak{M}]$	80	0.68	1.38
		Im $[\mathfrak{M}]$	80	2.07	

**Table 2.** Details of fits with statistical uncertainties only. From left to right we report the lattice ensemble, the fit name, the observables included in the analysis, the number of datapoints and finally the partial and total  $\chi^2$ .



**Figure 3.** Upper plots:  $\sqrt{\delta_i}$  contributions for each datapoint of the fine ensemble. The red and yellow lines highlight the  $4\sigma$  and  $3\sigma$  cut respectively. Lower plots: PDFs from fits *fine-stat* (orange) and *fine-stat-3 $\sigma$*  (green), normalized to the former.



**Figure 4.** Data for the real part of the pseudo-ITD at their original scale  $z^2$  and evolved at the common scale  $z_0^2$ .

table 2, showing how, upon removing the outliers, the remaining points, coming from a wide range of momentum  $p$  and Euclidean separation  $z_3$ , are fitted reasonably well. The PDFs resulting from the  $3\sigma$  cut are plotted in the lower plot of figure 3, normalized to the fit without any cuts: it is clear how, despite spoiling the total  $\chi^2$ , the problematic points do not seem to have a big impact on the final PDFs.

We conclude that, depending on the specific lattice ensemble we consider, quite a high number of small Ioffe-time points do not belong to the fitted distribution. In order to get reasonable  $\chi^2$  values, such points have to be removed from the fit. This highlights possible tensions between datapoints and may point out the presence of systematic effects. In order to avoid any underestimation of the PDFs error and to introduce back in the analysis all the available points, systematic uncertainties need to be quantified and implemented in the fit.

## 4 Systematic effects

### 4.1 Discussion

The high  $\chi^2$  values of the fits presented in the previous section might point out the presence of some tensions between datapoints. In the following, focusing on the case of the fine ensemble results, we will show that this is indeed the case, and we will investigate possible sources of systematic uncertainties and their numerical values.

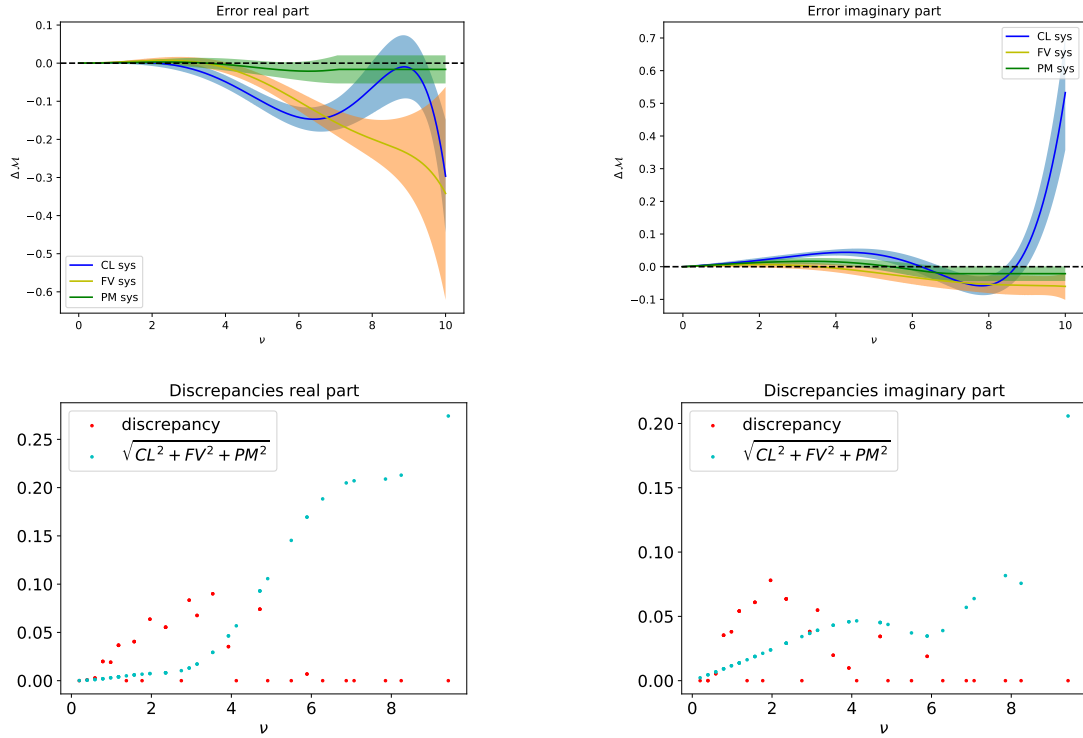
The matrix element defining the pseudo-ITD is a function of the Ioffe-time  $\nu$  and of the scale  $z^2$ . Points having the same Ioffe-time but different Euclidean separation can be related through eq. (2.9), which can be used to evolve each pseudo-ITD point up to a chosen reference scale  $z_0^2 = (0.7a)^2$ . Looking at figure 1 it is clear that, given this choice for  $z_0$ , the sign of the NLO correction of eq. (2.9) will be positive for every datapoint, so that the evolution increases the real part of the pseudo-ITD. Considering the imaginary part, the sign of the NLO evolution term is initially negative, and it turns positive at bigger values of  $\nu$ . Such effects can be seen in figure 4, where the pseudo-ITD points computed from the fine ensemble are plotted before (blue) and after evolution (red). After evolution, points having the same Ioffe time should have the same value. In practice, they should be

compatible within errors. Looking at the red points of figure 4, where each point is plotted with the corresponding statistical uncertainty, it is clear how, especially in the small Ioffe time region, this is not always the case: after evolution, some points having the same Ioffe time are not compatible between each other. Such discrepancies might be explained by the presence of systematic effects we are not accounting for.

A proper investigation of the systematic effects affecting the computation of the equal time correlators underlying the definition of pseudo-PDFs is a difficult and expensive task which would require to run different lattice simulations varying a set of parameters, like for example the lattice spacing, the lattice volume, the pion mass. Alongside systematic effects due to the lattice simulation, other sources of errors are those connected to the theoretical framework of the pseudo-PDFs approach, like the presence of higher twist effects and perturbative matching truncation effects. A detailed discussion of many of these uncertainties, together with a series of possible scenarios for their numerical values, can be found for example in ref. [27].

As mentioned in section 2, in ref. [32] additional pseudo-ITD points were computed starting from other two lattice ensembles, with pion mass similar to that of the fine one, but having different volume and lattice spacing, denoted as *big* and *coarse*, whose features are reported in table 1. Systematic uncertainties due to the continuum limit (CL) and finite volume (FV) can be directly estimated using these additional results as detailed in ref. [32]: the real and imaginary components of the pseudo-ITD are fitted to a polynomial as a function of the Ioffe-time  $\nu$ ; the difference between coarse and fine ensemble results is taken as an estimate for lattice spacing effects as a function of  $\nu$ , while the analogous difference considering the coarse and big ensembles gives an estimate for uncertainties due to finite lattice volume. Systematic effects due to the pion mass (PM) can be estimated in a similar way: as mentioned in section 2, in ref. [33] the data of the ensembles fine and 170 have been supplemented with additional pseudo-ITD results produced from a third ensemble having pion mass equal to 278 MeV, denoted as ensemble 280. The difference between polynomial fits for the ensembles fine and 280 is taken as an estimate for pion mass effects. These differences will be considered as three independent sources of correlated systematic, affecting each datapoint entering the analysis. They are shown in the upper plots of figure 5 as functions of the Ioffe-time, denoted as FV (finite volume), CL (continuum limit) and PM (pion mass).

It is important to understand whether or not these systematic uncertainties are enough to account for the discrepancies described at the beginning of the section. In the lower plots of figure 5 FV, CL and PM systematic effects are plotted for the relevant Ioffe-time values, together with the aforementioned discrepancies. Consistently with what observed previously, the latter seem to affect mostly low Ioffe-time points, which are also those for which the estimated systematics reach their minimum values. Therefore from figure 5 it follows that FV, CL and PM systematics cannot be considered responsible for the big contributions to the  $\chi^2$  noted in the fits of the previous section. In other words, they are likely not enough to account for the observed discrepancies affecting low Ioffe-time points. It should be noted that a study of more than 2 ensembles for each systematic error may be necessary for a more definitive conclusion.



**Figure 5.** Upper plots: finite volume (FV), continuum limit (CL) and pion mass (PM) systematics provided as functions of the ioffe-time  $\nu$  for the real (left) and imaginary (right) part of the matrix element. Lower plots: discrepancies between data having the same ioffe-time (red) together with the total FV, CL and PM systematic effects (blue).

Excited states contaminations might represent another possible source of systematic effects. Also missing higher orders in perturbation theory and higher twist effects could in principle be treated as additional systematic uncertainties. Unlike the case of the FV, CL and PM systematic uncertainties discussed above, we cannot estimate the size of such effects using the current lattice results. One could then follow the approach adopted in ref. [27], where different scenarios for the size of such systematics have been considered, and try to draw conclusions about their impact on the PDFs and on the fit quality. Here we will follow a different approach, trying to quantify an additional uncertainty which accounts for the unknown missing systematic effects, following a Bayesian approach as detailed in the following.

The figure of merit which is minimized during a Gaussian fit is defined as the probability of the data  $D$  given the model parameters  $\theta$ , namely the likelihood

$$\mathcal{P}(D|\theta) = e^{-\frac{1}{2}(D-T(\theta))^T \Sigma^{-1} (D-T(\theta))}. \quad (4.1)$$

where  $\Sigma$  is the covariance matrix of the data  $D$ , accounting for the known statistical and systematic uncertainties, and  $T(\theta)$  is the theoretical prediction, function of the model parameters. If we assume the presence of unknown systematic effect  $\Delta$  affecting the data-points  $D$ , eq. (4.1) can be modified as

$$\mathcal{P}(D, \Delta|\theta) = e^{-\frac{1}{2}(D+\Delta-T(\theta))^T \Sigma^{-1} (D+\Delta-T(\theta))}. \quad (4.2)$$

Assuming a Gaussian prior distribution  $\mathcal{P}(\Delta) = \exp\left[-\frac{1}{2}\Delta^T \hat{\Sigma}^{-1} \Delta\right]$  we can marginalize over  $\Delta$  getting

$$\int d\Delta \mathcal{P}(\Delta) \mathcal{P}(D, \Delta|\theta) \propto e^{-\frac{1}{2}(D-T(\theta))(\Sigma+\hat{\Sigma})^{-1}(D-T(\theta))}, \quad (4.3)$$

which defines the relevant likelihood to be minimized. Eq. (4.3) shows how the presence of unknown systematic effects can be accounted for by introducing in the likelihood an additional contribution to the covariance matrix, denoted by  $\hat{\Sigma}$ , which defines the prior probability distribution of these systematics. Its specific definition is of course arbitrary, and depends on the knowledge of the missing uncertainties we have.

This Bayesian approach, despite not providing a general method to estimate the missing systematics, allows to include in the analysis the partial information we may have about them. It should be noted that eq. (4.3) is obtained under the hypothesis that the unknown systematic uncertainty is gaussianly distributed. This is an assumption of the model, often done in the literature, which allows to greatly simplify the problem. However depending on the specific data considered it might not be the most realistic one. In the present work we will work using this gaussian assumption, leaving for future studies the investigation of different choices for the prior probability distribution. A gaussian Bayesian approach has already been applied in different physical problems, when the data are affected by unknown sources of systematics: in the case of global QCD analysis, in refs. [68, 69], a suitable covariance matrix  $\hat{\Sigma}$  was defined by mean of scale variations, in order to take into account the theoretical error due to missing higher orders, while in ref. [70] a similar approach was applied to cosmological data.

In our case, we only know the discrepancies observed at the beginning of this section, not described by continuum limit, finite volume and pion mass effects. We can look at such discrepancies as an indication of the minimal size of the systematic effects affecting the data and use them to construct a suitable  $\hat{\Sigma}$ : for each couple of points having a given Ioffe-time value, we will define the two corresponding diagonal components of  $\hat{\Sigma}$  as half of the distance between evolved points, setting the off diagonal elements to zero. Each point sharing the same Ioffe time value with at least another one will therefore be affected by an additional, uncorrelated systematic such that, after evolution, datapoints having the same Ioffe-time will be compatible between each other. Clearly, this global, uncorrelated systematic will be the dominant one for small Ioffe-time points, where most of the problematic points are, while for higher value of  $\nu$  lattice spacing, finite volume and mass effects will dominate.

## 4.2 Results

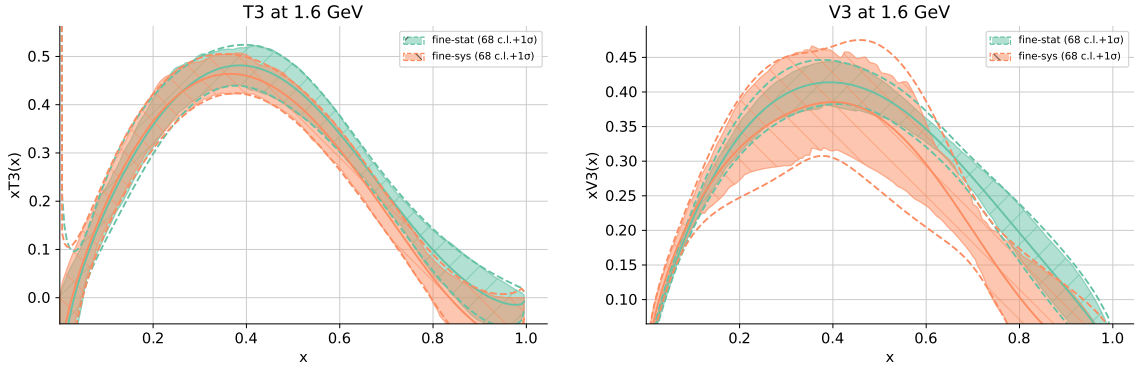
To sum up, in section 4.1, we have discussed and estimated four different source of systematics: the first three, accounting for finite volume, lattice spacing and pion mass effects, can be computed directly from the available lattice results as a function of the Ioffe-time  $\nu$ , and will be implemented in the fit as three independent sources of correlated systematics; the fourth one has been estimated using the size of the discrepancies observed between points having the same Ioffe-time, and will be considered as an additional uncorrelated uncertainty, in order to take into account the minimal size of all the remaining systematic effects

Ensemble	fit	Obs	$n_{\text{dat}}$	$\chi^2$	$\chi^2_{\text{tot}}$
fine	no cuts	Re $[\mathcal{M}]$	48	1.00	1.15
		Im $[\mathcal{M}]$	48	1.30	

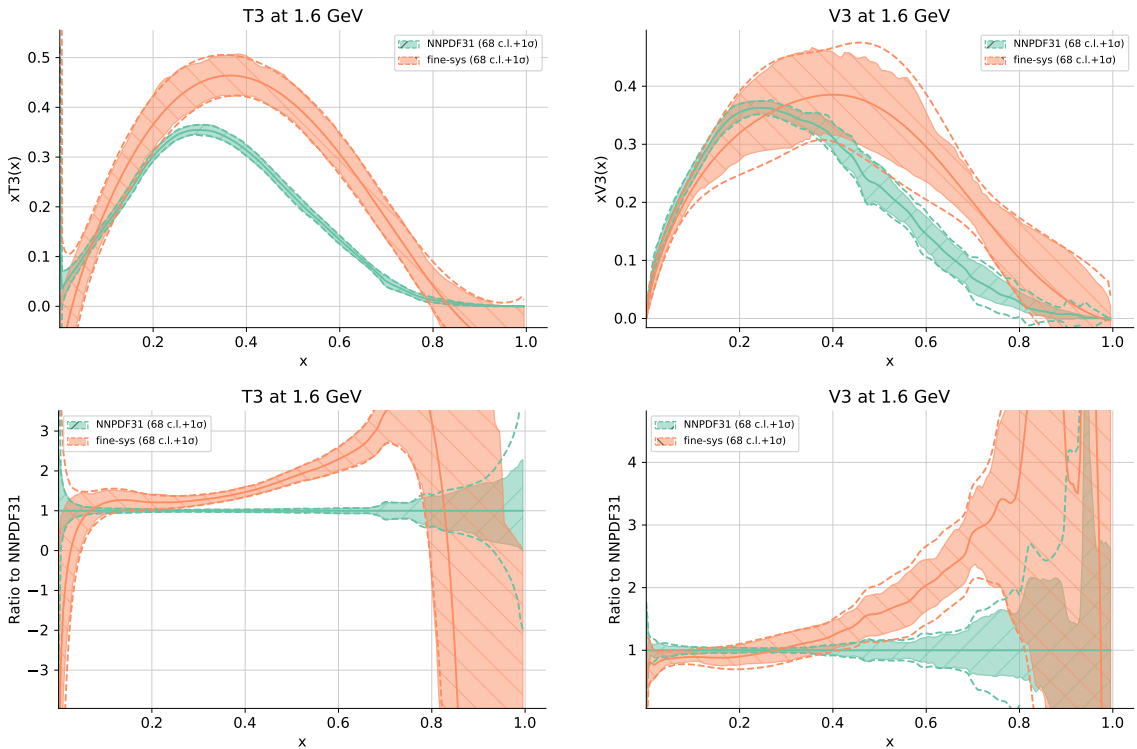
**Table 3.** Details of the fit with systematic uncertainties. From left to right we report the lattice ensemble, the fit name, the observables included in the analysis, the number of datapoints and finally the partial and total  $\chi^2$ .

we have not directly computed. As mentioned in section 2, such systematics enter the definition of the covariance matrix responsible for both replicas generation and the  $\chi^2$  definition, and therefore it has a central role in both the determination of the fit central value and its error band. The new fit is denoted as *fine-sys* and the resulting PDFs are plotted in figure 6, together with the results from the fit *fine-stat* presented section 3: the distribution  $T_3$  is only marginally affected by the introduction of the systematic errors, showing a mild down shift of its central value in the medium and large  $x$  regions; on the other hand both the central value and the error band of  $V_3$  change, with an overall down shift of the former and a sizable increase of the latter. The  $\chi^2$  values are reported in table 3: the average value per datapoint is now 1.15, showing a good fit quality. It should be noted that after the inclusion of systematic uncertainties in the analysis, the effect on the final result could be different depending on the specific situation we are considering. In other words, it is not always the case that the inclusion of new systematic effects leads to an increase of the final PDFs error. This can be seen for example in the case of the distribution  $T_3$  plotted in figure 6, from which it is clear how the error of the fit *fine-sys* has not increased with respect to the one of *fine-stat*. The reason for this can be traced back to the fact that the covariance matrix defined in eq. (2.12) enters both the Monte Carlo replicas generation and the  $\chi^2$  definition of eq. (2.11): while the former mostly controls the final PDFs error, the latter is responsible for the relative weights different points have in the analysis. Points affected by bigger errors will give smaller contributions to the  $\chi^2$  and therefore will count less in the fit. Each replica will be shifted by a certain amount, which takes into account both the new replicas distribution and the different weights of the data entering the  $\chi^2$ , so that the net effect on the final PDFs is non-trivial, and might consist in a global shift of the central value of the replicas distribution rather than in an increase of its error band.

Despite it is probably too early to draw comparisons between our results and phenomenological distributions, it is interesting to see how they look when plotted together: given the fact that nowadays  $V_3$  and  $T_3$  are very well constrained by experimental data, the discrepancies we observe between lattice and phenomenological results might be a good indication of the size of the systematic we are still missing, highlighting specific  $x$ -region where the lattice PDFs error might have been underestimated. In figure 7, our result *fine-sys* and the corresponding distributions from the NLO PDF sets NNPDF31 [1] are plotted together (orange and green curves respectively), both as absolute values (upper plots) and normalized to NNPDF31 (lower plots). Looking at results from *fine-sys*, in the case of both  $V_3$  and  $T_3$  the two distributions are compatible up to medium ( $\sim 0.25$  and  $\sim 0.45$ ) and for



**Figure 6.** PDFs from the fits *fine-stat* and *fine-sys*.



**Figure 7.** PDFs from the fits *fine-sys* compared with the corresponding distributions from NNPDF31. In the lower plots results are normalized to NNPDF31 PDFs.

large values of  $x$  ( $> 0.8$ ), showing a probable underestimation of the PDFs error for the intermediate  $x$  ranges.

## 5 Conclusions

In the present paper, we have considered the pseudo-ITD data produced in refs. [32, 33]. Using the position space factorization theorem relating such data to collinear PDFs, we have extracted two nonsinglet distributions within the NNPDF framework.



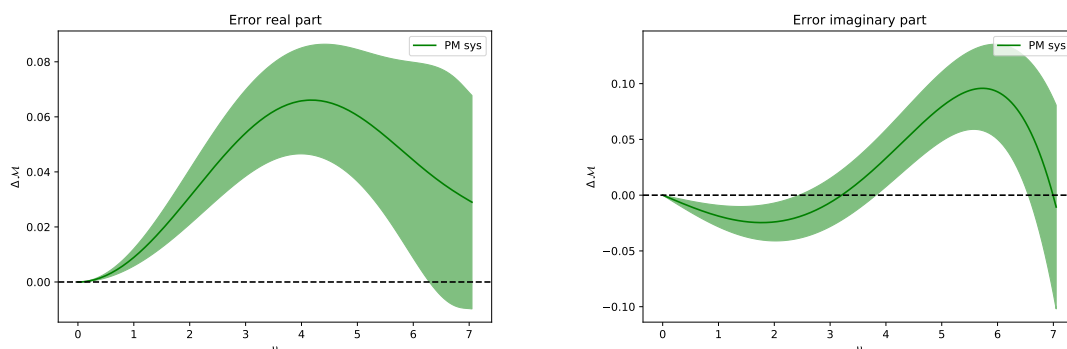
After extracting PDFs from different data sets and considering statistical uncertainties only, we have shown that in one of the cases considered, the fit quality appears to be really poor, pointing out the need for a detailed knowledge of the systematic effects. Using the results of refs. [32, 33] we have directly estimated those connected to finite volume, lattice spacing and pion mass effects. As for systematic uncertainties which cannot be directly computed from lattice results (like for example truncation effects and higher twist corrections), starting from the observed discrepancies between low Ioffe-time points we have used a Bayesian approach to introduce an additional systematic which allows us to mitigate the tensions between the problematic datapoints, using the partial pieces of information which are available to us.

The Bayesian approach however is not completely satisfying, since it relies on a partial knowledge of the missing uncertainties and requires to make a number of assumptions about them. More work has to be done to achieve a detailed knowledge of the systematic uncertainties in lattice simulations: without a stringent control over them, it is not possible to draw reliable conclusions and to make comparisons with phenomenological distributions.

Finally, we stress once more that the analysis performed in this paper is complementary to that presented in ref. [27], where quasi-PDFs matrix elements were considered instead, starting from the momentum space version of the factorization theorem. In both cases, results have been produced within the NNPDF environment, running the same machinery used for global QCD analysis over experimental data. The next logical step might be a global lattice QCD fit within this same framework, where data for multiple lattice observables coming from different simulations are simultaneously included in the analysis.

## Acknowledgments

We are grateful to E. R. Nocera for useful discussions and his interest in this work and to the organisers of 'Parton Distributions and Lattice Calculations' in Michigan in September 2019, which has been very useful for the present work. This work is supported by Jefferson Science Associates, LLC under U.S. DOE Contract #DE-AC05-06OR23177. KO was supported in part by U.S. DOE grant #DE-FG02-04ER41302 and in part by the Center for Nuclear Femtography grants C2-2020-FEMT-006, C2019-FEMT-002-05. AR was supported in part by U.S. DOE Grant #DE-FG02-97ER41028. JK was supported in part by the U.S. Department of Energy under contract DE-FG02-04ER41302, Department of Energy Office of Science Graduate Student Research fellowships, through the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program and is supported by U.S. Department of Energy grant DE-SC0011941. TG is supported by The Scottish Funding Council, grant H14027. The authors gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC) [71]. In addition, this work was made possible using results obtained at NERSC, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract #DE-AC02-05CH11231, as well as resources of the Oak Ridge Leadership



**Figure 8.** Pion mass (PM) systematic provided as functions of the ioffe-time  $\nu$  for the real (left) and imaginary (right) part of the matrix element.

Computing Facility (ALCC and INCITE) 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. The software libraries used on these machines were Chroma [72], QUDA [73, 74], QDP-JIT [75] and QPhiX [76, 77] developed with support from the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of Nuclear Physics, Scientific Discovery through Advanced Computing (SciDAC) program, and of the U.S. Department of Energy Exascale Computing Project.

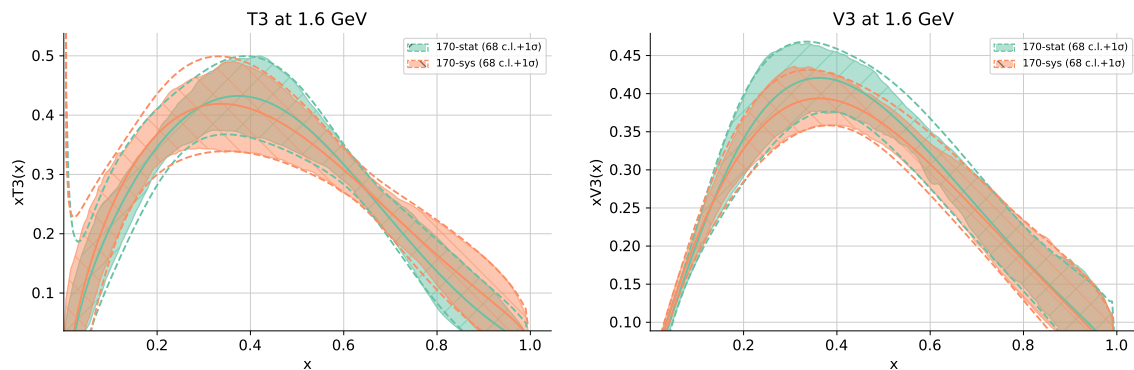
## A Pion mass dependence for 170 ensemble

Similarly to what done for the fine ensemble in section 4.1, the data for the ensemble 280 presented in ref. [33] can also be used to estimate pion mass effects for results concerning the ensemble 170. The corresponding polynomial curves are plotted in figure 8 as functions of the Ioffe-time.

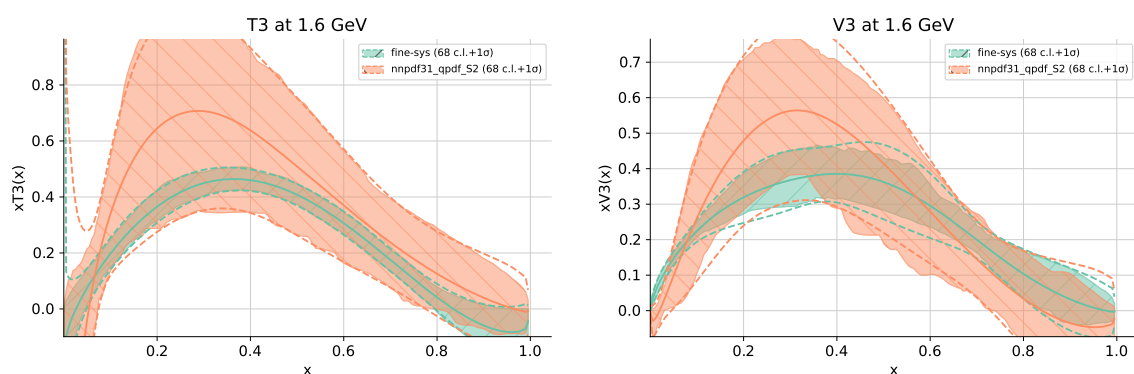
As in the case of the analysis for the fine ensemble, the curves in figure 8 are used to define a source of correlated systematic. The resulting PDFs, denoted as *170-sys*, are plotted in figure 9 together with the results for the ensemble 170 presented in section 3, where only statistical uncertainties have been considered. From figure 9 it is clear how introducing pion mass systematic effects in the analysis has very little impact on the distributions, the major effect being a mild down shift of the central value of  $V_3$  in the medium  $x$  region. We conclude that the mild pion mass dependence observed in pseudo-ITD data of ref. [33] has no sizable impact on the final PDFs.

## B Comparison with results from quasi-PDFs matrix elements

It is interesting to compare our best result *fine-sys* with the best result of ref. [27], denoted as *nnpdf31\_qpdf\_S2*. Both PDFs sets have been obtained using the same NNPDF methodology, the only difference being the input data (pesudo-ITD and quasi-PDFs data respectively) and the corresponding errors. For more details about the specific systematics uncertainties considered in the analysis for *nnpdf31\_qpdf\_S2* we refer to the original



**Figure 9.** PDFs from the fits *170-stat* and *170-sys*.



**Figure 10.** PDFs from the fits *fine-sys* compared with the corresponding distributions from the fit *nnpdf31\_qpdf\_S2*, presented in ref. [27].

publication [27]. Quasi-PDFs and pseudo-ITD results are plotted together in figure 10: both  $T_3$  and  $V_3$  distributions appear to be in good agreement, the main difference being a huge decrease in the PDFs error when considering results presented in this work. This difference can be partially traced back to the number of points included in the analysis: while in ref. [27] 16 points for quasi-PDFs matrix element were included, in the present work data corresponding to all momentum values are considered, for a total of 48 pseudo-ITD points. Clearly, having more points in the analysis allows to better constraint the fit results, giving final PDFs with smaller error. Given equivalent computational cost, the low momenta matrix elements, which are used in the pseudo-PDF approach, are exponentially more precise than the large momenta matrix elements, to which the quasi-PDF approach are restricted. The size of the statistical and systematic uncertainties affecting the points entering the two analyses is of course another reason for different PDFs error, however a detailed study of such differences is beyond the scope of this work. We leave a detailed comparison between the quasi- and pseudo-PDFs approaches for a future study.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [[arXiv:1706.00428](#)] [[INSPIRE](#)].
- [2] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev. D* **93** (2016) 033006 [[arXiv:1506.07443](#)] [[INSPIRE](#)].
- [3] S. Alekhin, J. Blümlein, S. Moch and R. Placakyte, *Parton distribution functions,  $\alpha_s$ , and heavy-quark masses for LHC run II*, *Phys. Rev. D* **96** (2017) 014011 [[arXiv:1701.05838](#)] [[INSPIRE](#)].
- [4] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J. C* **63** (2009) 189 [[arXiv:0901.0002](#)] [[INSPIRE](#)].
- [5] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [[arXiv:1412.7420](#)] [[INSPIRE](#)].
- [6] M. Constantinou et al., *Parton distributions and lattice QCD calculations: toward 3D structure*, [arXiv:2006.08636](#) [[INSPIRE](#)].
- [7] M. Constantinou, *The X-dependence of hadronic parton distributions: a review on the progress of lattice QCD*, in *38<sup>th</sup> international symposium on lattice field theory*, (2020) [[arXiv:2010.02445](#)] [[INSPIRE](#)].
- [8] A.V. Radyushkin, *Theory and applications of parton pseudodistributions*, *Int. J. Mod. Phys. A* **35** (2020) 2030002 [[arXiv:1912.04244](#)] [[INSPIRE](#)].
- [9] K. Cichy and M. Constantinou, *A guide to light-cone PDFs from lattice QCD: an overview of approaches, techniques and results*, *Adv. High Energy Phys.* **2019** (2019) 3036904 [[arXiv:1811.07248](#)] [[INSPIRE](#)].
- [10] C. Monahan, *Recent developments in  $x$ -dependent structure calculations*, *PoS(LATTICE2018)018* (2018) [[arXiv:1811.00678](#)] [[INSPIRE](#)].
- [11] X. Ji, Y.-S. Liu, Y. Liu, J.-H. Zhang and Y. Zhao, *Large-momentum effective theory*, [arXiv:2004.03543](#) [[INSPIRE](#)].
- [12] X. Ji, *Parton physics on a Euclidean lattice*, *Phys. Rev. Lett.* **110** (2013) 262002 [[arXiv:1305.1539](#)] [[INSPIRE](#)].
- [13] A.V. Radyushkin, *Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions*, *Phys. Rev. D* **96** (2017) 034025 [[arXiv:1705.01488](#)] [[INSPIRE](#)].
- [14] B.U. Musch, P. Hagler, J.W. Negele and A. Schafer, *Exploring quark transverse momentum distributions with lattice QCD*, *Phys. Rev. D* **83** (2011) 094507 [[arXiv:1011.1213](#)] [[INSPIRE](#)].
- [15] V. Braun, P. Gornicki and L. Mankiewicz, *Ioffe-time distributions instead of parton momentum distributions in description of deep inelastic scattering*, *Phys. Rev. D* **51** (1995) 6036 [[hep-ph/9410318](#)] [[INSPIRE](#)].
- [16] X. Ji, J.-H. Zhang and Y. Zhao, *Renormalization in large momentum effective theory of parton physics*, *Phys. Rev. Lett.* **120** (2018) 112001 [[arXiv:1706.08962](#)] [[INSPIRE](#)].
- [17] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu and S. Yoshida, *Renormalizability of quasiparton distribution functions*, *Phys. Rev. D* **96** (2017) 094019 [[arXiv:1707.03107](#)] [[INSPIRE](#)].

- [18] A.V. Radyushkin, *Quark pseudodistributions at short distances*, *Phys. Lett. B* **781** (2018) 433 [[arXiv:1710.08813](#)] [[INSPIRE](#)].
- [19] A. Radyushkin, *Nonperturbative evolution of parton quasi-distributions*, *Phys. Lett. B* **767** (2017) 314 [[arXiv:1612.05170](#)] [[INSPIRE](#)].
- [20] T. Izubuchi, X. Ji, L. Jin, I.W. Stewart and Y. Zhao, *Factorization theorem relating Euclidean and light-cone parton distributions*, *Phys. Rev. D* **98** (2018) 056004 [[arXiv:1801.03917](#)] [[INSPIRE](#)].
- [21] X. Ji, J.-H. Zhang and Y. Zhao, *More on large-momentum effective theory approach to parton physics*, *Nucl. Phys. B* **924** (2017) 366 [[arXiv:1706.07416](#)] [[INSPIRE](#)].
- [22] L. Del Debbio, T. Giani and C.J. Monahan, *Notes on lattice observables for parton distributions: nongauge theories*, *JHEP* **09** (2020) 021 [[arXiv:2007.02131](#)] [[INSPIRE](#)].
- [23] Y.-Q. Ma and J.-W. Qiu, *Exploring partonic structure of hadrons using ab initio lattice QCD calculations*, *Phys. Rev. Lett.* **120** (2018) 022003 [[arXiv:1709.03018](#)] [[INSPIRE](#)].
- [24] Y.-Q. Ma and J.-W. Qiu, *Extracting parton distribution functions from lattice QCD calculations*, *Phys. Rev. D* **98** (2018) 074021 [[arXiv:1404.6860](#)] [[INSPIRE](#)].
- [25] J. Karpie, K. Orginos, A. Rothkopf and S. Zafeiropoulos, *Reconstructing parton distribution functions from Ioffe time data: from Bayesian methods to neural networks*, *JHEP* **04** (2019) 057 [[arXiv:1901.05408](#)] [[INSPIRE](#)].
- [26] T. Izubuchi et al., *Valence parton distribution function of pion from fine lattice*, *Phys. Rev. D* **100** (2019) 034516 [[arXiv:1905.06349](#)] [[INSPIRE](#)].
- [27] K. Cichy, L. Del Debbio and T. Giani, *Parton distributions from lattice data: the nonsinglet case*, *JHEP* **10** (2019) 137 [[arXiv:1907.06037](#)] [[INSPIRE](#)].
- [28] X. Gao et al., *Valence parton distribution of the pion from lattice QCD: approaching the continuum limit*, *Phys. Rev. D* **102** (2020) 094513 [[arXiv:2007.06590](#)] [[INSPIRE](#)].
- [29] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato and F. Steffens, *Light-cone parton distribution functions from lattice QCD*, *Phys. Rev. Lett.* **121** (2018) 112001 [[arXiv:1803.02685](#)] [[INSPIRE](#)].
- [30] C. Alexandrou et al., *Systematic uncertainties in parton distribution functions from lattice QCD simulations at the physical point*, *Phys. Rev. D* **99** (2019) 114504 [[arXiv:1902.00587](#)] [[INSPIRE](#)].
- [31] J. Bringewatt, N. Sato, W. Melnitchouk, J.-W. Qiu, F. Steffens and M. Constantinou, *Confronting lattice parton distributions with global QCD analysis*, *Phys. Rev. D* **103** (2021) 016003 [[arXiv:2010.00548](#)] [[INSPIRE](#)].
- [32] B. Joó, J. Karpie, K. Orginos, A. Radyushkin, D. Richards and S. Zafeiropoulos, *Parton distribution functions from Ioffe time pseudo-distributions*, *JHEP* **12** (2019) 081 [[arXiv:1908.09771](#)] [[INSPIRE](#)].
- [33] B. Joó, J. Karpie, K. Orginos, A.V. Radyushkin, D.G. Richards and S. Zafeiropoulos, *Parton distribution functions from Ioffe time pseudodistributions from lattice calculations: approaching the physical point*, *Phys. Rev. Lett.* **125** (2020) 232003 [[arXiv:2004.01687](#)] [[INSPIRE](#)].
- [34] H.-W. Lin, J.-W. Chen, S.D. Cohen and X. Ji, *Flavor structure of the nucleon sea from lattice QCD*, *Phys. Rev. D* **91** (2015) 054510 [[arXiv:1402.1462](#)] [[INSPIRE](#)].

- [35] J.-W. Chen, S.D. Cohen, X. Ji, H.-W. Lin and J.-H. Zhang, *Nucleon helicity and transversity parton distributions from lattice QCD*, *Nucl. Phys. B* **911** (2016) 246 [[arXiv:1603.06664](#)] [[INSPIRE](#)].
- [36] C. Alexandrou et al., *Lattice calculation of parton distributions*, *Phys. Rev. D* **92** (2015) 014502 [[arXiv:1504.07455](#)] [[INSPIRE](#)].
- [37] C. Alexandrou et al., *Updated lattice results for parton distributions*, *Phys. Rev. D* **96** (2017) 014513 [[arXiv:1610.03689](#)] [[INSPIRE](#)].
- [38] C. Monahan and K. Orginos, *Quasi parton distributions and the gradient flow*, *JHEP* **03** (2017) 116 [[arXiv:1612.01584](#)] [[INSPIRE](#)].
- [39] J.-H. Zhang, J.-W. Chen, X. Ji, L. Jin and H.-W. Lin, *Pion distribution amplitude from lattice QCD*, *Phys. Rev. D* **95** (2017) 094514 [[arXiv:1702.00008](#)] [[INSPIRE](#)].
- [40] C. Alexandrou et al., *A complete non-perturbative renormalization prescription for quasi-PDFs*, *Nucl. Phys. B* **923** (2017) 394 [[arXiv:1706.00265](#)] [[INSPIRE](#)].
- [41] J. Green, K. Jansen and F. Steffens, *Nonperturbative renormalization of nonlocal quark bilinears for parton quasidistribution functions on the lattice using an auxiliary field*, *Phys. Rev. Lett.* **121** (2018) 022004 [[arXiv:1707.07152](#)] [[INSPIRE](#)].
- [42] J.-W. Chen et al., *Lattice calculation of parton distribution function from LaMET at physical pion mass with large nucleon momentum*, [arXiv:1803.04393](#) [[INSPIRE](#)].
- [43] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato and F. Steffens, *Transversity parton distribution functions from lattice QCD*, *Phys. Rev. D* **98** (2018) 091503 [[arXiv:1807.00232](#)] [[INSPIRE](#)].
- [44] H.-W. Lin et al., *Proton isovector helicity distribution on the lattice at physical pion mass*, *Phys. Rev. Lett.* **121** (2018) 242003 [[arXiv:1807.07431](#)] [[INSPIRE](#)].
- [45] Z.-Y. Fan, Y.-B. Yang, A. Anthony, H.-W. Lin and K.-F. Liu, *Gluon quasi-parton-distribution functions from lattice QCD*, *Phys. Rev. Lett.* **121** (2018) 242001 [[arXiv:1808.02077](#)] [[INSPIRE](#)].
- [46] Y.-S. Liu et al., *Nucleon transversity distribution at the physical pion mass from lattice QCD*, [arXiv:1810.05043](#) [[INSPIRE](#)].
- [47] J.R. Green, K. Jansen and F. Steffens, *Improvement, generalization, and scheme conversion of Wilson-line operators on the lattice in the auxiliary field approach*, *Phys. Rev. D* **101** (2020) 074509 [[arXiv:2002.09408](#)] [[INSPIRE](#)].
- [48] Y. Chai et al., *Parton distribution functions of  $\Delta^+$  on the lattice*, *Phys. Rev. D* **102** (2020) 014508 [[arXiv:2002.12044](#)] [[INSPIRE](#)].
- [49] H.-W. Lin, J.-W. Chen, Z. Fan, J.-H. Zhang and R. Zhang, *Valence-quark distribution of the kaon and pion from lattice QCD*, *Phys. Rev. D* **103** (2021) 014516 [[arXiv:2003.14128](#)] [[INSPIRE](#)].
- [50] C. Alexandrou et al., *Unpolarized and helicity generalized parton distributions of the proton within lattice QCD*, *Phys. Rev. Lett.* **125** (2020) 262001 [[arXiv:2008.10573](#)] [[INSPIRE](#)].
- [51] K. Orginos, A. Radyushkin, J. Karpie and S. Zafeiropoulos, *Lattice QCD exploration of parton pseudo-distribution functions*, *Phys. Rev. D* **96** (2017) 094503 [[arXiv:1706.05373](#)] [[INSPIRE](#)].



- [52] J. Karpie, K. Orginos and S. Zafeiropoulos, *Moments of Ioffe time parton distribution functions from non-local matrix elements*, *JHEP* **11** (2018) 178 [[arXiv:1807.10933](#)] [[INSPIRE](#)].
- [53] B. Joó et al., *Pion valence structure from Ioffe-time parton pseudodistribution functions*, *Phys. Rev. D* **100** (2019) 114512 [[arXiv:1909.08517](#)] [[INSPIRE](#)].
- [54] M. Bhat, K. Cichy, M. Constantinou and A. Scapellato, *Parton distribution functions from lattice QCD at physical quark masses via the pseudo-distribution approach*, [arXiv:2005.02102](#) [[INSPIRE](#)].
- [55] Z. Fan, R. Zhang and H.-W. Lin, *Nucleon gluon distribution function from 2 + 1 + 1-flavor lattice QCD*, [arXiv:2007.16113](#) [[INSPIRE](#)].
- [56] G.S. Bali et al., *Pion distribution amplitude from Euclidean correlation functions*, *Eur. Phys. J. C* **78** (2018) 217 [[arXiv:1709.04325](#)] [[INSPIRE](#)].
- [57] G.S. Bali et al., *Pion distribution amplitude from Euclidean correlation functions: exploring universality and higher-twist effects*, *Phys. Rev. D* **98** (2018) 094507 [[arXiv:1807.06671](#)] [[INSPIRE](#)].
- [58] R.S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu and D.G. Richards, *Pion valence quark distribution from matrix element calculated in lattice QCD*, *Phys. Rev. D* **99** (2019) 074507 [[arXiv:1901.03921](#)] [[INSPIRE](#)].
- [59] RQCD collaboration, *Light-cone distribution amplitudes of octet baryons from lattice QCD*, *Eur. Phys. J. A* **55** (2019) 116 [[arXiv:1903.12590](#)] [[INSPIRE](#)].
- [60] R.S. Sufian et al., *Pion valence quark distribution from current-current correlation in lattice QCD*, *Phys. Rev. D* **102** (2020) 054508 [[arXiv:2001.04960](#)] [[INSPIRE](#)].
- [61] A. Radyushkin, *One-loop evolution of parton pseudo-distribution functions on the lattice*, *Phys. Rev. D* **98** (2018) 014019 [[arXiv:1801.02427](#)] [[INSPIRE](#)].
- [62] J. Karpie, K. Orginos, A. Radyushkin and S. Zafeiropoulos, *Parton distribution functions on the lattice and in the continuum*, *EPJ Web Conf.* **175** (2018) 06032 [[arXiv:1710.08288](#)] [[INSPIRE](#)].
- [63] R.D. Ball et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, *Nucl. Phys. B* **838** (2010) 136 [[arXiv:1002.4407](#)] [[INSPIRE](#)].
- [64] V. Bertone, S. Carrazza and N.P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205 [[arXiv:1605.02070](#)] [[INSPIRE](#)].
- [65] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040 [[arXiv:1410.8849](#)] [[INSPIRE](#)].
- [66] N. Hansen, *The CMA evolution strategy: a tutorial*, [arXiv:1604.00772](#).
- [67] Z. Kassabov, *Reportengine: a framework for declarative data analysis*, *Zenodo*, (2019).
- [68] NNPDF collaboration, *A first determination of parton distributions with theoretical uncertainties*, *Eur. Phys. J. C* (2019) 79:838 [[arXiv:1905.04311](#)] [[INSPIRE](#)].
- [69] NNPDF collaboration, *Parton distributions with theory uncertainties: general formalism and first phenomenological studies*, *Eur. Phys. J. C* **79** (2019) 931 [[arXiv:1906.10698](#)] [[INSPIRE](#)].

- [70] J. Luis Bernal and J.A. Peacock, *Conservative cosmology: combining data with allowance for unknown systematics*, *JCAP* **07** (2018) 002 [[arXiv:1803.04470](#)] [[INSPIRE](#)].
- [71] D. Krause and P. Thörnig, *JURECA: modular supercomputer at Jülich supercomputing centre*, *J. Large-Scale Res. Facil. JLSRF* **4** (2018) A132.
- [72] SciDAC, LHPC and UKQCD collaborations, *The Chroma software system for lattice QCD*, *Nucl. Phys. B Proc. Suppl.* **140** (2005) 832 [[hep-lat/0409003](#)] [[INSPIRE](#)].
- [73] M.A. Clark, R. Babich, K. Barros, R.C. Brower and C. Rebbi, *Solving lattice QCD systems of equations using mixed precision solvers on GPUs*, *Comput. Phys. Commun.* **181** (2010) 1517 [[arXiv:0911.3191](#)] [[INSPIRE](#)].
- [74] R. Babich, M.A. Clark and B. Joo, *Parallelizing the QUDA library for multi-GPU calculations in lattice quantum chromodynamics*, in *SC 10 (Supercomputing 2010)*, (2010) [[arXiv:1011.0024](#)] [[INSPIRE](#)].
- [75] F.T. Winter, M.A. Clark, R.G. Edwards and B. Joó, *A framework for lattice QCD calculations on GPUs*, in *28<sup>th</sup> IEEE international parallel and distributed processing symposium*, *IEEE*, (2014) [[arXiv:1408.5925](#)] [[INSPIRE](#)].
- [76] B. Joó et al., *Lattice QCD on Intel® Xeon Phi™ coprocessors*, *Lect. Notes Comput. Sci.* **7905** (2013) 40.
- [77] B. Joó, D.D. Kalamkar, T. Kurth, K. Vaidyanathan and A. Walden, *Optimizing Wilson-Dirac operator and linear solvers for Intel® KNL*, *Lect. Notes Comput. Sci.* **9945** (2016) 415.