The impact of data from future lepton colliders on light hadrons fragmentation functions

Bin Zhou

INPAC, Shanghai Jiao Tong University

Based on JHEP 02, 003 (2025) with Jun Gao

Opportunities and Ideas at the QCD Frontier, April 8, 2025

Outline

Outline

Fragmentation functions

- Fragmentations of quarks and gluons into hadrons have been the central topic of QCD since only hadrons are observed experimentally
- In its simplest form, fragmentation functions (FFs) describe number density of the identified hadron wrt the fraction of momentum of the initial parton it carries, as measured in single inclusive hadron production, e.g., from single-inclusive annihilation (SIA), semi-inclusive DIS (SIDIS), pp collisions[1]

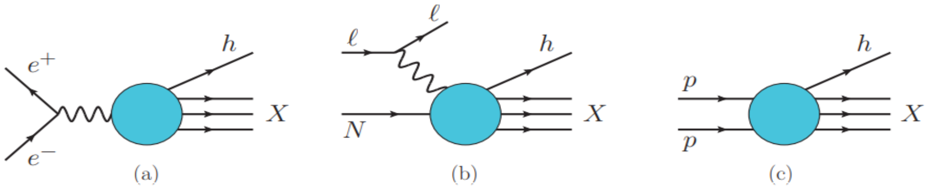


Figure 1: Representation of the scattering amplitude for processes whose QCD description involves FFs: Single-inclusive hadron production in (a) e^+e^- annihilation, (b) deep-inelastic lepton-nucleon scattering, (c) proton-proton scattering.

- As the FFs are related to a nonperturbative aspect of QCD, they cannot be calculated from the first principle of QCD at present and are in general extracted from fits to a variety of experimental data.

[1] A. Metz and A. Vossen, "Parton Fragmentation Functions", Prog. Part. Nucl. Phys. **91**, 136–202 (2016).

Current situations

Many studies have already been conducted on a comprehensive fit involving various data samples to extract FFs.

- DSS, AKK, NNFF, MAPFF, JAM, and NPC23, which combine data sets from SIA, SIDIS, or pp collisions
- HKNS, which only includes SIA data

} NLO in QCD

Furthermore, determinations of FFs have been made at next-to-next-to-leading order (NNLO) with SIA data only[1 2 3 4], at approximate NNLO with SIA and SIDIS data[5 6], and most recently at full NNLO accuracy in QCD for charged pions and kaons using a global analysis of SIA and SIDIS data[7].

In the present study, we focus on the fit based solely on the data from the lepton collider.

- the separation between quark and antiquark FFs
- the gluon FF

[1] V. Bertone et al. (NNPDF), "A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties", Eur. Phys. J. C **77**, 516 (2017).

[2] M. Soleymaninia et al., "First QCD analysis of charged hadron fragmentation functions and their uncertainties at next-to-next-to-leading order", Phys. Rev. D **98**, 074002 (2018).

[3] M. Soleymaninia et al., "Impact of unidentified light charged hadron data on the determination of pion fragmentation functions", Phys. Rev. D **99**, 034024 (2019).

[4] M. Soleymaninia et al., "Simultaneous extraction of fragmentation functions of light charged hadrons with mass corrections", Phys. Rev. D **103**, 054045 (2021).

[5] I. Borsa et al., "Towards a Global QCD Analysis of Fragmentation Functions at Next-to-Next-to-Leading Order Accuracy", Phys. Rev. Lett. **129**, 012002 (2022).

[6] R. Abdul Khalek et al. (MAP (Multi-dimensional Analyses of Partonic distributions)), "Pion and kaon fragmentation functions at next-to-next-to-leading order", Phys. Lett. B **834**, 137456 (2022).

[7] J. Gao et al., "Fragmentation functions of charged hadrons at next-to-next-to-leading order and constraints on proton PDFs", (2025).

Opportunities with future lepton colliders

The run of future lepton colliders, such as ILC[1], CEPC[2], and FCC–*ee*[3], will collect high-quality and diverse data on hadron production, sufficiently for precision determination of fragmentation functions alone

Particle	$E_{c.m.}$ (GeV)	Years	SR Power (MW)	Lumi. /IP ($10^{34}cm^{-2}s^{-1}$)	Integrated Lumi. /yr (ab^{-1}, 2 IPs)	Total Integrated L (ab^{-1}, 2 IPs)	Total no. of events
H*	240	10	50	8.3	2.2	21.6	4.3×10^6
			30	5	1.3	13	2.6×10^6
Z	91	2	50	192**	50	100	4.1×10^{12}
			30	115**	30	60	2.5×10^{12}
W	160	1	50	26.7	6.9	6.9	2.1×10^8
			30	16	4.2	4.2	1.3×10^8
$t\bar{t}$	360	5	50	0.8	0.2	1.0	0.6×10^6
			30	0.5	0.13	0.65	0.4×10^6

CEPC operation plans

Working point	Luminosity (10^{34} cm^{-2} s^{-1})	Tot. lum./year (ab^{-1})/year	Goal (ab^{-1})	Run time (years)
Z first two years	100	24	150	4
Z other years	200	48		
W	25	6	10	1–2
H	7.0	1.7	5	3
RF reconfiguration				1
$t\bar{t}$ 350 GeV (first year)	0.8	0.20	0.2	1
$t\bar{t}$ 365 GeV	1.4	0.34	1.5	4

FCC–*ee* operation plans

	$\int \mathcal{L}dt$ [fb^{-1}]			
$\sqrt{s}$	G20	H20	I20	Snow
250 GeV	500	2000	500	1150
350 GeV	200	200	1700	200
500 GeV	5000	4000	4000	1600

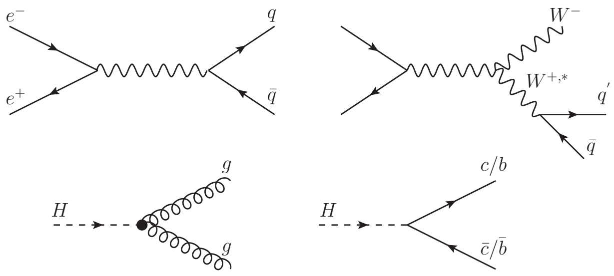
ILC operation plans

[1] A. Aryshev et al. (ILC International Development Team), "The International Linear Collider: Report to Snowmass 2021", (2022).

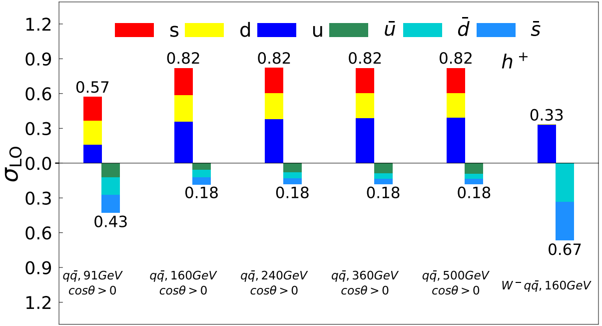
[2] W. Abdallah et al. (CEPC Study Group), "CEPC Technical Design Report – Accelerator (v2)", (2023).

[3] A. Abada et al. (FCC), "FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2", Eur. Phys. J. ST **228**, 261–623 (2019).

Opportunities with future lepton colliders



FFs–sensitive processes



LO cross section for various light quarks and antiquarks production

- producing qqbar samples at various energies with kinematic cuts, important for light quark flavor and charge separation
- heavy-quark enriched samples and gluon samples from Higgs hadronic decays
- further quark flavor and charge separation from W-boson production with hadronic decays

Outline

Pseudo-data generation at NLO in QCD

If $\sigma_i^{\rm th}$ is the theoretical cross-section for bin i of a given process, then the central values of the pseudo–data $\sigma_i^{\rm exp}$ is constructed by means of

$$\sigma_i^{\rm exp} = \sigma_i^{\rm th} \times (1 + r_i \cdot \delta_{{\rm tot},\,i}^{\rm exp} + s \cdot \delta_{\rm cor}^{\rm exp})$$

- $\sigma_i^{\rm th}$: obtained using FMNLO[1] with the NPC23 set as input.
- r_i, s: univariate Gaussian random numbers
- $\delta_{{\rm tot},i}^{\rm exp}$: defined as $\left((\delta_{{\rm stat},i}^{\rm exp})^2 + (\delta_{{\rm sys},i}^{\rm exp})^2\right)^{1/2}$, where statistical uncertainties are evaluated from the expected number of events per bin and assume the systematic uncertainties are the same as those in the measurements from SLD at Z–pole
- $\delta_{\rm cor}^{\rm exp}$: derived from the reference SLD measurement at Z–pole, 1%

Why Not Use General-Purpose Event Generators?

- Generators (e.g., PYTHIA, HERWIG) rely on phenomenological models (Lund string, cluster) for hadronization.
- Most generators implement LO parton showers (one-loop DGLAP evolution)

[1] C. Liu et al., "Automated calculation of jet fragmentation at NLO in QCD", JHEP **09**, 108 (2023).

Selection of data

- N×2
- $W^- W^{+,*}$ production
- CEPC: 1831 points in total

e^+e^- annihilation							
$\sqrt{s}$ (GeV)	luminosity (ab^{-1})			final state	kinematic cuts	hadrons	N_{pt}
	CEPC	FCC-*ee*	ILC				
91.2	60	150	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	132
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	65
160	4.2	-	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	168
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	83
161	-	10	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	168
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	83
240	13	5	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	186
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	92
250	-	-	2	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	186
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	92
350	-	0.2	0.2	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	198
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	98
360	0.65	-	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	198
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	98
365	-	1.5	-	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	198
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	98
500	-	-	4	$q\bar{q}$	$\cos(\theta) > 0$	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	198
				$c\bar{c}/b\bar{b}$	-	$\pi^\pm, K^\pm, p/\bar{p}$	98
W boson decay channels							
$\sqrt{s}$ (GeV)	# events (million)			final state	kinematic cuts	hadrons	N_{pt}
	CEPC	FCC-*ee*	ILC				
80.419	116	68	62	$W^- W^{+*} \to W^- q\bar{q}$	-	$\pi^+, \pi^-, K^+, K^-, p, \bar{p}$	120
	58	34	31	$W^- W^{+*} \to W^- c\bar{s}$			
Higgs boson decay channels							
$\sqrt{s}$ (GeV)	# events (million)			final state	kinematic cuts	hadrons	N_{pt}
	CEPC	FCC-*ee*	ILC				
125	0.23	0.09	0.07	gg	-	$\pi^\pm, K^\pm, p/\bar{p}$	77
	0.08	0.03	0.02	$c\bar{c}$			
	1.53	0.59	0.47	$b\bar{b}$			

A summary of the main features of the CEPC, FCC-*ee*, and ILC pseudo–data generated for the present study

Fit of FFs

Choosing a framework[1][2] for an analysis of fragmentation functions to identified charged hadrons, including charged pion, kaon and proton, using pseudo–data from future lepton colliders

parametrization of FFs to charged pion/kaon/ proton at initial scale (Q=5 GeV):

$$zD_i^h(z,Q_0) = z^{\alpha_i^h}(1-z)^{\beta_i^h}\exp\left(\sum_{n=0}^{m} a_{i,n}^h(\sqrt{z})^n\right)$$

parton-to-π^+	favored	α	β	a_0	a_1	a_2	d.o.f.
u	Y						5
$\bar{d} \simeq u$	Y	-	-		-	-	1
$\bar{u} = d$	N					x	4
$s = \bar{s} \simeq \bar{u}$	N	-				x	3
$c = \bar{c}$	N					x	4
$b = \bar{b}$	N					x	4
g	N		F				4

parton-to-K^+	favored	α	β	a_0	a_1	a_2	d.o.f.
u	Y					x	4
$\bar{s} \simeq u$	Y	-	-		-	x	1
$\bar{u}=d=\bar{d}=s$	N					x	4
$c = \bar{c}$	N					x	4
$b = \bar{b}$	N					x	4
g	N		F			x	3

parton-to-p	favored	α	β	a_0	a_1	a_2	d.o.f.
$u = 2d$	Y					x	4
$\bar{u} = \bar{d} = s = \bar{s}$	N				x	x	3
$c = \bar{c}$	N					x	4
$b = \bar{b}$	N					x	4
g	N		F			x	3

- a joint determination of FFs to charged pion, kaon and proton at NLO in QCD (63 parameters) including estimation of uncertainties with Hessian sets
- apply a strong selection criteria on the kinematics of fragmentation processes to ensure validity of leading-twist factorization and perturbative calculations ($z > 0.01$ and $E_h > 4$GeV)
- including theory uncertainties (residual scale variations) into the covariance matrix

[1] J. Gao et al., "Simultaneous Determination of Fragmentation Functions and Test on Momentum Sum Rule", Phys. Rev. Lett. **132**, 261903 (2024).

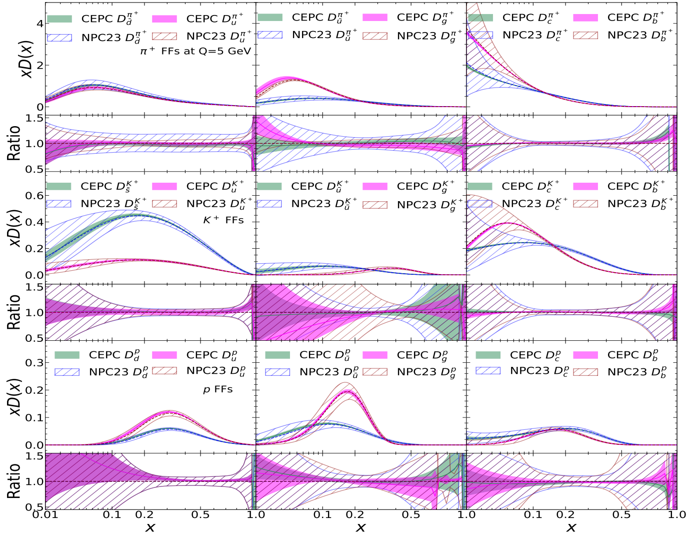
[2] J. Gao et al., "Global analysis of fragmentation functions to charged hadrons with high-precision data from the LHC", (2024).

Outline

The constraints to FFs from CEPC

We are only interested in the relative reduction of the FFs uncertainties once the data from future lepton colliders is included in the fit.

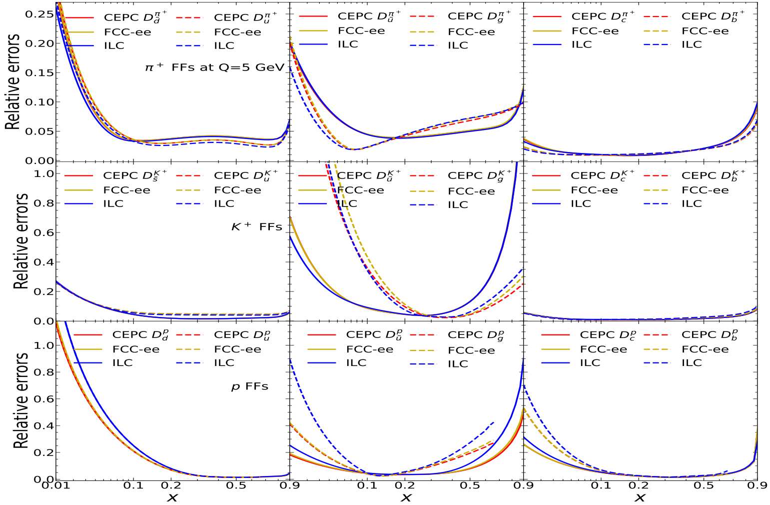
- the total χ^2 is 1659.2 for a total number of data points of 1831, resulting in $\chi^2/N_{\rm pt} = 0.91$
- we observe a marked reduction in the FFs uncertainties almost in all cases
- a slight increase of uncertainties for gluon FF to π^+ at small x due to the insufficient number of data points available after applying the kinematic selection in the small x region.



A comparison of the baseline NPC23 FFs with those based on CEPC pseudo–data for π^+, K^+, and p

The constraints to FFs from future lepton colliders

The pseudo–data from CEPC, FCC-*ee* and ILC have a comparable impact on the uncertainties associated with the FFs

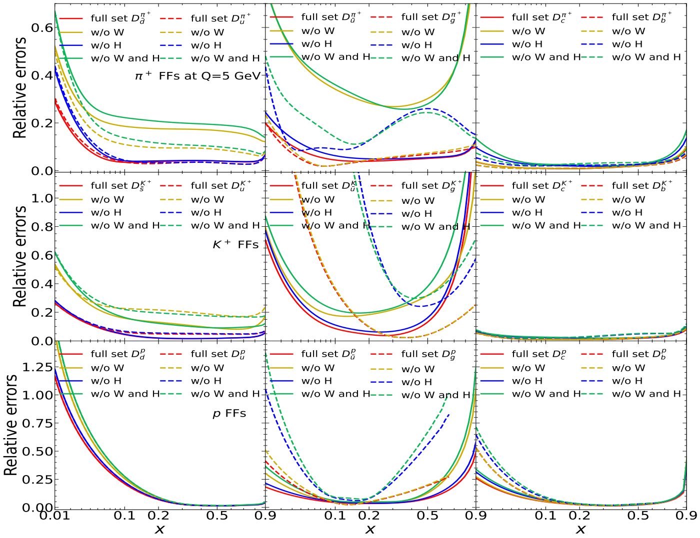


A comparison of the relative errors among the FFs obtained from CEPC, FCC-*ee*, and ILC measurements, respectively

Alternative fit

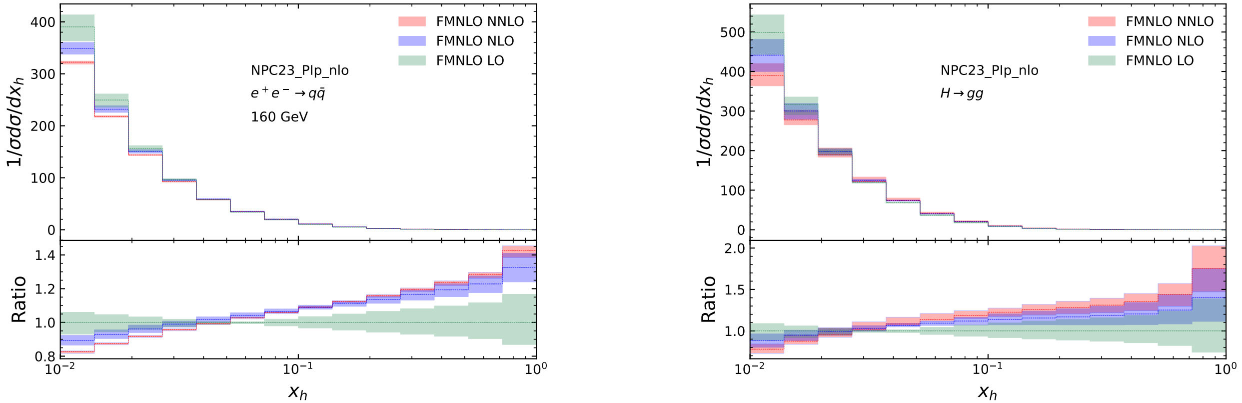
We utilize the FFs from CEPC as a case study to investigate the impact of individual data sets on the extracted FFs

- the $W^- W^{+,*}$ production data set has strong constraints on the FFs of light quarks
- the measurements from the decay of the Higgs boson exert a profound influence on the FFs of gluon
- the data sets on quark pair production at lepton colliders show a strong influence on the FFs of light quarks



FFs (positive charge) vs. momentum fraction

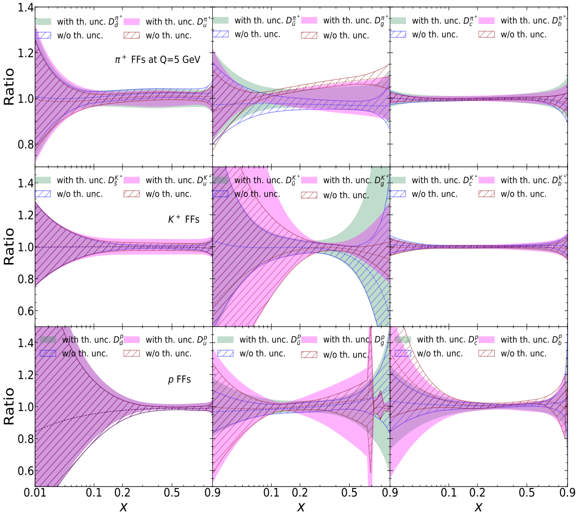
The effect of the higher-order QCD corrections

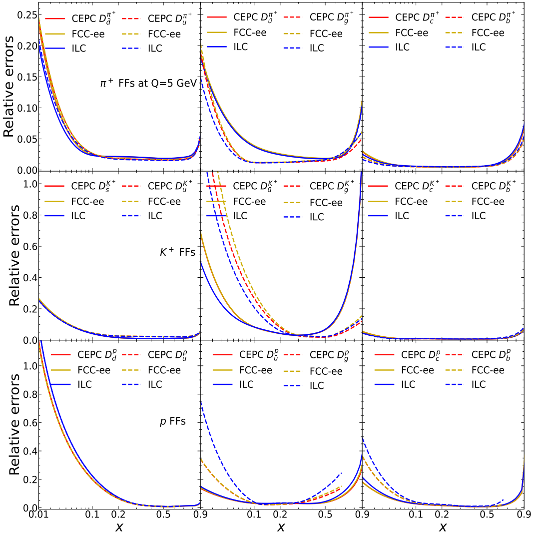


- The inclusion of NNLO corrections results in a further reduction in scale uncertainties.
- Furthermore, we also calculate the predictions using the NNLO set fragmenting into π^+ from NNFF1.0 as the input FFs and get similar results on the size of QCD corrections and scale uncertainties.

The effect of the higher-order QCD corrections

Alternative fits were conducted without including the theoretical uncertainties to discuss the effects of the higher-order corrections.

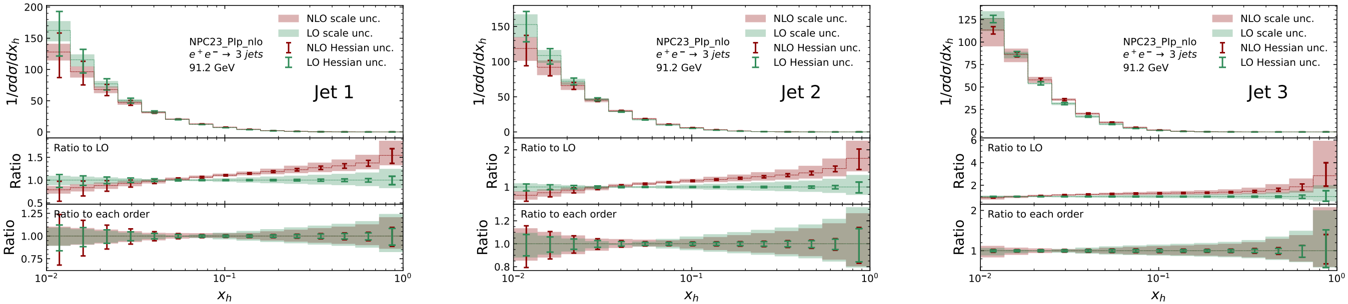




- A stronger constraint on the FFs when higher-order corrections are included in the fit

The constraints from three-jet production

The experiment is also capable of measuring the relevant observable in three-jet production.



A comparison between the scale uncertainties and the Hessian uncertainties for the distribution of the hadron energy fraction in each jet of three-jet events

- The NNLO corrections will be needed in the future to effectively include the data into a global analysis of FFs

Outline

Summary

- In this work, we study the constraining power of future lepton colliders on FFs to light charged hadrons from quarks and gluon
- We find that either CEPC, FCC-ee or ILC can significantly reduce the uncertainties of FFs in a wide kinematic range, compared to the NPC23 set
- We also discuss the impact of higher-order QCD corrections, and the potential constraints from measurements of three-jet production.
- Currently, hadron production cross sections at NNLO in QCD are available for SIDIS, SIA, and decays of the Higgs boson to gluons, and have been implemented into the FMNLO program

Thank you for listening!