Durham E-Theses

Event generation at NLO for photoproduction and diffraction

PETER MEINZINGER

How to cite:

MEINZINGER, PETER (2024) Event generation at NLO for photoproduction and diffraction. Doctoral thesis, Durham University.

Use policy

This work is licensed under a Creative Commons Attribution 3.0 (CC BY)

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Event generation at NLO for photoproduction and diffraction

Peter Richard Meinzinger

A Thesis presented for the degree of
Doctor of Philosophy

Institute for Particle Physics Phenomenology
Department of Physics
Durham University
United Kingdom

November 2024

Event generation at NLO for photoproduction and diffraction

Peter Richard Meinzinger

Submitted for the degree of Doctor of Philosophy
November 2024

Abstract: Precise predictions are key to the successful operation and data interpretation at collider experiments. While DIS is well understood and has reached very high accuracies, the precision goals of the upcoming EIC motivate revisiting lepton-hadron collisions. In this work, we implement photoproduction and diffraction which both make up significant fractions of the total cross-section at these colliders.

In the case of photoproduction, we achieve the first matched NLO simulation including beam remnants and multiple interaction modelling. Validating the calculation against data from the OPAL and ZEUS collaborations, we see satisfactory agreement. We then scrutinize the uncertainty associated with the photon PDFs by comparing eleven PDF sets and find that deviations between sets are of similar size as the Leading Order scale uncertainties. This leads to the conclusion that photon PDFs are the current bottleneck for precision photoproduction phenomenology for the EIC. We present the first fully-differential predictions for photoproduction at the EIC.

We furthermore conduct a comparative study of different implementations of photoproduction in general-purpose event generators and examine the differences between their predictions. There remain some open questions in the modelling, e.g. the transition to DIS, and future developments will benefit from past and future data from HERA and EIC, respectively.

Diffraction is described in the DIS regime by means of factorisation into the so-called Diffractive PDF and the matrix element. Implementing the corresponding terms, we again arrive at matched NLO simulations and validate our implementation against data from the H1 and ZEUS experiments and see excellent agreement. The underlying factorisation, however, breaks down in the photoproduction regime resulting in the calculation overshooting the data by approximately a factor of 2. We argue for a necessary suppression not only of the resolved but also the direct component and quantify these effects by a fit to data. We present predictions for diffractive DIS and diffractive photoproduction at the EIC where for the latter we apply the fitted suppression factors.

Contents

Abstract **3**

List of Figures **9**

List of Tables **17**

1 Introduction **25**

2 Background **27**

2.1 The EIC and HERA 27
2.2 Event classes in lepton-hadron colliders 31
  2.2.1 Deep-Inelastic Scattering 31
  2.2.2 Photoproduction 34
  2.2.3 Parton distributions for the photon 37
  2.2.4 Diffraction 40
2.3 Monte Carlo event generation 41
  2.3.1 Hard scattering 43
  2.3.2 Parton showers 43
  2.3.3 Hadronisation 44
  2.3.4 Multiple-parton interactions 45
  2.3.5 Beam remnants 46
2.4 Next-to-Leading Order and matching to the parton shower 47

3 Photoproduction 51

3.1 Equivalent Photons and their PDFs 51

3.1.1 Phase space handling 51

3.1.2 Equivalent Photon Approximation 53

3.1.3 Photon PDFs 54

3.1.4 NLO and matching to the parton shower 56

3.1.5 Multiple-parton interactions in photons and the beam remnants . 57

3.2 Validation of the framework 58

3.2.1 Comparison with LEP data 58

3.2.2 Comparison with HERA data 64

3.2.3 Photon PDF for precision phenomenology 71

3.3 Predictions for jet photoproduction at the EIC 71

3.4 Summary 79

4 Comparison of photoproduction in event generators 81

4.1 Event Generators 81

4.1.1 HERWIG 82

4.1.2 PYTHIA 83

4.1.3 SHERPA 84

4.1.4 Differences between generators 85

4.2 Comparisons to LEP 88

4.3 Comparisons to HERA 89

4.4 Predictions for EIC 93

4.5 Summary 96

5 Diffraction **99**

5.1 Theory framework for jet production in diffraction 99

5.2 Validation: simulation and data 104

5.2.1 Event generation 104

5.2.2 Experimental observables and datasets 105

5.3 Diffractive DIS 107

5.4 Diffractive photoproduction 111

5.4.1 Diffractive photoproduction at MC@NLO accuracy 111

5.4.2 Reggeon contribution 113

5.4.3 Modelling factorisation breaking in diffractive photoproduction . 113

5.5 Predictions for the EIC 119

5.5.1 Diffractive DIS 119

5.5.2 Diffractive photoproduction 124

5.6 Summary 124

6 Conclusion **129**

Bibliography **133**

List of Figures

2.1 Sketch of the layout of the planned EIC, based upon the existing RHIC at BNL. Taken from [8]. 28

2.2 Coverage in x and Q^2 in neutral current DIS of HERA and the EIC for different beam energies as used in an analysis of the EIC's data on PDF fitting. Taken from [12]. The variables will be discussed in Section 2.2.1. . . . 30

2.3 Sketch of the kinematics for inclusive neutral current DIS, taken from [14]. 31

2.4 Differential distributions of thrust τ_C for different $\langle Q\rangle \in \{15, 24, 58, 81\}$ GeV in deep-inelastic ep collisions as measured by H1 [20] compared to predictions by SHERPA with MEPS@NLO accuracy. 33

2.5 Differential distributions in virtuality Q^2 (top left), Bjorken variable x (top right), jet pseudorapidity η (bottom left), and jet transverse energy E_T (bottom right), comparing data from ZEUS [21] with N^{3}LO Fixed Order calculation by the NNLOJET collaboration [19], figures taken from the latter. 35

2.6 Sketch of an example Feynman diagram for a dijet event in DIS. . . . 36

2.7 Sketch of a $t\bar{t}H$ event in a hadron collider as simulated by an event generator. The dark red blob represents the hard interactions, the blue and red emissions the initial and final state radiation, respectively, and green blobs depict hadronisation and subsequent hadron decays. The purple blob depicts secondary scatters, and yellow lines soft photon radiation. Taken from [6]. 42

3.1 Schematic sketch of the phase space mappings between the different steps in the initial states, i.e. the Equivalent Photon Approximation (EPA), the Initial State Radiation (ISR), and the Matrix Element (ME). Each coordinate pair of Mandelstam-$s'^{(\prime)}$ and rapidity $y'^{(\prime)}$ is randomly sampled and the momenta are calculated as functions of these. Adapted from [90]. . . . 52

3.2 Spectrum of the photon flux at the Leading Logarithmic (LL) and Next to Leading Logarithmic (NLL) accuracy (left) and the ratio between the two (right). 54

3.3 Comparison of the gluon content of the photon for the SAS1M, SAS2M, SAL and CJKLLO PDF sets at $\mu_F^2 = 5\ \text{GeV}^2$. The vertical dotted black lines indicate the range of data, $x \in [0.001, 0.9]$, that was used for the latest fit in the SAL library. 55

3.4 Comparison of the light quarks and gluon contents for the SAS1M photon PDF and the PDF4LHC21_40_pdfas proton PDF at $\mu_F^2 = 5\ \text{GeV}^2$. 56

3.5 Differential dijet inclusive cross-section with respect to $\cos\theta^*$, defined in Eq. (3.2.4), comparing results of our SHERPA MC@NLO simulation at hadron-level with the LO simulation (top row) and with simulations at parton-level (bottom row) and with data from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. In the left and right panels the requirement $x_\gamma^\pm < 0.75$ and $x_\gamma^\pm > 0.75$ are applied and enhance resolved and direct contributions, respectively. . . 60

3.6 Distributions $x_\gamma^\pm$, collectively denoted as x_γ, in different bins of average transverse jet energy: $\bar{E}_T \in [5\,\text{GeV}, 7\,\text{GeV}]$ (top left), $\bar{E}_T \in [7\,\text{GeV}, 11\,\text{GeV}]$ (top right), $\bar{E}_T \in [11\,\text{GeV}, 25\,\text{GeV}]$ (bottom). Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. 61

3.7 Distributions of $|\Delta\eta|$ (top row), $|\eta_{\text{cntr}}|$ (middle row), and $|\eta_{\text{fwd}}|$ (bottom row) for requiring $x_\gamma < 0.75$ for exactly one (left column) or both (right column), comparing MC@NLO and LO. Results of the SHERPA simulation are compared with results from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. 62

3.8 Distributions of $|\Delta\eta|$ (top left), $|\eta_{\text{cntr}}|$ (top right), and $|\eta_{\text{fwd}}|$ (bottom) with the requirement $x_\gamma^\pm < 0.75$, comparing MC@NLO with and without MPI. Results of the SHERPA simulation are compared with results from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. 63

3.9 Distributions $\bar{E}_T$ for all values of $x_\gamma^\pm$ (top left), requiring x_γ^+ or $x_\gamma^- < 0.75$ (top right), and requiring $x_\gamma^\pm < 0.75$ (bottom). Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. 65

3.10 Differential dijet inclusive cross-section with respect to $\cos\theta^*$ for $x_\gamma^{\text{obs}} < 0.75$ (left) and $x_\gamma^{\text{obs}} > 0.75$ (right), comparing results of our SHERPA MC@NLO simulation with ZEUS Run 1 data [40]. 66

3.11 Differential dijet inclusive cross-section with respect to x_γ^{obs}, in different bins of leading jet transverse energy: $E_T \in [14\,\mathrm{GeV}, 17\,\mathrm{GeV}]$ (top left), $E_T \in [17\,\mathrm{GeV}, 25\,\mathrm{GeV}]$ (top right), $E_T \in [25\,\mathrm{GeV}, 35\,\mathrm{GeV}]$ (bottom left), $E_T \in [35\,\mathrm{GeV}, 90\,\mathrm{GeV}]$ (bottom right), comparing results of our SHERPA MC@NLO simulation with ZEUS Run 1 data [40]. 67

3.12 Differential single-jet inclusive photoproduction cross-section with respect to the pseudorapidity of the leading jet, $\frac{\mathrm{d}\sigma}{\mathrm{d}\eta}$, comparing results of our SHERPA simulation at parton (left) and hadron-level (right) with ZEUS Run 2 data [41]. 68

3.13 Distributions of the subleading jet pseudorapidity η^{jet2} with the requirement $x_\gamma^{\mathrm{obs}} < 0.75$ in three bins of leading jet pseudorapidity, $-1 < \eta^{\mathrm{jet1}} < 0$ (top left), $0 < \eta^{\mathrm{jet1}} < 1$ (top right) and $1 < \eta^{\mathrm{jet1}} < 2.4$ (bottom), comparing MC@NLO with and without MPI. Results of the SHERPA simulation are compared with results from ZEUS [40]. 69

3.14 Single-jet inclusive transverse energy spectra for the leading jet in different pseudorapidity bins of the k_T-jets: $0 < \eta < 1$ (top left), $1 < \eta < 1.5$ (top right), $1.5 < \eta < 2$ (bottom left), and $2 < \eta < 2.5$ (bottom right), comparing SHERPA MC@NLO results with ZEUS Run 2 data [41]. 70

3.15 Differential dijet inclusive cross-section with respect to the transverse momentum of the leading jet, with $0 < \eta^{(1)} < 1$ (left column) and $1 < \eta^{(1)} < 2.4$ (right column) and in different bins for $\eta^{(2)}$, comparing results of our SHERPA MC@NLO simulation at hadron-level including MPI effects with the LO simulation and with data [40] taken by ZEUS at HERA Run 1. . . 72

3.16 Distributions for $\cos\Theta^*$ at LEP (left) and for $\cos\theta^*$ at HERA (right), comparing SHERPA's LO simulations with data from OPAL [34] and ZEUS [40]. Scale uncertainties at LO are indicated by the pink band, while PDF uncertainties are shown with the blue hatched area. 73

3.17 Distributions of inclusive jet transverse energy E_T (top left), the x_γ (top right), inclusive jet pseudorapidity η (bottom left), and $\cos(\theta^*)$ (bottom right) of the SHERPA simulation with MC@NLO accuracy, compared with results at LO. 74

3.18 Event shapes observables: Transverse thrust major (upper left) and minor (upper right) distributions, C parameter (lower left) and charged particle multiplicity distribution N_{charged} inside the detector acceptance of $|\eta| < 4$ (lower right), all obtained with the SHERPA simulation at MC@NLO accuracy, compared with results at LO. 76

3.19 Distributions of $2 \to 1$ and $3 \to 2$ jet rates using the anti-k_T algorithm of the SHERPA simulation with MC@NLO accuracy, compared with results at LO. 77

3.20 Distributions of heavy quark jet transverse momentum p_T (left) and pseudorapidity η (right) of the SHERPA simulation with MC@NLO accuracy, compared with results at LO. 77

3.21 Distributions of x_γ of the SHERPA simulation at parton-level with LO accuracy and without MPI effects, averaged over 11 photon PDF sets. The PDF envelope shows the per-bin variation among the PDF sets. 78

4.1 Comparison of fixed-order parton level events between PYTHIA and SHERPA for a LEP-like setup. 86

4.2 Comparison of fixed-order parton level events including ISR parton shower and beam remnants between PYTHIA and SHERPA for an EIC-like setup. . 86

4.3 Comparison of the CJKLLO and SAS2M PDF sets for light quarks and the gluon at $\mu_{\text{F}}^2 = 5\ \text{GeV}^2$. 88

4.4 Distribution of average transverse energy of dijets, $\bar{E}_T$, from OPAL [34] for all $x_\gamma^\pm$ (top left), x_γ^+ or $x_\gamma^- < 0.75$ (top right) and $x_\gamma^\pm < 0.75$ (bottom), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 90

4.5 Distribution of pseudorapidity η of dijets from OPAL [34] for x_γ^+ or $x_\gamma^- < 0.75$ (left) and $x_\gamma^\pm < 0.75$ (right), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 91

4.6 Distribution of x_γ of dijets from OPAL [34] for bins of average jet transverse energy $\bar{E}_T \in [5\,\text{GeV}, 7\,\text{GeV}]$ (left) and $\bar{E}_T \in [11\,\text{GeV}, 25\,\text{GeV}]$ (right), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 91

4.7 Distributions of leading jet transverse energy E_T^{jet1} for $x_\gamma > 0.75$ (left) and $x_\gamma < 0.75$ (right) with jet pseudorapidities for the two leading jets in $1 < \eta^{\text{jet1}} < 2.4$ and $0 < \eta^{\text{jet2}} < 1$ from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 92

4.8 Distributions of sub-leading jet pseudorapidity η^{jet2} for $x_\gamma > 0.75$ (left) and $x_\gamma < 0.75$ (right) with leading jet pseudorapidity in $0 < \eta^{\text{jet1}} < 1$ from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 93

4.9 Distributions of x_γ for leading jet transverse energies in $14 < E_T^{\text{jet1}} < 17$ GeV (left) and $25 < E_T^{\text{jet1}} < 90$ GeV (right) from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 94

4.10 Predictions of transverse jet energy E_T (left) and x_γ (right) for jet production at the EIC, comparing to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. . . 95

4.11 Predictions of transverse thrust $T_\perp$ (left) and transverse sphericity $S_\perp$ (right), for jet production at the EIC, comparing to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 95

4.12 Predictions of charged particle multiplicity N_{charged} for events with a jet with $E_{\text{T}} < 6$ GeV at the EIC, comparing to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA. 96

5.1 Sketch of an diffractive DIS event for the production of two jets, j_1 and j_2, together with the remnants forming system X and p' forming the system Y. 100

5.2 Pomeron (left) and reggeon (right) fluxes as functions of $x_{I\!P}$ and t. . . 101

5.3 Integrated pomeron, $I\!P$, and reggeon, $I\!R$, fluxes with $t_{\text{cut}} = -1.6\,\text{GeV}^2$ as a function of $x_{I\!P,I\!R}$. 102

5.4 The H1 Fit B pomeron PDF and the GRVPI0 used for the reggeon at $\mu_F^2 = 10\ \text{GeV}^2$. 103

5.5 Convolution of the pomeron flux with the dominant gluon PDF from H1 Fit B at $\mu_F^2 = 10\ \text{GeV}^2$. 103

5.6 Differential DDIS cross-sections with respect to the photon-proton centre-of-mass energy W (top left), the photon virtuality Q^2 (top right), the leading jet transverse momentum $E_T^{*\text{jet1}}$ (bottom left), and average jet pseudorapidity $\langle\eta_{\text{jet}}^{\text{lab}}\rangle$ (bottom right) from [72]. 108

5.7 Differential DDIS cross-sections with respect to the average pseudorapidity $\langle\eta\rangle$ (top left) and the pseudorapidity difference $|\Delta\eta|$ (top right) of the jets as well as the virtuality Q^2 (bottom left) and the diffractive system's mass M_X (bottom right) from [136]. 109

5.8 Differential DDIS cross-sections with respect to momentum ratios $x_{I\!P}$ (top left), $z_{I\!P}$ (top right), the inelasticity y (bottom left), and the transverse momentum of the leading jet E_T^{jet1} (bottom right) from [136]. 110

5.9 Differential diffractive photoproduction cross-sections with respect to momentum ratios $x_{I\!P}$, $z_{I\!P}$, y and x_γ (top to bottom), obtained by H1 (left column, data from [136]) and ZEUS (right column, data from [137]). . . 112

5.10 Differential dijet Diffractive Photoproduction cross-sections with respect to leading jet transverse energy E_T^{jet1}, invariant mass M_X, difference in dijet pseudorapidity $|\Delta\eta|$ and average dijet pseudorapidity $\langle\eta\rangle$. Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from H1 [136]. 114

5.11 Distributions of leading jet transverse energy E_T^{jet1} (left column) and x_γ (right column) with and without reggeon contribution, compared to data by H1 from [72] (top) and [136] (middle) and from ZEUS [137] (bottom) in their respective definitions of fiducial phase space. 115

5.12 Differential distribution of x_γ in diffractive photoproduction. Results of the SHERPA simulation at parton- and hadron-level (p.l. and h.l. respectively) with MC@NLO accuracy are compared with results at LO and with data by H1 from [72] (top) and [136] (middle) and ZEUS [137] (bottom); with 7-point scale variations, and on linear scale in the left column and on logarithmic scale on the right column. 117

5.13 Differential diffractive photoproduction cross-sections with respect to momentum ratio x_γ as measured by H1 in [72] (top left) and [136] (top right) and x_γ^{obs} as measured by ZEUS [137] (bottom), compared to results of the SHERPA simulation at MC@NLO accuracy with and without a veto on γp interactions. 118

5.14 Differential diffractive photoproduction cross-sections with respect to momentum ratio x_γ as measured by H1 in [72] (top left) and [136] (top right) and x_γ^{obs} as measured by ZEUS [137] (bottom), compared to results of the SHERPA simulation at MC@NLO accuracy with the direct and resolved component scaled separately. 120

5.15 Predictions at LO and MC@NLO accuracy of SHERPA for diffractive DIS at the EIC for the leading jet transverse momentum $E_T^{(1)}$ (top left), average jet rapidity $\langle\eta\rangle$ (top right), diffractive mass M_X (bottom left), and dijet mass M_{jj} (bottom right). Phase space regions below the kinematic cuts are populated by the parton shower. 122

5.16 Predictions at LO and MC@NLO accuracy of SHERPA for diffractive DIS at the EIC for the momentum ratios $x_{I\!P}$, $z_{I\!P}$ (top row), inelasticity y, transverse thrust $T_\perp$ (middle row), thrust minor $T_{\perp,m}$ and transverse sphericity $S_\perp$ (bottom row). 123

5.17 Predictions at LO and MC@NLO accuracy of SHERPA of momentum ratio $x_{I\!P}$, $z_{I\!P}$ and x_γ (upper row, left to right), of inelasticity y in diffractive photoproduction at the EIC. 125

5.18 Predictions at LO and MC@NLO accuracy of SHERPA of event shapes transverse thrust $T_\perp$ and transverse thrust minor $T_{\perp m}$ (top row), transverse sphericity $S_\perp$ and diffractive mass M_X (middle row) and leading jet transverse momentum E_T^{jet1} and average jet pseudorapidity $\langle\eta\rangle$ (botto row) in diffractive photoproduction at the EIC. Phase space regions below the kinematic cuts are populated by the parton shower. 126

List of Tables

3.1 Photon PDF libraries included in SHERPA and their properties. 55

3.2 Inclusive cross-sections for one million events in a HERA Run 1 dijet photoproduction setup with two different proton PDFs and the same PDF for the photon. 73

4.1 Differences between the generators of default settings, specifically for photoproduction. 87

5.1 Scaling factors for the direct, R_{dir}, and resolved, R_{res}, component in diffractive photoproduction for the respective experimental data. 119

Declaration

The work in this thesis is based on research carried out at the Institute of Particle Physics Phenomenology, Department of Physics, at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification. This thesis is partly based on joint research, presented in the publications:

- Höche, S., Krauss, F. & Meinzinger, P.: Resolved photons in SHERPA. *Eur. Phys. J. C* **84**, 178 (2024). `https://doi.org/10.1140/epjc/s10052-024-12551-w`
- Meinzinger, P. & Krauss, F.: Hadron-level NLO predictions for QCD observables in photo-production at the Electron-Ion Collider. *Phys. Rev. D* **109**, 034037 (2024). `https://doi.org/10.1103/PhysRevD.109.034037`
- Helenius, I., Meinzinger, P., Plätzer, S. & Richardson, P.: Photoproduction in general-purpose event generators. arXiv:2406.08026.
- Krauss, F. & Meinzinger, P.: Hard Diffraction in Sherpa. *Eur. Phys. J. C* **84**, 894 (2024). `https://doi.org/10.1140/epjc/s10052-024-13295-3`

Copyright © 2024 Peter Richard Meinzinger.
The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof Frank Krauss for the many interesting discussions and insights into various phenomena in the field of High Energy Physics. The breadth of knowledge has been an inspiring experience, and I am very grateful for our discussions.

I would also like to thank the Institute of Particle Physics Phenomenology, which allowed me to pursue this degree and has fostered a lively environment, even if the start there in 2020 had been overshadowed by the Covid pandemic. Thanks to my colleagues, especially Lois Flower, Guillaume Rostagni, Hitham Hassan, Ery McPartland, and Malina Rosca for proofreading this thesis.

I'm also thankful to Joe Walker for the excellent atmosphere in our office, especially his appreciation for German proverbs; and complementary to working, the introduction to the Parkrun cult and the extensive conversations at pubs, be it in the US or the UK.

Last but not least, thank you to my family for their love and support throughout the years, especially Edith for being so open, supportive and thoughtful during all this time.

Not all those who wonder are lost.

Chapter 1

Introduction

The discovery of the quantum nature of elementary particles in the 20th century was one of the turning points in the history of physics. It sparked an enormous effort and great advances in the last century, with its tentative peak in the operation of the Large Hadron Collider (LHC), which is one of the largest scientific projects in human history.

This era of discovery in the field of High Energy Physics has allowed the establishment of the Standard Model (SM) of Particle Physics which successfully describes the data from the LHC and previous experiments to very high precision. However, even if the model is consistent with itself, there are underlying questions that remain unanswered. These include, among others, the baryon asymmetry and phenomena like dark matter, which would need to be embedded in some way in a complete theory.

Generally, this leads to the conclusion that the SM is only an effective theory, and that — just like Newtonian mechanics has been superseded by General Relativity — it will need to be extended into a greater theory.

There are two ways to look for this new theory, directly and indirectly. Ideally, physicists can find *direct* signals which cannot be described by the SM, e.g. the production of non-SM particles and observing these or their decay products. These SM extensions are commonly called Beyond the Standard Model (BSM) models and the advantage of this approach is the immediate confirmation of a specific model class; once the properties of a new particle are established, a clear picture of the type of extension of the SM would become clear. However, the plethora of potential signals to look out for combined with the limited energy reach of experiments do not allow for quick advances with this approach, and indeed no new signatures have been measured so far with the present machinery.

That is why a lot of effort has also been put into *indirect* searches, i.e., the precise calculation of predictions within the SM. While this would not immediately incorporate a new observation into a new theory, it exploits the enormous amounts of data which are seemingly in agreement with the SM, but which might exhibit small deviations due to higher order effects stemming from BSM couplings. It is at this point where the field of precision phenomenology comes into play. Collider experiments need simulations of their measurements and by pushing for ever higher accuracy in both the experiments and the predictions, the hope is to find hints of new theories in discrepancies between the data and the calculations.

While the LHC is still running and will be upgraded to the HL-LHC in a few years [1], there have been previous colliders like LEP and HERA, and in the next decade the electron-ion collider (EIC) will be built. Collisions at these experiments differ significantly from hadron-hadron colliders such as the LHC, and therefore need to be studied separately. In this work we aim at achieving better description of the event classes seen at these experiments, with a special focus on lepton-hadron colliders. To this end, we work in the framework of the SHERPA event generator [2, 3].

This work is organised as follows: in Chapter 2 we familiarise the reader with the collider experiments that will be most important in this context, and introduce different event classes and their characterisation. We will then provide a quick introduction into the field of Monte Carlo event generation, discussing how higher order corrections can be computed within the frameworks and how they are matched to the parton shower. Building upon that, in Chapter 3, we introduce the theory concepts relevant for photoproduction, validating their implementation in SHERPA against data from the OPAL and ZEUS experiments; we also discuss the precision bottleneck, the photon PDFs. We complete the chapter with predictions of jet photoproduction at the EIC. Photoproduction has, at Leading Order (LO), already been implemented in different event generators and in Chapter 4 we compare the photoproduction implementation among three general-purpose event generators, PYTHIA, HERWIG and SHERPA. Moving on to Chapter 5, the theoretical framework for diffractive events is introduced, validating the implementation against data from the H1 and ZEUS experiments. We discuss the factorisation breaking observed in diffractive photoproduction and complete the chapter by presenting predictions for the EIC. We summarise the findings of this work in Chapter 6.

Chapter 2

Background

This chapter will provide an introduction to the field of collider experiments and event generation as relevant for this work. We will assume that the reader is familiar with the Standard Model of Particle Physics and perturbation theory, and otherwise refer to the vast and extensive literature for an introduction into this field, e.g. [4–6].

2.1 The EIC and HERA

The Electron-Ion Collider (EIC) is a facility currently being commissioned in the USA [7,8]. It is a joint effort of the Nuclear Physics and High Energy Physics communities and will be built at the Brookhaven National Laboratory (BNL), in collaboration with Thomas Jefferson National Accelerator Facility (JLab). The operation will include different beam setups: one beam will use electrons, with energies ranging from 2.5 to 18 GeV, and the second beam will allow for different nuclei, ranging from protons to Uranium, with the proton reaching the highest energies of up to 275 GeV. The design aims to achieve very high luminosities, peaking at $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for the highest beam energies [8]. The facility builds up on the existing Relativistic Heavy Ion Collider (RHIC) and its schematic layout is shown in Fig. 2.1.

The scientific programme of the collider [8] aims at precision measurements of a variety of phenomena related to strong interactions and how a nuclear environment impacts them. In the Yellow Report [8], four different themes were identified.

First, nucleon properties will be examined, specifically the spin and the mass compositions. For the former, it has been found that the quarks and antiquarks only make up

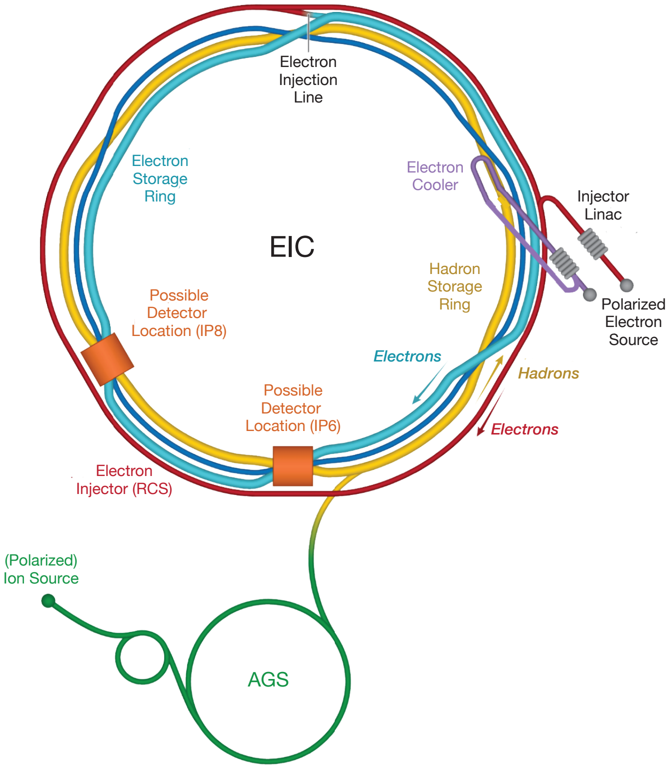

Electron
Injection
Line
Electron
Storage
Ring
Electron
Cooler
Injector
Linac
EIC
Hadron
Storage
Ring
Polarized
Electron
Source
Possible
Detector
Location (IP8)
Electrons
Hadrons
Possible
Detector
Location (IP6)
Electrons
Electron
Injector (RCS)
(Polarized)
Ion Source
AGS

Figure 2.1: Sketch of the layout of the planned EIC, based upon the existing RHIC at BNL. Taken from [8].

about 25% of the nucleon's total spin [9,10], and measurements of parton helicity distributions down to very small momentum fractions x will allow for an improved analysis of spin components. Similarly, the nucleon mass can only be partially explained by the quark masses and the EIC will investigate this problem with dedicated studies.

Another main topic is the multi-dimensional imaging of the nucleons, most importantly to examine the spatial distribution of partons in Generalized Parton Distributions (GPDs), and their transverse momentum in Transverse Momentum Dependent PDFs (TMDs). Going beyond traditional collinear PDFs, GPDs encode additionally the transverse spatial distribution of partons, while TMDs encode the longitudinal and transverse momentum distributions of partons. Key processes will be Deep-Virtual Compton Scattering (DVCS) and Semi-Inclusive DIS (SIDIS), respectively.

The third theme at the EIC will be the study of nuclei, as it is the first dedicated electron-nucleus collider. Open questions in this area are the effects of gluon saturation, nuclear PDFs and particle propagation through nuclear medium.

Last but not least, the EIC will focus on hadronisation. The transition from the perturbative regime to confined particles is one of the biggest open questions in QCD. Measurements of jet substructure and fits of fragmentation functions will provide a deeper understanding of the dynamics at the threshold of perturbation theory.

This ambitious programme rests, among others, on high-quality theory predictions, both in the form of analytic calculations and of Monte Carlo simulations, also known as event generators. While the former will typically provide higher accuracy for (semi-)exclusive measurements, the latter are still key by providing a sound baseline for the interpretation of global phenomena and inclusive measurements.

HERA is the only previous lepton-hadron collider, taking data from 1992 to 2007 [11]. Even though the LHC sparked a lot of progress in the measurement techniques and data interpretation since the end of the HERA operation, it is still instructive to study the latter. The asymmetric beam setup led to a few subtleties inherent to lepton-hadron colliders and HERA allows us to study these, as a benchmark for the EIC. For example, it has become customary to classify scattering events by the virtual mass squared, or virtuality, Q^2, of the photon exchanged between the incident electron and nucleon (or, in the case of the EIC, nucleus). Broadly speaking, essential parts of the HERA scientific programme, namely the determination of proton structure functions and PDFs, were nearly exclusively driven by the regime of non-zero Q^2 associated with deeply-inelastic scattering (DIS), also known as — somewhat out of fashion — electroproduction. The same will certainly be true for EIC-based efforts of improved PDF determination, too. For example, in [12] a study was conducted to examine the impact of the EIC on future PDF determination, with the kinematic reach of the data shown in Fig. 2.2.

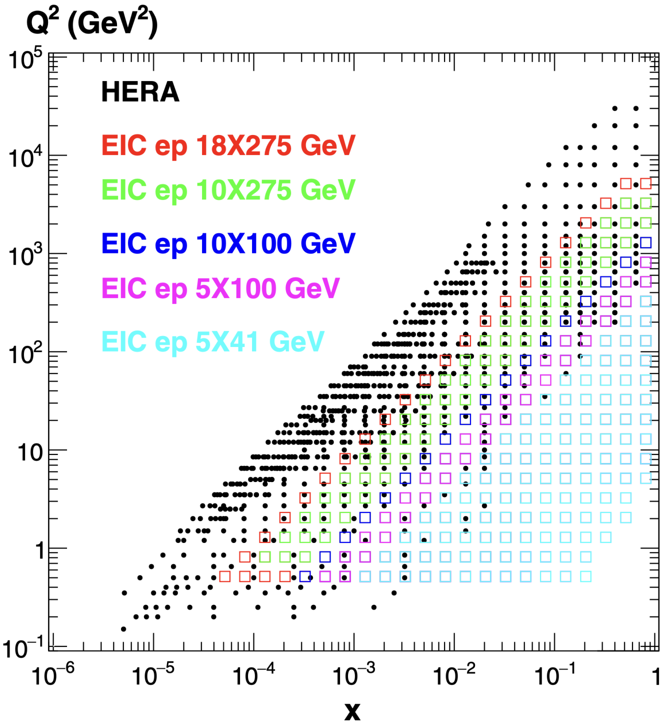


Figure 2.2: Coverage in x and Q^2 in neutral current DIS of HERA and the EIC for different beam energies as used in an analysis of the EIC's data on PDF fitting. Taken from [12]. The variables will be discussed in Section 2.2.1.

As the spectrum of the exchanged photons scales like $1/Q^4$, however, it is clear that the electron-proton cross-section is dominated by the photoproduction regime of relatively low $Q^2 \approx 0$. This necessitates the careful treatment of such photoproduction events, which so far have not attracted the same attention as, for example, event generation for the LHC, which has reached a very satisfying level of theoretical accuracy and maturity [13]. The same can be said about diffractive events, which at HERA amounted to a significant 10% of the total cross-section; we will review these three event classes more thoroughly in the next section.

2.2 Event classes in lepton-hadron colliders

Having two different beam particles, lepton-hadron colliders provide an interesting testing ground of Quantum Field Theory. Measurements at these type of colliders (in the limit of $Q^2 \gg 0$, see below) do not suffer from the same complications as measurements at hadron colliders, like Underlying Events, Pile-Up, etc., while still giving access to hadronic initial states. Therefore, to a large extent, the focus of their programmes has been the measurement of the hadronic structure, making use of the precisely measurable lepton and the inference of the underlying kinematics from it. This type of event is called DIS, and even if it has not been the focus of this work, an introduction would not be complete without it. These events provide important insights into hadronic structures, however, they constitute only a minority of the overall activity in these experiments. We will introduce DIS and the other two event categories, photoproduction and diffraction, in the following subsections.

2.2.1 Deep-Inelastic Scattering

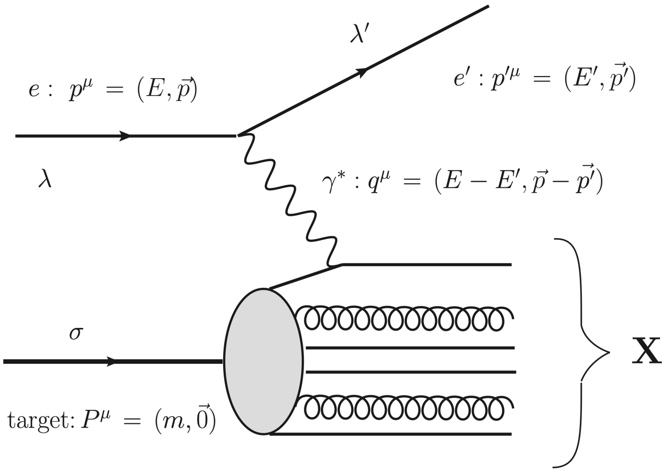

λ'
$e: \; p^\mu = (E, \vec{p})$
$e' : p'^\mu = (E', \vec{p}')$
λ
$\gamma^* : q^\mu = (E - E', \vec{p} - \vec{p}')$
σ
X
target: $P^\mu = (m, \vec{0})$

Figure 2.3: Sketch of the kinematics for inclusive neutral current DIS, taken from [14].

As mentioned above, DIS has seen great interest on both the theoretical and experimental sides. This is owed to the fact that the electron, in Neutral Current interactions, i.e. if the interactions is mediated through either a γ or Z boson, can be measured in the final state. Such an event is schematically depicted in Fig. 2.3. From the electron momentum alone one can already infer the kinematics for each event to a very high precision, where typically the inelasticity y, the virtuality Q and the Bjorken scaling variable x are used, defined as:

$$y = \frac{P \cdot q}{P \cdot p} \,, \quad Q^2 = -q^2 \,, \quad x = \frac{Q^2}{2P \cdot q} \,, \quad s = (P + p)^2 \tag{2.2.1}$$

The following equations hold true between these variables:

$$syx = Q^2 \quad \text{and} \quad M_X^2 = Q^2 \frac{1-x}{x} \tag{2.2.2}$$

with s the electron and proton beams' invariant mass squared and M_X the invariant mass of the system X, and therefore there are only two independent variables for a given collider setup which describe the events.

Calculations at Next-to-Leading Order (NLO) [15–17] for DIS have been available for some time now and many of the theory advances during the LHC era have already been applied to DIS calculations, namely Leading Order merging [18], and the current state-of-the-art being N^{3}LO calculations [19] in Fixed Order and MEPS@NLO in general-purpose event generation.

As an example of the currently available technology, we compare event shape distributions for different virtuality bins against MEPS@NLO predictions from SHERPA in Fig. 2.4. The uncertainties shown here and throughout the thesis are obtained by varying factorisation and renormalisation scales, μ_F and μ_R, independently up and down by factors of 2, omitting the two extreme variations. The thrust variable τ_C was defined as

$$\tau_C = 1 - \max_{\vec{n}_\mathrm{T}} \frac{\sum_i |\vec{p}_i \cdot \vec{n}_\mathrm{T}|}{\sum_i \vec{p}_i} \tag{2.2.3}$$

where the index i runs over all final state hadrons. H1 [20] used the measured data for a fit of the strong coupling constant α_S and similar studies will be possible at the EIC.

In terms of Fixed Order calculations, the highest accuracy to date is a N^{3}LO in [19], which is presented in Fig. 2.5. For most of the phase space, the calculation's accuracy exceeds the measurement's, however, this is not true in other parts of the phase space. The factorisation of the cross-section into PDFs and the partonic cross-section introduces the artificial factorisation scale μ_F into the equation. The boson exchange that needs to happen in a lepton-hadron scattering allows us to choose the virtuality of the exchanged boson as a generic scale. In certain parts of the phase space though, this is not the appropriate choice,

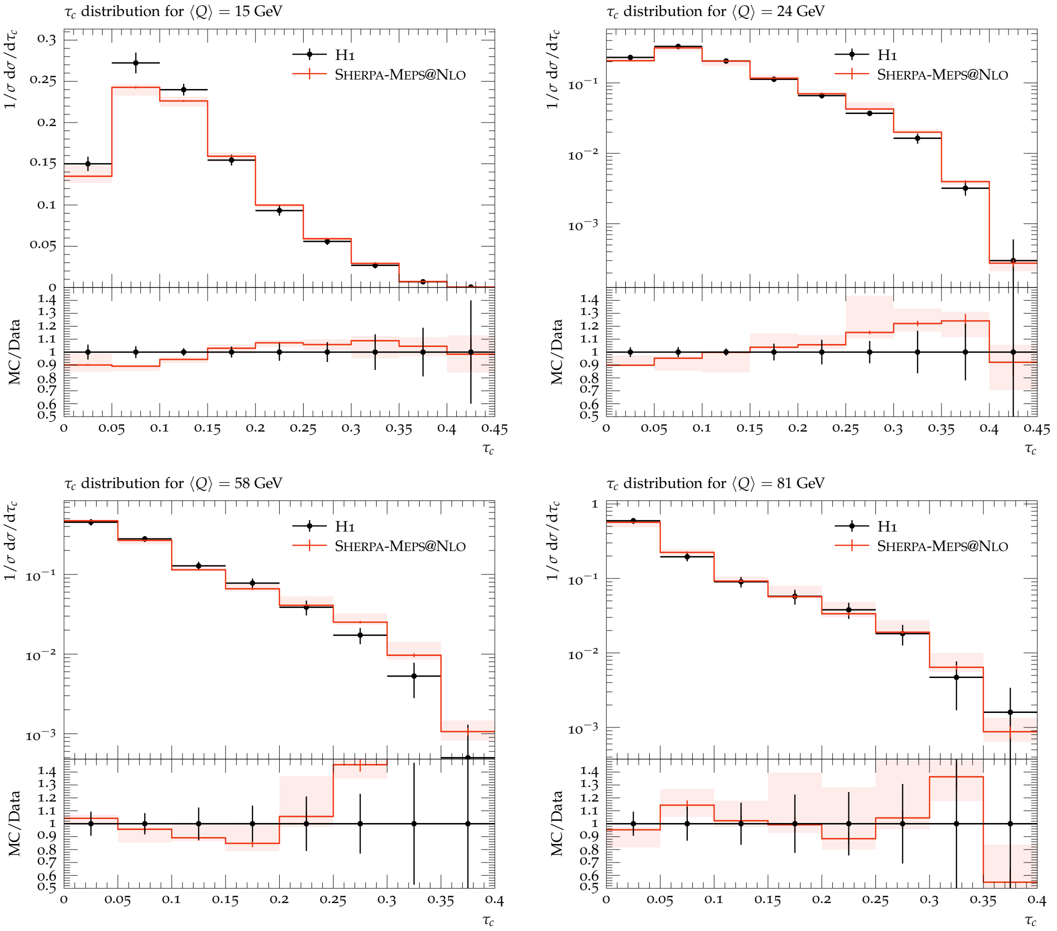


Figure 2.4: Differential distributions of thrust τ_C for different $\langle Q \rangle \in \{15, 24, 58, 81\}$ GeV in deep-inelastic *ep* collisions as measured by H1 [20] compared to predictions by SHERPA with MEPS@NLO accuracy.

and using an inadequate scale deteriorates the convergence of the perturbative expansion in the coupling. Looking at Fig. 2.5 we can identify these regions as the low-Q^2, low-x, low-E_T and forward region.

It is precisely this part of the phase space for which the virtuality is not the largest scale in the process. There are two ways to circumvent this problem to some extent. Firstly, one can choose the scale according to

$$\mu \propto \sqrt{Q^2 + H^2} \tag{2.2.4}$$

where H is a quantity calculated from the hadronic activity in the event, for example the leading jet transverse energy $E_T^{(1)}$ or the scalar sum of the transverse momenta of hadronic particles $H_{h,T}$. This approach works because part of the problem is that the problematic regions can be associated with contributions of the form given in Fig. 2.6. If the virtuality $Q^2 = -q^2$ is low, additional emissions can occur which stem from a propagator with $-k^2 > Q^2$. In that case, the hadronic scale H provides the better description of the underlying dynamics, and a scale choice as given in Eq. (2.2.4) interpolates between these two regimes.

This leads to another way to improve the calculation which is multi-leg merging, as proposed in [22] and applied to DIS in [18]. The parton shower in principle approximates the NLO real emission, but can only fill phase space below the factorisation scale. If the chosen scale is too low, the parton shower will not capture the correct correlation between the final states due to the breakdown of the underlying approximation. Splitting the phase space into a part filled by the parton shower and a complementary part filled by the matrix element, by means of a jet criterion, allows us to consistently include higher-multiplicity final states in the calculation. Additionally, employing a backwards clustering of the final state allows for unambiguously determining the highest scale in the process; for example, a trijet DIS process, $ej \to ejjj$, might be clustered to either $ej \to ej$, $\gamma^* j \to jj$ or $jj \to jj$, depending on which propagator has the hardest scale.

Ultimately though, at very low virtualities, splittings of the photon into partons have to take into account non-perturbative effects and this is where photoproduction comes into play.

2.2.2 Photoproduction

Photon-induced processes provide a rich testing ground for a wide range of physics effects. This is, on the one hand, because photons will couple to any electromagnetically

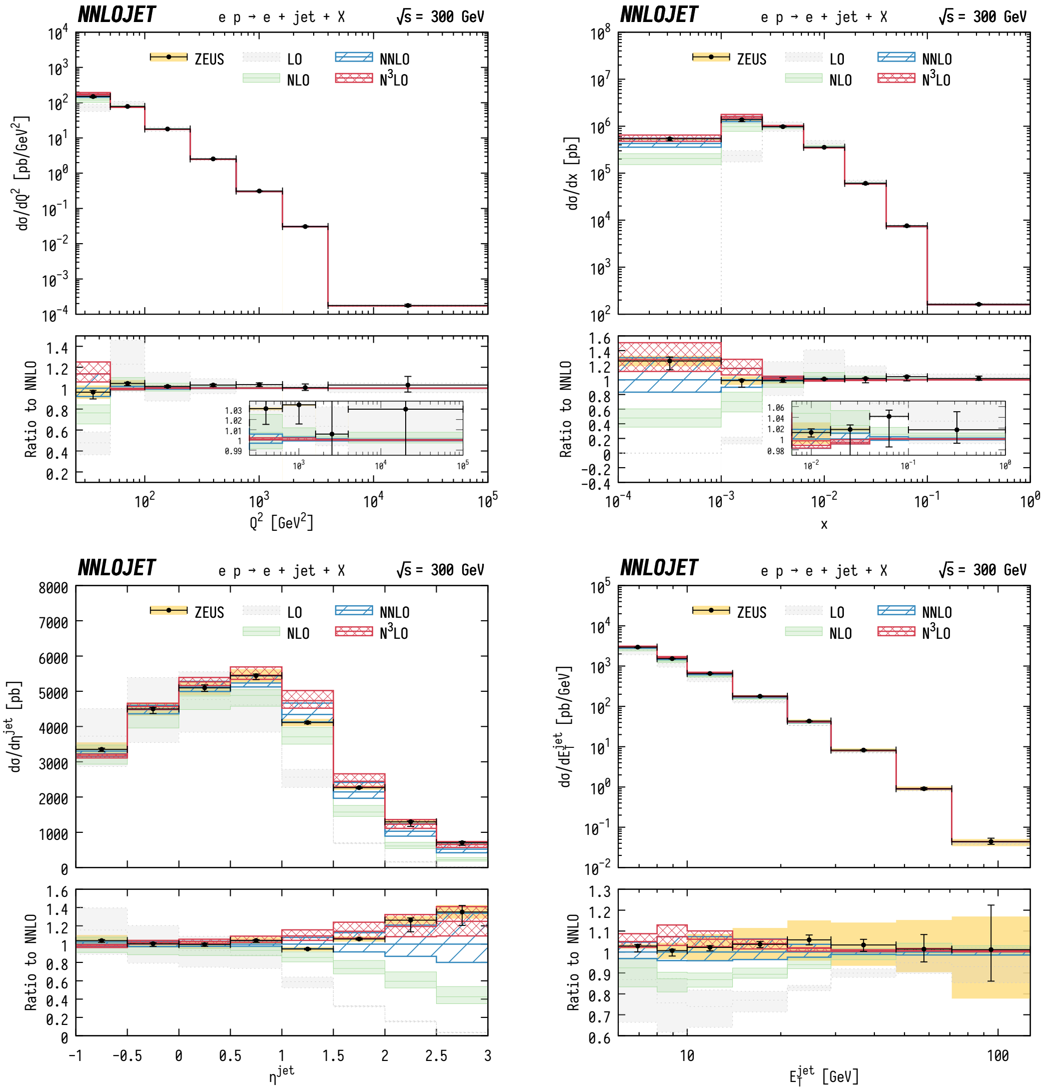


Figure 2.5: Differential distributions in virtuality Q^2 (top left), Bjorken variable x (top right), jet pseudorapidity η (bottom left), and jet transverse energy E_T (bottom right), comparing data from ZEUS [21] with N^{3}LO Fixed Order calculation by the NNLOJET collaboration [19], figures taken from the latter.

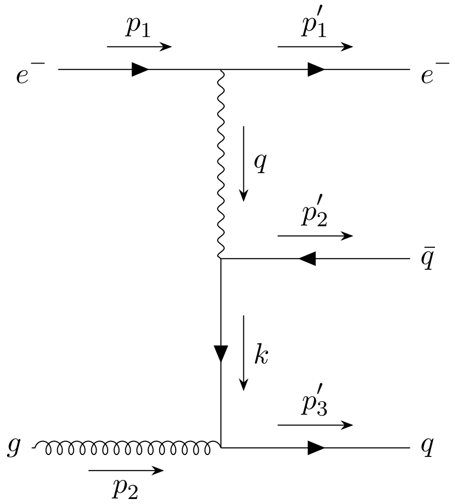


Figure 2.6: Sketch of an example Feynman diagram for a dijet event in DIS.

charged particle, resulting in a wide spectrum of accessible final states. On the other hand, photons have the quantum numbers of neutral vector mesons and their wave functions therefore have a sizeable hadronic component which lets them interact strongly, essentially fluctuating into states with quantum numbers of neutral vector mesons such as the ρ^0 or similar. At higher resolution scales, this non-perturbative component of the photon structure is further augmented by perturbative splittings of the photon into quark–anti-quark pairs — a feature that is not present in the more familiar proton PDFs. As a consequence, the corresponding γp collisions will not be characterised by a point-like photon interacting with partons, but rather look like hadron-hadron collisions involving the PDFs of photons, with parametrisations usually dating back two decades. This in itself suggests an interesting physics programme related to photoproduction, evidenced by the breadth and number of published analyses by the HERA collaborations, which cover a wide range of exclusive and inclusive final states [23–29].

The production of low-multiplicity final states in $\gamma\gamma$ collisions has been observed in many experiments [30–33], and yields interesting insights into the physics of hadrons and hadron resonances described by effective theories of the strong interactions. At increasing centre-of-mass energies of the colliding photons, new channels open up, and the production of jets has been studied, for example, at LEP [34–39]. Similarly, the photoproduction of various final states, including jets, has been analysed by the HERA experiments [23,28,40–44].

While the flux of quasi-real photons from relativistic charged particles has been understood for long time now, commonly known as the Equivalent Photon Approximation (EPA) [45–47], the results obtained in these experiments have allowed parametrising the parton content of the hadronic component of these photons in the form of parton distribution functions (PDFs), for example in the GRV [48], CJK [49, 50], SAL [51], and SaS [52, 53] sets. These PDF sets differ, for example, in the chosen parametrisation for the non-perturbative component, the inclusion of perturbative splittings and the starting scale.

In fact, based on these PDFs, satisfying agreement between data and calculations has been achieved, and a complete model for photon structure functions and high-energy photon interactions [54] has been encoded in PYTHIA [55]. There, the hard production of QCD final states at large scales, i.e. jets, is simulated in the usual way by dressing the hard parton-level matrix element with subsequent parton showers, the fragmentation of the resulting partons into hadrons during hadronisation, possibly including an underlying event. Recently, the model was extended to also include the perpendicular component of the photon momentum [56].

Pushing for higher accuracy, there have been a few predictions for inclusive jet-production at fixed-order at HERA [57–61] and the EIC [62], while recently a lot of attention has been paid to exclusive meson production processes and photoproduction at heavy-ion collisions (usually known as Ultra-Peripheral Collisions, or UPCs) [63–68].

Anticipation of increased precision requirements for successfully operating a possible future lepton collider such as FCC-ee or the planned electron-ion collider, EIC, motivates a revisit of the physics of photon-induced processes and to arrive at fully-differential predictions at NLO in QCD perturbation theory.

2.2.3 Parton distributions for the photon

We now turn to a more detailed discussion of the parton content inside photons. In this chapter and throughout this work we adopt the convention of writing $f_p^{(H)}$ for the distribution function for parton p in particle H.

When it comes to jet production at low photon virtualities one can distinguish three ways in which the photon can contribute to the scattering. For simplicity, we will focus on real photons, the generalisation to virtual photons has been attempted, but data for these cases is sparse. This section summarises the findings pioneered in [52, 54].

The most naive way would be for the photon to not undergo any splittings or interactions and interact in the hard process with the quarks; a second way would be for the photon to undergo a splitting into two charged fermions as part of the initial state evolution, i.e. at scales lower than probed by the hard process; and as a third and last way, by mixing with neutral vector mesons it could become hadron-like and therefore a purely non-perturbative QCD state. One can summarise the photon's wave function as

$$|\gamma\rangle = c_{\text{direct}}|\gamma_{\text{direct}}\rangle + \sum_V c_V|V\rangle + \sum_q c_q|q\bar{q}\rangle + \sum_l c_l|l\bar{l}\rangle \tag{2.2.5}$$

where V indicates appropriate vector mesons, and q and l quarks and leptons, respectively. For the purpose of jet production, the lepton contribution can be disregarded as their contribution will be negligible when compared to the other contributions. We will first turn to the vector meson and $q\bar{q}$ contributions. In line with the above equation the parton content of the photon can be approached with the ansatz

$$f_i^{(\gamma)}(x,\mu) = f_i^{(\gamma,\text{direct})}(x,\mu) + f_i^{(\gamma,\text{V})}(x,\mu;\mu_0) + f_i^{(\gamma,q\bar{q})}(x,\mu;\mu_0) \tag{2.2.6}$$

where the assumption has been made explicit that indeed the perturbative $|q\bar{q}\rangle$ states can be separated from the non-perturbative $|V\rangle$ states unambiguously by some scale μ_0. This scale also determines which vector mesons have to be taken into account in the sum in Eq. (2.2.5). We will call these two components the point-like and hadron-like component for $f_i^{(\gamma,q\bar{q})}$ and $f_i^{(\gamma,\text{V})}$, respectively. This ansatz extended the previous models, which only took into account the direct component and one or more vector meson contributions, the latter being commonly known as Vector Meson Dominance (VMD) models.

The first term in Eq. (2.2.6) is simply determined by

$$f_i^{(\gamma,\text{direct})}(x,\mu) = c_{\text{direct}}^2\delta_{i\gamma}\delta(1-x) \quad \text{with} \quad c_{\text{direct}}^2 = 1 - \sum c_V^2 - \sum c_q^2 - \sum c_l^2 \tag{2.2.7}$$

imposed by the unitarity of the physical state in Eq. (2.2.5). For practical purposes c_{direct} can be approximated by 1 [54].

The latter two terms in the decomposition constitute the contributions which typically are called resolved. For the rest of this work, we therefore redefine the photon PDF from Eq. (2.2.6) to the more common notion of

$$f_i^{(\gamma)}(x,\mu) = f_i^{(\gamma,\text{V})}(x,\mu;\mu_0) + f_i^{(\gamma,q\bar{q})}(x,\mu;\mu_0) \tag{2.2.8}$$

i.e. separating the direct interactions from the QCD-like ones. As mentioned before, the point-like component is perturbatively calculable with the boundary condition

$$f_i^{(\gamma,q\bar{q})}(x,\mu_0;\mu_0) = 0 \tag{2.2.9}$$

and the hadron-like component needs an ansatz which can be chosen as

$$f_i^{(\gamma,\mathrm{V})}(x,\mu;\mu_0) = \sum_V \frac{4\pi\alpha_{\mathrm{em}}}{f_V^2} f_i^{(V)}(x,\mu;\mu_0) \tag{2.2.10}$$

where α_{em} is the electromagnetic coupling. The vector mesons considered in the sum are usually ρ, ω, ϕ and J/ψ, however, it depends on the parameters used for the fitting. The coupling strengths f_V can be measured from elastic processes like $\gamma p \to Vp$. While physically meaningful, this approach does suffer from the fact that there are no measurements of the parton content of neutral vector mesons and therefore this ansatz still relies on many unknown parameters. One way around this problem would be to assume that the PDFs of these vector mesons can be approximated by the π^+ PDF; another ansatz is to assume a functional form and fit it to data.

At all scales, momentum must be conserved and hence the two components need to separately fulfil the usual momentum sum rule

$$\sum_i \int_0^1 \mathrm{d}x\, x\, f_i^{(\gamma,q\bar{q})}(x,\mu;\mu_0) = 1 \quad \text{and} \quad \sum_i \int_0^1 \mathrm{d}x\, x\, f_i^{(\gamma,\mathrm{V})}(x,\mu;\mu_0) = 1 \tag{2.2.11}$$

and additionally for the hadron-like component

$$\int_0^1 \mathrm{d}x\, 2\, v_q(x) = 2 \tag{2.2.12}$$

where $v_q(x)$ is the valence-like distribution and the equality is owed to the fact that the photon only mixes with mesons. Furthermore, in the resulting distributions quarks and antiquarks will always have the same PDFs as there is no term asymmetric with respect to the flavour in the evolution.

The resolved parton distributions obey the DGLAP equation which, due to the additional QED splitting, in this case reads

$$\frac{\partial f_i^{(\gamma)}(x,\mu)}{\partial \log\mu} = \frac{\alpha_{\mathrm{em}}(\mu)}{2\pi} P_{\gamma i}(x) + \frac{\alpha_S(\mu)}{2\pi} \sum_j \int \frac{\mathrm{d}x'}{x'} P_{ji}\left(\frac{x}{x'}\right) f_j^{(\gamma)}(x',\mu) \tag{2.2.13}$$

with P_{ab} being the DGLAP splitting kernels for $a \to bc$. This equation is an inhomogeneous differential equation, and it is instructive to view the separation into the vector meson and quark contribution, $f_i^{(\gamma,\mathrm{V})}$ and $f_i^{(\gamma,q\bar{q})}$, as the general and particular solutions to the equation, respectively.

This model has been extended to virtual photons by means of a dispersion relation [54], however, as we deal with real photons in our simulation we do not consider them here.

One major concern in this model is the dependence on the unphysical scale μ_0; in the final fit the distributions must not change with different choice of the cut-off between perturbative and non-perturbative regions. The authors of [52] used two different values of $\mu_0 = 0.6\,\mathrm{GeV}$ and $2\,\mathrm{GeV}$ of opposite extremes and arrived at significantly different distributions. A more diligent way would be to assume an interleaved evolution of the perturbative contributions with the mesonic resonances, and therefore a smooth transition between the two components instead of a hard cut-off at μ_0. However, this would pose additional mathematical hurdles in the fitting while still suffering from sparse data.

Given the kinematic ranges of the available data, especially the gluon content of the photon remains largely unconstrained and therefore suffers from large variations between different PDF fits. We will discuss this point and the phenomenology of photoproduction, including the quality of the fits and their uncertainties, in Chapters 3 and 4.

2.2.4 Diffraction

Diffractive events constitute a significant part of the cross-section at electron-hadron colliders such as HERA and the upcoming EIC, and at hadron-hadron colliders such as the currently operating LHC. They are characterised by the beam hadron scattering elastically or producing low-mass excitations of it and hence often show large gaps in the rapidity distribution of the overall activity. The production of jets in diffractive events offers additional insights into the dynamics of the strong interaction, more specifically its interplay between the hard perturbative scales of jet production and the soft non-perturbative scales of hadronic structures [69]. As before, the interaction of incident electrons and protons is dominantly mediated by a virtual photon; depending on its virtual mass $q^2 = -Q^2$ the interaction will either be classified as DIS (at large Q^2), or as photoproduction (at small Q^2). Just as explained in the previous section, the latter necessitates the inclusion of photon PDFs.

In general, the description of hard processes involving incident hadrons relies on the factorisation of the cross-section calculation into PDFs which encode the transition of the hadrons into the partons, i.e. quarks and gluons, and into the parton-level cross-sections. This factorisation picture has been proven to hold true for DIS [70], and has been extended for diffractive DIS [71], underpinning calculations for dijet production cross-sections at next-to [72] and next-to-next-to leading order (NNLO) in QCD [73]. Diffractive PDFs were extracted at NLO accuracy mainly from HERA data in [74]. It is customary to factorise them into the flux of an intermediate state and a PDF of this "particle"; these states are usually called pomerons and reggeons [75].

While successful for the description of diffractive DIS, it appears as if the factorisation assumption does not extrapolate to the case of diffractive photoproduction events. There are a number of underlying physics effects which may render factorisation invalid, including, for example, (i) additional scattering between the hadronic structure of the photon and the proton remnants, broadly speaking part of the unitarisation of the hard diffractive cross-section [76,77]; (ii) the impact of hadronisation effects on the emergence and survival of rapidity gaps necessary for the definition of diffractive events [78]. The apparent breakdown of factorisation in diffractive photoproduction also led to approaches where the different components of the cross-section, usually the resolved component where the photon assumes a hadronic structure, are rescaled to fit the data [79,80]. A similar rescaling approach was also taken by the authors of [81] who analysed differences between H1 and ZEUS data as well. Generally speaking, the resolved components were found to need a suppression by a factor of about 3, however, it remains unclear whether a rescaling of only the resolved component correctly reflects the process of factorisation breaking. We will turn to this problem and phenomenology of diffraction generally in Chapter 5.

2.3 Monte Carlo event generation

While collider experiments offer a plethora of data in the form of the detected particles, this information is not easily comparable to theoretical calculations. That is why the field of Monte Carlo event generation has seen extensive research and this brief introduction will follow the lines of [6,82].

Depending on the measurement and its observables, the comparison of theory and experimental data is complicated by various factors. First of all, the calculation of matrix elements becomes virtually impossible to do beyond the first few orders of perturbation theory in the strong coupling α_S. The number of amplitudes that has to be considered grows factorially with the order, immensely increasing the computational cost for their calculation.

Another major problem is the confinement of QCD. For hadrons in the initial state, it can be overcome by factorisation, i.e. the fitting of PDFs and their convolution with the matrix elements; in the final state, so-called fragmentation functions can be employed for (semi-)exclusive measurements of identified hadrons, however, for inclusive measurements it has not been solved from first principles. A third problem is the number of dimensions that has to be integrated over in the phase space. For n particles, the number of dimensions is $3n - 4$ and as this quickly becomes unfeasible for analytical techniques and even for numerical approaches like the trapezium rule or Simpson rule.

General purpose event generators aim to overcome all of these problems. A typical example of an event as modelled by an event generator is sketched in Fig. 2.7, and we will go through the main components in the following paragraphs.

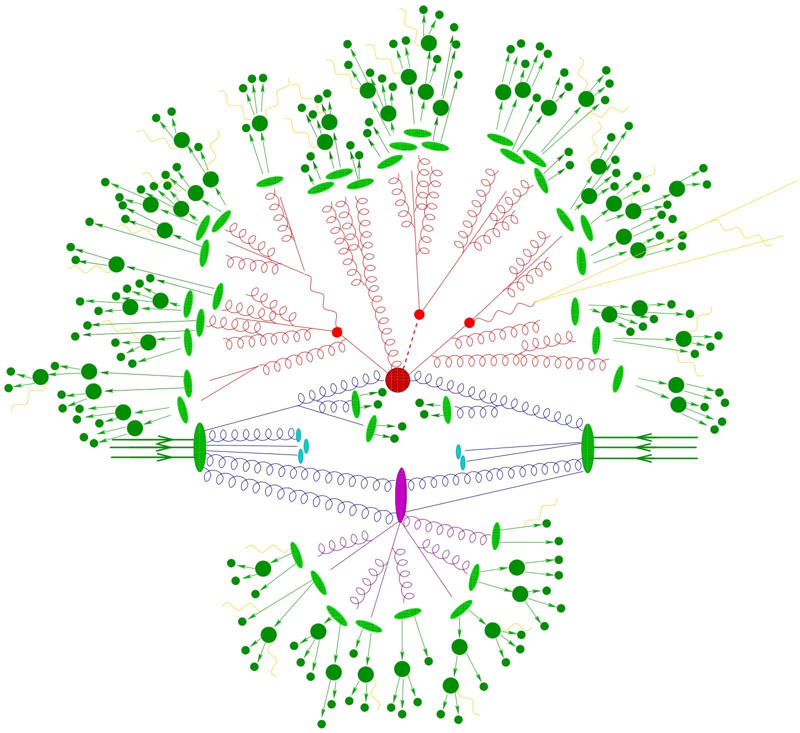


Figure 2.7: Sketch of a $t\bar{t}H$ event in a hadron collider as simulated by an event generator. The dark red blob represents the hard interactions, the blue and red emissions the initial and final state radiation, respectively, and green blobs depict hadronisation and subsequent hadron decays. The purple blob depicts secondary scatters, and yellow lines soft photon radiation. Taken from [6].

2.3.1 Hard scattering

Typically, the first and foremost step in the simulation is that of the signal, the hard scattering process. Taking a process at the LHC as an example, the factorisation reads

$$\sigma\left(pp \to X\right) = \sum_{a,b} \int \mathrm{d}x_1 f_a^{(p)}\left(x_1, \mu_F^2\right) \int \mathrm{d}x_2 f_b^{(p)}\left(x_2, \mu_F^2\right) \hat{\sigma}\left(ab \to X\right) \tag{2.3.1}$$

and the partonic cross-section $\hat{\sigma}$ is computed by

$$\hat{\sigma}\left(ab \to X\right) = \int_{\Phi(\mathcal{O})} \left[\prod_{i=1}^{N} \frac{\mathrm{d}^3 q_i}{(2\pi)^3 2E_i}\right] \delta^4\left(p_1 + p_2 - \sum_i q_i\right) \left|\mathcal{M}_{ab\to\{i\}}(p_1, p_2; \{q_i\})\right|^2 . \tag{2.3.2}$$

The computation can be split into two parts: the calculation of the phase space point, including the evaluation of the PDFs and the multidimensional integration over the momenta; and the computation of the matrix element squared, $|\mathcal{M}|^2$. The latter can get prohibitively complicated at two fronts, the inclusion of more legs and the computation of virtual corrections and the corresponding loops. The latter requires extensive algebraic efforts to evaluate the appearing integrals and reduce their complexity, and typically for fully-differential event generation, only one-loop amplitudes are considered through dedicated libraries [83–86].

Regarding additional legs, the computational complexity grows factorially with the number of legs. Different approaches have been used to compute these in matrix-element generators, for example COMIX [87], which relies on Berends-Giele recursion [88] to compute the tree-level amplitudes. With this approach there is no inherent limit on the number of legs, however, practically it is restricted by the increasing computational cost.

2.3.2 Parton showers

After yielding a parton-level matrix element, the parton shower is one of two essential steps to arrive at hadron-level predictions. They are the subject of extensive research, but for the purposes of this work we want to only give a brief summary of their properties. Their task is to evolve the hard process, happening at very high scales μ, e.g. at $\mu = m_Z = 91\ \mathrm{GeV}$, down to the boundaries of the perturbative expansion and where hadronisation then will provide a better description, $\mu = \mathcal{O}(\Lambda_{\mathrm{QCD}})$.

To derive the correct evolution expressions, one assumes unitarity of the parton branching and a Poissonian distribution of these; the latter has only one free parameter, the probability, which in case of a parton branching will not be constant but depend on kinematic variables. Even though there are a few more subtleties, the elegance of parton showers lies in the fact that they describe a wealth of data with only a few free parameters.

There are three main ingredients to describe a parton shower implementation:

First, the emission kernels K have to be determined which must be derived from the single emission matrix element, i.e. an approximation of the real correction to the Born matrix element at leading colour, and the kernels must reproduce the correct soft and collinear limits. For the latter, they will just reproduce the DGLAP splitting kernel, while for the former some intricacies arise due to interferences from sub-leading colour configurations.

Second, an evolution variable t has to be chosen. There are different possibilities like the invariant mass m_{ij}, the angle θ_{ij} or the relative transverse momentum $k_{\perp,ij}$ of the produced parton pair. This choice may influence the logarithmic accuracy and the inclusion of quantum coherence effects. This variable parametrises the evolution down to a given cut-off scale t_c which is free to choose and typically of the order of $1\ \mathrm{GeV}$.

The third and last component are the kinematics of individual splittings. The incoming parton is always on-shell, and after each step a physically meaningful configuration must be obtained, i.e. the produced pair must be put on-shell as well. This necessitates the inclusion of a so-called spectator in the splitting which is used to "borrow" momentum in order to maintain overall momentum conservation.

The emission kernel can then be used to derive a no-emission probability, the so-called Sudakov form factor Δ which we will introduce in a more technical way in Section 2.4. For the case of initial state evolution, the parton flux factor, given by a ratio of PDFs, must also be included in the Sudakov form factor.

2.3.3 Hadronisation

After the parton-showering, the second step from parton-level fixed order computations towards hadron-level event descriptions is the hadronisation. It takes over once the evolution down to the cut-off scale in the parton shower has terminated and translates the individual partons into hadrons. The exact process has not been understood from first principles yet, and since it also violates, e.g., colour conservation one has to resort to empirical models with many free parameters and fit them to data.

Generally, the different models can be subdivided into two approaches, the Lund string model and cluster fragmentation models.

The Lund string model is based on the picture of a quark–anti-quark pair being bound by a colour-less "string", while gluons are kinks on these strings. Depending on the energy in this system, they may either break up creating new colour-less quark–anti-quark pairs or they fall into a stable "yo-yo" motion, which is then interpreted as a meson.

Cluster fragmentation is based on the principle of pre-confinement, which observes that at each step the parton shower creates locally colour neutral parton pairs and that asymptotically these have universal properties, depending only on the shower cut-off scale t_c and on Λ_{QCD}. These systems form so-called primary clusters and "decay" non-perturbatively into smaller clusters or directly into hadrons.

Typically, these models are fitted to data from LEP, the only high-energy e^+e^- collider to date, due to the absence of other non-perturbative effects like beam remnants and multiple interactions.

2.3.4 Multiple-parton interactions

Even though not strictly necessary for hadron-level calculations, multiple-parton interactions (MPI) play a non-negligible role in the description of hadronic collisions.

If one applies the perturbative description to $2 \to 2$ scatterings in QCD, the partonic cross-section will approximately scale like $\mathrm{d}p_T^2/p_T^4$ with p_T the transverse momentum of the final state partons. This behaviour is regulated by phase space cuts, in this case some minimal transverse momentum $p_{T,\min}$. If this cut, however, is lowered to small values of $\mathcal{O}(1 - 10\,\mathrm{GeV})$, the integral of the partonic cross-section will exceed the total hadron-hadron cross-section σ_{hh}, i.e.

$$\int_{p_{T,\min}^2}^{\mu^2} \mathrm{d}p_T^2 \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_T^2} > \sigma_{hh} \,. \tag{2.3.3}$$

This observation is then interpreted as indication of multiple parton interactions for the given collisions, and after accounting for elastic and diffractive contributions to the total hadronic cross-section, one can calculate the expected number of partonic interactions N by

$$\langle N \rangle = \frac{1}{\sigma_{hh,\mathrm{ND}}} \int_{p_{T,\min}^2}^{\mu^2} \mathrm{d}p_T^2 \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_T^2} \tag{2.3.4}$$

where $\sigma_{hh,\mathrm{ND}}$ is the non-diffractive hadron-hadron cross-section. This quantity can be calculated from Regge theory, while the partonic $2 \to 2$ cross-section is known analytically. Knowing the expectation value $\langle N \rangle$, one can assume Poissonian statistics for the partonic interactions and therefore distribute individual scatters by their probability P_{MPI} using

$$P_{\mathrm{MPI}}(p_T) = \exp\left(-\frac{1}{\sigma_{hh,\mathrm{ND}}}\int_{p_T^2}^{\mu^2} \mathrm{d}\tilde{p}_T^2 \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\tilde{p}_T^2}\right) . \tag{2.3.5}$$

This picture can be refined by the inclusion of an impact parameter b, i.e. the spatial transverse distance between the colliding hadrons, which usually leads to a larger variation in the distributions of the partonic activities.

To finish this discussion, we point out that multiple interactions inherently go beyond the factorisation theorem and must therefore be considered as non-perturbative effects. The models will depend on several free parameters and must be tuned to data. As the scales involved in these interactions are rather small, the corresponding Bjorken-x extend to very small values, too, making the whole procedure highly dependent on the PDF fit which is used.

2.3.5 Beam remnants

As part of a complete event generation, the conservation of all quantum numbers has to be ensured and furthermore, the valence content of the beam particles has to be respected. We will outline the procedure for the proton in the following.

Given a list of partons extracted from the beam in the hard interaction and, potentially, the MPI, flavour has to be compensated in two steps. If the list contains a valence quark, e.g. the u quark, a corresponding diquark, e.g. a ud diquark, is created to arrive at the correct valence content. If no such quark is present, both the quark and the diquark have to be created. After this, so-called spectators have to be created which compensate the flavour of any remaining (anti-)quarks. Once a full set of flavour-neutral sea quarks and the correct valence content has been established, colour has to be compensated by creating additional gluons.

After this, the momentum has to be distributed and the distribution must fulfil two constraints: momentum must be conserved, i.e. the sum of parton momenta must add up to the beam momentum p_{beam}; and each (anti-)quark must have enough energy to be put on-shell later in the hadronisation, while for gluons we assume some minimal energy as well, of the order of Λ_{QCD}. For a given beam momentum and the momenta of extracted partons $\{p_i\}$, the remaining momentum is given by $p_{\mathrm{rem}} = p_{\mathrm{beam}} - \sum_i p_i$. If the total mass of the spectators is larger than the remaining energy, $\sum_s m_s > p^0_{\mathrm{rem}}$, then the on-shell condition cannot be accommodated, and the event has to be rejected. Otherwise, the momentum is distributed according to a power law, which approximates the small-x behaviour of the PDF, again ensuring that each parton can be put on-shell and hadronised later.

Additionally to the longitudinal momentum, transverse momentum has to be generated for the remnants, the so-called primordial k_T. This is randomly sampled from a Gaussian and again must conserve momentum, i.e. the total transverse momentum has to add up to $\vec{0}$. The mean and the width of the k_T distribution, as well as some other parameters that can be introduced, are subject to tuning, usually to DIS data.

While the effects of the beam remnants are not generally visible in LHC analyses, they can be detected in ep collisions. The modelling of these can lead to large effects in the photoproduction and diffraction regime, where the particles typically have very small energies. There, the phase space rejection during the remnant creation does lead to sizeable effects, as will be shown in Section 4.1.4.

2.4 Next-to-Leading Order and matching to the parton shower

A parton shower encodes the bremsstrahlung that scattered high energy particles exhibit by means of approximative matrix elements, as sketched out previously. These approximations are computed by expanding the emission of an additional particle in the soft-collinear limit. By then allowing for multiple such emissions the parton shower resums logarithms that appear in certain observables.

Additional emissions can also be accounted for by computing higher orders of the matrix element, they are then captured in the real corrections. These emissions will correctly describe correlations between highly energetic particles at large angles, i.e. in phase space complementary to the parton shower emissions, capturing potential interferences between different amplitudes. Obviously, higher order computations also take into account corrections to the total cross-section, unlike the parton shower which by definition leaves the total cross-section unchanged. Both procedures work by allowing an additional parton emission off a given process, and therefore a double-counting would be introduced when combining these two approaches in a naive way.

At LO, the cross-section is given by the Born contribution B, $\sigma^{(\mathrm{LO})} = \int \mathrm{d}\Phi_B B(\Phi_B)$, and the parton shower generates an emission according to

$$\sigma^{(\mathrm{LO+PS})} = \int \mathrm{d}\Phi_B B\left(\Phi_B\right) \left[\Delta(t_0, \mu_F^2) + \int_{t_0}^{\mu_F^2} \mathrm{d}\Phi_1 K\left(\Phi_1\right) \Delta(t, \mu_F^2)\right] \tag{2.4.1}$$

where the Sudakov form factor Δ has been introduced, defined as

$$\Delta\left(t, t'\right) = \exp\left[-\int_t^{t'} \mathrm{d}\Phi_1 K\left(\Phi_1\right)\right] \tag{2.4.2}$$

and where $K\left(\Phi_1\right)$ is the shower emission kernel. Here and throughout we ignore other dependencies like flavour for clarity.

At NLO though, infrared divergences in the real and the virtual correction have to be cancelled, and the cross-section is therefore given by

$$\sigma^{(\mathrm{NLO})} = \int \mathrm{d}\Phi_B \left[B\left(\Phi_B\right) + \tilde{V}\left(\Phi_B\right) + I^{(S)}\left(\Phi_B\right)\right] + \int \mathrm{d}\Phi_R \left[R\left(\Phi_R\right) - D^{(S)}\left(\Phi_R\right)\right] \tag{2.4.3}$$

where $B\left(\Phi_B\right)$, $\tilde{V}\left(\Phi_B\right)$ and $R\left(\Phi_R\right)$ are the Born, virtual and real matrix elements, respectively; and $D^{(S)}\left(\Phi_R\right)$ and $I^{(S)}\left(\Phi_B\right)$ are the subtraction terms and their integrated terms. For the latter, a subtraction technique needs to define an appropriate single-emission phase space Φ_1 to allow for the integration of the subtraction terms, i.e. it must fulfil $\Phi_R = \Phi_B \otimes \Phi_1$.

The difficulty now is to appropriately redefine the quantities in Eq. (2.4.1) such that the total cross-section is not changed, and the parton shower emission respects the kinematic pattern as given by the real correction R.

We start by splitting the real emission into two parts, a regular (hard) and singular (soft) part, H and S. This allows us to redefine Eq. (2.4.3) as

$$\begin{aligned}\sigma^{(\mathrm{NLO})} = &\int \mathrm{d}\Phi_B \left[B\left(\Phi_B\right) + \tilde{V}\left(\Phi_B\right) + I^{(S)}\left(\Phi_B\right)\right] + \\ &+ \int \mathrm{d}\Phi_R \left[S\left(\Phi_R\right) - D^{(S)}\left(\Phi_R\right)\right] + \int \mathrm{d}\Phi_R H\left(\Phi_R\right) \ .\end{aligned} \tag{2.4.4}$$

As the last term is completely regular it can be treated as a $N+1$ leading order matrix element. So far we are NLO accurate but have not yet included a parton shower emission. Because S captures the singular phase space, it is defined in the correct limits to allow a redefinition of the shower emission kernel as

$$K\left(\Phi_1\right) \rightarrow \frac{S\left(\Phi_B, \Phi_1\right)}{B\left(\Phi_B\right)} \tag{2.4.5}$$

and correspondingly the Sudakov form factor as

$$\Delta\left(t, t'\right) \rightarrow \bar{\Delta}\left(t, t'\right) = \exp\left[-\int_t^{t'} \mathrm{d}\Phi_1 \frac{S\left(\Phi_B, \Phi_1\right)}{B\left(\Phi_B\right)}\right] . \tag{2.4.6}$$

The separation of the regular and singular parts of the real correction is somewhat arbitrary, but it has been found to be particularly useful to identify the singular part with the subtraction term, i.e. to equate $S\left(\Phi_R\right) = D^{(S)}\left(\Phi_R\right) \Theta\left(\mu_F^2 - t\right)$. This allows the redefinition of the cross-section:

$$B\left(\Phi_B\right) \rightarrow \bar{B}\left(\Phi_B\right) = B\left(\Phi_B\right) + \tilde{V}\left(\Phi_B\right) + I^{(S)}\left(\Phi_B\right) + \int \mathrm{d}\Phi_1 \left[S\left(\Phi_B, \Phi_1\right) - D^{(S)}\left(\Phi_B, \Phi_1\right)\right] . \tag{2.4.7}$$

Taking together all these substitutions, the matching can be achieved according to

$$\sigma_{\mathrm{MC@NLO}}^{(\mathrm{NLO+PS})} = \int \mathrm{d}\Phi_B \bar{B}\left(\Phi_B\right) \left[\bar{\Delta}(t_0, \mu_F^2) + \int_{t_0}^{\mu_F^2} \mathrm{d}\Phi_1 \frac{S\left(\Phi_B, \Phi_1\right)}{B\left(\Phi_B\right)} \bar{\Delta}(t, \mu_F^2)\right] + \int \mathrm{d}\Phi_R H\left(\Phi_R\right) . \tag{2.4.8}$$

One can see this method as using the shower emission kernel for the subtraction, or, conversely, the subtraction terms for the shower evolution. The technique introduced here is commonly known as the MC@NLO matching, and it is used throughout this work; other methods exist with POWHEG [89] being the most common alternative to MC@NLO.

Chapter 3

Photoproduction

After a general introduction to photoproduction in Section 2.2.2 we will in this chapter turn to a more technical discussion of the calculation and the phenomenology at different colliders.

3.1 Equivalent Photons and their PDFs

For the simulation of photoproduction events in SHERPA, we use its existing EPA interface for the photon flux, with improved phase space handling for the initial states for a more efficient integration, and added relevant photon PDFs to SHERPA's internal PDF interface. The resulting code has been made publicly available as part of the release of SHERPA 3.0.

3.1.1 Phase space handling

In the following we detail the structures for efficient phase space sampling, using the most involved example of doubly resolved photon-photon collisions at lepton colliders, schematically depicted in Fig. 3.1. In the cases of direct photons, the corresponding ISR terms do not need to be taken into account and the corresponding phase space integration is simplified.

The two incoming leptons have momenta p_1 and p_2, and a (beam) c.m.-system characterised by the c.m.-energy squared s_{12} and its rapidity y_{12} in the lab system. The momenta of the photons emitted by the leptons, p_1' and p_2', create a (photon) c.m.-system characterised by its c.m.-energy squared s_{12}' and rapidity y_{12}' with respect to the beam system.

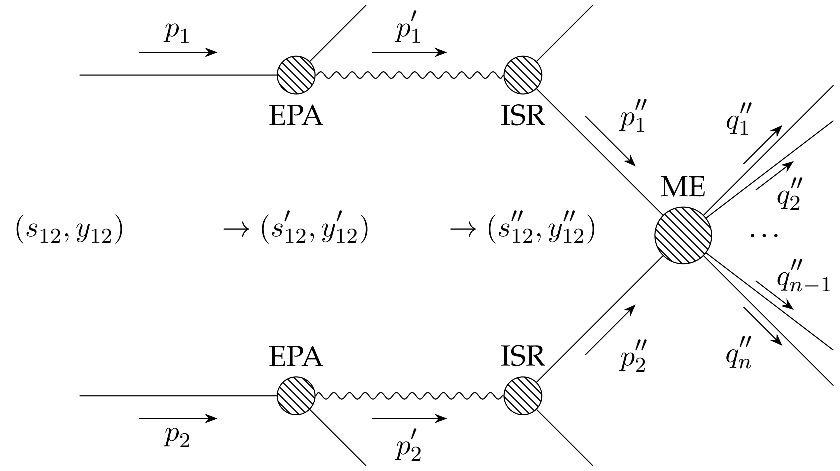


Figure 3.1: Schematic sketch of the phase space mappings between the different steps in the initial states, i.e. the Equivalent Photon Approximation (EPA), the Initial State Radiation (ISR), and the Matrix Element (ME). Each coordinate pair of Mandelstam-$s'^{(\prime)}$ and rapidity $y'^{(\prime)}$ is randomly sampled and the momenta are calculated as functions of these. Adapted from [90].

The partonic structure of the photons, as described by the PDFs, results in two partons with momenta p_1'' and p_2'' to finally enter the hard process which will result in final state particles with momenta q_i''. The hard scattering is characterised by a c.m.-energy of s_{12}'' and a rapidity y_{12}'' with respect to the photon system. This structure requires two nested integrations for the two successive "initial states" (photons and partons): first an integration over s_{12}' and y_{12}', with factors given by the EPA spectra, and then an integration over s_{12}'' and y_{12}'', with factors given by the PDFs, before adding the integration over the final state phase space over the outgoing momenta q_i''. Efficient integration over this complex phase space in SHERPA is facilitated through the multichannel method [91] with automatically generated integration channels that map out intrinsic structures such as s-channel resonances etc.

For each generation of a phase space point, the generated variables $\left(s_{12}', y_{12}'\right)$ are then transformed according to

$$x_{1,2} = \sqrt{\frac{s_{12}'}{s_{12}}} \exp(\pm y_{12}') \quad \text{and} \quad \mathrm{d}x_1 \mathrm{d}x_2 = \frac{1}{2s_{12}} \mathrm{d}s_{12}' \mathrm{d}y \,. \tag{3.1.1}$$

Accordingly, the full integral of the double-resolved component as exemplified in Fig. 3.1 reads

$$\begin{aligned}\sigma\left(ee\to eeX\right) =&\sum_{i,j}\int_0^1 \mathrm{d}x_1 f_\gamma^{(e)}\left(x_1\right)\int_0^1 \mathrm{d}x_2 f_\gamma^{(e)}\left(x_2\right)\\ &\int_0^1 \mathrm{d}x_1' f_i^{(\gamma)}\left(x_1',\mu_F\right)\int_0^1 \mathrm{d}x_2' f_j^{(\gamma)}\left(x_2',\mu_F\right)\hat{\sigma}\left(ji\to X\right)\\ =&\sum_{i,j}\int\frac{\mathrm{d}s_{12}'}{2s_{12}}\int\mathrm{d}y_{12}'\, f_\gamma^{(e)}\left(x_1\left(s_{12}',y_{12}'\right)\right) f_\gamma^{(e)}\left(x_2\left(s_{12}',y_{12}'\right)\right)\\ &\int\frac{\mathrm{d}s_{12}''}{2s_{12}'}\int\mathrm{d}y_{12}''\, f_i^{(\gamma)}\left(x_1'\left(s_{12}'',y_{12}''\right),\mu_F\right) f_j^{(\gamma)}\left(x_2'\left(s_{12}'',y_{12}''\right),\mu_F\right)\hat{\sigma}\left(ji\to X\right)\;.\end{aligned}\tag{3.1.2}$$

It is important to remember that as explained in Section 2.2.3 the direct components are separately computed by replacing $f_i^{(\gamma)}\left(x,\mu_F\right)\to\delta(1-x)$. Furthermore, we point out that the only physically meaningful quantity is the sum of all these contributions, i.e. only after taking into account every term as laid out in Eq. (2.2.6).

3.1.2 Equivalent Photon Approximation

The equivalent photon approximation encoded in the Weizsäcker-Williams formula [45–47] is based on the observation that quasi-virtual photons can be approximated through real photons for small virtualities $Q^2 < \Lambda_{\text{cut}}^2$. As photoproduction events are dominated by the interaction of low-virtuality photons, the differential cross-section can be substituted by $\mathrm{d}\sigma_{eX} = \sigma_{\gamma X}(Q^2 = 0) f_\gamma^{(e)}(x)\mathrm{d}x$. SHERPA uses an improved version of the spectrum, following [92], which introduces the term proportional to m_e^2 to the spectrum, and the photons are assumed to be collinear to the electron beam. The dependence of the photon spectrum on the photon virtuality is integrated out. This results in the following spectrum for electrons:

$$f_\gamma^{(e)}(x) = \frac{\alpha_{\text{em}}}{2\pi}\frac{\mathrm{d}x}{x}\left[\left(1+(1-x)^2\right)\log\left(\frac{Q_{\text{max}}^2}{Q_{\text{min}}^2}\right) - 2m_e^2x^2\left(\frac{1}{Q_{\text{min}}^2}-\frac{1}{Q_{\text{max}}^2}\right)\right]\;.\tag{3.1.3}$$

Here, x denotes the ratio of photon to electron energy, $\frac{E_\gamma}{E_e}$, and α_{em} is the electromagnetic coupling constant and the term proportional to $\log\left(\frac{Q_{\text{max}}^2}{Q_{\text{min}}^2}\right)$ will be called Leading Logarithmic (LL) and the full formulate Next-to-Leading Logarithmic (NLL). $Q^2_{\rm max}$ and $Q^2_{\rm min}$ denote the maximal and minimal photon virtuality, where the latter can be calculated from kinematic restrictions and is given by

$$Q^2_{\rm min} = \frac{m_e^2 x^2}{1-x} \ . \tag{3.1.4}$$

The maximal virtuality is given by the experimental setup and the maximal deflection angle of the electron, $\theta_{\rm max}$, below which the hard process is still considered to be photon-induced. It is given by

$$Q^2_{\rm max} = \min\Big(Q^2_{\rm min} + E_e^2(1-x)\theta^2_{\rm max},\ Q^2_{\rm max,fixed}\Big) \ . \tag{3.1.5}$$

The default choices are $\theta_{\rm max} = 0.3$ and $Q^2_{\rm max,fixed} = 3\ {\rm GeV}^2$, but they can be overwritten by the user, c.f. the SHERPA manual [93].

As shown in Fig. 3.2, the NLL correction accounts for approximately a 4% decrease in the flux.

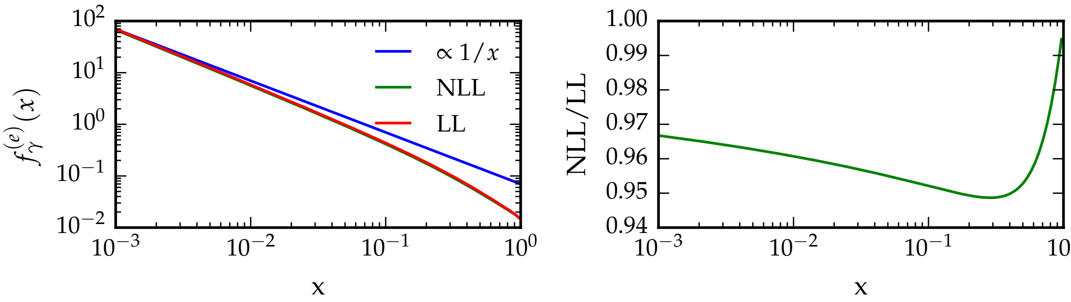


Figure 3.2: Spectrum of the photon flux at the Leading Logarithmic (LL) and Next to Leading Logarithmic (NLL) accuracy (left) and the ratio between the two (right).

3.1.3 Photon PDFs

To facilitate a comparison over different parametrisations, four PDF libraries have been included in SHERPA, see Table 3.1 for a summary. Currently, all PDFs are evaluated at virtuality $Q^2 = 0$. The extension to virtual photons, taking also into account longitudinal polarisations, will be introduced in a later release.

Table 3.1: Photon PDF libraries included in SHERPA and their properties.

Name	# sets	Virtual?	NLO?	# flavours	x-range	μ_F^2-range
GRV [48]	2	No	Yes	5	$[10^{-5}, 1]$	$[0.25, 10^6]$
SAL [51]	1	No	Yes	6	$[10^{-5}, 0.9999]$	$[2, 8 \cdot 10^4]$
SaS [52,53]	4	Yes	Yes	6	$[10^{-5}, 1]$	$[0.25, 10^6]$
CJK [49,50]	4	No	Yes	5	$[10^{-5}, 1]$	$[0.25, 2 \cdot 10^5]$

Fig. 3.3 shows the gluon PDF for different sets. The gluon has the largest PDF within the photon, but the exact content is largely unconstrained as can be seen by the large differences between the different sets spanning up to an order of magnitude for small x values. Additionally, we show a comparison between photon and proton PDFs in Fig. 3.4. Apart from the much lower values in the photon, the biggest difference is the quark content at large values of x; while the proton exhibits a valence structure, the photon has equal distributions for quarks and antiquarks, resembling the sea quark distributions in the proton.

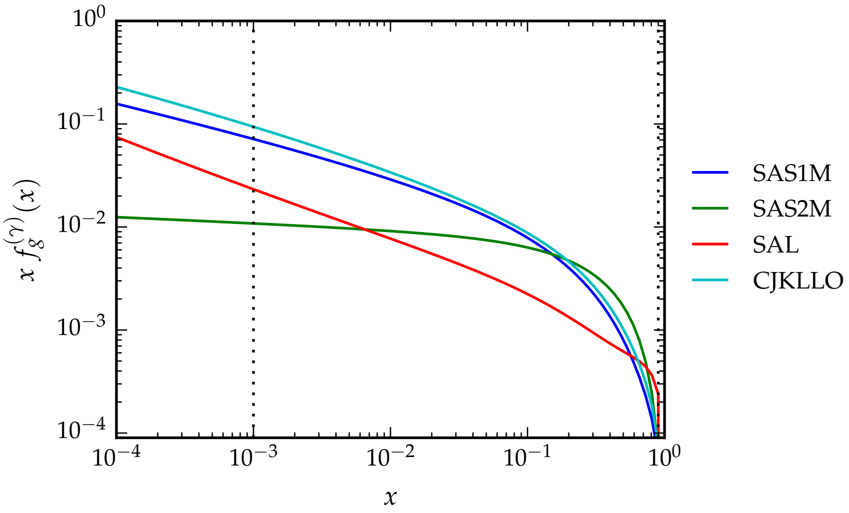


Figure 3.3: Comparison of the gluon content of the photon for the SAS1M, SAS2M, SAL and CJKLLO PDF sets at $\mu_F^2 = 5\ \mathrm{GeV}^2$. The vertical dotted black lines indicate the range of data, $x \in [0.001, 0.9]$, that was used for the latest fit in the SAL library.

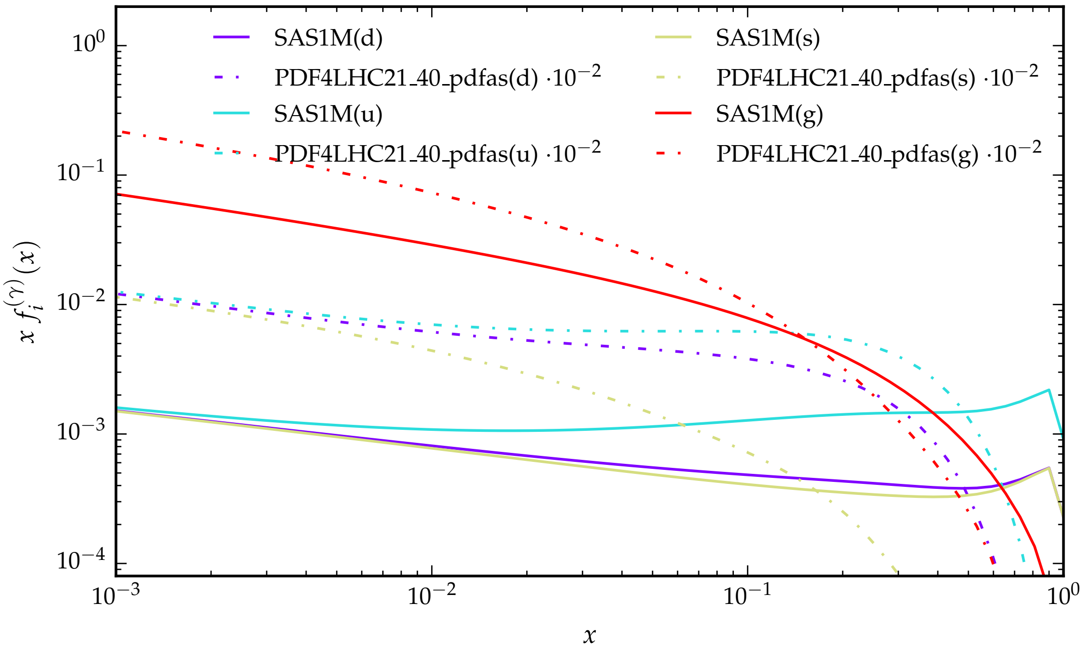


Figure 3.4: Comparison of the light quarks and gluon contents for the SAS1M photon PDF and the PDF4LHC21_40_pdfas proton PDF at $\mu_F^2 = 5\ \text{GeV}^2$.

3.1.4 NLO and matching to the parton shower

The procedure for a NLO calculation in photoproduction has been presented in [94]. We will follow the same line of arguments for the matching to the parton shower, and start by recalling (c.f. Section 2.2.3) the DGLAP evolution for the photon PDF to be

$$\frac{\partial f_i^{(\gamma)}}{\partial \log \mu^2} = \frac{\alpha_{\text{em}}}{2\pi} P_{i\gamma} + \frac{\alpha_S}{2\pi} \sum_j P_{ij} \otimes f_j^{(\gamma)} \,. \tag{3.1.6}$$

At Next-to-Leading-Order, there will be additional collinear divergences in the real-corrections, stemming from the photon splitting. These divergences appear only in the direct-photon component, but they can be subtracted using the splitting kernel as the corresponding term in the photon PDF evolution, $P_{i\gamma}$, has the same behaviour in the limit.

To ensure an exact cancellation between these terms, the PDF has to use the same factorisation scheme as the subtraction scheme, which in the case of SHERPA is the $\overline{\text{MS}}$ scheme.

This reduces the number of possible PDF sets directly available for the NLO calculation in SHERPA down to SAS1M and SAS2M from the SaS library. We note that additional PDFs can be made available easily by adding the respective factorisation scheme dependent correction terms [95]. Apart from this subtlety in the choice of PDFs, both NLO calculations and full simulations proceed in full analogy to the more familiar case of, e.g., proton-proton collisions.

3.1.5 Multiple-parton interactions in photons and the beam remnants

To model multiple interactions with hadron-like photons, one has to extend the model sketched in Section 2.3.4 in two ways. Firstly, due to the variable photon energy, the computation of the partonic and the total hadronic cross-section becomes dependent on the c.m.s.-energy of the system and has to be recalculated for each event. And secondly, one has to arrive at a parametrisation in Regge theory of the total hadronic cross-section of the photon. The former is a technicality and has been generalised in our model, while the latter needs data from experiments. Parameters have been extracted as a superposition of different neutral vector meson states, in analogy with the VMD model for the PDF, in [54].

Another ingredient in an accurate event simulation is the question of beam remnants. The photon does not have the typical notion of sea quarks and valence quarks; one could either see all quarks as sea quarks, as the quarks and anti-quarks come with the same distributions as is typical for sea quarks. Or one could see all quarks as valence partons, as a photon must contain at least one light quark–anti-quark pair.

In the implementation we use therefore a similar procedure is applied as in the remnant construction for a hadron, c.f. Section 2.3.5, however, with a few simplifications: given a number of partons extracted from the photon in the hard process and the MPI, we create a set of so-called spectators to compensate for flavour. During this process we ensure that at least one quark–anti-quark pair is present in the list of partons, i.e. we prohibit pure gluon states. In the latter case, a quark-antiquark pair is constructed from one of the light quark generations, i.e. any light-flavour quark pair can be "valence content". Once flavour has been compensated, we distribute the remaining beam momentum $p_{\mathrm{rem}} = p_{\mathrm{beam}} - \sum_{\mathrm{i}\in\mathrm{hard}} p_i$ according to a power law which has been fitted to the small-x behaviour of the photon PDFs; we use exponents of -1 for (anti-)quarks and -1.2 for gluons. Following that, the photon remnants are given — analogously to a hadron remnant — a small transverse momentum each, given by some non-perturbative intrinsic k_T distribution, ensuring overall momentum conservation.

3.2 Validation of the framework

We will now turn to the validation of our implementation, comparing results at MC@NLO precision with photoproduction data from the OPAL and ZEUS experiments. For each collider set-up and energy we generated samples of $5 \cdot 10^6$ events per component at MC@NLO accuracy. For the calculation of matrix elements we used AMEGIC [96] and COMIX [87] for the tree-level matrix elements and subtraction terms [97] and OPENLOOPS [98] for the one-loop matrix elements. We added the CSSHOWER parton shower [99] for the jet evolution, matched with the MC@NLO method [100, 101] as implemented in SHERPA [102]. Underlying event effects have been included through an implementation of the Sjostrand-van Zijl model [103, 104], adapted as sketched in Section 3.1.5, within SHERPA[1], and the partons were hadronised with the cluster fragmentation of AHADIC [105]. We consistently used the current default value for $\alpha_S(M_Z) = 0.118$ with three-loop running. As described in Section 3.1.4, the event generation must currently use PDFs based on the $\overline{\text{MS}}$ scheme, so the resolved-photon predictions were generated with, and averaged over, the SAS1M and SAS2M PDF sets. The factorisation scale and the renormalisation scale were both set to $\mu_F = \mu_R = H_T/2$, and we kept the 7-point variation for the scale uncertainty estimate.

We used the RIVET [106] framework with existing analyses implementing [34–36, 40, 41]. For each experiment, different components of the cross-section have to be summed over, for example,

$$\sigma_{\text{tot}} = \sigma_{\gamma\gamma} + \sigma_{j\gamma} + \sigma_{\gamma j} + \sigma_{jj} \tag{3.2.1}$$

for LEP and

$$\sigma_{\text{tot}} = \sigma_{\gamma j} + \sigma_{jj} \tag{3.2.2}$$

for HERA, where in both cases j denotes a photon or proton resolved through a PDF.

3.2.1 Comparison with LEP data

The OPAL analysis of photon-induced dijet production at 198 GeV c.m.-energy [34] offers the most differential observables, and we use it as the primary reference for our validation of photoproduction at LEP 2. To comply with experimental cuts, the k_T algorithm with a jet radius of $R = 1.0$ is used and cuts of $E_T > 4.5$ (2.5) GeV for the leading (subleading) jet are imposed.

---

[1] For details and a future tune to data we refer the reader to a forthcoming SHERPA manual.

$$x_\gamma^\pm = \frac{\sum\limits_{j=1,2} E^{(j)} \pm p_z^{(j)}}{\sum\limits_{i\in\mathrm{hfs}} E^{(i)} \pm p_z^{(i)}} \,, \tag{3.2.3}$$

are used in this analysis to disentangle direct, singly, and doubly resolved production. In their definition, the sum in the numerator is over the two jets, and the sum in the denominator is over all hadronic final state particles, thereby distilling the energy-fraction the jets have with respect to the overall photon energies.

This is exemplified by, e.g., the distribution of events in

$$\cos\Theta^* = \tanh\frac{\eta_1 - \eta_2}{2} \,, \tag{3.2.4}$$

an approximation of the angle between the two jets, and exhibited in Fig. 3.5. Apart from satisfying agreement with data, a few things are worth noticing here, which we will continue to observe also in the following: For $x_\gamma^\pm > 0.75$ the direct component dominates by about 1.5 orders of magnitude, with only small scale uncertainties, indicated by the light red band. Conversely, for $x_\gamma^\pm < 0.75$ the doubly-resolved component dominates with a significantly larger scale uncertainty, which, in this case, also includes factorisation scale uncertainties. Intuitively, the singly-resolved component in both cases ranges between the two other components. In addition, we observe that hadronisation effects reduce the cross-section in the unresolved domain, while the combination of hadronisation and multiple parton scattering increases it in the doubly-resolved regime. The visible effect in the latter suggests that a careful retuning of the MPI may further improve agreement with data.

We report that distributions in $x_\gamma^\pm$ for three different average dijet transverse energies $\bar{E}_T = \sum_{j=1,2} E_T^{(j)}/2$ experience a significant improvement in shape when going from Leading to Next-to-Leading Order, c.f. Fig. 3.6. However, in the transition region between doubly resolved to unresolved events, we notice a clear difference in shape: While for $x_\gamma^\pm < 0.6 - 0.7$ the prediction is relatively flat below the data, the underprediction at around $x_\gamma^\pm \approx 0.8$ persists at NLO. Apart from possibly insufficient photon PDFs — a point we will elucidate below — there are a number of possible explanations: First of all, as before, a retuning of MPI may come to the rescue and fill up the gap.

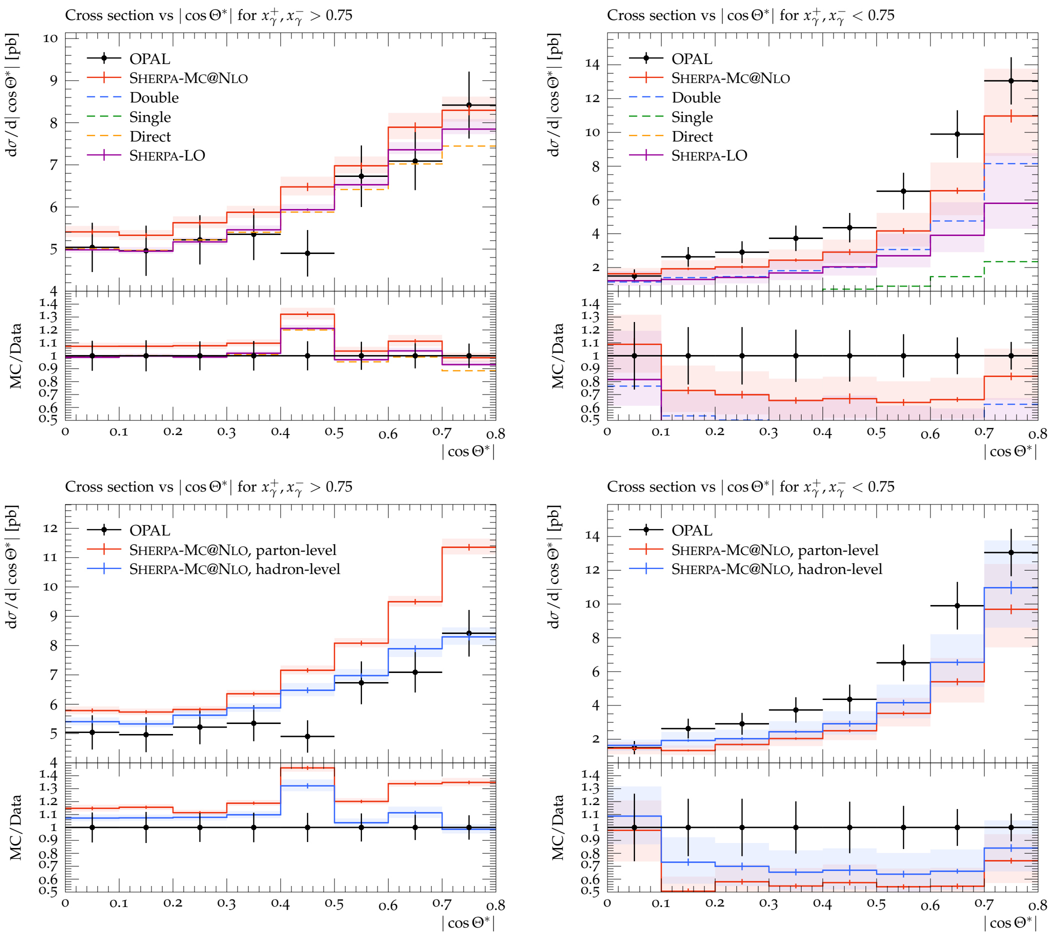


Figure 3.5: Differential dijet inclusive cross-section with respect to cos θ^*, defined in Eq. (3.2.4), comparing results of our SHERPA MC@NLO simulation at hadron-level with the LO simulation (top row) and with simulations at parton-level (bottom row) and with data from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34]. In the left and right panels the requirement $x_\gamma^\pm < 0.75$ and $x_\gamma^\pm > 0.75$ are applied and enhance resolved and direct contributions, respectively.

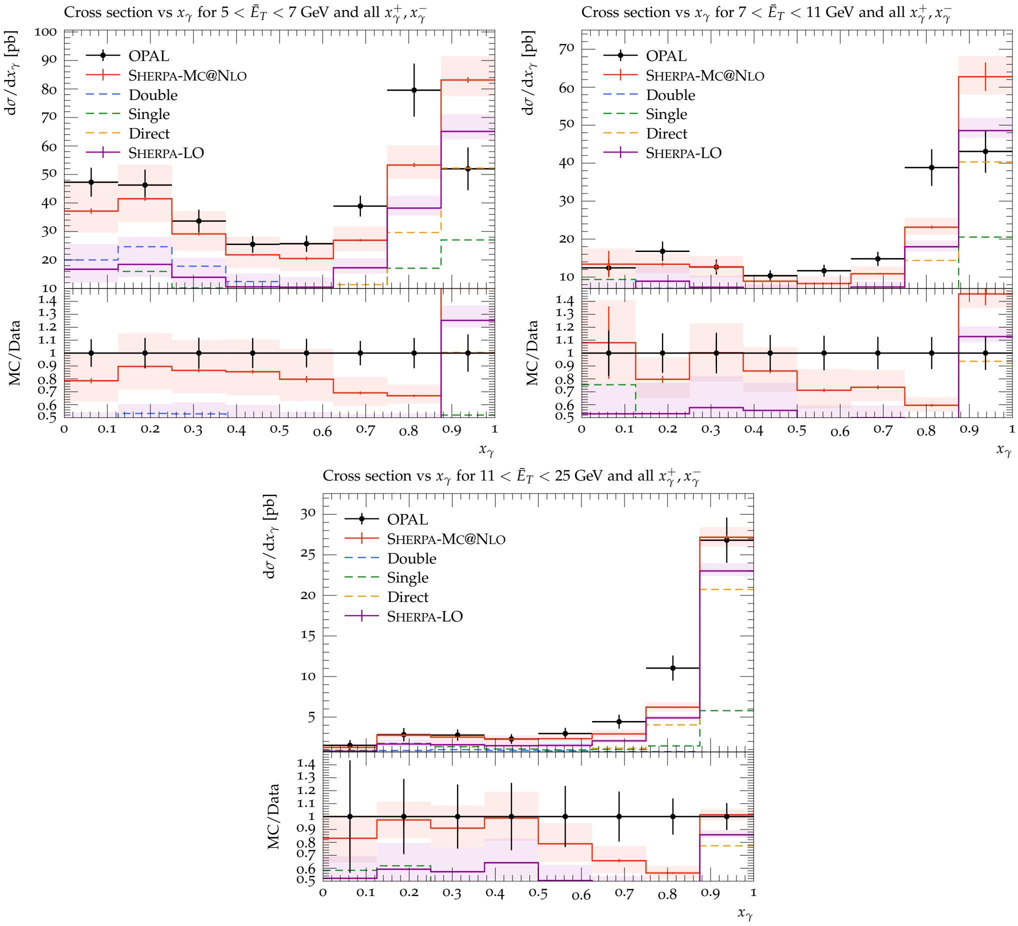
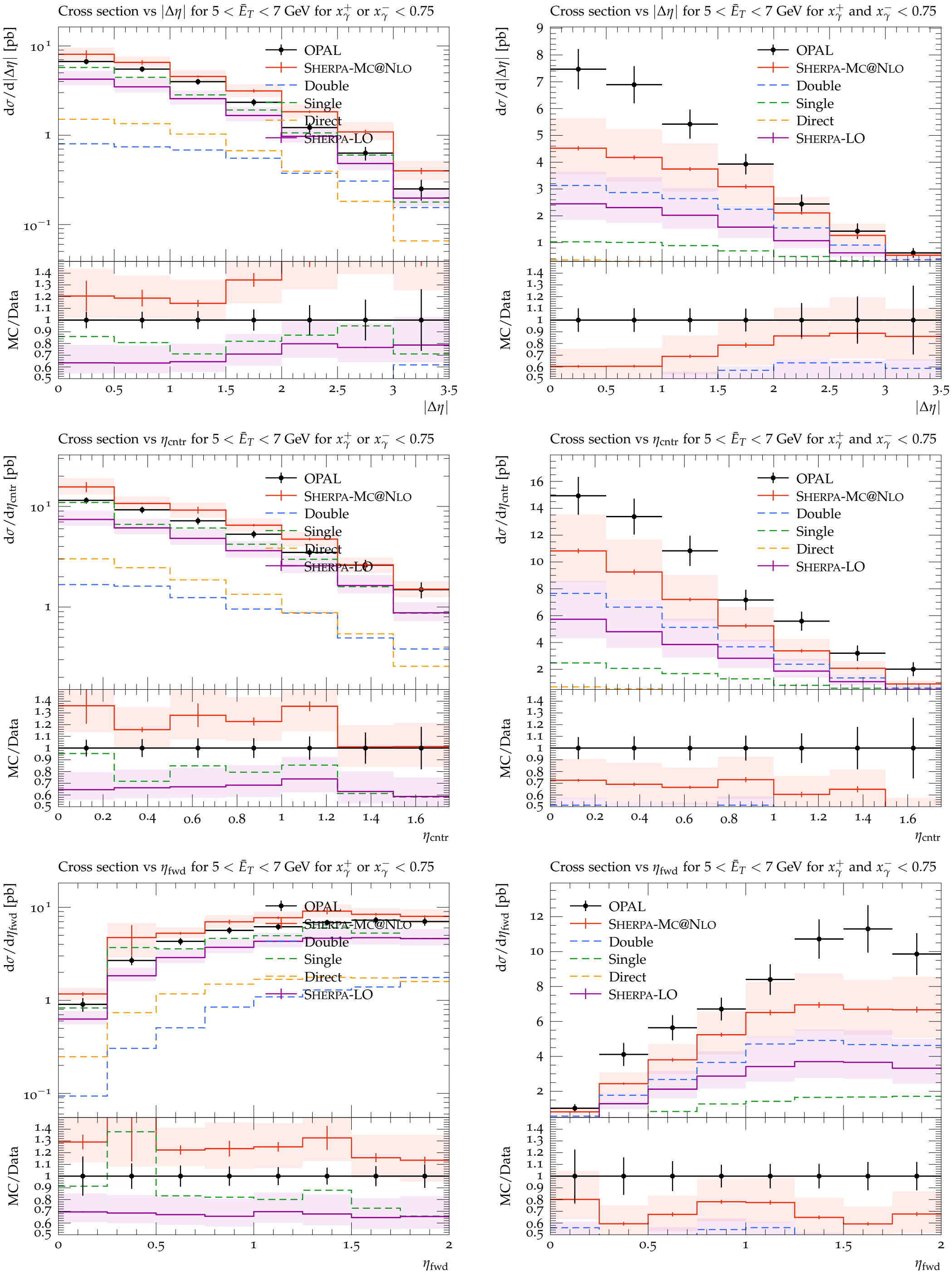


Figure 3.6: Distributions $x_\gamma^\pm$, collectively denoted as x_γ, in different bins of average transverse jet energy: $\bar{E}_T \in [5\,\mathrm{GeV}, 7\,\mathrm{GeV}]$ (top left), $\bar{E}_T \in [7\,\mathrm{GeV}, 11\,\mathrm{GeV}]$ (top right), $\bar{E}_T \in [11\,\mathrm{GeV}, 25\,\mathrm{GeV}]$ (bottom). Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from OPAL at an $e^- e^+$ c.m.-energy of 198 GeV [34].



Cross section vs $|\Delta\eta|$ for $5 < \bar{E}_T < 7$ GeV for x_γ^+ or $x_\gamma^- < 0.75$

Cross section vs $|\Delta\eta|$ for $5 < \bar{E}_T < 7$ GeV for x_γ^+ and $x_\gamma^- < 0.75$

Cross section vs η_{cntr} for $5 < \bar{E}_T < 7$ GeV for x_γ^+ or $x_\gamma^- < 0.75$

Cross section vs η_{cntr} for $5 < \bar{E}_T < 7$ GeV for x_γ^+ and $x_\gamma^- < 0.75$

Cross section vs η_{fwd} for $5 < \bar{E}_T < 7$ GeV for x_γ^+ or $x_\gamma^- < 0.75$

Cross section vs η_{fwd} for $5 < \bar{E}_T < 7$ GeV for x_γ^+ and $x_\gamma^- < 0.75$

dσ/d|Δη| [pb]
dσ/dη_cntr [pb]
dσ/dη_fwd [pb]
MC/Data
|Δη|
η_cntr
η_fwd
OPAL
SHERPA-MC@NLO
Double
Single
Direct
SHERPA-LO
10^1
1
10^-1
0 0.5 1 1.5 2 2.5 3 3.5
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
0 0.5 1 1.5 2
1 2 3 4 5 6 7 8 9
2 4 6 8 10 12 14 16
0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4

Figure 3.7: Distributions of $|\Delta\eta|$ (top row), $|\eta_{\mathrm{cntr}}|$ (middle row), and $|\eta_{\mathrm{fwd}}|$ (bottom row) for requiring $x_\gamma < 0.75$ for exactly one (left column) or both (right column), comparing MC@NLO and LO. Results of the SHERPA simulation are compared with results from OPAL at an $e^- e^+$ c.m.-energy of 198 GeV [34].

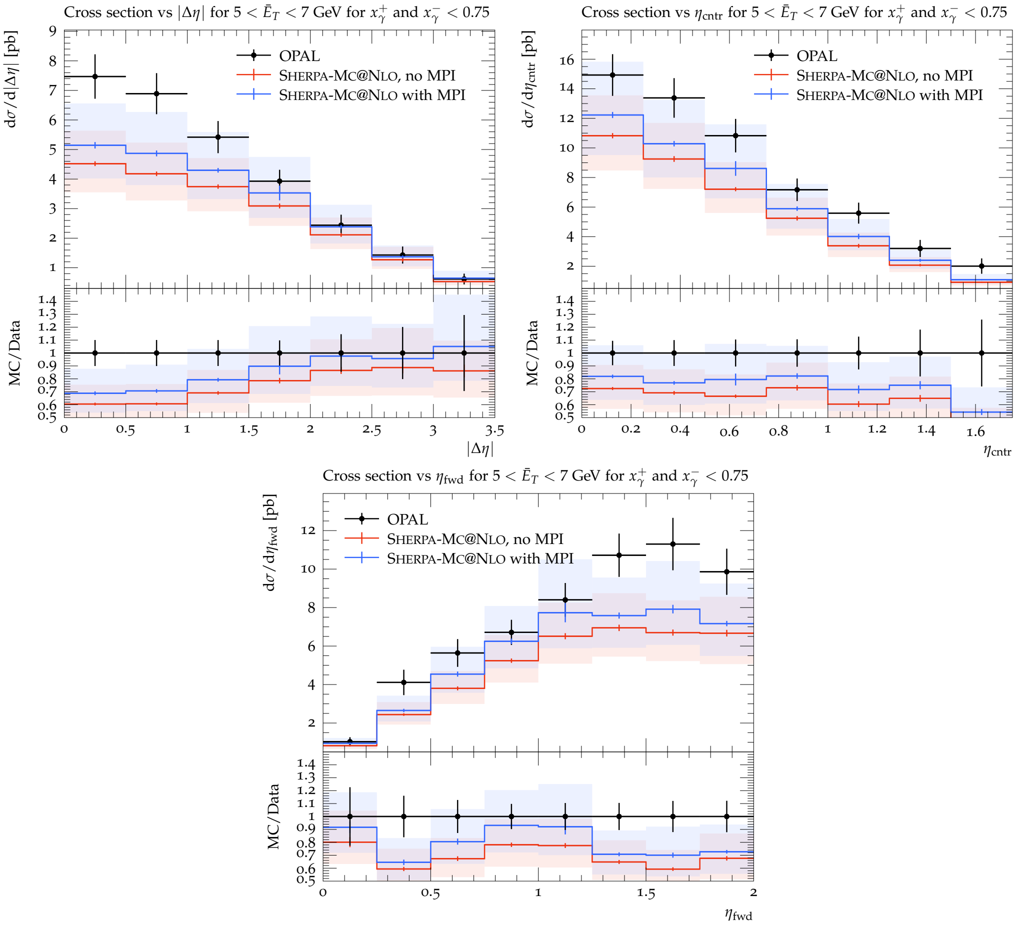


Figure 3.8: Distributions of $|\Delta\eta|$ (top left), $|\eta_{\text{cntr}}|$ (top right), and $|\eta_{\text{fwd}}|$ (bottom) with the requirement $x_\gamma^\pm < 0.75$, comparing MC@NLO with and without MPI. Results of the SHERPA simulation are compared with results from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34].

Secondly, this drop around $x_\gamma^\pm \approx 0.7$ could be attributed to the missing QED splitting kernel in the evolution of the parton shower. Including this term would impact the backwards evolution of the photonic initial state radiation leading to a photon being reconstructed as the initial state also in the case of a resolved process. This in turn would lead to less radiation being generated, therefore shifting the distribution of the resolved process towards larger $x_\gamma^\pm$ values. This splitting is included in PYTHIA and is part of the comparison in Section 4.2, but does not seem to improve the shape. The inclusion of this term in the evolution of the initial state showering in SHERPA is left for future work.

Finally, we should stress that our singly resolved events are described by the $2 \to 2$ scattering of on-shell photons with partons from the resolved photons, an approximation which is probably not entirely correct as virtual photons would lead to a DIS-like scattering of the resolved photon, thereby inducing somewhat different kinematics and scale choices.

Fig. 3.7 shows distributions of jet pseudorapidities and their differences. Again, the overall shape of the prediction is improved and the lowered NLO cross-section is countered by the inclusion of Multiple-Parton Interactions (MPI). In Fig. 3.8 we show a dedicated comparison of the effect of MPI in jet pseudorapidities which suggests that a tuning might at least partially fill the gap between the data and the simulation. As a last observable, we compare the distributions in average transverse jet energies in Fig. 3.9 and see overall good agreement for the different phase space, dominated by direct, single-resolved and double-resolved processes, respectively.

3.2.2 Comparison with HERA data

For the further validation of our implementation in electron-proton collisions we mainly rely on ZEUS data [41] taken at HERA Run 2. The kinematic cuts on the final states in the hard matrix element calculation are chosen to be a minimal transverse momentum of $p_T > 11(8)$ GeV for the (sub-)leading jets using the k_T clustering algorithm with radius $R = 1.0$ to safely capture the phase space cuts of $p_T^{(1)} > 14$ GeV and $p_T^{(2)} > 11$ GeV used in the analysis of the ZEUS Run 1 data [40], even after taking into account MPI. For the ZEUS Run 2 data [41], we choose $p_T > 13$ GeV to comply with the experimental cut of $E_T > 17$ GeV on the leading jet, with otherwise identical settings.

We use the same scale setting algorithm as before, and we also evaluate the theory uncertainties as a combination of scale and PDF uncertainties like in the case of electron-positron colliders in the section above. Defining, in analogy to the case at lepton colliders above, Eqs. (3.2.3) and (3.2.4), respectively,

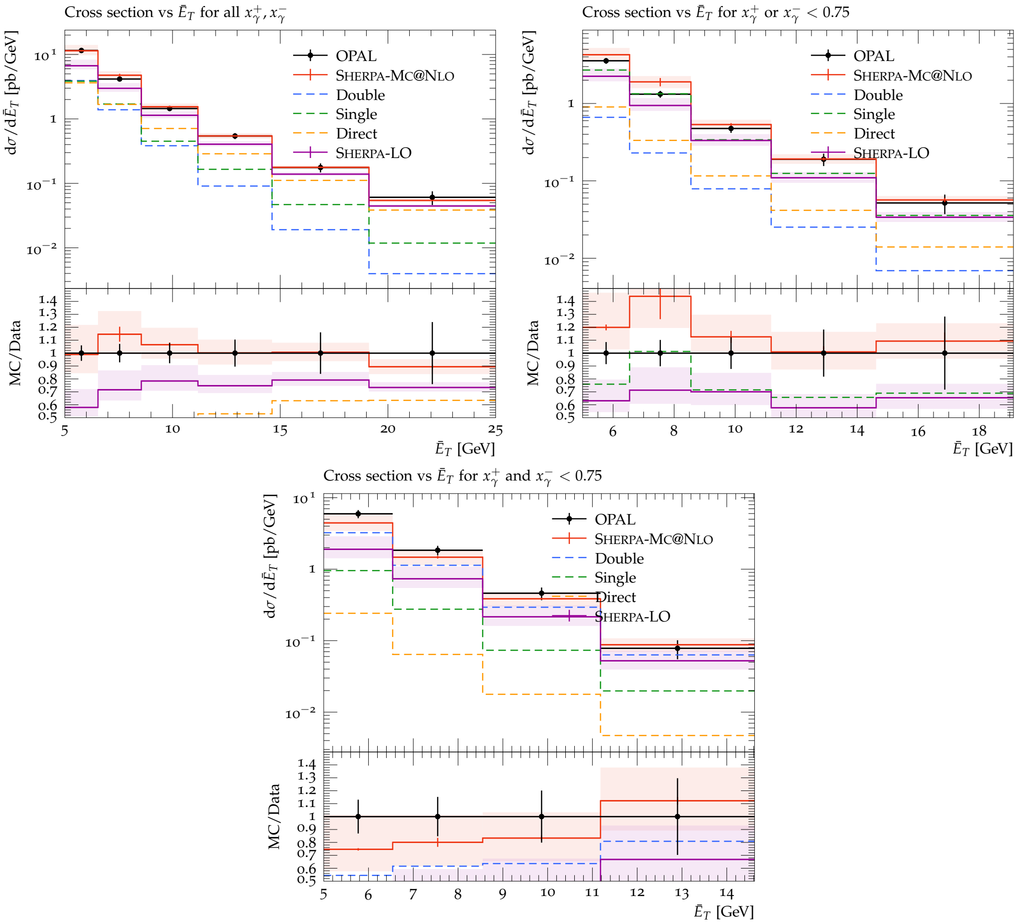


Figure 3.9: Distributions $\bar{E}_T$ for all values of $x_\gamma^\pm$ (top left), requiring x_γ^+ or $x_\gamma^- < 0.75$ (top right), and requiring $x_\gamma^\pm < 0.75$ (bottom). Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from OPAL at an e^-e^+ c.m.-energy of 198 GeV [34].

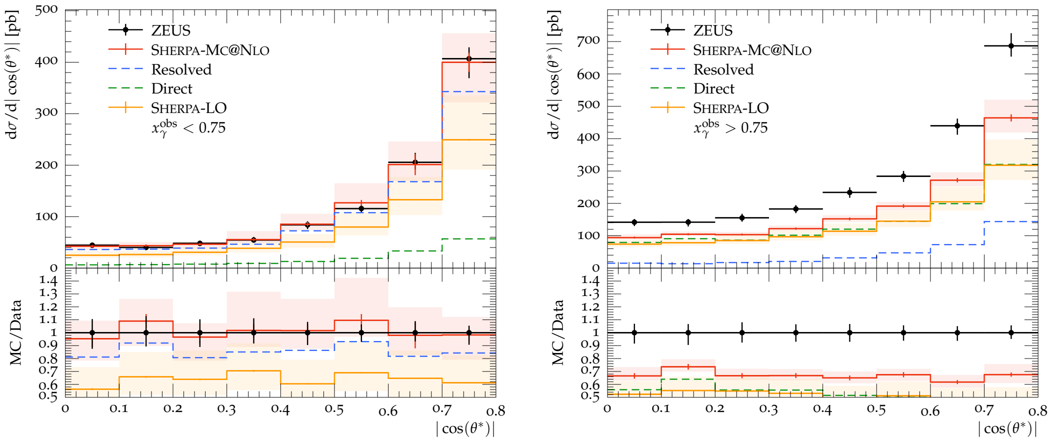


Figure 3.10: Differential dijet inclusive cross-section with respect to $\cos\theta^*$ for $x_\gamma^{\text{obs}} < 0.75$ (left) and $x_\gamma^{\text{obs}} > 0.75$ (right), comparing results of our SHERPA MC@NLO simulation with ZEUS Run 1 data [40].

$$x_\gamma^{\text{obs}} = \frac{E_T^{(1)} e^{-\eta^{(1)}} + E_T^{(2)} e^{-\eta^{(2)}}}{2yE_e}\,, \quad \text{and} \quad \cos\theta^* = \tanh\frac{\eta^{(1)} - \eta^{(2)}}{2}\,. \tag{3.2.5}$$

Here $^{(1,2)}$ labels the leading and sub-leading jet, E_e is the lepton energy, and y is the energy fraction of the photon with respect to the lepton. As before we observe that x_γ is excellently suited to disentangle unresolved and resolved photon interactions, c.f. Fig. 3.10. In the resolved regime we again observe satisfactory agreement with data, indicating that the summation over the different components is correct. Interestingly, as suggested by the right panel of the figure, it appears as if the direct component is not sufficient to fully recover the experimental cross-section.

We study this effect double-differentially in Fig. 3.11. While the calculation undershoots the highest x_γ^{obs} bin in every bin of E_T^{jet1}, the discrepancy is less pronounced for high jet energies.

Possible explanations, as before, are related to the missing "anomalous" $\gamma \to q\bar{q}$ splitting in the backwards evolution, or, possibly more relevant here, a failure of the strictly on-shell approximation of the incident photons inherent to the treatment through EPA.

With this caveat in mind we will now turn to a more differential analysis of QCD final states in photoproduction at HERA. In Fig. 3.12 we compare the parton- against the hadron-level results and as previously observed, the shape improves significantly through the

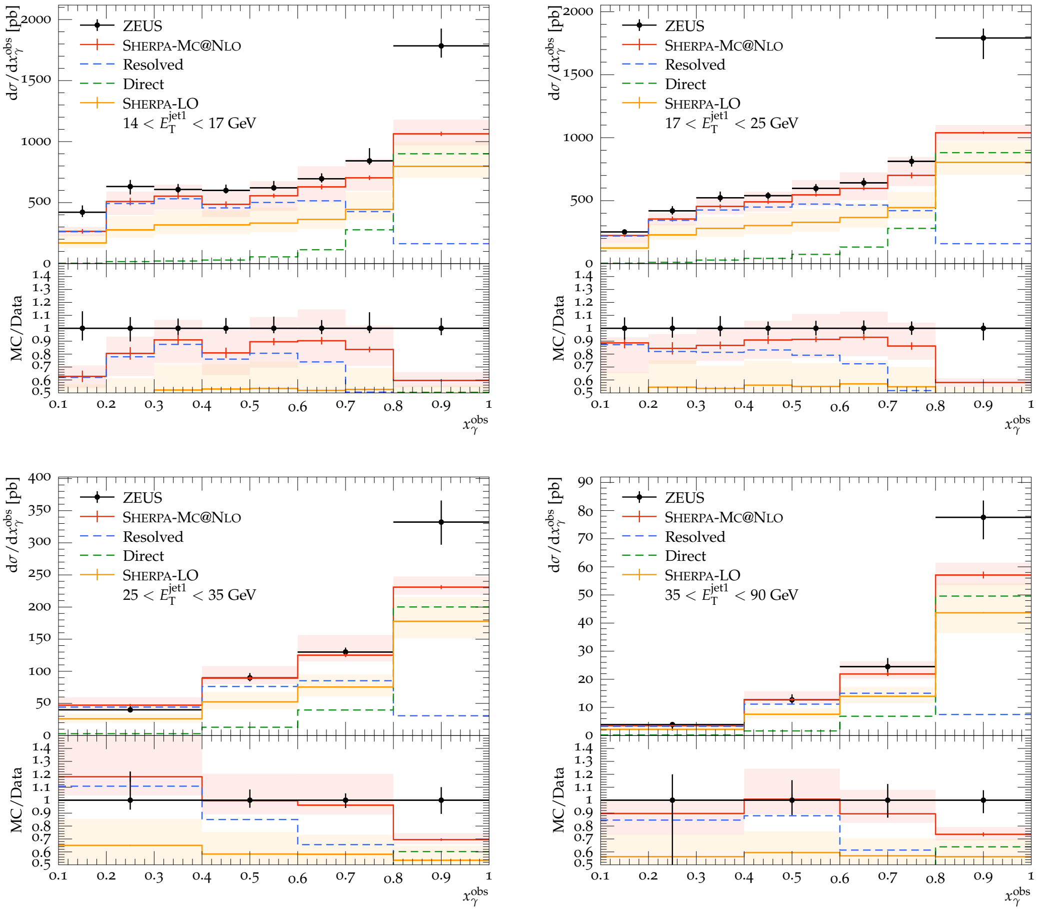


Figure 3.11: Differential dijet inclusive cross-section with respect to x_γ^{obs}, in different bins of leading jet transverse energy: $E_T \in [14\,\text{GeV}, 17\,\text{GeV}]$ (top left), $E_T \in [17\,\text{GeV}, 25\,\text{GeV}]$ (top right), $E_T \in [25\,\text{GeV}, 35\,\text{GeV}]$ (bottom left), $E_T \in [35\,\text{GeV}, 90\,\text{GeV}]$ (bottom right), comparing results of our SHERPA MC@NLO simulation with ZEUS Run 1 data [40].

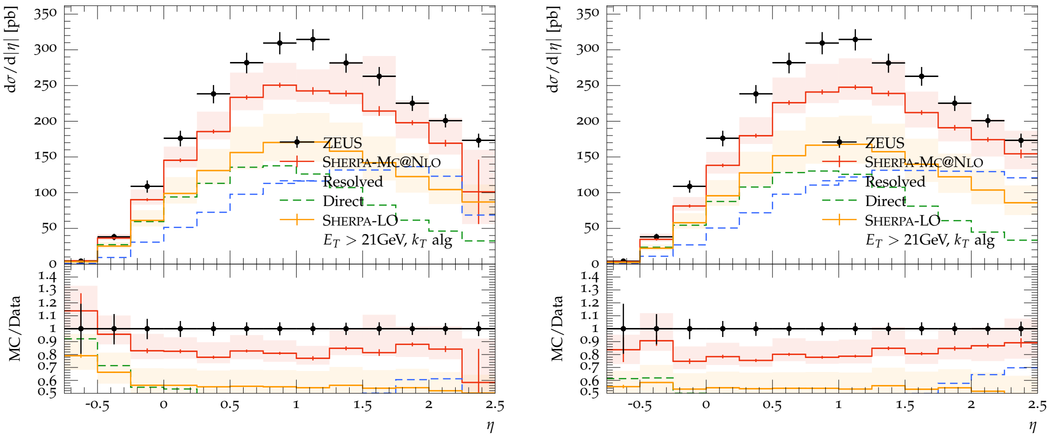


Figure 3.12: Differential single-jet inclusive photoproduction cross-section with respect to the pseudorapidity of the leading jet, $\frac{d\sigma}{d\eta}$, comparing results of our SHERPA simulation at parton (left) and hadron-level (right) with ZEUS Run 2 data [41].

combined effect of hadronisation and MPI. This is most visible in the phase space of $\eta < 0$, rendering the distribution of the simulation data flat compared to the experimental data. While there remains a discrepancy between simulation and data of around 10-20%, it might be explained by three observations. First, this analysis' cuts require only one jet in the final state which allows for contributions from the DIS region to leak into this measurement, for example if the scattered beam electron has not been correctly detected or identified; however, this should have been taken care of in the experimental analysis of the backgrounds. Secondly, the precise value of the strong coupling in the fit of the photon PDFs is not explicitly mentioned in the corresponding publications [52, 53]. An updated photon PDF fit would be performed with the current world average of α_S and might further reduce the discrepancy.

Lastly, the modelling of multiple-parton interactions for photon-proton interactions needs a fitting to the data for both proton-proton and photon-proton data. As neither have been tuned to data yet it can be suspected that the MPI will receive larger contributions, thereby improving overall agreement of simulation and data. We compare pseudorapidity distributions with and without MPI in Fig. 3.13; for rather central jets, the simulation agrees well with the data after the inclusion of MPI, while for forward jets the agreement deteriorates. The spectrum of the jet transverse energies E_T, displayed in Fig. 3.14 exhibits a slight shape in the distribution for forward jets and small E_T. The same effect can be seen in Fig. 3.15, where the simulation describes data well for central leading jets, but worsens

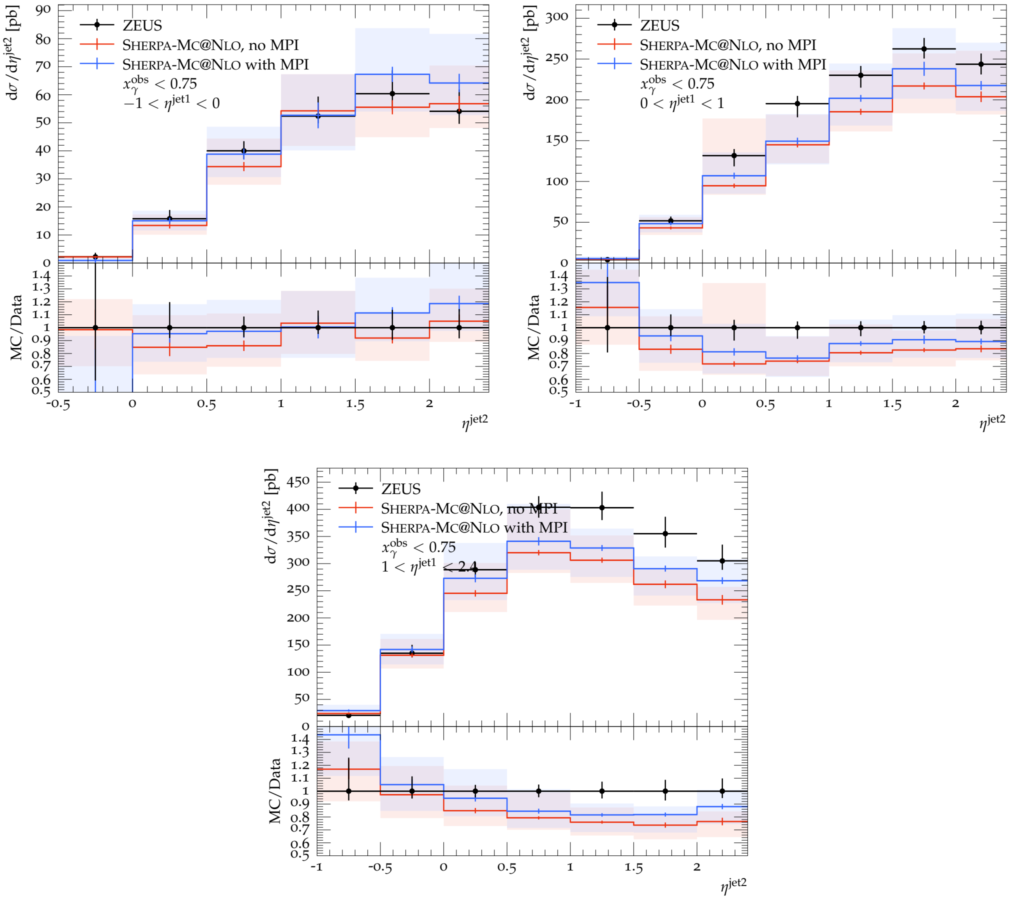


Figure 3.13: Distributions of the subleading jet pseudorapidity η^{jet2} with the requirement $x_\gamma^{\text{obs}} < 0.75$ in three bins of leading jet pseudorapidity, $-1 < \eta^{\text{jet1}} < 0$ (top left), $0 < \eta^{\text{jet1}} < 1$ (top right) and $1 < \eta^{\text{jet1}} < 2.4$ (bottom), comparing MC@NLO with and without MPI. Results of the SHERPA simulation are compared with results from ZEUS [40].

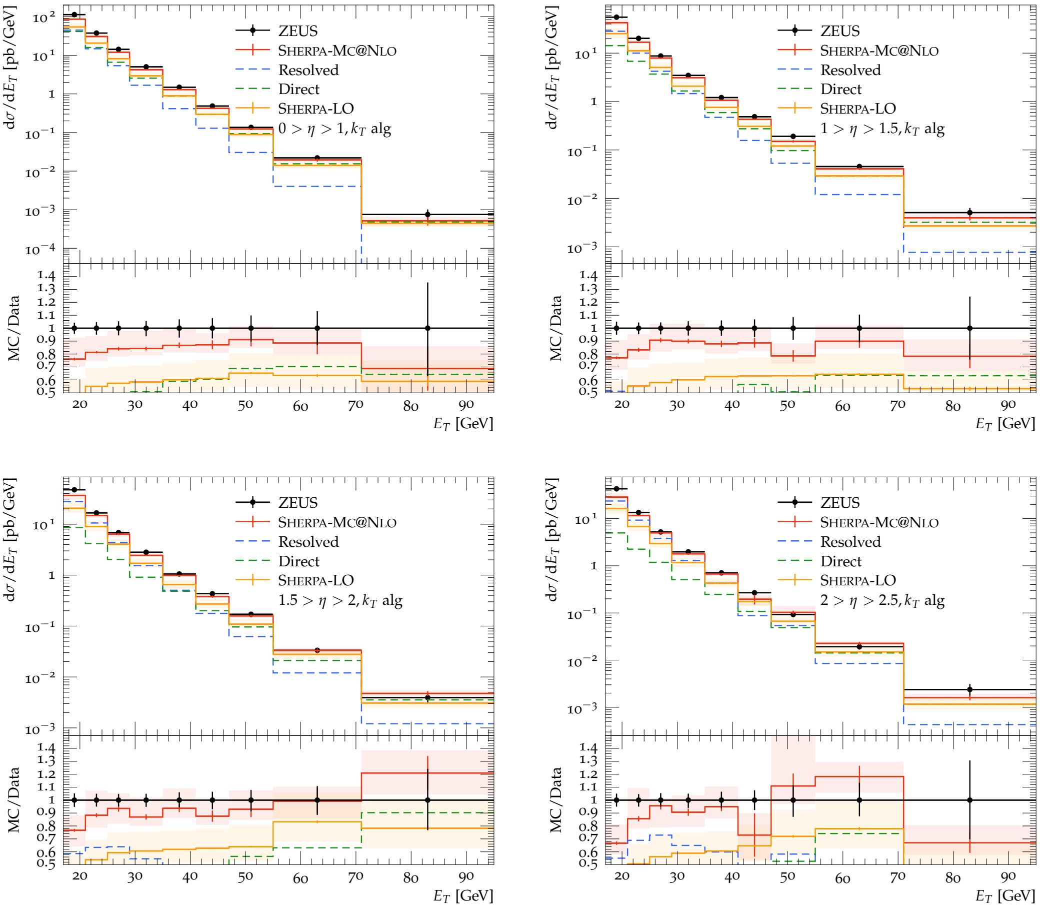


Figure 3.14: Single-jet inclusive transverse energy spectra for the leading jet in different pseudorapidity bins of the k_T-jets: $0 < \eta < 1$ (top left), $1 < \eta < 1.5$ (top right), $1.5 < \eta < 2$ (bottom left), and $2 < \eta < 2.5$ (bottom right), comparing SHERPA MC@NLO results with ZEUS Run 2 data [41].

as both go more forward. As this part of the phase space receives contributions from the MPI, we again suspect that the naive parameter choice underestimate the amount of additionally generated radiation and, probably more importantly, that merging of additional jets would improve the agreement as well.

3.2.3 Photon PDF for precision phenomenology

The predictions of photoproduction cross-sections and distributions in low-x_γ space exhibit large variations depending on the used photon PDF. In fact, these deviations can be as large or even larger in value than the estimate of higher-order corrections through the scale variations, especially for the simulation of photoproduction at lepton colliders, see Fig. 3.16. There we show the very inclusive $\cos\Theta^*$ and $\cos\theta^*$ distributions, obtained at leading order, at LEP and HERA, respectively., indicating LO scale and PDF uncertainties separately. For the latter we present the full range of results from all available leading order PDFs, and we observe that the PDF uncertainties alone are of equal size as the LO scale uncertainties. This underlines the need for a comprehensive retuning of the photon PDFs with available data, and including higher-order calculations and simulations.

It is also worth noting that a consistent simulation of photoproduction at hadron-lepton colliders necessitates a combined fit of the photon and proton PDFs. Depending on the proton PDF and its value for α_S the inclusive cross-section in a HERA Run 2 simulation gives deviations of about 20%, as can be seen in Table 3.2. This underlines the necessity for a systematic refit of photon PDFs to use in conjunction with modern proton PDFs. While no new data has been taken since the retiring of the HERA collider, a consistent fit to all the data with the updated values for α_S and including error estimates should increase the confidence in precision phenomenology for photoproduction.

3.3 Predictions for jet photoproduction at the EIC

Using our implementation we apply the calculation to the EIC. The beam setup is chosen to be an electron and a proton beam with energies of 18 and 275 GeV, respectively, in accordance with the highest energy scenario at the planned EIC facilities [8]. Similarly to studies in [109], particles are analysed in the laboratory frame using the anti-k_T jet algorithm with $R = 1.0$ and a minimum transverse energy of $E_T > 6$ GeV and a pseudorapidity $|\eta| < 4$, demanding one jet in the event.

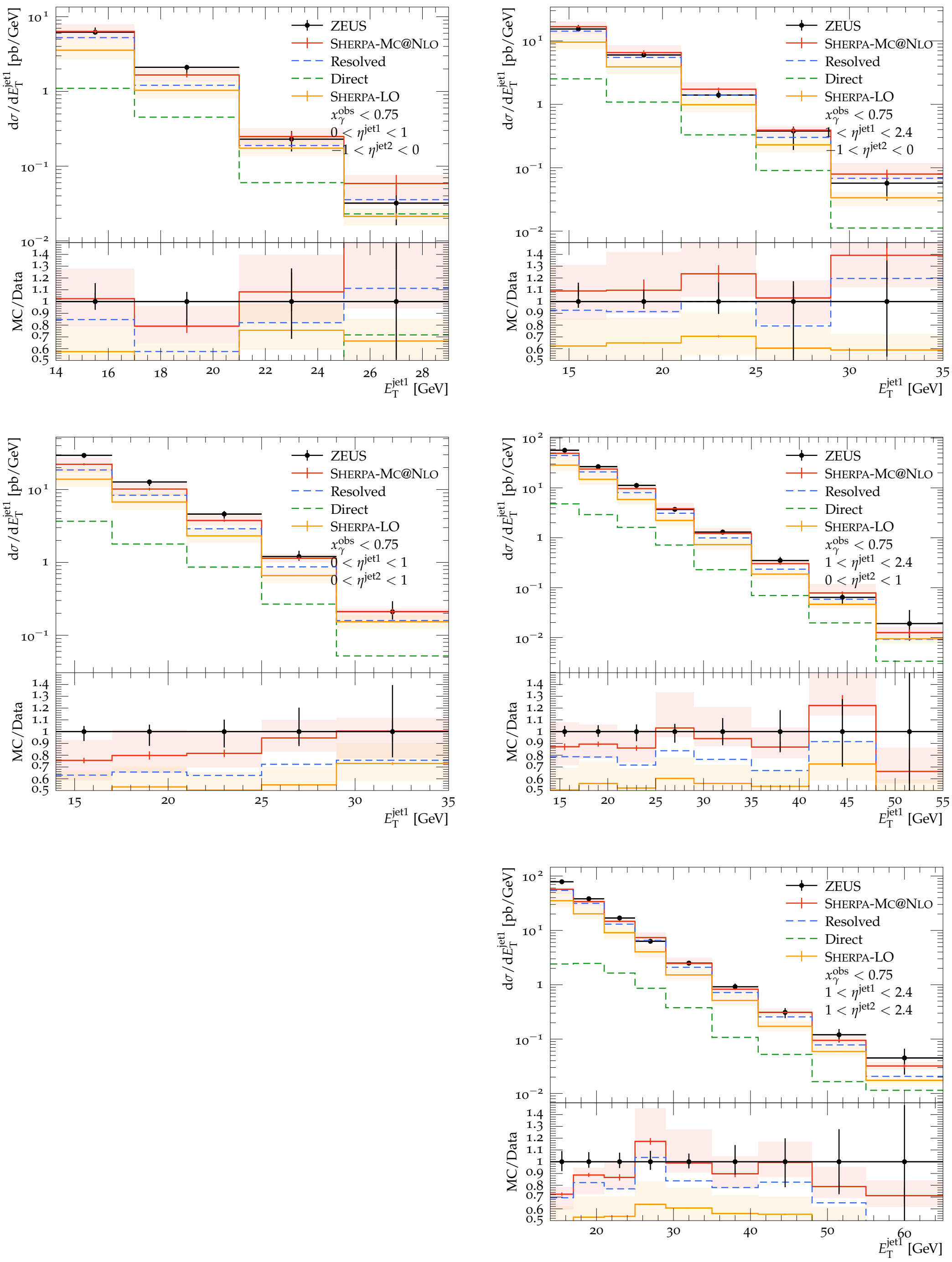


Figure 3.15: Differential dijet inclusive cross-section with respect to the transverse momentum of the leading jet, with $0 < \eta^{(1)} < 1$ (left column) and $1 < \eta^{(1)} < 2.4$ (right column) and in different bins for $\eta^{(2)}$, comparing results of our SHERPA MC@NLO simulation at hadron-level including MPI effects with the LO simulation and with data [40] taken by ZEUS at HERA Run 1.

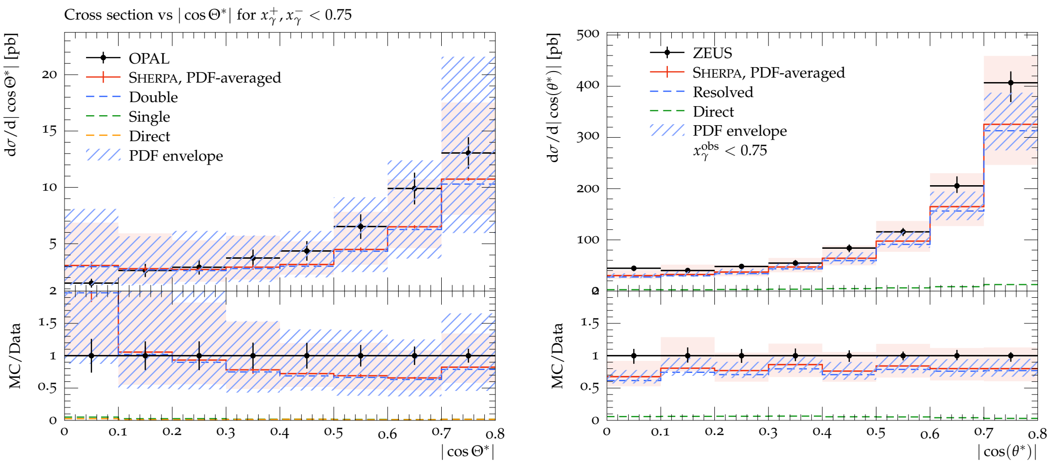


Figure 3.16: Distributions for $\cos\Theta^*$ at LEP (left) and for $\cos\theta^*$ at HERA (right), comparing SHERPA's LO simulations with data from OPAL [34] and ZEUS [40]. Scale uncertainties at LO are indicated by the pink band, while PDF uncertainties are shown with the blue hatched area.

Table 3.2: Inclusive cross-sections for one million events in a HERA Run 1 dijet photoproduction setup with two different proton PDFs and the same PDF for the photon.

PDF	NNPDF23_lo_as_0130_qed [107]	PDF4LHC21_40_pdfas [108]
α_S	0.13	0.118
Order	1	3
$\sigma(\gamma j \to jj)$ / nb	2.85 ± 0.02	2.303 ± 0.016
$\sigma(jj \to jj)$ / nb	2.151 ± 0.002	1.997 ± 0.002

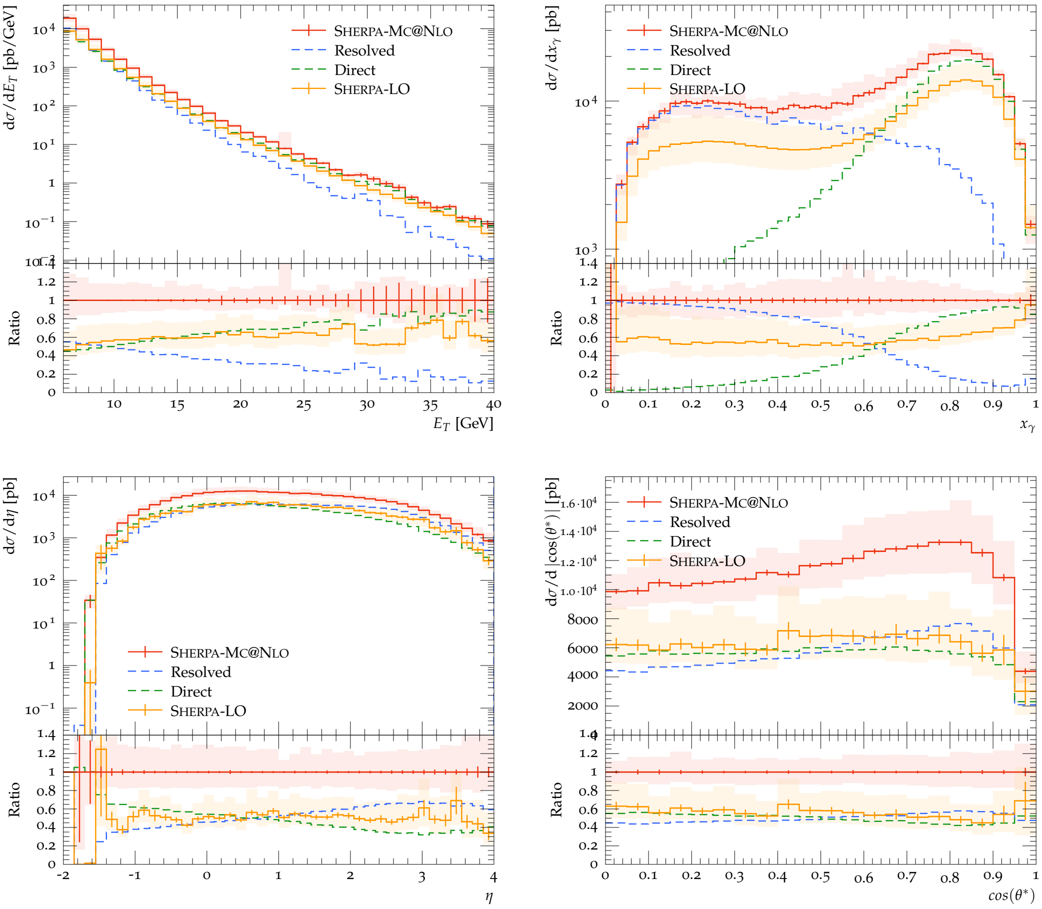


Figure 3.17: Distributions of inclusive jet transverse energy E_T (top left), the x_γ (top right), inclusive jet pseudorapidity η (bottom left), and $\cos(\theta^*)$ (bottom right) of the SHERPA simulation with MC@NLO accuracy, compared with results at LO.

In Fig. 3.17, we present the inclusive jet transverse momentum distribution, inclusive jet pseudorapidity, and the observables x_γ and $\cos(\theta^*)$ where the latter are defined in Eq. (3.2.5). It had previously been observed that the x_γ observable is especially suitable to distinguish different photon PDF parametrisations [40]. The spectrum of the jet transverse energies is driven by the resolved component at $E_T < 10$ GeV, with the direct component taking over with increasing jet E_T. In the x_γ-distribution the two components are more clearly separated: for low values of $x_\gamma \lesssim 0.6$ the resolved component dominates, however at $x_\gamma > 0.6$ the cross-section of the direct component exceeds the resolved one. This is significantly lower than the value of $x_\gamma > 0.75$ that had been used as a cut at HERA [40,41] and is a consequence of the lower collision energy, decreasing the contribution of the resolved component relative to the direct cross-section.

Overall we see a large K factor in these two observables of about 60%. This can be attributed to the asymmetric jet cuts which leads to a large part of the phase space being left unfilled in the Leading Order calculation. By means of the real correction this part of the phase space is filled up, however, owing to the low E_T cut on the jets and the Born-level order of these corrections, the cross-section does not receive a significant reduction of its scale uncertainty bands.

Moving on to event shapes, we looked at similar variables as in [69], studying distributions of transverse thrust and transverse thrust minor in the upper panel of Fig. 3.18. In both observables, the direct and the resolved component contribute approximately equal amounts throughout the whole parameter space. This is in contrast to the C-parameter distribution, depicted in Fig. 3.18 (lower left panel), where we can identify distinct regions where one of the two components dominates. While the direct component contributes mostly at the lowest value and in the tail of the distribution, the resolved component contribution reaches up to about 60% near the global maximum of $C \approx 0.1$. Looking at the hadron multiplicity distribution in Fig. 3.18 (lower right), we see — in agreement with expectations — that the low-multiplicity states are dominantly produced by direct photoproduction, while the high-multiplicity tail is determined by the resolved component. It is worth noting here that the effect of multiple-parton interactions is generally mostly negligible at the EIC collision energies; it does, however, play a significant role in high-multiplicity states, therefore still necessitating a tuning of the modelling.

As another typical event shape observable we present jet rates in Fig. 3.19 with the anti-k_T algorithm which have previously been used for fits of the strong coupling. As a last pair of observables, we look at the ratio of c- and b-quark jets to light-quark jets in transverse momentum and pseudorapidity distributions in Fig. 3.20. Heavy-quark jets are almost

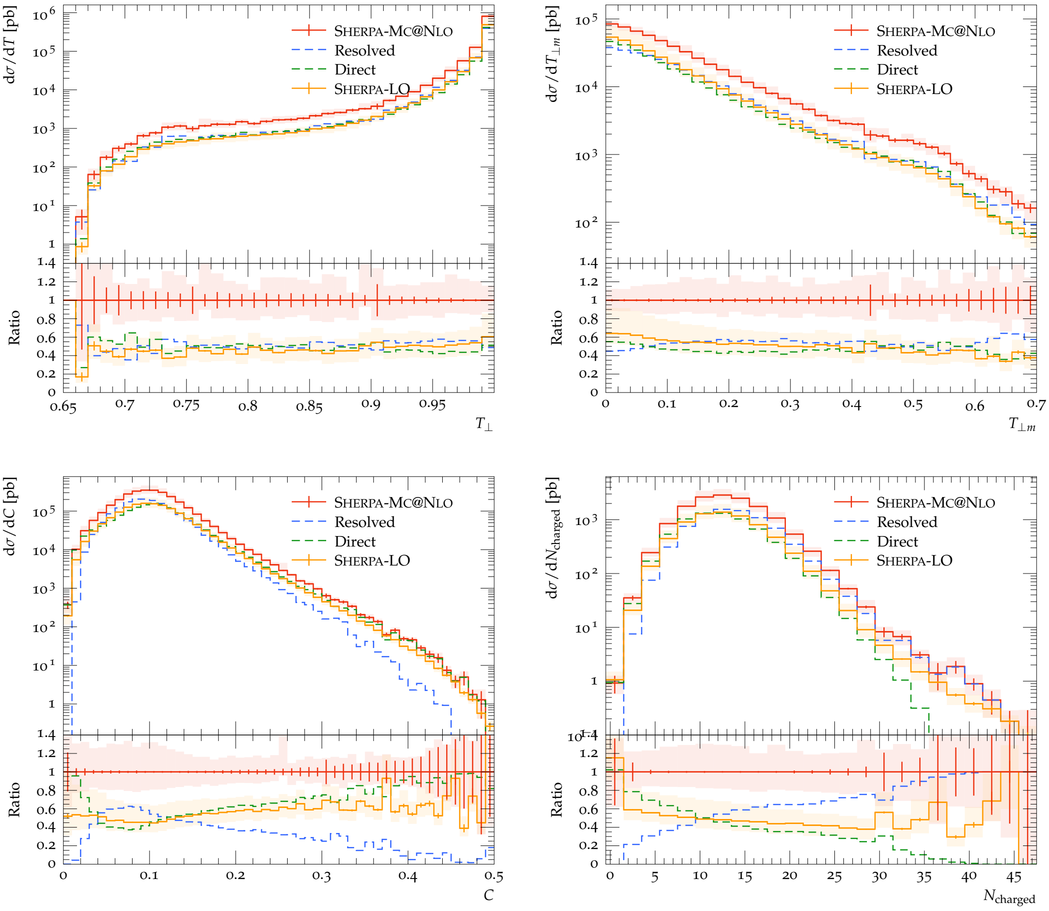


Figure 3.18: Event shapes observables: Transverse thrust major (upper left) and minor (upper right) distributions, C parameter (lower left) and charged particle multiplicity distribution N_{charged} inside the detector acceptance of $|\eta| < 4$ (lower right), all obtained with the SHERPA simulation at MC@NLO accuracy, compared with results at LO.

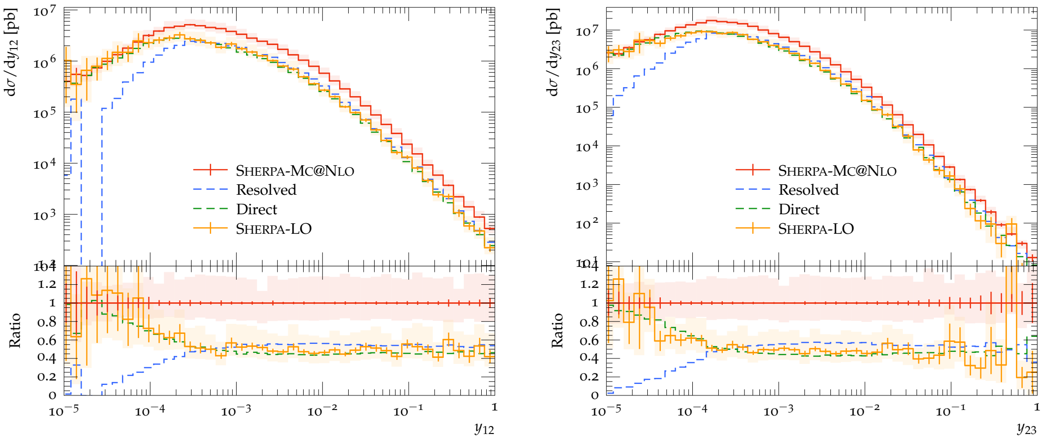


Figure 3.19: Distributions of 2 → 1 and 3 → 2 jet rates using the anti-k_T algorithm of the SHERPA simulation with MC@NLO accuracy, compared with results at LO.

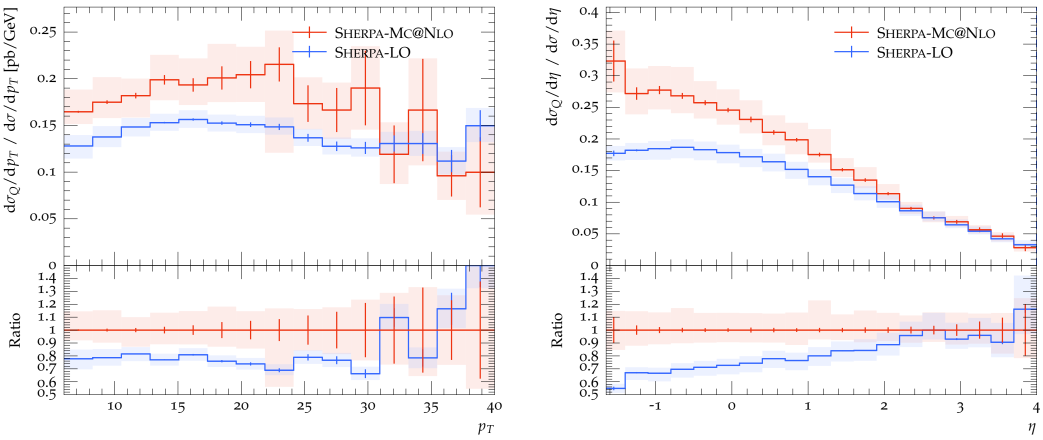


Figure 3.20: Distributions of heavy quark jet transverse momentum p_T (left) and pseudorapidity η (right) of the SHERPA simulation with MC@NLO accuracy, compared with results at LO.

exclusively produced through the direct process and in the region $\eta < 0$, where they make up almost a third of the overall activity. As expected, the heavy-quark jets are predominantly c-quark jets and their tagging will benefit from a cut on the x_γ value.

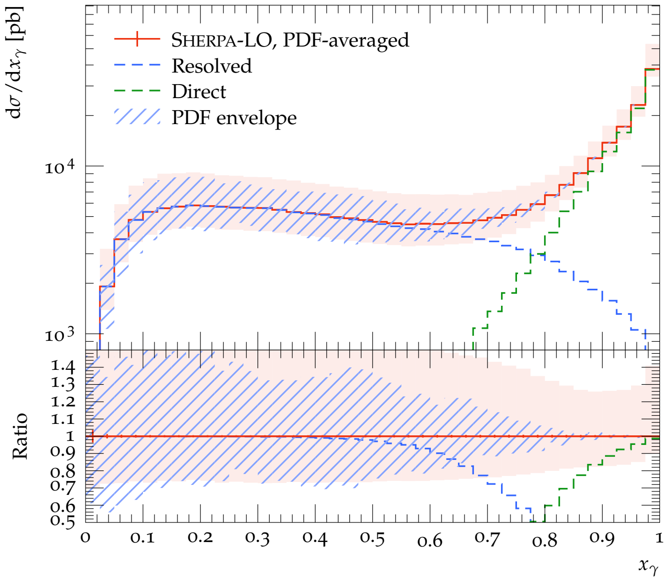


Figure 3.21: Distributions of x_γ of the SHERPA simulation at parton-level with LO accuracy and without MPI effects, averaged over 11 photon PDF sets. The PDF envelope shows the per-bin variation among the PDF sets.

As indicated above, the x_γ distribution is particularly important for tagging and distinguishing different event types, acting as an excellent discriminator of resolved and unresolved photon interactions. Given the relative cross-sections, it is also clearly important to improve the theoretical predictions obtained in resolved photoproduction. However, at the moment, this domain suffers from large uncertainties due to the somewhat outdated photon PDFs, obtained about 2 decades ago, and with proton PDFs from the same period for the description of photoproduction at HERA as we already discussed in Section 3.2.3. Similar to there, we estimate the uncertainty by averaging over 11 photon PDF sets each, including a 7-point scale variation, using the same settings as before in Fig. 3.21. Creating an envelope from the minimal/maximal values per bin among the different PDF sets, we find that the variations are of similar size as the scale uncertainties and reach up to 50% of the binned cross-section. This finding also suggests that a renewed fitting exercise of photon PDFs is of great importance for the upcoming EIC and the full exploitation and understanding of its data.

3.4 Summary

In this chapter we report on developments of the SHERPA event generator towards the first hadron-level simulation of photoproduction at NLO accuracy which have been published in the SHERPA 3.0 public release. We describe how the initial state phase space is treated to allow for a flexible yet efficient integration of different spectra and distributions, based on the equivalent photon approximation and the inclusion of a broad range of photon PDFs. We validate our model by comparison to data from the LEP and HERA experiments and find satisfactory agreement between the data and the simulation, noting that the non-perturbative aspects of the simulation, in particular intrinsic k_T and multiple-parton interactions, still require a comprehensive tuning to data. Apart from possible improvement in this sector of the simulation, our analysis of uncertainties underline the necessity for a refitting of the photon PDFs, especially in view of the anticipated precision of the planned electron-ion collider. We believe that this step – a first refitting of photon PDFs after 20 years – based on higher-order calculations, modern simulation tools at least at NLO accuracy, and recent proton PDFs will significantly improve the quality of our theoretical preparations for this new collider experiment

We also present the first fully differential hadron-level study at Next-to-Leading order accuracy for the photoproduction of inclusive QCD final states at the EIC. For our predictions, obtained with SHERPA, we assume beam energies of 18 (275) GeV for the electron (proton) beam and analyse the events with inclusive-jet as well as event-shape and heavy-jet observables. Photoproduction represents an interesting laboratory to study a large variety of sometimes subtle QCD effects, by providing access to a wide range of exclusive and inclusive production channels. Our results will also be instrumental for renewed efforts to fit PDFs for the photon in collinear factorisation.

We reiterate the importance of renewed analyses of the parton content of the photon for phenomenology at the EIC, arguing that an updated fit based on modern methodology including error estimates is urgently needed: a coherent and qualitatively satisfying theoretical description of resolved photoproduction necessitates a PDF fit based on modern proton PDFs, with consistent settings, for example for α_S.

Chapter 4

Comparison of photoproduction in event generators

In this chapter we present how photoproduction is simulated in the three main general-purpose event generators, scrutinising the implementations and comparing predictions.

4.1 Event Generators

To scrutinise the difference between different calculations of photoproduction, we focus first on the factorisation which for ep collisions is given by

$$\mathrm{d}\sigma(ep \to eX) = f_{\gamma}^{(e)}(x, Q^2) \otimes f_{i}^{(\gamma)}(x_{\gamma}, \mu_{\mathrm{F}}^2) \otimes f_{j}^{(p)}(x_{\mathrm{p}}, \mu_{\mathrm{F}}^2) \otimes \mathrm{d}\hat{\sigma}_{ij} \; , \tag{4.1.1}$$

where $f_{\gamma}^{(e)}(x, Q^2)$ is the photon flux from an electron at a given momentum fraction x and photon virtuality Q^2, $f_{i}^{(\gamma)}(x_{\gamma}, \mu_{\mathrm{F}}^2)$ the photon PDF which for direct interactions is a Delta distribution, and $\hat{\sigma}_{ij}$ perturbatively calculable hard coefficient function for a given hard process with initiating particles i and j. The photon flux can be computed using the equivalent photon approximation (EPA) [110]. For photons from a charged lepton l the leading-log (LL) approximation gives (c.f. Section 3.1.2)

$$f_{\gamma}^{(l)}(x) = \frac{\alpha_{\mathrm{em}}}{2\pi} \frac{1 + (1-x)^2}{x} \log\left[\frac{Q^2_{\mathrm{max}}}{Q^2_{\mathrm{min}}(x, m_l)}\right] . \tag{4.1.2}$$

The lower limit for the virtuality follows from kinematics and the upper limit is adjusted according to the experimental configuration.

In the case of $\gamma\gamma$ collisions in e^+e^- the corresponding factorisation reads (c.f. Section 3.1.1 and Fig. 3.1)

$$\sigma(ee \to eeX) = f_\gamma^{(e)}(x_1, Q_1^2) \otimes f_i^{(\gamma)}(x_\gamma^+, \mu_\mathrm{F}^2) \otimes f_\gamma^{(e)}(x_2, Q_2^2) \otimes f_j^{(\gamma)}(x_\gamma^-, \mu_\mathrm{F}^2) \otimes \mathrm{d}\hat{\sigma}_{ij} \,. \quad (4.1.3)$$

In this case there are four different contributions, direct-direct, direct-resolved, resolved-direct and resolved-resolved, that need to be accounted for. This factorised approach provides the starting point for all considered event generators over which further modelling of parton showers, beam remnants, MPI and hadronisation are included. Thus, even though the starting point is the same for all generators, the results will depend on the applied inputs and particular choices made in this modelling. Before comparing the different event generators to data, we will therefore describe the main features of the generators and the default inputs, and highlight some significant differences in the following subsections.

4.1.1 HERWIG

HERWIG [111–113] is a multi-purpose event generator, which has traditionally focused on accurate QCD simulations. It includes two parton shower modules, an angular ordered one based on coherent branching [114], and more recently a dipole parton shower [115, 116]. Both parton shower modules include full mass effects, spin correlations, and initial and final state radiation [117, 118].

Hard processes can be simulated using a wide range of built-in matrix elements at leading and next to leading order, and via the Matchbox module [112, 116] using external matrix element providers and an automated matching to NLO QCD either within the MC@NLO or POWHEG approaches. Multi-jet merging [119, 120] is available using the dipole shower. Photoproduction processes can be simulated in e^+e^-, ep and pp collisions including all direct and resolved components. The PDF sets and fluxes we currently provide are the SaS photon PDFs [121] with the 1D, 1M, 2D and 2M schemes, and photon fluxes following a Weizsäcker-Williams parametrization for electrons, and the Budnev approach [47] for protons. Hadronisation is modelled using the cluster hadronisation model [111], and we provide an eikonal MPI model, which currently however is limited to pp collisions and unfortunately cannot yet treat resolved photons due to technical issues. Similar restrictions currently apply to NLO matching using Matchbox, though both of these issues are currently being worked on and should be available with one of the next HERWIG releases.

To assess uncertainties, we can vary the scales of the hard process as well as the hard veto scales of the parton showers, which we perform in this study, additionally to varying the photon PDF sets. All of these variations turn out not to be significant, with the PDF set variation contributing the biggest uncertainty. The HERWIG parton showers always need to terminate on a valence quark of the incoming parton, and thus create a remnant which complements the parton which had been extracted. In case of the photon we will therefore always generate an anti-quark remnant, which might cause additional (collinear) shower activity even if radiation in the showers is switched off entirely.

4.1.2 PYTHIA

Simulations for photoproduction in ep and $\gamma\gamma$ collisions in $\mathrm{e}^+\mathrm{e}^-$ have been fully enabled from release 8.226 onwards and the current setup is described in the 8.3 manual [56]. Both hard and soft QCD processes can be generated, the former including processes like inclusive and diffractive jet production, and the latter non-diffractive and diffractive events and also elastic scattering. The implementation of resolved processes is based on the CJKL photon PDFs [49] which includes a hadron-like part with non-perturbative input separately from the perturbatively calculated anomalous contribution. In the generation of resolved processes these are not considered separately but the parton-shower algorithm includes the $\gamma \to q\bar{q}$ splitting that may collapse the resolved photon into an unresolved state at a perturbative scale corresponding to the anomalous part of the evolution equation.

The simulations include generation of MPI as long as the photons are in a resolved state and the MPI framework, where the QCD cross-sections are regulated with the screening parameter $p_{\mathrm{T},0}$, can be applied also for non-diffractive soft QCD processes without any phase-space cuts. The cross-section for the different soft processes are obtained from the SaS parametrizations in [121]. For proton PDFs the current default NNPDF23_lo_as_0130_qed has been applied and the value of α_S has been fixed accordingly to 0.130 at $\mu_{\mathrm{R}}^2 = M_{\mathrm{Z}}^2$. Here we have used PYTHIA version 8.310 released in July 2023.

In this study we have applied the default setup for photoproduction in PYTHIA 8.3. In the case of $\gamma\gamma$, a specific tune for $p_{\mathrm{T},0}$ has been applied based on single-inclusive charged-particle production data from OPAL at LEP [122]; in the case of γp, the $p_{\mathrm{T},0}^{\mathrm{ref}}$ in the standard parametrization has been adjusted to 3.0 GeV based on comparisons to single-inclusive charged particle production in photoproduction at HERA [123] which is in line also with the multiplicity distributions measured recently by ZEUS in [124]. Heavy quark masses are included apart from the $gQ \to gQ$ subprocess where massless expressions for the matrix element are applied also for charm and bottom quarks. The remnants are constructed by adding a minimal number of partons such that momentum and colour are conserved. In the case of resolved photons the flavour for a "valence" quark-antiquark pair is sampled according to relative fractions of the PDFs at the specific kinematics and part of the phase-space is cut out to allow room for the remnant partons. The scale variations have been performed by varying renormalisation and factorisation scales independently in the hard process generation by a factor of two, excluding the variations where the relative difference would be 4.

4.1.3 SHERPA

In the SHERPA event generator [2, 3], photoproduction of jets at MC@NLO accuracy has been recently implemented and validated against data [90] and has been published under version 3.0. Resolved processes can be calculated through interfaces to various different photon PDFs, where the default is the SaS PDF library [121].

We generated events at Leading Order and MC@NLO accuracy as described in [90]. We used AMEGIC [96] and COMIX [87] for tree-level matrix elements and subtraction terms [97], and OPENLOOPS [98] for one-loop matrix elements. For the calculation of the matrix element, we treated the b-quarks for ZEUS, and additionally the c-quark for OPAL and EIC runs, as massive and included them in the final state. While at LO all subprocesses are available, the initial state subtraction for massive quarks has not been implemented yet. The CSSHOWER parton shower [99] was used, combined with the MC@NLO method [100, 101] as implemented in SHERPA [102].

Multiple-parton interactions were modelled through an implementation of the Sjostrand-van Zijl model [103, 104] in which the total hadronic cross-section is calculated using Regge theory and the parametrizations for the photon are obtained as a superposition of light neutral vector mesons as proposed and parametrised in [104]. The tuning of the rewritten MPI modelling is currently work-in-progress as part of the new release. Particles were hadronised through the cluster fragmentation model in AHADIC [105] and this has already been tuned against data. The photon flux was modelled as computed in [92], which includes a correction for $x \to 1$ relevant for lepton-hadron colliders. For the PDFs, we used the SAS2M [52, 53] set for the photon and the PDF4LHC21_40_pdfas [108] set for the proton, and the value of $\alpha_S(M_Z)$ was kept at its default value of 0.118 with three-loop running. Factorisation and renormalisation scales were set to $\mu_\mathrm{F} = \mu_\mathrm{R} = H_\mathrm{T}/2$ and a 7-point scale variation in the matrix element and the shower was done as an uncertainty estimation.

4.1.4 Differences between generators

As a baseline for a broader comparison between the generators, we compared events of LEP-like setups with beam energies of 99 GeV from PYTHIA and SHERPA at the bare cross-section-level for each of the components. Due to technical limitations at the interplay between beam remnants, parton shower and hadronisation, this bare parton-level simulation was not possible within HERWIG.

As a first step we used the same photon flux, PDF, setting to $\alpha_S(M_Z) = 0.130$ with 1-loop running and computed only processes including light partons. In Fig. 4.1 we can see that indeed the results are, up to statistical fluctuations, in a perfect agreement. The considered observables include the energy and pseudorapidity distribution, and transverse momentum spectra of the outgoing partons for direct, single-resolved and double resolved contributions separately. In the case of the single-resolved contribution the direct photon has a positive p_z.

Based on this, we changed the beam configuration in the previous setup to an EIC-like setup with electron-proton beams with 18 and 275 GeV beam energies, respectively, and looked at photoproduction events with only resolved photons in Fig. 4.2. We used the same settings as before, however, now also modelling the initial state radiation and the beam remnants, including charm and bottom quarks in the processes and using the NNPDF23_lo_as_0130_qed PDF for the proton. The inclusion of beam remnants necessitates initial state radiation for the evolution down to small scales, hence these two cannot be further decomposed. The total cross-section in PYTHIA is larger than in SHERPA due to the way the remnants are created: in PYTHIA, forced splittings are combined with phase space rejection to ensure the correct creation of the valence quark content of the beam particle, which allows for an iterative procedure. In SHERPA, the remnant creation is purely based on phase space rejection which naturally will lead to more events being rejected, especially for charm and bottom quarks. In the subsequent modelling of the beam breakup, we can see that SHERPA creates more activity in the low-E_T and low-p_T region, whereas PYTHIA tends to have higher transverse momenta and energies in its particle spectrum.

In Table 4.1, we summarise the differences between the generators with respect to the simulation of photoproduction. Other differences are for example the fragmentation

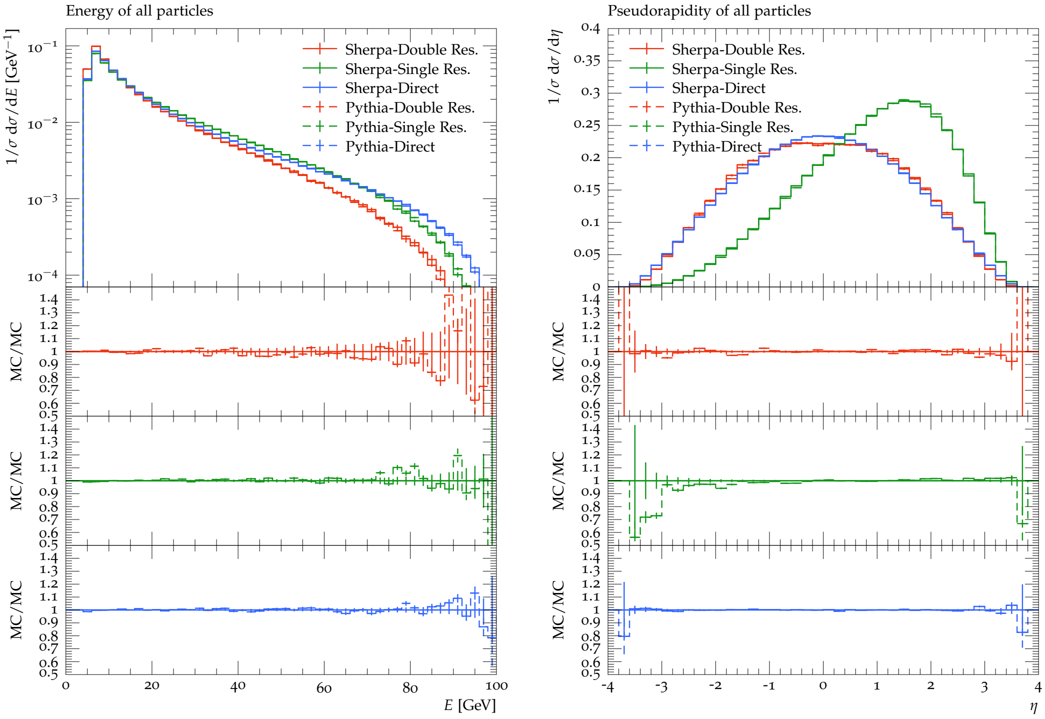


Figure 4.1: Comparison of fixed-order parton level events between PYTHIA and SHERPA for a LEP-like setup.

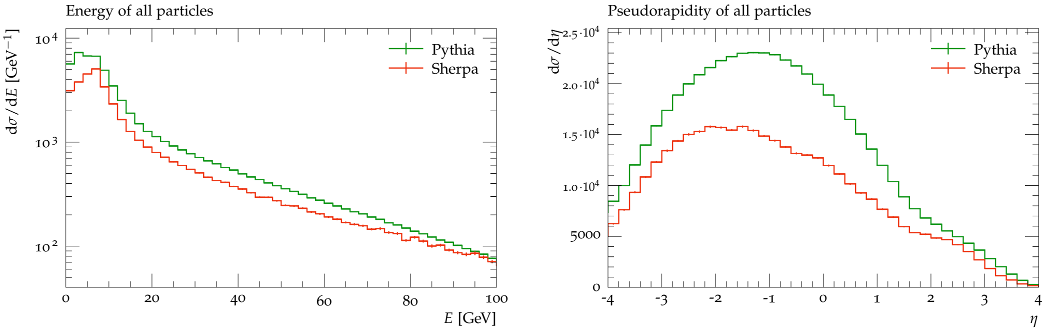


Figure 4.2: Comparison of fixed-order parton level events including ISR parton shower and beam remnants between PYTHIA and SHERPA for an EIC-like setup.

model and QED corrections, however, these will not be discussed here and can be found in the manuals [56,93,111].

Table 4.1: Differences between the generators of default settings, specifically for photoproduction.

Property	PYTHIA	SHERPA	HERWIG
Flux	LL	NLL	LL
$\alpha_S(M_Z^2)$	0.130, 1-loop	0.118, 3-loop	0.118, 2-loop
PDFs	CJKL	SAS2M	SAS2M
Remnants	forced splittings and PS rejection	PS rejection	forced splitting
Photon Splitting	yes	no	no
MPI tuning	preliminary tune	untuned	untuned

In terms of the perturbative part, the event generators model the flux to different accuracies where the difference is given by

$$f_{\mathrm{NLL}}(x) - f_{\mathrm{LL}}(x) = -\frac{\alpha_{\mathrm{em}}}{\pi} m_e^2 x \left(\frac{1}{Q^2_{\mathrm{min}}} - \frac{1}{Q^2_{\mathrm{max}}} \right) . \tag{4.1.4}$$

The correction comes with an overall negative sign and hence will lead to smaller total cross-sections. In differential observables, the correction will however only be sizeable in regions where $x \to 1$, c.f. Section 3.1.2 and Fig. 3.2 therein. Even though not specific to photoproduction we point out that the settings of α_S are different: while in SHERPA the current default is set according to the default proton PDF set and is close to the PDG world average [125], in PYTHIA it is used as a dynamical K factor, also in accordance with its default proton PDF set. Photon PDFs do not allow the modern standard procedure of setting α_S in accordance with the fit as this information is often not given in the corresponding publication and would in any case be in conflict with settings of modern proton PDFs. Hence, strictly speaking the factorisation of the cross-section is not fully consistent due to the different values of α_S used throughout event generation.

As a last point, we plot the parton distributions of the light partons for the two PDF libraries SAS2M and CJKL in Fig. 4.3. The two fits come to vastly different behaviour at small x, where the SAS2M sets reach an almost constant value and the CJKL sets behave like a power-law, with the most pronounced difference in the gluon distribution.

It is also worth pointing out that unlike the settings for α_S and the flux, which cover the perturbative region and are independent of any fitting, the remnant creation and the MPI

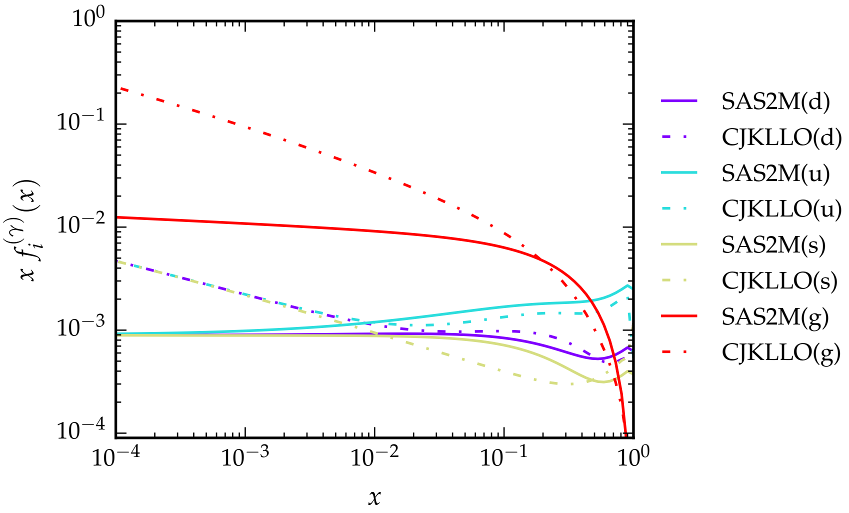


Figure 4.3: Comparison of the CJKLLO and SAS2M PDF sets for light quarks and the gluon at $\mu_{\mathrm{F}}^2 = 5\ \mathrm{GeV}^2$.

interplay with non-perturbative effects and rely on precise data by experiments like HERA and LEP. We also point out that all differences of PYTHIA with respect to SHERPA lead to an increase in the cross-section which will be a common observation in the following section.

4.2 Comparisons to LEP

For the comparison to LEP data, we used a dijet measurement from the OPAL collaboration [34]. At lepton colliders, the cross-section for photoproduction can be separated into $\sigma = \sigma_{\gamma\gamma\to X} + \sigma_{\gamma j\to X} + \sigma_{j\gamma\to X} + \sigma_{jj\to X}$, where j denotes a photon being resolved into partons. The analysis used data taken at $\sqrt{s} = 198$ GeV and clustered jets with the k_T algorithm with $R = 1$ demanding $E_T > 3$ GeV and $|\eta| < 2$ with at least two jets. To separate resolved from direct photoproduction processes, the analysis defined

$$x_\gamma^\pm = \frac{\sum_{j=1,2} E^{(j)} \pm p_z^{(j)}}{\sum_{i\in\mathrm{hfs}} E^{(i)} \pm p_z^{(i)}} \tag{4.2.1}$$

and associated values $x_\gamma^\pm < 0.75$ with resolved processes. For a parton-level $2 \to 2$ scattering these definitions would match with the momentum fractions in photons going positive and negative p_z in Eq. (4.1.3) but adding parton showers, MPI and hadronisation will smear the kinematics such that the division will be only approximative. In the numerator the sum runs over the two jets with the highest transverse momentum, p_T, and in the denominator over the complete hadronic final state. In Fig. 4.4 we present the average transverse energy of the dijets for different bins in x_γ. The LO prediction from SHERPA undershoots the total cross-section, however, the effect is more pronounced for resolved photons than for unresolved ones. Going to NLO accuracy, the simulation describes the data well within the errors. The large K factors hint at the real corrections and the filled-up phase space as the drivers of the improved description. PYTHIA tends to overshoot the data for regions dominated by the direct contribution but agrees well with the data for the resolved-resolved case. The HERWIG simulation typically lies between the SHERPA and PYTHIA results. In all cases the data is within the large uncertainties from scale variations. As discussed in Section 4.1.4, the large difference between the LO results from HERWIG, PYTHIA and SHERPA builds up from varying inputs.

The distributions in pseudorapidity η in Fig. 4.5 show a similar picture, however for double-resolved processes, i.e. $x_\gamma^\pm < 0.75$, all predictions have a slightly steeper fall-off as a function of η than seen in the data, leaning towards undershooting the forward region. One potential reason could be the weakly constrained gluon-content of the photon, which should be the leading contribution at low-E_T and forward jets.

Fig. 4.6 shows distributions of x_γ for low and high average jet transverse energies $\bar{E}_T$, respectively. We see good agreement for both the PYTHIA and the SHERPA-MC@NLO simulations; however, the transition from the resolved to the direct processes seems to be poorly modelled, as can be seen by the consistent undershoot at around $x_\gamma \approx 0.8$ for all predictions. Unlike HERWIG and SHERPA, the PYTHIA simulation does include the correct evolution of the photon PDF, i.e. the photon splitting $\gamma \to q\bar{q}$ is taken into account. The distribution still shows this shape in that case too, though. Potential reasons could be the poor quality of the photon PDF or insufficient tuning of the fragmentation and multiple-parton interactions modelling. Combined with the overshoot in the largest-x bin, we would expect these effects to increase the multiplicity of direct processes, hence shifting cross-section towards lower values of x.

4.3 Comparisons to HERA

For this comparison we used data taken by the ZEUS collaboration [40] during HERA Run 1. Here, the photoproduction cross-section can be decomposed into two parts, $\sigma = \sigma_{\gamma j} + \sigma_{jj}$, where again j denotes a parton resolved from within the photon or the proton. Similar to the previous analysis, jets were clustered with the k_T algorithm with $R = 1$ with

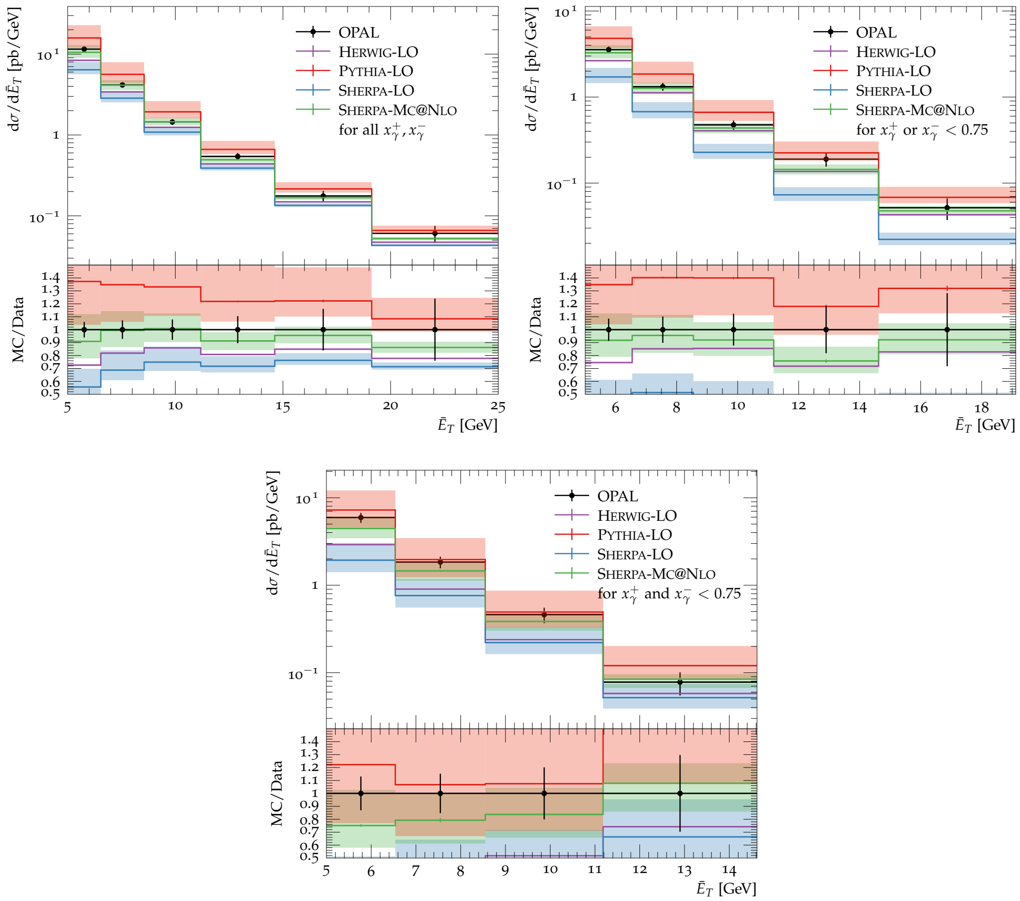


Figure 4.4: Distribution of average transverse energy of dijets, $\bar{E}_T$, from OPAL [34] for all $x_\gamma^\pm$ (top left), x_γ^+ or $x_\gamma^- < 0.75$ (top right) and $x_\gamma^\pm < 0.75$ (bottom), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

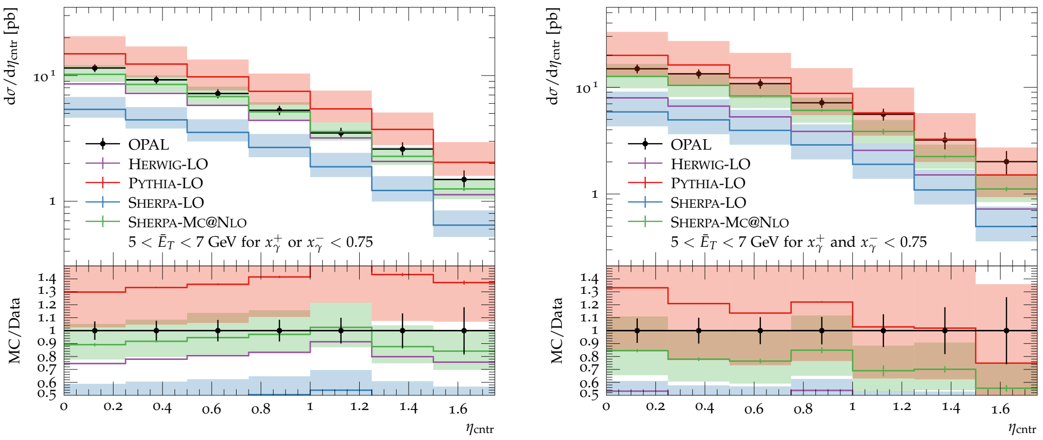


Figure 4.5: Distribution of pseudorapidity η of dijets from OPAL [34] for x_γ^+ or $x_\gamma^- < 0.75$ (left) and $x_\gamma^\pm < 0.75$ (right), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

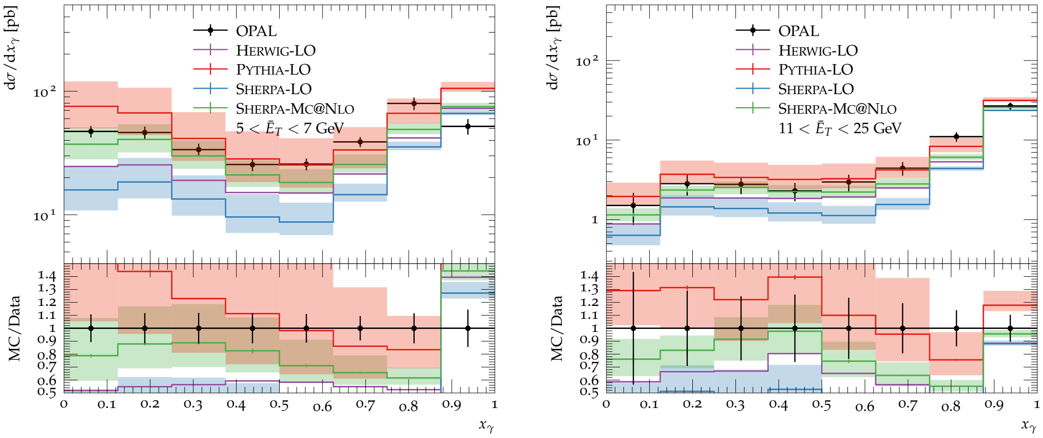


Figure 4.6: Distribution of x_γ of dijets from OPAL [34] for bins of average jet transverse energy $\bar{E}_T \in [5\,\text{GeV}, 7\,\text{GeV}]$ (left) and $\bar{E}_T \in [11\,\text{GeV}, 25\,\text{GeV}]$ (right), compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

acceptance cuts requiring $\eta \in [-1, 2.4]$ and $E_T > 14$ (11) GeV for the (sub-)leading jet, respectively. The analysis used a similar discriminant for the tagging of direct and resolved processes, defined as

$$x_\gamma^{\text{obs}} = \frac{\sum_{j=1,2} E_{\text{T}}^{(j)} \mathrm{e}^{-\eta^{(j)}}}{2yE_e} \,, \tag{4.3.1}$$

where the sum runs over the two highest E_{T} jets, E_e is the energy of the incoming electron and y the inelasticity.

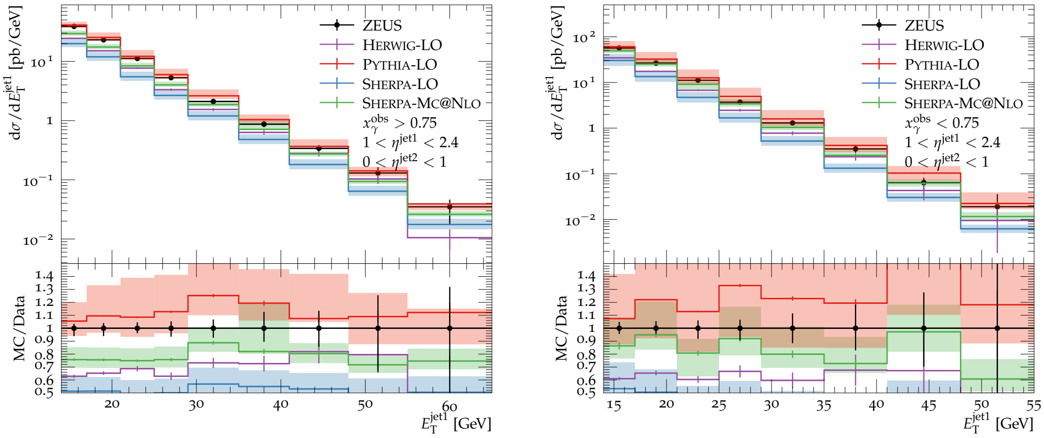


Figure 4.7: Distributions of leading jet transverse energy E_T^{jet1} for $x_\gamma > 0.75$ (left) and $x_\gamma < 0.75$ (right) with jet pseudorapidities for the two leading jets in $1 < \eta^{\text{jet1}} < 2.4$ and $0 < \eta^{\text{jet2}} < 1$ from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

In Fig. 4.7 we show the leading jet transverse energy E_T^{jet1} for direct and resolved processes. While the SHERPA-LO prediction stays below the data, the PYTHIA predictions agree within the error bars; the SHERPA-MC@NLO predictions describe the data well with the exception of the low-E_T phase space in the direct process, where it undershoots by about 20%. As the cuts on the pseudorapidity select the forward region, this observable is probably sensitive to additional radiation from underlying events, the used PDFs, and to the photon splitting in the parton shower. Fig. 4.8 shows a similar situation for the η-dependence, where for the resolved processes both PYTHIA and SHERPA-MC@NLO, agree with data but for direct processes SHERPA undershoots slightly.

We finish this section with the distributions in x_γ for low and high leading jet transverse

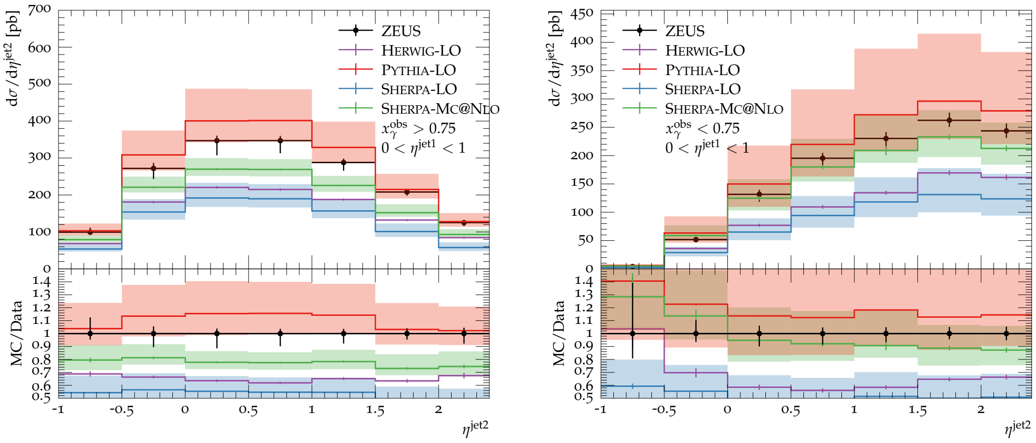


Figure 4.8: Distributions of sub-leading jet pseudorapidity η^{jet2} for $x_\gamma > 0.75$ (left) and $x_\gamma < 0.75$ (right) with leading jet pseudorapidity in $0 < \eta^{\text{jet1}} < 1$ from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

energy in Fig. 4.9. Unlike the modelling for LEP, there is no undershoot visible at the transition from direct to resolved processes. SHERPA undershoots the data in the region where both x_γ and E_T are small, however this can be attributed to the missing tuning of the MPI as the same region is fairly well described by PYTHIA. HERWIG is again compatible with the leading order simulation from SHERPA.

4.4 Predictions for EIC

For the planned EIC we present predictions for electron-proton beams with 18 and 275 GeV beam energies, respectively, similar to the study in [126]. We cluster jets with the anti-k_T algorithm with $R = 1.0$ and demand at least one jet with $E_\text{T} > 6$ GeV. Looking at inclusive (di-)jet observables in Fig. 4.10, we see a similar behaviour in the comparison between the generators, where SHERPA-LO yields the smallest cross-section and PYTHIA-LO the largest, while HERWIG delivers a significantly different shape of the x_γ distribution. The K factor in these observables is roughly 50%, again hinting at the real correction and the phase space driving the correction at NLO. The PYTHIA LO prediction deviates another 50% from the NLO-accurate prediction, which can partially be explained by the different PDF sets, α_S values and the other differences as pointed out in Section 4.1.4. This means that, going towards the highest possible precision, the perturbative accuracy needs to

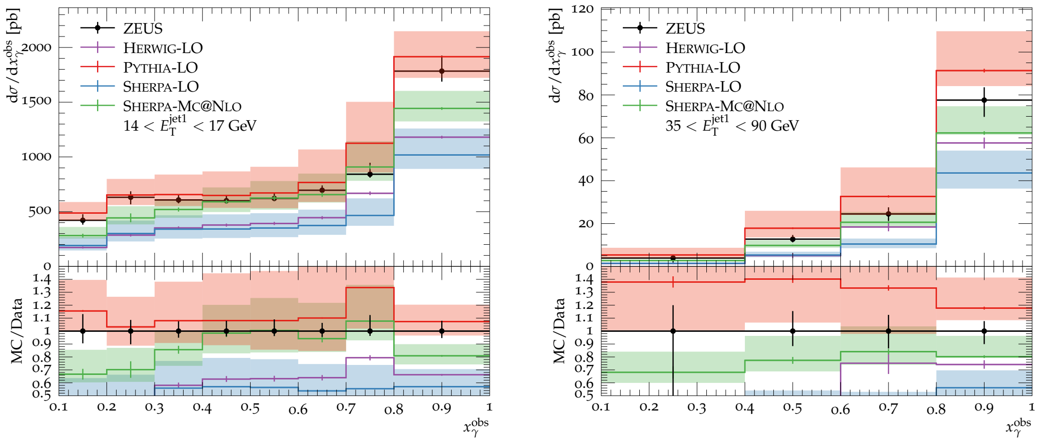


Figure 4.9: Distributions of x_γ for leading jet transverse energies in $14 < E_T^{\mathrm{jet1}} < 17$ GeV (left) and $25 < E_T^{\mathrm{jet1}} < 90$ GeV (right) from ZEUS [40], compared to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

be improved further and the non-perturbative effects need to be constrained by data. Publicly available RIVET analyses for the relevant data from HERA and LEP are crucial to reach the latter goal. While there has been some recent progress on porting the existing analyses to the RIVET framework, there still are some shortages related to MPI constraints and virtuality modelling.

To study the event shapes in more detail, we present predictions for transverse thrust $T_\perp$ and transverse sphericity $S_\perp$, defined as

$$T_\perp = \max_{\vec{n}_\mathrm{T}} \frac{\sum_i |\vec{p}_{\mathrm{T},i} \cdot \vec{n}_\mathrm{T}|}{\sum_i \vec{p}_{\mathrm{T},i}} \quad \text{and} \tag{4.4.1}$$

$$S_\perp = \frac{2\lambda_2}{\lambda_1 + \lambda_2} \tag{4.4.2}$$

in Fig. 4.11. Here n_T is the transverse-thrust axis that maximizes the quantity and $\lambda_{1,2}$ are the eigenvalues of the transverse linearised sphericity tensor $\mathbf{S}_{\alpha\beta}$ defined as

$$\mathbf{S}_{\alpha\beta} = \frac{1}{\sum_i |\vec{p}_{\mathrm{T},i}|} \sum_i \frac{1}{|\vec{p}_{\mathrm{T},i}|} \begin{pmatrix} p_{i,x}^2 & p_{i,x}p_{i,y} \\ p_{i,y}p_{i,x} & p_{i,y}^2 \end{pmatrix} \tag{4.4.3}$$

with i summing over the momenta in the final state and α, β over $1, 2$. In both observables,

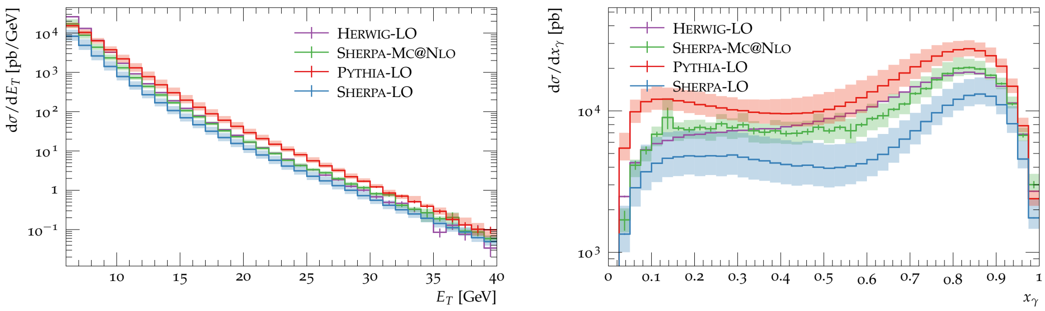


Figure 4.10: Predictions of transverse jet energy E_T (left) and x_γ (right) for jet production at the Eic, comparing to Leading Order simulations by Herwig, Pythia and Sherpa and Mc@Nlo-accurate simulations by Sherpa.

Pythia predicts slightly more isotropic events than Sherpa, which might again be caused by a larger number of MPI modelled within Pythia. Comparing the two Sherpa predictions, we again observe a sizeable K factor and a shift towards more isotropic events. While using a higher value for $\alpha_S(M_Z)$ in Pythia as a proxy for the K factor seems to work quite well and gives similar cross-sections, the uncertainties from scale variations are considerably smaller at NLO.

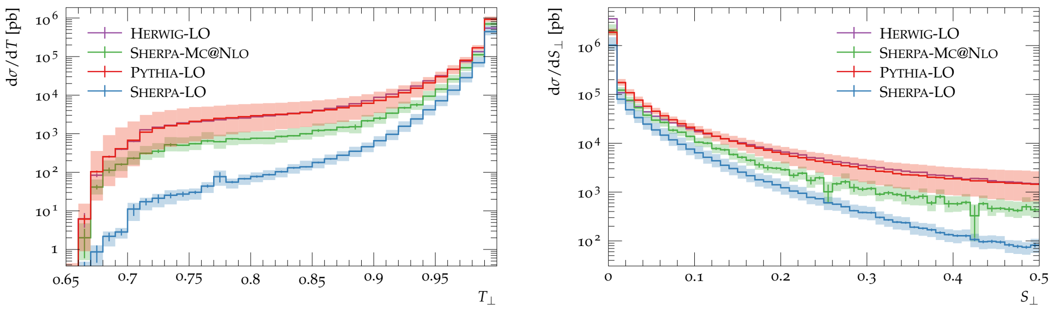


Figure 4.11: Predictions of transverse thrust $T_\perp$ (left) and transverse sphericity $S_\perp$ (right), for jet production at the Eic, comparing to Leading Order simulations by Herwig, Pythia and Sherpa and Mc@Nlo-accurate simulations by Sherpa.

As a last observable, in Fig. 4.12 we look at the charged particle multiplicity in the detector acceptance range $|\eta| < 4$ again for events that contain at least one jet with $E_\mathrm{T} > 6$ GeV and see a large disagreement between Pythia and Sherpa. Even though MPI do not play a huge role when studying high-p_T observables, such as jets, they do come into play when studying event structure in more detail. The result shows fewer hadrons being generated in SHERPA than in PYTHIA or HERWIG, and it means that a careful study of MPI and hadronisation is necessary to correctly simulate these observables. As discussed before, while the perturbative accuracy is under good control, corrections due to non-perturbative effects rely on data being made available.

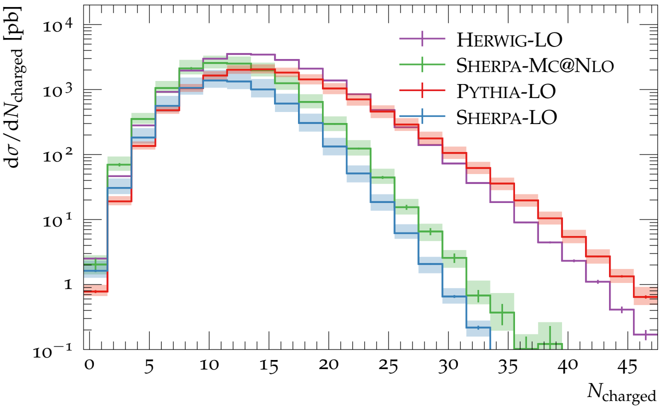


Figure 4.12: Predictions of charged particle multiplicity N_{charged} for events with a jet with $E_{\mathrm{T}} < 6$ GeV at the EIC, comparing to Leading Order simulations by HERWIG, PYTHIA and SHERPA and MC@NLO-accurate simulations by SHERPA.

4.5 Summary

We present a comparison of the three general purpose event generators, HERWIG, PYTHIA, and SHERPA, and contrast the results against experimental data from LEP and HERA colliders. While the starting point and theoretical ingredients are similar for each generator, the default inputs and differences in phenomenological modelling do result in significant differences for the considered observables. In particular, we compare to data for dijet photoproduction analysed by the OPAL and ZEUS collaborations, seeing an overall good agreement with the data for the SHERPA-MC@NLO and the PYTHIA simulations. At leading order, to which the HERWIG and PYTHIA simulations are currently limited, and consistent with SHERPA at leading order, shapes are roughly consistent with the data, however normalizations lack a significant K-factor for HERWIG while PYTHIA compensates with setting a higher value for α_S. Furthermore, we present predictions for the upcoming EIC for inclusive QCD observables and event shapes for events containing at least one high-p_{T} jet.

We find significant differences in observables sensitive to the underlying event modelling and conclude that having more experimental data in the RIVET framework would allow for further tuning of the parameters related to MPI generation such that discrepancies could be largely resolved.

Being the dominant production mechanism for hadronic final states, more work is needed in preparation for precision phenomenology for the EIC to understand the different regimes and a coherent modelling of these, including development of the relevant infrastructure in all general purpose event generators. Open questions remain such as the transition region between DIS and photoproduction at virtualities of $Q^2 \approx 1\ \mathrm{GeV}^2$ or the direct and resolved regimes of the photon.

Chapter 5

Diffraction

Building on the more general introduction to diffraction in Section 2.2.4, we introduce in this chapter the theoretical description of diffraction as well as different measurements at HERA with which we validate the implementation. We also present a phenomenological study for diffraction at the EIC.

5.1 Theory framework for jet production in diffraction

Diffractive events are characterised by the presence of an intact beam proton, or a low-mass excitation thereof in the final state. In [71], factorisation has been shown to hold for these type of processes with the introduction of so-called Diffractive PDFs (DPDF), $f_i^D(x, \mu_F, x_{I\!P}, t)$. In this factorised approach, the cross-section for the process $ep \to eXY$, where Y denotes the elastically scattered proton or its excitation, can be calculated through

$$\mathrm{d}\sigma_{ep\to eXY}(x_{I\!P}, t) = \sum_i f_i^D(x, \mu_F, x_{I\!P}, t) \otimes \hat{\sigma}_{ei\to eX}(x, \mu_F)\ . \tag{5.1.1}$$

The kinematic variables, are defined as momentum transfer $t = (p_p - p_{p'})^2$ and light-cone momentum fractions $x_{I\!P} = p_{I\!P}^+/p_p^+$ and $x = p_i^+/p_{I\!P}^+$ [127], c.f. Fig. 5.1 for a sketch of the kinematics. Due to the colour-disconnection of the final state systems X and Y, the events typically show a large rapidity gap (LRG). The tagging of such events can be experimentally achieved either through dedicated detection of the system Y or by vetoing activity in certain rapidity regions, hence selecting events with LRGs.

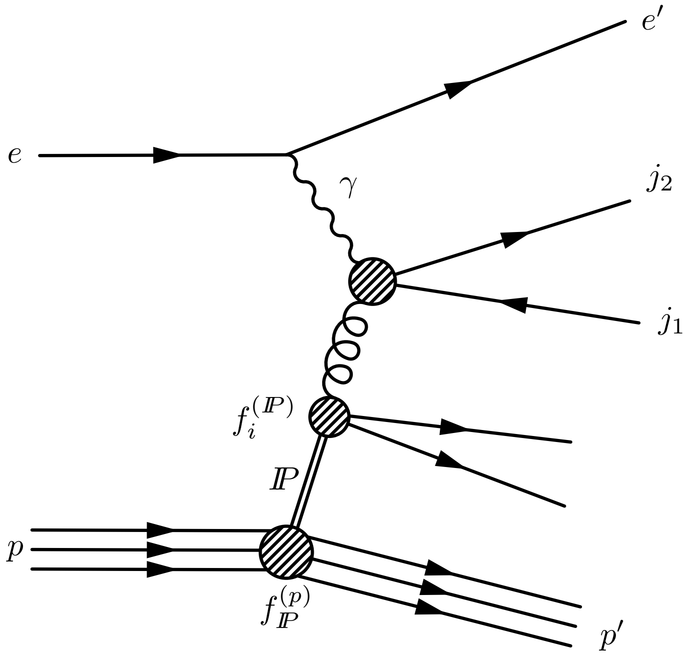

e'
e
γ
j_2
j_1
$f_i^{(I\!P)}$
$I\!P$
p
$f_{I\!P}^{(p)}$
p'

Figure 5.1: Sketch of an diffractive DIS event for the production of two jets, j_1 and j_2, together with the remnants forming system X and p' forming the system Y.

Such event topologies are usually attributed to the exchange of pomeron (and subleadingly a reggeon); motivated by this physics model the factorisation picture has been extended [75], to separate the DPDF further into products of flux factors and PDFs:

$$f_i^D\left(x,\mu_F,x_{I\!P},t\right)=f_{I\!P}^{(p)}\left(x_{I\!P},t\right)f_i^{(I\!P)}\left(x,\mu_F\right)+n_{I\!R}f_{I\!R}^{(p)}\left(x_{I\!P},t\right)f_i^{(I\!R)}\left(x,\mu_F\right) \tag{5.1.2}$$

Inspired by Regge theory, the flux factors assume the form

$$f_{I\!P,I\!R}^{(p)}\left(x_{I\!P},t\right)=A_{I\!P,I\!R}\frac{e^{B_{I\!P,I\!R}t}}{x_{I\!P,I\!R}^{2\alpha_{I\!P,I\!R}(t)-1}} \tag{5.1.3}$$

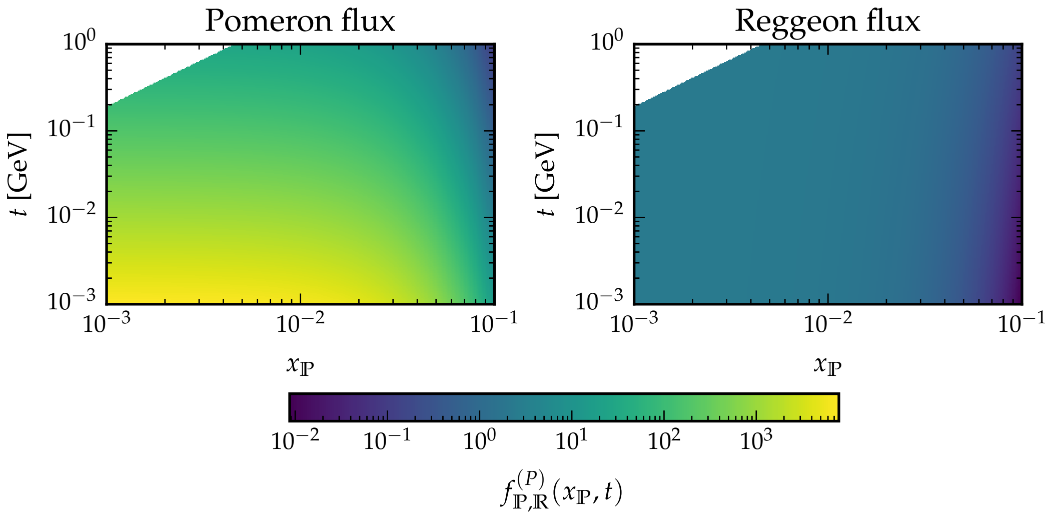


Figure 5.2: Pomeron (left) and reggeon (right) fluxes as functions of $x_{I\!P}$ and t.

with the parameters describing the trajectories $\alpha_{I\!P,I\!R}(t)=\alpha_{I\!P,I\!R}(0)+\alpha'_{I\!P,I\!R}\,t$ fitted to data. In a naive picture, the pomeron can be interpreted as a colour-neutral two-gluon state and the reggeon as a colour-neutral quark–anti-quark state. In Fig. 5.2 we plot the fluxes double-differentially in $x_{I\!P}$ and t. Integrating over t in the interval $[t_{\text{min}},\ t_{\text{cut}}]$ yields

$$f_{I\!P,I\!R}^{(p)}\left(x_{I\!P}\right)=\frac{A_{I\!P,I\!R}x_{I\!P}^{1.-2.\alpha_{I\!P,I\!R}(0)}}{B_{I\!P,I\!R}-2.\alpha'_{I\!P,I\!R}\log(x_{I\!P})}\left(e^{B_{I\!P,I\!R}t_{\text{min}}}x^{-2.\alpha'_{I\!P,I\!R}t_{\text{min}}}-e^{B_{I\!P,I\!R}t_{\text{cut}}}x^{-2.\alpha'_{I\!P,I\!R}t_{\text{cut}}}\right)\ . \tag{5.1.4}$$

The integration boundaries are given by $t_{\text{min}}=-\frac{m_p^2x_{I\!P}^2}{1-x_{I\!P}}$, with m_p the proton mass, and t_{cut}

the experimental boundary of the momentum transfer of the proton. We plot the corresponding integrated fluxes in Fig. 5.3, which shows that indeed the dominant contribution in the relevant phase space is the pomeron, exceeding the reggeon flux by up to almost four orders of magnitude. Furthermore, the pomeron flux can be approximated quite well by a function proportional to $x_{I\!P}^{-1.222}$.

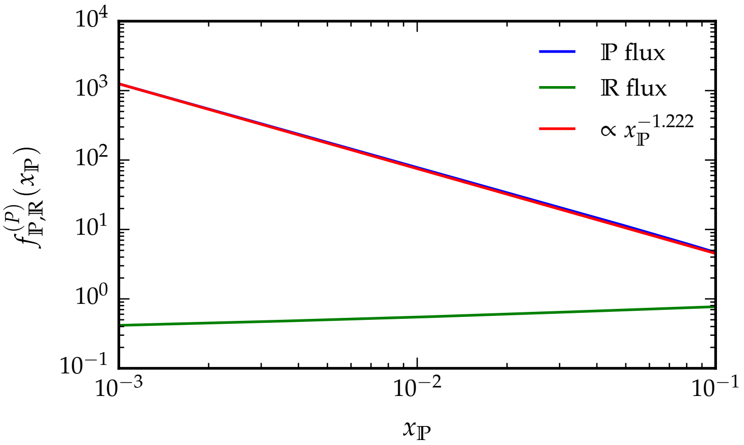


Figure 5.3: Integrated pomeron, $I\!P$, and reggeon, $I\!R$, fluxes with $t_{\text{cut}} = -1.6\,\text{GeV}^2$ as a function of $x_{I\!P,I\!R}$.

While the PDF for the pomeron $f_i^{(I\!P)}(x, \mu_F)$ must be fitted to data, the reggeon PDF is usually approximated by the pion PDF, $f_i^{(I\!R)}(x, \mu_F) \approx f_i^{(\pi^0)}(x, \mu_F)$ [128]. In Fig. 5.4 we show the PDFs used for our study for the pomeron and the reggeon; the former clearly shows a dominant gluon contribution up to high values of x while the quark contributions are about an order of magnitude smaller. In both fits, the same distributions describe the u- and the d-quark content. To quantify the dominating contribution, we convolve the pomeron flux with its gluon PDF and plot it double-differentially in x and $x_{I\!P}$ in Fig. 5.5. We see that the product depends more strongly on $x_{I\!P}$ than on x.

Depending on the virtuality of the exchanged photon, events can be further differentiated into the deep-inelastic scattering (DIS) or the photoproduction regime. In the latter case, the factorisation includes a photon flux and, in the resolved case, a photon PDF. The overall phase space setup and the implementation of the photoproduction events follows Section 3.1.

This results in the following factorisation formula for the cross-section for diffractive DIS

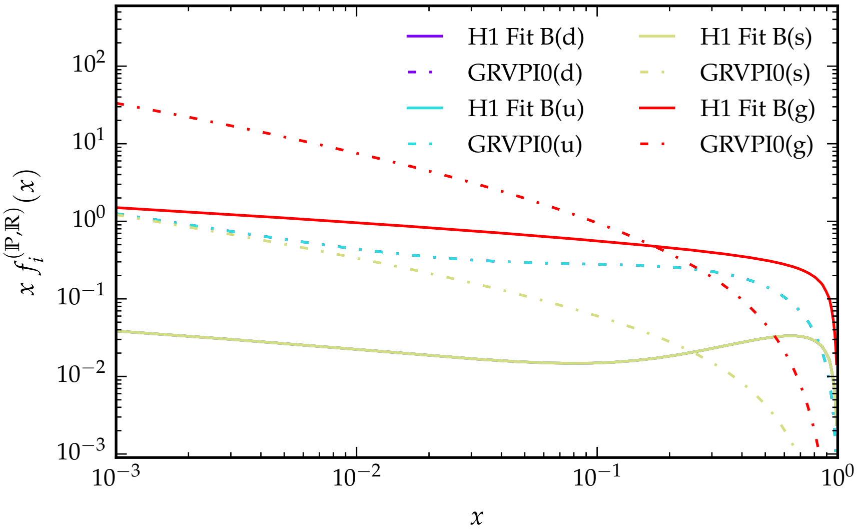


Figure 5.4: The H1 Fit B pomeron PDF and the GRVPI0 used for the reggeon at $\mu_F^2 = 10\ \mathrm{GeV}^2$.

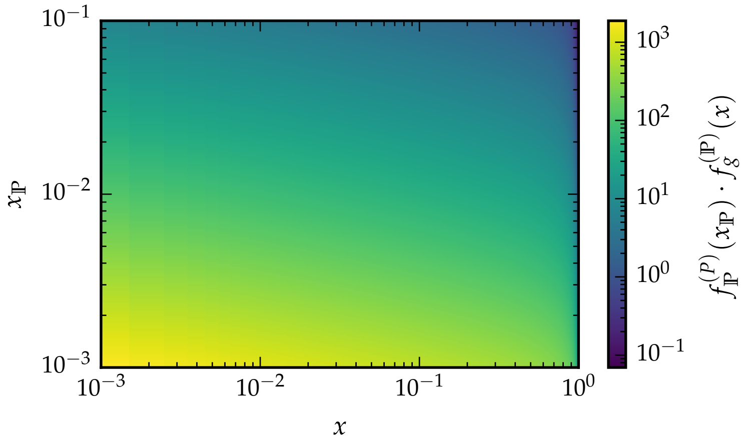


Figure 5.5: Convolution of the pomeron flux with the dominant gluon PDF from H1 Fit B at $\mu_F^2 = 10\ \mathrm{GeV}^2$.

$$\sigma^{(\mathrm{DDIS})}\left(ep \to eXY\right) = \int_0^{x_{I\!P,\max}} \mathrm{d}x_{I\!P} \int_{t_{\mathrm{cut}}}^{t_{\min}} \mathrm{d}t \int_0^1 \mathrm{d}x_i f_i^D\left(x_i, \mu_F, x_{I\!P}, t\right)\, \hat{\sigma}\left(ei \to eXY\right) \tag{5.1.5}$$

and for diffractive photoproduction

$$\begin{aligned}\sigma^{(\mathrm{DPHO})}\left(ep \to eXY\right) = &\int_0^1 \mathrm{d}x f_\gamma^{(e)}\left(x\right) \int_0^{x_{I\!P,\max}} \mathrm{d}x_{I\!P} \int_{t_{\mathrm{cut}}}^{t_{\min}} \mathrm{d}t \\ &\int_0^1 \mathrm{d}x_j f_j^{(\gamma)}\left(x_j, \mu_F\right) \mathrm{d}x_i f_i^D\left(x_i, \mu_F, x_{I\!P}, t\right)\, \hat{\sigma}\left(ji \to XY\right)\end{aligned} \tag{5.1.6}$$

where, as in Chapter 3, in the latter case $f_j^{(\gamma)}$ is replaced by a Delta distribution for direct photoproduction.

5.2 Validation: simulation and data

5.2.1 Event generation

Events are generated based on a pre-release of SHERPA [2, 129] v3.1; the code will be made available in a future release. The matrix element part of the simulation used AMEGIC [96,97] and COMIX [87] for tree-level matrix elements, OPENLOOPS [83,130–132] for loop matrix elements, and Catani-Seymour dipole subtraction [95, 133] automated in [97] for the treatment of infrared divergences. The matching to the parton shower [99] is achieved through the MC@NLO formalism [100] in its implementation presented in [102]. For the PDFs we used built-in interfaces to the SAS1M set [52, 53] for the photon and the H1 2007 Fit B set [74] for the pomeron. For the reggeon and the proton, we used the sets GRVPI0 [128] and PDF4LHC21_40_pdfas [108], respectively, interfaced through LHAPDF [134]. The parameters of the pomeron flux were taken from the fit in [74]. We calculated in the 3-flavour scheme in accordance with the PDF and additionally allowed for massive c-quarks at NLO and massive b-quarks at LO in the final state. We consistently used the current default value for $\alpha_S(M_Z) = 0.118$ with three-loop running, which is also in-line with the H1 2006 PDF. The factorisation scale and the renormalisation scale were set to $\mu_F = \mu_R = H_T/2$ for photoproduction events and to $\mu_F^2 = \mu_R^2 = \frac{1}{4}(Q^2 + H_{T,\mathrm{hadr}}^2)$ for DIS events, with $H_{T,\mathrm{hadr}}$ as the scalar sum over the transverse momenta of all hadronic particles, and were varied by factors of 2 in a 7-point scale variation to estimate the uncertainties. The partonic final states were hadronised with the cluster fragmentation model of AHADIC [105], tuned to LEP data. To conserve quantum numbers, remnants are created similarly to the description in Section 2.3.5; the only difference is that the valence content of the pomeron is assumed to be a two-gluon state and of the reggeon any two-quark state.

Diffractive events are simulated with an assumed intact beam proton. Low-mass excitations have been seen to account for an additional 20% in cross-section flat in phase space and can therefore be taken into account with an overall scaling of 0.83 [135]. Diffractive photoproduction is composed of two components, the direct and the resolved photon contributions, with the latter simulated through a photon PDF. However appealing this picture of combining two PDFs is — one for the photon, one for the pomeron or reggeon — the assumed factorisation underpinning it is expected to break down [71] as a consequence of additional soft interactions between the photon and the proton beam. To account for the suppression of these events, we generalised the multiple-interactions modelling in SHERPA to allow for this kind of interaction. Naturally, this argument can only be applied to the resolved component, and therefore it has been conjectured that factorisation might still hold for the direct component [77]. We will study this ansatz in Section 5.4.3.

5.2.2 Experimental observables and datasets

Our implementation of diffractive events in DIS and photoproduction was validated with data from the H1 [72, 136] and ZEUS [137] collaborations.

H1, JHEP05 (2015) 056 The data from the H1 collaboration in [136] detected the outgoing proton in the Very Forward Proton Spectrometer (VFPS) and measured dijet production in both the DIS, $4\,\mathrm{GeV}^2 < Q^2 < 80\,\mathrm{GeV}^2$, and photoproduction, $Q^2 < 2\,\mathrm{GeV}^2$, regime. To describe the kinematics of the diffractive exchange, the variable $x_{I\!P}$ was defined as

$$x_{I\!P} = 1 - \frac{E'_p}{E_p} \tag{5.2.1}$$

with $E_p^{(\prime)}$ the energy of the incoming (outgoing) proton; the acceptance of the VFPS yielded a range of $0.010 < x_{I\!P} < 0.024$. The partonic momentum fractions with respect to the diffractive exchange and, in the case of photoproduction, with respect to the electron and the photon have been defined as

$$z_{I\!P} = \frac{Q^2 + M_{12}^2}{Q^2 + M_X^2} \tag{5.2.2}$$

for diffractive DIS events and

$$z_{I\!P} = \frac{\sum_{j=1,2} (E + p_z)_j}{\sum_{i \in X} (E + p_z)_i}\ , \quad y = \frac{\sum_{i \in X} (E - p_z)_i}{2E_e}\ , \quad \text{and} \quad x_\gamma = \frac{\sum_{j=1,2} (E - p_z)_j}{\sum_{i \in X} (E - p_z)_i} \tag{5.2.3}$$

for diffractive photoproduction events, where M_{12} denotes the dijet invariant mass, M_X the invariant mass of X defined below; the index j runs over the leading jets and the index i over the final state particles in X. The momentum transfer t from the proton was required to be $|t| < 0.6\,\mathrm{GeV}^2$. The leading and subleading jets were clustered with the k_T-algorithm with $R = 1$ in the photon-proton rest frame and were required to have transverse energies of $E_T^* > 5.5\,\mathrm{GeV}$ and $4.0\,\mathrm{GeV}^2$, respectively, and lie within the pseudorapidity range of $-1 < \eta < 2.5$ in the laboratory frame. The inelasticity y was required to be within $0.2 < y < 0.7$. The invariant mass of the system X was calculated as

$$M_X^2 = \left(\sum_{i \in X} E_i \right)^2 - \left(\sum_{i \in X} \vec{p}_i \right)^2 . \tag{5.2.4}$$

H1, EPJC51 (2007) 549 A similar measurement was undertaken in [72], where the mass of the system Y was restricted to $M_Y^2 < 1.6\,\mathrm{GeV}^2$. For the tagging of diffractive events the LRG method was used. Similarly to before, dijets were measured in the DIS, $4\,\mathrm{GeV}^2 < Q^2 < 80\,\mathrm{GeV}^2$, and photoproduction, $Q^2 < 0.01\,\mathrm{GeV}^2$, regimes and jets were clustered as before, demanding $E_T^* > 5(4)$ GeV for the (sub)leading jet within pseudorapidity $-1 < \eta^{\mathrm{lab}} < 2$. The photon-proton c.m.s.-energy W was restricted to $165\,\mathrm{GeV} < W < 242\,\mathrm{GeV}$, the proton momentum transfer to $|t| < 1\,\mathrm{GeV}^2$ and the longitudinal momentum ratio to $x_{I\!P} < 0.03$. The kinematical variables were defined as

$$x_{I\!P} = \frac{Q^2 + M_X^2}{Q^2 + W^2}\ , \ \text{and} \quad z_{I\!P} = \frac{Q^2 + M_{12}^2}{Q^2 + M_X^2}\ , \tag{5.2.5}$$

---

[2] By the superscript '*' we denote quantities measured in the photon-proton rest frame.

with M_{12} the dijet mass, for diffractive DIS events and

$$x_{I\!P} = \frac{\sum_{i\in X}(E+p_z)_i}{2E_p}\ ,\ z_{I\!P} = \frac{\sum_{j=1,2}(E+p_z)_j}{2x_{I\!P}E_p}\ ,\ y = 1-\frac{E'_e}{E_e}\ ,\ \text{and}\ x_\gamma = \frac{\sum_{j=1,2}(E-p_z)_j}{2yE_e} \tag{5.2.6}$$

for diffractive photoproduction events, where $E_e^{(\prime)}$ denotes the incoming (outgoing) lepton's energy; M_X was defined as in the previous analysis, c.f. Eq. (5.2.4).

ZEUS, EPJC55 (2008) 177 The ZEUS collaboration measured dijet production in diffractive photoproduction [137] with photon virtualities of $Q^2 < 1\,\mathrm{GeV}^2$ and an inelasticity of $0.20 < y < 0.85$. The out-going proton was not detected, instead diffractive events were selected by requiring LRGs. To correct for proton-dissociative events, a fraction of 16% had been subtracted from the total cross-section. The following variables were defined:

$$x_{I\!P} = \frac{\sum_h (E+p_z)_h}{2E_p}\ ,\ z_{I\!P} = \frac{\sum_j E_T^{(j)}\mathrm{e}^{\eta^{(j)}}}{2x_{I\!P}E_p}\ ,\ y = \frac{\sum_h (E-p_z)_h}{2E_e}\ ,\ x_\gamma = \frac{\sum_j E_T^{(j)}\mathrm{e}^{-\eta^{(j)}}}{2yE_e}\ , \tag{5.2.7}$$

and

$$M_X^2 = \sum_h (E-p_z)_h\,(E+p_z)_h\ , \tag{5.2.8}$$

where the index h runs over reconstructed Energy Flow Objects in the main detectors, and it was required $x_{I\!P} < 0.025$. Jets were clustered with the k_T algorithm using $R = 1$ in the laboratory frame, with cuts of $E_T > 7.5(6.5)$ GeV for the (sub)leading jet and within pseudorapidity $-1.5 < \eta < 1.5$.

5.3 Diffractive DIS

Turning first to the Diffractive DIS measurements, we compare the SHERPA results, obtained at MC@NLO accuracy with dijet production data from two publications by the H1 collaboration [72, 136].

In Fig. 5.6 we focus on fairly inclusive observables that describe the overall kinematics of the events. They include very general observables such as the centre-of-mass energy of

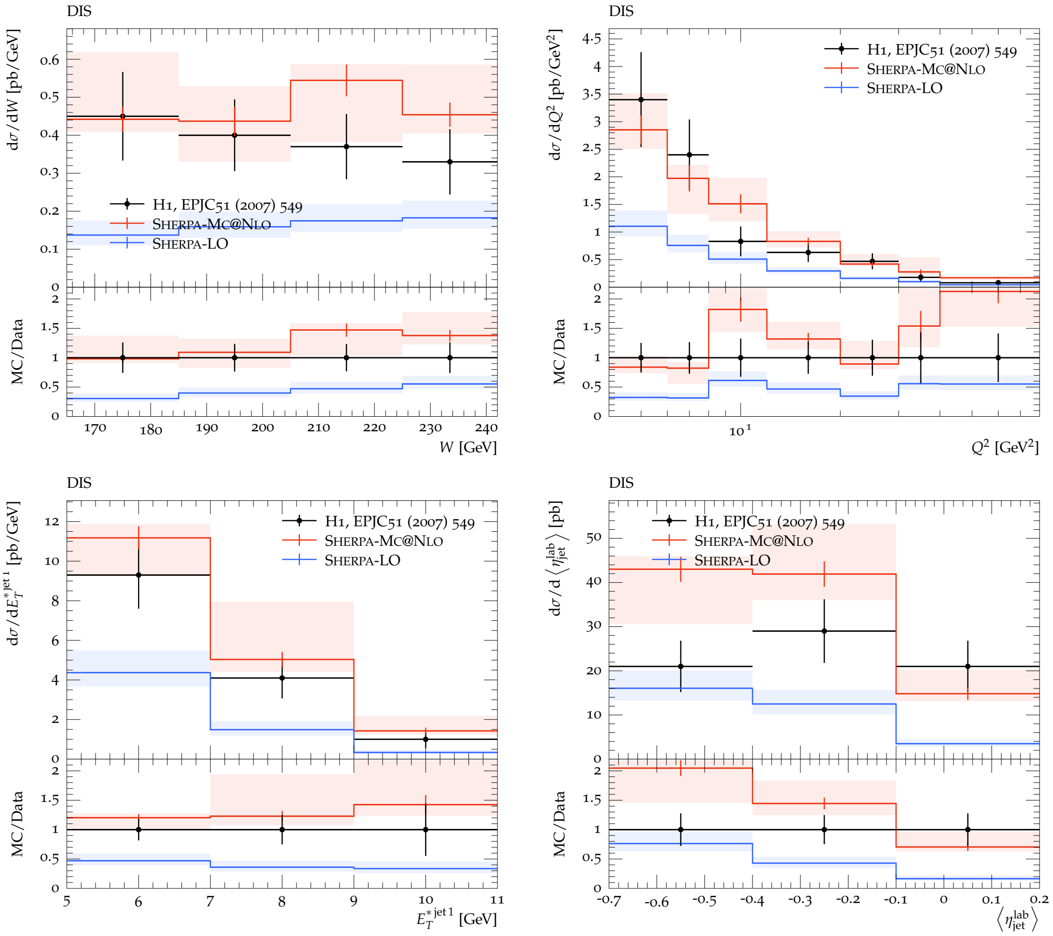


Figure 5.6: Differential DDIS cross-sections with respect to the photon-proton centre-of-mass energy W (top left), the photon virtuality Q^2 (top right), the leading jet transverse momentum $E_T^{*\mathrm{jet1}}$ (bottom left), and average jet pseudorapidity $\langle \eta_{\mathrm{jet}}^{\mathrm{lab}} \rangle$ (bottom right) from [72].

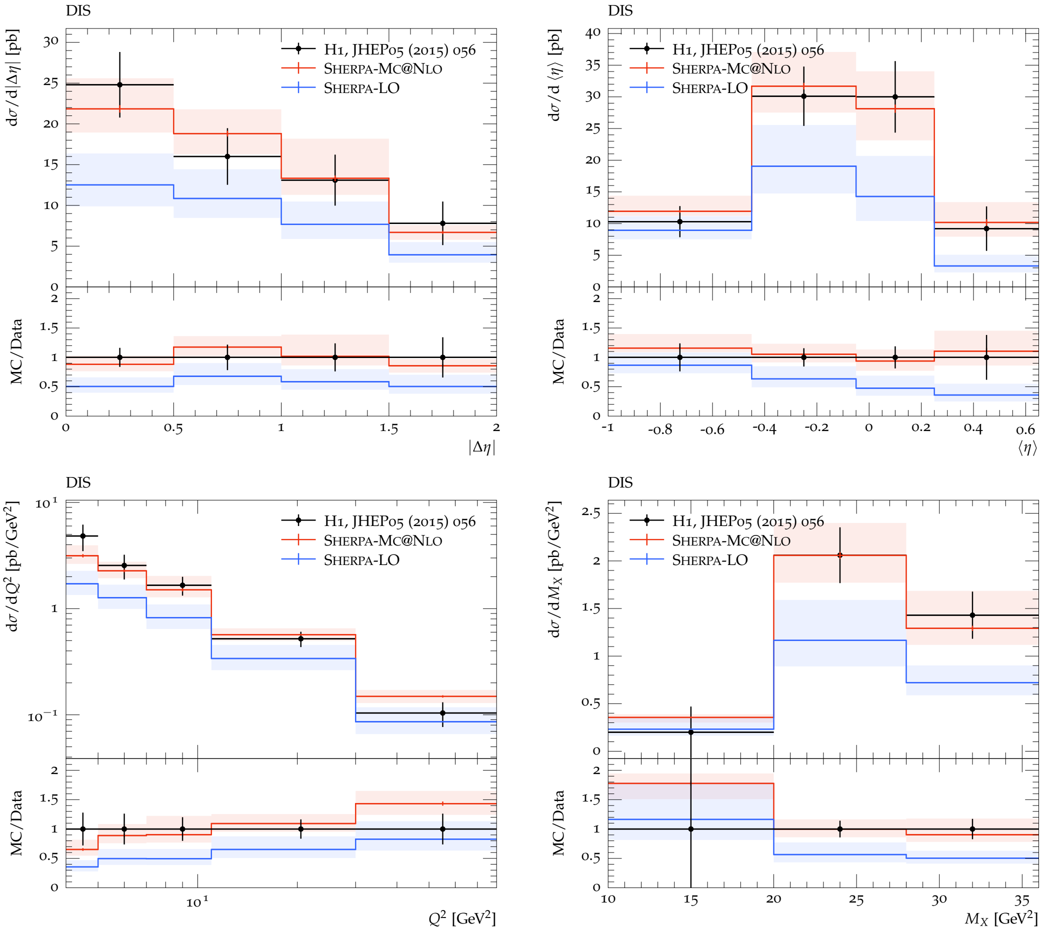


Figure 5.7: Differential DDIS cross-sections with respect to the average pseudorapidity $\langle\eta\rangle$ (top left) and the pseudorapidity difference $|\Delta\eta|$ (top right) of the jets as well as the virtuality Q^2 (bottom left) and the diffractive system's mass M_X (bottom right) from [136].

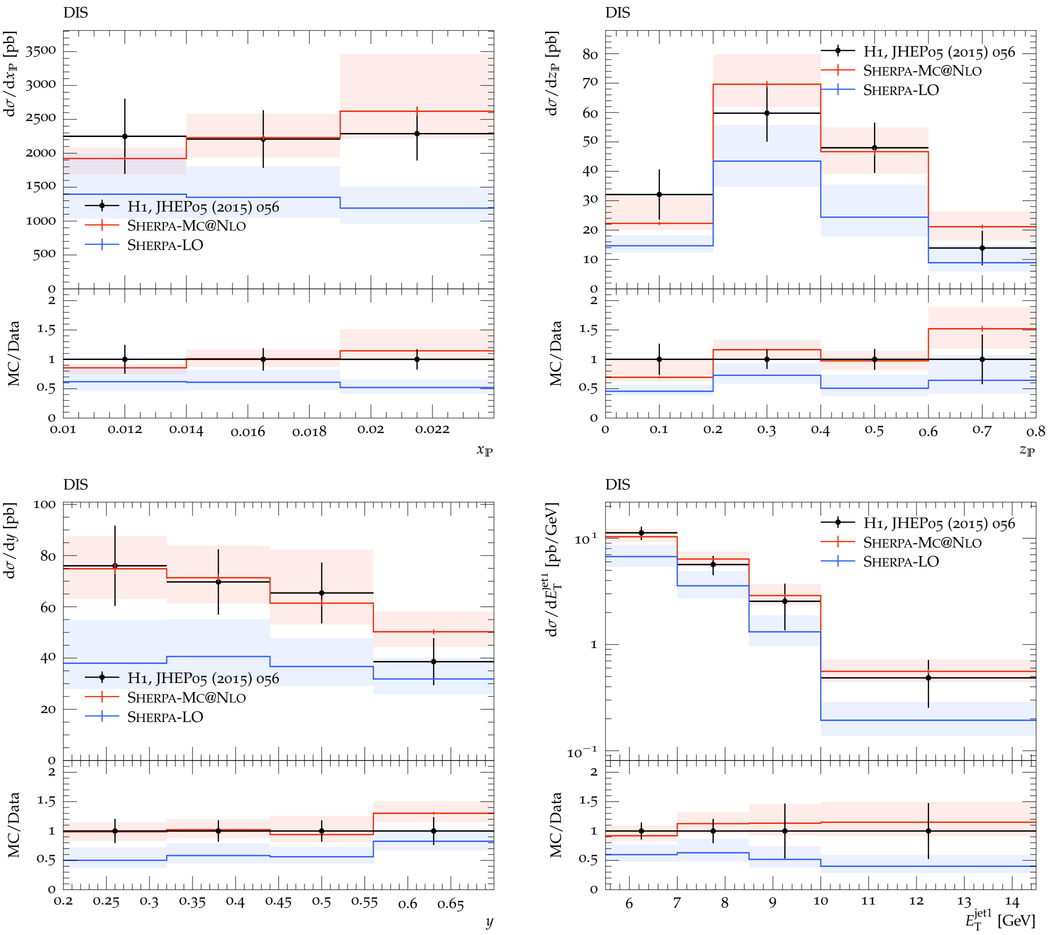


Figure 5.8: Differential DDIS cross-sections with respect to momentum ratios $x_{I\!P}$ (top left), $z_{I\!P}$ (top right), the inelasticity y (bottom left), and the transverse momentum of the leading jet E_T^{jet1} (bottom right) from [136].

the photon-proton system W, the photon virtuality Q^2, the transverse momentum of the leading jet $E_T^{*\text{jet1}}$ and the average jet pseudorapidity $\langle\eta_{\text{jet}}^{\text{lab}}\rangle$, all taken from [72]. Comparing against data from [136], we extend to similar inclusive observables in Fig. 5.7 such as pseudorapidity average and difference, $\langle\eta\rangle$ and $|\Delta\eta|$, virtuality Q^2 and the diffractive mass M_X; and observables that expose more of the underlying parton-level dynamics such as the momentum fractions $x_{I\!P}$ and $z_{I\!P}$ and the inelasticity y alongside the leading jet transverse momentum E_T^{jet1} in Fig. 5.8. We consistently observe excellent agreement of the simulation at MC@NLO accuracy with data, with K-factors ranging from about 1.5 up to 2-3 depending on the observable and the associated phase space region. With the exception of the W and y distributions, the K-factors are relatively flat and do not overly change the shape of the distributions. This confirms other findings which established good agreement of data and fixed-order calculations at NLO [73] and the corresponding K-factors. The increasing size of K-factors is readily understood and associated with the additional phase space made available due to the asymmetric cuts and additional parton-level channels with the largest enhancement of the cross-section seen in the forward region, as expected.

5.4 Diffractive photoproduction

5.4.1 Diffractive photoproduction at MC@NLO accuracy

Turning now to the description of diffractive photoproduction of dijets, we observe that the convincing agreement of simulation and data does not hold true anymore, supporting statements about the possible breakdown of factorisation in such processes [71]. To highlight this, let us first take a look at the momentum ratios $x_{I\!P}$, $z_{I\!P}$, y and x_γ. They have been analysed by both the H1 and the ZEUS collaboration in [136] and [137], respectively, and we display the results in the left and right column of Fig. 5.9. As already seen in previous NLO calculations [79], the calculation of the cross-section in our MC@NLO samples severely overestimates the data in diffractive photoproduction, by a factor of up to 2-3, while the LO predictions are in somewhat better agreement overall. In addition, we observe a sharp increase, amounting to a visible shape distortion, of the MC@NLO simulation with respect to the experiment, particularly at large values of $x_\gamma \gtrsim 0.75$, a regime usually associated with "direct" photoproduction, i.e. the photon acting as a point-like particle. By and large, however, the MC@NLO simulation describes data reasonably well in the "resolved" photoproduction regime of small $x_\gamma \lesssim 0.5$.

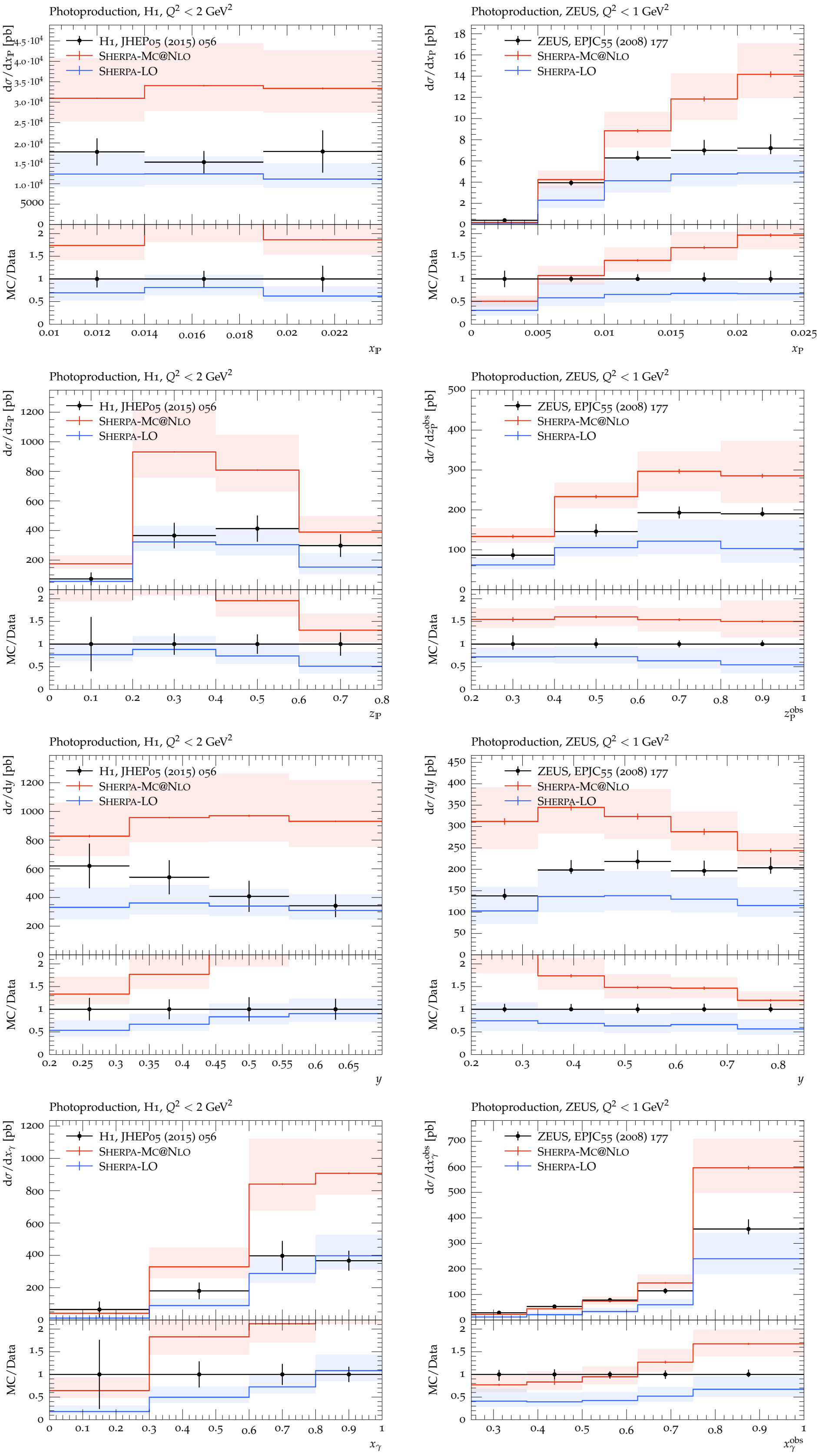


Figure 5.9: Differential diffractive photoproduction cross-sections with respect to momentum ratios $x_{\mathbb{P}}$, $z_{\mathbb{P}}$, y and x_γ (top to bottom), obtained by H1 (left column, data from [136]) and ZEUS (right column, data from [137]).

We note that differences in the shape and normalisation between the two different sets of data - and therefore between the two simulations - stem from the different phase spaces populated by the H1 and ZEUS analyses, mainly related to the difference in the range of Q^2 and the definition of the diffractively scattered proton.

Our findings so far of NLO predictions overshooting data by large factors, are in agreement also with an analysis of final state observables, such as the leading jet transverse energy, E_T^{jet1}, the diffractive invariant mass M_X, and the average pseudorapidity of jets $\langle\eta\rangle$ and the pseudorapidity difference between the two leading jets $|\Delta\eta|$, which we show in Fig. 5.10. Again the MC@NLO predictions have a K-factor of about 2 with respect to their leading-order counterparts, and they exceed data (taken from H1 [136]), again, by a factor of about 2. We observe that the shapes of the distributions do not agree between data and theory, especially in the pseudorapidity distributions and the momentum fractions.

To compensate for the massive difference in measured and calculated cross-sections, H1 applied a global factor of about 0.5 to the calculation [72,136,138], improving its agreement with the data. This rescaling does not improve the theory agreement with the ZEUS data, where the calculation even slightly undershoots the data for low x_γ; therefore a global scaling can not adequately be used to describe the factorisation breaking. We will turn to this problem in more detail in Section 5.4.3 by further examining the interplay seen in this observable. We conclude that a naive perturbative expansion without an appropriate modelling of factorisation breaking does not seem to converge for diffractive photoproduction.

5.4.2 Reggeon contribution

To further elucidate the components of the cross-section, we show in Fig. 5.11 the leading jet transverse energy, comparing the calculation with and without the reggeon contribution, in three different definitions of phase space. Depending on it, we find a sizeable positive contribution for the full range of the observable.

5.4.3 Modelling factorisation breaking in diffractive photoproduction

The overshoot of the cross-section hints at the breakdown of the factorisation and different models have been brought forward to explain the discrepancies, which we will review in the following. The variable x_γ has been used by the experiments as a discriminator

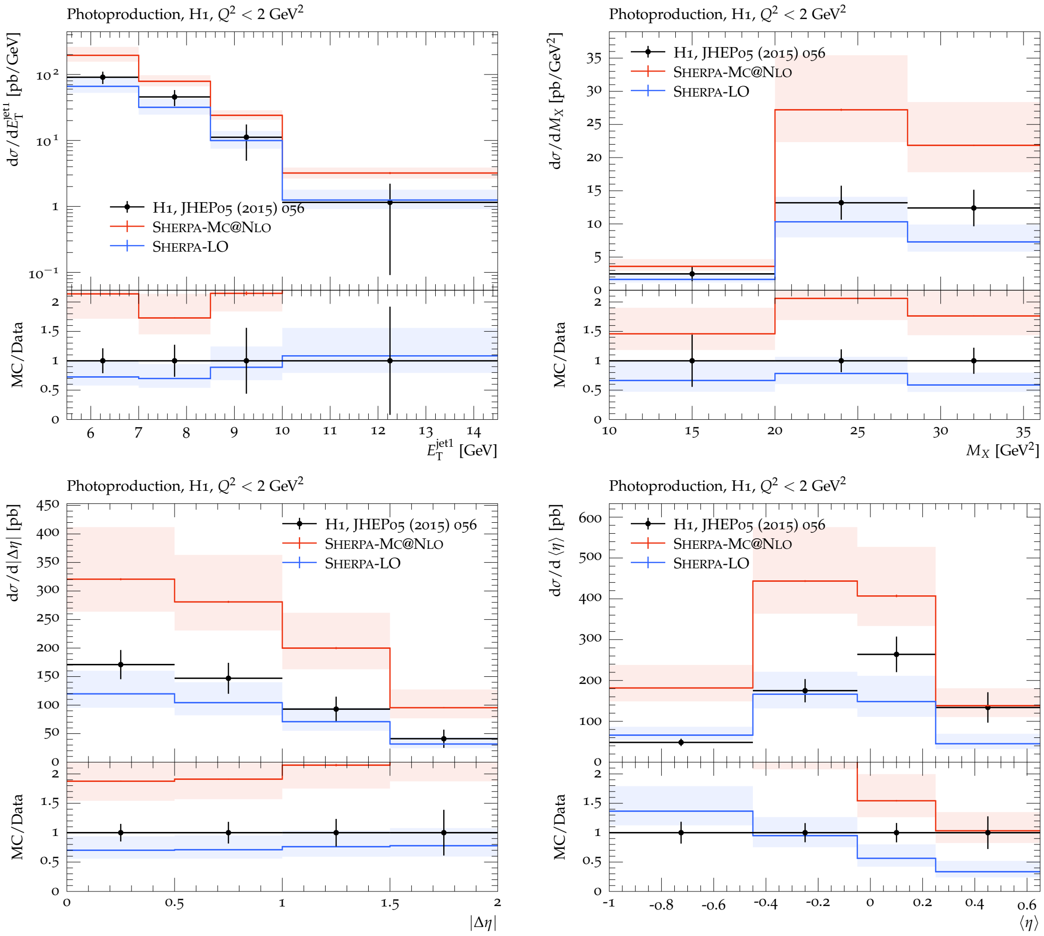


Figure 5.10: Differential dijet Diffractive Photoproduction cross-sections with respect to leading jet transverse energy E_T^{jet1}, invariant mass M_X, difference in dijet pseudorapidity $|\Delta\eta|$ and average dijet pseudorapidity $\langle\eta\rangle$. Results of the SHERPA simulation with MC@NLO accuracy are compared with results at LO and with data from H1 [136].

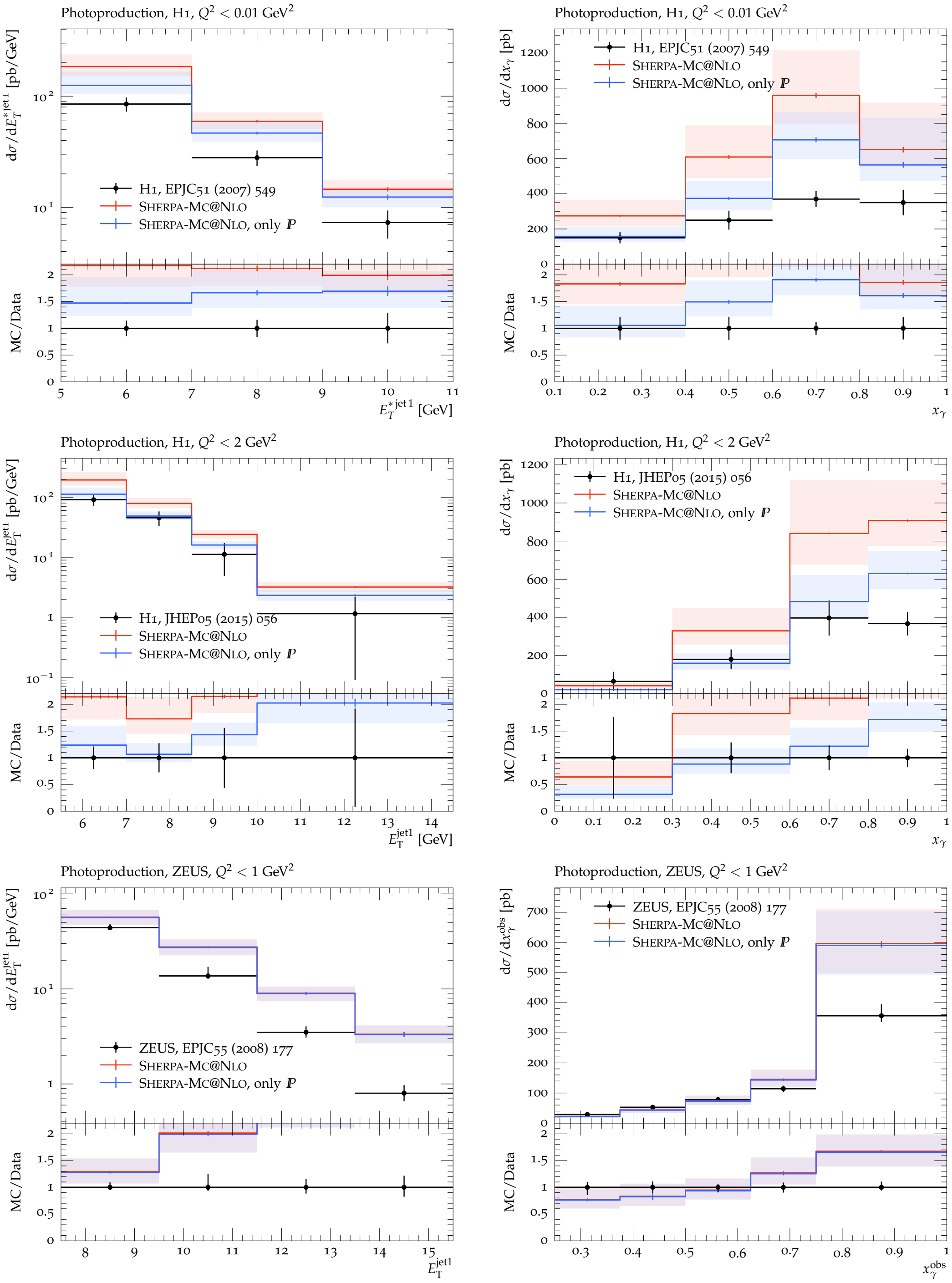


Figure 5.11: Distributions of leading jet transverse energy E_T^{jet1} (left column) and x_γ (right column) with and without reggeon contribution, compared to data by H1 from [72] (top) and [136] (middle) and from ZEUS [137] (bottom) in their respective definitions of fiducial phase space.

between the direct and resolved photon components and our calculation confirms that it indeed works well to discern the different contributions. Generally, a small value of x_γ will correspond to a dominant contribution by the resolved photon and its PDF, while values close to unity will mostly stem from direct contributions, i.e. the photon acting as a point-like particle. In this section we use the handle this observable provides on the dynamics of factorisation breaking and its interplay with the direct and/or resolved components.

In [78] it was argued that the factorisation breaking is a consequence of hadronisation, bin migration and NLO effects, which we exemplify in the distribution of x_γ in Fig. 5.12. While hadronisation certainly plays a big role in the bin migration between the two largest x_γ bins and the total cross-section is decreased, the overall overestimation of the total cross-section is still present after taking these effects into account. In fact, hadronisation is essential for the reconstruction of x_γ in the ZEUS analysis, as a significant number of events in the parton-level simulation end up in the region $x_\gamma > 1$. It is also somewhat amusing to note that hadronisation effects tend to reduce the overshoot of the simulation in the highest x_γ bins for the H1 analyses, while they tend to actually create it for the ZEUS analysis.

As mentioned previously, it has also been argued that the decrease of the cross-section is due to soft interactions between the resolved photon and the proton [71]. We implemented a simplistic generalised multiple-parton interactions modelling in SHERPA to veto events, which have an additional scattering between the photon and the proton and would thus destroy the rapidity gap. In Fig. 5.13 we show the effect of this rejection. Naturally this mechanism only applies to the resolved component and thus affects only those regions of the phase space where the resolved component dominates. We also recall that the resolved component can be further decomposed into the point- and hadron-like component, where the difference is that the backwards evolution would collapse the former to a photon and the latter to a meson-like state. Furthermore, it has been pointed out in [78] that the so-called anomalous component, i.e. splittings of $\gamma \to q\bar{q}$, would not exhibit further interactions with the proton. This would lead to a smaller suppression in the resolved component, depending on the size of these splittings in the radiation off these quarks[1]. We therefore expect that the generalised MPI modelling would only apply to meson-like states, whereas the point-like contribution would see a suppression mechanism similar to the direct component. A comprehensive study of the suppression mechanism will have to disentangle these two components; while the hadron-like states will undergo the MPI-

---

[1] These splittings are currently not included in the simulation in SHERPA and, as indicated, we leave the study of this effect to future work.

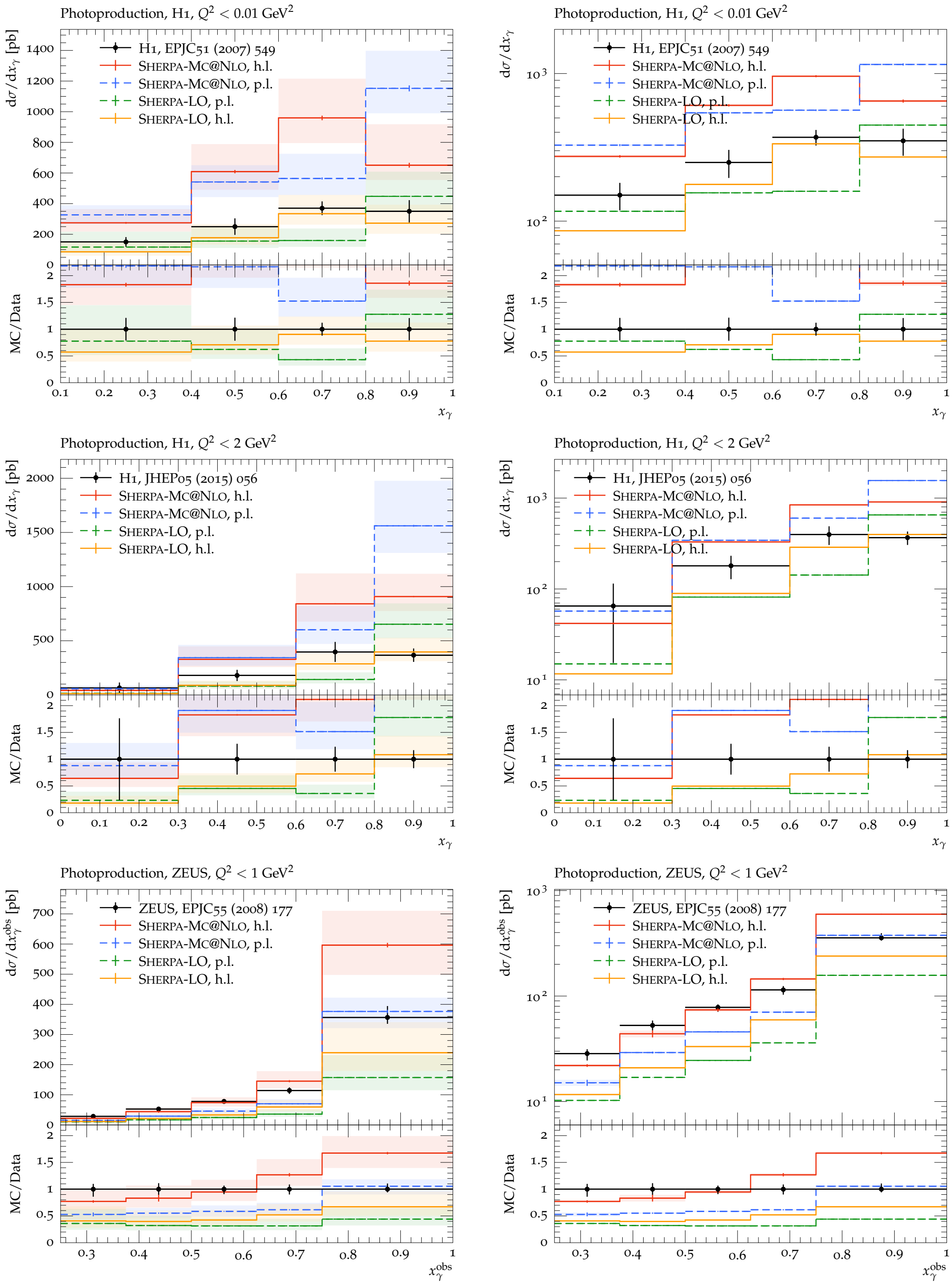


Figure 5.12: Differential distribution of x_γ in diffractive photoproduction. Results of the SHERPA simulation at parton- and hadron-level (p.l. and h.l. respectively) with MC@NLO accuracy are compared with results at LO and with data by H1 from [72] (top) and [136] (middle) and ZEUS [137] (bottom); with 7-point scale variations, and on linear scale in the left column and on logarithmic scale on the right column.

based breaking of the factorisation, the point-like state will have to interpolate between the direct and the hadron-like regimes. The implementation of this modelling and the details of the suppression of the point-like resolved component are left for future work. However, while our naive model depends on some assumptions about the impact parameter and other unconstrained parameters and some additional simplifications, which certainly deserve further investigation, it does not appear as if these effects alone can accommodate the observed large discrepancy of simulation and data.

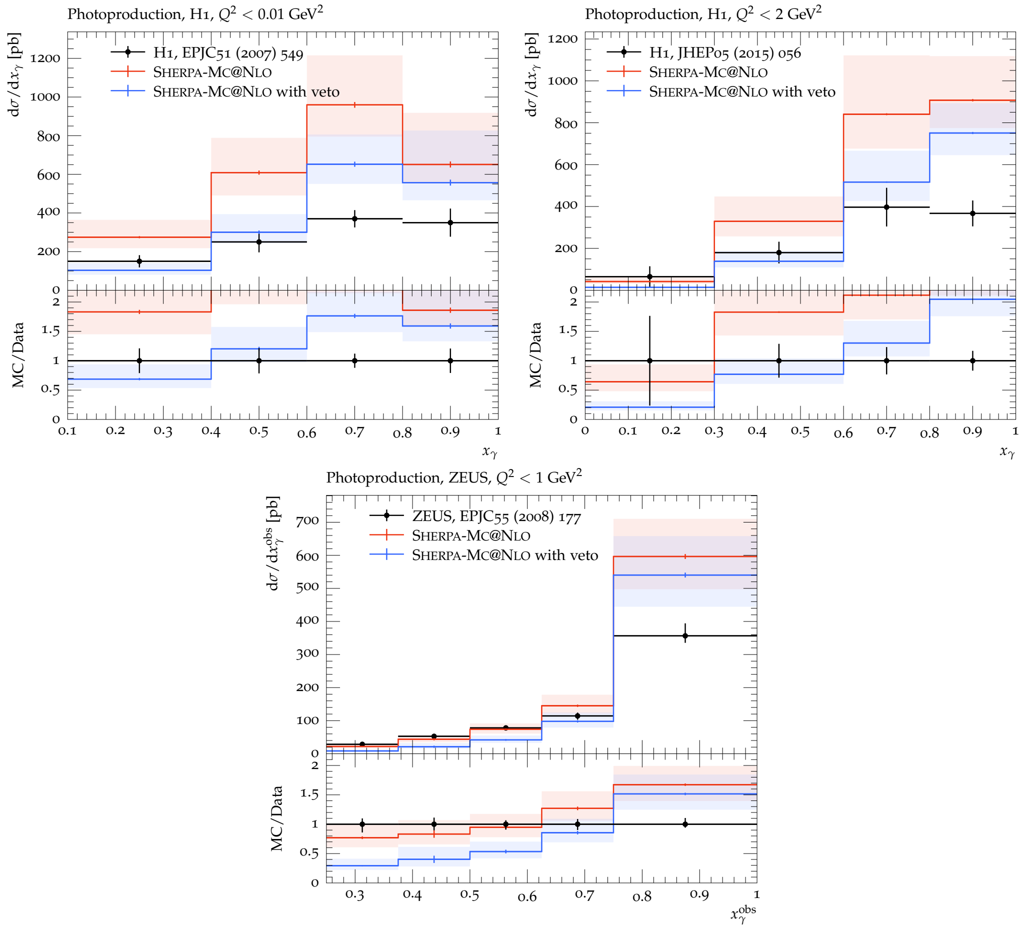


Figure 5.13: Differential diffractive photoproduction cross-sections with respect to momentum ratio x_γ as measured by H1 in [72] (top left) and [136] (top right) and x_γ^{obs} as measured by ZEUS [137] (bottom), compared to results of the SHERPA simulation at MC@NLO accuracy with and without a veto on γp interactions.

The authors of [81] conducted a study which found that, even though there is a slight dependence, the different phase space cuts are not the cause of the discrepancy in the suppression between the H1 and ZEUS measurements.

Table 5.1: Scaling factors for the direct, R_{dir}, and resolved, R_{res}, component in diffractive photoproduction for the respective experimental data.

	H1, EPJC51 (2007) 549 [72]	H1, JHEP05 (2015) 056 [136]	ZEUS, EPJC55 (2008) 177 [137]
R_{res}	0.4 ± 0.1	0.6 ± 0.3	1.3 ± 0.1
R_{dir}	0.4 ± 0.1	0.3 ± 0.2	0.5 ± 0.1

Turning back to the x_γ distributions, we point out that, while the data from ZEUS and H1 do not agree with each other on the overall size of the necessary suppression, they both show an overshoot for large x_γ values. This hints at factorisation breaking in the direct component too. In fact, the distinction between direct and resolved components of photoproduction cannot be maintained at NLO as the real correction to "direct" photoproduction, $\gamma j \to jjj$, cannot be distinguished from contributions of photon splitting combined with a two-jet matrix element, $(\gamma \to q\bar{q})_{\text{PDF}} \otimes (qj \to jj)_{\text{ME}}$, where j denotes any parton. This leads us to revisit the logic outlined in [72], to further elucidate the impact of factorisation breaking on the different photonic components. We fit two prefactors, one each for direct and resolved components in the simulation, to the data to quantify the effect of the suppression in the two components separately, with results shown in Table 5.1. While the scaling of the resolved component does vary among the different measurements, the direct component seems to support a somewhat universal suppression by a factor of 0.5. The data covers different cuts on the photon virtuality, hence the suppression seems to be independent of the kinematics at the electron-photon vertex.

In Fig. 5.14 we exhibit the results of the fit; we also did not observe any significant discrepancies between simulation and data in other distributions as a result of the rescaling.

5.5 Predictions for the EIC

5.5.1 Diffractive DIS

For the analysis of diffractive DIS events we implemented a routine for RIVET [106], loosely based on the measurement in [136], with the following phase space: The photon virtuality

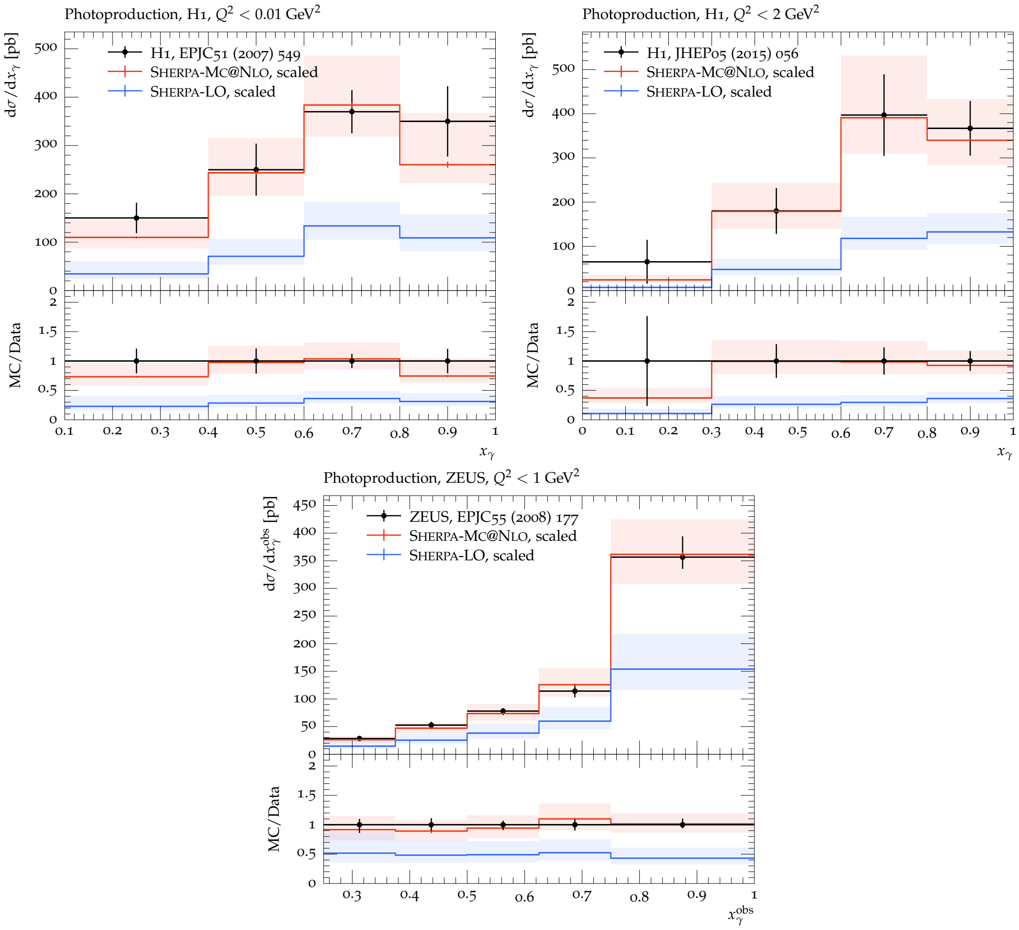


Figure 5.14: Differential diffractive photoproduction cross-sections with respect to momentum ratio x_γ as measured by H1 in [72] (top left) and [136] (top right) and x_γ^{obs} as measured by Zeus [137] (bottom), compared to results of the Sherpa simulation at Mc@Nlo accuracy with the direct and resolved component scaled separately.

was restricted to $4\,\mathrm{GeV}^2 < Q^2 < 110\,\mathrm{GeV}^2$, and we clustered jets in the lab frame using the k_T algorithm with $R = 1$ within pseudorapidity $|\eta| < 4$, demanding at least two jets with transverse energy E_T of at least 5.5 and 4 GeV, respectively. We assumed proton-tagging and reconstructed the $x_{I\!P}$ as

$$x_{I\!P} = 1 - \frac{E_p'}{E_p} \tag{5.5.1}$$

which had to satisfy $x_{I\!P} < 0.1$, to allow for more phase space for jet production compensating for the lower beam energies. The momentum transfer was restricted to $|t| < 0.6\,\mathrm{GeV}^2$, and we defined

$$z_{I\!P} = \frac{Q^2 + M_{12}^2}{Q^2 + M_X^2}\,. \tag{5.5.2}$$

In Fig. 5.15, we present inclusive observables like the leading jet transverse energy $E_T^{(1)}$, the average jet rapidity $\langle\eta\rangle$, diffractive mass M_X, and dijet mass M_{jj}. We note, again, significant K-factors between LO and MC@NLO predictions ranging up to values of about 5 in the forward regime, testaments to the lower energy scales of the processes we study here.

In addition to these observables we also display in Fig. 5.16 the distributions in the momentum fractions $x_{I\!P}$ and $z_{I\!P}$ and the inelasticity y as well as event shape distributions, namely transverse thrust $T_\perp$, transverse thrust minor $T_{\perp m}$ and transverse sphericity $S_\perp$. These observables are defined by

$$T_\perp = \max_{\vec{n}_\mathrm{T}} \frac{\sum_i |\vec{p}_{\mathrm{T},i} \cdot \vec{n}_\mathrm{T}|}{\sum_i \vec{p}_{\mathrm{T},i}}\,, \quad T_{\perp,m} = \max_{\vec{n}_\mathrm{T}} \frac{\sum_i |\vec{p}_{\mathrm{T},i} \times \vec{n}_\mathrm{T}|}{\sum_i \vec{p}_{\mathrm{T},i}}\,, \quad S_\perp = \frac{2\lambda_2}{\lambda_1 + \lambda_2}\,, \tag{5.5.3}$$

and n_T is the transverse-thrust axis that maximises the $T_\perp$ and $\lambda_{1,2}$ are the eigenvalues of the transverse linearised sphericity tensor $\mathbf{S}_{\alpha\beta}$

$$\mathbf{S}_{\alpha\beta} = \frac{1}{\sum_i |\vec{p}_{\mathrm{T},i}|} \sum_i \frac{1}{|\vec{p}_{\mathrm{T},i}|} \begin{pmatrix} p_{i,x}^2 & p_{i,x}p_{i,y} \\ p_{i,y}p_{i,x} & p_{i,y}^2 \end{pmatrix}. \tag{5.5.4}$$

While the events are broadly dominated by dijet kinematics, the event shape distributions indicate a non-negligible contribution from three-jet events. Clearly, the data taken at the EIC will complement the HERA data in the high-x region [139] in updated fits to the DPDFs.

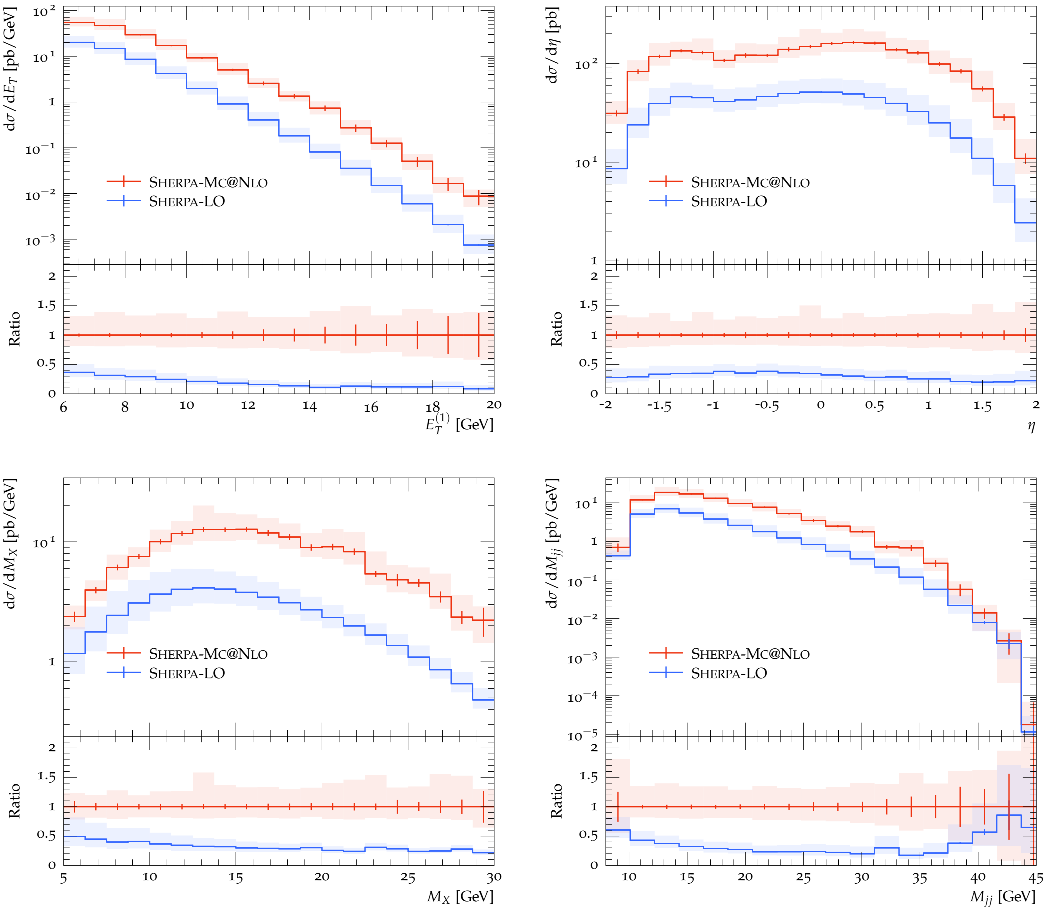


Figure 5.15: Predictions at LO and MC@NLO accuracy of SHERPA for diffractive DIS at the EIC for the leading jet transverse momentum $E_T^{(1)}$ (top left), average jet rapidity $\langle\eta\rangle$ (top right), diffractive mass M_X (bottom left), and dijet mass M_{jj} (bottom right). Phase space regions below the kinematic cuts are populated by the parton shower.

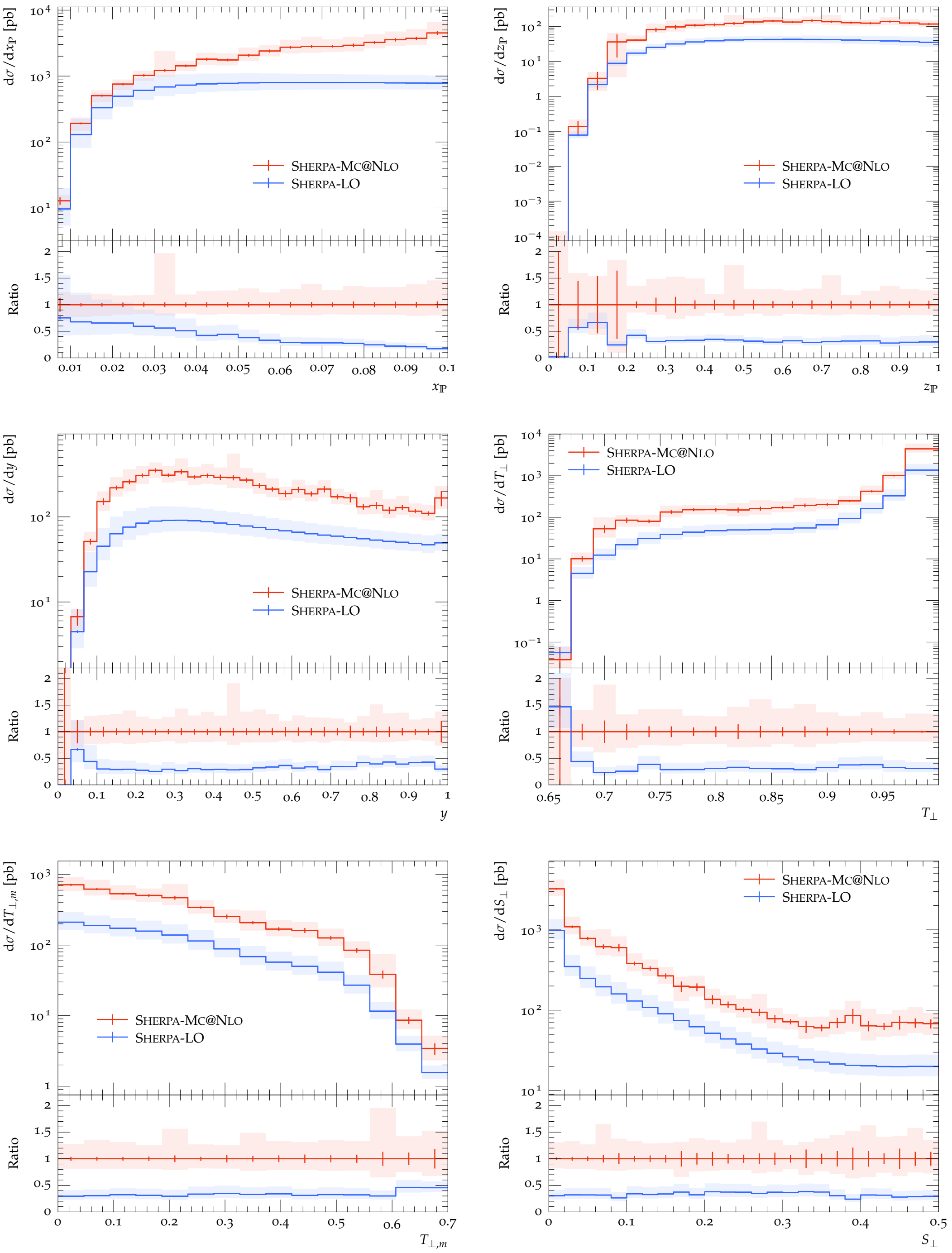


Figure 5.16: Predictions at LO and MC@NLO accuracy of SHERPA for diffractive DIS at the EIC for the momentum ratios $x_{I\!P}$, $z_{I\!P}$ (top row), inelasticity y, transverse thrust $T_\perp$ (middle row), thrust minor $T_{\perp,m}$ and transverse sphericity $S_\perp$ (bottom row).

5.5.2 Diffractive photoproduction

Measuring diffractive photoproduction at the EIC will shed new light on factorisation breaking, and expand on some of the findings highlighted in Section 5.4.3. To obtain predictions for this process, we calculated an average over the suppression factors for direct and resolved photon processes from Table 5.1: $R_{\text{res}}^{(\text{EIC})} = 0.8 \pm 0.2$ and $R_{\text{dir}}^{(\text{EIC})} = 0.4 \pm 0.1$ for the resolved and direct component, respectively. Additionally to the scale uncertainty, we obtained envelopes for the uncertainties from the suppression factors. We used the same cuts and settings as for the predictions for diffractive DIS, see the previous subsection Section 5.5.1, restricting the virtuality to $Q^2 < 4\,\text{GeV}^2$ and defined different momentum fractions as

$$x_{\mathbb{P}} = 1 - \frac{E_p'}{E_p}\,, \quad z_{\mathbb{P}} = \frac{\sum_j E_T^{(j)} \mathrm{e}^{\eta^{(j)}}}{2x_{\mathbb{P}}E_p}\,, \quad y = 1 - \frac{E_e'}{E_e}\,, \quad x_\gamma = \frac{\sum_j E_T^{(j)} \mathrm{e}^{-\eta^{(j)}}}{2yE_e} \tag{5.5.5}$$

In Fig. 5.17, we display the momentum fractions $x_{\mathbb{P}}$, $z_{\mathbb{P}}$ and x_γ and inelasticity y and observe K factors of about 2.5 for all these variables, approximately constant throughout. We complement these by jet observables as well as event shapes in diffractive photoproduction in Fig. 5.18. Uncertainties due to the fitting to factorisation breaking and due to scale choices are of comparable size, and the K factors between the MC@NLO and LO accuracy again reach values of about 5. This is a consequence of low scales present in the process and new channels opening up. The theoretical description hinges on the understanding of the factorisation breaking mechanism and MC@NLO predictions will provide a baseline for the data taken at the EIC. The event shapes exhibit an anticipated effect, namely that the limited phase space leads to lower multiplicities and an even stronger dominance of dijet events compared to the DDIS events in the previous chapter, manifest in the sharper peak at low values of $1 - T_\perp$, $T_{\perp m}$, and $S_\perp$.

5.6 Summary

Diffraction played an important role at HERA, making up 10% of the total cross-section. We simulate (hard) diffractive jet production at Next-to-Leading-Order in SHERPA for electron-proton collision in both the DIS and photoproduction regimes, matching our calculation to the parton shower. This results in the first fully differential hadron-level calculation of hard diffraction at MC@NLO accuracy and provides an important accuracy standard for future EIC predictions.

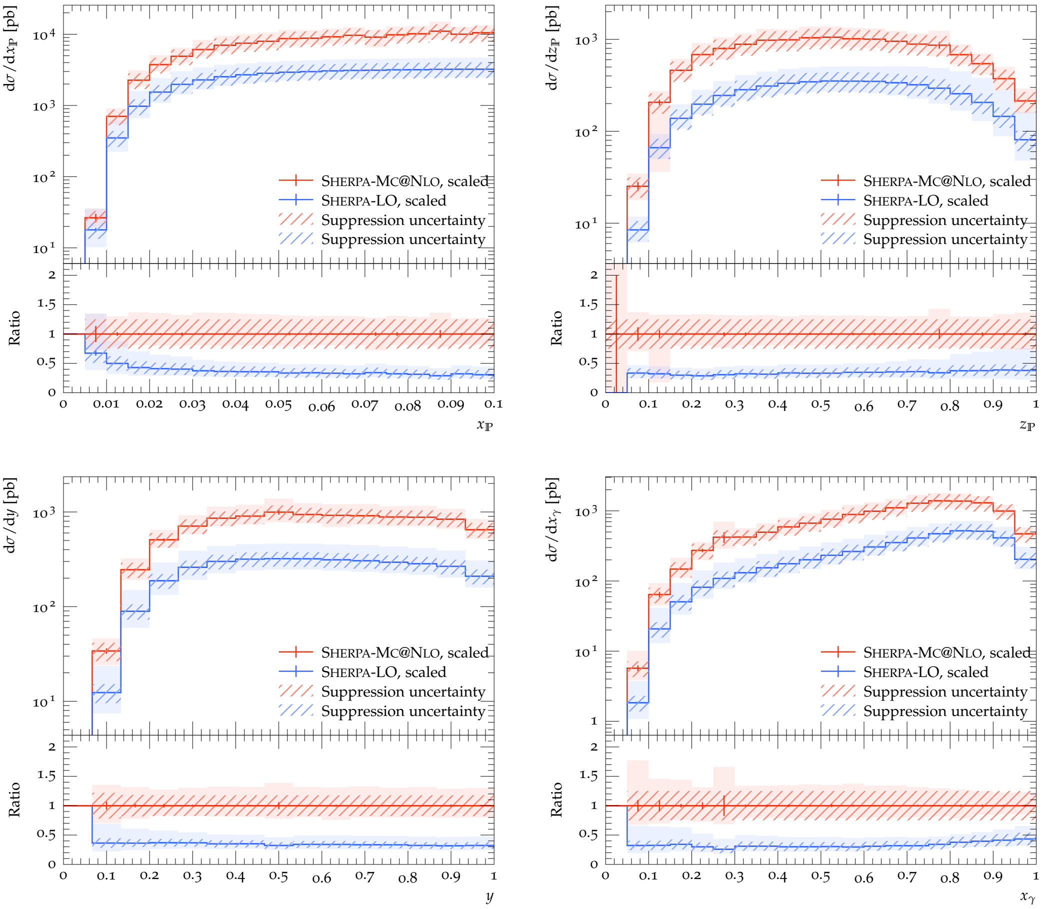


Figure 5.17: Predictions at LO and MC@NLO accuracy of SHERPA of momentum ratio $x_{I\!P}$, $z_{I\!P}$ and x_γ (upper row, left to right), of inelasticity y in diffractive photoproduction at the EIC.

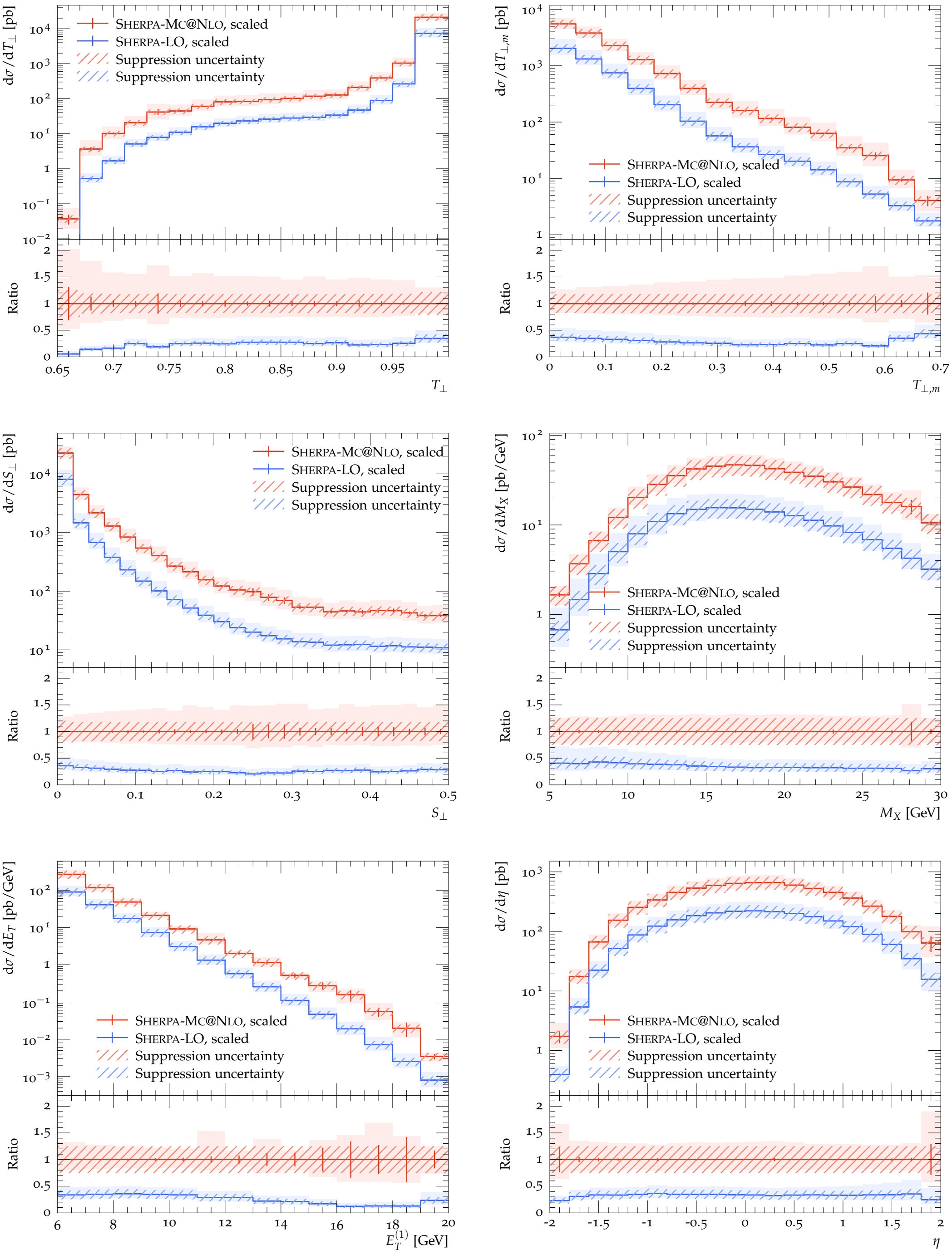


Figure 5.18: Predictions at LO and MC@NLO accuracy of SHERPA of event shapes transverse thrust $T_\perp$ and transverse thrust minor $T_{\perp m}$ (top row), transverse sphericity $S_\perp$ and diffractive mass M_X (middle row) and leading jet transverse momentum E_T^{jet1} and average jet pseudorapidity $\langle\eta\rangle$ (botto row) in diffractive photoproduction at the EIC. Phase space regions below the kinematic cuts are populated by the parton shower.

Validating our simulation against data from the H1 and ZEUS collaborations we see excellent agreement for Diffractive DIS setups. In the Diffractive Photoproduction regime, we observe significant discrepancies compared to the data, confirming the findings in previous Fixed Order calculations. We review and discuss different ansätze to explain the differences in view of our hadron-level simulations and conclude that none of the proposed solutions suffices to conclusively clarify the mechanism of factorisation breaking in this regime. We argue that the factorisation breaking happens in fact in *both* the direct and the resolved component in diffractive photoproduction. A coherent mechanism would need to take into account the suppression of the real correction to the "direct" component at NLO. We quantify the suppression in this component by fitting the two components to the data.

Lastly, we present predictions for Diffractive DIS and Diffractive Photoproduction at the EIC, where for the latter we estimated the suppression due to factorisation breaking by means of the fits to the H1 and ZEUS data. Data taking at the EIC will provide more insights into the exact mechanism of factorisation breaking and a thorough comparison to theory predictions will determine the exact nature of the corresponding mechanism. With this understanding it should be possible to apply this to hadron colliders, hence allowing to study this phenomenon on the basis of the vast data taken at the LHC.

Chapter 6

Conclusion

Lepton-hadron colliders provide invaluable insights into the partonic content of the probed hadron. As a result, Deep-Inelastic Scattering has seen large interest in the High Energy Physics community. In the collider experiments, these events are subdominant, with most hadronic activity stemming from photoproduction events. A third event class is constituted by diffractive events, which are defined as scatterings in which the proton stays intact.

The construction of the Electron-Ion Collider in the US and its ambitious precision targets make it necessary to revisit these phenomena and provide state-of-the-art simulations for them. In this work, the latter two event classes, photoproduction and diffraction, are implemented in the SHERPA event generator.

In Chapter 3 we present phenomenology of photoproduction. Jet photoproduction necessitates the inclusion of the hadronic component of the photon that comes into play at low virtualities. Photon Parton Distribution Functions must therefore be included in the calculation of the cross-section and are interfaced to SHERPA. We generalize the multiple-interactions model to allow for additional interactions with these resolved photons. We extend the calculation to NLO and are able to produce the first hadron-level matched NLO simulation of photoproduction, which we validate against data from the OPAL and ZEUS experiments at LEP and HERA, respectively.

Comparing the LO and NLO simulation, we see a large improvement with respect to the data in the shape and the overall cross-section. While the agreement of the NLO simulation with the OPAL data is excellent, the ZEUS data is somewhat undershot in double-differential distributions of x_γ and η. We attribute this to the missing tune of MPI, the correct treatment of photon splittings in the initial-state shower and — most importantly — to the parton content of the photon and potential associated systematic errors. Our study is complemented with an examination of the photon PDF uncertainties at LO, comparing all available PDF sets in SHERPA, where we find uncertainties of similar size as the scale uncertainties at LO. We point out that new fits will be crucial to photoproduction phenomenology at the EIC. Following that we provide predictions for the EIC for jet photoproduction and analyse the events with inclusive-jet as well as event-shape and heavy-jet observables.

Moving on, in Chapter 4 we compare three general-purpose event generators, HERWIG, PYTHIA and SHERPA, with respect to photoproduction. Between PYTHIA and SHERPA differences are scrutinised and attributed to the differing treatments of beam remnants, the value used for α_S, the flux, the photon splittings in the initial-state shower and the choice of photon PDF. We compare the three generators for LEP, HERA and EIC setups, and therefore see significant differences between them. In view of the EIC, we argue that more work is needed to properly understand the different regimes, like the transition region between photo- and electroproduction and the interplay of direct and resolved processes. Making the data from past colliders accessible in the RIVET framework would expedite the studies of these effects and the comparison of different approaches.

Last but not least, in Chapter 5 we study diffraction by implementing the relevant flux and PDFs in SHERPA. While diffractive DIS has been well understood, the factorisation of the cross-section breaks down in diffractive photoproduction. With matched NLO accuracy, we again are able to provide the first hadron-level simulations of this event class. Validating the simulation against data measured at HERA, we see excellent agreement between theory and experiment.

In diffractive Photoproduction, the breakdown of factorisation is usually suspected to be caused by additional scatterings between the resolved photon and the proton. We therefore extend the MPI modelling to supplement the NLO accurate simulation, by allowing interactions between the resolved photons and the original proton and using these interactions to veto events that do not leave the proton intact. We also summarise findings from other studies and argue that contrary to some other attempts, the factorisation breaking has to be assumed to happen in both the resolved and the direct component. Essentially, the two modes are indistinguishable at NLO and therefore the direct component must also be rescaled. To arrive at the exact mechanism, one relies on experimental data which the EIC will hopefully provide in the next decades.

The simulation is then applied to an EIC setup in both diffractive DIS and diffractive photoproduction regimes. For the latter, we rescaled the direct and resolved component based on a fit of suppression factors to the HERA data. We examined inclusive jet observables, event shapes, and momentum ratios used in previous studies.

Going beyond the work presented here, the phenomenology of lepton-hadron colliders can be improved further in several regards. One of the most interesting questions will be understanding the parton content of virtual photons, and connected to that, the transition between the electroproduction (i.e. DIS) and the photoproduction regime. For this purpose, the picture of the parton content in the photon must be extended, with the primary obstacle being the presence of an additional scale, the photon virtuality, and its interference with the factorisation scale in the evolution equations.

However, as a first step, the phenomenology of photoproduction currently suffers from outdated PDF fits, meaning that a new parametrisation using up-to-date techniques will significantly improve the quality of any theoretical calculation and allow for a better estimation of PDF uncertainty. Our studies showed that the currently available fits are somewhat disparate, making any theory calculation highly dependent on the chosen PDF set.

Lastly, diffraction is an interesting phenomenon at hadron colliders as well and understanding the factorisation breaking in diffractive photoproduction at the EIC will allow also for more informed studies at the LHC.

In summary, while hadron-level predictions are an important baseline for the interpretation of data taken at the EIC and potentially other future experiments, the theoretical description of both diffraction and photoproduction has to be improved in the future.

Bibliography

[1] I. Zurbano Fernandez et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, .

[2] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [`0811.4622`].

[3] SHERPA collaboration, E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [`1905.09127`].

[4] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and collider physics*, vol. 8. Cambridge University Press, 2, 2011, 10.1017/CBO9780511628788.

[5] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995, 10.1201/9780429503559.

[6] J. Campbell, J. Huston and F. Krauss, *The Black Book of Quantum Chromodynamics : a Primer for the LHC Era*. Oxford University Press, 2018, 10.1093/oso/9780199652747.001.0001.

[7] A. Accardi et al., *Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all*, *Eur. Phys. J. A* **52** (2016) 268, [`1212.1701`].

[8] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447, [`2103.05419`].

[9] EUROPEAN MUON collaboration, J. Ashman et al., *A Measurement of the Spin Asymmetry and Determination of the Structure Function g(1) in Deep Inelastic Muon-Proton Scattering*, *Phys. Lett. B* **206** (1988) 364.

[10] R. L. Jaffe and A. Manohar, *The G(1) Problem: Fact and Fantasy on the Spin of the Proton*, *Nucl. Phys. B* **337** (1990) 509–546.

[11] B. Warmbein, *End of an era: HERA switches off*, *CERN Courier* **47** (2007) .

[12] N. Armesto, T. Cridge, F. Giuli, L. Harland-Lang, P. Newman, B. Schmookler et al., *Impact of inclusive electron ion collider data on collinear parton distributions*, *Phys. Rev. D* **109** (2024) 054019, [`2309.11269`].

[13] A. Buckley et al., *General-purpose event generators for LHC physics*, *Phys. Rept.* **504** (2011) 145–233, [`1101.2599`].

[14] Y. V. Kovchegov and E. Levin, *Quantum Chromodynamics at High Energy*, vol. 33. Oxford University Press, 2013, 10.1017/9781009291446.

[15] E. Mirkes and D. Zeppenfeld, *Dijet production at HERA in next-to-leading order*, *Phys. Lett. B* **380** (1996) 205–212, [`hep-ph/9511448`].

[16] Z. Nagy and Z. Trocsanyi, *Multijet cross-sections in deep inelastic scattering at next-to-leading order*, *Phys. Rev. Lett.* **87** (2001) 082001, [`hep-ph/0104315`].

[17] Z. Nagy and Z. Trocsanyi, *Three-jet event-shapes in lepton-proton scattering at next-to-leading order accuracy*, *Phys. Lett. B* **634** (2006) 498–503, [`hep-ph/0511328`].

[18] T. Carli, T. Gehrmann and S. Hoeche, *Hadronic final states in deep-inelastic scattering with Sherpa*, *Eur. Phys. J. C* **67** (2010) 73–97, [`0912.3715`].

[19] J. Currie, T. Gehrmann, E. W. N. Glover, A. Huss, J. Niehues and A. Vogt, *N^3LO corrections to jet production in deep inelastic scattering using the Projection-to-Born method*, *JHEP* **05** (2018) 209, [`1803.09973`].

[20] H1 collaboration, A. Aktas et al., *Measurement of event shape variables in deep-inelastic scattering at HERA*, *Eur. Phys. J. C* **46** (2006) 343–356, [`hep-ex/0512014`].

[21] ZEUS collaboration, S. Chekanov et al., *Forward jet production in deep inelastic ep scattering and low-x parton dynamics at HERA*, *Phys. Lett. B* **632** (2006) 13–26, [`hep-ex/0502029`].

[22] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, [`0903.1219`].

[23] ZEUS collaboration, S. Chekanov et al., *Exclusive photoproduction of J / psi mesons at HERA*, *Eur. Phys. J. C* **24** (2002) 345–360, [`hep-ex/0201043`].

[24] ZEUS collaboration, J. Breitweg et al., *Elastic and proton dissociative ρ^0 photoproduction at HERA*, *Eur. Phys. J. C* **2** (1998) 247–267, [`hep-ex/9712020`].

[25] ZEUS collaboration, J. Breitweg et al., *Measurement of elastic Upsilon photoproduction at HERA*, *Phys. Lett. B* **437** (1998) 432–444, [`hep-ex/9807020`].

[26] ZEUS collaboration, M. Derrick et al., *Observation of hard scattering in photoproduction at HERA*, *Phys. Lett. B* **297** (1992) 404–416.

[27] H1 collaboration, T. Ahmed et al., *Total photoproduction cross-section measurement at HERA energies*, *Phys. Lett. B* **299** (1993) 374–384.

[28] H1 collaboration, C. Adloff et al., *Elastic photoproduction of J / psi and Upsilon mesons at HERA*, *Phys. Lett. B* **483** (2000) 23–35, [`hep-ex/0003020`].

[29] H1 collaboration, I. Abt et al., *Measurement of inclusive jet cross-sections in photoproduction at HERA*, *Phys. Lett. B* **314** (1993) 436–444.

[30] BABAR collaboration, J. P. Lees et al., *Measurement of the $\gamma^\star\gamma^\star \to \eta'$ transition form factor*, *Phys. Rev. D* **98** (2018) 112002, [`1808.08038`].

[31] ALICE, LHCB collaboration, S. Ragoni, *New measurements on diffractive vector mesons*, *PoS* **LHCP2021** (2021) 085, [`2109.03066`].

[32] ALICE collaboration, S. R. Klein, *ρ photoproduction in ALICE*, in *Low-x Workshop 2021*, 12, 2021, `2112.07827`.

[33] STAR collaboration, Z. Tu, *Exclusive J/ photoproduction off deuteron in d+Au ultra-peripheral collisions at STAR*, *PoS* **HardProbes2020** (2021) 100, [`2009.04860`].

[34] OPAL collaboration, G. Abbiendi et al., *Dijet production in photon-photon collisions at s(ee)**(1/2) from 189-GeV to 209-GeV*, *Eur. Phys. J. C* **31** (2003) 307–325, [`hep-ex/0301013`].

[35] OPAL collaboration, G. Abbiendi et al., *Inclusive Jet Production in Photon-Photon Collisions at s(ee)**(1/2) from 189 to 209-GeV*, *Phys. Lett. B* **658** (2008) 185–192, [`0706.4382`].

[36] L3 collaboration, P. Achard et al., *Inclusive jet production in two-photon collisions at LEP*, *Phys. Lett. B* **602** (2004) 157–166, [`hep-ex/0410012`].

[37] OPAL collaboration, G. Abbiendi et al., *Total hadronic cross-section of photon-photon interactions at LEP*, *Eur. Phys. J. C* **14** (2000) 199–212, [`hep-ex/9906039`].

[38] OPAL collaboration, K. Ackerstaff et al., *Inclusive production of charged hadrons and K0(S) mesons in photon-photon collisions*, *Eur. Phys. J. C* **6** (1999) 253–264, [`hep-ex/9808009`].

[39] OPAL collaboration, G. Abbiendi et al., *Inclusive production of D*+- mesons in photon-photon collisions at S**(1/2)(ee) = 183-GeV and 189-GeV and a first measurement of F**gamma(2,c)*, *Eur. Phys. J. C* **16** (2000) 579–596, [`hep-ex/9911030`].

[40] ZEUS collaboration, S. Chekanov et al., *Dijet photoproduction at HERA and the structure of the photon*, *Eur. Phys. J. C* **23** (2002) 615–631, [`hep-ex/0112029`].

[41] ZEUS collaboration, H. Abramowicz et al., *Inclusive-jet photoproduction at HERA and determination of alphas*, *Nucl. Phys. B* **864** (2012) 1–37, [`1205.6153`].

[42] ZEUS collaboration, M. Derrick et al., *Measurement of the F2 structure function in deep inelastic e+ p scattering using 1994 data from the ZEUS detector at HERA*, *Z. Phys. C* **72** (1996) 399–424, [`hep-ex/9607002`].

[43] ZEUS collaboration, J. Breitweg et al., *ZEUS results on the measurement and phenomenology of F(2) at low x and low Q**2*, *Eur. Phys. J. C* **7** (1999) 609–630, [`hep-ex/9809005`].

[44] H1 collaboration, C. Adloff et al., *Measurement and QCD analysis of neutral and charged current cross-sections at HERA*, *Eur. Phys. J. C* **30** (2003) 1–32, [`hep-ex/0304003`].

[45] C. F. von Weizsacker, *Radiation emitted in collisions of very fast electrons*, *Z. Phys.* **88** (1934) 612–625.

[46] E. J. Williams, *Nature of the high-energy particles of penetrating radiation and status of ionization and radiation formulae*, *Phys. Rev.* **45** (1934) 729–730.

[47] V. M. Budnev, I. F. Ginzburg, G. V. Meledin and V. G. Serbo, *The Two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation*, *Phys. Rept.* **15** (1975) 181–281.

[48] M. Gluck, E. Reya and A. Vogt, *Photonic parton distributions*, *Phys. Rev. D* **46** (1992) 1973–1979.

[49] F. Cornet, P. Jankowski, M. Krawczyk and A. Lorca, *A New five flavor LO analysis and parametrization of parton distributions in the real photon*, *Phys. Rev. D* **68** (2003) 014010, [`hep-ph/0212160`].

[50] F. Cornet, P. Jankowski and M. Krawczyk, *A New 5 flavor NLO analysis and parametrizations of parton distributions of the real photon*, *Phys. Rev. D* **70** (2004) 093004, [`hep-ph/0404063`].

[51] W. Slominski, H. Abramowicz and A. Levy, *NLO photon parton parametrization using ee and ep data*, *Eur. Phys. J. C* **45** (2006) 633–641, [`hep-ph/0504003`].

[52] G. A. Schuler and T. Sjostrand, *Low and high mass components of the photon distribution functions*, *Z. Phys. C* **68** (1995) 607–624, [`hep-ph/9503384`].

[53] G. A. Schuler and T. Sjostrand, *Parton distributions of the virtual photon*, *Phys. Lett. B* **376** (1996) 193–200, [`hep-ph/9601282`].

[54] G. A. Schuler and T. Sjostrand, *A Scenario for high-energy gamma gamma interactions*, *Z. Phys. C* **73** (1997) 677–688, [`hep-ph/9605240`].

[55] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [`hep-ph/0603175`].

[56] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, [`2203.11601`].

[57] M. Klasen and G. Kramer, *Inclusive two jet production at HERA: Direct and resolved cross-sections in next-to-leading order QCD*, *Z. Phys. C* **76** (1997) 67–74, [`hep-ph/9611450`].

[58] S. Frixione and G. Ridolfi, *Jet photoproduction at HERA*, *Nucl. Phys. B* **507** (1997) 315–333, [`hep-ph/9707345`].

[59] G. Kramer and B. Potter, *Low Q**2 jet production at HERA in next-to-leading order QCD*, *Eur. Phys. J. C* **5** (1998) 665–679, [`hep-ph/9804352`].

[60] M. Klasen, T. Kleinwort and G. Kramer, *Inclusive Jet Production in γp and $\gamma\gamma$ Processes: Direct and Resolved Photon Cross Sections in Next-To-Leading Order QCD*, *Eur. Phys. J. direct* **1** (1998) 1, [`hep-ph/9712256`].

[61] B. W. Harris and J. F. Owens, *Photoproduction of jets at HERA in next-to-leading order QCD*, *Phys. Rev. D* **56** (1997) 4007–4016, [`hep-ph/9704324`].

[62] V. Guzey and M. Klasen, *Inclusive and Diffractive Dijet Photoproduction at the Electron–Ion Collider in NLO QCD*, *Acta Phys. Polon. Supp.* **16** (2023) 7–A11, [`2303.05182`].

[63] C.-H. Chang, R. Li and J.-X. Wang, *J/psi polarization in photo-production up-to the next-to-leading order of QCD*, *Phys. Rev. D* **80** (2009) 034020, [`0901.4749`].

[64] K. J. Eskola, C. A. Flett, V. Guzey, T. Löytäinen and H. Paukkunen, *Exclusive J/ψ photoproduction in ultraperipheral Pb+Pb collisions at the CERN Large Hadron Collider calculated at next-to-leading order perturbative QCD*, *Phys. Rev. C* **106** (2022) 035202, [`2203.11613`].

[65] K. J. Eskola, C. A. Flett, V. Guzey, T. Löytäinen and H. Paukkunen, *Next-to-leading order perturbative QCD predictions for exclusive J/ψ photoproduction in oxygen-oxygen and lead-lead collisions at energies available at the CERN Large Hadron Collider*, *Phys. Rev. C* **107** (2023) 044912, [`2210.16048`].

[66] K. J. Eskola, C. A. Flett, V. Guzey, T. Löytäinen and H. Paukkunen, *Predictions for exclusive $\varUpsilon$ photoproduction in ultraperipheral Pb $+$ Pb collisions at the LHC at next-to-leading order in perturbative QCD*, *Eur. Phys. J. C* **83** (2023) 758, [`2303.03007`].

[67] S. P. Jones, A. D. Martin, M. G. Ryskin and T. Teubner, *Exclusive J/ψ and Υ photoproduction and the low x gluon*, *J. Phys. G* **43** (2016) 035002, [`1507.06942`].

[68] S. P. Jones, A. D. Martin, M. G. Ryskin and T. Teubner, *Probes of the small x gluon via exclusive J/ψ and Υ production at HERA and the LHC*, *JHEP* **11** (2013) 085, [`1307.7099`].

[69] P. Newman and M. Wing, *The Hadronic Final State at HERA*, *Rev. Mod. Phys.* **86** (2014) 1037, [`1308.3368`].

[70] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization of Hard Processes in QCD*, *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, [`hep-ph/0409313`].

[71] J. C. Collins, *Proof of factorization for diffractive hard scattering*, *Phys. Rev. D* **57** (1998) 3051–3056, [`hep-ph/9709499`].

[72] H1 collaboration, A. Aktas et al., *Tests of QCD factorisation in the diffractive production of dijets in deep-inelastic scattering and photoproduction at HERA*, *Eur. Phys. J. C* **51** (2007) 549–568, [`hep-ex/0703022`].

[73] D. Britzger, J. Currie, T. Gehrmann, A. Huss, J. Niehues and R. Žlebčík, *Dijet production in diffractive deep-inelastic scattering in next-to-next-to-leading order QCD*, *Eur. Phys. J. C* **78** (2018) 538, [`1804.05663`].

[74] H1 collaboration, A. Aktas et al., *Measurement and QCD analysis of the diffractive deep-inelastic scattering cross-section at HERA*, *Eur. Phys. J. C* **48** (2006) 715–748, [`hep-ex/0606004`].

[75] G. Ingelman and P. E. Schlein, *Jet Structure in High Mass Diffractive Scattering*, *Phys. Lett. B* **152** (1985) 256–260.

[76] A. B. Kaidalov, V. A. Khoze, A. D. Martin and M. G. Ryskin, *Factorization breaking in diffractive dijet production*, *Phys. Lett. B* **559** (2003) 235–238, [`hep-ph/0302091`].

[77] A. B. Kaidalov, V. A. Khoze, A. D. Martin and M. G. Ryskin, *Unitarity effects in hard diffraction at HERA*, *Phys. Lett. B* **567** (2003) 61–68, [`hep-ph/0306134`].

[78] A. B. Kaidalov, V. A. Khoze, A. D. Martin and M. G. Ryskin, *Factorization breaking in diffractive dijet photoproduction at HERA*, *Eur. Phys. J. C* **66** (2010) 373–376, [`0911.3716`].

[79] M. Klasen and G. Kramer, *Factorization breaking in diffractive dijet photoproduction*, *Eur. Phys. J. C* **38** (2004) 93–104, [`hep-ph/0408203`].

[80] M. Klasen and G. Kramer, *Review of factorization breaking in diffractive photoproduction of dijets*, *Mod. Phys. Lett. A* **23** (2008) 1885–1907, [`0806.2269`].

[81] R. Zlebcik, K. Cerny and A. Valkarova, *Factorisation breaking in diffractive dijet photoproduction at HERA?*, *Eur. Phys. J. C* **71** (2011) 1741, [`1102.3806`].

[82] S. Höche, *Introduction to parton-shower event generators*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295, 2015, `1411.4085`, DOI.

[83] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, [`1111.5206`].

[84] V. Hirschi and O. Mattelaer, *Automated event generation for loop-induced processes*, *JHEP* **10** (2015) 146, [`1507.00020`].

[85] J. Campbell and T. Neumann, *Precision Phenomenology with MCFM*, *JHEP* **12** (2019) 034, [`1909.09117`].

[86] A. Denner, J.-N. Lang and S. Uccirati, *Recola2: a one-loop matrix-element generator for BSM theories and SM effective field theory*, *PoS* **RADCOR2017** (2017) 019, [`1712.04754`].

[87] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, [`0808.3674`].

[88] F. A. Berends and W. T. Giele, *Recursive Calculations for Processes with n Gluons*, *Nucl. Phys. B* **306** (1988) 759–808.

[89] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [`hep-ph/0409146`].

[90] S. Hoeche, F. Krauss and P. Meinzinger, *Resolved photons in Sherpa*, *Eur. Phys. J. C* **84** (2024) 178, [`2310.18674`].

[91] R. Kleiss and R. Pittau, *Weight optimization in multichannel Monte Carlo*, *Comput. Phys. Commun.* **83** (1994) 141–146, [`hep-ph/9405257`].

[92] S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, *Improving the Weizsacker-Williams approximation in electron - proton collisions*, *Phys. Lett. B* **319** (1993) 339–345, [`hep-ph/9310350`].

[93] "Sherpa Project webpage." `https://sherpa-team.gitlab.io/`.

[94] S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, *Heavy quark production*, *Adv. Ser. Direct. High Energy Phys.* **15** (1998) 609–706, [`hep-ph/9702287`].

[95] S. Catani and M. H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys. B* **485** (1997) 291–419, [`hep-ph/9605323`].

[96] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: A Matrix element generator in C++*, *JHEP* **02** (2002) 044, [`hep-ph/0109036`].

[97] T. Gleisberg and F. Krauss, *Automating dipole subtraction for QCD NLO calculations*, *Eur. Phys. J. C* **53** (2008) 501–523, [`0709.2881`].

[98] F. Buccioni, J.-N. Lang, J. M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866, [`1907.13071`].

[99] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038, [`0709.1027`].

[100] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029, [`hep-ph/0204244`].

[101] S. Frixione, P. Nason and B. R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, *JHEP* **08** (2003) 007, [`hep-ph/0305252`].

[102] S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, [`1111.1220`].

[103] T. Sjostrand and M. van Zijl, *A Multiple Interaction Model for the Event Structure in Hadron Collisions*, *Phys. Rev. D* **36** (1987) 2019.

[104] G. A. Schuler and T. Sjostrand, *Hadronic diffractive cross-sections and the rise of the total cross-section*, *Phys. Rev. D* **49** (1994) 2257–2267.

[105] G. S. Chahal and F. Krauss, *Cluster Hadronisation in Sherpa*, *SciPost Phys.* **13** (2022) 019, [`2203.11385`].

[106] C. Bierlich et al., *Robust Independent Validation of Experiment and Theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026, [`1912.05451`].

[107] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [`1207.1303`].

[108] PDF4LHC WORKING GROUP collaboration, R. D. Ball et al., *The PDF4LHC21 combination of global PDF fits for the LHC Run III*, *J. Phys. G* **49** (2022) 080501, [`2203.05506`].

[109] ATHENA collaboration, J. Adam et al., *ATHENA detector proposal — a totally hermetic electron nucleus apparatus proposed for IP6 at the Electron-Ion Collider*, *JINST* **17** (2022) P10019, [`2210.09048`].

[110] V. M. Budnev, I. F. Ginzburg, G. V. Meledin and V. G. Serbo, *The Two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation*, *Phys. Rept.* **15** (1975) 181–281.

[111] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639–707, [`0803.0883`].

[112] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196, [`1512.01178`].

[113] G. Bewick et al., *Herwig 7.3 Release Note*, `2312.05175`.

[114] S. Gieseke, P. Stephens and B. Webber, *New formalism for QCD parton showers*, *JHEP* **12** (2003) 045, [`hep-ph/0310083`].

[115] S. Platzer and S. Gieseke, *Coherent Parton Showers with Local Recoils*, *JHEP* **01** (2011) 024, [`0909.5593`].

[116] S. Platzer and S. Gieseke, *Dipole Showers and Automated NLO Matching in Herwig++*, *Eur. Phys. J. C* **72** (2012) 2187, [`1109.6256`].

[117] J. Bellm, K. Cormier, S. Gieseke, S. Plätzer, C. Reuschle, P. Richardson et al., *Top Quark Production and Decay in Herwig 7.1*, `1711.11570`.

[118] P. Richardson and S. Webster, *Spin Correlations in Parton Shower Simulations*, *Eur. Phys. J. C* **80** (2020) 83, [`1807.01955`].

[119] S. Plätzer, *Controlling inclusive cross sections in parton shower + matrix element merging*, *JHEP* **08** (2013) 114, [`1211.5467`].

[120] J. Bellm, S. Gieseke and S. Plätzer, *Merging NLO Multi-jet Calculations with Improved Unitarization*, *Eur. Phys. J. C* **78** (2018) 244, [`1705.06700`].

[121] G. A. Schuler and T. Sjostrand, *Gamma gamma and gamma P events at high-energies*, in *Two-Photon Physics from DAPHNE to LEP200 and Beyond*, 3, 1994, `hep-ph/9403393`.

[122] OPAL collaboration, G. Abbiendi et al., *Inclusive production of charged hadrons in photon-photon collisions*, *Phys. Lett. B* **651** (2007) 92–101, [`hep-ex/0612045`].

[123] I. Helenius, *Photon-photon and photon-hadron processes in Pythia 8*, *CERN Proc.* **1** (2018) 119, [`1708.09759`].

[124] ZEUS collaboration, I. Abt et al., *Azimuthal correlations in photoproduction and deep inelastic ep scattering at HERA*, *JHEP* **12** (2021) 102, [`2106.12377`].

[125] PARTICLE DATA GROUP collaboration, P. A. Zyla et al., *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

[126] P. Meinzinger and F. Krauss, *Hadron-level NLO predictions for QCD observables in photo-production at the Electron-Ion Collider*, *Phys. Rev. D* **109** (2024) 034037, [`2311.14571`].

[127] J. C. Collins, *Light cone variables, rapidity and all that*, `hep-ph/9705393`.

[128] M. Gluck, E. Reya and A. Vogt, *Pionic parton distributions*, *Z. Phys. C* **53** (1992) 651–656.

[129] SHERPA collaboration, E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [`1905.09127`].

[130] A. Denner and S. Dittmaier, *Reduction of one loop tensor five point integrals*, *Nucl. Phys. B* **658** (2003) 175–202, [`hep-ph/0212259`].

[131] A. Denner and S. Dittmaier, *Reduction schemes for one-loop tensor integrals*, *Nucl. Phys. B* **734** (2006) 62–115, [`hep-ph/0509141`].

[132] A. Denner and S. Dittmaier, *Scalar one-loop 4-point integrals*, *Nucl. Phys. B* **844** (2011) 199–242, [`1005.2076`].

[133] S. Catani, S. Dittmaier, M. H. Seymour and Z. Trocsanyi, *The Dipole formalism for next-to-leading order QCD calculations with massive partons*, *Nucl. Phys. B* **627** (2002) 189–265, [`hep-ph/0201036`].

[134] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132, [`1412.7420`].

[135] F. D. Aaron et al., *Measurement of the cross section for diffractive deep-inelastic scattering with a leading proton at HERA*, *Eur. Phys. J. C* **71** (2011) 1578, [`1010.1476`].

[136] H1 collaboration, V. Andreev et al., *Diffractive Dijet Production with a Leading Proton in ep Collisions at HERA*, *JHEP* **05** (2015) 056, [`1502.01683`].

[137] ZEUS collaboration, S. Chekanov et al., *Diffractive photoproduction of dijets in ep collisions at HERA*, *Eur. Phys. J. C* **55** (2008) 177–191, [`0710.1498`].

[138] H1 collaboration, F. D. Aaron et al., *Diffractive Dijet Photoproduction in ep Collisions at HERA*, *Eur. Phys. J. C* **70** (2010) 15–37, [`1006.0946`].

[139] N. Armesto, P. R. Newman, W. Slominski and A. M. Stasto, *Diffractive longitudinal structure function at the Electron Ion Collider*, *Phys. Rev. D* **105** (2022) 074006, [`2112.06839`].