Measurement and NNLO QCD analysis of jet production in deep inelastic scattering at ZEUS

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik
und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von
Florian Lorkowski

Hamburg
2023

Abstract

A measurement of inclusive jet cross sections in high-Q^2 neutral current deep-inelastic scattering is performed using data collected by the ZEUS detector at the HERA $e^{\pm}p$ collider. The data were taken in the years 2004 to 2007 at a centre-of-mass energy of 318 GeV and correspond to an integrated luminosity of $347 \, \mathrm{pb}^{-1}$. Jets are reconstructed using the $k_{\perp}$-algorithm in the Breit reference frame. They are measured as a function of the squared four-momentum transfer Q^2 and the transverse momentum of the jets in the Breit frame $p_{\perp,\mathrm{Breit}}$. The measurement is performed in the kinematic region of $150 \, \mathrm{GeV}^2 < Q^2 < 15000 \, \mathrm{GeV}^2$ and $0.2 < y < 0.7$, with y being the inelasticity. Jets are required to fulfil the conditions $7 \, \mathrm{GeV} < p_{\perp,\mathrm{Breit}} < 50 \, \mathrm{GeV}$ and $-1 < \eta_{\mathrm{lab}} < 2.5$, where η_{lab} is the pseudorapidity in the laboratory frame.

The measured jet cross sections are compared to previous measurements and perturbative QCD predictions. The measurement is used in a next-to-next-to-leading-order QCD analysis to simultaneously determine the parton distribution functions of the proton and the strong coupling constant. A significant improvement in accuracy is observed compared to similar determinations. The physical scale dependence of the strong coupling is investigated.

Zusammenfassung

Eine Messung der inklusiven Jet-Wirkungsquerschnitte bei hohen Q^2 in tiefinelastischer Streuung wird durchgeführt unter Verwendung von Daten des ZEUS Detektors am HERA $e^{\pm}p$ Beschleuniger. Die Daten wurden in den Jahren 2004 bis 2007 bei einer Schwerpunktsenergie von 318 GeV aufgezeichnet und entsprechen einer integrierten Luminosität von $347 \, \mathrm{pb}^{-1}$. Jets werden mit dem $k_{\perp}$-Algorithmus im Breit-Bezugssystem rekonstruiert. Sie werden gemessen als Funktion des Quadrats des Viererimpulsaustauschs Q^2 und des Transversalimpulses im Breit-System $p_{\perp,\mathrm{Breit}}$. Die Messung wird durchgeführt in der kinematischen Region $150 \, \mathrm{GeV}^2 < Q^2 < 15000 \, \mathrm{GeV}^2$ und $0.2 < y < 0.7$, wobei y die Inelastizität ist. Jets werden betrachtet in der Region $7 \, \mathrm{GeV} < p_{\perp,\mathrm{Breit}} < 50 \, \mathrm{GeV}$ und $-1 < \eta_{\mathrm{lab}} < 2.5$, mit der Pseudorapidität im Laborsystem η_{lab}.

Die gemessenen Wirkungsquerschnitte werden mit vorhergehenden Messungen sowie mit störungstheoretischen QCD-Vorhersagen verglichen. Die Messdaten werden benutzt um gleichzeitig die Parton-Verteilungsfunktionen des Protons und die starke Kopplungskonstante zu bestimmen. Eine signifikant verbesserte Präzision wird erzielt im Vergleich zu ähnlichen Bestimmungen. Die physikalische Skalen-Abhängigkeit der starken Kopplung wird untersucht.

Contents

Abbreviations

Theory and calculations

MC	Monte Carlo (see p. 3)
NC	Neutral current (see p. 9)
CC	Charged current (see p. 9)
DIS	Deep inelastic scattering (see p. 11)
PHP	Photoproduction (see p. 11)
QED	Quantum electrodynamics (see p. 11)
QPM	Quark-parton model (see p. 13)
QCD	Quantum chromodynamics (see p. 15)
PDF	Parton distribution function (see p. 24)
NLO	Next-to-leading order (see p. 42)
NNLO	Next-to-next-to-leading order (see p. 42)
ISR	Initial-state radiation (see p. 66)
FSR	Final-state radiation (see p. 66)

Detector components and systems

CAL	High-resolution uraninum scintillator calorimeter (see p. 33)
RCAL	Rear section of CAL (see p. 36)
BCAL	Barrel section of CAL (see p. 36)
FCAL	Forward section of CAL (see p. 36)
EMC	Electromagnetic section of RCAL/BCAL/FCAL (see p. 36)
HAC	Hadronic section of RCAL/BCAL/FCAL (see p. 36)
HES	Hadron-electron separator layer of RCAL EMC (see p. 36)

Tracking detectors

CTD	Central tracking detector (see p. 33)
MVD	Microvertex detector (see p. 33)
SRTD	Small-angle rear tracking detector (see p. 33)
STT	Straw-tube tracker (see p. 33)

Luminosity and polarisation measurement

PCAL	Photon calorimeter (see p. 37)
SPEC	Luminosity spectrometer (see p. 38)
TPOL	Transverse polarimeter (see p. 33)
LPOL	Longitudinal polarimeter (see p. 33)

Trigger system

FLT	First-level trigger (see p. 38)
SLT	Second-level trigger (see p. 38)
TLT	Third-level trigger (see p. 39)

Symbols

ϕ	Azimuthal angle (see p. 26)
x_{Bj}	Bjorken scaling variable (see p. 10)
s	Centre-of-mass energy squared (see p. 10)
E_e	Energy of lepton beam (see p. 32)
E_P	Energy of proton beam (see p. 32)
E_e'	Energy of scattered lepton (see p. 10)
E_p'	Energy of scattered hadronic system (see p. 61)
μ_{f}	Factorisation scale (see p. 25)
k	Four-momentum of the incoming lepton (see p. 9)
k'	Four-momentum of the scattered lepton (see p. 9)
P	Four-momentum of the incoming hadron (see p. 9)
P'	Four-momentum of the scattered hadronic system (see p. 9)
p'	Four-momentum of the hadronic system from the hard scattering (see p. 13)
q	Four-momentum of the exchanged virtual boson (see p. 10)
y	Inelasticity (see p. 10)
L	Integrated luminosity (see p. 2)
σ	Interaction cross section (see p. 2)
ξ	Momentum fraction of struck quark (see p. 13)
Q^2	Momentum transfer or boson virtuality (see p. 10)
f_i	Parton momentum density (see p. 14)
η	Pseudorapidity (see p. 26)
θ	Polar angle of the scattered lepton (see p. 10)
γ	Polar angle of the scattered hadronic system (see p. 61)
Λ_{QCD}	QCD scale (see p. 10)
μ_{r}	Renormalisation scale (see p. 23)
α_{s}	Strong coupling (see p. 23)
$p_\perp$	Transverse momentum (see p. 27)

CHAPTER 1

Introduction

The ultimate aim of any domain of science is to describe the natural world accurately and precisely. Among these fields, particle physics stands out for its remarkable ability to explain phenomena within its realm. Since it is the most fundamental domain of physical science, many observations can be accurately explained from first principles. Consequently, predictions can be made and tested to an astonishing degree of precision, which is exemplified by the electron magnetic moment,[1] the Lamb shift in hydrogen[2] or the fine-structure constant.[3] The most outstanding achievement of this field to date is the Standard Model of particle physics.[4] It describes all known non-gravitational interactions of subatomic particles. Some of its achievements include the prediction of electroweak unification,[5–7] the existence of the Higgs boson,[8–11] the explanation of the short range of the strong interaction[12] as well as the prediction of the properties of many particles.[13–17] In principle, almost any observed physical phenomenon can be explained, though sometimes very indirectly, by either the Standard Model or the theory of general relativity,[18] which describes gravity at macroscopic length scales.

Nevertheless, the Standard Model is not a complete description of nature. So far, it could not be unified with general relativity, such that no entirely consistent theory of gravity at microscopic length scales exists. There are also several phenomena that the Standard Model cannot explain, such as dark matter,[19] the smallness of neutrino masses,[20] the amount of matter-antimatter asymmetry in the universe[21] or the strong CP-problem.[22] Due to these issues, many modern studies focus on theories beyond the Standard Model.[23]

However, the Standard Model has made many accurate predictions, which is not by coincidence. Any generalised theory must necessarily reproduce the Standard Model in some limiting case. No matter what future research may uncover, this model will remain an accurate description of an extensive range of physical phenomena. Therefore, precision measurements of the Standard Model are highly relevant and will remain so in the future.

One aspect of the Standard Model that remains under intense investigation is the strong interaction. This interaction gives rise to many complex physical phenomena, including the proton's existence, composition and properties. Collider experiments involving protons are therefore ideally suited to study the strong interaction. Electron-proton collisions are particularly useful, as the electron serves as a well-understood probe of the proton. Using electrons as a probe also produces a much cleaner experimental picture than in, e.g. proton-proton collisions.

The HERA accelerator at DESY was the world's only high-energy electron-proton collider so far.[24,25] This circular collider accelerated beams of electrons or positrons and protons in opposing directions, which made it possible to observe collisions at a centre-of-mass energy of

up to $318\,\text{GeV}$. This thesis presents a measurement of jet cross sections in electron-proton and positron-proton collisions. Jets are experimental signatures that are created by strongly interacting particles. The measurement is performed using data from the ZEUS detector, which was one of the two general-purpose particle detectors at HERA.[25] The analysis is the most important still missing jet measurement left at HERA. It uses about two-thirds of the entire dataset collected at ZEUS – the most extensive dataset that was not yet entirely analysed in this regard. Within this thesis, the cross sections are used to determine the strong coupling constant, one of the fundamental parameters of the Standard Model. As such, the measurement provides valuable information for further studies of the strong interaction.

There have been previous attempts at performing this analysis, but they have not led to a publication.[26,27] This work supersedes those attempts and improves upon their procedures by adding additional corrections, a multidimensional matrix-unfolding method and a more careful and complete estimation of systematic uncertainties.

1.1 Cross-section measurements in collider physics

Being the most fundamental branch of physics, particle physics finds itself in a difficult situation concerning the comparison of theory and experiment. Theories are constructed at the most fundamental level, yet experiments must be designed using macroscopic detectors. This leaves few relevant quantities that are accessible via both approaches. The interaction cross section σ is one such quantity. The cross section for a given process is defined according to

$$N = \sigma \cdot L, \tag{1.1}$$

where L is the integrated luminosity and N is the number of times the specific process occurred. The luminosity is a measure of the intensity and the focus of the incoming particles. For the remainder of this section, it is assumed to be known.

In an idealised, perfect detector that unambiguously reconstructs the entire interaction, equation (1.1) would be enough to measure the cross section. The measurement would consist of simply counting the number of occurred events N. For a real detector, however, three kinds of complications occur.

1. **Detector effects**   A real detector is never perfect. For example, particles might lose energy or get scattered as they pass through inactive material such as support structures or electronics before being measured. Some particles might not get measured at all, for example, because their energy is too low or they escape through gaps in the detector, such as the beam pipe. Particles can get misidentified because their measured properties resemble another particle. As with any measurement apparatus, the recorded quantities in a particle detector are subject to statistical fluctuations. Generally, a measured distribution might differ significantly in shape and normalisation from its underlying detector-independent counterpart.

2. **Statistical nature of the scattering process**   The scattering of particles is described by quantum mechanics and is, therefore, probabilistic. The expected number of occurrences of a process is given by equation (1.1), but the actual number of occurrences is random. Thus, there is always a statistical uncertainty associated with the final result. The actual number of times an event occurs is distributed according to a Poisson distribution with an expectation value of N.

3. **Background processes** By the time the final-state particles reach the detector, what has happened before is often ambiguous. Different physical processes might lead to identical final-state particles. Particles that are crucial in identifying a given process can get misreconstructed. Interactions unrelated to the nominal collision can cause signals in the detector. Therefore, a completely different process might resemble the investigated process.

Due to these effects, the actual number of events that occurred is not directly measurable. While statistical fluctuations can only be dealt with by estimating an uncertainty, detector effects and background contributions can and have to be corrected to obtain an accurate result.

In particle physics, this correction is usually done using event samples generated by a Monte Carlo (MC) simulation. A MC simulation consists of the event generation and the detector simulation. The event generation simulates the collision of the incoming particles and all subsequent decays before the measurement in the detector. The detector simulation describes the interaction of the final-state particles with active and inactive components of the detector.

The quantities computed by a MC event generator are usually called generator-level or truth-level quantities. They correspond to the underlying detector-independent quantities that are to be measured. After processing these MC samples in the detector simulation, they are called detector-level quantities. They correspond to the measured data from the real detector.

The event generation uses the most accurate available theoretical models for the interaction, and the detector simulation uses thoroughly established principles from atomic and solid-state physics. However, due to approximations in practical calculations and uncertainties on their input parameters, it has to be assumed that the simulation is not entirely correct. Deriving corrections from MC simulations and applying them to the data tends to introduce biases. Care must be taken to minimise these biases and to quantify the remaining uncertainties. A large fraction of the effort of making a particle physics measurement goes towards verifying and correcting the event generation and detector simulation in the MC sample.

Another aspect that makes MC simulations useful is that the number of events is accessible both at generator and detector level. If the detector simulation describes the real detector sufficiently well, the relation between these two quantities will, within uncertainties, be the same in data and MC, even if the event generation is not entirely correct. This makes it possible to estimate the number of events at truth level in the data from the measured number of events.

Thus, a measurement of a cross section consists of the following steps. Here, it is assumed that the data from the detector is already available and that MC samples have been generated and processed by the detector simulation.

1. Reconstruct events from measured data and perform the event selection. The event selection removes events that fall outside the desired analysis phase space at detector level or are otherwise problematic. This could include, for example, phase space regions where the detector resolution is particularly bad, regions that are not well described by the MC event generation or regions with a large contribution from background events.

2. Apply the same reconstruction and event selection to the detector-level MC samples. At this point, it is possible to compare various distributions between data and MC to determine how well the simulation describes the data. If the agreement is unsatisfactory, correction procedures must be applied, or the event selection must be refined.

3. Apply generator-level event selection to the MC samples. Here, a different selection is employed. Events are only discarded if they fall outside the analysis phase space. This step constitutes the definition of the measured cross sections.

4. At this point, three samples are available: detector-level data, detector-level MC and generator-level MC. These samples are then used as input to a so-called unfolding procedure that estimates the truth-level data distributions, from which the cross sections can be derived.

5. During the first two steps, additional selection criteria and correction procedures are introduced, which have uncertainties associated with them. The influence of these uncertainties on the final result is quantified, e.g. by changing numerical values in the procedure or by using an alternative procedure. The sum of these variations will become part of the systematic uncertainty of the measured cross sections.

After discussing the theoretical framework of the measurement, the ZEUS detector and related measurements in chapters 2, 3 and 4, the MC samples employed in this thesis are introduced in chapter 5. The event reconstruction and selection at detector level (step 1) are described in chapters 6 and 7, respectively. Differences between data and MC are discussed and corrected (step 2) in chapter 8. Event selection at generator level (step 3) is presented in section 7.4. The unfolding (steps 4) is performed in chapter 9. The estimation of uncertainties (step 5) is described in chapter 10. Finally, the measured cross sections are presented in chapter 11.

1.2 Interpretation of a cross-section measurement

The measured cross sections refer to processes almost exclusively found in particle colliders. As such, they often have no direct application in any applied science or engineering. Their practical value lies in their ability to discriminate and constrain different models of nature. These models can then potentially be used to make predictions for technically relevant scenarios or to gain insights into other domains of physics. The measured cross sections are defined in such a way that they can be calculated from theoretical models. By doing so with different models or different model parameters and comparing the predictions to the measurement, it is possible to exclude specific models or constrain their parameter ranges.

This fact leads to another important use of cross section measurements. They can be used to determine the parameters within a given model. The model parameters are iteratively adjusted, and new theoretical cross sections are computed at each step. The set of parameters whose corresponding cross section shows the best agreement with the measurement is the preferred choice of model parameters, given the considered experimental data.

The cross sections measured in this thesis are expected to be described by the Standard Model. In chapter 12, they will be used to determine the strong coupling constant, which is the most relevant parameter of the Standard Model for the investigated scattering process. At the same time, this is also a test of the Standard Model. If no value of the strong coupling constant can be found that gives a satisactory description of the data, this could hint at physics beyond the Standard Model. Furthermore, every measurement within the Standard Model is simultaneously a constraint on new physics models. Any model that predicts deviations from the Standard Model in the experimentally verified regimes can be excluded.

CHAPTER 2

# Theory of deep inelastic scattering at HERA

The desire to understand the structure of matter is as old as mankind itself. The idea of a smallest, indivisible unit of matter dates back at least to ancient Greece.[28] The first empirical hints of such an object were described by Daniel Bernoulli in 1739. He found that the behaviour of gases could be very well explained by describing them as a collection of submicroscopic particles that move according to the laws of statistical mechanics.[29] In the following two centuries, the existence of atoms and later electrons was thoroughly established. Another significant advancement was achieved by Ernest Rutherford in 1911, as he analysed the results of the Geiger-Marsden experiment in which α-particles were scattered on a thin gold foil. He observed that the scattering cross section does not correspond to that of an extended atom, as was expected at the time, but rather to a point-like object at its centre, thus discovering the nucleus. After further observations of the scattering of α-particles, Rutherford concluded that there is a smallest constituent of the nucleus, the hydrogen ion, or as Rutherford called it, the proton.[30,31]

Striving for an even deeper understanding of the structure of matter, further experiments would soon be carried out, in which the scattering off protons was investigated. Due to their point-like nature, electrons are an ideal probe to investigate the proton. The most powerful of these experiments to date were located at the HERA accelerator, in which beams of protons and electrons were accelerated and brought to collision.

2.1 The Standard Model

The Standard Model of particle physics is a theory that describes all currently known elementary particles and all of their interactions, except gravity.[4] An overview of the Standard Model is given in figure 2.1. The model describes 17 elementary particles: twelve fermions, four vector bosons and one scalar boson. These particles interact via the electromagnetic, the weak nuclear and the strong nuclear interactions.[17] In the figure and the remainder of this thesis, natural units will be employed with the convention $\hbar = c = 1$.

Fermions are particles with half-integer spin. According to the spin-statistics theorem, they respect the Pauli exclusion principle. This means that two identical fermions cannot occupy the same state, which is an important property related to the stability of matter.[32–34] All elementary fermions are of spin $\frac{1}{2}$. They can be categorised into three generations. Each generation contains one up-type quark with electric charge $+\frac{2}{3}e$, one down-type quark with charge $-\frac{1}{3}e$, a charged lepton with charge $-e$ and an electrically neutral neutrino, where e is the elementary charge. Each fermion has a corresponding antiparticle with an identical mass,
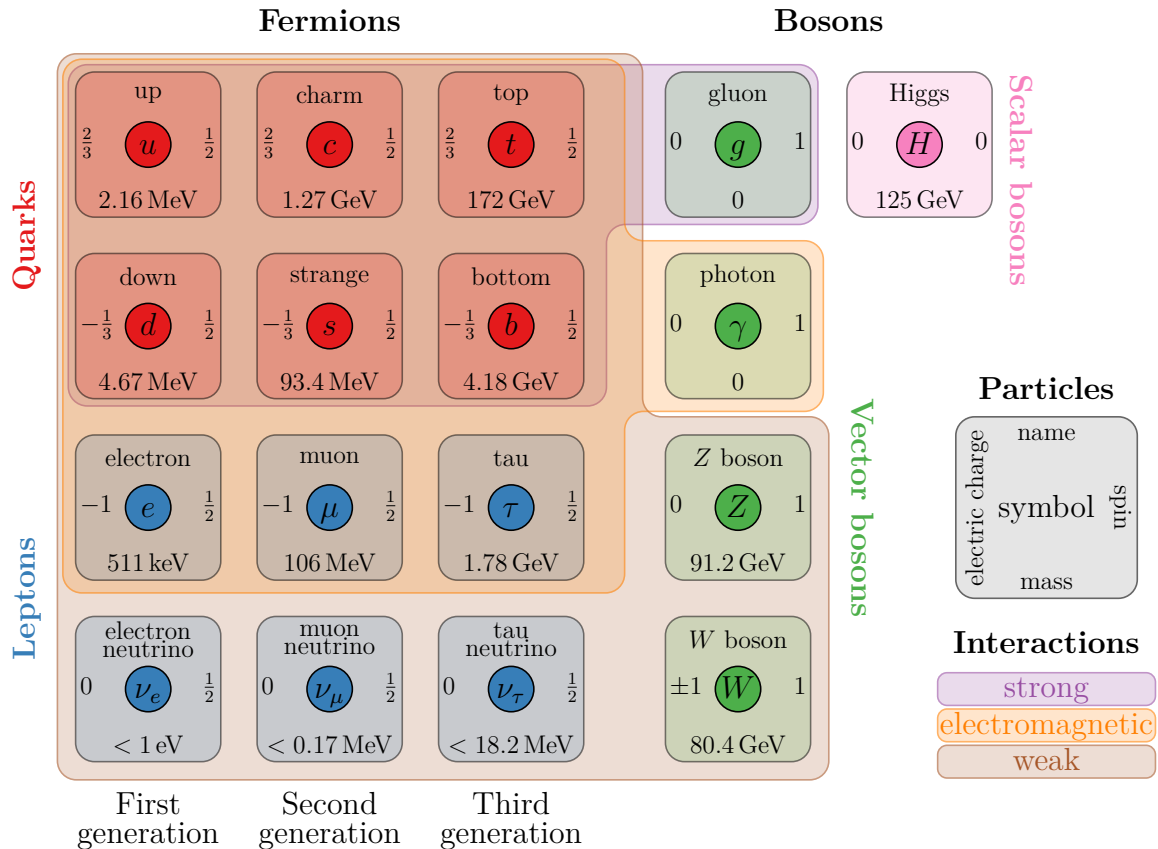
Figure 2.1: An overview of the Standard Model of particle physics. The elementary particles are shown alongside their spin, their electric charge in units of the elementary charge and their mass.[17] The shaded areas indicate the interactions between the particles as mediated by the respective bosons.

but its electric charge and other quantum numbers are the opposite compared to the regular particle. Each quark exists in three variants, distinguished by their so-called colour charge. The most relevant fermions for this thesis are the quarks and antiquarks and the electron and its antiparticle, the positron. The six types of quarks are referred to as different quark flavours. Similarly, the six lepton types are known as different lepton flavours.

Bosons are particles with integer spin. They can be distinguished into scalar bosons with spin zero, vector bosons with spin one or tensor bosons with higher spin. The Standard Model contains four types of elementary vector bosons. They are also referred to as gauge bosons since their existence as mediators of the interactions of the Standard Model is predicted by the corresponding gauge theories. The scalar Higgs boson was introduced to explain the masses of the heavy gauge bosons.

The simplest interaction of the Standard Model is electromagnetism, which is also the only long-range interaction in this model. This interaction is based on a $U(1)_{\mathrm{EM}}$ gauge symmetry acting on the electric charge. It is mediated by the photon and affects all particles with an electric charge. The weak interaction is mediated by the W and Z bosons and affects all fermions. It couples differently to left- and right-handed particles, making it the only interaction in the Standard Model that violates parity and charge-parity symmetry.[35] It is also

the only interaction that can change the flavour of a particle, e.g. it can change an electron into a neutrino.

At energies above $\sim 246\,\text{GeV}$, the electromagnetic and weak interactions unify into the electroweak interaction.[5–7] This interaction is described by a $SU(2)_\text{L} \times U(1)_Y$ gauge theory, where $SU(2)_\text{L}$ acts on the weak isospin of left-handed particles and right-handed antiparticles, and $U(1)_Y$ acts on the weak hypercharge Y. At energies below this scale, the electroweak symmetry is spontaneously broken to $U(1)_\text{EM}$, leading to the emergence of the photon and the weak bosons. In a gauge invariant theory, gauge bosons are expected to be massless, but the weak bosons were found to be among the most massive particles in the Standard Model. The Higgs mechanism was introduced to resolve this issue.[8–11] This required the existence of the scalar Higgs field and the associated Higgs boson.

The strong interaction is mediated by the gluons and only affects the quarks. It is governed by a $SU(3)_\text{c}$ gauge symmetry, acting on the colour charge. Quarks carry a colour charge that can take the values red, green and blue, while antiquarks can have the colours anti-red, anti-green and anti-blue. Correspondingly, the quarks and antiquarks transform under the fundamental and conjugate representations of the $SU(3)$ group, respectively. Gluons transform under the adjoint representation, meaning they carry a combination of a colour and an anti-colour charge, leading to eight distinct gluons since the colour-neutral combination does not exist. When multiple quarks form a composite particle, it is assigned a colour according to additive colour mixing of elementary colour theory. Quarks only appear in colourless, i.e. white bound states called hadrons. If a hadron is made from an even number of quarks, it will have an integer spin and thus be a boson. Such a particle is called a meson. A hadron made from an odd number of quarks and behaves fermionically is called a baryon. The most important hadron is the proton, which is made from two up quarks and one down quark.[4] The charm, bottom and top quarks are known as the heavy quark flavours, while the up, down and strange quarks are the light flavours.

2.2 Kinematics of lepton-hadron scattering

In the Standard Model, the scattering of leptons off hadrons is mediated by a gauge boson. If this boson is an electrically neutral photon or a Z boson, the process is referred to as neutral current (NC). In charged current (CC) processes, a charged $W^\pm$ boson is exchanged. In the following, the interactions of the lepton are assumed Standard-Model-like via a single boson. No assumptions are made about the hadronic interaction. The general graph of a lepton-hadron scattering process $e(k) + P(P) \rightarrow e'(k') + P'(P')$ is shown in figure 2.2. A lepton e interacts with a hadron P via the exchange of a boson γ. The final state consists of the scattered lepton e' and the hadronic final state P', consisting of one or more particles. In the context of this thesis, the incoming and scattered leptons are usually electrons or positrons, and the exchanged boson is usually a photon. The formulae described in this section also apply when considering neutrinos as external particles or weak-boson exchange.

The final-state four-momenta k' and P' possess eight components, four of which are constrained by energy-momentum conservation, one is constrained by requiring the scattered lepton to be on its mass-shell, and one corresponds to the azimuthal symmetry around the beam axis. The hadronic system may consist of multiple particles, so it cannot be assigned a definite mass. Two degrees of freedom remain necessary to describe the event. In early lepton-hadron scattering experiments, the initial-state hadron was usually at rest in the laboratory frame. Observables were the polar angle θ and the energy E'_e of the scattered lepton.[30] In modern
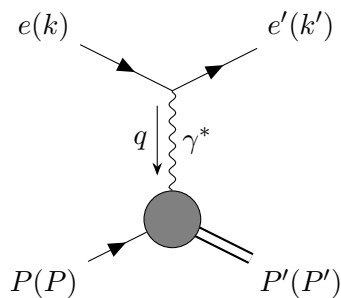
Figure 2.2: General graph of a lepton-hadron scattering process at leading order in electroweak theory on the lepton side. Time runs from left to right.

collider experiments, both particles start with non-zero and non-equal momenta. The angle and energy of both the scattered lepton and the hadronic system are measurable.

Instead of energies or angles, the event is more universally described by a set of Lorentz invariant quantities. For lepton-hadron scattering, the following four quantities are commonly used.[36] The centre-of-mass energy $\sqrt{s}$ is defined as

$$\sqrt{s} = \sqrt{(k + P)^2}. \tag{2.1a}$$

At accelerator experiments, the centre-of-mass energy is known and constant, as it only depends on the initial-state momenta.

The momentum transfer or boson virtuality Q^2 is defined as

$$Q^2 = -q^2 = -(k - k')^2. \tag{2.1b}$$

The negative sign is introduced as q^2 is always negative. Especially interesting for the investigation of the structure of the hadron is the regime where Q^2 is large compared to the so-called QCD scale $\Lambda_{\text{QCD}}^2 \approx (200\,\text{MeV})^2$. The QCD scale is the energy above which the strong interaction weakens, eventually leading to asymptotic freedom, see section 2.4.4. In this region, the exchanged boson acts as a point-like probe of the hadron, such that it can be probed at length scales below about 1 fm.

The Bjorken scaling parameter x_{Bj} and the inelasticity y are defined as

$$x_{\text{Bj}} = \frac{Q^2}{2P \cdot q}, \tag{2.1c}$$

$$y = \frac{P \cdot q}{P \cdot k}. \tag{2.1d}$$

As discussed in the next section, x_{Bj} has a clear physical interpretation in the quark-parton model, one of the simplest models of lepton-hadron scattering. This simple correspondence is lost, however, when considering corrections due to the strong interaction. In the fixed-target frame, y is the fraction of the lepton's energy transferred to the hadron. Neglecting the masses of the initial-state particles, these four quantities are connected by the relation $Q^2 = x_{\text{Bj}}ys$, which leaves again two quantities necessary to describe an event. Analysing their expressions, one finds that these variables can take values in the range $x_{\text{Bj}}, y \in [0, 1]$ and $Q^2 \in [0, s]$.

Based on the boson virtuality, two kinematic regimes can be distinguished. In the photoproduction (PHP) regime $Q^2 \lesssim \Lambda_{\text{QCD}}^2$ and the exchanged photon is almost real. In contrast, if $Q^2 \gg \Lambda_{\text{QCD}}^2$, the process is referred to as deep inelastic scattering (DIS). In this regime, the interaction can be interpreted as a hard scattering of the point-like boson on the hadron.[37]

2.3 Regimes of electron-proton scattering

The following section highlights important steps in the historical understanding of the proton structure. As higher energy measurements became available, increasingly sophisticated models of the proton could be constructed. In the following, the term 'electron' will denote both electrons and positrons.

2.3.1 Elastic scattering on a point-like proton

In very low energy electron-proton scattering experiments, the proton appears like a point-like elementary particle that can be described by the laws of quantum electrodynamics (QED).[38] The corresponding Feynman graph is shown in the left panel of figure 2.3. The mass of the final-state proton P' is known, such that only one degree of freedom remains for each event. For the kinematic variables, one finds

$$x_{\mathrm{Bj}} = \frac{Q^2}{2P \cdot q} = \frac{Q^2}{2P \cdot (P' - P)} = \frac{Q^2}{-(P'-P)^2 - \underbrace{P^2}_{=m_P^2} + \underbrace{P'^2}_{=m_P^2}} = \frac{Q^2}{-q^2} = 1, \qquad (2.2)$$

where m_P is the mass of the proton. The scattering cross section σ is often described using the leptonic and hadronic tensors $L^{\mu\nu}$ and $H^{\mu\nu}$

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{\pi\alpha^2}{Q^4 s^2} L^{\mu\nu} H_{\mu\nu}, \qquad (2.3)$$

with $\alpha = \frac{e^2}{4\pi}$ and e being the elementary charge.[36] In the theory of QED, expressions for the leading contributions to the leptonic and hadronic tensors can be directly derived from first principles

$$L^{\mu\nu} = 2k^\mu k^\nu - (k^\mu q^\nu + k^\nu q^\mu) + (k \cdot q)g^{\mu\nu}, \qquad (2.4a)$$

$$H^{\mu\nu} = Q^2 \left[\frac{1}{2}\left(\frac{q^\mu q^\nu}{q^2} - g^{\mu\nu} \right) + \frac{1}{P \cdot q}\left(P^\mu + \frac{1}{2}q^\mu \right)\left(P^\nu + \frac{1}{2}q^\nu \right) \right] + \mathcal{O}\left(\frac{m_P^2}{Q^2} \right). \qquad (2.4b)$$

Combining these expressions leads to the scattering cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{2\pi\alpha^2}{Q^4} Y_+ + \mathcal{O}\left(\frac{m_P^2}{Q^2} \right), \qquad (2.5)$$

with $Y_\pm = 1 \pm (1-y)^2$. The proton mass may not be neglected in fixed-target or low-energy collider experiments. In these scenarios, target mass corrections and higher-twist terms must be included. A convenient way to treat target mass corrections is through the Nachtmann variable, which is related to x_{Bj} but includes mass terms of the proton.[39] This thesis focuses on the high-energy regime, where these effects can be neglected. Therefore, terms involving the proton mass are not evaluated explicitly here.

2.3.2 Elastic scattering on a structured proton

As higher-energy experiments were conducted, it became increasingly apparent that the proton is not a point particle. Starting in 1956, Robert Hofstadter published a series of measurements in which he scattered electrons of a few hundred MeV on stationary protons.[40] The results
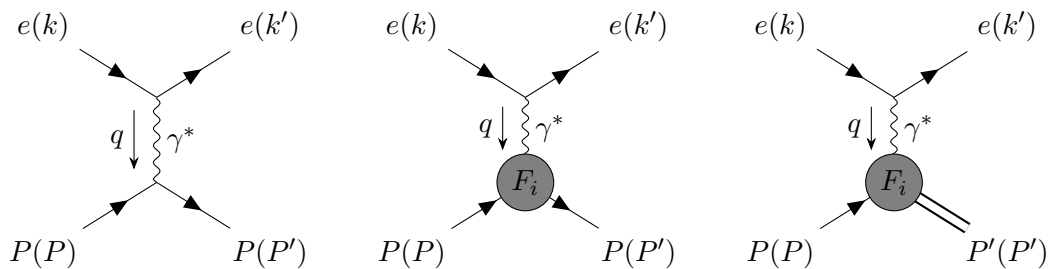
Figure 2.3: Electron-proton scattering in different regimes. All graphs are shown at leading order in electroweak theory on the electron side. Left: elastic scattering on a point-like proton. Centre: elastic scattering on a structured proton. Right: inelastic scattering on a structured proton.

showed a clear deviation from the predictions for a proton with a point-like charge and a point-like magnetic moment. This suggested that the proton has a non-trivial internal structure. A generalised form of equation (2.5) can be derived by using the most general expression for the hadronic tensor that is allowed by Lorentz- and gauge invariance and current and parity conservation

$$
H^{\mu\nu} = Q^2 \left[W_1(Q^2) \left(\frac{q^\mu q^\nu}{q^2} - g^{\mu\nu} \right) + \frac{W_2(Q^2)}{P \cdot q} \left(P^\mu + \frac{1}{2} q^\mu \right) \left(P^\nu + \frac{1}{2} q^\nu \right) \right] + \mathcal{O}\left(\frac{m_P^2}{Q^2} \right),
$$
(2.6)

with W_1, and W_2 being the structure functions of the proton.[36] They are real scalar functions of the available Lorentz-invariant quantities and parameterise the interactions of the proton. The corresponding cross section can be written as

$$
\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{Q^4} \left(y^2 F_1(Q^2) + (1-y) F_2(Q^2) \right).
$$
(2.7)

The redefined structure functions $F_{1/2}$ are functions of $W_{1/2}$. Comparing to equation (2.5), one finds that the model of a point-like proton predicts $F_1 = 1/2$ and $F_2 = 1$. The graph corresponding to elastic electron-proton scattering with a structured hadron is depicted in the middle panel of figure 2.3.

2.3.3 Inelastic scattering

Another leap in energy was achieved by the Stanford Linear Accelerator Center, which managed to accelerate electrons to a $\mathcal{O}(10\,\mathrm{GeV})$ by 1969. For the first time, inelastic scattering could be observed, in which the proton broke apart,[41] as illustrated in the right panel of figure 2.3. As the hadronic system P' may now consist of multiple particles, its invariant mass becomes an additional degree of freedom, such that two quantities are necessary to describe the event. Equation (2.2) no longer holds, such that in general $x_{\mathrm{Bj}} \neq 1$. The cross section is double differential and equation (2.7) becomes

$$
\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{Bj}} \mathrm{d}Q^2} = \frac{4\pi\alpha^2}{x_{\mathrm{Bj}} Q^4} \left(x_{\mathrm{Bj}} y^2 F_1(x_{\mathrm{Bj}}, Q^2) + (1-y) F_2(x_{\mathrm{Bj}}, Q^2) \right).
$$
(2.8)

At these energies, terms involving the proton mass and higher-twist terms can be neglected.[36] This also includes terms involving the mass of the scattered hadronic system, as they only appear together with the proton mass.
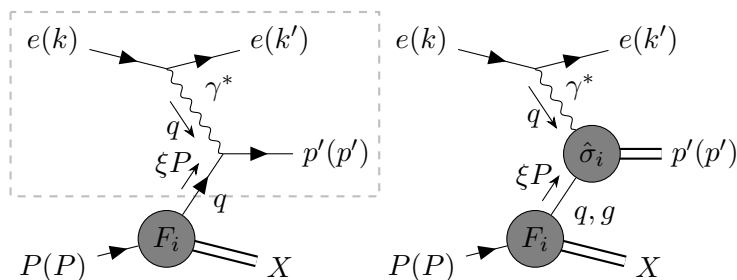
Figure 2.4: Models of inelastic electron-proton scattering. Left: quark-parton model. The grey box indicates the electron-quark interaction. Right: general graph that is also valid for quantum chromodynamics.

2.3.4 The quark-parton model

A theoretical model of inelastic electron-proton scattering was developed, which postulated the existence of point-like quarks within the proton.[42–44] In this model, the electron scatters elastically on one of the quarks, which carries a fraction ξ of the total momentum of the proton. This model is now known as the quark-parton model (QPM) and is depicted in the left panel of figure 2.4. In this picture, one can distinguish the scattered quark p' and the proton remnant X, which collectively form the hadronic system $P' = p' + X$.

In this model, one finds for the kinematic variables that

$$x_{\mathrm{Bj}} = \frac{Q^2}{2P \cdot q} = \xi \underbrace{\left(\frac{Q^2}{2\xi P \cdot q} \right)}_{=1} = \xi, \tag{2.9}$$

where the last step is analogous to equation (2.2), except that the proton mass is replaced by the quark mass, which can be neglected.

The electron-quark interaction in the QPM is identical to the elastic scattering of an electron on a point-like proton. Thus, the cross section of this sub-process is given by equation (2.5) when replacing the initial-state proton with the quark. To obtain an expression in terms of the proton, the replacement $P \to \xi P$ has to be made, which leads to $x_{\mathrm{Bj}} \to \xi$ and $s \to \xi s$. As mentioned in section 2.1, the quark only carries a fraction e_i of the elementary charge, such that the coupling on the quark side needs to be modified. In equation (2.8), one of the factors α corresponds to the electron and one to the quark. The latter one is modified as $\alpha \to e_i^2 \alpha$. In terms of the incoming proton, the QPM cross section therefore reads

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2} \right)_{\xi,i} = \frac{2\pi\alpha^2}{Q^4} Y_+ e_i^2. \tag{2.10}$$

This expression is valid for a single quark of flavour i and at a fixed value of ξ. Assuming that the quark flavour is not observable, all possible flavours must be summed over to obtain the full electron-proton cross section. The momentum fraction ξ constitutes an additional degree of freedom, such that two variables are necessary to describe the event. Therefore the full electron-proton cross section becomes double differential and reads

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{Bj}}\mathrm{d}Q^2} = \frac{2\pi\alpha^2}{x_{\mathrm{Bj}}Q^4} Y_+ \sum_{i \in \{q, \bar{q}\}} e_i^2 x_{\mathrm{Bj}} f_i(x_{\mathrm{Bj}}). \tag{2.11}$$

The probability of finding a quark i carrying a fraction ξ of the proton momentum is parameterised by the parton density distribution $f_i(\xi)$.[36] Comparing equations (2.11) and (2.8), the QPM predicts

$$F_2(x_{\mathrm{Bj}}, Q^2) = \sum_i e_i^2 x_{\mathrm{Bj}} f_i(x_{\mathrm{Bj}}), \tag{2.12a}$$

$$F_1(x_{\mathrm{Bj}}, Q^2) = F_2(x_{\mathrm{Bj}}, Q^2)/(2x_{\mathrm{Bj}}). \tag{2.12b}$$

2.3.5 Electroweak corrections to the quark-parton model

So far, only a photon was considered as the exchanged boson between the electron and the quark. If Q^2 is comparable to the Z boson mass M_Z^2, the Standard Model also permits the exchange of a Z boson and the interference of the two processes. Taking this effect into account leads to a parity-violating term $\sim \varepsilon^{\mu\nu\kappa\lambda} q_\kappa q_\lambda$ in the leptonic and hadronic tensors (2.6) and correspondingly an additional structure function in the cross section.[36] The electron-proton cross section may thus be written in the most general form

$$\frac{\mathrm{d}^2\sigma^\pm}{\mathrm{d}x_{\mathrm{Bj}}\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{x_{\mathrm{Bj}}Q^4}\Big(Y_+ F_2(x_{\mathrm{Bj}}, Q^2) \mp Y_- x_{\mathrm{Bj}} F_3(x_{\mathrm{Bj}}, Q^2) - y^2 F_{\mathrm{L}}(x_{\mathrm{Bj}}, Q^2)\Big), \tag{2.13}$$

where $\sigma^\pm$ refers to the scattering of positrons or electrons on protons. Here, the so-called longitudinal structure function $F_{\mathrm{L}} = F_2 - 2x_{\mathrm{Bj}}F_1$ was introduced, which vanishes in the QPM, as shown in equation (2.12b).

In the QPM including electroweak corrections, one finds

$$F_2(x_{\mathrm{Bj}}, Q^2) = \sum_{i \in q} \left(\int_0^1 \left(\xi f_i(\xi) + \xi \bar{f}_i(\xi)\right) \delta(\xi - x_{\mathrm{Bj}})\mathrm{d}\xi \right)\left(e_i^2 + \mathcal{O}\left(\frac{Q^2}{M_Z^2}\right)\right), \tag{2.14a}$$

$$x_{\mathrm{Bj}} F_3(x_{\mathrm{Bj}}, Q^2) = \sum_{i \in q} \left(\int_0^1 \left(\xi f_i(\xi) - \xi \bar{f}_i(\xi)\right) \delta(\xi - x_{\mathrm{Bj}})\mathrm{d}\xi \right)\mathcal{O}\left(\frac{Q^2}{M_Z^2}\right), \tag{2.14b}$$

$$F_{\mathrm{L}}(x_{\mathrm{Bj}}, Q^2) = 0. \tag{2.14c}$$

Note that the sums only run over the regular quarks. Antiquarks are taken into account via the $\bar{f}_i$ terms. In the QPM, F_2 is proportional to the sum of the quark and antiquark densities, while $x_{\mathrm{Bj}}F_3$ is proportional to their difference. Terms of order Q^2/M_Z^2 arise due to the interference of the diagrams corresponding to photon and Z exchange, and terms of order $(Q^2/M_Z^2)^2$ arise due to the diagram corresponding to pure Z exchange. The terms related to the Z boson also involve the weak isospin of the quarks and the electron, which is how this interaction violates charge-parity symmetry.[17,36]

The parton density distributions $f_i(\xi)$ are not predicted by the QPM. However, some constraints may be derived in the form of so-called sum rules. Since the proton is expected to contain two up-valence quarks u_{v} one down-valence quark d_{v}, the relations $\int f_{u_{\mathrm{v}}}(\xi)\mathrm{d}\xi = 2$ and $\int f_{d_{\mathrm{v}}}(\xi)\mathrm{d}\xi = 1$ should hold.[36] The momentum sum rule predicts that the total momentum of all quarks sums up to the total proton momentum

$$\sum_{i \in \{q, \bar{q}\}} \int_0^1 \xi f_i(\xi)\mathrm{d}\xi \overset{!}{=} 1. \tag{2.15}$$

Experimentally, it was found that this rule does not hold.[36] The expression on the left turned out to be less than one and depended on the energy at which it was measured. This indicates
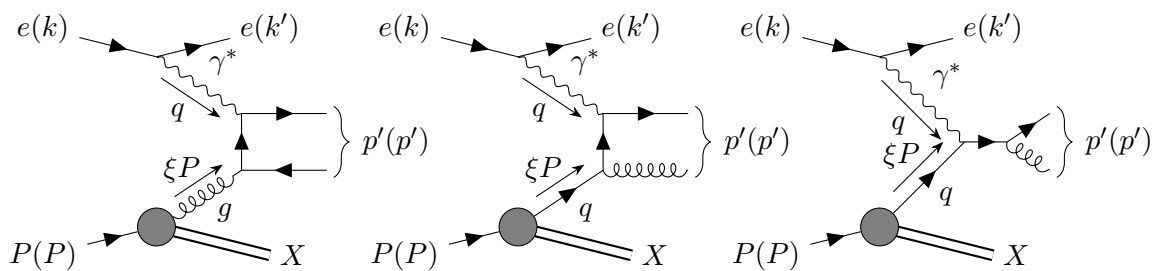
Figure 2.5: Leading tree-level contributions to electron-hadron scattering that include QCD effects. Left: boson-gluon-fusion process. Middle and right: QCD Compton process.

that there are additional components of the proton that are not described by the original QPM.

2.3.6 Inelastic scattering including effects from quantum chromodynamics

In 1979, the PETRA e^+e^- collider at DESY confirmed the existence of the gluon, which turned out to be this missing contribution.[45–48] The gluon is a gauge boson predicted by the theory of quantum chromodynamics (QCD). It has no electric or weak charge and thus does not couple directly to the boson coming from the electron. The leading contribution involving a gluon occurs via an intermediate quark, as shown in figure 2.5. Quarks and gluons are collectively referred to as 'partons'.

The general graph of electron-proton scattering, including QCD corrections, is shown in the right panel of figure 2.4. Here, p' is the hadronic final state from the boson-parton interaction. The generalised boson-parton vertex $\hat{\sigma}_i$ allows for modifications of the coupling due to QCD corrections. These effects can also lead to p' consisting of multiple particles. This is conceptually different from the case of elastic electron-proton scattering discussed before, where the generalised boson-proton vertex F_i was introduced to describe the structure of the proton.

The mass of the hadronic final state becomes a third degree of freedom in each event. The kinematic quantities Q^2, x_{Bj} and y are no longer sufficient to describe the event since only two of them are independent. The simple relation (2.9) between x_{Bj} and ξ is lost and replaced by

$$x_{\text{Bj}} = \frac{Q^2}{2P \cdot q} = \frac{\xi Q^2}{2\xi P \cdot q} = \frac{\xi Q^2}{\underbrace{(q + \xi P)^2}_{=p'} - q^2 - \xi^2 \underbrace{P^2}_{=0}} = \frac{\xi Q^2}{p'^2 + Q^2} = \frac{\xi}{1 + \frac{p'^2}{Q^2}}. \tag{2.16}$$

When considering QCD corrections, all three structure functions given in equation (2.14) gain additional terms proportional to the QCD coupling. They also gain contributions not involving $\delta(\xi - x_{\text{Bj}})$, such that the structure functions at any value of x_{Bj} are sensitive to the momentum densities at all values of ξ.

2.4 Perturbative quantum chromodynamics

The following section introduces the theory of QCD. Deriving this theory with complete theoretical rigour is beyond the scope of this thesis.[39,49] Instead, key ideas in the development from non-relativistic quantum mechanics to quantum field theory and eventually QCD are

highlighted. It is demonstrated that QED and QCD can be derived in a similar manner. The phenomenological differences, however, are vast and QCD is a rich theory with a wide range of unique features.

2.4.1 Path to quantum field theory

The classical energy-momentum relation of a single free particle of mass m is $E = \frac{p^2}{2m}$ with E and p being the energy and momentum of the particle. Quantising this relation leads to the well-known Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(\vec{x}, t) = -\frac{1}{2m}\nabla_{\vec{x}}^2\psi(\vec{x}, t), \tag{2.17}$$

that describes the evolution of the wave function $\psi(\vec{x}, t)$ and forms the basis of non-relativistic quantum mechanics.[50,51]

Being a first-order differential equation in time and second-order in space, this equation is clearly not Lorentz covariant. An equation describing relativistic quantum mechanics can be derived analogously from the relativistic energy-momentum relation $E^2 = m^2 + p^2$. The quantised version of this relation is the Klein-Gordon equation

$$\left(\partial_\mu\partial^\mu - m^2\right)\psi(x) = 0, \tag{2.18}$$

with $x = (t, \vec{x})$ being the spacetime coordinate.[52,53] This equation turns out to give the correct description of scalar particles, i.e. particles without spin.

To derive a first-order differential equation for relativistic quantum mechanics, Paul Dirac attempted to take the square root of the Klein-Gordon operator. He realised this is not possible for a scalar wave function, but instead, it required a wave function of at least four components. The operator that acts on the wave function becomes a 4×4-matrix. The Dirac equation reads

$$\left(i\gamma^\mu\partial_\mu - m\right)\psi(x) = 0, \tag{2.19}$$

where the Dirac matrices γ^μ are four complex 4×4-matrices and ψ is now a four-component Dirac spinor.[54] To simplify notation, the spinor indices of ψ and γ^μ are usually not written explicitly. This equation correctly describes spinor particles, i.e. particles with spin $\frac{1}{2}$. The four components correspond to the two spin-states of the particle and its antiparticle.

The Dirac equation describes both particles and antiparticles. This implies that particles may be created or annihilated. Being a one-particle equation, the Dirac equation cannot describe these processes. To describe systems with a variable number of particles, the framework of quantum field theory is necessary.[38]

2.4.2 Quantum electrodynamics

The Lagrangian density corresponding to the Dirac equation is

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi, \tag{2.20}$$

with $\bar{\psi} = \psi^\dagger\gamma^0$.[38] This expression exhibits a global $U(1)$ symmetry, i.e. the equation remains invariant when multiplying ψ by unitary 1×1-matrix $\psi \to \exp(i\theta)\psi$ with constant θ. The theory of QED can be derived by promoting this global symmetry to a local one $\theta = \theta(x)$. For the Lagrangian to remain invariant, the derivative ∂_μ has to be replaced by the covariant

derivative $D_\mu = \partial_\mu + ieA_\mu$. A massless vector field $A_\mu(x)$ is introduced, that transforms as $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\theta(x)$. A kinetic term for this vector field must be added to complete the theory. This term can be derived from the classical relativistic Maxwell equations $\partial_\mu F^{\mu\nu} = 0$ with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Ignoring muons and τ-leptons, the complete Lagrangian density for QED thus reads[36]

$$\mathcal{L}_{\mathrm{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi. \tag{2.21}$$
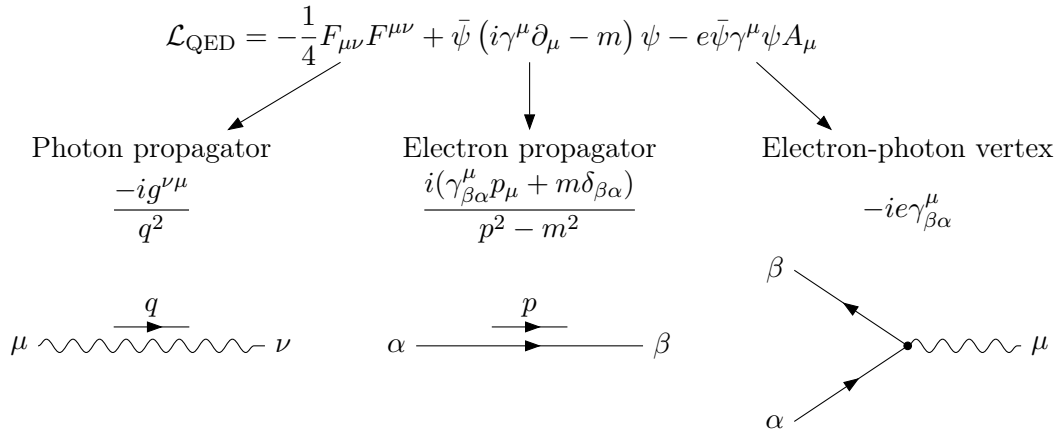
In this equation, the spinor field ψ represents electrons and positrons, and the vector field A_μ represents photons. The first term describes the propagation of free photons, and the second term that of electrons. Via the additional field term in the covariant derivative, the second term also describes the interaction of photons and electrons. When ignoring this interaction term, the theory can be solved exactly and trivially, as all particles remain in their initial states of motion and can never interact with each other. Taking the interaction term into account is not possible exactly and analytically. In many physically relevant situations, the coupling e is numerically small, such that the interaction term can be treated as a small perturbation of the free theory. This allows performing approximate computations in the interacting theory by expanding expressions in powers of the coupling. The Feynman diagram formalism gives a convenient framework to perform this expansion.[38]

In QED, Feynman diagrams may contain only three elements corresponding to the three terms in the Lagrangian: the photon propagator, the electron propagator and the electron-photon vertex. The corresponding graphical representations and mathematical expressions are given in figure 2.6. Each diagram that is built from these elements corresponds to one possible way in which particles may interact. A scattering process is defined by the set of incoming and outgoing particles. According to the superposition principle of quantum mechanics, the transition amplitude for a particular scattering process is constructed by summing all possible diagrams with the corresponding set of external particles. For any given process, there is an infinite number of such diagrams. The diagrams containing more powers of the coupling will be numerically smaller and therefore contribute less to the amplitude. An observable can thus be calculated approximately by considering only those diagrams that contain low powers of the coupling. In the Feynman diagram picture, this corresponds to only considering diagrams containing a small number of vertices. All terms contributing at a given order should be considered to ensure the cancellation of large corrections.

2.4.3 Quantum chromodynamics

The theory of QCD can be derived similarly to QED, except that a local $SU(3)$ gauge symmetry is imposed rather than a $U(1)$ gauge symmetry. This symmetry acts on an additional degree of freedom called 'colour'. Correspondingly, the wave function is promoted to a vector in colour space $\psi = (\psi_{\mathrm{red}}, \psi_{\mathrm{green}}, \psi_{\mathrm{blue}})$. The gauge transformation now reads $\psi \to \exp(ig\frac{\lambda^a}{2}\theta^a(x))\psi$, with the QCD coupling constant g and eight independent phases $\theta^a(x)$. The generators of the colour-group λ^a are eight Hermitian 3×3-matrices that satisfy the commutation relation $[\lambda^a, \lambda^b] = 2if^{abc}\lambda^c$ with f^{abc} being the structure constants of $SU(3)$.[36] To simplify notation, the quark-colour indices of ψ and λ^a are usually not written explicitly. QCD is called a non-Abelian gauge theory because the generators do not commute. This property contributes significantly to the rich phenomenology and increased complexity of QCD compared to QED.

When defining the covariant derivative, eight vector fields A_μ^a need to be introduced $D_\mu = \partial_\mu + ig\frac{\lambda^a}{2}A_\mu^a$. These fields are called gluons, and unlike the electrically neutral photon in

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu$$

Photon propagator

$$\frac{-ig^{\nu\mu}}{q^2}$$

Electron propagator

$$\frac{i(\gamma^\mu_{\beta\alpha}p_\mu + m\delta_{\beta\alpha})}{p^2 - m^2}$$

Electron-photon vertex

$$-ie\gamma^\mu_{\beta\alpha}$$

Figure 2.6: The graphical and mathematical representation of the QED Feynman rules. The indices $\mu, \nu \in \{0, 1, 2, 3\}$ are spacetime indices and $\alpha, \beta \in \{1, 2, 3, 4\}$ are spinor indices.

QED, they also carry colour charges. The gluon-colour charge is a combination of a quark-colour charge and an antiquark-colour charge, excluding the colour-neutral combination. The gluon-colour charge can thus take eight different values rather than three. The covariant derivative becomes a matrix in colour space. The field strength tensor takes the form $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$. Notably, this expression contains an additional quadratic term not present in its QED analogue. The QCD Lagrangian reads

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \sum_f \bar{\psi}_f\left(i\gamma^\mu D_\mu - m_f\right)\psi_f, \tag{2.22}$$

where the sum over f runs over the six flavours of quarks. The corresponding Feynman rules are given in figure 2.7.[36] The quadratic term in the field strength tensor leads to the appearance of three- and four-gluon vertices.

An important phenomenological property of QCD is colour-confinement. As discussed in the next section, the gluon field between colour-charged particles does not decrease with distance. Therefore, particles that carry a colour charge cannot be isolated at large distances. Partons always appear as part of colour-neutral bound states called hadrons. There are two types of hadrons. Mesons are formed from a quark and an antiquark with opposite colour charges, e.g. red and anti-red. Baryons are formed from three quarks of different colours, i.e. red, green and blue. In both cases, the overall colour charge of the hadron vanishes. In addition to these so-called valence quarks, hadrons contain many gluons, which carry a colour and an anti-colour charge, allowing them to continuously form and annihilate further quark-antiquark pairs.

2.4.4 Renormalisation and running of the strong coupling

One of the most distinctive features of QCD is the fact that the strong coupling decreases with energy. This feature makes QCD qualitatively different from QED and is responsible for many of its phenomenological properties. It leads to colour confinement at low energies and asymptotic freedom at high energies.[12,55] This makes perturbative QCD only applicable at sufficiently high energies, where the coupling becomes small. This behaviour can be derived by renormalising QCD in perturbation theory.

Gluon propagator

$$\frac{-ig^{\nu\mu}\delta^{ba}}{q^2}$$

Quark propagator

$$\frac{i(\gamma^{\mu}_{\beta\alpha}p_{\mu} + m\delta_{\beta\alpha})\delta_{ji}}{p^2 - m^2}$$

μ, a ⁓⁓⁓ q ⁓⁓⁓ ν, b α, i ——p—→ β, j

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}(\partial_{\mu}A^a_{\nu} - \partial_{\nu}A^a_{\mu})(\partial^{\mu}A^{a\nu} - \partial^{\nu}A^{a\mu}) + \sum_f \bar{\psi}_f\left(i\gamma^{\mu}\partial_{\mu} - m_f\right)\psi_f$$

$$-g\sum_f \bar{\psi}_f\frac{\lambda^a}{2}\gamma^{\mu}\psi_f A^a_{\mu} - \frac{1}{2}g(\partial_{\mu}A^a_{\nu} - \partial_{\nu}A^a_{\mu})f^{abc}A^{b\mu}A^{c\nu} - \frac{1}{4}g^2 f^{abc}f^{ade}A^b_{\mu}A^c_{\nu}A^{d\mu}A^{e\nu}$$

Quark-gluon vertex

$$-i\frac{g}{2}\lambda^a_{ji}\gamma^{\mu}_{\beta\alpha}$$

Three-gluon vertex

$$-gf^{abc}\big(g^{\mu\nu}(k_1 - k_2)^{\rho}$$
$$+g^{\nu\rho}(k_2 - k_3)^{\mu}$$
$$+g^{\rho\mu}(k_3 - k_1)^{\nu}\big)$$

Four-gluon vertex

$$-ig^2\big(f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\lambda} - g^{\mu\lambda}g^{\nu\rho})$$
$$+f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\lambda} - g^{\mu\rho}g^{\nu\lambda})$$
$$+f^{ace}f^{dbe}(g^{\mu\lambda}g^{\nu\rho} - g^{\mu\nu}g^{\rho\lambda})\big)$$
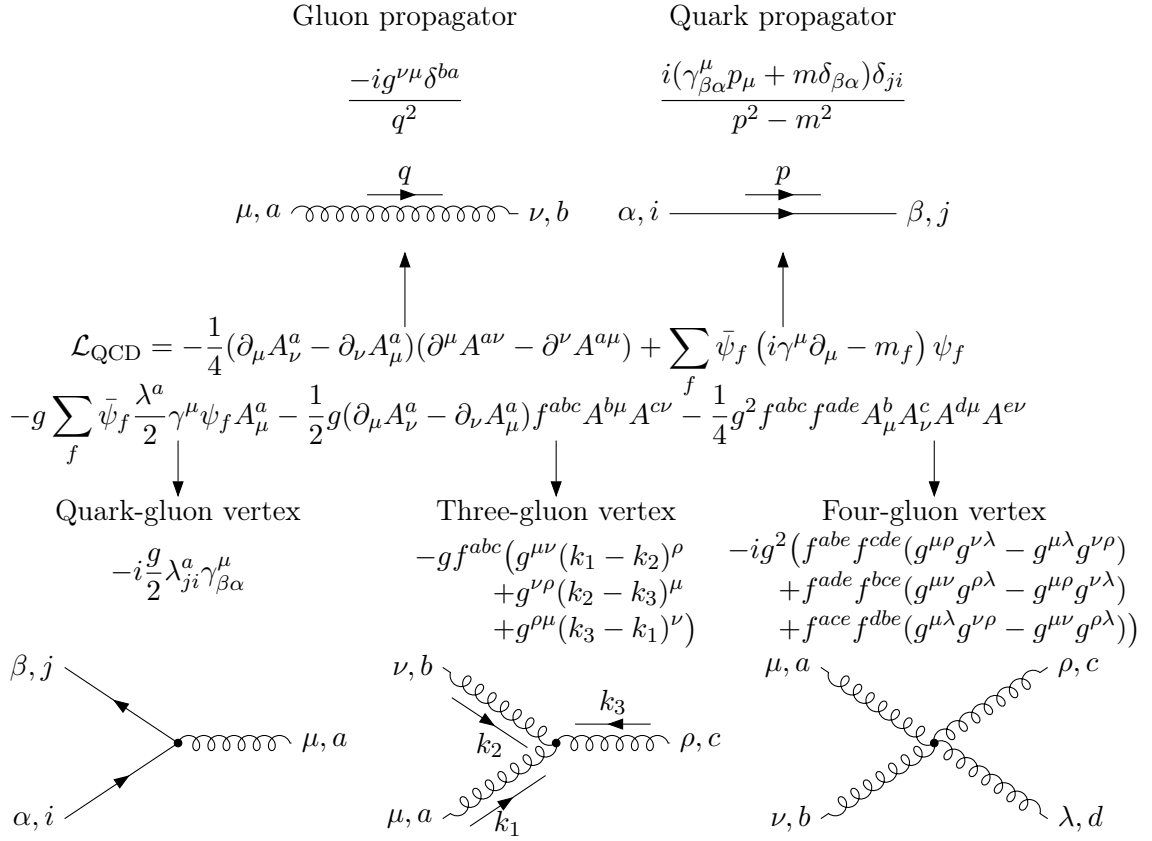
Figure 2.7: The graphical and mathematical representation of the QCD Feynman rules. The indices $\mu, \nu, \rho, \lambda \in \{0, 1, 2, 3\}$ are spacetime indices, $\alpha, \beta \in \{1, 2, 3, 4\}$ are spinor indices, $i, j \in \{\text{red}, \text{green}, \text{blue}\}$ are quark-colour indices and $a, b, c, d, e \in \{1, 2, \ldots, 8\}$ are gluon-colour indices.

Strategy for renormalisation

The following computation is performed in the limit of massless quarks. In massless QCD, there are only three independent quantities that can be renormalised: the quark and gluon field and the coupling.[a] The bare quantities are related to the renormalised ones as

$$\psi_{(\mathrm{b})} = \sqrt{Z_{\psi}}\psi_{(\mathrm{r})}, \tag{2.23a}$$

$$A^{a,\mu}_{(\mathrm{b})} = \sqrt{Z_A}A^{a,\mu}_{(\mathrm{r})}, \tag{2.23b}$$

$$g_{(\mathrm{b})} = Z_g\mu^{\varepsilon}g_{(\mathrm{r})}. \tag{2.23c}$$

The calculation is performed using dimensional regularisation in $D = 4 - 2\varepsilon$ dimensions. Dimensional analysis of the bare Lagrangian (2.22) shows that the bare coupling $g_{(\mathrm{b})}$ must have mass dimension ε. The factor μ^{ε}, where μ is an arbitrary constant of mass dimension 1, is introduced in equation (2.23c) to ensure that the renormalised coupling is dimensionless. The renormalisation factors Z must be chosen to keep the results of the computations finite. They can be computed using perturbation theory.

---

[a]In principle, the gauge fixing parameter ξ is also renormalised. This is not relevant for the present discussion, and it does not add anything to the theory, since $Z_{\xi} = Z_A$.[56]

By definition, the bare coupling $g_{(b)}$ is independent of the scale μ. Therefore, the scale dependence of the renormalised coupling $g_{(r)}$ must be such that the right-hand side of equation (2.23c) is also independent of μ. In the following, the factor Z_g is explicitly computed at one-loop order in perturbation theory and the scale dependence of the coupling is extracted.

One-loop perturbative calculation

Writing the Lagrangian (2.22) in terms of the renormalised quantities yields

$$
\begin{aligned}
\mathcal{L}_{\text{QCD}} =& -\frac{1}{4}(F^2)^{(\text{gg})}_{(b)} - \frac{1}{4}(F^2)^{(\text{ggg})}_{(b)} - \frac{1}{4}(F^2)^{(\text{gggg})}_{(b)} + \bar{\psi}_{(b)}i\gamma^\mu\partial_\mu\psi_{(b)} - \bar{\psi}_{(b)}g_{(b)}\frac{\lambda^a}{2}A^\mu_{(b)}\gamma_\mu\psi_{(b)} \\
=& -\frac{1}{4}(F^2)^{(\text{gg})}_{(r)} - \frac{1}{4}(F^2)^{(\text{ggg})}_{(r)} - \frac{1}{4}(F^2)^{(\text{gggg})}_{(r)} + \bar{\psi}_{(r)}i\gamma^\mu\partial_\mu\psi_{(r)} - \bar{\psi}_{(r)}g_{(r)}\frac{\lambda^a}{2}A^\mu_{(r)}\gamma_\mu\psi_{(r)} \\
& -\frac{1}{4}\delta_{gg}(F^2)^{(\text{gg})}_{(r)} - \frac{1}{4}\delta_{ggg}(F^2)^{(\text{ggg})}_{(r)} - \frac{1}{4}\delta_{gggg}(F^2)^{(\text{gggg})}_{(r)} \\
& + \delta_{qq}\bar{\psi}_{(r)}i\gamma^\mu\partial_\mu\psi_{(r)} - \delta_{qqg}\bar{\psi}_{(r)}g_{(r)}\frac{\lambda^a}{2}A^\mu_{(r)}\gamma_\mu\psi_{(r)},
\end{aligned}
$$

where the symbols $(F^2)^{(\text{gg})}$, $(F^2)^{(\text{ggg})}$ and $(F^2)^{(\text{gggg})}$ are used as shorthands for the two-, three- and four-gluon terms in $F^a_{\mu\nu}F^{a\mu\nu}$. The first part of the renormalised Lagrangian looks identical to the bare one, but written in terms of the renormalised quantities. The second part contains additional interactions, the so-called counterterms. The coefficients of the counterterms can be expressed in terms of the renormalisation constants by analysing the terms that they multiply. One finds $\delta_{gg} = Z_A - 1$, $\delta_{qq} = Z_\psi - 1$, $\delta_{qqg} = Z_g Z_\psi \sqrt{Z_A} - 1$, $\delta_{ggg} = Z_g Z_A^{3/2} - 1$ and $\delta_{gggg} = Z_g^2 Z_A^2 - 1$. Infinities that appear in the calculation can be absorbed into these coefficients such that the calculated observables remain finite.[56,57]

When calculating observables in the renormalised theory, additional vertices appear, corresponding to the counterterms. Since their coefficients always involve the coupling, they should be treated as interactions, even if they contain just two fields. The counterterm Feynman rules are given in figure 2.8.

To order $g_{(r)}^3$, the corrections to the gluon propagator, quark propagator and quark-gluon vertex are

$$
= \quad i\Pi^{\mu\nu}(q) \qquad +i\delta_{gg}(q^\mu q^\nu - q^2 g^{\mu\nu}), \qquad (2.24a)
$$

$$
= \quad i\Sigma(q) \qquad + \quad i\delta_{qq}\gamma^\mu q_\mu, \qquad (2.24b)
$$

$$
= \quad ig_{(r)}\Gamma^\mu(q) \qquad - \quad ig_{(r)}\delta_{qqg}\frac{\lambda}{2}\gamma^\mu, \qquad (2.24c)
$$

where the grey circles refer to all possible one-loop corrections without counterterms. The functions Π, Σ and Γ can be calculated using perturbation theory, and their divergent parts

Gluon counterterm
$$i\delta_{gg}(q^\mu q^\nu - q^2 g^{\mu\nu})$$

Quark counterterm
$$i\delta_{qq}\gamma^\mu p_\mu$$

$$\mathcal{L}_{\text{QCD}}^{\text{(CT)}} = -\frac{1}{4}\delta_{gg}(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu}) + \delta_{qq}\bar{\psi}i\gamma^\mu\partial_\mu\psi$$

$$-g\delta_{qqg}\bar{\psi}\frac{\lambda^a}{2}\gamma^\mu\psi A_\mu^a - \frac{1}{2}g\delta_{ggg}(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)f^{abc}A^{b\mu}A^{c\nu} - \frac{1}{4}g^2\delta_{gggg}f^{abc}f^{ade}A_\mu^b A_\nu^c A^{d\mu}A^{e\nu}$$

Quark-gluon counterterm
$$-ig\delta_{qqg}\frac{\lambda_{ji}^a}{2}\gamma_{\alpha\beta}^\mu$$

Three-gluon counterterm
$$-g\delta_{ggg}f^{abc}\big(g^{\mu\nu}(k_1 - k_2)^\rho$$
$$+g^{\nu\rho}(k_2 - k_3)^\mu$$
$$+g^{\rho\mu}(k_3 - k_1)^\nu\big)$$

Four-gluon counterterm
$$-ig^2\delta_{gggg}\big(f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\lambda} - g^{\mu\lambda}g^{\nu\rho})$$
$$+f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\lambda} - g^{\mu\rho}g^{\nu\lambda})$$
$$+f^{ace}f^{dbe}(g^{\mu\lambda}g^{\nu\rho} - g^{\mu\nu}g^{\rho\lambda})\big)$$
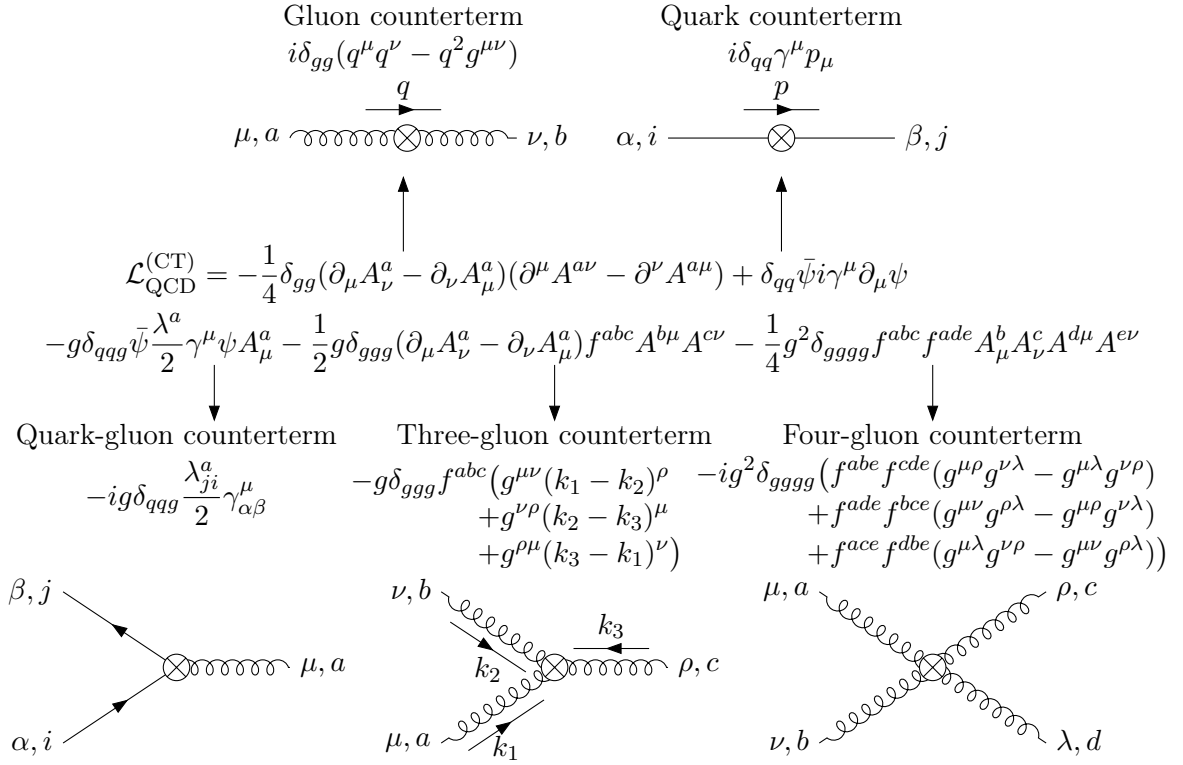
Figure 2.8: The graphical and mathematical representation of the QCD counterterm Feynman rules. The indices are labelled as in figure 2.7.

can be absorbed into the counterterm coefficients δ. These terms are explicitly calculated at one-loop order in appendix A. The results are

$$i\Pi^{\mu\nu}(q) = i\left(\frac{2}{3}n_{\text{f}} - \frac{5}{3}C_{\text{A}}\right)f(q)(q^\mu q^\nu - q^2 g^{\mu\nu}), \tag{2.25a}$$

$$i\Sigma(q) = iC_{\text{F}}f(q)q^\mu\gamma_\mu, \tag{2.25b}$$

$$ig_{(\text{r})}\Gamma^{\mu,a}(q) = -ig_{(\text{r})}(C_{\text{F}} + C_{\text{A}})f(q)\frac{\lambda^a}{2}\gamma^\mu + (\text{UV finite}), \tag{2.25c}$$

$$\text{with } f(q) = \frac{g_{(\text{r})}^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \gamma_{\text{E}} + \log(4\pi) - \log\left(\frac{-q^2}{\mu^2}\right)\right),$$

where $C_{\text{F}} = \frac{N^2-1}{2N}$ and $C_{\text{A}} = N$ are the Casimir operators of the $SU(N)$ group and n_{f} is the number of active quark flavours, i.e. the number of quark flavours whose mass is low enough to contribute at the investigated energy.[39,49]

Inserting equations (2.25) into (2.24) and absorbing their divergent parts in the counterterms yields

$$\delta_{gg}^{\overline{\mathrm{MS}}} = -\frac{g_{(\mathrm{r})}^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \gamma_{\mathrm{E}} + \log(4\pi)\right) \cdot \left(\frac{2}{3}n_{\mathrm{f}} - \frac{5}{3}C_{\mathrm{A}}\right), \tag{2.26a}$$

$$\delta_{qq}^{\overline{\mathrm{MS}}} = -\frac{g_{(\mathrm{r})}^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \gamma_{\mathrm{E}} + \log(4\pi)\right) \cdot C_{\mathrm{F}}, \tag{2.26b}$$

$$\delta_{qqg}^{\overline{\mathrm{MS}}} = -\frac{g_{(\mathrm{r})}^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \gamma_{\mathrm{E}} + \log(4\pi)\right) \cdot (C_{\mathrm{F}} + C_{\mathrm{A}}). \tag{2.26c}$$

In the 'minimal subtraction' scheme (MS), only the divergent terms involving $1/\varepsilon$ are absorbed into the counterterms. Here, the 'modified minimal subtraction' scheme ($\overline{\mathrm{MS}}$) is used. In this scheme, also the constant terms involving γ_{E} and $\log(4\pi)$ are absorbed. This definition is convenient, as these terms appear together in the calculation. Note that the counterterms are proportional to $g_{(\mathrm{r})}^2$. Thus, the expressions in equation (2.24) are a consistent expansion up to $\mathcal{O}(g_{(\mathrm{r})}^3)$.

The coupling renormalisation factor is then given by

$$Z_g^{\overline{\mathrm{MS}}} = \frac{\delta_{qqg}^{\overline{\mathrm{MS}}} + 1}{(1 + \delta_{qq}^{\overline{\mathrm{MS}}})\sqrt{1 + \delta_{gg}^{\overline{\mathrm{MS}}}}} = 1 + \delta_{qqg}^{\overline{\mathrm{MS}}} - \delta_{qq}^{\overline{\mathrm{MS}}} - \frac{1}{2}\delta_{gg}^{\overline{\mathrm{MS}}} + \mathcal{O}(g_{(\mathrm{r})}^4) = 1 - \frac{g_{(\mathrm{r})}^2}{4\pi\varepsilon}\frac{b_0}{2} + g_{(\mathrm{r})}^2 C^{\overline{\mathrm{MS}}}, \tag{2.27}$$

where $b_0 = (11C_{\mathrm{A}} - 2n_{\mathrm{f}})/12\pi$ and $C^{\overline{\mathrm{MS}}}$ is a constant involving γ_{E} and $\log(4\pi)$.

The scale dependence of the renormalised coupling is found by demanding that the bare coupling is independent of the scale

$$0 \overset{!}{=} \mu\frac{\mathrm{d}g_{(\mathrm{b})}}{\mathrm{d}\mu} = \mu\frac{\mathrm{d}\big(g_{(\mathrm{r})}\mu^\varepsilon Z_g\big)}{\mathrm{d}\mu}$$

Inserting equation (2.27) and solving for the scale dependence of the strong coupling yields

$$\mu\frac{\mathrm{d}g_{(\mathrm{r})}}{\mathrm{d}\mu} = -\frac{g_{\mathrm{r}}^3}{4\pi}b_0 + \mathcal{O}(\varepsilon) + \mathcal{O}(g_{(\mathrm{r})}^5),$$

or equivalently, for the β-function of QCD

$$\beta(\alpha_{\mathrm{s}}) = \mu^2\frac{\mathrm{d}\alpha_{\mathrm{s}}}{\mathrm{d}\mu^2} = -\alpha_{\mathrm{s}}^2 b_0 + \mathcal{O}(\alpha_{\mathrm{s}}^3), \qquad \text{with } b_0 = \frac{11C_{\mathrm{A}} - 2n_{\mathrm{f}}}{12\pi} \tag{2.28}$$

where $\alpha_{\mathrm{s}} = \frac{g_{(\mathrm{r})}^2}{4\pi}$. In QCD, $C_{\mathrm{A}} = 3$ and $n_{\mathrm{f}} \leq 6$, such that b_0 is positive. This equation, therefore, predicts that the strong coupling decreases with increasing energy. This ultimately leads to asymptotic freedom at high energies and colour confinement at low energies.[12]

At one-loop order in perturbation theory, this differential equation can be solved exactly, and an analytic expression for the running of the coupling may be derived

$$\alpha_{\mathrm{s}}(\mu^2) = \frac{\alpha_{\mathrm{s}}(\mu_0^2)}{1 + \alpha_{\mathrm{s}}(\mu_0^2)b_0\log\left(\frac{\mu^2}{\mu_0^2}\right)} = \frac{1}{b_0\log\left(\frac{\mu^2}{\Lambda_{\mathrm{QCD}}^2}\right)}. \tag{2.29}$$
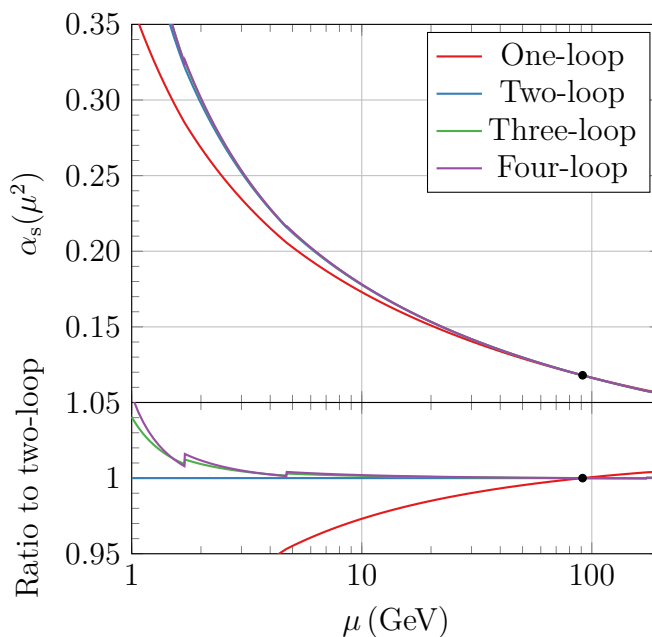
Figure 2.9: The running of the strong coupling computed at up to four-loop order in the zero-mass variable-flavour-number scheme. The ratio is shown relative to the two-loop calculation, which is the highest order without discontinuities. All curves were calculated starting from $\alpha_s(M_Z^2) = 0.118$. When passing a quark-mass threshold, the curves exhibit a small kink and, starting at three-loop order, a discontinuity. The quark-mass thresholds are set identically at each order in QCD.

Usually, $\alpha_s(\mu^2)$ is parameterised either by fixing its value at an arbitrary reference scale $\alpha_s(\mu_0^2)$ or using the QCD scale Λ_{QCD}. It is conventional to quote the strong coupling at the mass of the Z boson: $\alpha_s(M_Z^2)$.[58]

This derivation is performed in the zero-mass fixed-flavour-number scheme. In this scheme, all quarks are treated as massless, but individual quark flavours only contribute to the calculation when the considered scale is larger than their mass.[59] Consequently, $n_f(\mu)$ is a step function that increases at the quark-mass thresholds. This scheme is valid only if the considered scale is far away from any of the quark masses, where threshold effects become relevant. When passing a quark-mass threshold and adjusting n_f, the value of b_0 changes, leading to a kink in the $\alpha_s(\mu^2)$ curve. Furthermore, at higher orders in perturbation theory, this transition of the QCD scheme can lead to a discontinuity. At the mass of the $n_f + 1$-th quark, the theory with $n_f + 1$ active quarks may be related to the theory with n_f active quarks as

$$\alpha_s^{(n_f+1)}(m_{f+1}^2) = \alpha_s^{(n_f)}(m_{f+1}^2)\left(1 + \sum_{n=1}^{\infty} c_n \alpha_s^{(n_f)}(m_{f+1}^2)^n\right). \tag{2.30}$$

It turns out that $c_1 = 0$.[17,60] Therefore, discontinuities are only present at three-loop order and higher. The running of $\alpha_s(\mu^2)$ is shown in figure 2.9 up to four-loop order.

The scale μ is called the renormalisation scale μ_r. It is an arbitrary parameter introduced during the renormalisation procedure. If the perturbative expansion was performed to all orders, any dependence on μ_r in the computed quantities would cancel. However, since the series is truncated at a fixed order, computed observables gain a scale dependence. Truncating

the perturbative series introduces an uncertainty in predictions due to missing higher-order terms. To minimise this uncertainty, the perturbative calculation should be performed to as many orders as possible.

2.4.5 Factorisation in lepton-hadron collisions

The central assumption of the QPM is that the lepton-hadron cross section can be factored into the probability $f_i(\xi)$ of finding a given quark in the proton and the cross section of the lepton-quark interaction, see equation 2.11. This assumption also holds when considering QCD effects, but the parton density distributions $f_i(\xi)$ no longer correspond directly to a probability. The straightforward generalisation of the QPM formula is

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{Bj}}\mathrm{d}Q^2} = \sum_{i\in\{q,\bar{q},g\}} \int_{x_{\mathrm{Bj}}}^1 \mathrm{d}\xi f_i(\xi)\hat{\sigma}_i\left(\frac{x_{\mathrm{Bj}}}{\xi}, Q^2, \mu_{\mathrm{r}}^2, \alpha_{\mathrm{s}}(\mu_{\mathrm{r}})\right), \tag{2.31}$$

where i runs over all partons in the proton and $\hat{\sigma}_i$ is the lepton-parton matrix element, which is calculable in perturbative QCD.[61] The parton density distributions $f_i(\xi)$ belong to the domain of low-energy QCD, where the coupling becomes large, and perturbation theory is not applicable. Currently, the $f_i(\xi)$ can only be determined from fits to experimental data or, to some extent, from lattice QCD computations.[62] Instead of $f_i(\xi)$, these determinations are usually performed in terms of the parton distribution functions (PDFs) $\xi f_i(\xi)$.

Three types of divergences can occur when evaluating $\hat{\sigma}_i$. Ultraviolet divergences can occur when a loop momentum goes to infinity, causing the integral to diverge. They can be dealt with by renormalisation, as shown in the previous section. Infrared divergences can occur when the momentum of a particle goes to zero, causing the propagator to diverge. Collinear divergences can occur when two massless particles become collinear, which also causes the propagator to diverge. According to the Bloch-Nordsieck and Kinoshita-Lee-Nauenberg theorems, infrared and collinear divergences in the final state cancel when considering both real and virtual corrections. The mass factorisation theorem states that infrared and collinear divergences in the initial state can be factored out of the matrix element and into the PDFs.[63] A procedure similar to renormalisation can be used on the PDFs to cancel these divergences. As a result, the matrix elements and the PDFs gain a dependence on a newly introduced scale μ_{f}, known as the factorisation scale.[36]

Mathematically, the factorisation scale is the scale at which the PDFs are matched to the lepton-parton matrix elements. Parton radiation above the scale μ_{f} is treated fully perturbatively in the matrix elements, while radiation below this scale is absorbed into the evolution of the PDFs. Equation (2.31) becomes

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{Bj}}\mathrm{d}Q^2} = \sum_{i\in\{q,\bar{q},g\}} \int_{x_{\mathrm{Bj}}}^1 \mathrm{d}\xi f_i(\xi,\mu_{\mathrm{f}}^2)\hat{\sigma}_i\left(\frac{x_{\mathrm{Bj}}}{\xi}, Q^2, \mu_{\mathrm{r}}^2, \mu_{\mathrm{f}}^2, \alpha_{\mathrm{s}}(\mu_{\mathrm{r}}^2)\right). \tag{2.32}$$

The dependence of the PDFs on μ_{f} is predicted by perturbative QCD. This is similar to the μ_{r} dependence of α_{s}, discussed in the previous section. This evolution is given by the DGLAP equations

$$\mu_{\mathrm{f}}\frac{\partial f_i(\xi,\mu_{\mathrm{f}}^2)}{\partial \mu_{\mathrm{f}}} = \sum_{j\in\{q,\bar{q},g\}} \int_\xi^1 \frac{\mathrm{d}\xi'}{\xi'} P_{ij}\left(\frac{\xi}{\xi'}, \alpha_{\mathrm{s}}(\mu_{\mathrm{f}}^2)\right) f_j(\xi',\mu_{\mathrm{f}}^2). \tag{2.33}$$

Here, i and j run over all partons and P_{ij} are the splitting functions. The splitting function $P_{ij}(z)$ gives the probability for a parton j to split into a parton i containing a fraction z of the original partons momentum. The remaining partons resulting from the splitting are not considered explicitly.[36] The splitting functions can be calculated using perturbation theory.

This system of coupled differential equations can be used to evolve the PDFs from one scale to another. Thus, it is sufficient to describe the PDFs at a single scale $\xi f_i(\xi, \mu_{f,0}^2)$, where $\mu_{f,0}^2$ is known as the parameterisation scale.

2.5 Jets

In high-energy QCD, the final state of the lepton-hadron interaction is described in terms of partons. As described in the previous section, these partons are experimentally unobservable as they will fragment into a collection of hadrons. Since these hadrons are moving approximately in the same direction, they will form a bundle. By summing up the momenta of all hadrons in such a bundle, the momentum of the original parton can be approximately reconstructed. This allows to measure partons experimentally, even though they are not directly observable. A collection of particles moving in a similar direction is called a 'jet'. Jets are an essential tool for comparing theoretical predictions and experimental measurements of QCD processes.

Jets can be constructed from simulated partons from a QCD calculation, simulated hadrons from MC simulations or experimentally measured particles or energy deposits. In either case, the input to the jet clustering procedure is a collection of four-momenta corresponding to some type of object. One distinguishes parton-level, hadron-level and detector-level jets depending on the object type used. Since these are different jet definitions, they must be kept strictly separate. However, there is a strong correspondence between jets constructed at different levels that allows them to be correlated. This makes jets a valuable tool for comparing theoretical predictions and experimental measurements.

2.5.1 Jet algorithms

It is essential to clearly define what constitutes a jet, to make precise comparisons of theoretical predictions and experimental observations. For this purpose, various jet algorithms have been developed that construct jets from groups of input objects.

The results of jet algorithms can be visualised in the η-ϕ-plane. Here, ϕ is the azimuthal angle around the collision axis and $\eta = -\log(\tan(\theta/2))$ is the pseudorapidity with θ being the polar angle with respect to the collision axis. While η can take any value in $(-\infty, \infty)$, values outside $|\eta| \lesssim 3$ are not of interest here, as they are not experimentally accessible in the ZEUS detector due to being very close to the beam pipe. A possible set of inputs and two ways to cluster them are shown in the η-ϕ-plane in figure 2.10.

Historically, the so-called cone algorithms were one of the first widely used classes of jet algorithms. These algorithms place circles in the η-ϕ-plane, within which particles are contained. The four-momenta of all particles within each circle are then added to construct the jet axis corresponding to that circle. The circles are then moved around until the jet axes coincide with the centres of the corresponding circles. One example of such an algorithm is the SISCone algorithm.[65] A possible output of this algorithm is shown in the left panel of figure 2.10.

Cone algorithms are often infrared- or collinear unsafe. An infrared unsafe algorithm is one whose output might change significantly when one of the particles in the jet radiates an additional soft particle. Collinear unsafe algorithms are affected similarly by radiation that is
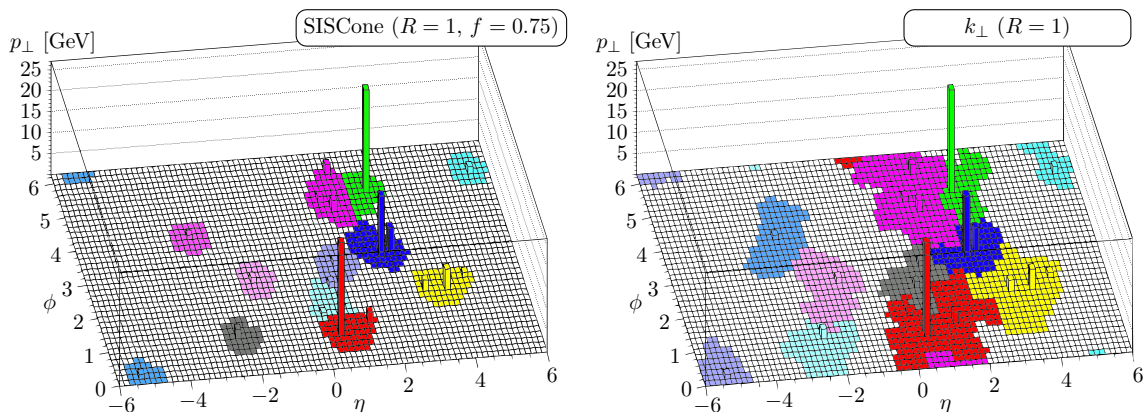
Figure 2.10: Comparison of different jet clustering algorithms.[64] The individual cells indicate the location and energy of the input particles, and the colours indicate the constructed jets. The same input particles were clustered in each panel using a different jet algorithm.

almost collinear to the particle axis. An algorithm affected by either of these problems will lead to unstable output sensitive to small input fluctuations. This makes it challenging to compare theoretical predictions and experimental measurements meaningfully.[66] The SIScone algorithm is an exception, as it is both infrared and collinear safe.

Another class of algorithms was developed that is both infrared- and collinear-safe. These so-called sequential clustering algorithms proceed by grouping particles based on their momenta. The most relevant such algorithm used at HERA is the $k_\perp$-algorithm. It works as follows:[67]

1. For each particle i, calculate the distance to the collision axis as
$$d_i = p_{\perp,i}^2 R^2.$$

2. For each pair of particles i and j, calculate the distance between particles as
$$d_{ij} = \min\left(p_{\perp,i}^2, p_{\perp,j}^2\right)\left(\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2\right).$$

3. Find the smallest value in $\{d_i, d_{ij}\}$.
   a) If the smallest value is d_i: Identify particle i a jet and remove it from the list of inputs.
   b) If the smallest value is d_{ij}: Combine particles i and j, i.e. add their combination to the list of inputs and remove i and j as individual inputs.

4. Repeat until no more input particles are left.

The value R in the first step is a constant that can be used to control the angular size of the jets. At HERA, usually $R = 1$. The term in parentheses in the expression for d_{ij} is the separation of the two particles in the η-ϕ-plane. This algorithm will start by merging particles of small energy or small angular separation. This makes it insensitive to infrared or collinear radiation. A possible output of this algorithm is shown in the right panel of figure 2.10.

In step 3b, two particles are merged into one. There are multiple ways to perform this merging. The simplest way is to add the four-momenta of the two particles. This is sometimes called the E-scheme and will lead to massive jets. The construction of massless jets is often desired for consistency with theoretical predictions of massless quarks. This can be achieved via the $p_\perp$-weighted scheme, in which the combined particle is defined as $p_\perp = p_{\perp,i} + p_{\perp,j}$, $\eta = (\eta_i p_{\perp,i} + \eta_j p_{\perp,j})/p_\perp$, $\phi = (\phi_i p_{\perp,i} + \phi_j p_{\perp,j})/p_\perp$ and $m = 0$.[67]

Sequential clustering algorithms are computationally more demanding and only came into widespread use due to advances in hardware and software.[68] Other sequential clustering algorithms are available, such as the anti-$k_\perp$-algorithm.[64] This algorithm is similar to the $k_\perp$-algorithm, except that the $p_\perp^2$ terms in step 1 and 2 are replaced by $1/p_\perp^2$. This leads to jets becoming much more circular, which is desirable at hadron-hadron colliders, where many collisions can occur at each bunch crossing, leading to much more background activity in the detectors. At HERA, it has been shown that both algorithms lead to a similar physics performance.[69]

2.5.2 The Breit frame of reference

In this thesis, jet production in DIS will be used to study QCD. Single-jet events can arise via the QPM-like process. Here, and in the remainder of this thesis, the term 'QPM-like process' refers to single-jet production in DIS, including higher-order QCD and electroweak corrections such as PDF evolution, collinear gluon emission, loop corrections or photon radiation.[70] The leading contribution to this process is of zeroth order in QCD, $\mathcal{O}(\alpha^2 \alpha_\mathrm{s}{}^0)$, in the matrix elements. First-order and higher-order QCD effects in the matrix elements will result in corrections to the single-jet process and additional contributions from events with two or more jets. This makes it difficult to probe QCD effects, as they will only manifest as a relatively small correction to the distributions produced by the leading-order QPM-like process. It is convenient to suppress the single-jet process to probe QCD effects more directly. In this case, the leading contribution to the jet distribution will be from dijet events, which are of first order in QCD in the matrix elements.

A convenient way to suppress single-jet events is to consider the events in the Breit reference frame.[36] This frame is defined such that the exchanged boson and the interacting parton from the hadron collide head-on along the z-axis, and the parton reverses its longitudinal momentum. Thus, the four-momentum of the exchanged boson is $q^\mu = (0, 0, 0, -Q)$ and that of the incoming parton is $\xi P^\mu = (Q/2, 0, 0, Q/2)$. Single-jet and dijet production are shown in the Breit frame in figure 2.11. In the case of single-jet production, the outgoing quark has a momentum of $p'^\mu = (Q/2, 0, 0, -Q/2)$, i.e. by construction, it has a vanishing transverse momentum in the Breit frame $p'_{\perp,\mathrm{Breit}} = 0$. In the dijet event, both jets will possess an equal and non-zero transverse momentum in the Breit frame.

Therefore, requiring jets to have a minimum transverse momentum in the Breit frame rejects single-jet events while retaining events with two or more jets. When applying this definition, the leading-order contribution to jet production will be of first order in QCD. This will allow for a precise study of QCD effects while minimising the influence of the modelling of the QPM process.

2.5.3 Inclusive jets

Various observable distributions can be constructed from jets. The simplest class of observables originates from dijet events, which are the leading-order QCD contribution to jet production. The two corresponding diagrams are shown in figure 2.5. An event is classified as a dijet event if it contains at least two jets of significant transverse momentum in the Breit frame. Possible observables in such an event are the invariant mass of the dijet system $p'^2 = (p_1 + p_2)^2$, the average transverse momentum $\overline{p_{\perp,\mathrm{Breit}}} = (p_{\perp,\mathrm{Breit},1} + p_{\perp,\mathrm{Breit},2})/2$, the difference of pseudorapidities $\eta^* = |\eta_{\mathrm{Breit},1} - \eta_{\mathrm{Breit},2}|/2$ and the momentum fraction of the interaction parton $\xi = x_{\mathrm{Bj}}(1 + p'^2/Q^2)$.[71] These quantities give insights not only into the occurrence of
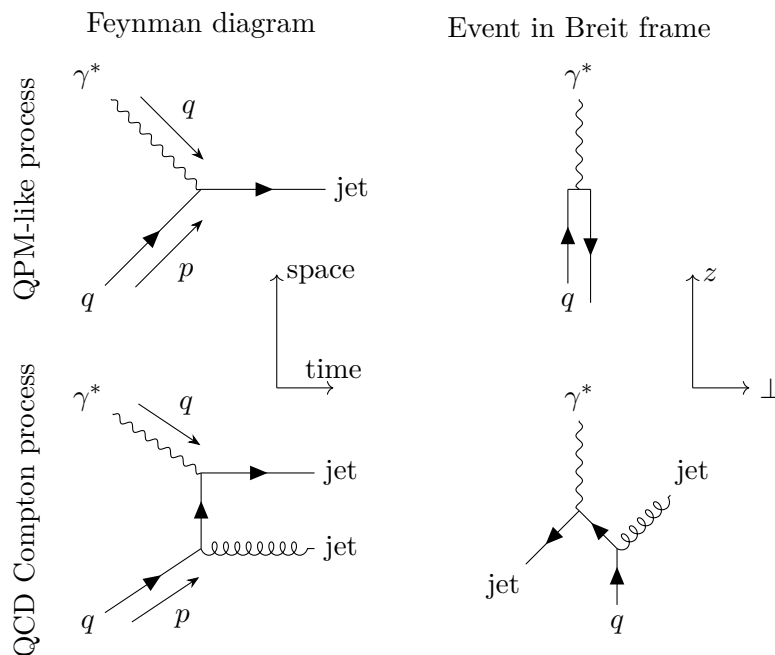
Figure 2.11: Comparison of the QPM and QCD Compton processes in the Breit reference frame. The left side shows the Feynman diagram of each process, and the right side shows the event views in the Breit frame. The scattered quark in the QPM process will satisfy $p_{\perp,\mathrm{Breit}} = 0$, while the two jets in the QCD Compton process will have $p_{\perp,\mathrm{Breit}} > 0$.

dijet events themselves but also into the relations between the two leading jets in the event. Analogous quantities can be constructed for trijet events, which are events containing three or more significant jets.[72]

Constructing observables based on a single jet can be less straightforward. Since the leading-order QCD contribution to jet production involves two jets, observables need to be insensitive to the presence of additional jets. Consider, for example, the pseudorapidity of the leading jet $\eta_{\mathrm{Breit},1}$ as an observable. The leading jet is that with the highest transverse momentum $p_{\perp,\mathrm{Breit}}$. If two jets have a similar transverse momentum, then small fluctuations in the reconstruction of $p_{\perp,\mathrm{Breit}}$ can change the identification of the leading jet and thus significantly change $\eta_{\mathrm{Breit},1}$. Therefore, this observable is unsuitable, as it is sensitive to small fluctuations.

A more suitable class are the so-called inclusive jet observables. For these observables, each jet is considered individually. If an event contains two jets of significant transverse momentum, both jets will contribute to the inclusive jet distributions. By definition, the presence of an additional jet does not change the contributions made by the other jets. This is conceptually different from the dijet or trijet observables, where each event can contribute at most once to the distributions. When considering inclusive jets, the information about the relations between jets within an event is lost, but additional information is gained regarding each individual jet. Furthermore, in dijet studies, one of the jets may fall outside the acceptance region, so the entire event must be discarded. In an inclusive jet analysis, the other jet can still contribute, leading to single-jet events even in the Breit frame. Therefore, inclusive jet studies are complementary to dijet and trijet studies.

Another advantage of inclusive jet observables is a higher precision of the theoretical calculations. Since only one jet needs to be considered explicitly, the remainder of the event can be integrated over, potentially simplifying the calculation.

One complication that arises in event-based inclusive jet studies is the treatment of statistical uncertainties. Since events can contribute multiple times to the distributions, the observables are no longer Poisson-distributed. This leads to increased statistical uncertainties and statistical correlations. Implications of this fact are discussed in detail in section 9.3.3.

One needs to distinguish inclusive jets in the Breit frame and inclusive jets in the laboratory frame. When considering inclusive jets in the Breit frame, one refers to all jets that have significant transverse momentum in the Breit frame, such that the leading-order contribution comes from dijet events. This criterion is not present when considering inclusive jets in the laboratory frame, such that the leading-order contribution is from single-jet events produced via the QPM-like process. In this thesis, the term 'inclusive jets' will always refer to inclusive jets in the Breit frame.

# CHAPTER 3

Experimental setup

The data used in this thesis were collected by the ZEUS experiment at the HERA lepton-hadron accelerator at the DESY national research centre in Hamburg, Germany. Before HERA, the DESY institute was home to three smaller circular accelerators.[25] The DESY accelerator, which gave the institute its name, was active from 1965 to 1976 and allowed to perform fixed-target experiments with an electron beam of 6 GeV. DESY has significantly contributed to validating the theory of QED.[73] The second accelerator, DORIS, was one of the first storage ring facilities worldwide and was in operation from 1974–1990. It could collide electron and positron beams at 3.5 GeV each, which was later upgraded to 5.6 GeV. DORIS enabled significant advancements towards understanding heavy quarks by studying the B-mesons' properties and placing further bounds on the top quark's mass.[74] The third accelerator at DESY was an even larger electron-positron storage ring with a length of 2.3 kilometres. PETRA operated in collider mode from 1979 to 1986, accelerating the beams to 19 GeV and later 23.5 GeV. At these energies, the gluon could be observed experimentally for the first time,[75] which played a significant role in establishing the theory of QCD and, with it, the Standard Model of particle physics. After the end of their initial use as particle colliders, many of the facilities at DESY have been reused as pre-accelerators, test-beam facilities or sources of synchrotron radiation.

3.1 The HERA collider

The fourth and largest particle accelerator at the DESY institute was HERA. This machine consisted of two storage rings of length 6.3 kilometres, one for electrons or positrons and another one for protons. Thus, HERA became the world's first and so far only high-energy lepton-hadron collider.[25] Its unique design allows the collection of valuable data that is complementary to that taken at other accelerator facilities. Similar to its predecessors, HERA's main strengths lie in the study of QCD phenomena, such as the strong coupling or the masses of the charm and bottom quarks.[76] The data collected at HERA allows the study of the proton structure at unprecedented precision. Its most significant achievement is the precise determination of the structure functions and PDFs.[77] HERA also allows the study of diffractive processes in which the proton does not break apart.[78] Furthermore, HERA is well suited to test the electroweak sector of the Standard Model at high energies and was the first facility to enable the study of electroweak unification.[79]

The electron and proton beams collided along the accelerator ring at two locations. The two general-purpose particle detectors, ZEUS and H1, were installed at these locations. Additionally,
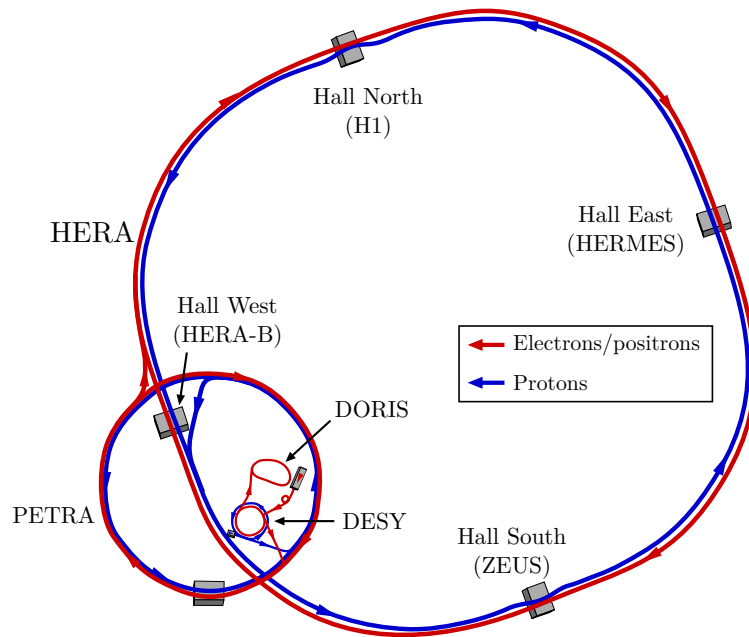
Figure 3.1: Diagram of the DESY, DORIS, PETRA and HERA accelerators.

there was the HERMES experiment, which used only the polarised electron beam to conduct fixed target experiments for studies of the spin structure of nucleons. The HERA-B experiment used only the proton beam for the study of B-mesons.[80]

Both beam types were pre-accelerated using a linear accelerator and upgraded versions of the DESY and PETRA accelerators before being injected into the HERA rings. An overview of the path of the particles through the accelerators is given in figure 3.1.

The operation of HERA can be divided into two major periods. The HERA I period lasted from 1992–2000. In this period, the proton beam was accelerated to $E_P = 820\,\text{GeV}$ and the lepton beam to $E_e = 27.5\,\text{GeV}$, resulting in a centre-of-mass energy of $\sqrt{s} \approx 300\,\text{GeV}$. Towards the end of this period, the proton beam energy was raised to $E_P = 920\,\text{GeV}$, resulting in a centre-of-mass energy of $\sqrt{s} \approx 318\,\text{GeV}$. An instantaneous luminosity, the time derivative of the integrated luminosity $\frac{\mathrm{d}L}{\mathrm{d}t}$, of $1.4 \cdot 10^{31}\text{cm}^{-1}\text{s}^{-1}$ was achieved. In the HERA II period, which lasted from 2003–2007, the instantaneous luminosity was increased to $7 \cdot 10^{31}\text{cm}^{-1}\text{s}^{-1}$.[81]

The leptons and hadrons are kept on circular trajectories using superconducting magnets installed along the HERA ring. The energy of a charged lepton in a magnetic field has two spin states, either its spin points in the same direction as the magnetic field or in the opposite. Depending on the lepton type, one of these states is preferred as it has a slightly lower energy. Therefore, as the leptons continuously emit synchrotron radiation, their spins tend to become aligned with the axis of the magnetic field, which is perpendicular to the axis of the beams. This process is known as the Sokolov-Ternov effect,[82] and it leads to the initially unpolarised lepton beam at HERA becoming transversely polarised over about 40 minutes. Due to their higher mass, hadrons are much less affected by this effect. For the HERA II period, spin rotators were installed close to the experiments. These turned the beams into longitudinally polarised ones close to the experiments. Along the ring, two polarimeters were installed to measure the degree of polarisation. The transverse polarimeter (TPOL) measured

the polarisation in the sections where the lepton beam was transversely polarised, and the longitudinal polarimeter (LPOL) measured it in the longitudinally polarised sections.[83,84]
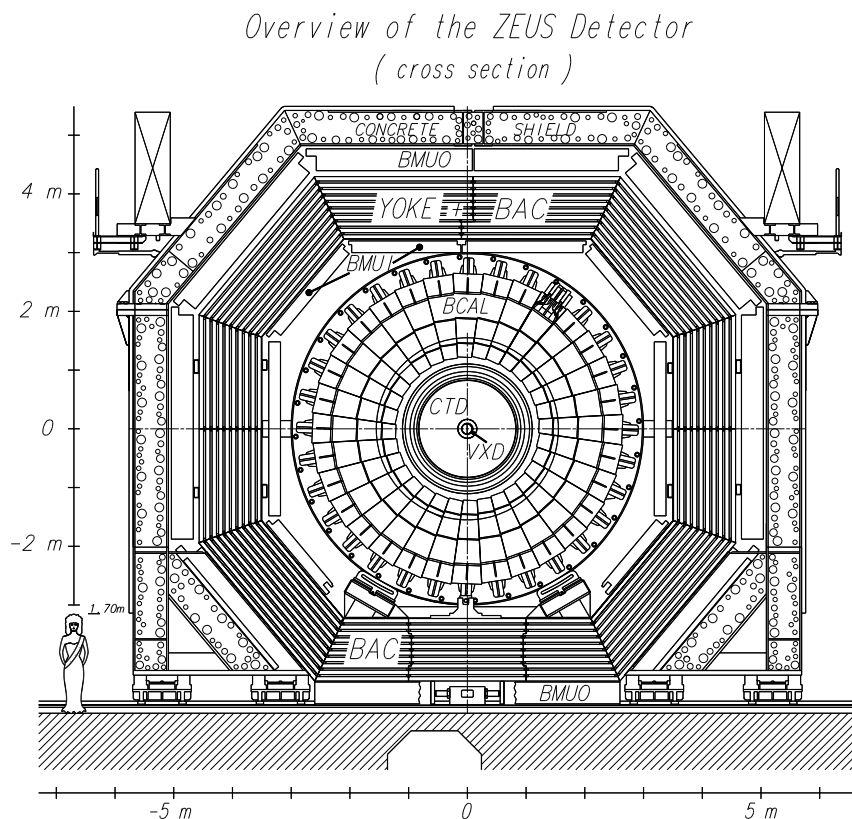
3.2 The ZEUS detector

The ZEUS detector was a general-purpose particle detector with an almost hermetic 4π coverage of the solid angle around the electron-proton collision region. It was twelve metres wide, eleven metres tall, 20 metres long and weighed about 3600 tonnes. The geometry of the detector is described in the ZEUS coordinate system. The system's origin is located at the nominal interaction point in the centre of the detector. The z-axis points in the proton beam's direction of motion also called the 'forward' direction. The direction of motion of the electron beam is referred to as the 'backward' direction. In low-Q^2 events, which are the majority of observed events, the scattered electron is scattered in the backward region. With rising Q^2, it is scattered in an increasingly forward region of the detector. The hadronic system is usually found in the forward section of the detector. The x-axis points towards the centre of the HERA ring, and the y-axis completes the right-handed coordinate system. A schematic view of the cross section of the detector in the y-z- and the x-y-planes are shown in figures 3.2 and 3.3. Due to the asymmetric nature of electron-proton collisions, the detector design is asymmetric with respect to the x-y-plane. The forward region of the detector ($z > 0$) is more strongly optimised for the detection of hadronic objects, while in the backwards region mostly leptons are of interest.[85] The following description applies to the HERA II period.

The innermost part of the detector was the microvertex detector (MVD).[87] The MVD was surrounded by the central tracking detector (CTD).[88] Both of these components were
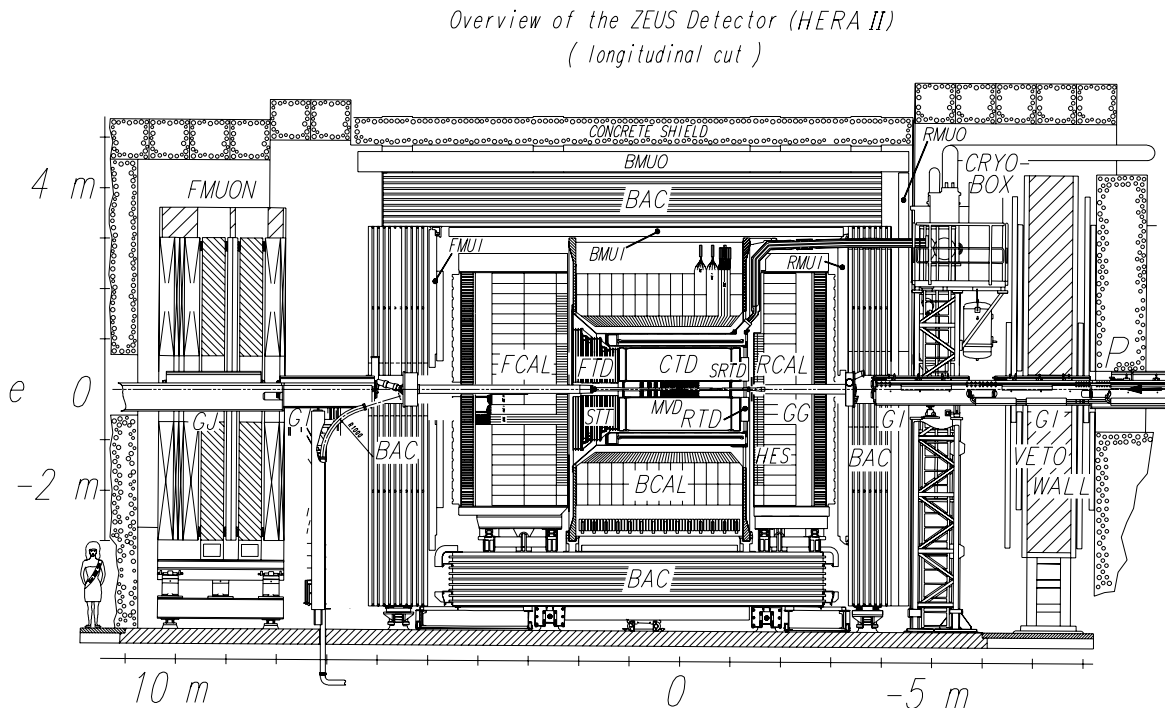
Figure 3.2: Schematic view of a cross section of the ZEUS detector in the y-z-plane.[86] Shown is the detector during the HERA II phase, as planned at the end of HERA I. The most relevant components are shown as they were at the end of the HERA II phase.

Figure 3.3: Schematic view of a cross section of the ZEUS detector in the x-y-plane during the HERA I phase.[86]

contained within the thin magnetic solenoid, that created a magnetic field of 1.43 T. The tracking efficiency was enhanced in the forward direction by the straw-tube tracker (STT) and in the backward direction by the small-angle rear tracking detector (SRTD).[89,90] The next layer of the detector was the high-resolution uraninum scintillator calorimeter (CAL), one of the distinctive features of the ZEUS detector.[91,92] The outermost section of the detector consisted of the backing calorimeter and the muon chambers.[93] The rear end of the detector was covered by an iron wall, which shielded the detector from radiation from the proton beam. The information from its integrated detectors was used to reject events produced by particles that made it through the shielding.[85]

3.2.1 Tracking detectors

The purpose of a tracking detector is to reconstruct the trajectories of charged particles. Using timing information, their energy loss over distance $\frac{\mathrm{d}E}{\mathrm{d}x}$ and the curvature of the trajectories in a magnetic field, it is possible to gain insights about a particle's type, electric charge and momentum.

The primary tracking detector at ZEUS was a drift chamber called the CTD. A drift chamber consists of an array of approximately parallel wires within a conductive enclosure. The wires are held at a high voltage compared to the chamber's walls, creating an electric field around the wires. The chamber is filled with a carefully selected mixture of gases, such that charged particles passing through the chamber can ionise atoms of the gas. The resulting
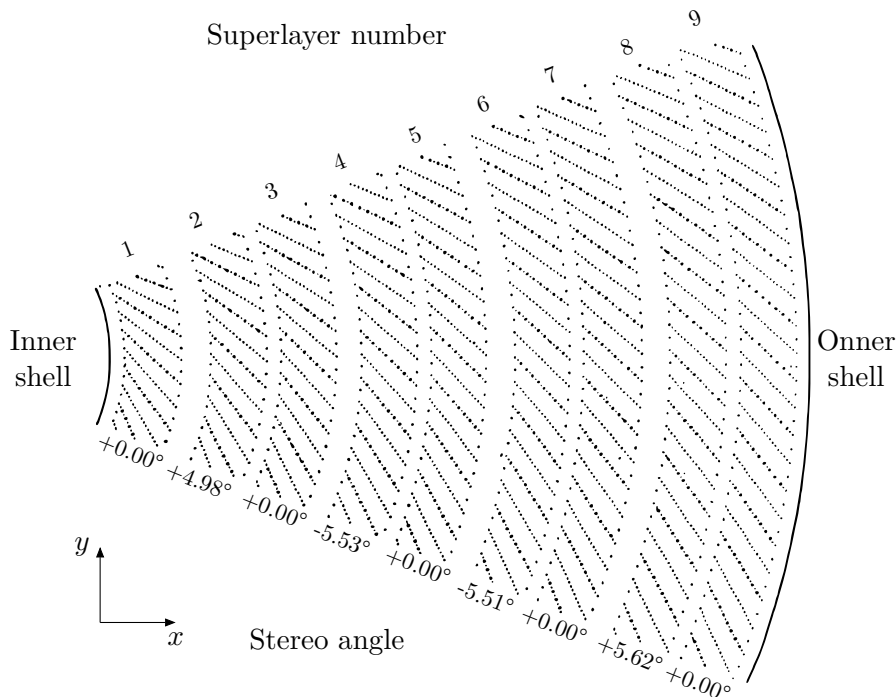
Figure 3.4: Schematic view of one octant of the cross section of the CTD in the x-y-plane. Each dot indicates one of the 4608 sense wires.[85]

free electrons are accelerated by the electric field of the wires, leading to a cascade of further ionisation. This avalanche of electrons causes a measurable pulse in the nearest wire. The particle's trajectory can be reconstructed by combining the information from every wire.

The CTD at ZEUS consisted of 72 cylindrical drift-chamber layers, organised in nine so-called superlayers, as shown in figure 3.4. If all wires were exactly parallel to the beam axis, no information on the longitudinal position of the detected particles could be extracted. To circumvent this limitation, the sense wires were slightly tilted relative to the beam axis in some superlayers, as indicated by the stereo angle.[85,94]

The CTD was supplemented by the MVD, which measured trajectories close to the primary vertex. These trajectories are especially relevant for studying short-lived particles, such as heavy quark flavours, but also contribute to the momentum measurement of longer-lived particles. The MVD consisted of a barrel and a forward section containing three cylindrical and four planar layers of silicon strip sensors, respectively. A silicon strip sensor is a solid-state device capable of measuring the position of particles passing through it. It consists of narrow silicon strips that are doped to act as diodes. A voltage is applied in the reverse direction such that no current flows. Charged particles passing through these strips will cause a detectable ionising current. The particle's trajectory can be reconstructed by stacking multiple sensors behind each other.[87]

The CTD covered particles with polar angles in the range $15° < \theta < 164°$. Trajectories close to the beam pipe were reconstructed by the dedicated STT and SRTD, covering the polar angles $5° < \theta < 23°$ and $162° < \theta < 172°$ respectively. The CTD achieved a resolution of the position of hits of $\approx 200\,\mu m$ in the x-y-plane and of $\approx 2\,mm$ in z. The MVD had a spatial resolution of $\approx 20\,\mu m$, which improved the resolution of primary vertex reconstruction over the

CTD alone. For tracks that pass through all nine superlayers and the MVD, the momentum resolution during the HERA II period was $\sigma(p_\perp)/p_\perp = 0.0029p_\perp \oplus 0.0081 \oplus 0.0012/p_\perp$ with $p_\perp$ in GeV.[88]

3.2.2 Calorimeter

After passing through the tracker, particles enter the calorimeter section of the detector, which aims to absorb and measure their energy. This is relevant for neutral particles that cannot be detected in the tracker but also for charged particles, as the calorimeter tends to provide a better energy resolution at high energies. The ZEUS calorimeter is a sampling calorimeter, meaning that each cell is constructed as a tower of alternating plates of a scintillator and absorber material. As a particle traverses the absorber plate, it will initiate a particle shower, which deposits its energy in the scintillator plate. A scintillator is a material that emits photons of a particular wavelength upon excitation. These photons are then converted to electrical signals by photomultipliers.[85]

The ZEUS calorimeter used depleted uranium as the absorber material. The natural radiation of the uranium was used as a stable and well-understood background signal which allowed a precise and uniform calibration of every channel. A very uniform response to both electromagnetic and hadronic showers was achieved by tuning the thickness of the uranium and scintillator plates.[85] The ZEUS calorimeter is called a compensating calorimeter because it achieved an equal response to electromagnetic and hadronic showers. This property improves the resolution of the energy measurement and sets the ZEUS detector apart from many other experiments.[95]

The CAL was divided into three components: the forward, barrel and rear calorimeter (FCAL, BCAL and RCAL) as shown in figure 3.5. These components covered ranges of the polar angle of $2.2° < \theta < 39.9°$, $36.7° < \theta < 129.1°$ and $128.1° < \theta < 176.5°$ respectively.[91,92]

Showers initiated via electromagnetic objects tend to occur early in the calorimeter and are narrower than hadronic showers. To provide a high-precision measurement of both types of showers, each of the three CAL components was further subdivided into an electromagnetic and one or two hadronic sections (EMC and HAC). The cells in the EMC sections were finer and located in front of the HAC cells, i.e. closer to the interaction point. The left part of figure 3.6 shows the decomposition of the RCAL into cells. The right part of the figure illustrates how each cell was built as a tower, on the example of an FCAL module.[85]

Electrons and hadrons close to the rear beam pipe were further distinguished using the hadron-electron separator (HES).[96] While isolated electromagnetic and hadronic objects are straightforward to identify using the CAL alone, electrons that overlap with hadronic showers are much more challenging to identify. The HES was a highly segmented detector made from silicon diodes and was inserted into the RCAL close to the maximum of the electromagnetic showers. This also improved the spacial resolution of the calorimeter for electromagnetic showers.

The HAC cells of the FCAL and RCAL had a dimension $20\,\text{cm} \times 20\,\text{cm}$, while the EMC cells had a size of $20\,\text{cm} \times 10\,\text{cm}$ in the RCAL and $20\,\text{cm} \times 5\,\text{cm}$ in the FCAL. In the BCAL, the cells were wedge-shaped and slightly rotated in the azimuthal plane. This ensured photons cannot escape undetected through gaps between the modules. At the inner edge of the hadronic section of the BCAL, HAC cells had a size of $24.4\,\text{cm} \times 27.1\,\text{cm}$ and EMC cells measured $4.9\,\text{cm} \times 23.3\,\text{cm}$. This small cell size allowed an especially precise reconstruction of particle positions in the central region of the detector. Under test beam conditions, the energy resolution of the CAL was determined to be $\sigma(E)/E = 0.18/\sqrt{E} \oplus 1\%$ for electrons
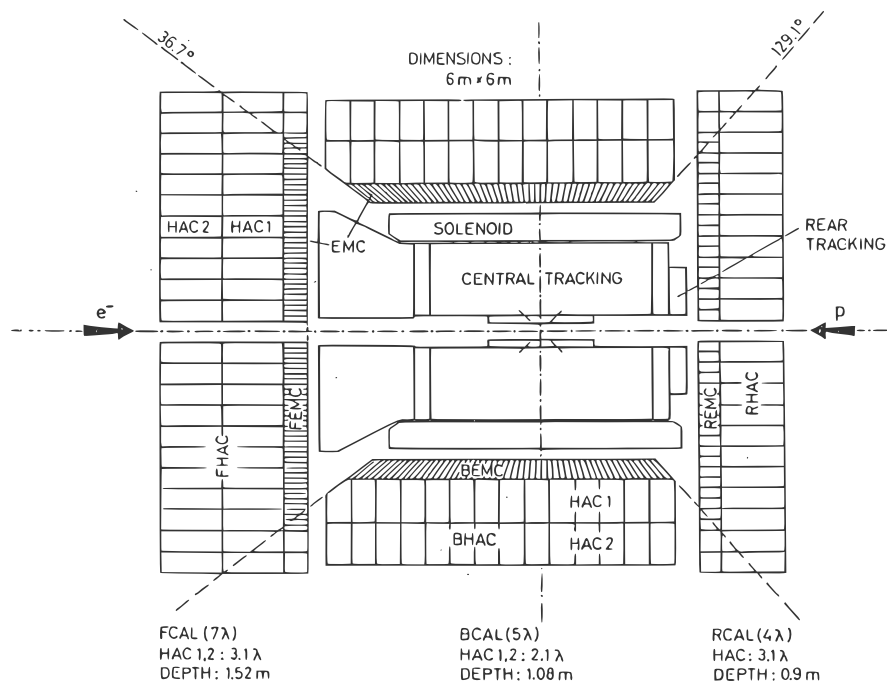
Figure 3.5: Schematic view of the cross section of the CAL in the r-z-plane from the HERA I period.[85] For the HERA II period, the locations of some components were slightly adjusted.

and $\sigma(E)/E = 0.35/\sqrt{E} \oplus 1\%$ for hadrons with E in GeV.[97,98] The latter made the CAL one of the most precise hadronic calorimeters ever used in a particle detector.[95]

3.2.3 Luminosity system

Since the luminosity is a direct ingredient to every measured cross section, its accurate determination is essential to every particle detector. During the HERA II period, the ZEUS detector used two independent systems to measure instantaneous luminosity. Both relied on measuring the number of photons produced in the bremsstrahlung (Bethe-Heitler) process $ep \rightarrow ep\gamma$ by the colliding beams. This QED process has a large and precisely known cross section.

These photons were measured about 100 m away from the interaction point. They were naturally separated from the electron and proton beams, as the beams were kept on a circular trajectory within the HERA accelerator by bending magnets. As the photons passed through a thin window in the beam pipe, a fraction of them converted into e^+e^- pairs. An additional magnet separated these electrons, positrons and unconverted photons. The number of photons was measured by the lead-scintillator photon calorimeter (PCAL) and the number of e^+e^- pairs by the luminosity spectrometer (SPEC).[99,100] During the HERA II period, the uncertainty of the luminosity measurement was between 1.8% and 2.5%.

3.2.4 Trigger system

Bunch crossings at HERA occurred every 96 ns, corresponding to a rate of about 10 MHz. Less than 1% of bunch crossings resulted in interactions. High-Q^2 DIS events occurred only with a rate of a few Hertz, which is also the rate at which the data acquisition system can
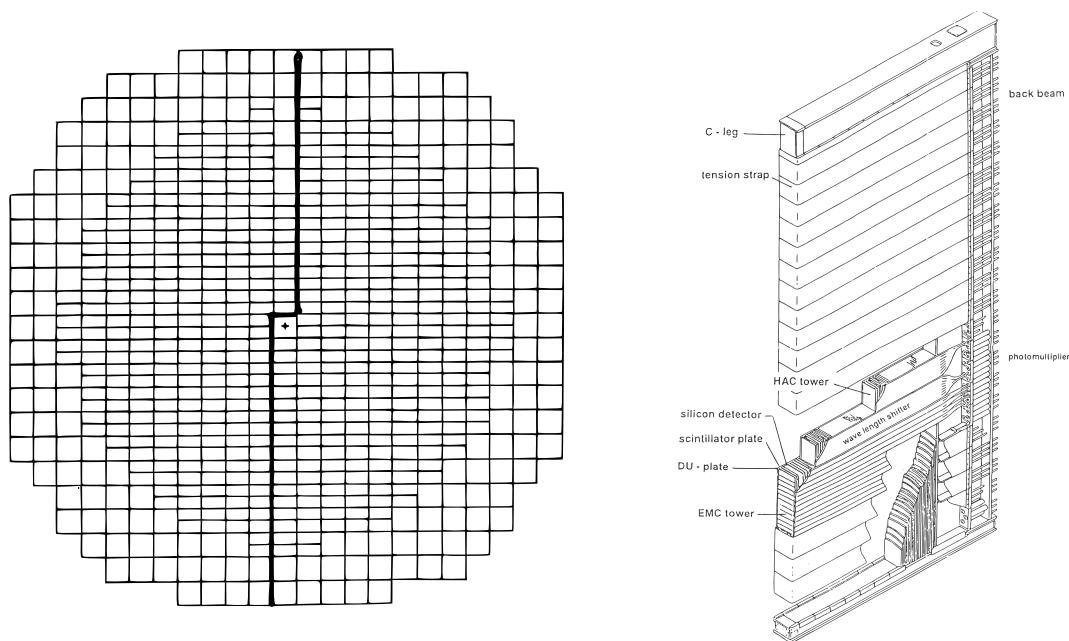
Figure 3.6: Left: schematic view of the RCAL, as seen from the interaction point. Larger squares indicate the HAC cell, while the smaller rectangles represent the EMC cells. The FCAL is built similarly, except for a slightly different number of HAC cells and the EMC cells being half the size. Right: schematic view of a module constituting one column of the FCAL.[85]

store event information. The ZEUS detector used a three-level trigger system to decide which events should be stored for analysis.[85,101]

The task of the first-level trigger (FLT) was to reduce the input rate to about 1 kHz. Each detector component had its own local FLT logic that provided a limited set of lower-resolution information to the global FLT. The global FLT made its decision based on simple features of the event, such as isolated energy deposits in the CAL and the number of tracks in the CTD. While the processing of the FLT was ongoing, the complete event data was stored in a pipeline, which gave the trigger system about 5 μs to make a decision. The FLT used programmable logic devices to allow for fast reconfiguration of the trigger conditions.[85]

If an event was selected by the FLT, its information was forwarded to the second-level trigger (SLT). Each detector component also had its own local SLT logic that could run a more sophisticated analysis based on the complete information at full resolution. At this level, timing information was used to reject events inconsistent with the bunch crossing time. Furthermore, events were selected using requirements based on energy deposits and tracks, such as the primary vertex position. These requirements rejected about 97% of beam-gas background events that made it through the FLT while retaining more than 95% of DIS events with $Q^2 > 100 \, \text{GeV}^2$. The SLT reduced the event rate to about 100 Hz.[85]

The third-level trigger (TLT) consisted of multiple computers at a data centre, running a reduced version of the full offline analysis software. This allowed for a complex decision-making process involving e.g. more elaborate track finding and vertex fitting algorithms. The TLT also computed kinematic quantities such as Q^2, x_{Bj} and $E_\perp$ and used them to select events. The output rate of the TLT was about 3-5 Hz. Events selected by the TLT were stored to tape for full offline analysis later on.[85]

CHAPTER 4

Context of measurement

The theory of QCD can describe a diverse range of observations and remains under intense experimental and theoretical investigation. In the zero-mass scheme of perturbative QCD, the theory contains just a single free parameter, the strong coupling constant, $\alpha_s(M_Z^2)$. At hadron colliders, non-perturbative QCD is mostly described by the PDFs. These are essential inputs for many practical applications involving QCD. Therefore, precisely determining these quantities is highly desirable for any analysis involving QCD effects. Even processes involving QED or electroweak final states can be significantly affected by QCD corrections and uncertainties.[102]

One of the primary motivations in constructing the HERA collider were detailed studies of QCD. Since only a single proton is involved, the experimental environment is much cleaner than at, e.g. proton-proton colliders. The PDFs enter calculations only once instead of twice, leading to reduced uncertainties on the predictions. The perhaps even more important consequence is that the PDFs can be accurately determined from the data. Because of this, the HERA inclusive DIS data have become the basis of every recent PDF determination.[77,103–108]

4.1 HERA inclusive deep inelastic scattering data

Throughout its operation, HERA has provided different beam configurations. Protons were accelerated to $460\,\mathrm{GeV}$, $575\,\mathrm{GeV}$, $820\,\mathrm{GeV}$ or $920\,\mathrm{GeV}$ and collided with electrons or positrons at $27.5\,\mathrm{GeV}$. The H1 and ZEUS collaborations have provided precise measurements of the low-Q^2 NC DIS, high-Q^2 NC DIS and high-Q^2 CC DIS processes under all of these conditions. In this context, 'low-Q^2' refers to the region $3\,\mathrm{GeV}^2 \lesssim Q^2 \lesssim 100\,\mathrm{GeV}^2$ and 'high-Q^2' refers to $Q^2 \gtrsim 150\,\mathrm{GeV}^2$. More than 40 individual datasets were recorded, corresponding to an integrated luminosity of about $1\,\mathrm{fb}^{-1}$, split evenly between the two experiments. Together, these datasets span the range $Q^2 \in (0.045, 50000)\,\mathrm{GeV}^2$ and $x_{\mathrm{Bj}} \in (6 \times 10^{-7}, 0.65)$ for the NC process and $Q^2 \in (200, 50000)\,\mathrm{GeV}^2$ and $x_{\mathrm{Bj}} \in (0.013, 0.40)$ for the CC process.

To simplify the consistent treatment of these measurements, they were averaged and combined into a single set of large datasets.[77] Since the H1 and ZEUS measurements do not use identical phase space definitions, data points were first translated to one of two common (Q^2, x_{Bj}) grids. For this translation, predictions of the cross sections at the measured and translated (Q^2, x_{Bj}) were needed. These were obtained from fits to the data points themselves, which reduced the dependence of the results on the assumed model.

Cross sections at the grid points were then determined by averaging the measurements while considering their statistical and systematic uncertainties. The uncertainties were propagated to the averaged values, which required careful consideration of their correlation between data
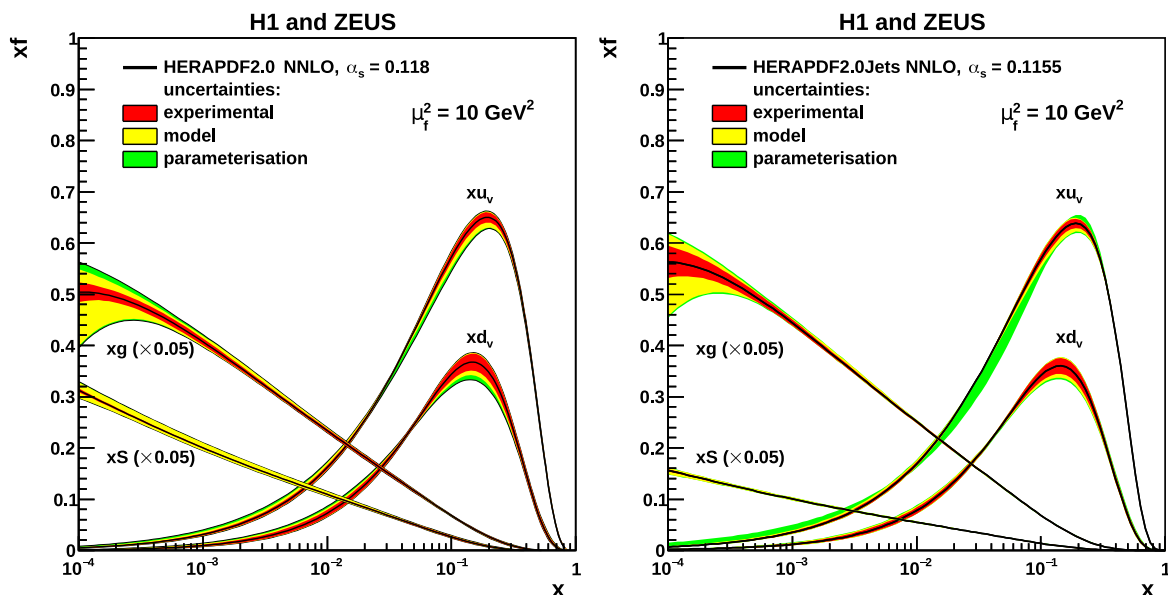
Figure 4.1: Comparison of the HERAPDF2.0 NNLO (left)[77] and Jets NNLO (right)[103] PDF sets. The former was determined at an externally chosen value of $\alpha_s(M_Z^2)$, while the latter uses the value preferred by the jet data. The coloured bands show the different contributions to the PDF uncertainty; see chapter 12 for details. The variable x in the figures corresponds to the momentum fraction ξ in the notation of this thesis. Note that the sea-quark distribution is defined as $xS = 2x(\bar{U} + \bar{D})$ in the left panel, and as $xS = x(\bar{U} + \bar{D})$ in the right panel.

points and different datasets. Since the two experiments have different detectors and used different reconstruction methods, the uncertainties calibrated each other to some extent, which led to a significant reduction of the uncertainties of the combined dataset. The result of this combination are seven datasets corresponding to different centre-of-mass energies, different lepton types and to the NC or CC process. In total, this combination contains 1307 averaged data points that were determined from 2927 measured points. Collectively, these points are referred to as the 'HERA inclusive DIS dataset'.[77]

A similar combination was performed using 13 datasets on heavy quark production in the kinematic region $Q^2 \in (2.5, 2000)\,\text{GeV}^2$ and $x_{\text{Bj}} \in (3 \times 10^{-5}, 5 \times 10^{-2})$.[76] This combined dataset is especially valuable for determining the masses of the charm and bottom quarks.

These datasets were used to determine the HERAPDF2.0 PDF sets at next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) in QCD. These determinations use HERA data exclusively and thus serve as a valuable benchmark for other PDF determinations. Since every other PDF determination also uses these datasets, the impact of additional data can be evaluated by comparing the determined PDFs to HERAPDF. Furthermore, datasets from other experiments could potentially be biased by physics beyond the Standard Model, which would bias the determined PDFs. This is less likely in the HERA environment,[109] such that the HERAPDF sets can be used as a reference in new-physics searches. Multiple variants of the HERAPDF sets are determined corresponding to different datasets used as input, different model parameters, different parameterisations of the PDFs and different QCD schemes. The nominal HERAPDF2.0 NNLO PDF set is shown in figure 4.1.

4.2 HERA jet data

Jet measurements at HERA also profit from the clean experimental environment and well-understood PDF uncertainties. At the same time, suitably defined jet measurements are directly sensitive to the strong coupling. This makes them ideally suited for determining this coupling, possibly in conjunction with the PDFs.

Jet measurements in PHP were extensively used to study QCD and determine the PDFs and $\alpha_\mathrm{s}(M_Z^2)$.[110] PHP events in the laboratory frame behave similar to DIS events in the Breit frame. Therefore, they are also directly sensitive to $\alpha_\mathrm{s}(M_Z^2)$. Jet data in PHP are complementary to jet data in DIS and put further constraints, especially on the gluon distribution. Due to their high statistics, these jets in PHP can be measured at high precision and in higher $p_\perp$ regions than are accessible in DIS. However, theoretical predictions of the PHP process are more challenging and have larger uncertainties than their DIS counterparts. Since predictions for jet production in PHP are currently only available at NLO in QCD, these data are not used in the latest NNLO determinations of the PDFs and $\alpha_\mathrm{s}(M_Z^2)$.

Similar to the inclusive DIS data, data on jet production in DIS are differentiated into the low-Q^2 and high-Q^2 region. While low-Q^2 data profit from higher statistics and thus lower experimental uncertainties, they suffer from increased theoretical uncertainties. An overview of existing jet measurements at HERA is given in table 4.1.

The DIS datasets were used in the HERAPDF2.0Jets analyses, which are combined determinations of the PDFs and $\alpha_\mathrm{s}(M_Z^2)$. The HERAPDF2.0 analyses without jets were performed at fixed values of $\alpha_\mathrm{s}(M_Z^2)$ because the inclusive DIS data alone do not sufficiently constrain $\alpha_\mathrm{s}(M_Z^2)$, see figure 4.2. Including the jet data additionally allows determining $\alpha_\mathrm{s}(M_Z^2)$ from the data, leading to a more consistent set of PDFs.

Table 4.1 also indicates which jet datasets were used in the HERAPDF2.0Jets analyses. The H1 HERA II dataset on low-Q^2 inclusive jet, dijet and trijet production was not yet published at the time of the NLO analysis. Trijet datasets had to be excluded from the NNLO analysis since no NNLO predictions are available. Not all HERA I datasets could be used, as they are subject to unknown and non-trivial statistical correlations. During the HERA II period, H1 managed to avoid this issue by performing combined analyses using a matrix-unfolding approach in which these correlations are also determined. Some datasets were excluded from the analyses due to their large theoretical uncertainties.

The $\alpha_\mathrm{s}(M_Z^2)$ values determined by the HERAPDF2.0Jets analyses are

$$\alpha_\mathrm{s}(M_Z^2) = 0.1183 \pm 0.0009 \text{ (exp./fit)} \pm 0.0005 \text{ (mod./par.)}^{+0.0037}_{-0.0030} \text{ (scale)} \pm 0.0012 \text{ (hadr.)}$$

at NLO and

$$\alpha_\mathrm{s}(M_Z^2) = 0.1156 \pm 0.0011 \text{ (exp./fit)}^{+0.0001}_{-0.0002} \text{ (mod./par.)} \pm 0.0029 \text{ (scale)}$$

at NNLO.[77,103] The experimental/fit uncertainty is the uncertainty of the fit and includes the uncertainties of the experimental input. At NNLO, it also includes the hadronisation uncertainty that is given separately at NLO. The model/parameterisation uncertainty expresses the remaining uncertainties of the analysis procedure, and the scale uncertainty corresponds to the uncertainty of the theoretical predictions. These uncertainties are discussed in detail in chapter 12.

The HERAPDF2.0Jets NNLO PDF set is compared to the determination without jets in figure 4.1. The main contribution of the jet data is to constrain the value of $\alpha_\mathrm{s}(M_Z^2)$. Since this value is highly correlated to the PDFs, their shape is also different, especially that of the

Year	93	94	95	96	97	98	99	00	03	04	05	06	07
L_{HERA} [pb^{-1}]	1	6	12	17	36	8	46	66	7	78	205	204	63

H1	**PHP**	inclusive jets				[111]									
		dijets		112	[113, 114]			[115]							
		trijets													
	low-Q^2	inclusive jets				[116]		[117]*,**				[118]**		
		dijets						[117]				[118]**		
		trijets						[117]				[118]		
	high-Q^2	inclusive jets				[119]		[120]*,**			[121]*,**, [118]**		
		dijets									[121]*,**		
		trijets									[121]*		
ZEUS	**PHP**	inclusive jets	122	123								[124]		
		dijets	125	126	[127–129]			[130]						
		trijets			[131]									
	low-Q^2	inclusive jets											
		dijets						[132]					
		trijets											
	high-Q^2	inclusive jets			[133]*,**		[134]			this thesis		
		dijets			[135]		[71]*,**			[71]*,**		
		trijets								preliminary[72,136]		

Table 4.1: Overview of jet measurements in PHP and in low-Q^2 ($3 \,\text{GeV}^2 \lesssim Q^2 \lesssim 100 \,\text{GeV}^2$) and high-$Q^2$ ($Q^2 \gtrsim 150 \,\text{GeV}$) NC DIS in the Breit frame at HERA. Each column indicates one year of running. The approximate luminosity delivered by the HERA accelerator for each year is listed and emphasised by the column width.[137] This allows judging of the relative significance of the different run periods. The useable luminosity collected for analysis is smaller and differs between the experiments. The markers '*' and '**' indicate the datasets used in the HERAPDF2.0Jets analyses at NLO and NNLO, respectively.[77,103]

gluon PDF. Another contribution of the jet data is a slight reduction of the uncertainty of the gluon PDF. These PDF sets differ mostly because of the updated value of $\alpha_\text{s}(M_Z^2)$. If PDFs were determined from the inclusive data alone, but at the updated $\alpha_\text{s}(M_Z^2)$ value, they would be similar to the determination including jets. However, since this updated $\alpha_\text{s}(M_Z^2)$ value is only possible due to the jets, the change of the PDFs should be considered as a contribution of the jets.

This thesis adds a measurement of the inclusive jet cross sections using data collected from 2004 to 2007 at ZEUS. This is the most relevant jet dataset still missing from HERA since high-Q^2 DIS jet datasets are currently the most precise way to constrain $\alpha_\text{s}(M_Z^2)$. This dataset contributes significantly to reducing uncertainties as it uses about two-thirds of the entire available luminosity. To make the data usable with previous datasets, the statistical correlations to the ZEUS HERA II dijet dataset[71] also need to be determined.
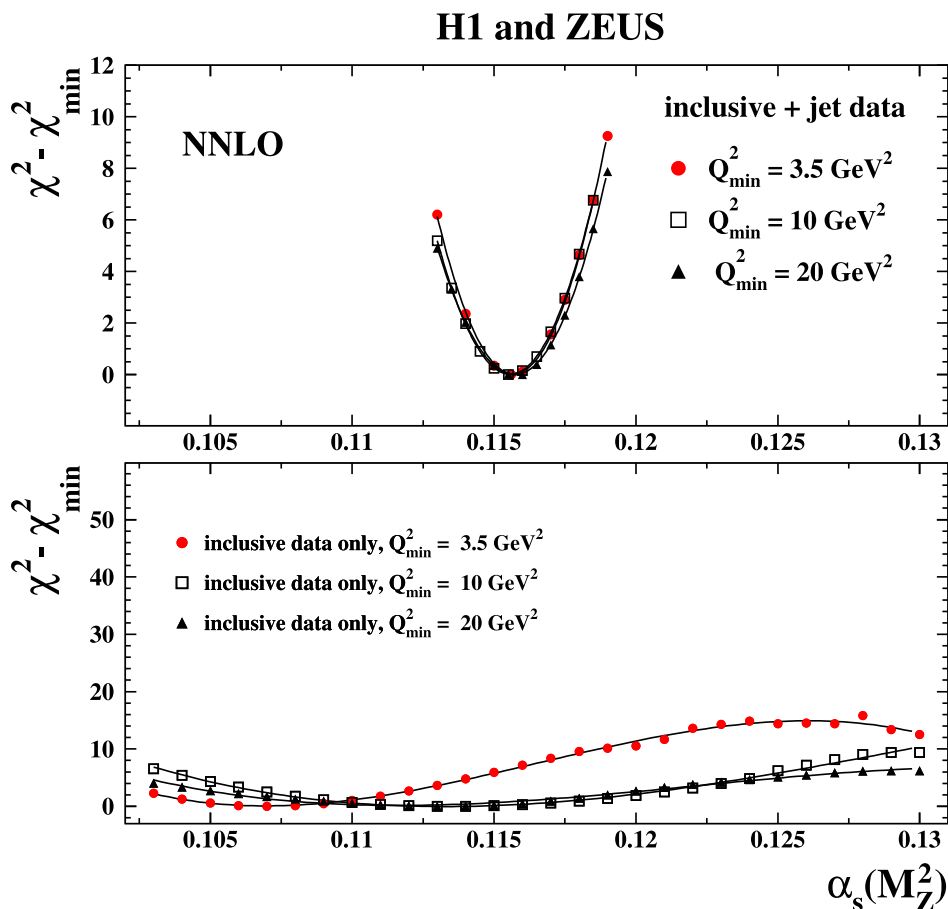
Figure 4.2: The χ^2 of the fit as a function of $\alpha_s(M_Z^2)$, as determined from inclusive DIS data alone (bottom) and including jet data (top).[103] In case of the inclusive data alone, the minimum is broad and depends significantly on the choice of Q_{min}^2, indicating that these data alone are not well suited to determine $\alpha_s(M_Z^2)$. Including jet data, the minimum is much more narrow and independent of Q_{min}^2.

4.3 Other measurements sensitive to the strong coupling

While HERA is the only collider experiment designed for the study of DIS, various fixed target experiments exist that allow the study of electron-, muon- or neutrino-induced DIS.[138–140] Apart from DIS processes, essentially any observable that involves QCD interactions can be used to constrain $\alpha_s(M_Z^2)$.

One such observable are trijet cross sections at e^+e^- collisions. A possible Feynman diagram for this process is shown in the left panel of figure 4.3. Also suitable are so-called jet shape observables, which study the distribution of particles within each jet. The datasets from the LEP[141,142] and PETRA[143] colliders have been reanalysed at NNLO QCD accuracy and are thus suitable to be compared to other state-of-the-art measurements.

Another well-suited type of experiment to study QCD are proton-proton and proton-antiproton collisions, e.g. at the LHC and Tevatron colliders.[144] Commonly studied processes are jet production,[145] top quark production[146] or the production of electroweak bosons.[147] Selected Feynman diagrams are shown in figure 4.3. These measurements profit from their large datasets and high experimental precision.
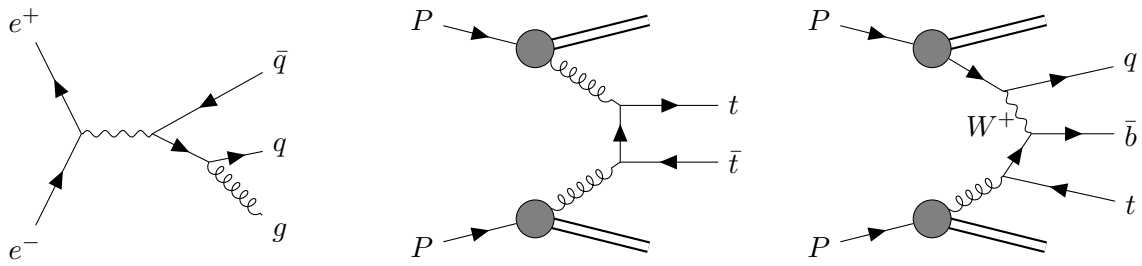
Figure 4.3: Processes involving QCD interactions. Left: trijet production at e^+e^- collisions. Middle: top quark pair production in *pp* collisions. Right: single top quark production in *pp* collisions.

Mesons consisting of a quark and its own antiquark are called quarkonium. Analyses of the masses of heavy quarkonium states are also sensitive to $\alpha_{\mathrm{s}}(M_Z^2)$.[148,149] Another possible observable is the hadronic decay of τ leptons.[150] This determination probes the strong coupling at the low energy $m_\tau = 1.78\,\mathrm{GeV}$ and is therefore especially suited to study the scale dependence of the $\alpha_{\mathrm{s}}(\mu^2)$, see section 12.2. Fits of the electroweak sector of the Standard Model are sensitive to $\alpha_{\mathrm{s}}(M_Z^2)$ since many of the involved processes also involve QCD interactions.[151]

4.4 Further determinations of the strong coupling

The measurements mentioned in the previous section are often performed as single-parameter fits of the strong coupling or include a few additional parameters, such as some of the quark masses.[121,143,146] At lepton-hadron and hadron-hadron colliders, this involves taking a fixed PDF set at different values of $\alpha_{\mathrm{s}}(M_Z^2)$. It has been argued that this approach introduces a systematic bias in the determination.[17,152] Preferably, the PDFs should be fitted at the same time. Such simultaneous fits are also often used to determine $\alpha_{\mathrm{s}}(M_Z^2)$.[145,153]

Combined determinations of the PDFs and $\alpha_{\mathrm{s}}(M_Z^2)$ have been performed by various groups. Some of the latest determinations are ABMP16,[105] CT18,[106] NNPDF3.1,[154] or MSHT20.[155] In addition to the HERA inclusive DIS and heavy quark datasets, these determinations use additional datasets, such as

- fixed target DIS data, including data on muon-proton collisions,
- neutrino induced semi-inclusive DIS $\nu p \to \nu\mu\mu + X$,
- Drell-Yan and W boson hadro-production processes $pp \to \gamma/Z/W \to l\bar{l}/l\nu$ at fixed target or collider experiments,
- various processes involving weak boson production, jet production and weak boson+jet production at proton-proton and proton-antiproton collisions,
- measurements of single-top production in proton-proton and proton-antiproton collisions,
- data on $t\bar{t}$ production in proton-proton collisions,
- data on proton-ion collisions

and potentially more. Due to their large number of additional datasets, these determinations can achieve greater precision than is possible from HERA data alone.

Currently, the most precise determinations of $\alpha_{\mathrm{s}}(M_Z^2)$ are obtained from lattice QCD calculations.[156] These simulations discretise spacetime on a lattice and solve the QCD field

equations numerically. These calculations are mathematically well-defined since the finite lattice spacing and the finite lattice volume naturally introduce ultraviolet and infrared momentum cut-offs.[157] The QCD continuum is recovered by studying the dependence of the results of these simulations on the parameters of the lattice and then extrapolating them to an infinitely large lattice with infinitesimally small spacing. Most importantly, these simulations are non-perturbative and thus allow the study of effects that are theoretically not accessible otherwise, such as the masses and decay rates of hadrons or properties of the quark-gluon plasma. Since these calculations are computationally very expensive, they are currently limited to relatively small physical systems on the order of the proton size.[158] Parameters, such as the strong coupling constant, can be determined from these simulations by fitting them to experimental inputs, usually the decay rates of hadrons involving heavy quarks.

The Particle Data Group collects many determinations of $\alpha_{\mathrm{s}}(M_Z^2)$ and combines them into the so-called world average value. At the time of writing, this average value is[17]

$$\alpha_{\mathrm{s}}(M_Z^2) = 0.1179 \pm 0.0009.$$

This value is computed as the unweighted average of the lattice QCD determination, $\alpha_{\mathrm{s}}(M_Z^2) = 0.1182 \pm 0.0008$, and the average of all remaining determinations, $\alpha_{\mathrm{s}}(M_Z^2) = 0.1176 \pm 0.0010$.

CHAPTER 5

Calculations of observables in particle physics

In particle physics, theoretical predictions of physical observables are used for two distinct purposes. Fixed-order perturbative calculations aim to calculate observables at the highest possible accuracy achievable from first principles. To this end, they use sophisticated physical and mathematical techniques to compute a large number of Feynman diagrams, often taking into account contributions up to NNLO in perturbation theory in QCD and even higher orders in QED.[16,159] These predictions present the most accurate models of particle physics available. Comparing them to data allows the validation of state-of-the-art theoretical models and the determination of model parameters. These comparisons are the primary motivation for performing a measurement, as physical conclusions can be drawn from them.

The other essential type of theoretical calculations are MC simulations. These consist of fixed-order matrix elements and various additional components. They serve a different purpose than fixed-order calculations. The samples generated from MC simulations are essential to performing precise particle physics measurements. They are used to understand the data, especially how they are affected by detector effects and to derive corrections. They are also used to model backgrounds. For this approach to be valid, the MC samples must describe the data accurately, even if no suitable theoretical model is available. Therefore, MC samples are often reweighted or otherwise modified to improve the agreement with the data. At this point, MC samples are no longer predictions, but rather a tool used to perform a measurement.

5.1 Fixed-order calculations

A computation that considers only those Feynman diagrams containing the lowest possible number of vertices for a particular final state is referred to as a leading-order calculation. Depending on the investigated process, the leading-order contribution can already contain higher powers of the coupling. For example, the leading-order contribution to dijet production in DIS is $\mathcal{O}(\alpha_s^1)$, while the leading-order contribution to trijet production is $\mathcal{O}(\alpha_s^2)$. A NLO calculation also considers diagrams that contain one more power of the coupling. At NLO, two types of higher-order diagrams are possible, so-called real and virtual corrections. Real correction diagrams describe the emission of additional particles. Virtual correction diagrams involve loops. One more power of the coupling constant may be achieved by the interference of the real corrections amongst themselves or by interference of the virtual corrections with the leading-order diagram. As an example, figure 5.1 shows the leading-order and NLO QCD diagrams that contribute to the boson-gluon fusion process. The number of diagrams increases rapidly with the considered order. At NNLO QCD, hundreds of diagrams need to
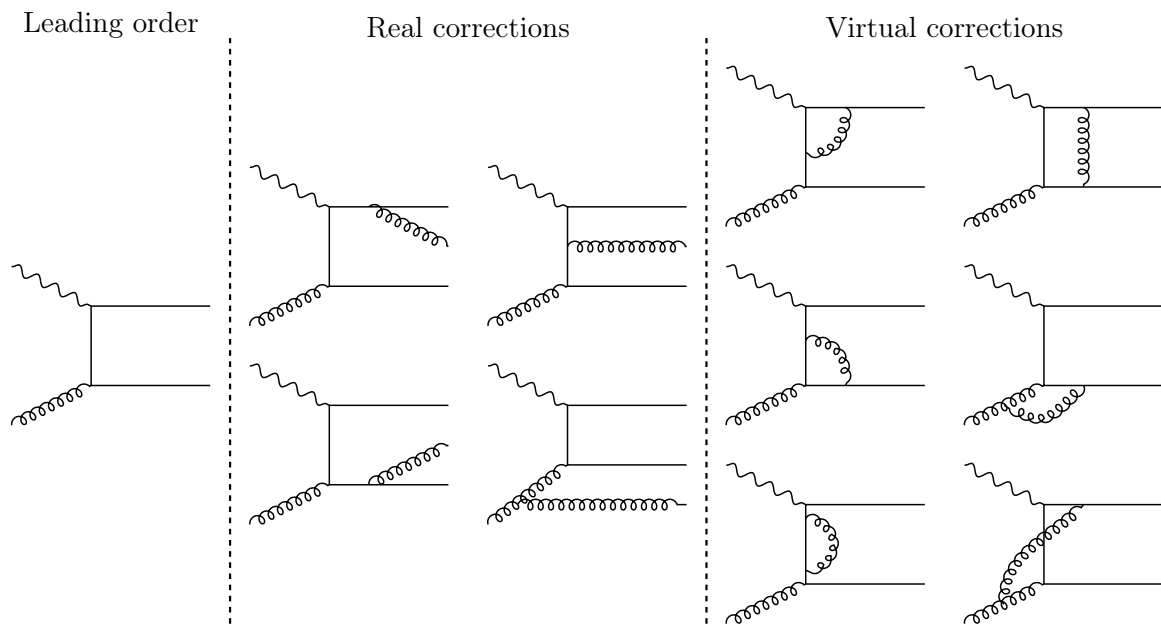
Figure 5.1: Leading-order ($\mathcal{O}(\alpha_\text{s}^1)$) Feynman graph for the boson-gluon fusion process as well as real and virtual QCD corrections at NLO. Each diagram will contribute twice since the direction of the quark line can be chosen either way. Loop corrections of the propagators of the external particles do not need to be considered explicitly here, as they can be absorbed in the theory parameters during renormalisation.

be considered. In addition to these QCD diagrams, higher-order diagrams in QED and the weak interaction should also be considered. Currently, many processes have been calculated at NNLO QCD accuracy.[159,160]

Scale dependence of cross section predictions When truncating the power series at a fixed order, the renormalisation-scale dependence no longer cancels exactly. A similar argument applies to the PDFs, which leads to a dependence on the factorisation scale. Consequently, computed observables depend on these scales. Since these dependencies cannot be avoided, it is crucial to choose sensible values for the scales and consider the uncertainty associated with these choices. The scales are usually set to a value related to the physical process under consideration. In jet production in DIS, the factorisation scale is often set to the momentum transfer $\mu_\text{f}^2 = Q^2$, as this is how the PDFs are defined. The renormalisation scale is usually set to a combination of the momentum transfer and the transverse momentum of the jet, e.g. $\mu_\text{r}^2 = Q^2 + p_\perp^2$. The uncertainty associated with these choices is estimated by varying both scales independently up and down, usually by a factor of two, and recording the change of the computed observable.[159] The scale uncertainty is often one of the dominant uncertainties of a theoretical prediction. Since the scale dependence decreases with increasing perturbative order, it is desirable to use predictions of the highest possible order to obtain accurate results.

At NNLO, the cross section can be written as

$$\hat{\sigma} = \alpha_\text{s}{}^n \, \hat{\sigma}^{(\text{LO})} + \alpha_\text{s}{}^{n+1} \, \hat{\sigma}^{(\text{NLO})} + \alpha_\text{s}{}^{n+2} \, \hat{\sigma}^{(\text{NNLO})} + \mathcal{O}\big(\alpha_\text{s}{}^{n+3}\big), \tag{5.1}$$

where n is the lowest order of α_s at which the cross section is non-zero and the $\hat{\sigma}$ on the left side of the equation is the lepton-parton matrix element introduced in equation (2.32). In

this equation, all quantities depend explicitly on μ_{r}^2. The dependence of the cross sections coefficients on μ_{r}^2 can be written explicitly as

$$\hat{\sigma}^{(\mathrm{LO})}(\mu_{\mathrm{r}}^2) = \hat{\sigma}^{(00)},$$

$$\hat{\sigma}^{(\mathrm{NLO})}(\mu_{\mathrm{r}}^2) = \hat{\sigma}^{(10)} + \log\left(\frac{\mu_{\mathrm{r}}^2}{\mu_0^2}\right)\hat{\sigma}^{(11)},$$

$$\hat{\sigma}^{(\mathrm{NNLO})}(\mu_{\mathrm{r}}^2) = \hat{\sigma}^{(20)} + \log\left(\frac{\mu_{\mathrm{r}}^2}{\mu_0^2}\right)\hat{\sigma}^{(21)} + \log^2\left(\frac{\mu_{\mathrm{r}}^2}{\mu_0^2}\right)\hat{\sigma}^{(22)},$$

where the $\hat{\sigma}^{(ij)}$ do not depend on μ_{r}^2. A similar expansion can be made with respect to μ_{f}^2, except that this case involves the splitting functions.[160] This expansion shows that there is only one scale-dependent term in the NLO expression. Therefore, the predictions will necessarily have a relatively large scale dependence. At NNLO further scale-dependent terms arise, which can partially cancel this behaviour, leading to a reduced scale dependence of the NNLO predictions.

Treatment of infrared and collinear divergences Another way to express the terms in equation (5.1) is to separate them into real and virtual contributions

$$\hat{\sigma}^{(\mathrm{LO})} = \int_n \mathrm{d}\hat{\sigma}^{(\mathrm{Born})},$$

$$\hat{\sigma}^{(\mathrm{NLO})} = \int_n \mathrm{d}\hat{\sigma}^{(\mathrm{V})} + \int_{n+1} \mathrm{d}\hat{\sigma}^{(\mathrm{R})},$$

$$\hat{\sigma}^{(\mathrm{NNLO})} = \int_n \mathrm{d}\hat{\sigma}^{(\mathrm{VV})} + \int_{n+1} \mathrm{d}\hat{\sigma}^{(\mathrm{RV})} + \int_{n+2} \mathrm{d}\hat{\sigma}^{(\mathrm{RR})},$$

where $\int_n$ represents an integral over the n-particle phase space, and the superscripts 'V' and 'R' indicate virtual and real corrections to the tree level matrix element. Starting at NLO, all terms are individually infrared or collinearly divergent while their sum remains finite at every order. Since these terms are usually too complicated to be integrated analytically, these divergences have to be regularised in such a way that they can be cancelled numerically.

One method to accomplish this is to use subtraction terms. As an example, consider the following integral

$$I = \int_0^1 \frac{\mathrm{d}x}{x}x^\varepsilon F(x) - \frac{1}{\varepsilon}F(0),$$

where $F(x)$ is a known function over the $(n+1)$-particle phase space.[161,162] The variable x represents the energy of an emitted particle or the angle between two particles. For $x \to 0$, the integral diverges, which is regularised by the x^ε term. The second term represents the virtual correction to the n-particle process, which is also divergent for $\varepsilon \to 0$. To cancel these divergences, the integral is split into the regions $[0, \delta]$ and $[\delta, 1]$. The first integral is approximated by a constant function, leading to

$$I = \int_\delta^1 \frac{\mathrm{d}x}{x}F(x) + F(0)\log(\delta).$$

If δ is sufficiently small, the δ-dependence of the two terms cancels. This alone can be enough to evaluate the expression. However, computing this integral for small δ can lead to numerical problems. To mitigate those, the term $\int_\delta^1 \frac{\mathrm{d}x}{x} E(x)$ is added and subtracted from the expression

$$I = \int_\delta^1 \frac{\mathrm{d}x}{x} (F(x) - E(x)) + \int_\delta^1 \frac{\mathrm{d}x}{x} E(x) + F(0)\log(\delta).$$

The function $E(x)$ must approach $F(x)$ close to the pole, so the first integral remains finite. Furthermore, $E(x)$ needs to be simple enough to be integrated analytically. The lower limit of the second integral should produce $-F(0)\log(\delta)$ to cancel the third term in the expression. If that is the case, the limit $\delta \to 0$ can be taken, and I can be computed numerically. The theory predictions used in this thesis use the 'antenna subtraction' method to construct the terms corresponding to $E(x)$.[159,163]

5.2 Parton showers and hadronisation

Fixed-order calculations are well-defined and highly accurate. They are, however, insufficient to compare to experimental results. For this purpose, higher-order and non-perturbative final-state effects must also be modelled. This is done using phenomenological models for parton showering and hadronisation. These models must be tuned to describe experimental data and are thus much less well-defined than perturbative calculations.

5.2.1 Parton showers

Before and after the hard lepton-parton interaction, the incoming and outgoing partons may emit further partons. In the limit of soft or collinear radiation, the probability for these processes increases rapidly, leading to many such emissions.

Parton showers solve the DGLAP equation (2.33) iteratively, leading to a series of $1 \to 2$ branchings. Each daughter particle can undergo further branchings. This procedure is known as the leading logarithmic approximation to parton showers.[39,164] Repeated application of the branching probability leads to the appearance of an exponential factor, known as the Sudakov form factor, that can be used to determine when a subsequent branching should occur. In the final-state shower, partons are evolved from the hard scale of the matrix element down to the pre-defined hadronisation scale. In the initial state, showers are evolved backwards from the matrix element to a cut-off scale. Furthermore, the initial-state shower has to take the PDFs into account, to ensure that the shower evolution is consistent with the PDF evolution. Since initial- and final-state parton showers are treated separately, interference effects between them cannot be taken into account. Furthermore, the leading logarithmic approximation is designed to describe soft and collinear emissions. The simulation of radiation of hard partons at large angles might be much less accurate.[165,166]

An alternative to the leading logarithmic approximation is the colour-dipole model. In this model, pairs of partons are treated as dipoles of the colour field. Gluons are emitted directly from this field rather than from the partons. New dipoles will form between the initial partons and the newly created gluon. From these dipoles, further gluons may be emitted. Gluons may split into a quark-antiquark pair. Colour dipoles can, in good approximation, be considered independent from each other. In the colour-dipole model, initial- and final-state parton showering are not distinguished. Instead, parton radiation is assumed to also originate from the dipole between the final-state parton and the proton remnant.[167]
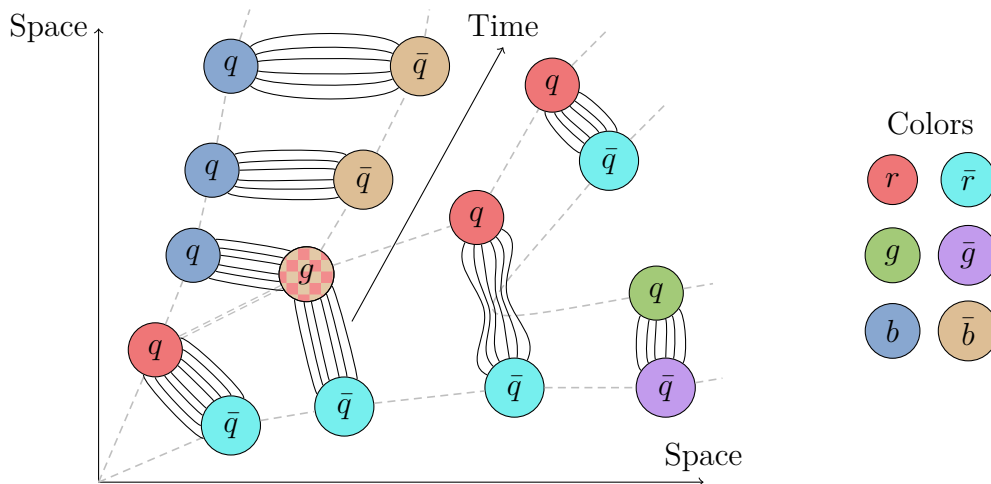
Figure 5.2: Schematic depiction of a quark-antiquark pair fragmenting into three separated hadrons, according to the Lund string model. With advancing time, the quark-antiquark pair moves apart, which leads to the breaking of the string connecting them and, thus, the creation of further quark-antiquark pairs. The lines between the partons represent the gluon field which is mediated by virtual gluons, while the gluon node indicates an almost real gluon.[169]

5.2.2 Hadronisation

The theory of perturbative QCD predicts the production of partons in the final state of the lepton-hadron interaction. Due to colour confinement, these partons are not observable as isolated particles. Instead, they will fragment and form colourless hadrons. As hadronisation describes the production of low-energy hadrons, it cannot be described using perturbative QCD calculations.

Simulations of scattering processes at HERA usually use the Lund string model to approximate fragmentation. In this model, partons are connected by colour strings. This description was later shown to be reasonable using lattice QCD simulations, which showed that the QCD field lines tend to form narrow tubes between coloured objects.[168] The energy density of these strings is found to be constant, i.e. the energy that is stored in a string grows linearly with its length. If the partons move far enough apart, the energy stored in the string becomes so large that it is energetically favourable to create a new quark-antiquark pair from the vacuum. At this point, the string will break, and new strings are formed between the original and the newly created pair of partons. Schematically, this process is depicted in figure 5.2. The strings consist of virtual gluons. The emission of almost real gluons can be taken into account in this picture as localised excitations of the string that act as additional nodes. These gluons are modelled explicitly using their four-momentum and undergo further splittings.[169] The energy to create the additional quark-antiquark pairs is taken from the kinetic energy of the partons to which the fragmenting string is connected. The process repeats until the remaining energy of the partons is too small to create further pairs. The combination of the remaining partons into colourless hadrons is described by fragmentation functions. These functions are, to some extent, the analogue to PDFs, as they describe the non-perturbative transition from partons to hadrons.[170,171]

The main alternative to the Lund string model is the cluster fragmentation model. In this model, all gluons are first split into quark-antiquark pairs. Quarks are then grouped into

clusters based on their colour, and fragmentation proceeds for each cluster similarly to the Lund string model. In the Lund string model, fragmentation happens over large distances and involves many particles simultaneously. In the cluster model, fragmentation is more localised and involves fewer particles at a time.[172] When tuned to experimental data, both models can describe fragmentation well.

The term 'hadronisation' refers to fragmentation and the subsequent decay of unstable excited particles. There is no clear theoretical distinction between parton showering, fragmentation and decays. The creation of quark-antiquark pairs during fragmentation can, to some extent, be considered a form of parton showering, or vice versa. Similarly, the decay of certain unstable particles can also be treated as part of the fragmentation. In practical calculations, these regimes are separated by cut-off parameters that are tuned to describe experimental data.

5.3 Monte Carlo simulations

MC simulations are an essential tool in high energy physics and are widely used to help understand the data and to derive corrections for it. These simulations consist of two major steps: event generation and detector simulation. During event generation, the interactions of the colliding particles are simulated. A fixed-order matrix element is used to describe the hard boson-parton interaction. In addition, MC simulations include models of initial- and final-state radiation, fragmentation from partons to hadrons, and decays of unstable hadrons until a set of long-lived and spatially separated particles are produced. At HERA, matrix elements are often used at leading order, and parton showering is described in the leading logarithmic approximation.

These particles are then used in the detector simulation that describes the response of a particular detector to a given set of particles. This simulation describes the interactions of the final-state particles with all active and inactive detector components, such as producing hits in the tracker or initiating showers and depositing energy in the calorimeter.

After the event is processed in the detector simulation, it is subjected to the same reconstruction and analysis algorithms as the data samples. At this point, the MC sample closely resembles the data, and the two can be directly compared. One aspect that makes MC simulations so valuable is that, apart from the detector-level information, the complete information from the event generator is also available. Comparing these so-called 'generator-level' or 'truth-level' particles to their 'detector-level' counterparts helps to understand the detector response and correct for it in the data.

5.3.1 Event generators

This analysis uses a series of MC chains. Each chain comprises a set of program packages that perform different parts of the event generation. In the following, the capabilities of each package will be briefly described. The next section explains how these packages are combined to form the complete MC chains.

Lepto event generator and parton showering   This analysis uses the Lepto package for two purposes. First, it is capable of generating boson-parton matrix elements for lepton-hadron collisions. It can simulate the $\mathcal{O}(\alpha_s^0)$ QPM-like process as well as the $\mathcal{O}(\alpha_s^1)$ QCD Compton and boson-gluon fusion processes.[165] The latter two processes constitute the leading-order contribution to inclusive jet production in the Breit frame. Like all event generators, Lepto

uses a PDF to describe the parton content of the hadron. While no higher-order QED effects are taken into account in the matrix elements, loops of virtual particles are accounted for via the scale dependence of the QED coupling constant $\alpha(\mu^2)$.

Secondly, Lepto can approximate higher orders through the emission of additional partons from the initial- or final-state parton of the interaction. For this purpose, the leading logarithmic approximation is used.

Ariadne parton showering   The Ariadne package implements the colour-dipole model. This model can be used as an alternative description of parton showering.[167]

Herwig event generator, parton showering and hadronisation   The Herwig event generator is capable of generating matrix elements for various processes, such as DIS, PHP and Higgs production in lepton-hadron interactions. It also simulates parton showering using an angular-ordered coherent branching algorithm as well as hadronisation using the cluster model.[173]

Rapgap event generator and parton showering   Rapgap is another event generator capable of generating various processes at lepton-hadron colliders. It can also simulate parton showering via a leading logarithmic approach.[166]

Heracles event generator   Like Lepto, the Heracles event generator calculates leading-order boson-parton matrix elements for lepton-hadron collisions. Additionally, leading-order electroweak corrections can be taken into account. This includes single photon bremsstrahlung from the initial- or final-state lepton, self-energy corrections of the exchanged boson, vertex corrections of the lepton-boson vertex as well as the exchange of a Z boson and photon-Z interference.[174] Heracles is also able to include radiative corrections on the quark side. These settings were, however, turned off for all samples used in this thesis.

Jetset hadronisation   The event generators discussed so far aim to describe the final state in terms of partons. The formation of hadrons from these partons is described by Jetset, formerly part of the Pythia package.[170] This package uses the Lund string model. Furthermore, Jetset simulates subsequent decays of unstable particles.

5.3.2 Monte Carlo chains for signal and background samples

To perform a full MC simulation, the packages described above are combined to form so-called 'MC chains'. Each chain consists of one or more packages to simulate the boson-parton matrix element, parton emission and hadronisation.

Two high-Q^2 NC DIS signal MC chains are defined, which will be referred to as 'Ariadne' and 'Lepto'. These samples will be used to unfold the measured events to truth level. Two more samples are generated, corresponding to the most relevant background contributions: the 'low-Q^2 DIS' and 'PHP' samples.

After event generation, all samples are processed by the detector simulation based on Geant 3.21, which simulates interactions with active and passive detector components.[175] Simulated events are subjected to the same trigger configurations as the real data on a run-by-run basis.

The samples are taken from the ZEUS common ntuple database.[176] Table 5.1 gives an overview of the samples.

Sample	(internal name)	Number of events	Luminosity [pb^{-1}]
Ariadne	(`ariadne_high_Q2_NC_`$\mathbb{P}$)	32.88×10^6	28.95×10^6
Lepto	(`mps_high_Q2_NC_`$\mathbb{P}$)	31.68×10^6	24.99×10^6
PHP	(`PHP_HER_{dir,res}_4D_`$\mathbb{P}$)	13.86×10^6	1.414×10^3
Low-Q^2 DIS	(`Ariadne_Low_Q2_NC_DIS_`$\mathbb{P}$)	238.6×10^6	371.3

Table 5.1: An overview of MC samples at detector-level. The large luminosity of the Ariadne, Lepto and PHP samples is because they contain subsamples in low-statistics regions. The internal name indicates the name in the ZEUS common ntuple database.[177] The $\mathbb{P}$ stands for any of the four run periods `0304p`, `05e`, `06e` and `0607p`. For the Lepto sample, there was a bug in one of the initial samples. The corrected sample `mps_high_Q2_NC_fixlst_0607p` is used instead.

Ariadne   The Ariadne chain is the central MC chain used in this analysis. It consists of HERACLES 4.6.6 to simulate the boson-parton matrix element, including all corrections mentioned above and using the CTEQ 5D PDF set.[104] The emission of further partons is simulated using ARIADNE 4.12 and hadronisation using JETSET 7.410.

Events were generated using this chain for $Q^2 > 100\,\mathrm{GeV}^2$. Since the cross section falls steeply with Q^2, this sample contains only a few events in the very high-Q^2 region. To ensure high statistical precision, even in this low-statistics region, further subsamples were generated with higher Q^2 cuts. The subsample with the highest cut was generated in the region $Q^2 > 50000\,\mathrm{GeV}^2$. To combine these subsamples into one consistent sample, events must be assigned a generator-level weight. This weight is computed as

$$w(Q^2) = \frac{L_{\mathrm{data}}}{\sum_i L_{\mathrm{MC},i}}, \tag{5.2}$$

where L is the integrated luminosity, and the sum runs over all MC samples whose phase space includes the event being weighted. Events at very high Q^2, no matter from which subsample they originate, will be assigned a small weight to account for the increased number of samples contributing to that region. This way, the overall normalisation is preserved, and the Q^2 distribution does not exhibit discontinuities. The generator-level weight is discussed in section 5.3.4.

Lepto   The Lepto chain is used to estimate the systematic effect of the choice of MC generator. It is identical to the Ariadne chain, except for the parton showering, which is instead simulated using LEPTO 6.5.1. Similarly to the Ariadne sample, the Lepto sample consists of multiple subsamples that are weighted and combined to form a single consistent sample.

Low-Q^2 DIS   The Ariadne and Lepto signal MC samples were generated in the region $Q^2 > 100\,\mathrm{GeV}^2$, which contains the measurement region of this analysis $Q^2 > 150\,\mathrm{GeV}^2$. An additional sample was generated to estimate the contribution from events generated outside the phase space region covered by the signal MC sample but have migrated into the measurement region. This sample was generated using the Ariadne chain described above, in the phase space region $4\,\mathrm{GeV}^2 < Q^2 < 100\,\mathrm{GeV}^2$. The cuts ensure this sample does not overlap with the signal MC samples. This sample will be referred to as the low-Q^2 DIS sample.

Photoproduction PHP events were generated for $Q^2 < 4\,\text{GeV}^2$ using the HERWIG 5.9 generator. This choice ensures that no events are missed or double-counted when combining this sample with the low-Q^2 DIS sample. Since HERWIG already simulates hadronisation using the cluster model, the use of JETSET is not necessary. The PHP sample consists of multiple subsamples with different cuts on $p_\perp$ and $E_\perp$ of the hadronic system.

5.3.3 Chains for hadron-level corrections

For some of the corrections that will be applied to the data, additional MC samples are required. These corrections are applied after unfolding, so no detector simulation is required for these samples. The need for these corrections will be motivated later.

These samples are newly generated for this analysis. Using configuration files from the ZEUS common ntuple database ensures that the new samples use consistent settings compared to the existing samples. It is verified that they produce consistent results.

Real QED radiation correction A sample is generated using matrix elements from HERACLES, including all mentioned corrections. Parton showering is performed using RAPGAP and hadronisation using JETSET. A second sample is generated using this chain, in which higher-order QED effects are turned off within HERACLES. These two samples are necessary to derive the QED radiation correction, which will be described in section 9.5.3.

As far as possible, the samples are generated with similar settings to the existing MC samples from the ZEUS database. In particular, the sample including QED radiation uses the HERACLES options `ISNC31 = ISNC32 = ISNC33 = 1`, `INC31 = INC32 = INC33 = 16` as well as `LPARIN(4) = 1` and `LPARIN(7) = 2`. These options enable collinear initial- and final-state radiation, non-collinear radiation and vertex corrections on the electron side.[174,178] The sample at Born level is generated by setting all of these options to 0, thus removing these contributions. Self-energy corrections are taken into account in both samples via the running of the electromagnetic coupling, which is enabled using the `LPARIN(3) = 3` option. Corrections to the quark side are turned off in both samples.

Polarisation correction The chain from the QED radiation correction, including higher-order QED effects, is used to generate four more samples with polarised lepton beams. Each sample corresponds to the average polarisation of one of the four run periods in the data. Comparing these four samples to the unpolarised one is required for the polarisation correction described in section 8.1.3. Polarised samples are generated using HERACLES by adjusting the `POLARI` input value from its initial value of 0 to the value corresponding to each run period, as given in table 6.1.

Virtual electroweak correction A MC chain is constructed using matrix elements and parton showering from LEPTO and hadronisation using JETSET. Two samples are generated using this chain. In the first sample, only photon exchange between the lepton and the hadron is considered. The second sample also considers Z boson exchange and photon-Z interference. These two samples are necessary for the weak-boson correction described in section 9.5.2. In LEPTO, the simulated process is controlled via the `LST(23)` option.[165] A value of 1 corresponds to considering only photon exchange. Setting this option to 4 enables Z exchange and the interference term.

5.3.4 Weights of Monte Carlo samples

Most signal and background MC chains comprise multiple, overlapping subsamples. To consistently combine these samples, each event must be assigned a generator-level weight, as shown in equation (5.2). This weight must be considered in all distributions where such a MC sample is used. When applied to the MC, the term 'number of events' actually refers to the sum of the event weights. For this reason, the 'number of events' can take non-integer values. This also makes it possible to have high statistical precision in low-statistics regions, as bins with a low 'number of events' can still have many entries with a small weight each.

Another use of the generator-level weight is to allow the reweighting of the MC samples. The MC event generators have known deficiencies, e.g. they overestimate the number of events in some regions of the phase space and underestimate them in others. A simple and practical way to correct these deficiencies is to adjust the generator-level weights of the events in the MC samples accordingly. Various such corrections will be applied in chapter 8.

In addition to the generator-level weight w_{gen}, each MC event also possesses a detector-level weight w_{rec}. When determining event counts at generator level, only the generator-level weight needs to be used. When determining detector-level counts, both weights must be included, i.e. events are assigned a weight of $w_{\text{gen}} \cdot w_{\text{rec}}$. This additional detector-level weight allows applying reweightings to the detector simulation. E.g. if there is a misrepresentation of events in some phase space region due to effects in the detector simulation, the detector-level weight can be adjusted to correct for it. Since this is a detector effect, it should not affect the generator-level distributions and therefore requires a separate weight.

Similarly, events in the data sample can be assigned a weight that can be adjusted to correct for known effects. Depending on the investigated effect, it is sometimes easier to reweight the data than the MC samples. There is no distinction between the generator-level and detector-level weights in the data since generator-level quantities are unknown. When doing so, the 'number of events' may take non-integer values, even in the data. In this thesis, this will only be done for one correction, with weights close to one. Therefore, the 'number of events' in the data will still be approximately integer-valued, see section 8.1.3.

CHAPTER 6

Event reconstruction

The detector measures a set of raw signals, the most important of which correspond to energy deposits in the CAL and hits in the CTD and MVD. To perform a meaningful physics analysis, particles must be reconstructed from these signals. In NC processes at HERA, an accurate reconstruction of the scattered beam electron is essential since the computation of the kinematic quantities relies heavily on this information. For the present analysis, the reconstruction of the hadronic final state is also crucial since those particles will be used to reconstruct jets.

6.1 Data samples

This analysis is based on data recorded at the ZEUS detector during the HERA II period. In this period, protons were accelerated to 920 GeV and electrons to 27.5 GeV, leading to a centre-of-mass energy of 318 GeV. The data are separated into four run periods with stable data-taking conditions, as shown in table 6.1. Each run period consists of many individual runs, typically lasting a few hours each. The luminosity of each run is determined from the PCAL if available and from the SPEC if not, see section 3.2.3.

During the HERA II period, longitudinally polarised electron beams were delivered to the experiments as described in section 3.1. The average polarisation of each run period is given in the table. The polarisation of each run is determined either from the LPOL or the TPOL, depending on which one was active longer during the run, see section 3.1. Combining all run periods, the data sample is, on average, almost unpolarised. The `03p` period and parts of the `04p` period are not used in the analysis since it is found that the trigger rates are not as well described by the MC as the other periods and the luminosity of these periods is small.

Run period	Luminosity [pb^{-1}]	Run numbers	Average polarisation
04p	12.9	50500–51245	0.32
05e	135.1	52258–57123	-0.06
06e	55.2	58207–59947	0.08
0607p	143.8	60005–62639	0.03
Total	347.0	50500–62639	0.01

Table 6.1: Overview of run periods. The last letter of the run period name indicates whether the lepton beam consisted of electrons or positrons.

6.2 Reconstruction of the scattered electron

In an NC DIS analysis, the scattered electron plays a central role, e.g. in the determination of kinematic quantities, rejection of background and for the construction of the Breit frame that is used for the clustering of jets. At ZEUS, two algorithms are commonly used to identify electron candidates.

The so-called EM algorithm combines information from the CAL, HES, CTD and SRTD.[179] It starts by grouping energy deposits from CAL cells into electromagnetic clusters, which are taken as electron candidates. Candidates within the CTD acceptance region must have a track associated with them. Each of these candidates is then rated according to several criteria, such as the fraction of energy found in the electromagnetic rather than the hadronic parts of the CAL or the isolation in the η-ϕ-plane. A total score is then derived by combining all individual ratings. Candidates are accepted as electrons if they exceed a specific score derived from MC studies. For high-energy electrons within the CTD acceptance region, which is most electrons used in this analysis, the measured energy is taken from the CAL and the position from the CTD track, extrapolated to the CAL surface, as these components have the best resolution of these quantities. If available, the position measurement is enhanced using information from the HES or SRTD. EM candidates are accepted as electrons for this analysis if their combined score exceeds 0.001.

Sinistra, the other commonly used algorithm for finding electrons at ZEUS, employs a neural network.[180] This algorithm analyses the energy distribution in the CAL to distinguish between electromagnetic and hadronic objects. A score between 0 and 1 is then assigned to each object based on its probability of being an electron. Afterwards, the measured energy is corrected for energy loss due to inactive material and for the small non-uniformity of the response of the CAL. If available, information from the HES and SRTD are also used. Sinistra candidates are considered to be of acceptable quality if their score exceeds 0.9. The Sinistra algorithm is optimised only for the central and rear regions of the detector, i.e. for electrons found in the RCAL and BCAL. Therefore, it is less suited for reconstructing very high-Q^2 events, in which the electron is found in the FCAL.

In this thesis, scattered electrons are identified using the EM algorithm, as it is better suited for the very high-Q^2 region.[181] To estimate the systematic effect of this choice, a separate study is performed using the Sinistra algorithm, as described in section 10.4.

6.3 Kinematic variables

As explained in section 2.2, an inclusive DIS event is usually characterised by the kinematic variables Q^2, x_{Bj}, y and s. As these quantities play a central role in defining cross sections, it is essential to determine them precisely. Neglecting masses of the initial-state particles and higher-order QED radiation, these quantities are related as $Q^2 = x_{\mathrm{Bj}} y s$. Since s only depends on the beam energies, it is constant. Therefore, only two of these kinematic quantities are independent and must be determined for each event.

The general graph for jet production in NC DIS $e + P \rightarrow e' + p' + X$ is shown in figure 6.1. Here, P is the initial-state proton, e is the initial-state electron, e' is the final-state electron, p' is the hadronic final state from the hard scattering, and X is the proton remnant. The struck parton takes a fraction ξ of the momentum of the incoming proton. All external particles are assumed as massless, which is a valid approximation for the present analysis. The following derivation is performed in the QPM, where the hadronic final state p' consists of a single
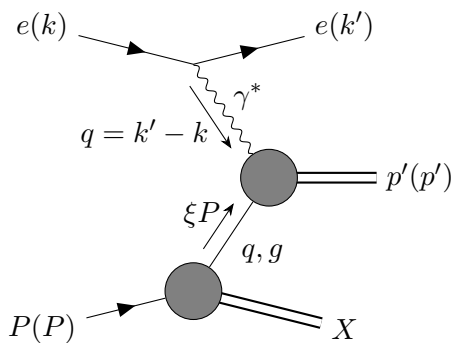
Figure 6.1: Graph of NC DIS at leading-order in the electroweak interaction on the electron side.

particle and is thus also massless. This is not a valid approximation when considering QCD effects. Implications of a mass of the hadronic system are discussed in the next section. In the QPM, the initial- and final-state momenta can be parameterised as

$$
k = \begin{pmatrix} E_e \\ 0 \\ 0 \\ -E_e \end{pmatrix}, \quad P = \begin{pmatrix} E_P \\ 0 \\ 0 \\ E_P \end{pmatrix}, \quad k' = \begin{pmatrix} E'_e \\ E'_e \sin(\theta)\cos(\phi) \\ E'_e \sin(\theta)\sin(\phi) \\ E'_e \cos(\theta) \end{pmatrix}, \quad p' = \begin{pmatrix} E'_p \\ E'_p \sin(\gamma)\cos(\psi) \\ E'_p \sin(\gamma)\sin(\psi) \\ E'_p \cos(\gamma) \end{pmatrix}. \quad (6.1)
$$

For now, the process is considered only at leading order in QED on the electron side. The consequences of radiative corrections are discussed in section 6.5.

The six final-state quantities E'_e, E'_p, θ, γ, ϕ and ψ are measurable in the detector. The proton remnant X is not practically measurable, as a significant part of it escapes into the beam pipe.

Four-momentum conservation of the electron-parton system demands

$$
k + \xi P = k' + p'. \quad (6.2)
$$

One of these four constraints relates the angles ϕ and ψ, which correspond to rotations around the beam axis and do not affect the kinematic quantities due to the azimuthal symmetry of the system. After eliminating ξ, two relations remain between the relevant observable quantities E'_e, E'_p, θ and γ

$$
2E_e = E'_e(1 - \cos(\theta)) + E'_p(1 - \cos(\gamma)), \quad (6.3a)
$$

$$
0 = E'_e \sin(\theta) - E'_p \sin(\gamma). \quad (6.3b)
$$

Thus, only two of these four observable quantities are independent and sufficient to describe the final state and compute the kinematic quantities. This corresponds to the earlier statement that only two of the kinematic variables are independent.

The simplest method to reconstruct the kinematic quantities is known as the electron method, as it uses only the properties E_e' and θ of the final-state electron. It can be obtained directly, by inserting expressions (6.1) into the definitions of the kinematic quantities (2.1)

$$Q_{\mathrm{EL}}^2 = 2 E_e E_e' \left(1 + \cos(\theta) \right), \tag{6.4a}$$

$$y_{\mathrm{EL}} = 1 - \frac{E_e'}{2 E_e} \left(1 - \cos(\theta) \right), \tag{6.4b}$$

$$x_{\mathrm{Bj,EL}} = \frac{E_e}{E_P} \frac{E_e'(1 + \cos(\theta))}{2 E_e - E_e'(1 - \cos(\theta))}. \tag{6.4c}$$

Using equations (6.3), it is possible to derive alternative methods using any other two of the four relevant observables.[182,183] Since the detector resolution depends on the measured quantity and the detector region, each method has a different resolution depending on the investigated phase-space region. The optimal choice of method depends on the analysis and on the desired treatment of QED corrections.

A particularly useful method is the double-angle method, which exclusively relies on the angles θ and γ. This method is, therefore, approximately independent of any bias in the energy measurement of the calorimeter for the scattered electron or the hadronic system from the hard scattering

$$Q_{\mathrm{DA}}^2 = \frac{4 E_e^2 \sin(\gamma) \left(1 + \cos(\theta) \right)}{\sin(\theta) + \sin(\gamma) - \sin(\theta + \gamma)}, \tag{6.5a}$$

$$y_{\mathrm{DA}} = \frac{\sin(\theta) \left(1 - \cos(\gamma) \right)}{\sin(\theta) + \sin(\gamma) - \sin(\theta + \gamma)}, \tag{6.5b}$$

$$x_{\mathrm{Bj,DA}} = \frac{E_e}{E_P} \frac{\sin(\gamma)}{\sin(\theta)} \frac{1 + \cos(\theta)}{1 - \cos(\gamma)}. \tag{6.5c}$$

A useful related quantity is the double-angle energy. This is the energy of the scattered electron E_e', as reconstructed from the angles θ and γ

$$E_{\mathrm{DA}} \coloneqq E_e'(\theta, \gamma) = \frac{2 E_e \sin(\gamma)}{\sin(\theta) + \sin(\gamma) - \sin(\theta + \gamma)}. \tag{6.5d}$$

The double-angle method requires knowledge of the hadronic scattering angle γ. This angle is determined from the momentum of the hadronic final state from the hard scattering p'

$$\cos(\gamma) = \frac{p_{\mathrm{x}}'^2 + p_{\mathrm{y}}'^2 - (E_p' - p_{\mathrm{z}}')^2}{p_{\mathrm{x}}'^2 + p_{\mathrm{y}}'^2 + (E_p' - p_{\mathrm{z}}')^2}, \tag{6.5e}$$

where p' is the sum over the momenta of all particles that contribute to the hadronic final state from the hard scattering.

Other methods for kinematic reconstruction are also widely used. E.g. the Jacquet-Blondel method uses only observables from the hadronic final state E_p' and γ. It has been shown in previous analyses that the double-angle method has the best resolution in the phase space considered in this analysis.[80,182]

It is instructive to invert equations (6.4) or (6.5) to obtain a more intuitive understanding of the kinematic quantities. The final-state observables can be written as functions of the kinematic quantities as

$$\begin{aligned} E_e' &= E_P x_{\mathrm{Bj}} y + E_e(1 - y), & \cos(\theta) &= \frac{E_P x_{\mathrm{Bj}} y - E_e(1 - y)}{E_P x_{\mathrm{Bj}} y + E_e(1 - y)}, \\ E_p' &= E_P x_{\mathrm{Bj}}(1 - y) + E_e y, & \cos(\gamma) &= \frac{E_P x_{\mathrm{Bj}}(1 - y) - E_e y}{E_P x_{\mathrm{Bj}}(1 - y) + E_e y}. \end{aligned} \tag{6.6}$$
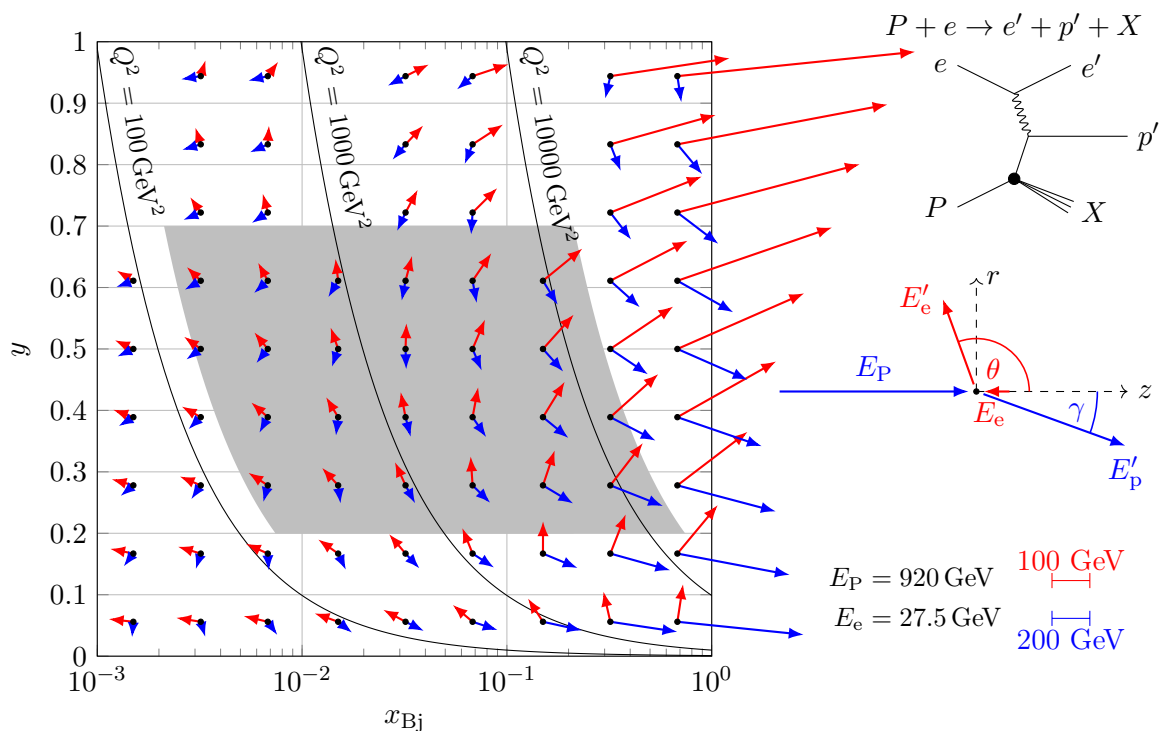
Figure 6.2: The final-state quantities E'_e, E'_p, θ and γ are shown as a function of the kinematic quantities Q^2, x_{Bj} and y for the beam parameters used HERA. The relations are valid at leading order in the QPM. The grey area indicates the phase space that is investigated in this analysis.

These relations are illustrated in figure 6.2. For large x_{Bj}, both the electron and the hadronic final state are mostly scattered in a forward direction and y indicates what fraction of the total energy available is taken by the electron. This relation flips at small x_{Bj}, and the energy of the hadronic final state becomes proportional to y, while both objects are scattered in a more backward direction. At $x_{\mathrm{Bj}} = E_e/E_P \approx 0.03$, the electron and the hadronic final state have the same energy and are scattered back-to-back, with y determining the polar angle relative to the beam axis.

6.4 Application to a massive hadronic system

In the previous section, the electron method (6.4) and the double-angle method (6.5) have been derived under the assumption that the hadronic final state from the hard scattering p' is massless. When considering inclusive jet production in the Breit frame, this system comprises at least two non-collinear particles. While all component particles may be assumed massless, their sum can be rather massive, easily exceeding the energy of the incoming electron. In the following section, it will be shown that the formulae derived above under the massless assumption are also applicable to the massive case.

For the electron method, this proof is simple. Its derivation merely consisted of inserting the parameterisations of k, P and k' from equation (6.1) into the definition of the kinematic

quantities (2.1). Therefore, it is independent of the hadronic final state p', and the same derivation applies to the massive case.

A formal derivation of the double-angle method in the massive case is rather complex. Instead, the following argumentation will start from the massless expressions (6.5) and show that they are also valid in the massive case. In the massive case, the parameterisation of p' has to be adjusted in equation (6.1)

$$
p' = \begin{pmatrix} E'_p \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \sqrt{p'^2_p + m^2} \\ -p'_p \sin(\gamma^{(\mathrm{true})}) \\ 0 \\ p'_p \cos(\gamma^{(\mathrm{true})}) \end{pmatrix}, \tag{6.7}
$$

where $m^2 = p'^2$ is the mass of the hadronic system. Four-momentum conservation now leads to

$$
p' = k + \xi P - k' = \begin{pmatrix} E_e + \xi E_P - E'_e \\ -E'_e \sin(\theta) \\ 0 \\ -E_e + \xi E_P - E'_e \cos(\theta) \end{pmatrix}. \tag{6.8}
$$

Each event is characterised by the six quantities E'_e, θ, ξ, p'_p, γ and m. Combining the above equations yields three constraints on these quantities. Thus, only three of these quantities are independent and sufficient to describe the event.

Consider an event described in terms of E'_e, θ and ξ. According to the electron method, the kinematic quantities Q^2, x_{Bj} and y are completely determined by E'_e and θ, i.e. they are independent of ξ. When picking $\xi = x_{\mathrm{Bj}}$, the event corresponds to the QPM and $m^2 = 0$. In this case, the double-angle formulae will yield the correct kinematic quantities according to the derivation in the previous section. What remains to be shown is that this is also the case of other values of ξ.

In the double-angle method, the kinematic quantities are described by θ and γ. Since θ is considered fixed, any ξ dependence of the kinematic quantities must be via γ. The quantity γ is defined via equation (6.5e). Note that γ only depends on p'_x, p'_y and $E'_p - p'_z$. According to equation (6.8), the first two of these quantities are independent of ξ and in the third one, the ξ-dependence cancels. Thus, γ is independent of ξ, and so are the kinematic quantities computed using the double-angle method.

To summarise, an event is completely described by E'_e, θ and ξ. The true kinematic quantities Q^2, x_{Bj} and y do not depend on ξ. The kinematic quantities computed using the double-angle method do not depend on ξ. For $\xi = x_{\mathrm{Bj}}$, the kinematic quantities according to the double-angle method are identical to the true ones. Therefore, the kinematic quantities computed using the double-angle method must be identical to the true ones for any value of ξ. The same argument can be applied to show that $E_{\mathrm{DA}} \equiv E'_e$.

This argumentation claims that γ is independent of ξ. Physically, this seems like a contradiction since the properties of the hadronic system clearly must depend on the momentum of the incoming quark. This apparent contradiction is because, in the massive case, the quantity γ loses its physical interpretation as the angle of the hadronic system. By inserting equation (6.7) into (6.5e), it is easy to show that $\gamma \neq \gamma^{(\mathrm{true})}$.

Therefore, the correctness of the double-angle method in the massive case depends crucially on the exact form of equation (6.5e). If $\cos(\gamma)$ where defined, e.g. as $p'_z / \sqrt{p'^2_x + p'^2_y + p'^2_z}$, which
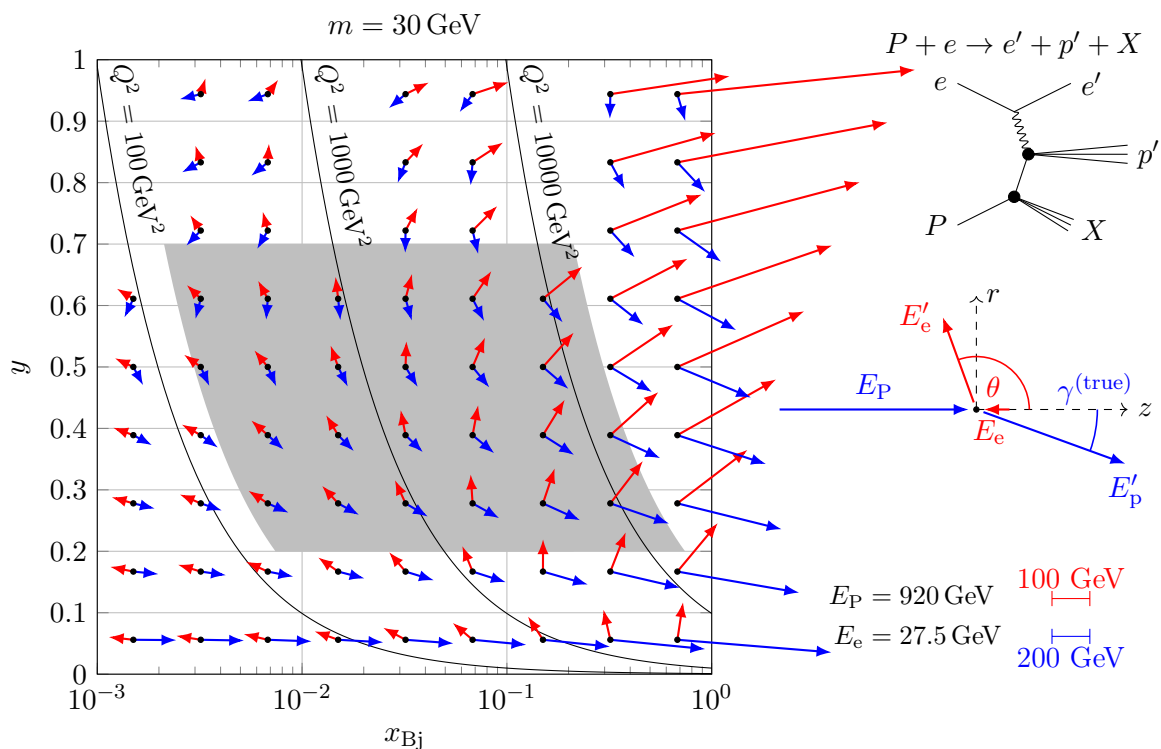
Figure 6.3: The final-state quantities E'_e, E'_p, θ and $\gamma^{(\mathrm{true})}$ in the case of a massive hadronic system with $m = 30\,\mathrm{GeV}$. For details, see the caption of figure 6.2.

yields $\cos(\gamma^{(\mathrm{true})})$, then the kinematic quantities would no longer correspond to the true ones. As can be seen, the fact that the double-angle method works in the massive case is non-trivial and is often glossed over when discussing the double-angle method in the context of multijet measurements.[72,80,184–186]

The inverse relations (6.6) that express the observables in terms of the kinematic quantities need to be modified for a massive hadronic system. The relations for E'_e and $\cos(\theta)$ remain unchanged, as they are derived from the electron method. Inserting those expressions into the momentum-conservation relation (6.8) leads to generalisations of the expressions for the hadronic system. Expressed in terms of x_{Bj}, y and m, they read

$$\xi = x_{\mathrm{Bj}} + \frac{m^2}{sy}, \tag{6.9a}$$

$$E'_p = E_P x_{\mathrm{Bj}}(1-y) + E_e y + \frac{m^2}{4E_e y}, \tag{6.9b}$$

$$\cos(\gamma^{(\mathrm{true})}) = \frac{E_P x_{\mathrm{Bj}}(1-y) - E_e y + \frac{m^2}{4E_e y}}{\sqrt{\left(E_P x_{\mathrm{Bj}}(1-y) - E_e y + \frac{m^2}{4E_e y}\right)^2 - m^2}}. \tag{6.9c}$$

Figure 6.3 shows these relations for a typical value $m = 30\,\mathrm{GeV}$. Compared to the massless case, the hadronic system moves in a more forward direction. The effect is most significant in the low-x_{Bj}, low-y region.
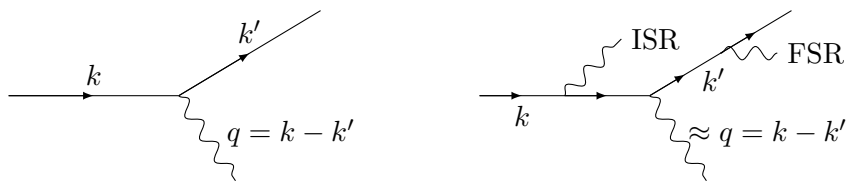
Figure 6.4: Feynman graph of the electron-boson-interaction at QED Born level (left) and including radiative corrections of the electron (right). For the right graph, the momenta are labelled according to the definitions used in this analysis.

6.5 Treatment of radiative effects

When considering single- or multijet production without real corrections, all presented methods of kinematic reconstruction are equivalent, up to detector resolution effects. This is no longer true when considering higher-order QED corrections to this process. Most importantly, photon radiation from the electron line can affect these methods, as it modifies the momentum-conservation relation (6.2). Such radiation is predominantly collinear and is referred to as initial-state radiation (ISR) or final-state radiation (FSR), depending on whether it originates from the incoming or outgoing electron. The different methods of kinematic reconstruction are derived using the leading-order version of the momentum-conservation relation. These methods are no longer equivalent when this relation does not hold. Therefore, the methods must be distinguished and used consistently also at MC generator level.

The Feynman graphs of the electron-boson vertex in the absence and presence of QED radiative effects are shown in figure 6.4. Collinear FSR and bremsstrahlung are not problematic, as they are reconstructed as part of the electron. Therefore, it does not affect the reconstruction of the event.

Collinear ISR photons require more careful treatment. They reduce the energy of the incoming electron and thereby affect the kinematic reconstruction. These photons escape into the beam pipe and are, therefore, usually not directly measurable in the detector, which makes it difficult to correct for the energy loss. This means that there is no simple way to define k and k', such that they are observable in the detector and that $q = k' - k$ is the momentum of the exchanged boson.

This issue also affects comparisons between measurements and theoretical predictions. The available predictions for jets in DIS correspond to QED Born level. In this definition, the process is considered at leading-order QED, i.e. the left panel of the figure. In the data and the signal MC, on the other hand, higher-order QED radiative effects are present. The approach chosen at HERA is to apply a correction to the measured cross sections to transform them to QED Born level. This procedure is described in section 9.5.3. This means there is some freedom in defining the intermediate cross sections, i.e. the ones derived by the unfolding but before the QED correction.

For this analysis, it is chosen to use the momentum of the electron beam before ISR as the initial-state electron k, the scattered electron including any potential FSR as final-state electron k' and to keep equations (2.1) as the definition of the kinematic quantities. This choice corresponds to using the electron method at generator level. It is shown in the right panel of figure 6.4.

This definition implies that in the presence of ISR, the quantity q no longer corresponds exactly to the momentum of the exchanged boson. This momentum is required for constructing

the Breit frame. Thus, this definition introduces a slight bias in the reconstruction of Q^2 and the Breit frame boost. In the data, this bias is kept to a minimum by removing events with large amounts of ISR from the event sample at detector level, see section 7.1. The QED-radiation correction that is applied to the unfolded cross sections accounts for this definition. After applying this correction, the cross sections correspond to QED Born level with the correct and consistent definitions of Q^2 and the Breit frame.

It is also possible to develop a method of kinematic reconstruction that uses three of the observables E'_e, E'_p, θ and γ and does not rely on knowledge of the initial-state-electron momentum. Such a method would not be sensitive to ISR and should allow for a more precise measurement of the momentum of the exchanged boson.[182,187] It would be interesting to explore such a method in more detail in future analyses.[a]

The interacting quark line can also radiate collinear photons. Since the coupling of fermions to the photon is proportional to their squared electric charge, radiative effects on the quark side are a subdominant effect compared to the electron side. Similarly to the electron side, FSR from the quark is not problematic, as it is reconstructed as a part of the jet. This is especially true due to the large radius of the constructed jets, see next section, and due to the almost equal response of the CAL to electromagnetic and hadronic objects. Collinear ISR from the quark also affects its momentum. Since the parton momentum is inherently unknown, none of the reconstruction methods explicitly rely on it. Therefore, they are independent of ISR from the quark. Radiative corrections to the quark side are not further considered in this thesis.

6.6 Reconstruction of detector-level jets

Particles of the hadronic final state are visible in the detector as energy deposits in the CAL. Electrically charged particles have a track associated with them in the CTD and MVD. The energy measured in the CAL has a better resolution for particles of high transverse momentum. Charged particles of lower transverse momentum are usually better measured by the CTD and MVD. A dedicated algorithm was developed at ZEUS to use the available information optimally. It combines the information from the CAL and the CTD to form so-called energy flow objects. Due to the non-uniformity of the tracker, energy flow objects require significantly more complex correction procedures. Therefore, in this analysis, jets are reconstructed from CAL energy deposits only. This is possible due to the high resolution and uniformity of the calorimeter of the ZEUS detector.

Jet candidates are reconstructed using the $k_\perp$-algorithm with radius parameter $R = 1$, as described in section 2.5.1. This radius corresponds to an opening angle of the jets of 115°. Choosing such a large jet size is convenient as it makes the reconstruction insensitive to soft large-angle radiation within the jets. The input to the algorithm consists of the massless four-momenta calculated from the position and measured energy of each CAL cell. Cells associated with the scattered electron are excluded, as well as cells directly adjacent to the forward beam pipe, as these are more likely to be biased by the proton remnant and by particles scattered off the beam elements. The $p_\perp$-weighted scheme is used to combine particles. This

---

[a]As a side note: such a method could potentially even be used to study DIS in proton-proton collisions, where one proton radiates a photon that then splits into an e^+e^--pair. One of these electrons could then engage in a DIS interaction with a proton from the other beam. Since the energy of this electron is inherently unknown, the study of this process is only possible using a method for kinematic reconstruction that does not rely on this information.

guarantees that the clustered jets are massless, which is necessary for consistency with previous measurements and theoretical predictions.[71,121,133,159]

In this analysis, the clustering is usually, but not always, performed in the Breit frame of reference that was described in section 2.5.2. In these cases, the four-momenta computed from the CAL cells are boosted to the Breit frame, where the $k_\perp$-algorithm is applied. The clustered jet candidates are then boosted back to the laboratory frame. Knowing the four-momentum of the exchanged boson is necessary to compute the boost matrix. This momentum is determined from the electron method, as discussed in the previous section.

CHAPTER 7

Event selection

In this analysis, inclusive jet production in NC DIS events is investigated. First, a set of NC DIS events is selected. Afterwards, the high-energy jets of each event are identified and collectively form the inclusive jet sample.

7.1 Signal and background characteristics

The Feynman graphs for the leading-order contributions to this analysis are given in figure 7.1. Processes contributing to the analysis contain at least two jets due to using the Breit frame, as described in section 2.5.2.

An NC DIS event is characterised by the scattered electron in the final state. The scattering angle of the electron is shown as a function of the kinematic variables in figures 6.2 and 6.3. The phase space for this high-Q^2 analysis is selected such that the electron is scattered at moderate angles, not too close to either beam pipe. Therefore, the scattered electron is expected to be measurable in the detector as an energy deposit in the CAL, usually with a CTD track pointing to it. At lower Q^2 values, the electron is found in the rear region of the detector. In higher Q^2 events, the electron is scattered in an increasingly forward direction. Since the cross section drops steeply as a function of Q^2, the majority of electrons are expected to be found in the RCAL and BCAL. Only at the highest Q^2 values, greater than about $10000\,\mathrm{GeV}^2$, can the electron be scattered towards the FCAL.

The transverse momentum of the scattered electron is balanced by that of the hadronic system. The proton remnant is not entirely measurable in the detector as it escapes into the
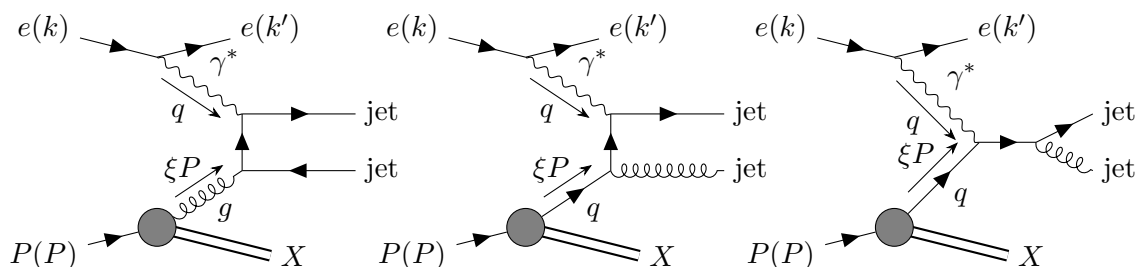
Figure 7.1: Feynman diagrams of boson-gluon fusion (right) and QCD Compton (middle and right) processes. In all cases, the measurable final state consists of the scattered electron and two jets.

Period 06*p*; Run 61747; Event 107771   Period 05*e*; Run 52262; Event 22318
Recorded 2007/01/01 03:45:57   Recorded 2004/12/06 09:00:50
$Q^2 = 184\,\mathrm{GeV}^2, x_{\mathrm{Bj}} = 0.740 \times 10^{-2}, y = 0.246$ $Q^2 = 13055\,\mathrm{GeV}^2, x_{\mathrm{Bj}} = 0.253, y = 0.508$
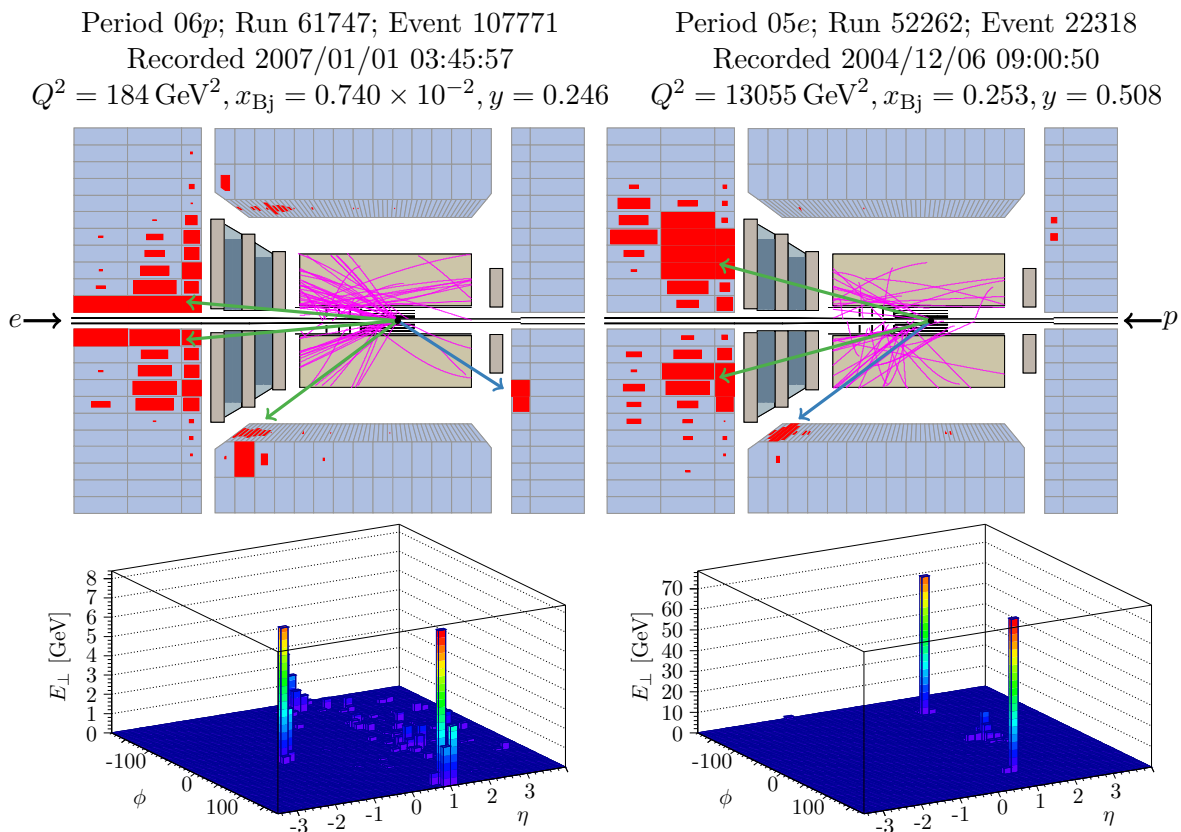
Figure 7.2: Event display of two signal events.[188] The left panels show a trijet event, with a Q^2 value close to the lower end of the analysis phase space and the right panels show a dijet event close to the upper end. The upper panels visualise the ZEUS detector in the z-r-plane. The beige boxes represent the CTD with tracks shown in pink. The blue areas correspond to the CAL, with the red boxes indicating the amount of energy registered in each cell. The reconstructed scattered DIS electron is shown as a blue arrow and reconstructed jets as green arrows. The lower panels show the transverse energy distribution in the η-ϕ-plane. The electrons and high-energy jets are clearly visible as energy clusters. The low-energy jets in the event shown on the left are spread over a larger area and, therefore, less easily visible.

beam pipe. Parts of the proton remnant are sometimes detectable in the FCAL cells adjacent to the beam pipe. Figure 7.2 shows the appearance in the detector of two typical signal events, one at the lower end and one at the higher end of the considered Q^2 range.

Photoproduction Background is expected from misreconstructed PHP events, especially at low Q^2. PHP events can have the same Feynman graph as DIS events, but they occur at very low Q^2, below $1\,\mathrm{GeV}^2$ according to the ZEUS convention. The electron is deflected at a very small angle and escapes into the beam pipe. In the detector, sometimes a different electron, a photon or a neutral pion decaying into two photons is misidentified as the scattered electron, such that the PHP event resembles the signature of a DIS event.

A helpful quantity to reject PHP events is $E - p_{\mathrm{z}}$, the difference of the energy and the momentum along the z-axis summed over all particles in the event. Particles travelling along the $+z$ direction, such as the initial-state proton or the final-state proton remnant, satisfy

$E = p_z$ and thus do not contribute to this difference. In the initial state, this quantity is determined entirely by the beam electron for which $E - p_z = 2E_e$. Due to momentum conservation, the same relation should hold for the final state. If the scattered electron escapes detection, this momentum balance will not be satisfied in the final state, thus making it possible to distinguish PHP and DIS events, even when the scattered electron is not correctly measured.[189]

After applying the event selection, a small amount of PHP events is expected to remain in the data sample. A dedicated PHP MC sample is used to estimate this contribution and subtract it from the data when determining the cross sections.

Charged-current deep inelastic scattering Another type of background is expected from misreconstructed CC DIS events. In these events, the exchanged boson is a W boson and the final-state lepton is a neutrino. Since the neutrino is not measurable in the detector, a different electron, photon or pion can be misidentified as the scattered electron. CC DIS contributes significantly only when Q^2 becomes comparable to the mass of the W boson M_W^2, but since the scattered electron is misidentified, they can contribute as a background at any Q^2 value. CC DIS events can be rejected using transverse and longitudinal momentum conservation. As the neutrino carries a significant momentum and escapes detection, the overall momentum of the event will be unbalanced, which can be used to distinguish CC and NC DIS events.

Cosmic rays Cosmic rays are high-energy particles that move through space, such as protons originating from interstellar or intergalactic objects. When these particles interact with the Earth's atmosphere, they will induce showers of further particles. These showers can contain muons capable of depositing energy in the detector. Since these cosmic muons sometimes cannot be distinguished from particles originating in the beam interaction, they bias the reconstruction of the event or cause events that originate entirely from the cosmic ray. Similar to CC DIS events, cosmic ray interactions will affect the transverse momentum balance and can be rejected similarly. They can also be rejected by timing requirements since cosmic rays are usually inconsistent with the bunch crossing time and interact with the top sections of the detector earlier than with the bottom ones.

Beam-gas interactions The beam pipes and the interaction region are kept under a high vacuum to let the beams move undisturbed through the beam pipes. However, this vacuum is imperfect. A few molecules remain within the beam pipes and interaction region. If such a molecule interacts with one of the beams, it can resemble a signal event in the detector.

Beam-gas interactions often occur outside the interaction region and cause tracks in the detector that cannot be associated with any vertex. The fraction of tracks originating from the primary vertex can be used at trigger level to reject such interactions. They can also be suppressed using trigger-level timing information to reject events inconsistent with the bunch crossing time. Furthermore, depending on the momentum of the beam-gas molecule and with which beam it interacts, they can also cause deviations in the transverse or longitudinal momentum balances, thus giving further opportunity to distinguish them from signal events.

Events with significant initial-state radiation As discussed in section 6.5, radiative corrections to the electron line, especially ISR, can affect the reconstruction of the kinematic quantities. It is helpful to reject events with high energetic ISR photons, to minimise the influence of this effect. Furthermore, the final cross sections will be quoted after correcting

them to leading order QED, where no ISR exists. Therefore, rejecting such events from the data sample reduces the dependence on the modelling of this effect. As the ISR photon travels in the $-z$ direction and escapes detection, it will influence the longitudinal momentum balance. This makes rejecting ISR events possible, using a requirement on $E - p_{\mathrm{z}}$.

Low-Q^2 DIS The measurement focusses on the region $Q^2 > 150\,\mathrm{GeV}^2$. The signal MC samples were generated in the region $Q^2 > 100\,\mathrm{GeV}^2$ and are thus able to describe migrations from $Q^2 \in (100, 150)\,\mathrm{GeV}^2$ into the signal region. However, there is also a non-negligible amount of events that migrate from $Q^2 < 100\,\mathrm{GeV}^2$ into the signal region. Since these misreconstructions often cannot be distinguished from real signal events, rejecting them during event selection is difficult. A dedicated low-Q^2 DIS MC sample is used to estimate the contribution from these events to the data sample. To determine cross sections this contribution will be subtracted from the data.

7.2 Selection of deep inelastic scattering events

The event selection relies on finding a well-reconstructed electromagnetic energy deposit consistent with being the scattered electron. The electron is used to reconstruct the kinematic quantities. Requiring a high-quality electron candidate helps to reject PHP, CC DIS and beam-gas interaction events.

In the figures shown in this section, the a priori corrections discussed in section 8.1 are already applied since these are independent of the details of the event selection. The other corrections discussed in chapter 8 are not yet applied.

Apart from the data and the two NC DIS signal MC samples, Ariadne and Lepto, also background MC samples for PHP and low-Q^2 DIS are shown. The MC samples are discussed in section 5.3.2. Each sample is normalised according to the MC luminosity. In some regions of the phase space, the NC DIS MC samples alone will not describe the data. Therefore, it is instructive to construct a combined MC sample as $(\mathrm{Ariadne} + \mathrm{Lepto})/2 + \mathrm{PHP} + \mathrm{Low}\text{-}Q^2\,\mathrm{DIS}$, that is expected to be more representative of the data. Since cross sections will be determined using both the Ariadne and Lepto MC chains, it is relevant to consider their average instead of just the two predictions individually. In the figures, this combined sample is scaled up by about 2%, to match the normalisation of the data sample.

In the figures, each histogram is shown as if the corresponding cut is applied as the last DIS cut, i.e. only events that are kept by all other DIS cuts described in this section are shown. This representation is practical because the data sample is well-defined only after applying the event selection. E.g. the Q^2 distribution has little physical meaning before applying cuts on the quality of the DIS electron. Furthermore, this representation allows directly assessing the effect of each cut on its own.

7.2.1 Phase space

The kinematic range that is analysed is defined as $150\,\mathrm{GeV}^2 < Q^2_{\mathrm{DA}} < 15000\,\mathrm{GeV}^2$ and $0.2 < y_{\mathrm{DA}} < 0.7$. For calculation of the kinematic quantities via the double-angle method, the hadronic final state is defined using the CorAndCut algorithm, which is based on clusters of CAL cells.[190] The cut in Q^2 restricts the measurement to a well-understood detector region. The scattered electron is usually reconstructed in the BCAL or RCAL and is within the acceptance region of the CTD. The upper cut in y is necessary to ensure a good detector
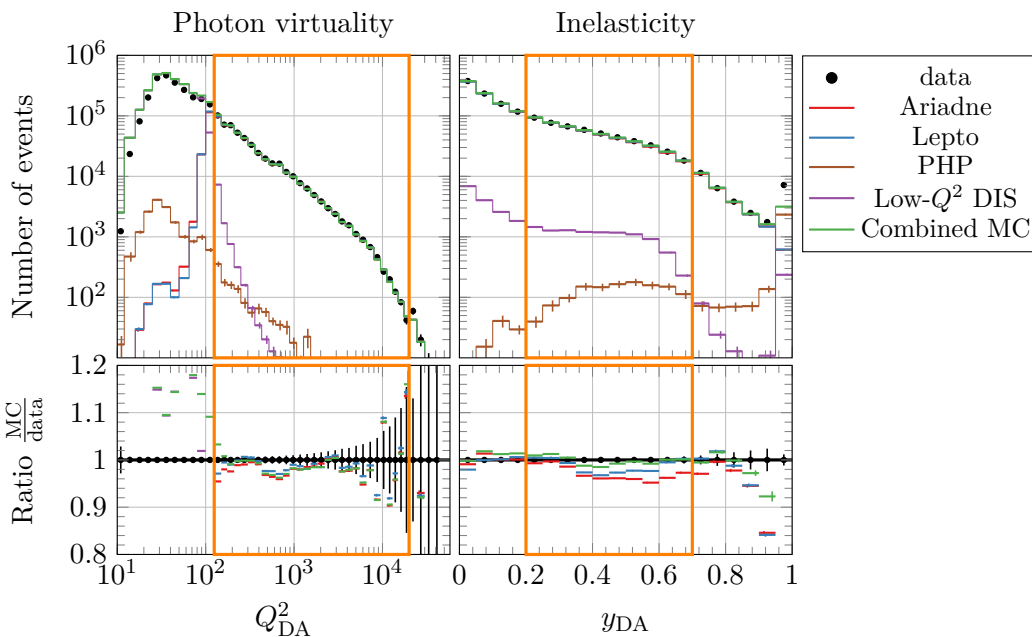
Figure 7.3: Distributions of the quantities Q^2_{DA} and y_{DA} in the data and MC samples. The orange frame shows the region that is retained by the depicted cut. Only events that are kept by all other DIS cuts are shown. Thus, events outside the orange frame are removed only by the cut shown in each histogram. The events inside the orange frame constitute the inclusive DIS sample and are identical in each histogram. The error bars show the statistical uncertainty. The error bars of the MC samples are offset horizontally for visual clarity. The 'combined MC' sample is a linear combination of the other MC samples, expected to correspond the closest to the data, see text.

acceptance of jets, and the lower cut limits the size of non-perturbative effects.[80] This is the same phase space region used in the corresponding analysis from H1, allowing for easy comparison between the two measurements.[121]

7.2.2 Quality cuts

To use the MC simulations for the acceptance correction, they must describe the data as closely as possible. Regions of the phase space not well described by the MC, e.g. due to specific background processes not being simulated or due to an inaccurate detector simulation, should be removed from the event sample. When performing the acceptance correction, these regions are accounted for via an extrapolation from the MC predictions.

Trigger bits

In the data, events were only stored if at least one FLT, SLT and TLT trigger bit was active, as discussed in section 3.2.4. Triggers are also simulated and applied to the simulated events to ensure that the MC describes the data. Since not all triggers are simulated accurately, it is necessary to select a subset of well-described triggers and reject events in both data and MC for which none of these triggers are active.

The selection of trigger bits is made similar to previous high-Q^2 NC DIS and jet analyses at ZEUS.[72,80,189] It is required that at least one of the FLT bits 28, 30, 36, 39, 40, 41, 43, 44, 46, 47 or 50 is active. These bits check for different combinations of energy deposits in different parts of the CAL. Some of them include requirements on the isolation of the energy deposit or on the global number of tracks. At SLT level, at least one of the triggers SPP1, DIS7, EXO1, EXO2 or EXO3 is required. These triggers introduce requirements on the total transverse momentum and the global $E - p_z$. They also contain timing vetos to reject events inconsistent with the bunch crossing time. For TLT, the bit DIS3 is required. This bit refines the requirements of the previous levels and adds additional requirements on the primary vertex reconstruction and the scattered electron candidate. The trigger rates for all mentioned triggers and all run periods are shown in appendix B.1.

Electron quality

The event reconstruction relies heavily on the final-state electron. It is, therefore, crucial to have an accurate and well-modelled measurement of the electron. The first electron identified by the EM algorithm is used as the candidate DIS electron. This electron is required to have a score of at least 0.001, as explained in section 6.2. Furthermore, the DIS electron is required to have an energy of at least $10\,\text{GeV}$.[71,191] This cut restricts the analysis to a well-measured phase space region. It also helps to reject PHP events in which a photon is misidentified as the electron.

Electrons that are not isolated are often poorly reconstructed. Additionally, if the DIS electron candidate is not isolated, there is an increased chance that it is a misidentified non-DIS electron. The sum of all energy deposits in a cone of $\Delta R = 0.8$ around the electron is calculated. The event is rejected if more than $10\,\%$ of this energy is not associated with the electron.[71] This requirement also helps to reject events in which a jet is misidentified as an electron. Distributions for the electron quality cuts are shown in figure 7.4.

Electron-track quality

If the electron is reconstructed in the acceptance region of the CTD, $0.3 < \theta < 2.85$, it is required to have a track associated with it.[191] This requirement helps to reject photons and neutral pions in the CAL that are misidentified as electrons. In the nominal event selection described in this section, the electron is always reconstructed in this region and thus always required to have a track. This distinction only becomes relevant for studying the jet-energy scale, for which the cut on the inelasticity is removed, as discussed in section 10.1.

The track matched to the electron is required to have a momentum of at least $3\,\text{GeV}$.[71,191] This value is notably below the $10\,\text{GeV}$ that is required for the electron itself, to account for FSR and bremsstrahlung photons, that are reconstructed as part of the electron cluster, but do not leave a track. The track is extrapolated to the CAL, and the distance between the track and the energy deposit corresponding to the reconstructed electron at the CAL surface is required to be less than $10\,\text{cm}$.[71,191] These two cuts reject events in which the electron track is mismatched. Distributions of the tracking cuts are shown in figure 7.5.

DIS selection

The longitudinal energy-momentum imbalance $E - p_z$ is used to reject PHP events, NC DIS events, events from beam-gas interactions and events with significant ISR. For NC DIS events, one expects $E - p_z = 2E_e = 55\,\text{GeV}$. Events are selected if they satisfy $38\,\text{GeV} < E - p_z <$
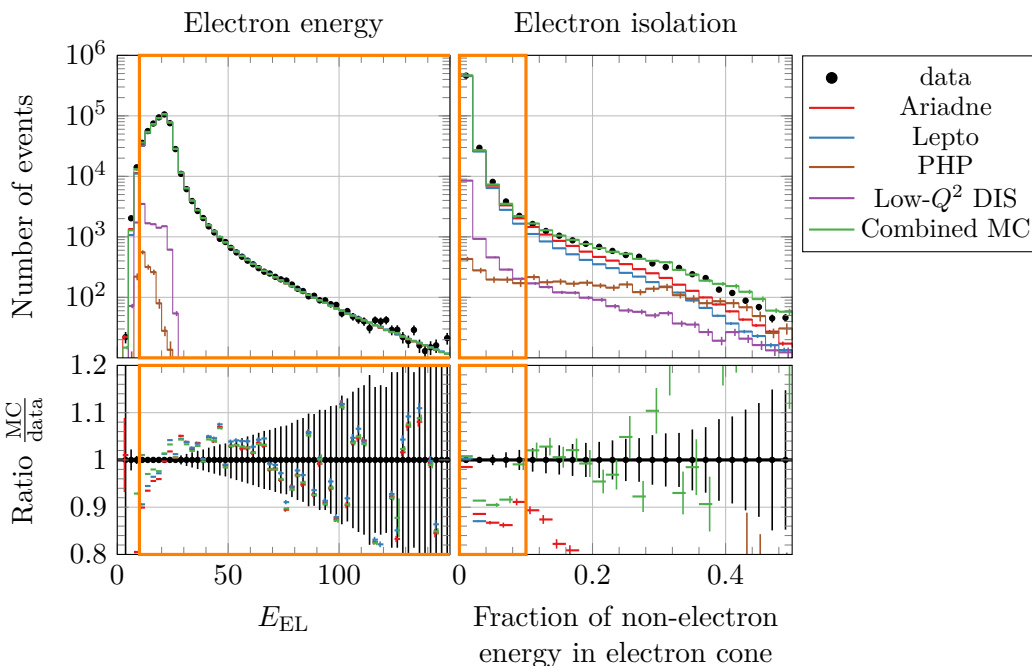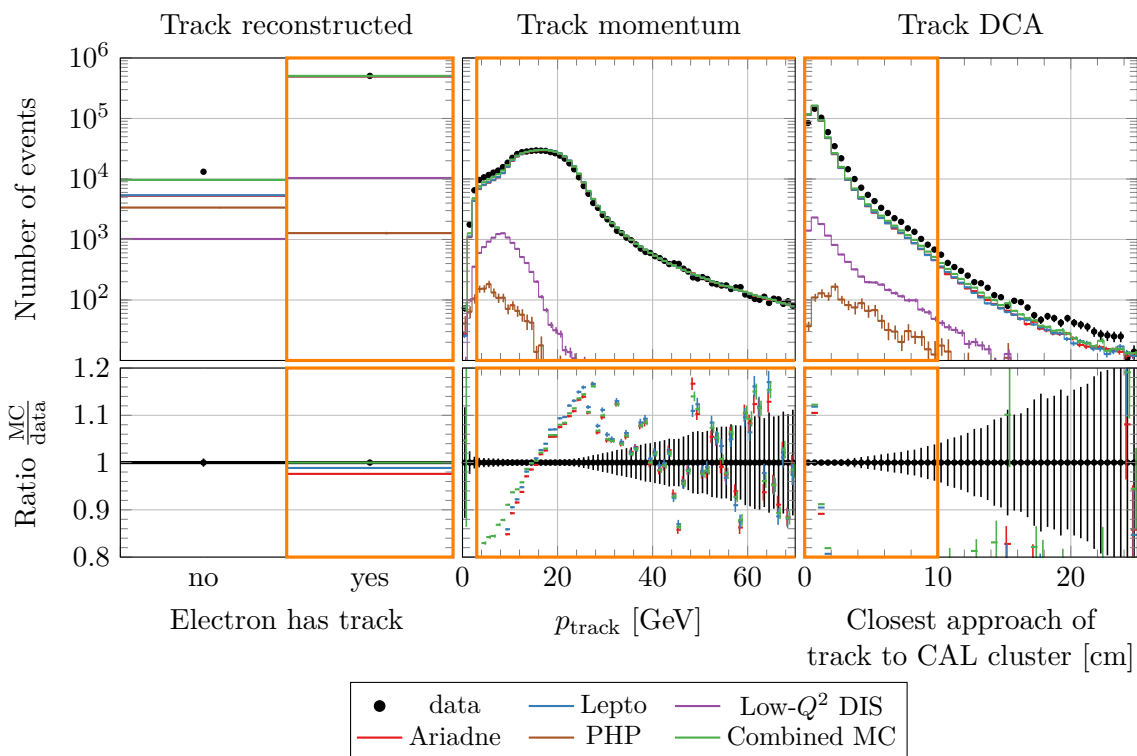
**Figure 7.4:** Distributions for electron quality cuts. For details on the figure, see the caption of figure 7.3. The electron-energy distribution is generally well described by the MC, with some deviations in the steeply rising and falling regions around the peak that are difficult to model accurately. The electron isolation is reasonably described only by the combined MC sample, especially in the region rejected by the cut, indicating that background contributions significantly affect this region. The uncertainty of the electron-energy cut will be covered by the electron-energy scale uncertainty. The uncertainty of the electron-isolation cut will be shown to be negligible in section 10.5.

65 GeV.[71,191] Especially above the upper bound, the MC deviates from the data, indicating the presence of effects that are not modelled. The lower bound can be translated into a cut on ISR events. It corresponds to rejecting events in which the ISR photon has an energy of more than 8.5 GeV.

For NC DIS events, the total reconstructed transverse momentum is expected to vanish. Significant deviations from zero can occur for background events from CC DIS, cosmic rays or beam-gas interactions. Such events are suppressed by requiring that $p_\perp/\sqrt{E_\perp} < 2.5\,\sqrt{\text{GeV}}$, where $p_\perp = \sqrt{\sum_i (p_{\text{x},i}^2 + p_{\text{y},i}^2)}$ is the vectorial sum and $E_\perp = \sum_i \sqrt{p_{\text{x},i}^2 + p_{\text{y},i}^2}$ is the scalar sum of the transverse momenta of all reconstructed particles, according to the CORANDCUT algorithm.[71,190] The normalisation factor $\sqrt{E_\perp}$ is inserted to account for the resolution of the CAL. Distributions for the DIS selection cuts are shown in figure 7.6.

Poorly measured regions of the detector

Due to the geometry of the detector, some regions of the CAL have a lower acceptance or are otherwise poorly understood. This leads to a reduced resolution which is often badly described by the MC. Thus, events in which the scattered electron is reconstructed in such an area should be rejected.

Figure 7.5: Distributions for electron tracking cuts. For details on the figure, see the caption of figure 7.3. The description of the tracking distributions by the MC is not ideal. In the next chapter, multiple corrections will be applied that improve the agreement slightly. The distributions are shown after corrections in figure B.8. Systematic uncertainties will be assigned to cover the remaining disagreement between data and MC close to the cut boundaries.

The CAL comprises three sections, as described in section 3.2.2. The regions where these parts are joined are the so-called super-crack regions and located at $-143.5\,\text{cm} < z < -98.5\,\text{cm}$ and $164\,\text{cm} < z < 216.5\,\text{cm}$.[72,80,189] Electrons reconstructed in this region are measured by the edge of the BCAL. It is likely that parts of the induced showers pass through the BCAL and thereby escape detection, creating a bias in the electron energy measurement. To avoid this bias, events in which the DIS electron is reconstructed in this region are rejected. Only the inner cut boundaries, $-98.5\,\text{cm}$ and $164\,\text{cm}$, are clearly defined, while the outer boundaries are arbitrarily placed between the BCAL and RCAL or FCAL, see figure 7.8.

A support pipe is located in the upper half of the RCAL. The event is rejected if the DIS electron is reconstructed in the RCAL at $y > 80\,\text{cm}$ and $|x| < 12\,\text{cm}$.[72,80,189] The need for this cut becomes clear in figure 7.7, which shows the two-dimensional spatial distribution of electrons reconstructed in the RCAL

The outer region of the RCAL is screened by the BCAL, which reduces the energy of measured particles. If the DIS electron is found within RCAL and more than $175\,\text{cm}$ from the beam axis, the event is rejected.[72,80,189] Distributions corresponding the each of these cuts are shown in figure 7.8.
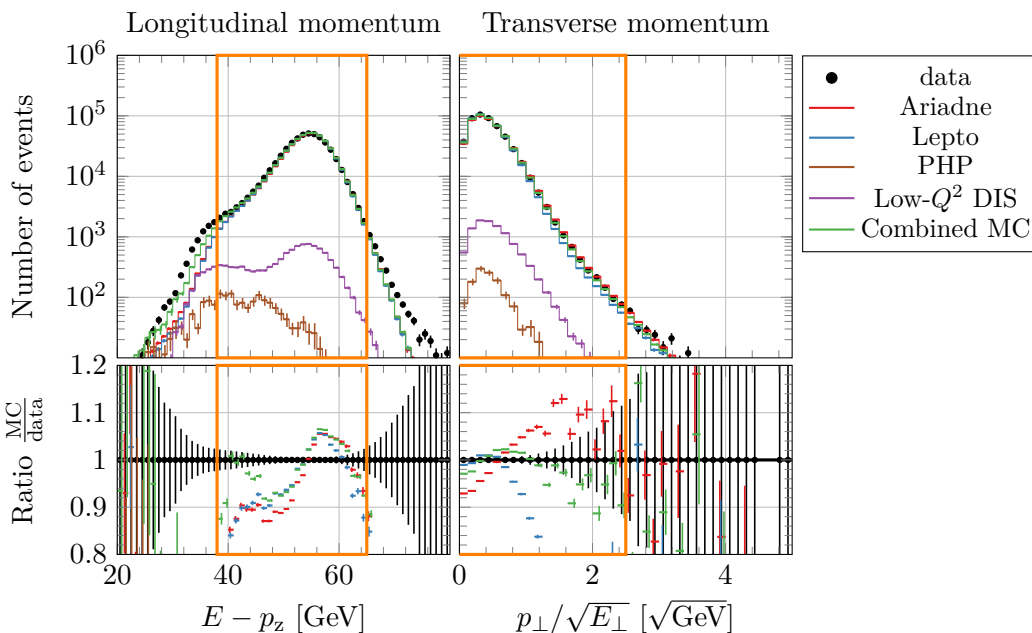
Figure 7.6: Distributions for further DIS selection cuts. For details on the figure, see the caption of figure 7.3. The signal MC samples alone do not describe these distributions satisfactorily, especially in the regions removed by the cuts. The agreement is much improved when compared to the combined MC sample. This demonstrates that these distributions are significantly affected by background and that these cuts are useful in rejecting non-NC-DIS events. In the transverse momentum distribution, modelling differences between Ariadne and Lepto can be observed, but their average gives a good description of the data. The difference between Ariadne and Lepto will be accounted for as a systematic uncertainty. A dedicated systematic uncertainty will be assigned to each of these cuts to cover the remaining disagreement between data and MC close to the cut boundaries.

Other quality cuts

Most collision events occur at a longitudinal position of $|z| \lesssim 10\,\mathrm{cm}$, a length which is determined mainly by the proton bunch length. Due to imperfections in the accelerator, particle bunches are usually surrounded by smaller, so-called satellite bunches. Interactions can also occur between nominal bunches of one beam and satellite bunches of the other, usually happening at $|z| \gtrsim 50\,\mathrm{cm}$. Since all event kinematics are affected by the vertex position, it is desirable for the vertex to be located close to $z = 0$. A large fraction of events that do not originate from interactions of the nominal bunches can be rejected by requiring that the primary vertex of each interaction is located at $|z| < 30\,\mathrm{cm}$.[71,191]

The position of the event vertex is determined from the tracks of charged particles, as measured in the MVD and CTD. A fit is then performed to determine the primary and possibly secondary vertex positions. To ensure this fit is possible, at least one well-defined track matched to the primary vertex is required. A track is considered well-defined if it passes through at least three superlayers and has a transverse momentum of $p_\perp > 0.2\,\mathrm{GeV}$. The sample quality is further improved by requiring χ^2/degree of freedom < 10 for this fit. The distributions of the vertex-related cuts are shown in figure 7.9.[72,80,189]
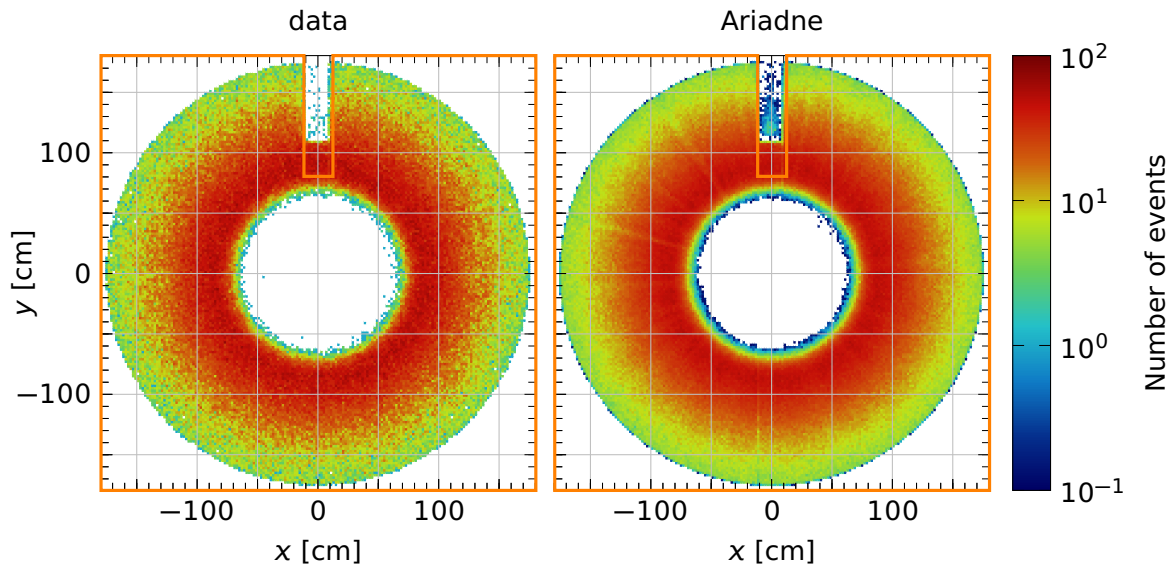
Figure 7.7: Spatial distribution of electrons reconstructed in the RCAL after applying all but the RCAL chimney cut. The hole that is caused by the support pipe is visible. The orange outline indicates the region that is retained by the cut. The left panel shows the distribution in the data and the right panel that in the Ariadne MC sample. The region that is removed by the cut is not well-modelled.
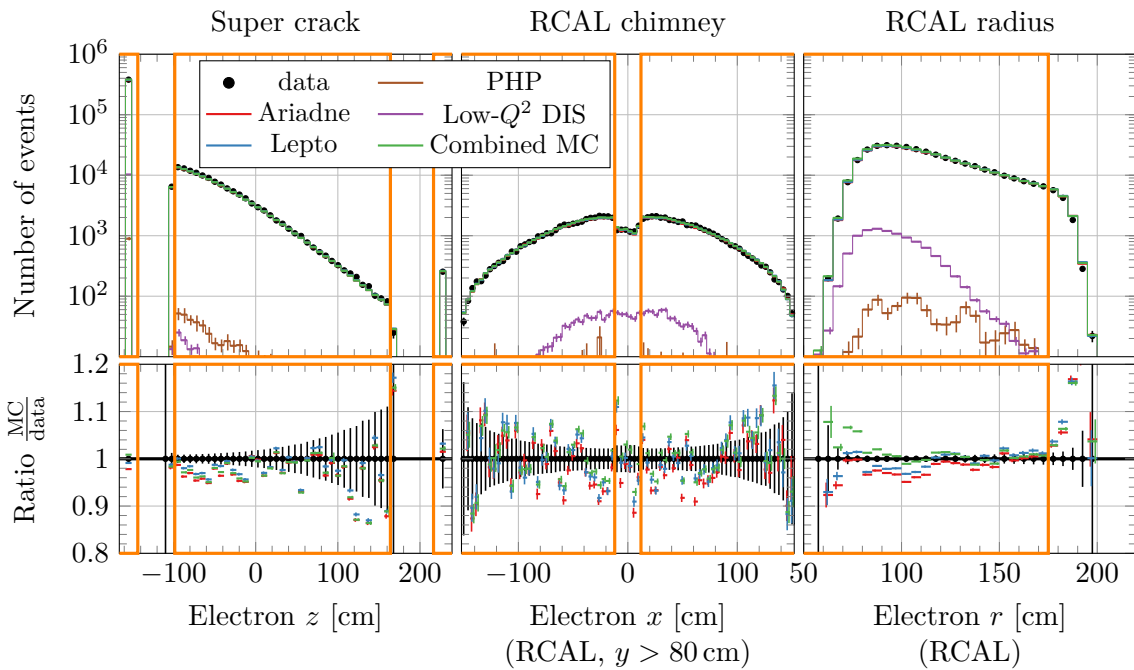

Figure 7.8: Distributions for cuts on poorly measured detector regions. For details on the figure, see the caption of figure 7.3. All three distributions are reasonably described by the MC.
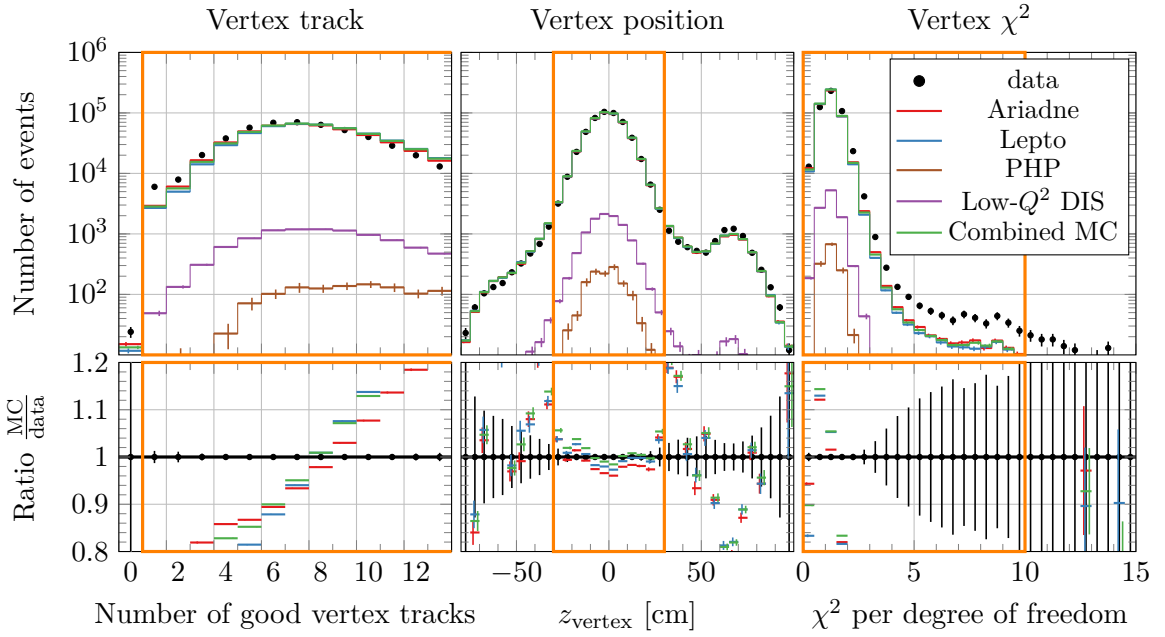
Figure 7.9: Distributions for vertex-related cuts. For details on the figure, see the caption of figure 7.3. Similar to the electron tracking cuts shown in figure 7.5, there is a notable deviation between data and MC. The agreement improves after applying the corrections described in the next chapter. The distributions are shown after corrections in figure B.9. For the vertex position cut, a systematic uncertainty will be introduced. The efficiencies of the other two cuts are sufficiently high to neglect their uncertainties.

The Feynman graphs of diffractive QED Compton events are given in figure 7.10. These events are not part of the cross section definition and are not modelled by any of the MC samples used. Therefore, they should be rejected from the data sample. Such an event is characterised by two isolated electromagnetic energy deposits in the CAL, one of which is the scattered electron, the other one the photon. The photon can be misreconstructed as an electron candidate. The two clusters balance the momentum of each other, and the event usually contains no other energy deposits since the proton remains intact. An event is classified as QED Compton and thus rejected if there is another electron candidate that has an angular
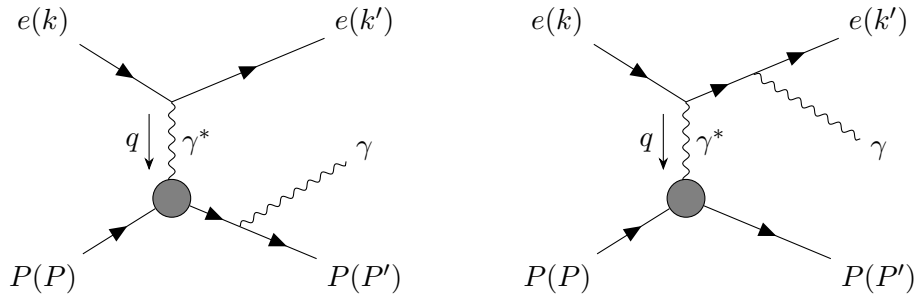
Figure 7.10: Diffractive QED Compton graphs in which a non-collinear photon is emitted from the final-state proton (left) or electron (right). The proton remains intact. Two more graphs exist in which the photon is emitted from the initial-state particles.

separation $\Delta\phi > 3$ to the nominal electron, whose momentum is within 20% of the nominal electron's momentum and if there is no more energy than $3\,\text{GeV}$ in the CAL, that is not associated with either of the two electron candidates.[72,80,189] According to this definition, about 0.6% of data events are classified as QED Compton events, while less than 0.02% of events are misidentified as such in the MC samples.

7.3 Selection of single inclusive jets

The inclusive jet sample is selected based on the inclusive DIS sample. Even though they are reconstructed from the same events, the two samples are of a very different nature. The inclusive DIS sample is a set of events, whereas the inclusive jet sample is a set of jets, each of which is associated with an event. An event with multiple jets can thus contribute multiple times, introducing statistical and systematic correlations within the inclusive jet sample.

The starting point for jet reconstruction is the corrected inclusive DIS sample, whose selection is described in the previous section and which includes the corrections described in sections 8.1 and 8.2. In the figures shown in this section, these corrections are already applied. The jet corrections from section 8.3 are not yet applied.

Massless jet candidates are clustered in the Breit reference frame based on CAL cells as described in section 6.6. Jet candidates are subjected to a series of cuts. Candidates that pass all cuts are then referred to as jets.
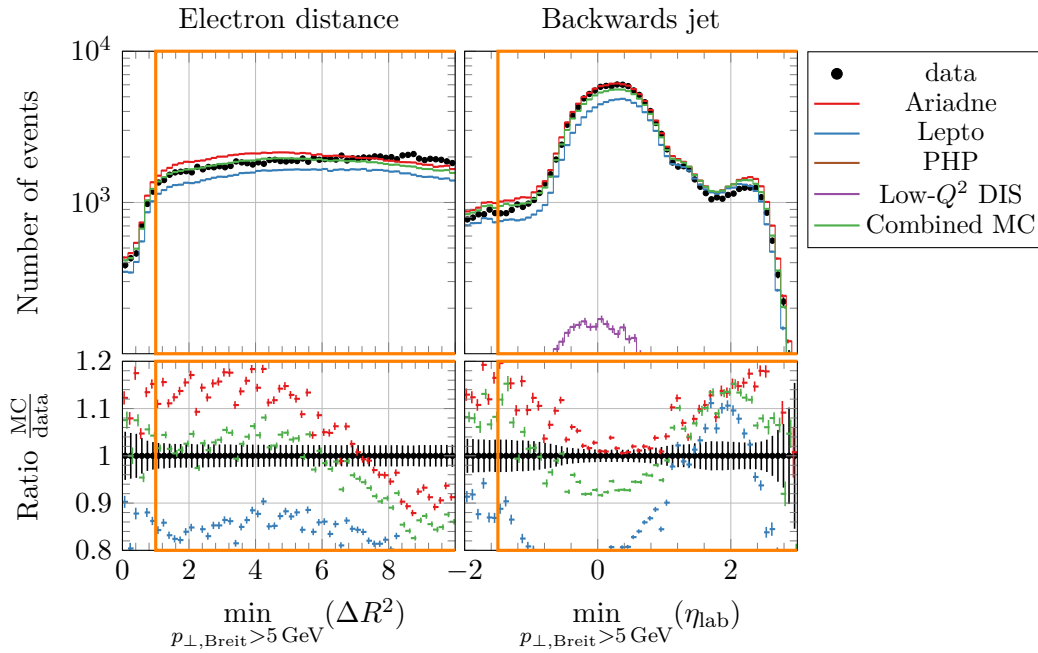
Figure 7.11: Distributions for jet veto cuts. For details on the figure, see the caption of figure 7.3. The figures are shown after applying the inclusive DIS selection. Each cut is shown as if it is applied as the last jet veto cut. The jet corrections described in section 8.3 improve the agreement between data and MC significantly. Corrected distributions are shown in figure B.10. The uncertainty due to the electron-distance cut becomes negligible after corrections, and the uncertainty due to the backwards-jet cut is covered by the model uncertainty.

Before applying the jet cuts, a series of jet vetos is applied. A veto is a specific condition under which the entire event is rejected. An event is rejected if any jet candidate with $p_{\perp,\text{Breit}} > 5\,\text{GeV}$ is closer than $\Delta R = 1$ to the DIS electron.[71] In such events, the electron and the jet may overlap, which leads to a poor reconstruction of both of them. Events are also rejected if any jet candidate with $p_{\perp,\text{Breit}} > 5\,\text{GeV}$ is found at $\eta_{\text{lab}} < -1.5$, as such candidates are often misreconstructed ISR photons from the electron. Distributions of the angular separation between the DIS electron and the jet candidates, as well as the most backward jet candidate, are given in figure 7.11.

After applying the jet veto cuts, the main jet cuts are applied. Each jet is required to have a transverse momentum in the Breit frame between 7 and 50 GeV. The lower bound removes candidates from the QPM-like process, that are not interesting for a study of QCD. The upper bound restricts the measurement to a region of sufficient statistics. Each jet is required to have a transverse momentum in the laboratory frame of at least 3 GeV.[71] This requirement rejects low-energy jet candidates that have a poor resolution in the CAL. Finally, each jet is required to be in the region $-1 < \eta_{\text{lab}} < 2.5$.[71] Here, the upper bound removes candidates that point in a very forward direction, as there is a high probability that they partially escaped into the beam pipe. The lower cut also removes a region of low statistics. Distributions for these quantities are shown in figure 7.12.
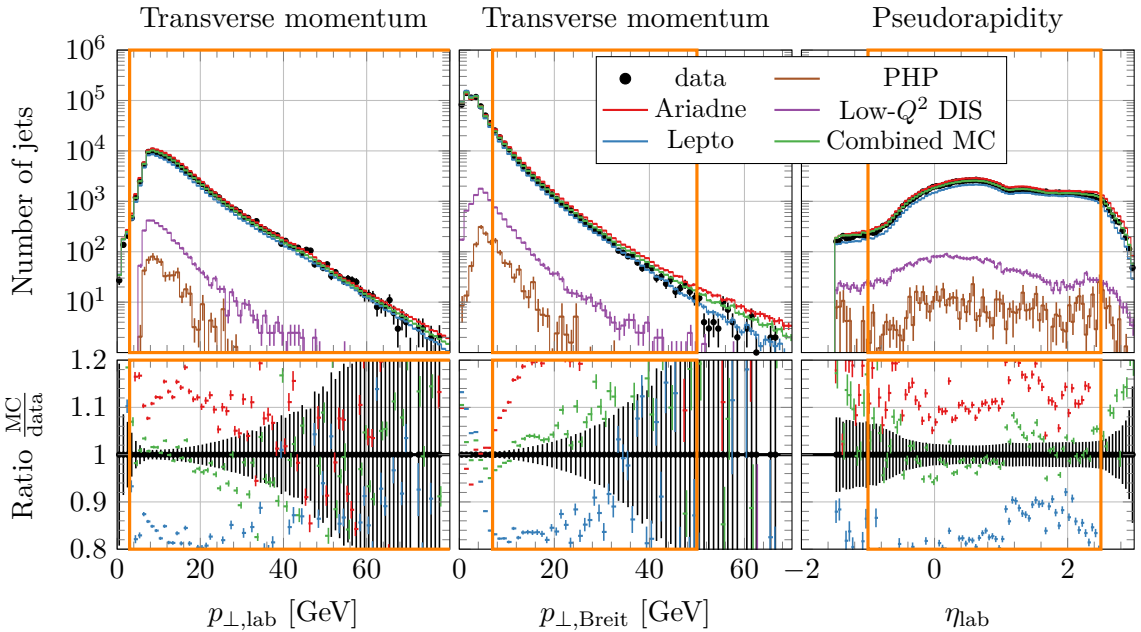
Figure 7.12: Distributions for jet cuts. For details on the figure, see the caption of figure 7.3. The figures are shown after the jet veto cuts. Each cut is shown as if it is applied as the last jet cut, i.e. only jet candidates that are kept by all other cuts are shown. The jet corrections described in section 8.3 improve the agreement between data and MC significantly. Corrected distributions are shown in figure B.10. A systematic uncertainty will be assigned to the $p_{\perp,\text{lab}}$ cut.

7.4 Hadron-level selection

The hadron-level event selection is crucial as it defines the cross sections of the measurement. Section 11.1 gives the complete cross section definition. The hadron-level event samples are also used for some of the corrections described in the next chapter. The hadron-level selection is more straightforward than the one at detector level, since it only consists of phase space cuts and does not contain any quality cuts.

7.4.1 Inclusive DIS selection

The kinematic variables are defined at hadron level in the presence of QED radiation on the electron side as discussed in section 6.5. Events are selected if they fall into the analysis phase space $150 \, \text{GeV}^2 < Q^2 < 15000 \, \text{GeV}^2$ and $0.2 < y < 0.7$.

Distributions of these quantities are shown in figure 7.13. The figures are shown after the a priori corrections discussed in section 8.1. Of those three corrections, only the first affects the hadron level.

The MC event generators do not use the same definition of the kinematic variables in the presence of QED radiative effects. Therefore, even though the samples were generated with a sharp cut of $Q^2 > 100 \, \text{GeV}^2$, the cut is not exact according to the definition used in this analysis. This is not problematic since the analysis phase space is sufficiently far away from this cut.

7.4.2 Jet selection

Hadron-level jets are clustered similarly to detector-level jets, as described in section 6.6. The input to the clustering algorithm consisted of generator-level hadrons, charged leptons and photons with a lifetime of more than $10 \, \text{ps}$, which corresponds to a flight time of $3 \, \text{mm}$. This roughly means that weakly decaying particles made from light quark flavours, such as kaons or lambda baryons, are considered stable. Strongly and electromagnetically decaying particles and weakly decaying hadrons involving heavy quarks, such as B or D mesons, are considered unstable. Neutrinos are not used during jet construction. This definition corresponds most closely to the measured data at detector level. Particles with a longer lifetime are likely to interact with the inner components of the detector and thus cannot decay naturally. Particles with a shorter lifetime might decay into neutrinos which escape undetected.

These particles are made massless by setting the absolute value of their three-momentum equal to their energy. This corresponds to the reconstruction at detector level since the calorimeter cells measure the energy of the particles, but not their momenta. The massless four-momenta corresponding to each particle are boosted to the Breit frame. There, the $k_\perp$-clustering algorithm is applied using the $p_\perp$-weighted combination scheme and a distance parameter of $R = 1$.

Inclusive jets are selected if they belong to the analysis phase space $7 \, \text{GeV} < p_{\perp,\text{Breit}} < 50 \, \text{GeV}$ and $-1 < \eta_{\text{lab}} < 2.5$. The distributions of these two cuts are shown in figure 7.14.
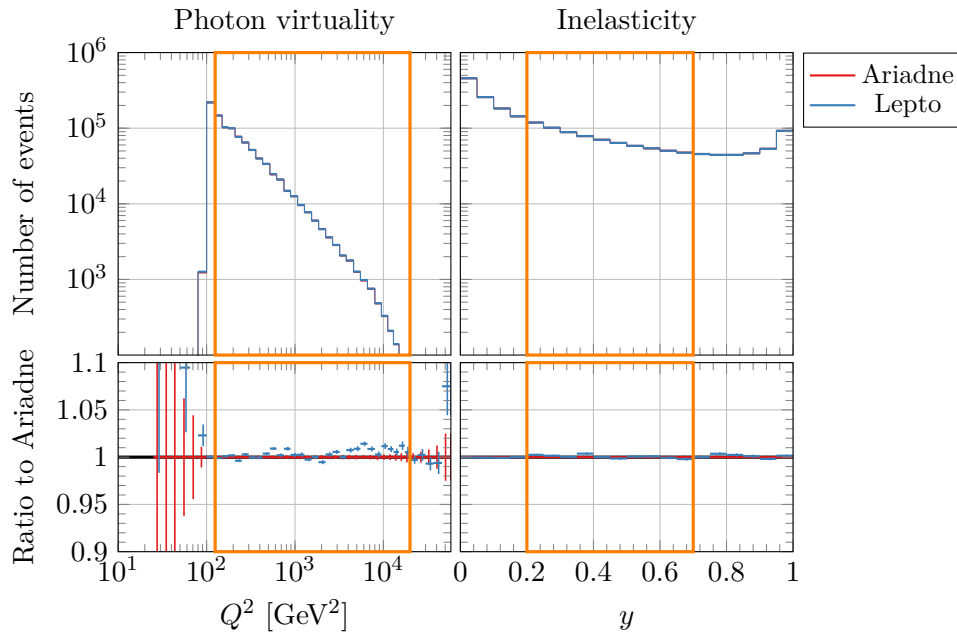
Figure 7.13: Distributions for inclusive DIS cuts at hadron level. The histogram for each cut is shown as if the other cut is already applied. The two MC chains are in excellent agreement in both distributions because they are based on the same matrix element generator.
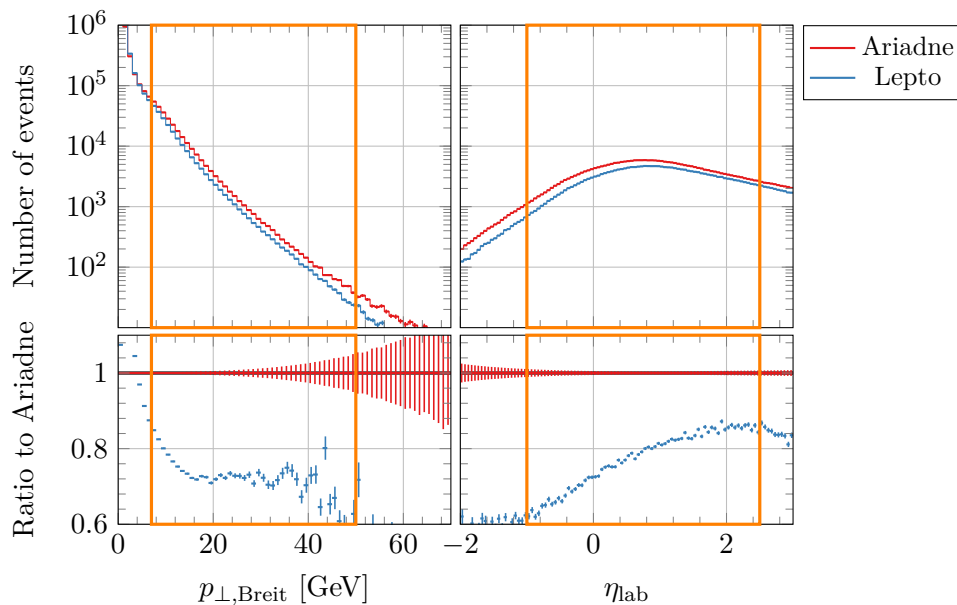
Figure 7.14: Distributions of inclusive jet cuts at hadron level. For details on the figure, see the caption of figure 7.3. The samples differ significantly due to the different models for parton showering. The agreement will improve significantly after applying the jet reweighting described in section 8.3. These distributions are shown after corrections in figure B.11.

CHAPTER 8

Corrections

Using MC samples that accurately describe the data is essential to a precise measurement. Due to deficiencies in the MC models, the initial description of the data is not ideal. Since many of these deficiencies are well understood, targeted corrections can be applied.

In this chapter, corrections are divided into three types. A priori corrections are corrections that can be applied without comparing data and MC as there is a single and unique way to apply them. Since these corrections are independent of the event selection, they are already applied in the figures shown in the previous chapter. DIS corrections are applied to the inclusive DIS sample. Most of them are derived by comparing the data and MC samples and deriving correction factors that are then applied to the MC. Some of them also involve the MC hadron level. Jet corrections work similarly to DIS corrections, except that they are derived based on the inclusive jet sample. All three types of corrections include corrections to the data, the detector simulation and the MC event generation.

All corrections are derived separately for each run period `04p`, `05e`, `06e` and `0607p`. Corrections to the MC event generation are applied separately to the Ariadne and Lepto MC chains. For corrections to the detector simulation, common factors are derived and applied to both chains. Corrections are not applied to background MC samples, as the effect would be negligible, and some corrections do not apply to all background samples.

8.1 A priori corrections

8.1.1 Longitudinal structure function

The longitudinal structure function correction is a correction to the MC event generator. Both signal MC chains used in this analysis neglect this contribution in their matrix elements, i.e. they set $F_{\mathrm{L}} \equiv 0$. This influences mostly the large-y region, according to equation (2.13). Within this region, the structure function plays a role, especially at low x_{Bj}, where it can affect the cross section by up to 10%.[192]

Two MC samples are generated, one including the longitudinal structure function and one excluding it. Cross sections are computed from each sample, and the ratio of those is used to reweight the signal MC sample at generator level, as a function of Q^2 and x_{Bj}.[189] The most notable effect of this correction, a change in the distribution of the inelasticity, is shown in figure 8.1.
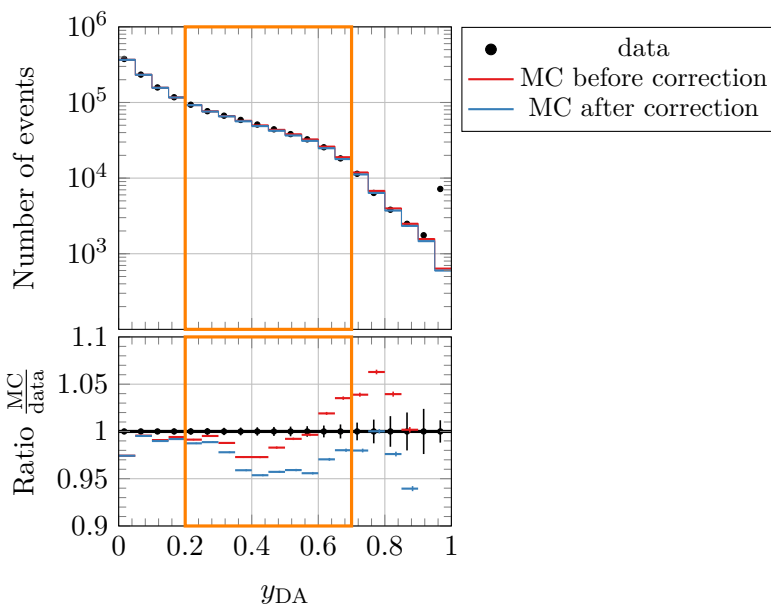
Figure 8.1: Distribution of the inelasticity at detector-level before and after applying the correction for the longitudinal structure function. The Ariadne and Lepto MC samples are combined for this figure. The correction improves the agreement of the shape of the MC distributions with the data.

8.1.2 FLT30 trigger efficiency

The FLT30 trigger efficiency correction is a correction to the detector simulation. During parts of the `0607p` run period, the FLT30 bit was misconfigured in the data.[189] Affected are 122 runs in the run range 60226 to 60778. The desired configuration, and the one that was used in the MC, was `(RCAL_EMC_TH>15000 or RCAL_EMC>3992)`, where `RCAL_EMC_TH` and `RCAL_EMC` are FLT estimates of RCAL energy deposits in MeV. The actual configuration used in the data was `RCAL_EMC_TH>15000`, which means that some events were mistakenly discarded from the data while still being present in the MC samples.

This discrepancy can be corrected by applying the undesired configuration also to the MC, thus losing some events but improving the description of the data. For all events in the affected runs, FLT30 is turned off, if `RCAL_EMC_TH<15000`. This does not necessarily mean that the event is rejected, as it might still be triggered by other FLT bits. The trigger rate of FLT30 for the affected run range is shown in figure 8.2.

8.1.3 Polarisation

The polarisation correction is a correction to the data at detector level. During the HERA II period, the lepton beam provided by the HERA accelerator was longitudinally polarised, as described in section 3.1. The QCD interaction does not depend on the polarisation, but the inclusive DIS cross section does. The goal of this analysis is to measure cross sections for unpolarised beams so the data are corrected for polarisation. This is the only reweighting applied to the data instead of the MC since the two employed MC chains already use unpolarised beams.
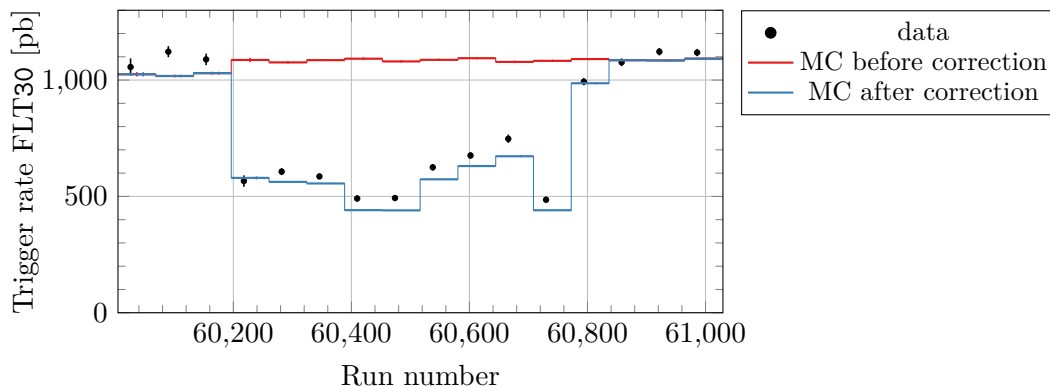
Figure 8.2: Trigger rate of FLT30 during part of 0607p run period before and after trigger configuration correction. The Ariadne and Lepto MC samples are combined for this figure. After applying the correction, the drop in trigger rate in the data is described by the MC samples.

Several sets of MC samples are generated as described in section 5.3.3, two using unpolarised beams, one for an electron and one for a positron beam, and four using the average polarisations of the `04p`, `05e`, `06e` and `0607p` run periods and the corresponding lepton type. The ratio of the unpolarised to polarised cross section is used to determine correction factors for each period. To reduce binning effects, a quadratic polynomial of $\log(Q^2)$ is fitted to each ratio and is then used to reweight the data distributions at detector level as a function of Q^2_{DA}. The ratios are shown in figure 8.3.

The effect on the cross section can be significant for the individual run periods, especially for the `04p` period, where the correction reaches a size of more than 10% at high Q^2. When considering all run periods together, the different polarisations cancel almost completely, such that the total effect of this correction is below 0.5% in the entire phase space.
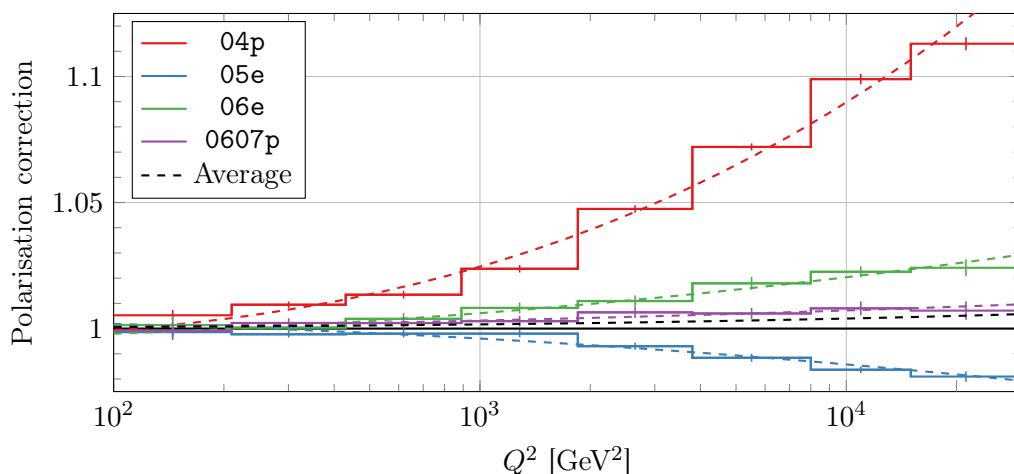
Figure 8.3: Polarisation correction factors ($\sigma_{\mathrm{unpolarised}}/\sigma_{\mathrm{polarised}}$). The dashed lines show the fitted polynomial. The black dashed line denotes the luminosity weighted average over all run periods.

8.2 DIS corrections

8.2.1 Vertex position

The primary vertex position is chosen by the detector simulation. Nevertheless, the vertex position correction should be treated as a correction to the event generator, which is equal for both signal MC samples. As described in section 7.2.2, the distribution of the longitudinal position of the primary vertex shows multiple peaks. The so-called 'satellite peaks' at $|z| \approx 60$ cm are not well described by the MC and are discarded for this analysis. Nevertheless, they must be reasonably described by the MC, as they affect the efficiency of the corresponding cut and, thereby, the normalisation of the measured cross sections.[72,80,189]

The ratio of the vertex position distributions between data and MC is computed at detector level. This ratio is then used to reweight the MC at generator level as a function of the true position of the vertex. The overall normalisation of the MC samples is maintained. To reduce binning effects, linear interpolation of the reweighting factors is performed between the two closest bin centres. Since the vertex position is determined by the detector simulation, common correction factors are derived for both signal MC chains. The corrected distributions are shown in figure 8.4.

8.2.2 Track-matching efficiency

The event selection contains three cuts related to the electron track: track existence, track momentum and distance of closest approach to CAL cluster. These cuts remove events in which no track or a wrong track is assigned to the DIS electron and events in which a photon is misidentified as the DIS electron. Neither of these effects is described by the signal MC
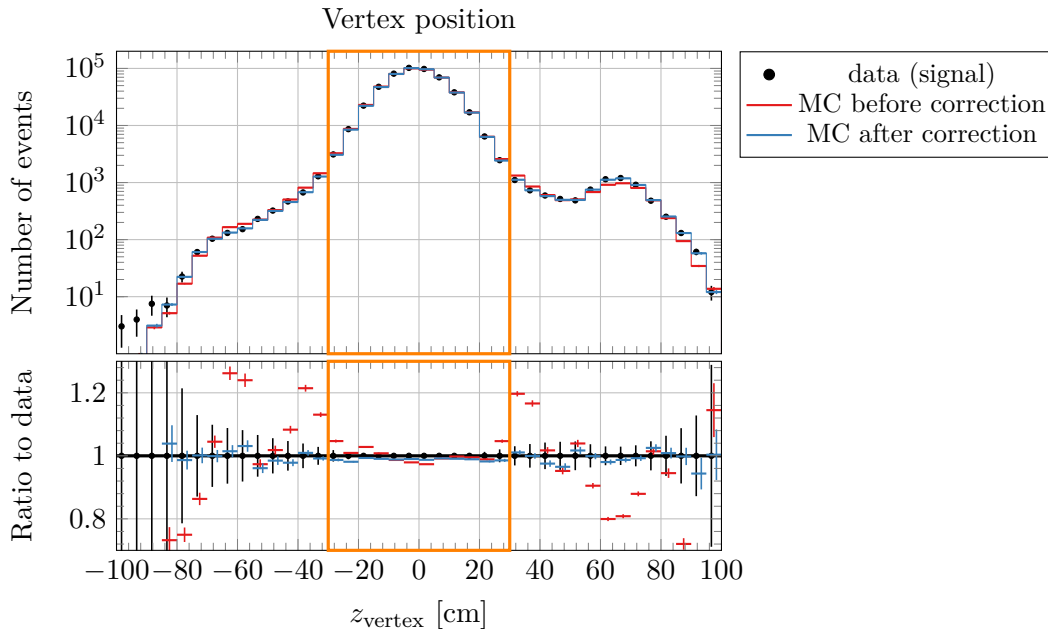
Figure 8.4: Distribution of longitudinal position of the primary vertex position before and after the corresponding correction. After correction, the MC samples describe the shape of the data very well.

samples. That the detector simulation does not describe the efficiency of matching tracks to the DIS electron was shown in a previous analysis.[189] Since the signal MC does not simulate isolated photon emissions, the rate of photon misidentification is not described either.

After applying these cuts, events in which the DIS electron candidate is a misidentified photon are removed from the event sample. Therefore, and since these events are not part of the cross section definition, it is not problematic that they are not described by the signal MC. The efficiencies of these cuts will differ in data and MC, and the MC should explicitly not be corrected for this effect.

Events in which a wrong track is matched to the DIS electron are, however, part of the cross section definition, and they must be adequately described by the MC. Therefore, when designing a correction, care has to be taken to correct for the rate of mismatching a track to the DIS electron but not for the rate of misidentification of photons.

Before reweighting the signal MC to the data, the background MC samples are subtracted from the data. The PHP background MC describes isolated radiation of neutral pions that decay into two photons. This sample describes part of the photon misidentification rate and ensures it is not erroneously corrected. Another contribution to the photon misidentification rate comes from the direct emission of isolated photons, which is not described by any of the considered MC samples. To account for these events, the correction described below is only applied halfway. A systematic uncertainty will be introduced to cover the effect of applying this correction either fully or not at all, see section 10.7.

The track-matching efficiency is defined as the fraction of events that pass the tracking cuts, given that they already passed all remaining DIS cuts that are listed in section 7.2

$$\text{Track matching efficiency} = \frac{\text{Number of events that pass all DIS cuts}}{\text{Number of events that pass all non-tracking cuts}}. \quad (8.1\text{a})$$

The events in the MC that pass the tracking cuts are then reweighted according to the ratio of the track matching efficiency in signal data and signal MC. Events that are rejected by the tracking cuts are reweighted using the ratio of the track matching inefficiency

$$w_{\text{tracking cut passed}} = \frac{\text{Track matching efficiency}_{\text{data}}}{\text{Track matching efficiency}_{\text{MC}}}, \quad (8.1\text{b})$$

$$w_{\text{tracking cut rejected}} = \frac{1 - \text{Track matching efficiency}_{\text{data}}}{1 - \text{Track matching efficiency}_{\text{MC}}}. \quad (8.1\text{c})$$

Reweighting events that are rejected by a cut is not strictly necessary, but it is convenient, as it preserves the overall normalisation of the signal MC. Since this is a correction to the detector simulation, the reweighting is applied to the detector-level weight, see section 5.3.4. To apply this correction halfway, events are reweighted by $\sqrt{w}$ instead of w.

As the track-matching efficiency depends on the position of the reconstructed electron in the detector, this correction is applied as a function of θ. To reduce binning effects, linear interpolation of the reweighting factors is performed between the two closest bin centres. The effect of this correction on the track matching efficiency is shown in figure 8.5.

The halfway application procedure is checked using a dedicated PHP MC sample that includes the direct emissions of isolated photons. It is estimated that the contribution to the data sample is about 10^{-3} before the tracking cuts and 10^{-4} afterwards. Therefore, this contribution does not affect the track-matching efficiency significantly. The systematic uncertainty assigned to this correction will cover the uncertainty associated with neglecting this contribution.
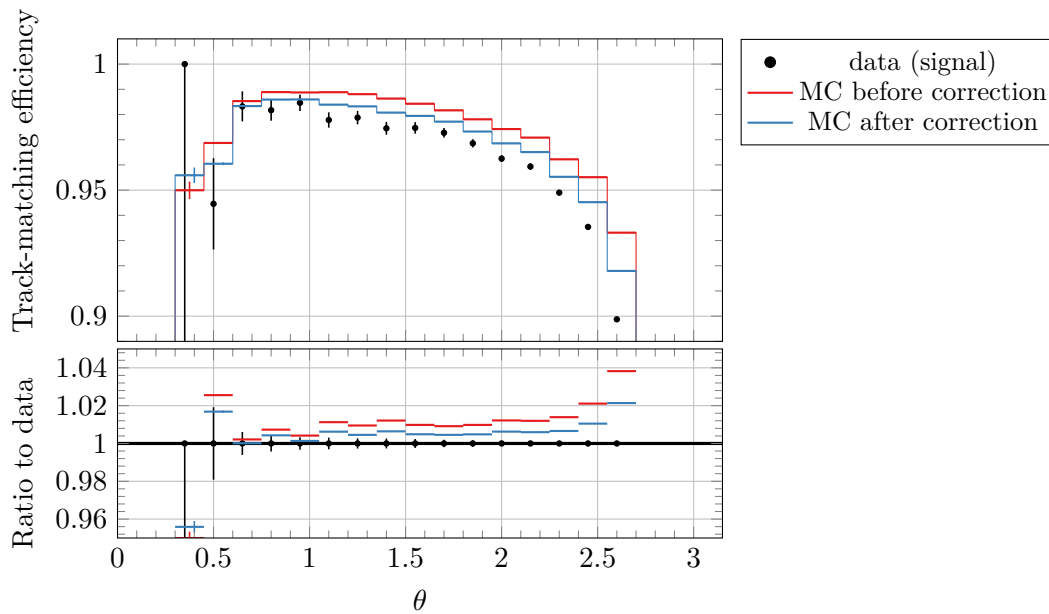
Figure 8.5: Track-matching efficiency before and after correction. The Ariadne and Lepto MC samples are combined for this figure. Since the correction is applied halfway, the corrected MC distribution is in the middle between the data and the uncorrected MC.

8.2.3 Track-veto efficiency

The track-veto-efficiency correction is a correction to the trigger simulation of the detector simulation. At the FLT, most bits have a track veto associated with them. A track veto is a condition under which the corresponding trigger bit does not activate, even if its remaining conditions are fulfilled. This condition is related to the total number of tracks and the track class. The track class is a function of the total number of tracks and the number of tracks matched to the primary vertex. This definition helps reject beam-gas interaction, see section 7.1. It is known that the detector simulation does not describe the efficiencies of these track vetos and, therefore, the corresponding trigger bits exactly.[72,80,189]

The FLT bits used in this analysis use four different types of track veto, including the absence of a track veto.

FLT bits	Track veto condition
30	none
36, 39, 41, 43, 46, 47	Track class = 2
28, 40, 50	Track class = 2 or (Track class = 8 and at least 26 tracks)
44	Track class = 2 or Track class = 8

If the listed condition is met, the corresponding FLT bits cannot be active. FLT30 is the only bit that does not have a track veto, i.e. it can always be active. Its efficiency is, therefore, independent of the description of the track vetos by the MC.

Based on these track-veto types, events are split into four distinct categories.

	Condition	FLT bits that can be active
A	Track class = 2	30
B	Track class = 8 and at least 26 tracks	30, 36, 39, 41, 43, 46, 47
C	Track class = 8 and less than 26 tracks	28, 30, 36, 39, 40, 41, 43, 46, 47, 50
D	None of the above	28, 30, 36, 39, 40, 41, 43, 44, 46, 47, 50

Category A is the strictest, as all three track-veto conditions are fulfilled, and only FLT30 can be active. Category D is the loosest, where each FLT bit might be active.

The straightforward way to apply a correction would be to define the track-veto efficiency as

$$\text{Track-veto eff. } i = \frac{\text{Number of category } i \text{ events that pass all non-FLT DIS cuts}}{\text{Number of events that pass all non-FLT DIS cuts}},$$

where i is A, B, C or D. The ratio of the efficiencies in data and MC could then be used to reweight the MC at detector level. Here 'non-FLT DIS cuts' refers to all cuts presented in section 7.2, except for the cut on the FLT bits. This procedure is, however, not possible since events that were rejected by the FLT got discarded during data taking and are therefore not available for analysis. Thus, the above definition of the track-veto efficiency cannot be applied to the data.

Instead, the track-veto efficiency is estimated based on all events for which FLT30 is active

$$\text{Track-veto eff. } i = \frac{\text{Number of category } i \text{ events that pass all DIS cuts and FLT30 is active}}{\text{Number of events that pass all DIS cuts and FLT30 is active}}.$$

Here, the track-veto efficiency is assumed to be independent of FLT30. A weight is then determined as a function of the FLT track multiplicity for each category as the ratio of the track-veto efficiency in data and MC

$$w_i = \frac{(\text{Track-veto eff. } i)_{\text{data}}}{(\text{Track-veto eff. } i)_{\text{MC}}}. \tag{8.2}$$

These weights are then used to reweight the MC samples separately for each event category i. Since this is a correction for the detector simulation, it is applied to the detector-level weight, see section 5.3.4. To reduce binning effects, linear interpolation of the reweighting factors is performed between the two closest bin centres. Events for which FLT30 is active are not reweighted since those are triggered independently of the track veto.

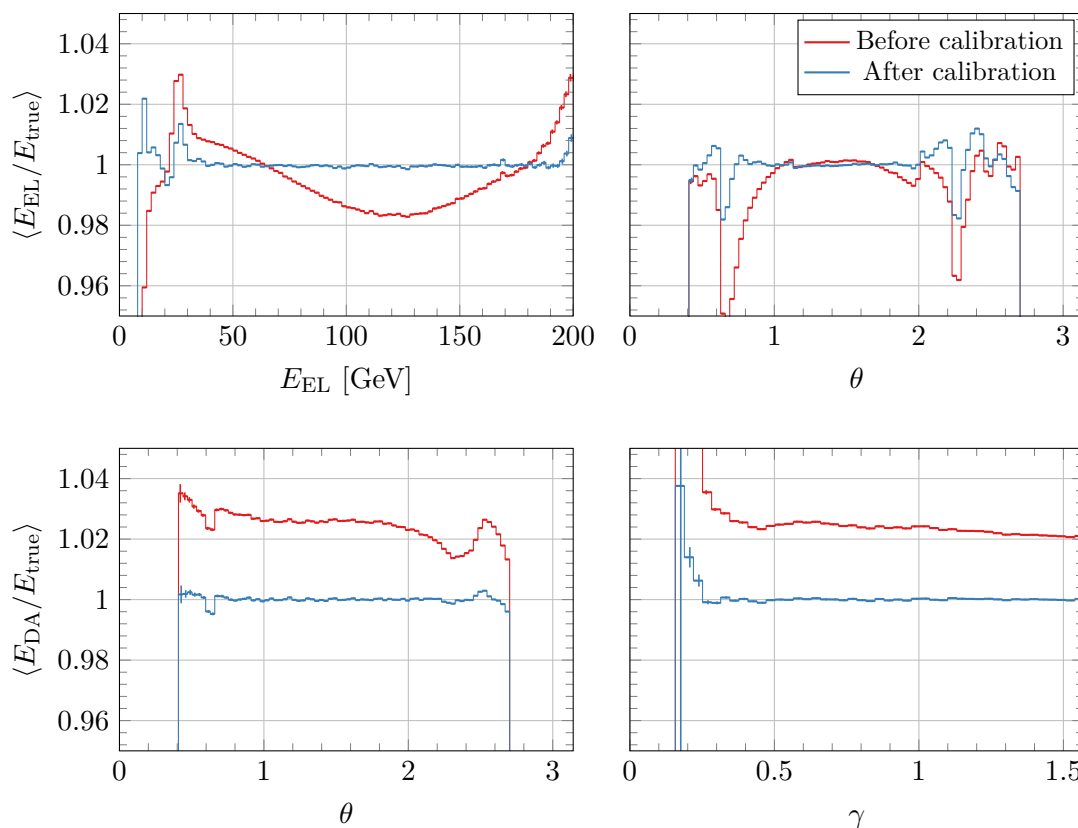### 8.2.4 Electron-energy calibration

The electron calibration is a correction to the event reconstruction and thus applies both to the data and the detector simulation. The DIS electron plays a significant role in this analysis. Therefore, the reconstructed detector-level electron must match the generator-level electron as closely as possible. Due to detector effects, such as the energy-scale calibration or the finite cell size of the calorimeter, the reconstructed electron energy has a systematic bias that can be corrected.

In this analysis, the electron-energy scale is corrected by comparing the DIS electron at detector and generator level in the MC. From this comparison, the average ratio of generator-level to detector-level energy is determined and then applied to the energy of the detector-level electrons in both data and MC. Since this is a correction of a detector effect, it is expected to have the same size in both MC samples. To minimise fluctuations, the Ariadne and Lepto samples are combined to derive the correction factors. The correction is applied independently

to the directly measured electron energy E_{EL} and the energy reconstructed from the double-angle method E_{DA}. The directly measured energy E_{EL} is used in the inclusive DIS selection, described in section 7.2. It is corrected as a two-dimensional function of the quantities corresponding to the electron method of reconstruction, i.e. E_{EL} and θ. The double-angle energy E_{DA} is used to construct the Breit frame and for the relative jet calibration, which is discussed in section 8.3.1. It is corrected as a function of the input quantities of the double-angle method, i.e. θ and γ.

For the determination of this correction, the event selection is modified. The cut $E_{\mathrm{EL}} > 10\,\mathrm{GeV}$ is replaced by a cut on the generator-level electron energy $E_{\mathrm{true}} > 10\,\mathrm{GeV}$. Similarly, the cut on the longitudinal energy $E - p_{\mathrm{z}}$ is replaced by a cut on the corresponding generator-level quantity. These changes help to include events in the selection that are otherwise rejected due to an incorrect measurement of the electron energy. Additionally, a cut on angular distance between the reconstructed and generated electron is introduced, $\sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.1$, where $\Delta\phi$ and $\Delta\eta$ and are the differences between the detector level and generator level electrons. This cut rejects events in which a different particle is misidentified as the DIS electron.

The average ratio of the reconstructed to the generated electron energy is shown in figure 8.6. Before the correction, the bias of the energy measurement was about 0.3% for the electron method and 2.1% for the double-angle method. After the correction, the bias is less than 0.1% for both methods.

Figure 8.6: Average ratio of reconstructed to generated electron energy for electron method and double-angle method as a function of the input variables to each method before and after calibration.

8.3 Jet corrections

Jet corrections are applied to the inclusive jet samples in data and MC. The inclusive jet samples are reconstructed and selected based on the corrected inclusive DIS sample. I.e. when applying the following corrections, all selection cuts from the previous chapter and all corrections described in this chapter so far are already applied.

This analysis uses three jet corrections. The jet-energy scale describes the relationship between the truth-level energy of a jet and the reconstructed energy of the same jet. Before corrections, the reconstructed energy might be biased compared to the true energy, and this bias might not be described by the MC.

The relative jet-energy scale correction is a correction to the detector simulation. It corrects for the fact that the simulated detector response is not necessarily the same as the calibration of the real detector, i.e. the bias in jet-energy reconstruction might be different in data and MC. After applying this relative jet-scale correction, the jet-energy scale in the MC will be correct relative to the data, i.e. the bias in jet energy reconstruction will be the same in data and MC.

Removing this bias in the reconstruction is the purpose of the absolute jet-energy-scale correction. This is a correction to the jet reconstruction and applies both to data and MC. This correction is similar to the electron calibration discussed in section 8.2.4. After applying the absolute jet-scale correction, the detector-level jet energy is, on average, equal to the corresponding hadron-level energy.

Finally, the jet reweighting is a correction to the MC event generators. It changes the jet distributions in the MC to match those in the data.

8.3.1 Relative jet-energy scale

The response of the detector to hadronic objects is not necessarily correctly described by the detector simulation. The purpose of the relative jet-energy-scale correction is to improve this description of the detector response, such that the MC can be used to derive the absolute jet correction, as described in the next section.

Due to conservation of momentum, the vectorial sum of the transverse momenta of all jets and the DIS electron vanishes. For events containing just one jet, consider the ratio

$$\frac{p_{\perp,\mathrm{lab,jet}}}{p_{\perp,\mathrm{DA}}},$$

where $p_{\perp,\mathrm{lab,jet}}$ is the transverse momentum of the single jet and $p_{\perp,\mathrm{DA}}$ is the transverse momentum of the double-angle electron. At parton level and in the absence of final-state radiation, momentum conservation demands that this ratio equals one. At detector level, the ratio can deviate from one due to hadronisation effects, large-angle radiation and since the detector has a different response to electrons and jets. Thanks to the uranium calorimeter employed at the ZEUS detector, this difference is minimised but not negligible.

The deviation of this ratio alone could be used to derive a relative and absolute jet-energy-scale correction. For this analysis, a different approach is chosen.[80] Rather than requiring this ratio to be one, which would only be reasonable in the absence of hadronisation and radiative effects, this ratio is required to be equal in data and MC. This approach is also valid in the presence of the mentioned effects, as long as they are described by the MC.

For the relative jet-scale correction, this ratio is determined at detector level in data and MC. To apply the correction, the energy of each jet in the MC is rescaled by the double
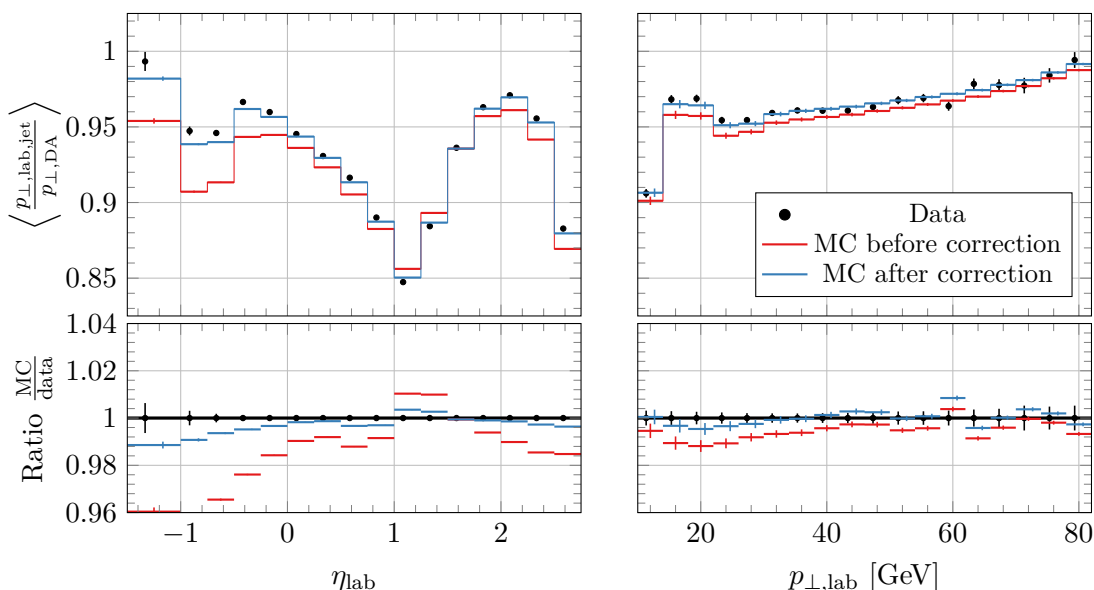
Figure 8.7: Average of the ratio $\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}$ as a function of η_{lab} and $p_{\perp,\text{lab}}$ before and after the correction. A double differential distribution of $\left\langle \frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}} \right\rangle$ as a function of η_{lab} and $p_{\perp,\text{lab}}$ is given in figure B.4. After the correction, the jet-energy scale agrees in data and MC to about 1%.

ratio $\left(\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}\right)_{\text{data}} \Big/ \left(\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}\right)_{\text{MC}}$, such that after the calibration $\left(\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}\right)_{\text{data}} = \left(\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}\right)_{\text{MC}}$. Afterwards, a dedicated absolute energy-scale correction will be applied.

As the detector response is expected to depend on the location of the jets within the detector, correction factors are determined as a function of η_{lab} of each jet. For the determination of the correction factors, the event selection is modified. The lower cut on the inelasticity $y_{\text{DA}} > 0.2$ is removed to increase the number of jets in the forward direction, i.e. with large η_{lab}. Figure 7.3 shows that the data is well described by the MC also in this region.

Since the determination of this correction uses single jet events, the inclusive jet selection discussed in section 7.3 is not used. Instead, jets are reconstructed in the laboratory frame, as this is a correction to the detector geometry. An event is classified as a single jet event if the leading jet has a transverse momentum in the laboratory of at least 10 GeV. All other jets are required to have a transverse momentum of no more than 5 GeV. Additionally, the vectorial sum of the transverse momenta of all remaining jets is required to be less than 5 GeV.

For this single-jet sample, the average of the ratio $\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}$ is determined in bins of η_{lab}. Since this is a correction to the detector simulation and the detector simulation is identical for both MC chains, the Ariadne and Lepto samples are combined to derive common correction factors. The distribution used to derive these factors is shown in figure 8.7. To apply the correction, the double ratio, as shown in the lower left panel, is used to scale the jet energies in the MC. A linear interpolation between the two adjacent bin centres is performed to reduce binning effects. The corrected distribution is also shown in the figure.

8.3.2 Absolute jet-energy scale

Jets can lose energy before being measured as they pass through inactive material in the detector. Therefore, the reconstructed jet energy deviates systematically from the true energy
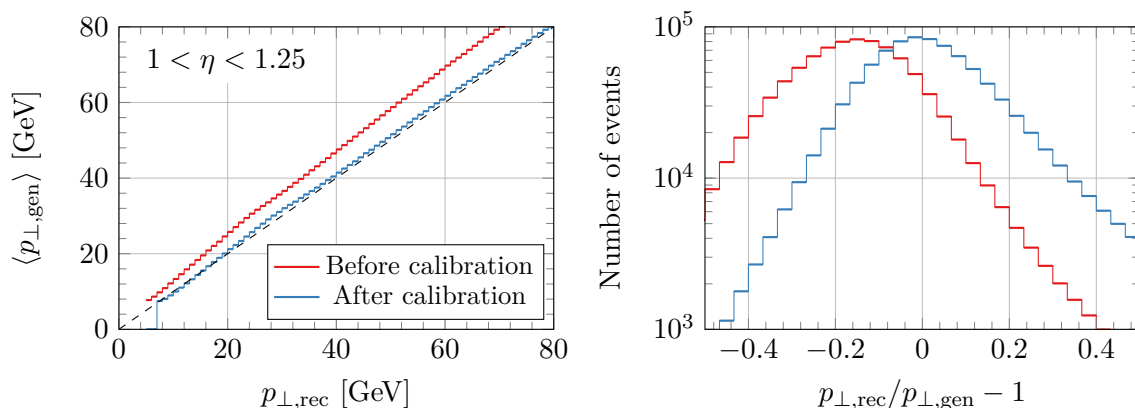
Figure 8.8: The left panel shows the average hadron-level jet energy as a function of the detector-level energy in one bin of η_{lab} before and after the calibration. The dashed line shows the diagonal $\langle p_{\perp,\mathrm{gen}} \rangle = p_{\perp,\mathrm{rec}}$. The right panel depicts the bias and resolution of the jet-energy reconstruction.

of each jet. After the relative jet-energy-scale calibration, this bias is similar in data and MC. The purpose of the absolute jet-energy-scale correction is to compensate for this bias of the energy measurement. The corresponding correction factors are derived by comparing MC hadron level and MC detector level. Therefore, the relative jet-energy-scale calibration must be applied beforehand, such that the detector simulation is representative of the data.[80]

For the absolute jet-scale correction, it is necessary to reconstruct both detector-level and hadron-level jets. Jets at the two levels are then matched to each other based on their distance in the η-ϕ-plane. Using a purely geometric matching condition rather than one based on the energy helps avoid biases in the calibration. Since this is a correction for a detector effect, the entire procedure is performed in the laboratory frame. For the same reason, the Ariadne and Lepto MC chains are combined to determine common correction factors.

A modified jet selection is used for the calibration. At both levels, jets are constructed in the laboratory frame and are required to fulfil $p_{\perp,\mathrm{lab}} > 5\,\mathrm{GeV}$. This requirement rejects many low-energy jet candidates, especially at detector level, that could otherwise get mismatched to high-energy hadron-level jets. Jets are also required to belong to the region $-1.25 < \eta_{\mathrm{lab}} < 2.75$. This range is slightly larger than the one used in the analysis to avoid cut-off effects at the edges of the analysis phase space.

After reconstruction, detector-level and hadron-level jets are matched to each other, according to the following algorithm. Between each pair of detector- and hadron-level jets, the angular distance $\Delta R = \sqrt{(\phi_{\mathrm{rec}} - \phi_{\mathrm{gen}})^2 + (\eta_{\mathrm{rec}} - \eta_{\mathrm{gen}})^2}$ is computed in the laboratory frame. The pair for which this distance is the smallest is considered a match if $\Delta R < 0.7$. The matching repeats until no more suitable pairs are found. All remaining jets at both levels are considered unmatched and ignored. Using this procedure, more than 99% of the detector-level jets with a transverse momentum of more than $10\,\mathrm{GeV}$ are matched.

The average hadron-level jet energy is determined as a function of the detector-level energy. This correction is performed in several bins of η_{lab} to account for differences in the detector geometry. The result in one such bin is depicted in the left panel of figure 8.8. The remaining bins are shown in figure B.5. It can be seen that the relation is in very good approximation a linear function. Therefore, it is parameterised as $\langle p_{\perp,\mathrm{gen}} \rangle = a + b \cdot p_{\perp,\mathrm{rec}}$, where the parameters a and b are determined from a fit to the described distributions, separately in each η_{lab} bin.

To apply the calibration, the four-momentum of each detector-level jet candidate is rescaled by a factor of $(a + b \cdot p_{\perp,\mathrm{rec,lab}})/p_{\perp,\mathrm{rec,lab}}$, where a and b are chosen according to the η_{lab} of the jet. The correction tends to increase the energy of the jet candidates, especially at low energies. To prevent very-low-energy jet candidates from becoming jets in the analysis, candidates with $p_{\perp,\mathrm{rec,lab}} < 3\,\mathrm{GeV}$ are not corrected. The influence of the η_{lab}-binning is reduced by linearly interpolating both a and b between the closest bin centres. Since the four-momenta in the Breit frame are related to the four-momenta in the laboratory frame, this correction can be propagated to the Breit frame by applying the same factor there. The correction notably reduces the bias of the jet-energy measurement, as can be seen in the right panel of figure 8.8. This bias is reduced from more than 10% to less than 1.5%. The resolution does not change significantly.

8.3.3 Jet reweighting

As seen in section 7.3, the leading-order MC predictions fail to describe the jet distributions in the data accurately. To use the MC for the acceptance correction, it is, therefore, necessary to reweight the MC distributions. After applying the relative and absolute jet-energy-scale corrections, detector-level jets correspond on average to hadron-level jets of the same energy. This makes it possible to derive correction factors for the MC hadron-level distributions by comparing the detector-level distributions of data and MC. The reconstruction of MC hadron-level quantities is described in section 7.4. Since this is a correction to the MC event generators, it is derived and applied separately to the Ariadne and Lepto MC chains.

In this analysis, two approaches to jet reweighting are considered. One is based on reweighting events based on their jet multiplicity, and the other is based on reweighting each jet individually. The multijet reweighting procedure is used as the nominal correction in this analysis. The inclusive jet reweighting is used to estimate the systematic uncertainty of this choice, see section 10.3.

Multijet reweighting

The idea of multijet reweighting is to sort jet events into three categories: single-jet events, dijet events and events with three or more jets. Each event category is then reweighted separately as a function of $p_{\perp,\mathrm{Breit}}$.

The definition of the single-jet, dijet and trijet categories does not need to agree with any definition previously used. For this correction, the following definitions are used. Only jets in the region $-1 < \eta_{\mathrm{lab}} < 2.5$ and $7\,\mathrm{GeV} < p_{\perp,\mathrm{Breit}} < 50\,\mathrm{GeV}$ are considered. At detector level, additionally $p_{\perp,\mathrm{lab}} > 3\,\mathrm{GeV}$ is required. If three or more such jets exist, and the invariant mass of the three leading jets in the Breit frame satisfies $M_{\mathrm{jjj}} > 40\,\mathrm{GeV}$, the event is classified as a trijet event. If that is not the case, but exactly two jets exist, and their invariant mass satisfies $M_{\mathrm{jj}} > 20\,\mathrm{GeV}$, the event is classified as a dijet event. If this is also not the case, but there is at least one jet, the event is classified as a single-jet event. For the single-jet category, the requirement on the transverse momentum in the Breit frame is relaxed to $p_{\perp,\mathrm{Breit}} > 4\,\mathrm{GeV}$. If all of these tests fail, the event is not classified as a jet event and is ignored for this correction.

For each event category, a histogram is created with bins in $p_{\perp,\mathrm{Breit}}$. For the dijet and trijet categories, the average transverse momentum of the jets, $\overline{p_{\perp,\mathrm{Breit}}}$, is used. A second-degree polynomial is then fitted to the ratio of this histogram in data and each MC chain. To apply the reweighting, jets are reconstructed at hadron level, such that hadron-level jet quantities can be determined. Events are classified into single-jet, dijet or trijet events using the same definition
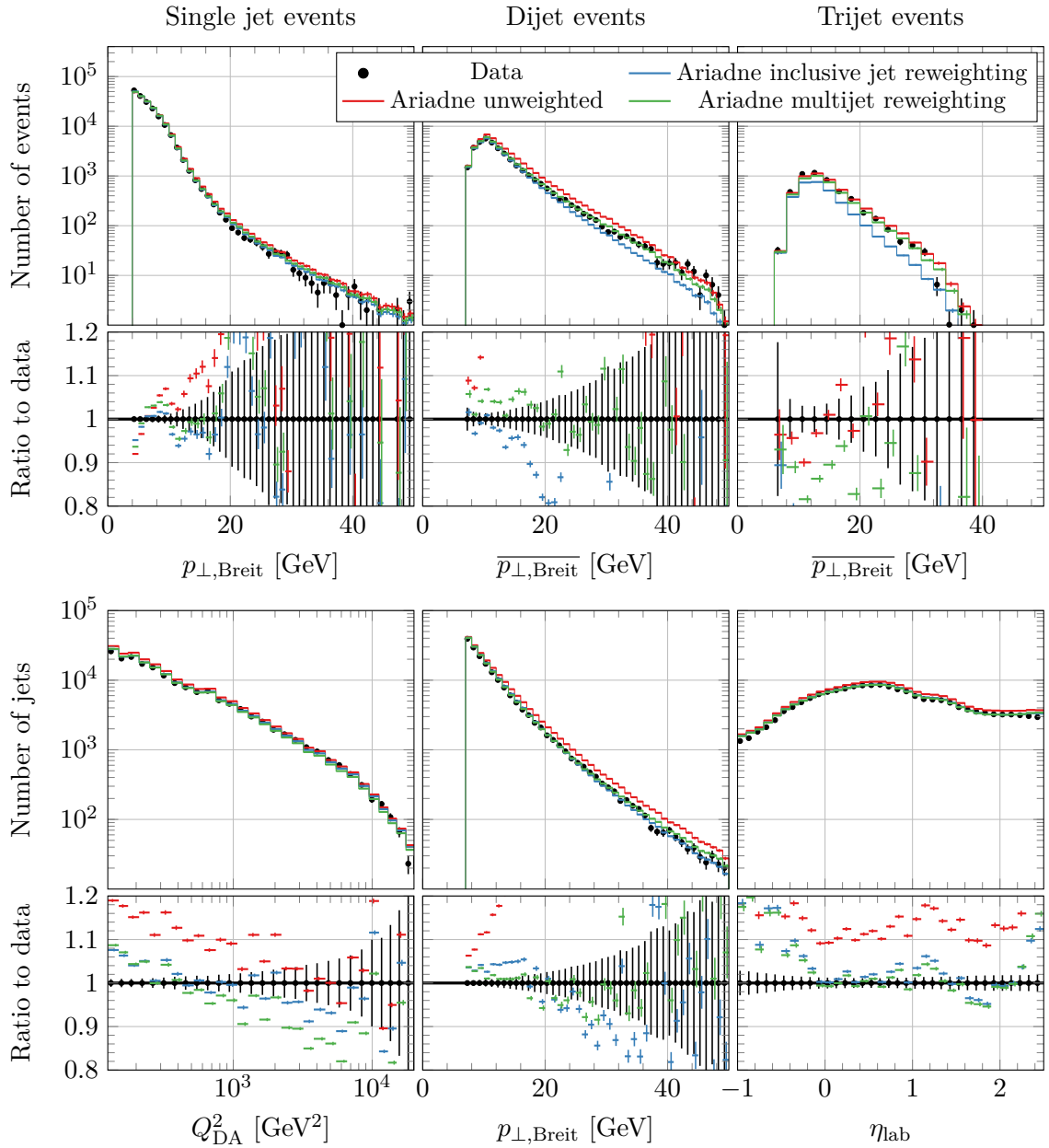
Figure 8.9: Multijet (upper row) and inclusive jet (lower row) distributions before and after multijet and inclusive jet reweighting of the Ariadne MC chain. Both methods show similar performance in improving the agreement between data and MC. The equivalent figure for the Lepto MC chain is given in figure B.6.

used at the detector level but using the corresponding hadron-level quantities. Afterwards, the fitted function of the appropriate category is used to reweight the MC event at hadron level.

The jet distributions before and after the multijet reweighting are shown in figure 8.9 for the Ariadne MC chain. The equivalent figure for the Lepto chain is given in figure B.6 in the appendix.

Inclusive jet reweighting

For the inclusive jet reweighting, each jet is reweighted individually. This means, in addition to the event weights, each jet also has an additional weight associated with it. For this approach to jet reweighting, jets must be matched at detector and hadron level to propagate the adjusted weights to the detector level.

To derive the correction, a two-dimensional histogram with bins in Q^2_{DA} and $p_{\perp,\mathrm{Breit}}$ is filled using the inclusive jet sample. A second-degree polynomial of $\log(Q^2_{\mathrm{DA}})$ and $p_{\perp,\mathrm{Breit}}$ is fitted to the ratio of this histogram in data and MC. When applying the reweighting, jets are reconstructed at hadron level, and each jet is reweighted individually according to the fitted function.

After that, generator- and detector-level jets are matched to each other in the laboratory frame using the same algorithm described in section 8.3.2, but with a modified cut of $\Delta R < 1$. The matching is performed before applying jet cuts at both levels to account for migrations. For each matched pair of jets, the weight of the hadron-level jet is copied to the detector-level jet. Unmatched hadron-level jets can be considered inefficiencies of the detector and are irrelevant for the reweighting. Unmatched detector-level jets are more problematic, as it is unclear which weight should be assigned to them. About 30% of the unmatched jets arise from jet splitting, i.e. a single hadron-level jet is reconstructed as multiple detector-level jets. Since the jet clustering is performed in the Breit frame, the jets may be very close together in the laboratory frame. To account for this effect, all remaining detector-level jets are matched to the closest hadron-level jet if they are closer to them than $\Delta R = 1$ in the laboratory frame. As this matching is only used to copy the weights, a one-to-one correspondence is not necessary. After applying this jet matching, less than 3% of detector-level jets remain unweighted. The weights of these remaining jets are not changed.

The distributions before and after the inclusive jet reweighting are shown in figure 8.9. It can be seen that both methods of jet reweighting give similar results and that the difference between the two methods is of similar size to the remaining disagreement between data and MC. In this analysis, multijet reweighting is used as the nominal correction method. Inclusive jet reweighting is used as a variation to estimate the remaining systematic uncertainty of this correction.

CHAPTER 9

Determination of cross sections

Detector-level distributions depend on the details of the accelerator, the detector, the event selection and the applied corrections. Therefore, they cannot be straightforwardly compared to theory predictions, measurements from different experiments or even different measurements from the same experiment. Interaction cross sections at hadron level are a detector independent observable that can be derived by correcting the detector-level distributions for all experiment-specific details.

9.1 The corrected data sample

The determination of cross sections uses the MC samples. For this approach to be valid, the MC samples must describe the data well in all variables used in the analysis.

The signal MC alone does not describe the data completely since contributions from other physical processes are also present in the data sample. The selection criteria introduced in chapter 7 aim to remove these contributions as much as possible. However, a small number of background events is expected to remain in the final data sample and needs to be accounted for when deriving hadron-level cross sections. In this analysis, the two dominant sources of background that remain after the inclusive jet selection are from low-Q^2 DIS and PHP events, see section 7.1. For both of these processes, the Feynman graph can be identical to the high-Q^2 NC DIS signal process, which leads to a similar signature in the detector.

Distributions of the inclusive DIS sample, including all DIS and jet corrections described in the previous chapter, are given in figures B.7, B.8 and B.9. Similar to the figures in chapter 7, these figures also show the contributions from the two background sources and the combined MC sample.

The contribution of low-Q^2 DIS is a few per cent at the lower end of the considered Q^2 range and drops off rapidly with increasing Q^2. The contribution from PHP is usually below 1%. Outside the selected phase space, the background contributions sometimes increase to more than 10%. The combined MC sample describes the shape of the data well within the selected phase space region.

Distributions of the jet veto cuts and the inclusive jet cuts are shown in figure B.10, including all corrections. Compared to uncorrected distributions from figures 7.11 and 7.12, the agreement between data and MC is significantly improved in both shape and normalisation.

Distributions for the inclusive DIS and inclusive jet selection at hadron level are given in figure B.11. Compared to figures 7.13 and 7.14, the agreement between the two samples in the DIS quantities is degraded because the jet reweighting affects the two samples differently. The

agreement in the jet quantities is significantly improved regarding normalisation and shape of the $p_{\perp,\text{Breit}}$ distribution.

9.2 Bin definition

The choice of the size of the bins of a cross section measurement needs to ensure that each bin contains sufficient data statistics and that bins are not smaller than the resolution of the considered variables. If a bin does not contain enough data events, its statistical uncertainty will be large, and no reliable cross section can be extracted. If the size of a bin is smaller than the detector resolution of a specific variable, the measured cross section will be strongly biased by the surrounding bins, which again prevents a precise measurement of the bin under consideration.

In this analysis, the definition of the bins is taken from the corresponding measurement of the H1 collaboration.[121] While the ZEUS and H1 detectors have somewhat different resolutions, this choice is also acceptable for the ZEUS detector, as will be shown in the next section. The kinematic range of DIS events is $150\,\text{GeV}^2 < Q^2 < 15000\,\text{GeV}^2$ and $0.2 < y < 0.7$. Jets are considered in the region $7\,\text{GeV} < p_{\perp,\text{Breit}} < 50\,\text{GeV}$ and $-1 < \eta_{\text{lab}} < 2.5$. The measurement is performed double-differentially in Q^2 and $p_{\perp,\text{Breit}}$ with the exact bin boundaries listed in table 9.1. This choice allows a direct comparison of the measured cross sections between the two experiments.

Label	Q^2 [GeV^2]
1	150 – 200
2	200 – 270
3	270 – 400
4	400 – 700
5	700 – 5000
6	5000 – 15000

Label	$p_{\perp,\text{Breit}}$ [GeV]
a	7 – 11
b	11 – 18
c	18 – 30
d	30 – 50

Table 9.1: Definition of the Q^2 and $p_{\perp,\text{Breit}}$ bins used for the measurement. See section 6.5 for the hadron-level definitions of Q^2 and $p_{\perp,\text{Breit}}$.

9.3 Bin-by-bin correction

A particle detector can never be perfect. The energies and positions of particles can get distorted by interactions with inactive detector material. Particles might not be measured at all because their energy is too low or they escape through gaps in the detector, such as the beam pipes. A particle can get misidentified during reconstruction because its measurement is ambiguous. Due to these reasons, corresponding detector and hadron-level distributions tend to differ significantly in shape and normalisation.

Mathematically, this process is called 'folding'. The detector-level distributions can be described by a convolution of the distributions at hadron level with the detector response function. To obtain results at hadron level, which can be compared to other measurements or theoretical predictions, this operation has to be reversed in a process called 'unfolding'.

Different approaches to unfolding exist, such as iterative Bayesian unfolding or matrix unfolding with regularisation.[193,194] The simplest type of unfolding is called the bin-by-bin

correction and is widely used at HERA.[71,133,179] This method works by determining the acceptance of each bin from the MC sample according to

$$\text{Acceptance} = \frac{N_{\text{rec}}}{N_{\text{gen}}}, \tag{9.1a}$$

where N_{rec} is the number of jets found in the bin at detector level and N_{gen} is the number of jets found in the corresponding bin at hadron level in the MC sample. Here, it is assumed that there is a conceptual one-to-one correspondence between bins at detector and hadron level. The acceptance relates the number of generated and reconstructed jets and thereby serves as a simplified model of the detector. The number of jets at truth level in the data is then determined by dividing the number of jets in each detector-level bin in the data by the acceptance.

9.3.1 Bin quality

The bin-by-bin correction is only applicable if the number of jets that migrate between bins and across the phase space boundaries is sufficiently small and well described by the MC. Whether this is the case can be determined by considering the efficiency and purity of each bin. These quantities are defined as

$$\text{Efficiency} = \frac{N_{\text{rec\&gen}}}{N_{\text{gen}}}, \tag{9.1b}$$

$$\text{Purity} = \frac{N_{\text{rec\&gen}}}{N_{\text{rec}}}, \tag{9.1c}$$

where $N_{\text{rec\&gen}}$ is the number of jets that belong to the same bin at detector and hadron level. Thus, the efficiency is the fraction of the hadron-level jets that are correctly reconstructed in the same bin at detector level. The purity is the fraction of detector-level jets that were also generated in the same bin.

For single-differential distributions at HERA, efficiency and purity are usually required to be above 40%.[80] Figure 9.1 shows these quantities for the Q^2 and $p_{\perp,\text{Breit}}$ bins. The purity is typically in the range of $50\% - 60\%$, indicating that a large fraction of the reconstructed jets is correctly measured. The efficiency is smaller, especially in the low-Q^2 and low-$p_{\perp,\text{Breit}}$ bins, where it is almost as low as 30%. This indicates that many hadron-level jets are either not reconstructed at all or not reconstructed in the correct bin.

To understand this effect, a study on the efficiency is performed. This study starts by counting the number of hadron-level jets, N_{gen}, in each bin of the analysis phase space. Then, one by one, additional requirements are imposed at detector level up to the full detector-level selection. At each step new counts $N_{\text{gen\&partial rec}}$ are determined, and the partial efficiency $N_{\text{gen\&partial rec}}/N_{\text{gen}}$ is computed. Comparing the partial efficiency at the different steps allows determining which part of the selection is responsible for the low efficiency.

The intermediate steps considered are:

1. A DIS electron candidate above the score threshold is required to exist. At this stage, no additional requirements, e.g. on the energy or isolation, are imposed, except for those already included in the electron finder;

2. The event has to belong to the analysis phase space in y_{DA} at detector level;
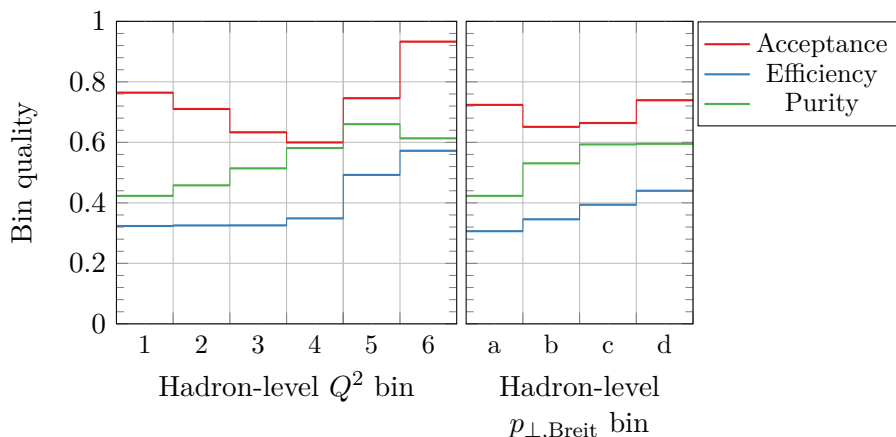
Figure 9.1: Acceptance, efficiency and purity for the single differential distributions in Q^2 and $p_{\perp,\mathrm{Breit}}$. The bins are labelled according to table 9.1.

3. The event has to belong to the desired phase space in Q^2_{DA} at detector level. For the efficiency as a function of Q^2, this means the event has to belong to the same Q^2 bin as at hadron level. As a function of $p_{\perp,\mathrm{Breit}}$, this only means that the event belongs to the analysis phase space in Q^2 because migrations in Q^2 within the phase space should not affect the efficiency in $p_{\perp,\mathrm{Breit}}$;

4. The event is required to pass all DIS cuts, from section 7.2;

5. The event is required to pass the jet veto cuts, and each hadron-level jet must be matched to a detector-level jet. Detector-level jets are required to belong to the analysis phase space in $p_{\perp,\mathrm{Breit}}$ and in η_{lab};

6. The matched detector-level jet has to belong to the same $p_{\perp,\mathrm{Breit}}$ bin as the hadron-level jet. Similar to the third step, this requirement does not apply to the efficiency as a function of Q^2, as that should not be affected by migrations in $p_{\perp,\mathrm{Breit}}$ within the phase space.

At this point, the full detector-level selection is applied, and the partial efficiency determined from the last step is equal to the regular efficiency discussed above. The results for the partial efficiency are shown in figure 9.2.

A detector-level DIS electron candidate exists for almost all hadron-level jets. When requiring that the event belongs to the analysis phase space in y, a notable drop in the partial efficiency is observed of almost 30% in some bins. This indicates large migrations in y out of the phase space. Additionally requiring that the event belongs to the correct region in Q^2 causes another significant drop in the low-Q^2 region, indicating a large migration between bins in that region. At high Q^2, the detector resolution improves, and the drop in efficiency is much smaller. As a function of $p_{\perp,\mathrm{Breit}}$, the drop due to migrations in Q^2 is much smaller since migrations within the phase space do not contribute. In the low-Q^2 region, where the efficiency is especially low, migrations in Q^2 and y are the dominant sources of inefficiency, contributing about 50% to the overall efficiency. As a function of $p_{\perp,\mathrm{Breit}}$, the efficiency drops by about 15% when requiring DIS cuts to be passed. The fact that this drop is smaller when considering the efficiency as a function of Q^2 indicates that the DIS cuts especially remove events in which Q^2 is poorly reconstructed. Especially in the lowest $p_{\perp,\mathrm{Breit}}$ bin, there is another significant drop when
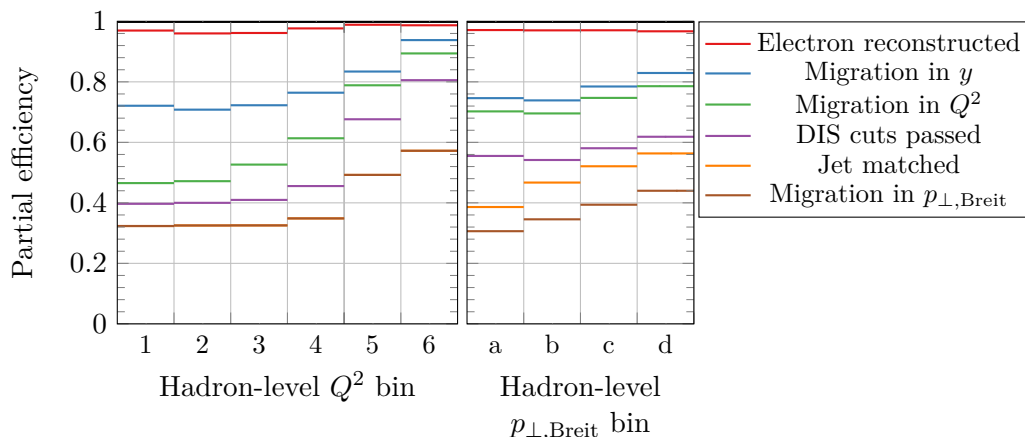
Figure 9.2: Partial efficiency for the single differential distributions in Q^2 and $p_{\perp,\mathrm{Breit}}$ for different stages of the detector-level reconstruction. The bins are labelled according to table 9.1.

requiring each hadron-level jet to be matched to a detector-level jet. This is due to the worse reconstruction efficiency of low-$p_{\perp,\mathrm{Breit}}$ jets.

In summary, as a function of Q^2 and $p_{\perp,\mathrm{Breit}}$, the efficiency of the DIS selection is about 40% and 60% and for the jet selection about 80% and 50%, respectively. It has been extensively checked in previous analyses that the MC describes the inclusive DIS data well, such that the low efficiency of this part of the event selection is acceptable.[80,189,195] The efficiency of the jet selection is sufficiently high to justify using the bin-by-bin correction.

A similar study is performed on the purity and is shown in figure B.12. Also, here, migrations in Q^2 and y are found to be a significant contribution. The corresponding double-differential figures are given in appendix B.4. These scenarios have lower efficiency and purity, as migrations can occur along two axes. However, they are not significantly lower since many sources of inefficiency and impurity are shared between the Q^2- and $p_{\perp,\mathrm{Breit}}$-distributions.

9.3.2 Cross section determination

Rearranging equation (9.1a) and taking background contributions into account yields an equation for the corrected number of jets in a given bin

$$N_{\mathrm{data,truth}} = \frac{N_{\mathrm{data}} - N_{\mathrm{background}}}{\mathrm{Acceptance}}, \tag{9.2}$$

where N_{data} is the number of jets in the data sample and $N_{\mathrm{background}}$ is the number of jets in the background MC samples at detector level. Section 9.4 gives more details on the background contribution.

When propagating the statistical uncertainty of the signal MC to the cross sections, the correlations of N_{rec} and N_{gen} in equation (9.1a) need to be taken into account. In previous ZEUS analyses, this was often done by instead expressing the acceptance in terms of the uncorrelated quantities $N_{\mathrm{rec\&gen}}$, $(N_{\mathrm{rec}} - N_{\mathrm{rec\&gen}})$ and $(N_{\mathrm{gen}} - N_{\mathrm{rec\&gen}})$, in which case simple error propagation can be used.[72,80] In this analysis, a different but equivalent approach is used. For each bin, the covariance of N_{rec} and N_{gen} is computed as $\sigma_{\mathrm{rec,gen}} = \sum w_{\mathrm{rec}} w_{\mathrm{gen}}$, where
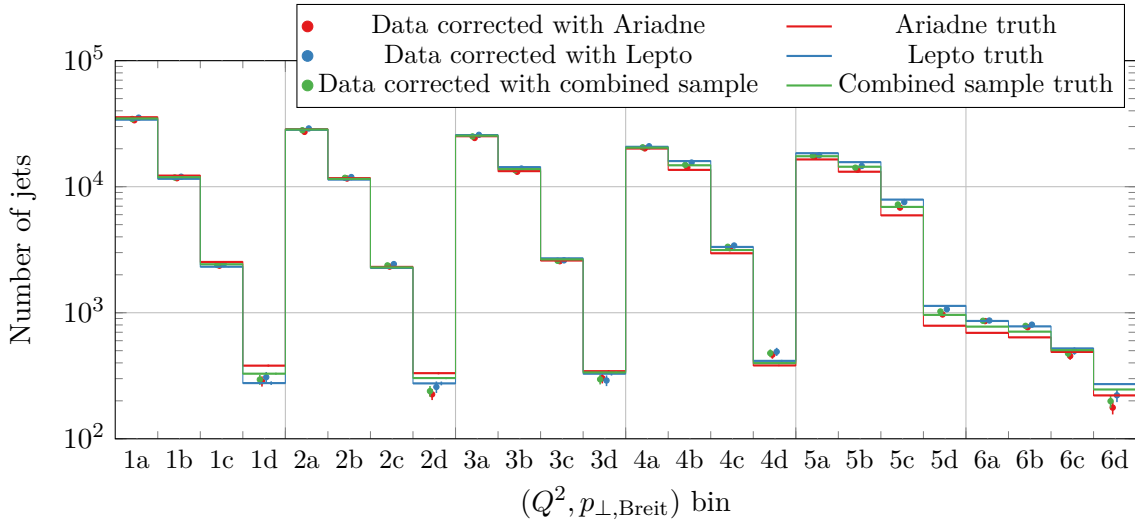
Figure 9.3: Bin-by-bin corrected jet distribution. The data are shown, bin-by-bin corrected with the Ariadne, Lepto or the combined MC sample. The corresponding MC hadron-level distributions are also shown. The bins are defined in table 9.1. There are six groups of four bins each. Each group represents one Q^2 bin. The individual bins within each group represent the $p_{\perp,\text{Breit}}$ bins.

the sum runs over matched pairs of detector-level and hadron-level jets with weights w_{rec} and w_{gen} respectively. The relative uncertainty of the acceptance is then given as

$$\delta_{\text{Acceptance}} = \sqrt{\delta_{\text{rec}}^2 + \delta_{\text{gen}}^2 - 2\delta_{\text{rec}}\delta_{\text{gen}}\frac{\sigma_{\text{rec,gen}}}{\delta_{\text{rec}}N_{\text{rec}}\delta_{\text{gen}}N_{\text{gen}}}}, \qquad (9.3)$$

with $\delta_{\text{rec/gen}}$ being the relative statistical uncertainty of the jet counts in the detector-level and hadron-level MC bins, respectively. In the low-Q^2 region, where data statistics is large, the statistical uncertainty from the MC sample makes up about 30% of the total statistical uncertainty. At high Q^2, this number decreases to less than 10%.

Ideally, the corrected cross sections do not depend on the MC model used to derive them. In practice, however, even after applying all corrections from the previous chapter, the cross sections still depend on the MC model to some extent. Since the MC models are not entirely correct, this will bias the result. To reduce this bias, the Ariadne and Lepto signal MC samples are combined to perform this correction. This reduces the dependence on each model. A systematic uncertainty will be introduced in section 10.2 to estimate the remaining bias.

The corrected number of jets is shown in figure 9.3. It can be seen that the MC samples still differ in some bins. The bin-by-bin corrected data points depend slightly on the MC model used. Their difference, however, is much smaller than the difference between the MC models. This indicates that the acceptance does not significantly depend on the MC hadron-level distribution, which is desirable for the bin-by-bin correction. Cross sections can be derived from the corrected number of jets using equation (1.1). Cross sections are discussed, including systematic uncertainties in chapter 11.

9.3.3 Correlations

When determining cross sections and using them in further analyses, such as the QCD fit described in chapter 12, it is important to treat correlations consistently. In this analysis, two

types of statistical correlation need to be considered when using the bin-by-bin correction method. Firstly, in an inclusive jet analysis, events may contain multiple jets and thus contribute multiple times to the measurement. Therefore, the statistical uncertainty of the different bins of the inclusive jet measurement will be correlated. Furthermore, the statistical uncertainty of each individual bin will be increased.

Secondly, the cross sections determined in this analysis will be analysed together with a previous dijet measurement at ZEUS. Since dijet production is closely related to inclusive jet production in the Breit frame and the two analyses use similar data samples and jet selection criteria, there will be considerable overlap between the jets contributing to both analyses. Thus, the statistical uncertainties will also be correlated across the two measurements. For both of these cases, a statistical correlation matrix needs to be determined.

Poisson statistics

Before discussing correlations, it is useful to state the relevant statistical laws. Event counts in particle physics measurements are usually treated as Poisson distributed. This model is justified if

1. Events occur independently of each other, i.e. the occurrence of an event in one collision does not affect the probability of occurrence in the next collision.

2. The average rate at which events occur is constant.

3. Two events may not occur simultaneously. At each collision, an event either occurs or does not occur.

When N Poisson-distributed events are observed with weight w_i, the expectation value μ and variance σ^2 are given by

$$\mu = \sum_{i=1}^{N} w_i, \tag{9.4a}$$

$$\sigma^2 = \sum_{i=1}^{N} w_i^2. \tag{9.4b}$$

The square root of the variance gives the statistical uncertainty. If $w_i \equiv 1$, these expressions reduce to $\mu = \sigma^2 = N$. When considering two observables, A and B, that may occur simultaneously with weights $w_{A,i}$ and $w_{B,i}$, their covariance is given by

$$\sigma_{AB} = \sum_{i=1}^{N} w_{A,i} w_{B,i}. \tag{9.4c}$$

A useful related quantity is the correlation between two observables, which is defined as

$$\rho_{AB} = \frac{\sigma_{AB}}{\sqrt{\sigma_A^2 \sigma_B^2}}. \tag{9.4d}$$

When considering multiple observables, e.g. event counts in multiple bins, it is convenient to promote these quantities to vectors and matrices

$$\vec{\mu} = \sum_{i=1}^{N} \vec{w}_i, \tag{9.5a}$$

$$V = \sum_{i=1}^{N} \vec{w}_i \vec{w}_i^{\mathsf{T}}, \tag{9.5b}$$

$$\vec{\sigma} = \sum_{A=1}^{\dim(\vec{\mu})} \vec{e}_A \sqrt{V_{AA}}, \tag{9.5c}$$

$$C = \sum_{A,B=1}^{\dim(\vec{\mu})} \vec{e}_A \vec{e}_B^{\mathsf{T}} \frac{V_{AB}}{\sigma_A \sigma_B}, \tag{9.5d}$$

where $\vec{e}_A$ is the unit vector in direction A. Now, $\vec{\mu}$ are the expectation values of all the observables, V is the covariance matrix, $\vec{\sigma}$ are the statistical uncertainties, and C is the correlation matrix. The correlation matrix can be considered a version of the covariance matrix normalised to the statistical uncertainties. The outcome of a measurement is specified either by the expectation values and covariance matrix, $\vec{\mu}$ and V or by the expectation values, their statistical uncertainties and the correlation matrix, $\vec{\mu}$, $\vec{\sigma}$ and C.

Correlations within inclusive jet measurement

The assumptions of a Poisson distribution are fulfilled for any observable that is defined at the event level, e.g. inclusive DIS events, single-jet events and dijet events. For inclusive jets, however, the third part of the definition is not fulfilled since multiple jets can be produced in the same collision event. Poisson statistics may still be employed if, instead of counting inclusive jets, n-jet events are counted, as n-jet events are individually Poisson distributed for any n. Each event is assigned a weight of $w_i = n$ to account for the number of jets. This approach also allows a determination of the correlations between bins, as is demonstrated by the following two examples.

Consider m n-jet events with all jets belonging to the same bin, i.e. $w_i \equiv n$. The expectation value will be $\mu = m \cdot n$ with statistical uncertainty $\sigma = \sqrt{m \cdot n^2}$, according to Poisson statistics (9.4). Now, consider $\frac{m}{2}$ $2n$-jet events, with all jets belonging to the same bin, i.e. $w_i \equiv 2n$. The expectation value remains unchanged. But now the statistical uncertainty is $\sigma = \sqrt{\frac{m}{2} \cdot (2n)^2} = 2\sqrt{m \cdot n^2}$. This example illustrates that the statistical uncertainty increases when more jets originate from the same event.

As a second example, consider m_1 dijet events, with the jets belonging to two different bins, accompanied by m_2 dijet events in which both jets belong to the first of these two bins and m_3 dijet events in which both jets belong to the second bin, i.e. $\vec{w}_{1,i} \equiv \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$, $\vec{w}_{2,i} \equiv \left(\begin{smallmatrix}2\\0\end{smallmatrix}\right)$ and $\vec{w}_{3,i} \equiv \left(\begin{smallmatrix}0\\2\end{smallmatrix}\right)$. The expectation value and covariance matrix are $\vec{\mu} = \left(\begin{smallmatrix}m_1+2m_2\\m_1+2m_3\end{smallmatrix}\right)$ and $V = m_1 \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)^{\mathsf{T}} + m_2 \left(\begin{smallmatrix}2\\0\end{smallmatrix}\right)\left(\begin{smallmatrix}2\\0\end{smallmatrix}\right)^{\mathsf{T}} + m_3 \left(\begin{smallmatrix}0\\2\end{smallmatrix}\right)\left(\begin{smallmatrix}0\\2\end{smallmatrix}\right)^{\mathsf{T}} = \left(\begin{smallmatrix}m_1+4m_2 & m_1\\m_1 & m_1+4m_3\end{smallmatrix}\right)$. The statistical uncertainties are $\vec{\sigma} = \left(\begin{smallmatrix}\sqrt{m_1+4m_2}\\\sqrt{m_1+4m_3}\end{smallmatrix}\right)$. The covariance matrix contains off-diagonal entries, indicating that the two bins are correlated. The correlation matrix is $C = \left(\begin{smallmatrix}1 & \rho_{12}\\\rho_{12} & 1\end{smallmatrix}\right)$ with $\rho_{12} = \frac{m_1}{\sqrt{(m_1+4m_2)(m_1+4m_3)}}$. This expression agrees with the physical intuition that the two bins are uncorrelated if $m_1 \to 0$ and fully correlated if $m_2, m_3 \to 0$.
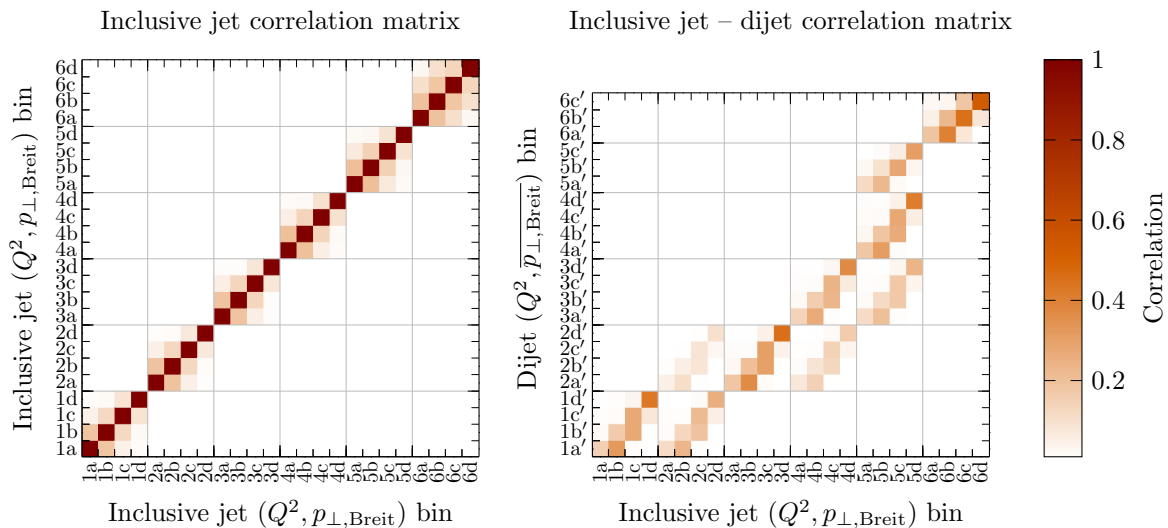
Figure 9.4: Inclusive jet correlation matrix (left) and inclusive jet – dijet correlation matrix (right) with bin-by-bin correction. The bins of the inclusive jet measurement are labelled according to table 9.2. The large bins indicate Q^2 bins, and the smaller ones are $p_{\perp,\text{Breit}}$ bins. The dijet bins are labelled similarly but using the bin definitions from that measurement.[71] A block structure arises because correlations can only occur between jets at the same Q^2. In the case of the inclusive jet – dijet correlation matrix, this structure is distorted due to the different bin boundaries of the two measurements. The numerical values of these matrices are given in table C.3.

Using equations (9.5) allows to determine the correct statistical uncertainties and correlations of inclusive jet distributions. This procedure is applied to the 24 inclusive jet bins at detector level, and the correlation matrix is derived. According to equation (9.4d), the bin-by-bin correction cancels in the correlation matrix. Therefore, the same matrix is also applicable to the hadron-level cross sections. The determined correlation matrix is shown in the left panel of figure 9.4. By definition, the matrix is symmetric, and entries along the diagonal are 100%. Only positive correlations can occur due to events producing multiple jets. The determined correlations between bins are small and do not exceed 20%.

Correlations to dijet measurement

A single event can produce inclusive jets and simultaneously be a dijet event. Since inclusive jets in the Breit frame and dijets are closely related, this will be the case for a significant fraction of events, such that correlations need to be taken into account also between the two measurements.

The dijet event selection is recreated.[80] Then, the correlation matrix between the two measurements is determined at detector level. This is done similarly to the previous sections by including both the inclusive jet bins and the dijet bins in the collection of observables $\vec{w}_i$, i.e. $\vec{w}_i$ will now contain the 24 inclusive jet bins and the 22 dijet bins. The corresponding correlation matrix will comprise three sub-matrices: the inclusive jet correlation matrix, the inclusive jet – dijet correlation matrix and the dijet correlation matrix. The latter will be a diagonal matrix, as dijet events are not statistically correlated with each other.

The run periods of the two measurements do not overlap completely. The impact of the overlap on the correlations can be quantified using equations (9.4). The set of all events is decomposed into the overlap region and the two regions that are only present in one of the analyses. For two observables A and B with partially overlapping run periods

$$
\begin{aligned}
\rho_{AB} &= \frac{\sigma_{AB}}{\sqrt{\sigma_A^2 \sigma_B^2}} \\[2ex]
&= \frac{\displaystyle\sum_{\text{all events}} w_{A,i} w_{B,i}}{\sqrt{\displaystyle\sum_{\text{all events}} w_{A,i}^2} \sqrt{\displaystyle\sum_{\text{all events}} w_{B,i}^2}} \\[2ex]
&= \frac{\displaystyle\sum_{\text{overlap}} w_i^{(A)} w_i^{(B)}}{\sqrt{\displaystyle\sum_{\text{overlap}} w_{A,i}^2 + \sum_{A\text{ only}} w_{A,i}^2} \sqrt{\displaystyle\sum_{\text{overlap}} w_{B,i}^2 + \sum_{B\text{ only}} w_{B,i}^2}} \\[2ex]
&= \frac{\displaystyle\sum_{\text{overlap}} w_{A,i} w_{B,i}}{\sqrt{\displaystyle\sum_{\text{overlap}} w_{A,i}^2} \sqrt{\displaystyle\sum_{\text{overlap}} w_{B,i}^2}} \sqrt{\frac{\displaystyle\sum_{\text{overlap}} w_{A,i}^2}{\displaystyle\sum_{\text{overlap}} w_{A,i}^2 + \sum_{A\text{ only}} w_{A,i}^2}} \sqrt{\frac{\displaystyle\sum_{\text{overlap}} w_{B,i}^2}{\displaystyle\sum_{\text{overlap}} w_{B,i}^2 + \sum_{B\text{ only}} w_{B,i}^2}} \\[2ex]
&= \rho_{AB,\text{overlap}} \sqrt{\frac{L_{\text{overlap}}}{L_A}} \sqrt{\frac{L_{\text{overlap}}}{L_B}}.
\end{aligned}
\tag{9.6}
$$

In the last step, it is used that the sum over a set of events is proportional to the integrated luminosity of those events. Since the expression in the sum is identical in the numerator and the denominator, the sums may be replaced by the luminosities. For the present case, this means that the correlations determined for the overlap of the two analyses should be scaled by $L_{\text{overlap}}/\sqrt{L_{\text{inclusive jet}} L_{\text{dijet}}} \approx 80\%$, where $L_{\text{inclusive jet}}$, L_{dijet} and L_{overlap} are the integrated luminosities of the data samples used in the inclusive jet measurement, the dijet measurement and the overlap of the two respectively. The inclusive jet – dijet correlation matrix is shown after applying this scaling in the right panel of figure 9.4. Correlations are notably larger than within the inclusive jets and can reach more than 55%.

9.4 Matrix unfolding

An alternative to the bin-by-bin correction method is the matrix unfolding approach. In the following, the hadron-level distribution will be denoted as $\vec{x}$, i.e. x_j is the number of jets in hadron-level bin j and the dimension of $\vec{x}$ is the number of hadron-level bins. Similarly, the detector-level distribution is denoted as $\vec{y}$. A central quantity of the matrix unfolding approach is the migration matrix A. Its elements A_{ij} give the probability for a jet originating in hadron-level bin j to be reconstructed in detector-level bin i. The matrix A thus serves as a model for the detector response. The j-th column of A gives the shape of the detector-level distribution induced by jets in the j-th hadron-level bin. These quantities can be related via the equation

$$
\vec{y} = A \cdot \vec{x}.
\tag{9.7}
$$

With matrix unfolding, the bins at hadron- and detector-level do not need to correspond to each other. Furthermore, a different number of bins can be used at each level. In this analysis, a two-dimensional unfolding is performed. Thus the vectors $\vec{x}$ and $\vec{y}$ refer to two-dimensional distributions. To simplify notation, these two-dimensional distributions are mapped onto a one-dimensional axis, such that the elements can be referred to using a single index and drawn in a one-dimensional histogram.

The basic approach of matrix unfolding is to take the detector-level distribution $\vec{y}$ from the data, the migration matrix A from the MC and use those to determine the truth-level distribution $\vec{x}$ in the data. If $\dim(\vec{x}) = \dim(\vec{y})$, i.e. the number hadron-level and detector-level bins are equal, A is a square matrix. In this case, equation (9.7) can be solved directly by matrix inversion. While mathematically correct, this approach can lead to strong anti-correlations and large fluctuations of the determined truth-level distribution.[193] In matrix unfolding, these fluctuations are reduced by choosing a larger number of bins at detector level than at hadron level. In this case, the equation can be solved by a linear transformation, as described below.

Not all jets that are identified at detector level belong to the signal region at truth level. Background contributions arise when a detector-level jet is misidentified and has no hadron-level counterpart. They also include jets that are generated outside the analysis phase space and have migrated to the inside. Another possible source are jets from different physical processes, such as PHP events that do not belong to the hadron-level phase space by definition. For the unfolding, background contributions are taken into account by defining

$$\vec{y} = \vec{y}_{\text{data}} \odot \left(\vec{1} - \vec{b} \oslash \vec{y}_{\text{MC}} \right), \tag{9.8}$$

where $\vec{1}$ is a vector, all of whose elements are one and $\odot$ and $\oslash$ are element-wise (Hadamard) multiplication and division. The vector $\vec{y}_{\text{data}}$ represents the distribution recorded in the data. The vector $\vec{b}$ represents the background that is estimated from the MC and should be subtracted from the data before unfolding. Subtracting the absolute number of background jets introduces biases to the unfolded result. These biases are reduced by subtracting the relative number of background jets instead. This is achieved by normalising $\vec{b}$ to the MC detector-level distribution $\vec{y}_{\text{MC}}$ and subtracting it multiplicatively. Technically, three types of background can be distinguished:

$$\vec{b} = \vec{b}_{\text{low-}Q^2\text{ DIS}} + \vec{b}_{\text{PHP}} + \vec{b}_{\text{unfolding}}. \tag{9.9}$$

The vectors $\vec{b}_{\text{low-}Q^2\text{ DIS}}$ and $\vec{b}_{\text{PHP}}$ represent the contributions from background processes and are determined from the corresponding background MC samples. The vector $\vec{b}_{\text{unfolding}}$ will be referred to as the 'unfolding background' and represents misreconstructions and migrations from outside the analysis phase space, that are described by the signal MC, e.g. migrations in y or $p_{\perp,\text{Breit}}$.

Unmatched jets also exist at hadron level. These missing jets occur when the event or the jet is discarded at detector level due to inefficiencies in the reconstruction. They also occur when a jet migrates out of the analysis phase space. Missing jets are considered by computing

$$\vec{x}_{\text{data}} = \vec{x} \odot \left(\vec{1} + \vec{m} \oslash \vec{x}_{\text{MC}} \right), \tag{9.10}$$

where $\vec{x}$ is the distribution determined from solving equation (9.7) and $\vec{x}_{\text{data}}$ is the corrected estimate for the truth-level distribution in the data. The vector $\vec{m}$ is the distribution of signal jets at MC hadron level that cannot be matched to a detector-level jet in the phase

space. Similarly to the background contribution, the missing jets are normalised to the MC hadron-level distribution $\vec{x}_{\mathrm{MC}}$ and added multiplicatively.

Here, the convention $\sum_i A_{ij} = 1$ is used, meaning that A only describes jets that are reconstructed at all at detector level. Alternatively, one could use the convention $\sum_i A_{ij} \leq 1$, so that A already accounts for the fact that some jets are not reconstructed. In that case, the missing jets $\vec{m}$ could be absorbed into A.

The unfolded distribution $\vec{x}$ is then determined by minimising the expression

$$\left(\vec{y} - A \cdot \vec{x}\right)^{\mathsf{T}} \cdot V^{-1} \cdot \left(\vec{y} - A \cdot \vec{x}\right), \tag{9.11}$$

where $V = V_{yy} + V_{bb}$ is the covariance matrix. It consists of a contribution from the detector-level distribution, V_{yy}, and one determined from the background, V_{bb}.[194] The detector-level distribution $\vec{y}_{\mathrm{data}}$ and the corresponding covariance matrix V_{yy} are determined from the data as explained in section 9.3.3. The distributions $\vec{m}$, $\vec{b}$, $\vec{x}_{\mathrm{MC}}$, $\vec{y}_{\mathrm{MC}}$ and A are taken from the MC samples and their statistical uncertainties are computed similarly.

In the MC sample, jets are clustered at detector and hadron level and matched to each other using the algorithm described in section 8.3.2, with a distance cut of $\Delta R > 0.9$. Matched pairs are used to fill the response matrix R. The element R_{ij} is the number of jets generated in bin j and reconstructed in bin i. The migration matrix is determined by normalising the columns of the response matrix $A_{ij} = R_{ij}/(\sum_i R_{ij})$. Unmatched hadron-level jets are filled into the vector $\vec{m}$ and unmatched detector-level jets into $\vec{b}_{\mathrm{unfolding}}$.

The uncertainties of bins in the unfolded distribution are correlated. Correlations can arise from two possible sources. Similarly to the bin-by-bin correction, jets are positively correlated if they originate from the same event. These input correlations are described by the matrix V_{yy}. Secondly, primarily negative correlations arise due to the finite detector resolution, leading to migrations of jets between bins, as described by the migration matrix A. Matrix unfolding naturally considers both types of correlation and determines the correlations alongside the unfolded distribution.

The solution for the unfolded distribution $\vec{x}$ and its covariance matrix V_{xx} is given by[194]

$$\vec{x} = D_{xy} \cdot \vec{y}, \tag{9.12a}$$

$$V_{xx} = D_{xy} \cdot V \cdot D_{xy}^{\mathsf{T}}, \tag{9.12b}$$

with

$$D_{xy} = (\partial_{\vec{y}} \vec{x})^{\mathsf{T}} = (A^{\mathsf{T}} \cdot V^{-1} \cdot A)^{-1} \cdot A^{\mathsf{T}} \cdot V^{-1}. \tag{9.12c}$$

If A is a square matrix, i.e. the same number of bins is used at detector and hadron level, this expression reduces to matrix inversion.

Another relevant quantity in matrix unfolding is the condition number of the matrix A. The condition number is a real number greater than one. It is a measure for how sensitive the solution of an equation of type (9.7) is to uncertainties in its input, i.e. the condition number describes how uncertainties of $\vec{y}$ will be amplified when determining $\vec{x}$. If the condition number is close to one, the uncertainty of $\vec{x}$ will not be much greater than that of $\vec{y}$. If the condition number is large, however, the uncertainty in $\vec{y}$ will be amplified and the result $\vec{x}$ can become sensitive to small fluctuations of $\vec{y}$. In this case, the problem is said to be ill-conditioned, and the unfolded distribution can exhibit large fluctuations.[145] In such cases, a regularisation term can be added to the unfolding procedure, which can dampen fluctuations at the cost of introducing a bias to the result. Such a regularisation procedure can distort features of

the truth-level distribution, e.g. if it contains a peak or a step. The condition number of the migration matrix used in the remainder of this section is about 12, which is considered small enough not to need a regularisation term. Using more bins at detector level than at hadron level can also be considered a form of regularisation, as it prevents fluctuations in the unfolded distributions that are not present at detector level.

The uncertainties determined by the unfolding procedure are a combination of statistical uncertainties of data, signal MC and background MC as well as systematic uncertainties due to jets belonging to the same event at theory level and systematic uncertainties due to migrations during reconstruction. With matrix unfolding, these contributions cannot be clearly separated. In the following, the sum of these uncertainties will be referred to as 'unfolding uncertainty'.

9.4.1 Choice of binning scheme

It is important to supply sufficient information to the unfolding procedure to reduce fluctuations in the determined distribution. Therefore, significantly more bins should be used at detector level than for the measurement. Furthermore, previous analyses have found that it can be advantageous to use finer hadron-level bins during the unfolding than are desired in the measurement, as this gives additional flexibility to the unfolding algorithm.[196] Cross sections in this finer binning tend to be strongly correlated. These correlations will be reduced when combining these bins to determine cross sections in the measurement binning.

The choice of hadron- and detector-level binning is optimised using a MC study. One of the Ariadne and Lepto MC samples is used as pseudo-data and unfolded with the other sample. The unfolded distribution is compared to the corresponding MC truth. Various aspects of the binning are, in turn, modified, and each change is kept if it improves the agreement between unfolded and truth distributions.

It is found that subdividing the Q^2 bins at hadron level has little effect on the result. However, subdividing the $p_{\perp,\mathrm{Breit}}$ bins reduces the dependence of the unfolded result on the MC model. It is determined that the optimal choice at hadron level is to divide each $p_{\perp,\mathrm{Breit}}$ bin into two, except for the lowest one. In the low-$p_{\perp,\mathrm{Breit}}$ region, the energy resolution of the reconstructed jets is limited by the CAL. To ensure sufficient information is supplied to the unfolding algorithm, bins at detector level are subdivided into two or three bins, depending on whether they were already subdivided at hadron level.

The fraction of missing and background jets in the signal region is reduced by adding overflow bins in Q^2 and $p_{\perp,\mathrm{Breit}}$ and an underflow bin in $p_{\perp,\mathrm{Breit}}$. As the signal MC samples are generated with the cut $Q^2 > 100\,\mathrm{GeV}^2$, all events that originate from below this threshold need to be considered background and subtracted using the dedicated low-Q^2 DIS background MC sample. Below the measurement region, this background becomes sizeable and therefore, no underflow bin is added in Q^2.

The final bin definitions used at hadron and detector level are given in table 9.2. In the following, the three types of binning will be called detector-level binning, hadron-level binning and measurement binning.

9.4.2 Inputs and results

Like the bin-by-bin correction, the matrix unfolding is performed using a combination of the Ariadne and Lepto signal MC samples to mitigate the model dependence. The remaining systematic uncertainty is estimated in section 10.2. In the following, the term '(signal) MC' will always refer to this combined sample.

Measurement		Hadron level		Detector level	
Label	Q^2 [GeV^2]	i	Q^2 [GeV^2]	i	Q^2 [GeV^2]
1	150 – 200	0	150 – 200	0 1	150 – 175 175 – 200
2	200 – 270	1	200 – 270	2 3	200 – 230 230 – 270
3	270 – 400	2	270 – 400	4 5	270 – 320 320 – 400
4	400 – 700	3	400 – 700	6 7	400 – 500 500 – 700
5	700 – 5000	4	700 – 5000	8 9 10	700 – 1000 1000 – 2500 2500 – 5000
6	5000 – 15000	5	5000 – 15000	11	5000 – 15000
		6	15000 – ∞	12	15000 – ∞

Measurement		Hadron level		Detector level	
Label	$p_{\perp,\mathrm{Breit}}$ [GeV]	j	$p_{\perp,\mathrm{Breit}}$ [GeV]	j	$p_{\perp,\mathrm{Breit}}$ [GeV]
		0	4.5 – 7	0	4.5 – 7
a	7 – 11	1	7 – 11	1 2	7 – 9 9 – 11
b	11 – 18	2 3	11 – 14 14 – 18	3 4 5	11 – 13 13 – 15 15 – 18
c	18 – 30	4 5	18 – 24 24 – 30	6 7 8	18 – 21 21 – 25 25 – 30
d	30 – 50	6 7	30 – 39 39 – 50	9 10 11	30 – 36 36 – 45 45 – 50
		8	50 – ∞	12	50 – ∞

Table 9.2: Bin definitions used during matrix unfolding. The measurement bins, shown in the left columns, are identical to table 9.1 and are labelled by a number and a letter. The hadron-level binning is given in the middle columns. The two-dimensional binning is mapped onto a one-dimensional axis by assigning each $(Q^2, p_{\perp,\mathrm{Breit}})$-bin a one-dimensional index as $9i + j$. Bins in the one-dimensional axis are indexed from 0 to 62. The detector-level binning is shown in the right columns. The bins are mapped onto a one-dimensional axis as $13i + j$, which leads to bin indices ranging from 0 to 168. Similar to figure 9.3, the blocks of bins correspond to the Q^2 bins. The individual bins within these blocks represent the $p_{\perp,\mathrm{Breit}}$ bins.

Detector-level distributions   The detector-level distributions are given in figure 9.5. It can be seen that the background distribution $\vec{b}$ is dominated by jets from the signal MC that cannot be assigned to a hadron level bin. Its contribution is about 15% in the central region of the phase space and increases towards the edges, especially in the under- and overflow bins. Within statistical uncertainties, the sum of the three MC contributions describes the data reasonably, achieving a χ^2 per degree of freedom[164] of about 1.2 in the signal region.
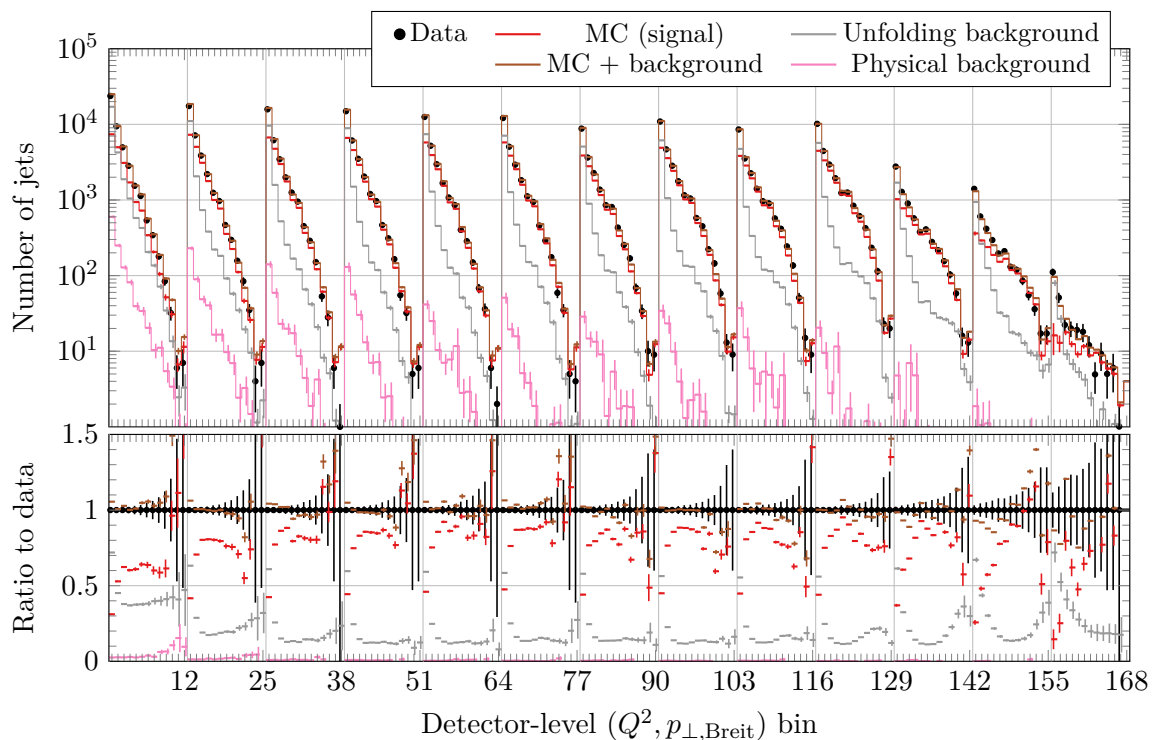
Figure 9.5: Distributions of jets at detector level in data $\vec{y}_{\text{data}}$, signal MC $\vec{y}_{\text{MC}}$ and signal + background MC $\vec{y}_{\text{MC}} + \vec{b}$, unfolding background $\vec{b}_{\text{unfolding}}$ and the contribution from background processes $\vec{b}_{\text{low-}Q^2 \text{ DIS}} + \vec{b}_{\text{PHP}}$. The bin indices are defined in table 9.2. The large bins indicate Q^2 bins, and the smaller ones are $p_{\perp,\text{Breit}}$ bins.

The matrix unfolding method naturally takes migrations between the unfolded bins into account. Since a two-dimensional unfolding of Q^2 and $p_{\perp,\text{Breit}}$ is performed, migrations along both directions will be considered. It is also possible to include further dimensions in the unfolding to reduce the dependence on the correctness of the MC model. For this purpose, the variables y and η_{lab} are well suited, as they are used to limit the phase space. Also, if under- and overflow bins in these quantities are considered, the fraction of missing and background jets in the signal region could be further reduced. However, as shown in figure 9.5, some bins already have just ten or fewer entries in the data. If bins were subdivided further, or lower statistics underflow and overflow bins were added, the data statistics would be insufficient for a reliable unfolding. Therefore, no further dimensions are added to the unfolding.

Detector-level covariance matrix The covariance matrix V_{yy} corresponding to the data distribution is shown in figure 9.6. An exact block structure arises because, by definition, jets from events with different Q^2 cannot be correlated. Each block in the figure corresponds to one Q^2 bin. The structure within each block represents the $p_{\perp,\text{Breit}}$ correlations in that Q^2 bin.

Migration matrix The final input to the unfolding procedure is the migration matrix as determined from the signal MC sample. This matrix is shown in figure 9.7. Also, here, a distinct structure is visible. The structure arises due to the overlap of the detector-level and hadron-level bins in table 9.2 and is slightly distorted due to the detector resolution.
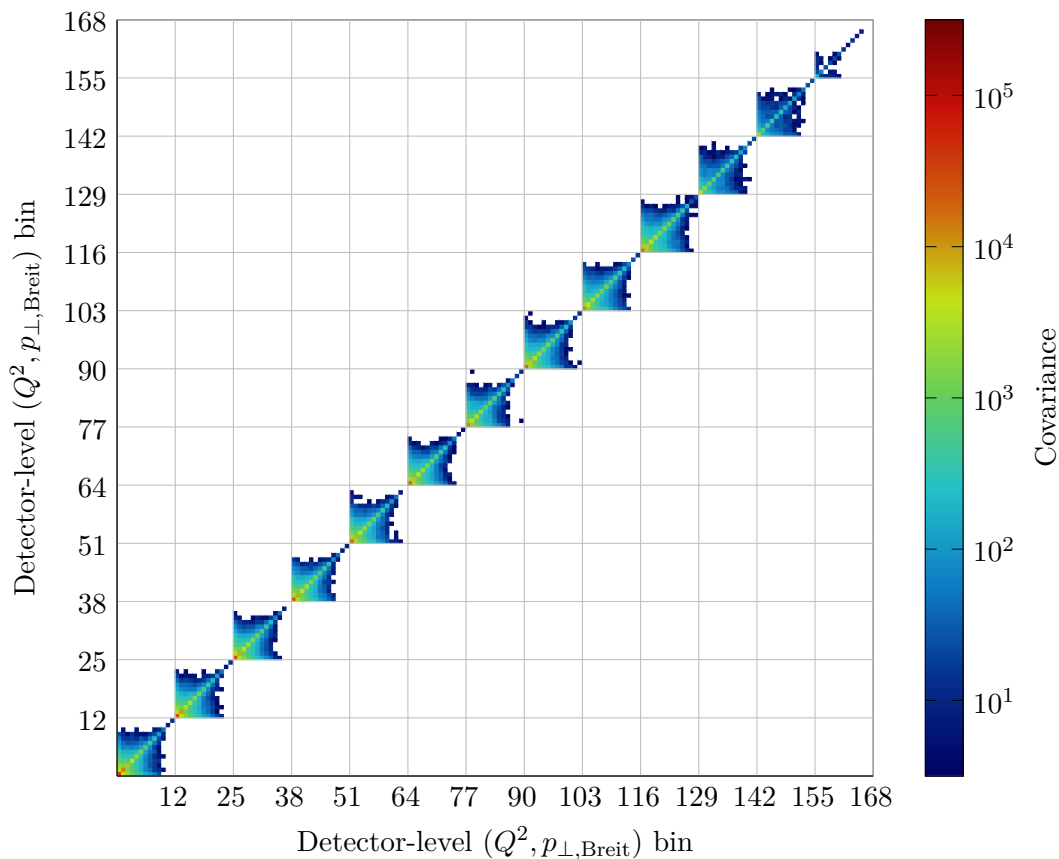
Figure 9.6: Covariance matrix of the data distribution V_{yy}. The bins are defined according to table 9.2.

Unfolded distributions in hadron-level binning   The unfolded distribution is shown in figure 9.8, compared to the corresponding distribution from the MC. It can be seen that matrix unfolding and the bin-by-bin correction method give consistent results in the signal region. In some overflow bins where the number of jets gets small, the matrix unfolding method fails to determine estimates of the hadron-level distribution. For bin values that are compatible with zero and thus do not contribute to the detector-level distribution, matrix unfolding can lead to very small or even slightly negative values. Within uncertainties, these bins are compatible with the shape created by the surrounding bins. This is not problematic as long as these bins do not belong to the measurement phase space.

The figure also shows that the uncertainties from matrix unfolding appear significantly larger than those from the bin-by-bin correction. This is primarily due to the fine hadron-level binning, in which bins are strongly anti-correlated when using matrix unfolding. The uncertainties appear large, as the uncertainties in the figure only represent the diagonal entries of the covariance matrix. When combining the bins to derive distributions in the measurement binning, correlations will be reduced, and uncertainties will appear smaller. Another reason is that the unfolding uncertainty from the matrix unfolding includes additional systematic contributions that are not included in the statistical uncertainty of the bin-by-bin correction. The correlation matrix in the hadron-level binning is given in figure 9.10. The anti-correlation between neighbouring $p_{\perp,\text{Breit}}$ bins frequently exceeds -50% and can reach as high as -95%.
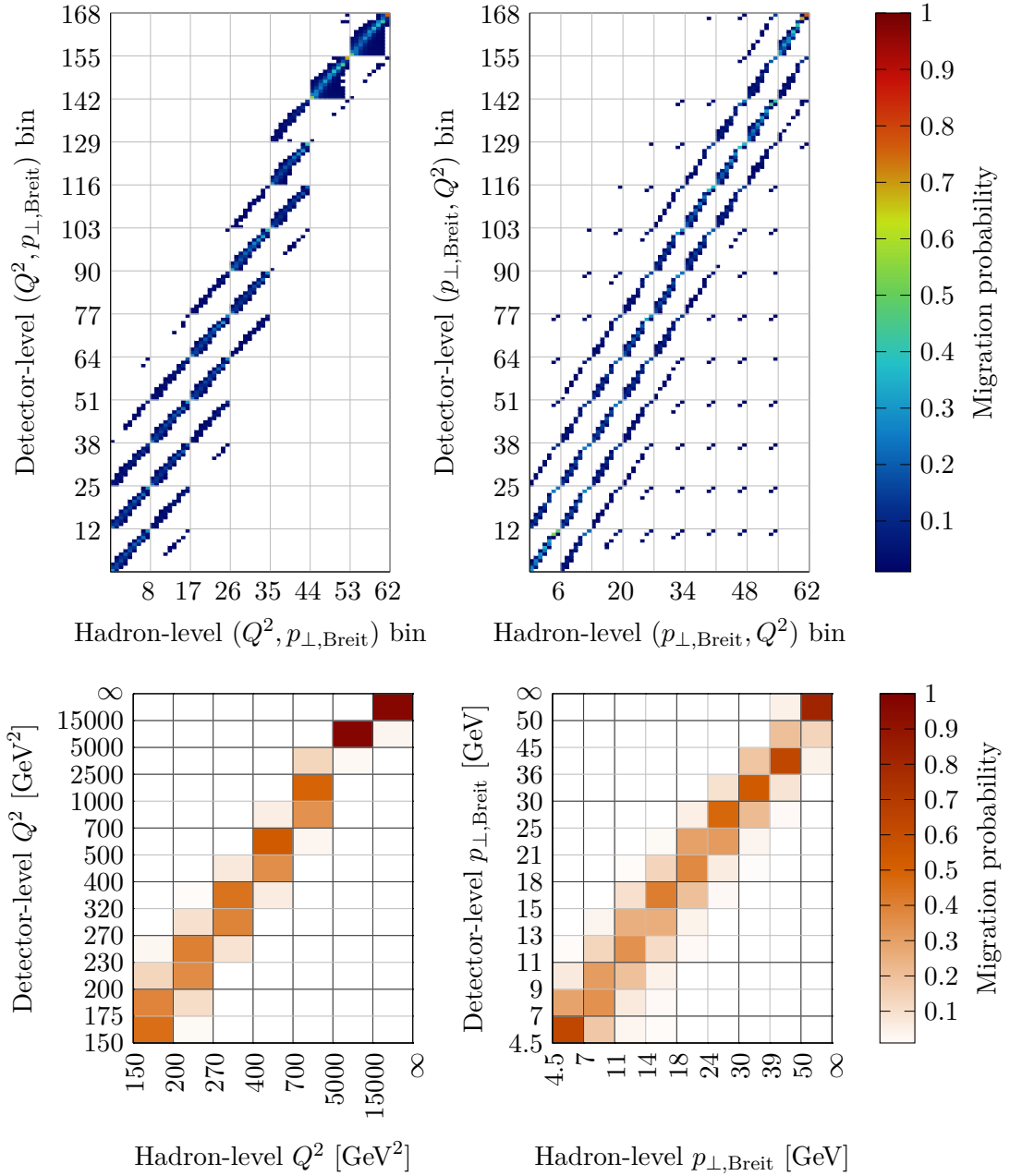
Figure 9.7: Migration matrix A from hadron to detector level. In the upper left panel, the detector-level and hadron-level bins are defined according to table 9.2. The upper right panel shows the same data, but the two-dimensional axes are mapped to one dimension as $13j + i$ and $7j + i$ at detector and hadron level, respectively. Consequently, the large bins correspond to $p_{\perp,\text{Breit}}$ bins and the small bins to Q^2. This way of drawing the matrix makes it easier to evaluate migrations in $p_{\perp,\text{Breit}}$, while the left panel is more suitable to judge migrations in Q^2. In the lower row, the response matrix is integrated over $p_{\perp,\text{Breit}}$ (left) or Q^2 (right) before deriving the migration matrix. These two figures are thus simplified versions of those in the upper row. They show only migrations in Q^2 or $p_{\perp,\text{Breit}}$, respectively.
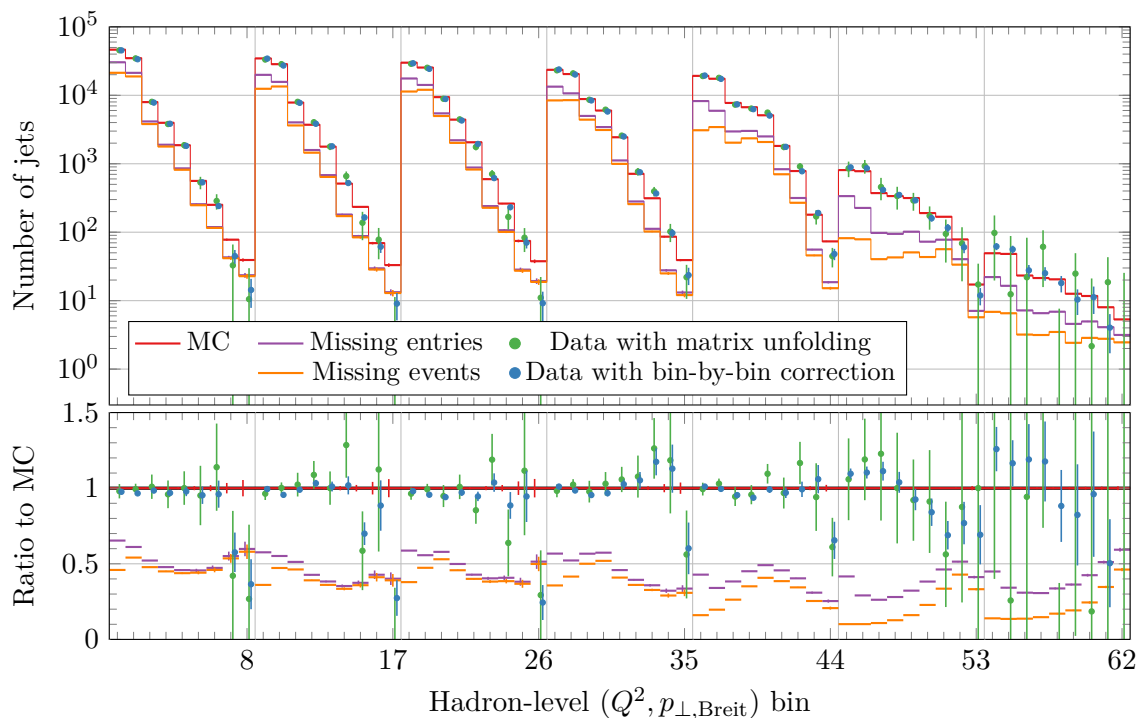
Figure 9.8: Distributions of jets at hadron level in the unfolded data $\vec{x}_{\text{data}}$, the signal MC $\vec{x}_{\text{MC}}$ and the missing entries $\vec{m}$. The bin indices are defined in table 9.2. The large bins indicate Q^2 bins, and the smaller ones are $p_{\perp,\text{Breit}}$ bins. Additionally shown is the contribution from missing hadron-level jets, that are missing because their event is rejected at detector level. For reference, the same distribution is derived using the bin-by-bin correction method.

Taking into account the correlations, the matrix unfolded distribution results in χ^2 per degree of freedom of about 0.9 compared to the MC truth in the signal region, which is consistent with the value at detector level. The bin-by-bin correction yields an increased value of 4.6. This is mostly due to the smaller uncertainty estimates of this method. The χ^2 per degree of freedom at hadron level is not necessarily expected to be 1, since data and MC might disagree. Instead, it should be compared to the corresponding value at detector level. This comparison is known as the bottom-line test.[164]

The figure also shows that the contribution from unmatched hadron-level jets is significant and makes up $30\% - 60\%$ of all jets. Many of the unmatched hadron-level jets are unmatched due to the corresponding event being rejected during detector-level reconstruction, e.g. since no under- and overflow bins were added in y and η_{lab}, small migrations close to the boundaries can cause jets to become unmatched. Only a small fraction of unmatched jets is due to deficiencies in the jet reconstruction. A similar effect is observed when using the bin-by-bin correction, where a low efficiency of $25\% - 45\%$ is observed, see figure B.14. As the efficiency represents the fraction of hadron-level jets matched to a detector-level jet in the same bin, the efficiency is, by definition, smaller than the matching rate.

Unfolded distributions in measurement binning   The hadron-level bins are combined to derive the measurement distribution shown in figure 9.9. It is again found that both unfolding methods yield consistent results. Again in the measurement binning, the uncertainties from

matrix unfolding appear larger, though the difference is much smaller than in the hadron-level binning. The most noticeable difference in the size of uncertainties is observed in bins 6a and 6b, though these are also the most strongly anti-correlated bins, as shown in the correlation matrix in figure 9.10.

In the measurement binning, the χ^2 per degree of freedom is 3.5 with matrix unfolding and 7.2 with the bin-by-bin correction. Both of these values are larger than their hadron-level counterparts. This is because there is a slight disagreement between data and MC. The disagreement is less prominent in the finer hadron-level binning because of the larger uncertainties. The difference is larger for matrix unfolding since the anti-correlations lead to an even stronger reduction of uncertainties when combining bins.

9.4.3 Correlations to dijet measurement

As discussed in section 9.3.3, it is vital to determine statistical correlations to make the measured cross sections usable in further studies. Matrix unfolding naturally takes correlations into account, and the correlation matrix presented in the previous section is sufficient to make the inclusive jet measurement usable on its own. To make the measurement also usable in conjunction with the previous ZEUS dijet measurement, correlations between the two measurements need to be determined.

It would be possible to extend the measurement vectors $\vec{x}$, $\vec{y}$, A, ... by additional dijet bins and unfold the two measurements simultaneously. This approach would be similar to the one taken for the bin-by-bin correction. However, for the current case, this method is not applicable. If the two measurements were unfolded simultaneously, the obtained correlation matrix would refer to matrix unfolded measurements, in which migrations between bins are explicitly considered during the unfolding. This is not the case for the previous dijet measurement, as that was performed using a bin-by-bin correction. The importance of this distinction is further demonstrated by the fact that the inclusive jet correlation matrix determined for the bin-by-bin correction in figure 9.4 is qualitatively different from the one for matrix unfolding in figure 9.10.

Rather than appending all of the dijet bins to the measurement vectors, only a single dijet bin is added. When repeating the unfolding, a single row of the inclusive jet – dijet correlation matrix is determined, namely that row that describes the correlations between all the inclusive jet bins and the single added dijet bin. The entire inclusive jet – dijet correlation matrix can be constructed by repeating this procedure for every dijet bin. Using this method, migrations of dijets between bins are not explicitly considered during the unfolding since only a single dijet bin is considered at every step. This makes it possible to derive a consistent correlation matrix between the matrix unfolded inclusive jets and the bin-by-bin corrected dijets.

The determined correlation matrix is given in figure 9.11, scaled by 80%, similar to the previous section. The structure is similar to the corresponding matrix for the bin-by-bin correction, as shown in figure 9.4, but the strength of the correlations is reduced, not exceeding 40%.

9.4.4 Formulation as template fit

A template fit describes a distribution as a linear combination of several component distributions. The coefficients are determined in a linear fit. This technique is often used to describe detector-level distributions in the data as a sum of component distributions from the MC. Matrix
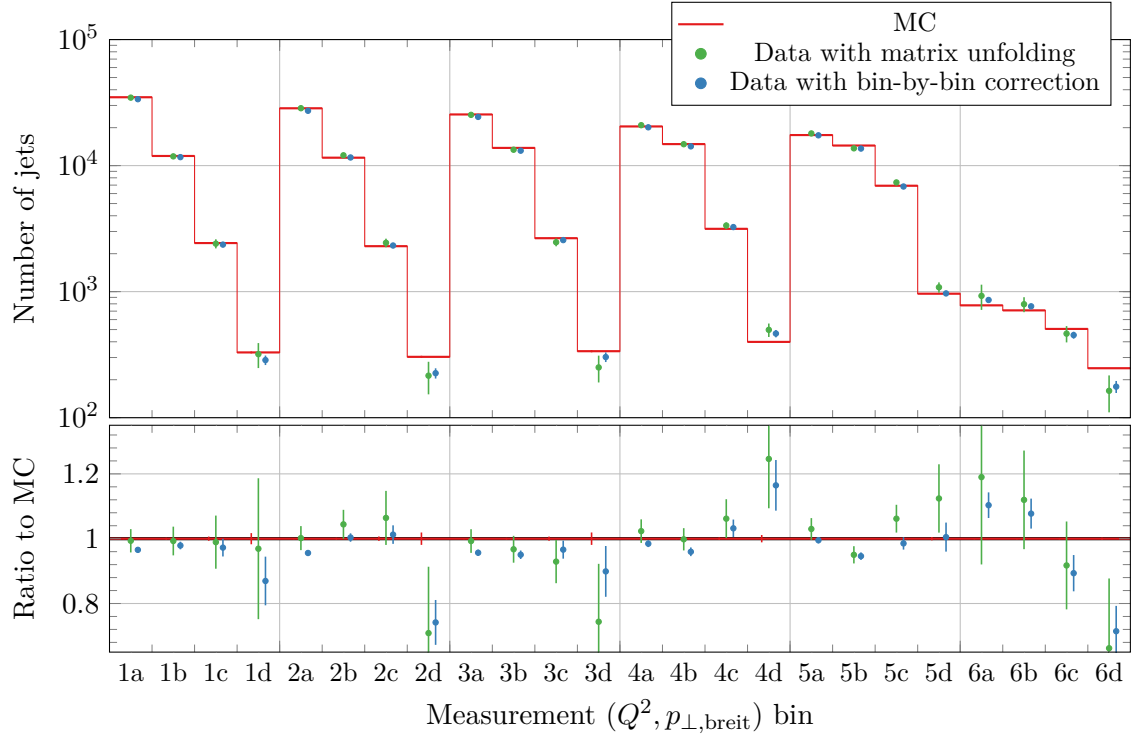
Figure 9.9: Distributions of jets at hadron level in the measurement binning. The data are corrected to hadron level using matrix unfolding and the bin-by-bin correction method and are compared to the MC truth distribution. The bin labels are defined in table 9.2. The large bins indicate Q^2 bins, the smaller ones $p_{\perp,\mathrm{Breit}}$.
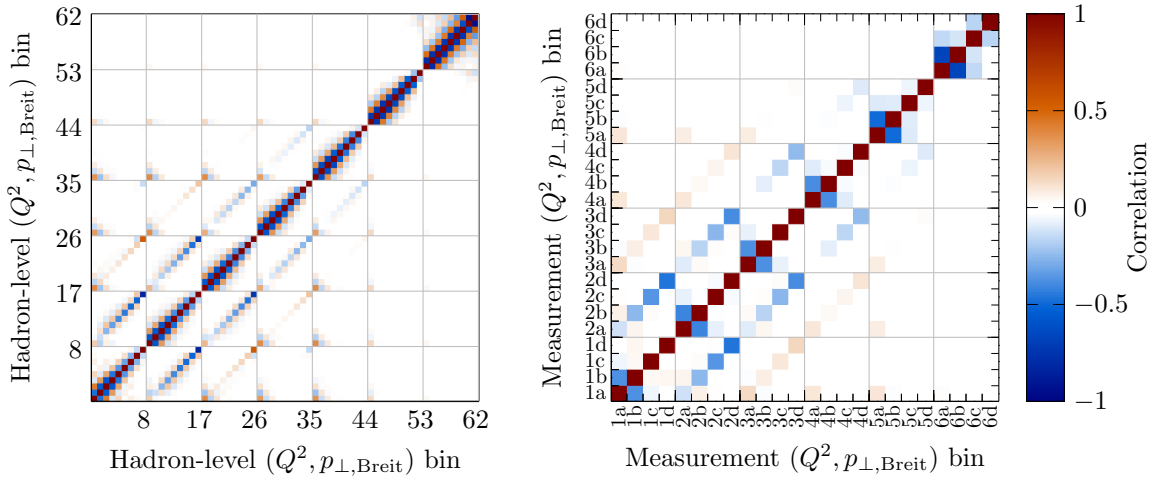
Figure 9.10: Correlation matrices of the unfolding uncertainty in the hadron-level binning (left) and in the measurement binning (right). The bin labels are defined in table 9.2. The numerical values of the correlation matrix in the measurement binning are given in table C.6.
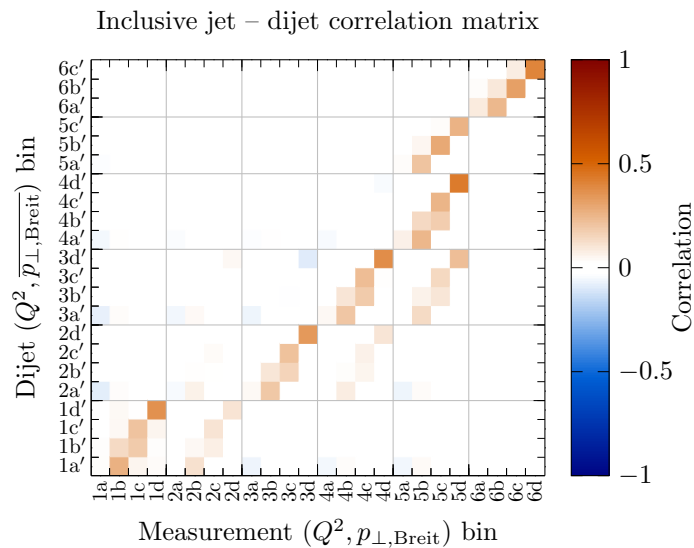
Figure 9.11: Correlation matrix between matrix unfolded inclusive jets and bin-by-bin corrected dijets. The bins of the inclusive jet measurement are labelled according to table 9.2. The large bins indicate Q^2 bins, and the smaller ones are $p_{\perp,\mathrm{Breit}}$ bins. The dijet bins are labelled similarly but using the bin definitions from that measurement.[71] The numerical values of this matrix are given in table C.6.

unfolding is often presented in this context, which helps understand the method in general and how the unfolding behaves in a specific situation.

As explained earlier in this section, each column of the response matrix corresponds to the shape of the detector-level distribution induced by jets in the corresponding hadron-level bin, i.e. when scaling each column by the corresponding hadron-level bin content and adding up all the columns, the full detector-level distribution is obtained. Rewriting equation (9.7) in this context yields

$$\vec{y} = \sum_j \vec{A}_j x_j, \tag{9.13}$$

with $\vec{A}_j = \sum_i A_{ij}\vec{e}_i$. This equation highlights the similarity of matrix unfolding and template fits. The j columns of the response matrix $\vec{A}_j$ are the component distributions used in the template fit. The hadron-level bin contents x_j are the coefficients of the components and are determined in the fit.

The effect of matrix unfolding can be understood by comparing the different components $\vec{A}_j x_j$ before and after the fit to the data, i.e. before and after the unfolding. Before the unfolding, the x_j are taken from the MC, and the sum of templates is the distribution from the signal MC. After the fit, they represent the unfolded distribution from the data. At this point, the sum of templates should agree with the data.

In this analysis, the hadron-level distribution used during the unfolding consists of 63 bins, as shown in table 9.2. This means there are 63 individual templates, each corresponding to one $(Q^2, p_{\perp,\mathrm{Breit}})$ bin. Drawing all of these templates at once would not give a clear and helpful picture. To obtain a manageable number of graphs, templates can be combined to form a smaller number of meta-templates. Two separate comparisons are made. In the first, templates are combined such that each meta-template corresponds to one hadron-level Q^2 bin,

leading to seven meta-templates. For the second comparison, nine meta-templates are formed, corresponding to one hadron-level $p_{\perp,\mathrm{Breit}}$ bin each. These two variants of the comparison are shown in figures 9.12 and 9.13.

Comparing these meta-templates before and after unfolding does not illustrate the full effect of the unfolding. Because the unfolding happens in both Q^2 and $p_{\perp,\mathrm{Breit}}$, the shape of the meta-templates will change during the unfolding, which is not the case for a regular template fit. Nevertheless, understanding the behaviour of the meta-templates in both cases can help understand the behaviour of the unfolding.

The structure of these graphs can be explained by the correspondence between the hadron-level and detector-level bins. The large bins in figure 9.12 correspond to Q^2 bins at detector level. The first two detector-level Q^2 bins are dominated by the $Q^2 \in (150, 200)\,\mathrm{GeV}^2$ meta-template. This is because the first two detector-level Q^2 bins range from 150–175 GeV^2 and 175–200 GeV^2 and thereby correspond to the first hadron-level Q^2 bin, see table 9.2. The next two detector-level Q^2 bins are dominated by the $Q^2 \in (200, 270)\,\mathrm{GeV}^2$ meta-template and so on. The same applies to figure 9.13, where the large bins correspond to detector-level $p_{\perp,\mathrm{Breit}}$ bins. Here, the first meta-template dominates only the first detector-level bin since the bin boundaries are identical at hadron level and detector level. The second meta-template corresponds to the second and third detector-level bin, and so on.

As expected, the figures show that the detector resolution is better in Q^2 than in $p_{\perp,\mathrm{Breit}}$. The Q^2 meta-templates contribute significantly only to a few detector-level bins. Their contribution drops off rapidly when moving away from this region. Especially in the high-Q^2 region, meta-templates contribute almost exclusively to their corresponding detector-level bins. The $p_{\perp,\mathrm{Breit}}$ meta-templates contribute to a broader range of detector-level bins. This indicates that migrations in $p_{\perp,\mathrm{Breit}}$ occur more frequently and over greater distances than in Q^2. The same effect is visible in the migration matrices in figure 9.7.

When comparing the sum of the templates to the data, it is clear that the unfolding improves the agreement in both shape and normalisation. Before the unfolding, the shapes of the Q^2 and $p_{\perp,\mathrm{Breit}}$ distributions deviate systematically from the data. After unfolding, they are essentially identical and agree within statistical uncertainty. This is especially visible in the $p_{\perp,\mathrm{Breit}}$ overflow bin of figure 9.13. In figures 9.12 and 9.13, this effect is small since the combined sample from Ariadne and Lepto agrees well with the data. In appendix B.5, the analogous figures are shown using only the Ariadne sample, where the effect of the unfolding is much more visible.
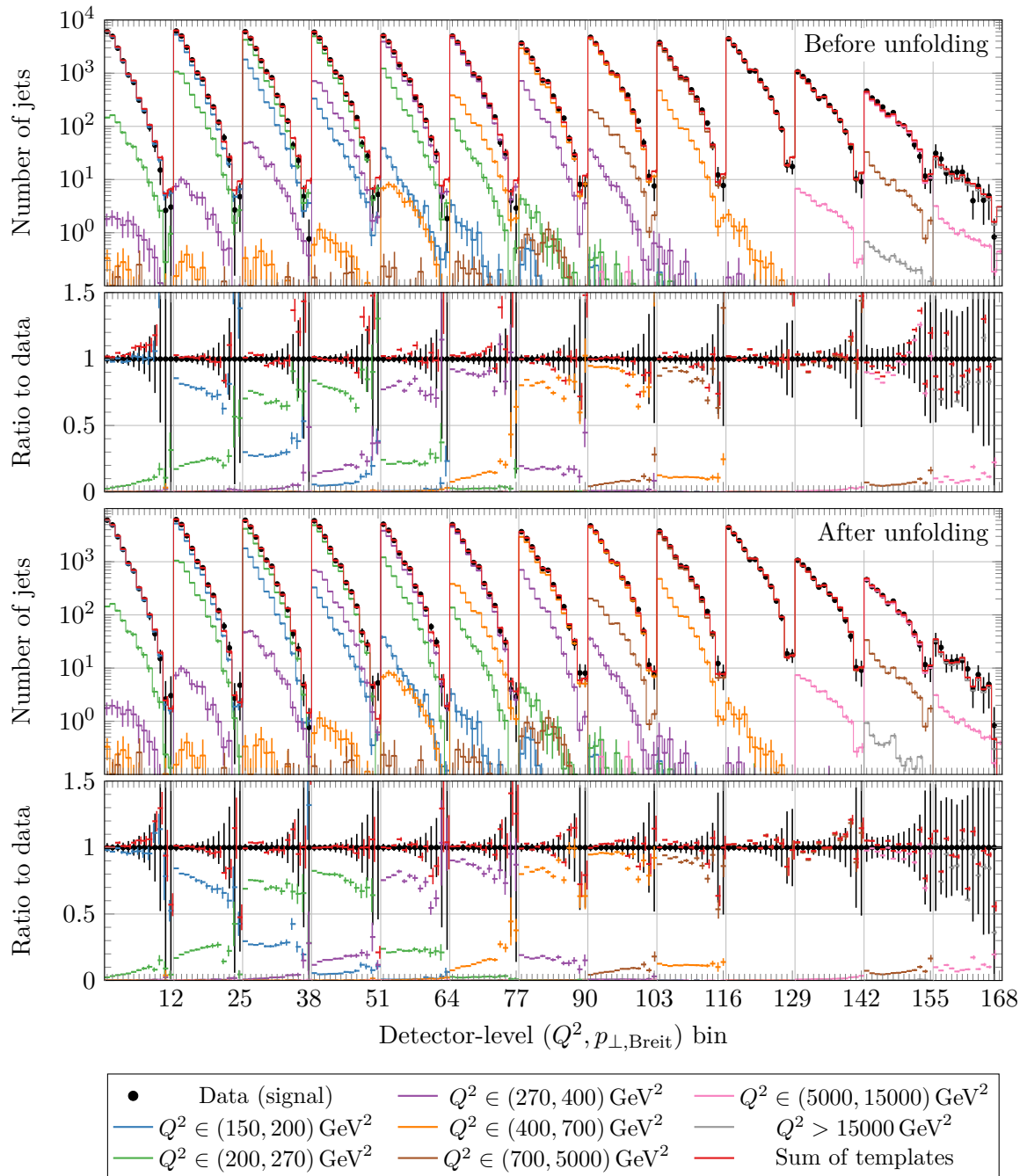
Figure 9.12: Matrix unfolding as template fit in Q^2. The data are compared after subtracting the background to the Q^2 meta-templates before (top panel) and after (bottom panel) the unfolding. The bin indices are defined in table 9.2. The large bins indicate Q^2 bins, and the smaller ones are $p_{\perp,\text{Breit}}$ bins.
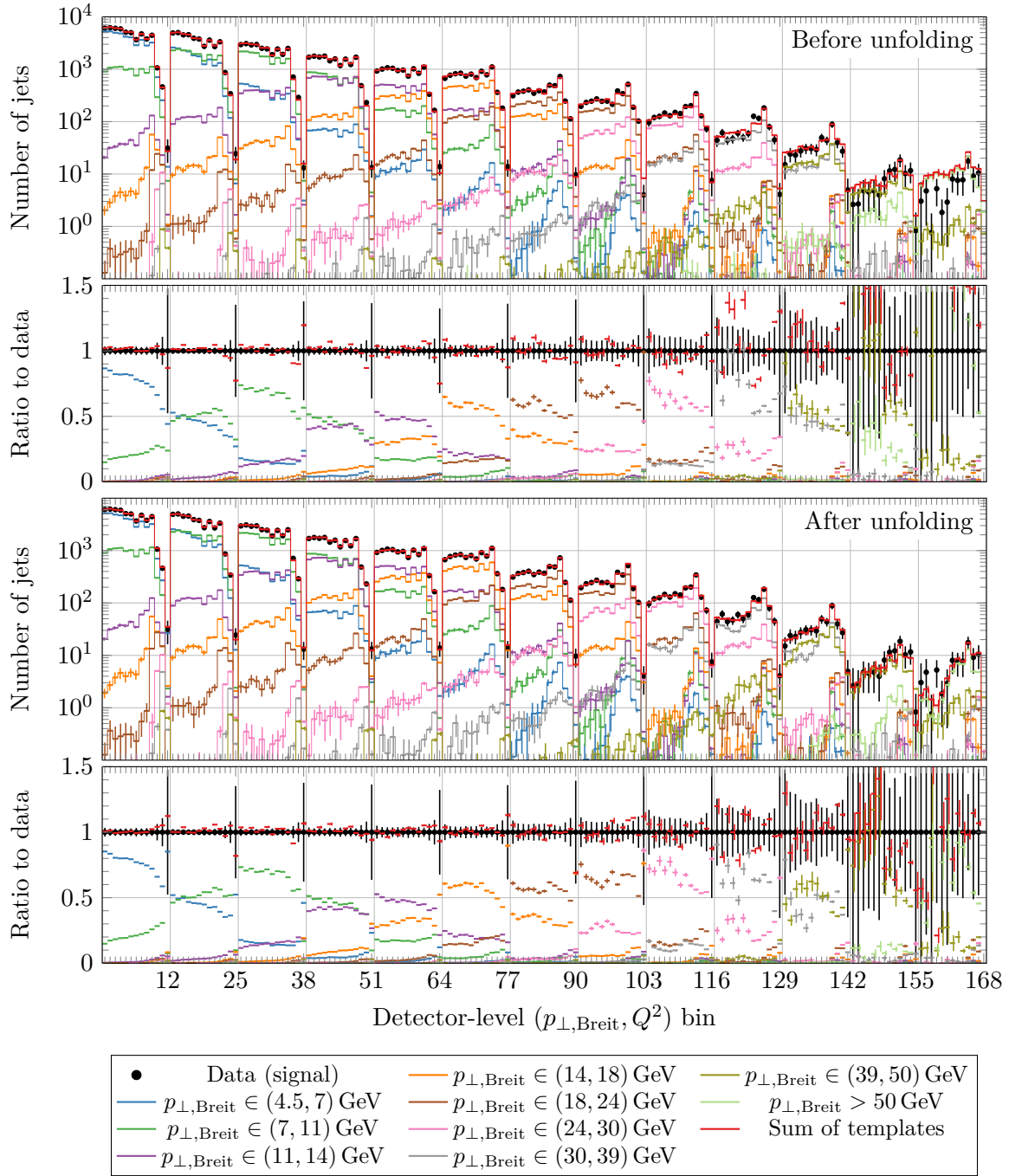
Figure 9.13: Matrix unfolding as template fit in $p_{\perp,\text{Breit}}$. For details see the caption of figure 9.12. To better show the shape of $p_{\perp,\text{Breit}}$ distribution, the bin indices are defined differently from the ones different in table 9.2. The two-dimensional distributions are mapped onto a one-dimensional axis as $13j + i$ rather than $13i + j$. Consequently, the large bins now indicate $p_{\perp,\text{Breit}}$ bins, and the smaller ones are Q^2 bins.

9.5 Theory corrections

The unfolded cross sections correspond directly to the data. These cross sections cannot be compared directly to fixed-order perturbative QCD predictions because not all effects present in the data are included in these predictions. To make a comparison, correction factors are derived and applied to the measured or the predicted cross sections. In the case of the present measurement, three corrections need to be derived: hadronisation, weak-boson exchange and QED radiation. By convention, the first two are applied to the theory predictions, whereas the last is applied to the data.

9.5.1 Hadronisation

The most notable effect in which data and theory differ is hadronisation, see section 5.2.2 and 7.4.2. Only hadrons are observable in the data so that the measured cross sections necessarily correspond to hadron jets. On the other hand, perturbative QCD can only describe partons. The relation between partons and hadrons is described by the hadronisation correction.

In this analysis, the size of the hadronisation correction is estimated using the same Ariadne and Lepto MC samples used for the unfolding. In each sample, two sets of jets are reconstructed at generator level, one from hadrons and one from partons. The correction factors are then defined as the ratio of the number of jets of the two jet types $c_\mathrm{had} = \frac{N_\mathrm{hadron\ jets}}{N_\mathrm{parton\ jets}}$ in a given bin. These factors are determined separately from the Ariadne and Lepto MC chains. The average of the correction factors is used as the nominal correction, and half the difference between them is used as an uncertainty of the correction.

Both Ariadne and Lepto rely on a leading-order matrix element calculation and an initial- and final-state parton shower to approximate higher orders. Compared to this approach, the theory calculations produce a relatively small number of partons (up to four at NNLO) whose hadronisation does not necessarily correspond to the hadronisation observed in the MC samples. It is assumed that the uncertainty of the hadronisation correction covers not only the difference between the two parton showering models used in Ariadne and Lepto but also the difference between the parton showering approach itself and the partons from the fixed-order calculation.

The predicted number of jets of each type is given in figure 9.14 for the two samples. The ratio panel shows the hadronisation correction. The predictions for parton and hadron jets differ significantly between the two MC chains, but hadronisation corrections determined from each sample are similar. The correction factors are below 1, indicating that there are fewer hadron jets than parton jets. This can be explained by the weak decays during hadronisation, in which neutrinos are produced. Neutrinos are, by definition, not part of the hadron-level jets. Furthermore, if hadrons are emitted too far away from the jet axis, they will not be part of the hadron-level jet. Therefore, hadron-level jets tend to have slightly lower energy and are more likely to be rejected by $p_{\perp,\mathrm{Breit}}$ requirements. In most bins, the effect is about 3%, but in the lowest $p_{\perp,\mathrm{Breit}}$ bins, it can reach up to 11%. These correction factors are applied to the perturbative QCD predictions to obtain predictions for hadron jets.

9.5.2 Weak-boson exchange

The theory predictions to which the measured cross sections will be compared consider only a photon as the exchanged DIS boson. This is a good approximation at low Q^2 values. However,
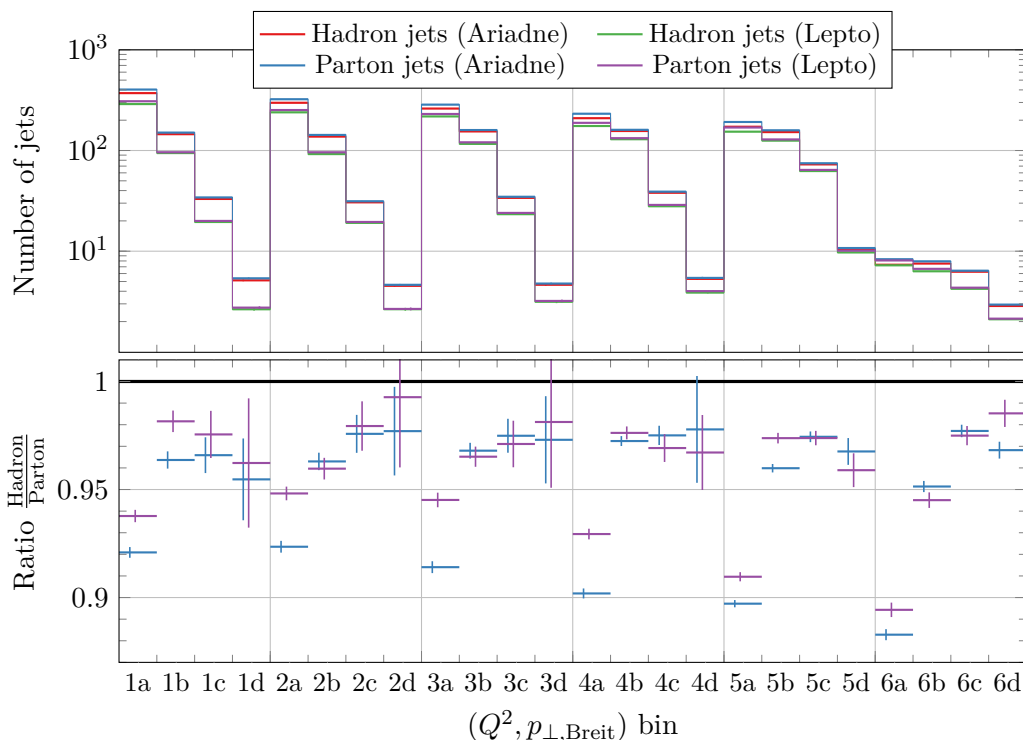
Figure 9.14: Jet predictions using the Ariadne and Lepto MC chain for hadron and parton jets. The average of the two graphs in the ratio panel gives the hadronisation correction. The uncertainty of the correction is given by half the difference between these two graphs. The numerical values for this correction and the associated uncertainty are given in table C.4.

when the boson virtuality becomes comparable to the mass of the Z boson, the exchange of a Z boson and the interference of the photon and Z boson processes can no longer be neglected.

Additional MC samples are generated as described in section 5.3.3 to account for this effect. In one sample, contributions from the Z boson are disabled, similar to the theoretical predictions. Two more samples are generated, including these contributions, one for e^-p and one for e^+p interactions. Due to the parity-violating nature of the weak interaction, the inclusive DIS cross section depends on the type of lepton beam, see section 2.3.5.

The predictions for the inclusive DIS cross section from each of these three samples is shown in figure 9.15. The unpolarised jet cross section is expected to depend on the lepton charge only via the inclusive DIS cross section. Therefore, this correction is applied only as a function of Q^2 and not as a function of $p_{\perp,\mathrm{Breit}}$. An additional graph in the figure shows the average of the cross sections, including electroweak corrections, weighted according to the luminosities of the electron and proton beam data samples used in this analysis. This ratio is $190.3/156.7 \approx 6/5$, see table 6.1.

The correction factors $c_{Z^0} = \frac{N_{\mathrm{photon\ and\ }Z}}{N_{\mathrm{photon\ only}}}$ are given in the ratio panel of the figure. As expected, the influence of electroweak effects is negligible at low Q^2. The correction is only significant in the highest two Q^2 bins, where it reaches a size of 36% for the individual samples and 15% for the average. The correction factors used in the analysis are given by the graph for the weighted average in the ratio panel. Since electroweak effects are well understood, no uncertainty is assigned to this correction.[159] A sufficient number of MC events is generated, such that the statistical uncertainty of these correction factors becomes negligible.
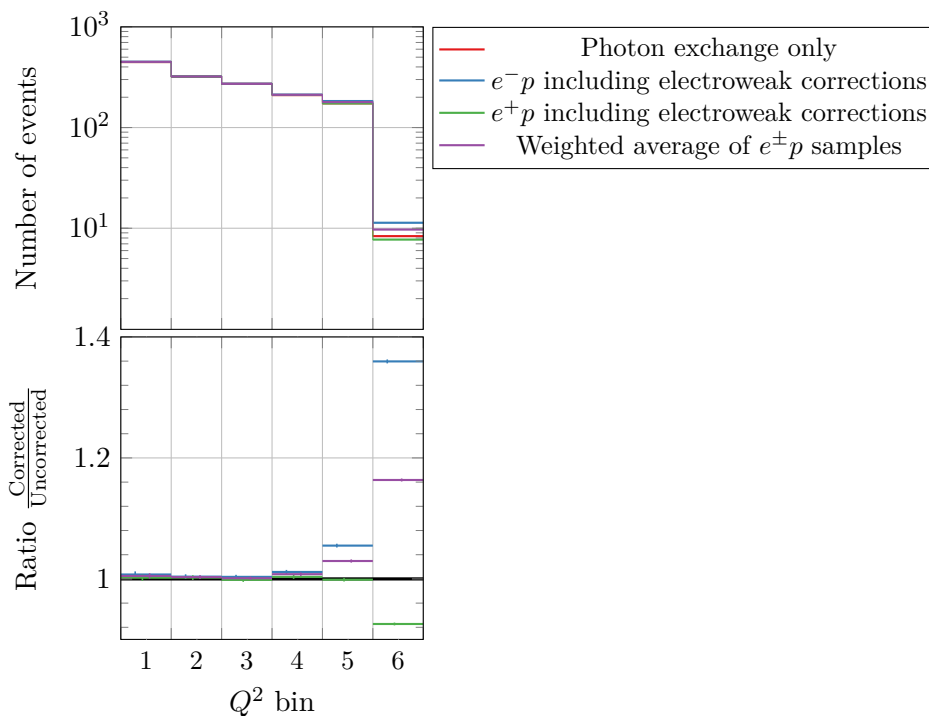
Figure 9.15: Inclusive DIS predictions calculated including and excluding electroweak corrections. The ratio panel gives the electroweak correction. The numerical values for this correction are given in table C.4.

9.5.3 QED radiation

As discussed in section 6.5, the treatment of radiative effects in the cross section definition is non-trivial. In the presence of radiative corrections, the definition of the kinematic quantities is no longer unique. The approach often chosen at HERA is to apply an artificial correction to the data to remove QED radiative effects. At this point, the cross sections are referred to as being at QED Born level. Since the theoretical prediction to which this measurement will be compared does not consider QED radiation, the same approach is taken in this analysis.

For this purpose, further MC samples are generated using the chain described in section 5.3.3. Two sets of samples are generated, one at QED Born level and one including QED radiation on the electron side. In both cases, one sample for e^-p and one for e^+p collisions is generated, which are then combined and weighted by the luminosities of the run periods with each lepton type in the data. From each sample, inclusive jet cross sections at hadron level are computed. For the sample including QED radiation, the electron method is used, as described in section 6.5. The predictions are shown in figure 9.16. The correction factors $c_{\mathrm{QED}} = \frac{N_{\mathrm{Born\ level}}}{N_{\mathrm{With\ radiation}}}$ are given in the ratio panel of the figure. The correction is sizeable, especially in the high-Q^2 region. The statistical uncertainty from generating the MC samples is used as a systematic uncertainty of the corrected cross sections.

In principle, a similar correction should be applied to QED radiation on the quark side. Since these effects are expected to be very small, this effect is neglected.

Alternative QED correction NLO electroweak corrections are available for the DIS process,[197] but they are not yet integrated into the NNLO QCD jet calculations. It is desirable
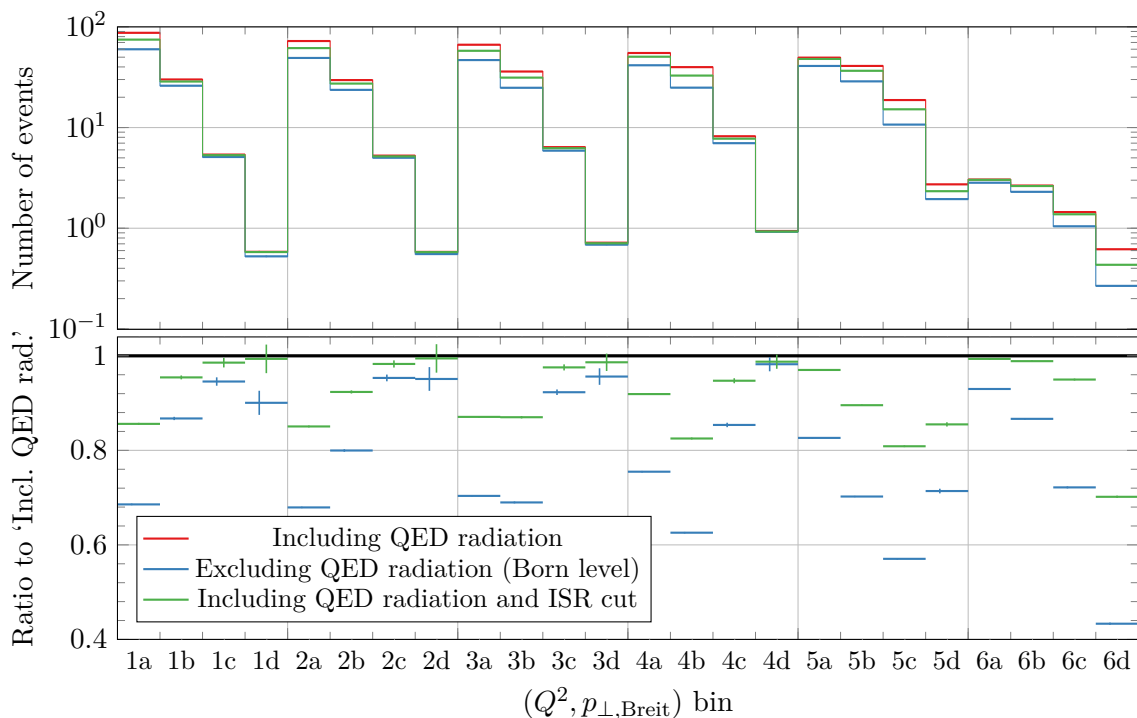
Figure 9.16: Inclusive jet predictions calculated at QED Born level and including QED radiative effects. The ratio panel gives correction factors. The numerical values for this correction are given in table C.4. Also shown are predictions including QED radiative effects and including a cut on hard ISR photons. These predictions are used for the alternative QED correction, as discussed in the text.

to treat QED radiation more rigorously, should more comprehensive theoretical predictions including these effects become available in the future. The correction factors for the QED Born level correction are provided explicitly, so it is possible to undo this correction.

The cross sections without this correction correspond to the data, including radiative corrections. These are vertex corrections and collinear initial-state, collinear final-state and non-collinear photon radiation from the electron or quark. However, not all of these effects were actually measured. At detector level, the cut $E - p_z \in (38, 65)\,\text{GeV}$ is applied, which, amongst other things, removes events with hard ISR photons from the electron. In the presence of ISR, the expected value for $E - p_z$ is $2(E_e - E_{\text{ISR}})$, where E_{ISR} is the energy of the collinear ISR photon. Therefore, the lower bound of this cut effectively amounts to a cut $E_{\text{ISR}} < 8.5\,\text{GeV}$. Consequently, events with hard ISR from the electron are absent in the data sample. These events are extrapolated from the MC during unfolding.

To reduce the dependence on the MC simulation, an alternative set of QED correction factors is determined. These factors correct the unfolded cross sections to a definition including higher-order QED effects and including the ISR cut. They are derived from the same sample that is used for the QED Born level correction and are defined as $c'_{\text{QED}} = \frac{N_{\text{With radiation and ISR cut}}}{N_{\text{With radiation}}}$. When applying these factors to the cross sections, the dependence on the MC samples cancels. This definition is slightly more challenging on the theoretical side, as this cut has to be integrated into the calculations. These alternative factors and their uncertainties are shown in figure 9.16. They are given numerically, alongside the corresponding cross sections in table C.7.

CHAPTER 10

Systematic uncertainties

Experimental systematic uncertainties of the measurement arise due to imperfect modelling of the data or the theory. This includes the description of the detector response by the detector simulation, and the description of the signal and background processes by the MC event generators. Corrections are introduced for many of these imperfections as described in the previous chapters. In these cases, the uncertainty due to imperfect modelling is replaced by a much smaller uncertainty due to an imperfect correction.

To quantify the influence of these imperfections on the measured cross sections, each known contribution or correction is varied within its uncertainty. The effect of each variation on the unfolded cross section is recorded and added up to form the total systematic uncertainty.

Using the bin-by-bin correction, the effect of each variation on the cross sections can be determined straightforwardly. The bin-by-bin correction is repeated with the modified data and MC samples, including the variation corresponding to a systematic uncertainty. The result is compared to the nominal cross sections, and the difference is used as the systematic uncertainty on the cross sections.

When using matrix unfolding, this simple approach is often not viable, as statistical fluctuations of the data and MC samples tend to be amplified and would therefore be double counted as part of the systematic uncertainty. It is possible to compensate for this effect by applying a smoothing procedure to the determined uncertainties.[145] In this analysis, a different approach is chosen. For most systematic uncertainties, instead of repeating the unfolding for the variation, the effect of the variation is propagated using bin-wise factors. In each bin

$$N_{\text{unfolded,variation}} \coloneqq N_{\text{unfolded,nominal}} \frac{N_{\text{data,variation}}}{N_{\text{data,nominal}}} \frac{N_{\text{rec,nominal}}}{N_{\text{rec,variation}}} \frac{N_{\text{gen,variation}}}{N_{\text{gen,nominal}}},$$

where N is the number of jets in a given bin, the subscripts 'data', 'rec' and 'gen' refer to data at detector level, MC at detector level and MC at hadron level respectively and $N_{\text{unfolded,nominal}}$ is the number of jets in the data at truth level, as determined by matrix unfolding using the nominal samples. Instead of repeating the matrix unfolding for each variation, $N_{\text{unfolded,variation}}$ is used. Using this formula, the relative systematic uncertainty for the matrix unfolded cross sections is the same as the relative uncertainty determined from the bin-by-bin correction.

This simplified approach is justified for most sources of systematic uncertainty since matrix unfolding is expected to have a similar uncertainty as the bin-by-bin correction. A notable exception is the model uncertainty, which is expected to decrease when using matrix unfolding. For this uncertainty, a dedicated treatment is employed for matrix unfolding. There are a few sources of systematic uncertainty, which are estimated using a reweighting of the samples. Since

events do not move between bins for these variations, statistical fluctuations will cancel when comparing the unfolded cross sections to the nominal ones. Therefore, the straightforward way to evaluate the systematic uncertainty discussed above can also be applied to matrix unfolding. In most cases, the uncertainties derived via both approaches are essentially identical.
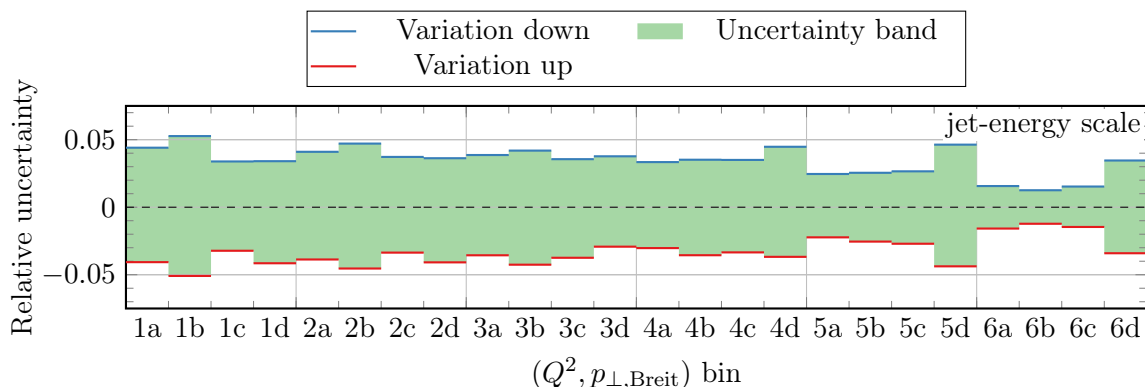
10.1 Jet-energy scale

The dominant systematic uncertainty on most jet measurements arises due to uncertainty in the description of the detector response to hadronic objects in the MC. Due to an imperfect reconstruction, the measured energy of each hadronic object does not necessarily correspond exactly to the true energy of the object. If the relation between true and measured energy is not described by the MC sample used for unfolding, there will be a systematic bias in the measured cross sections. The uncertainty in the description of this relation is called the jet-energy-scale uncertainty. Since the jet spectrum is steeply falling, the uncertainty on the jet-energy scale is amplified when determining its effect on the measured cross sections.

In this analysis, the jet-energy scale in the MC samples is calibrated as described in section 8.3.1. The result of this correction is shown in figure 8.7. After the correction, data and MC agree within about 1%. Therefore, a value of 1% is taken as the uncertainty on the jet-energy scale. The correction is applied to jets with $p_{\perp,\mathrm{lab}} > 10\,\mathrm{GeV}$. It has been shown in a previous analysis that for lower-energy jets, the uncertainty is 3%.[80]

The effect of the jet-energy-scale uncertainty on the measured cross sections is determined by repeating the unfolding with two modified MC samples. In these modified samples, the energies of the jets after the relative jet-energy calibration are modified up or down by 1% for jets with $p_{\perp,\mathrm{lab}} > 10\,\mathrm{GeV}$ and by 3% for jets of lower energy.

The resulting cross section uncertainty is given in figure 10.1. Due to the steeply falling jet spectrum, the 1% uncertainty on the jet-energy scale results in a cross section uncertainty of about 4% at low Q^2. At higher Q^2, the jet spectrum becomes more shallow, and the cross section uncertainty decreases to about 2%. A decrease in the jet energy in the MC downwards leads to an increase in the unfolded cross sections and vice versa.

Figure 10.1: Systematic uncertainty associated with the MC jet-energy scale. The entries 'Variation up/down' refer to the change of the MC jet energy, which changes the cross sections in the opposite direction. The bins are labelled according to table 9.1.

10.2 Model uncertainty

Ideally, the unfolded cross sections should not depend on the correctness of the MC event generator used for the unfolding. In practice, this is not the case since the model of the detector response used during the unfolding is necessarily a strong simplification of the complete detector response and integrates over a large number of MC distributions. If the MC does not describe the data in these distributions, the unfolded cross sections will be biased by the MC event generator. Many selection criteria and corrections applied in the previous chapters, most notably the jet reweighting from section 8.3.3, aim to make the MC samples more similar to the data and, thereby, reduce this bias.

Quantifying the remaining bias is the purpose of the model uncertainty. In this analysis, this is done by comparing the Ariadne and Lepto signal MC samples. Most importantly, this variation accounts for the uncertainty from the modelling of the parton-showering. At ZEUS, this uncertainty is considered the dominant contribution to the model uncertainty.[191,198,199] This approach also accounts for uncertainties in any other parameter that differs between the two samples. However, since all of these uncertainties are considered simultaneously, some may partially cancel. Also, the uncertainties in all parameters that are identical in both samples are neglected. For this thesis, these effects are considered subdominant. Some of them are accounted for in the other systematic uncertainties.

When the unfolding is performed using a bin-by-bin correction, each bin is unfolded independently, meaning that all distributions in the MC are integrated over. In the matrix unfolding approach, the Q^2 and $p_{\perp,\mathrm{Breit}}$ distributions are considered explicitly, which leads to a significantly reduced dependence on the correctness of these two distributions. After reweighting the MC samples to the data, this effect is expected to be reduced but still observable. Therefore, a dedicated treatment of the model uncertainty is necessary for the bin-by-bin correction and for matrix unfolding.

10.2.1 For the bin-by-bin correction

Cross sections are derived using the combination of the Ariadne and Lepto signal MC samples, see section 9.3.2. The model uncertainty is estimated by repeating the cross section determination twice, once using only the Ariadne sample and once using only Lepto. The difference between the nominal set of cross sections and these two alternative sets is then taken as the model uncertainty. Since the nominal cross sections were derived using a combination of the two samples, this model uncertainty will be approximately symmetric.

The relative uncertainty resulting from this procedure is shown in the first panel of figure 10.2. The individual components from each MC sample are also shown.

10.2.2 For matrix unfolding

When using matrix unfolding, the nominal cross section is defined similarly as for the bin-by-bin correction, i.e. the two MC samples are combined to determine the migration matrix used for unfolding. However, to evaluate the model uncertainty for the matrix unfolding, the approach described above cannot be easily used, as it would amplify and double count statistical fluctuations in the data, leading to a significant overestimation of the model uncertainty. Instead, the model uncertainty is estimated from a MC study. The idea is to perform a closure test, in which one MC sample is taken as pseudo-data and unfolded with the other sample. The
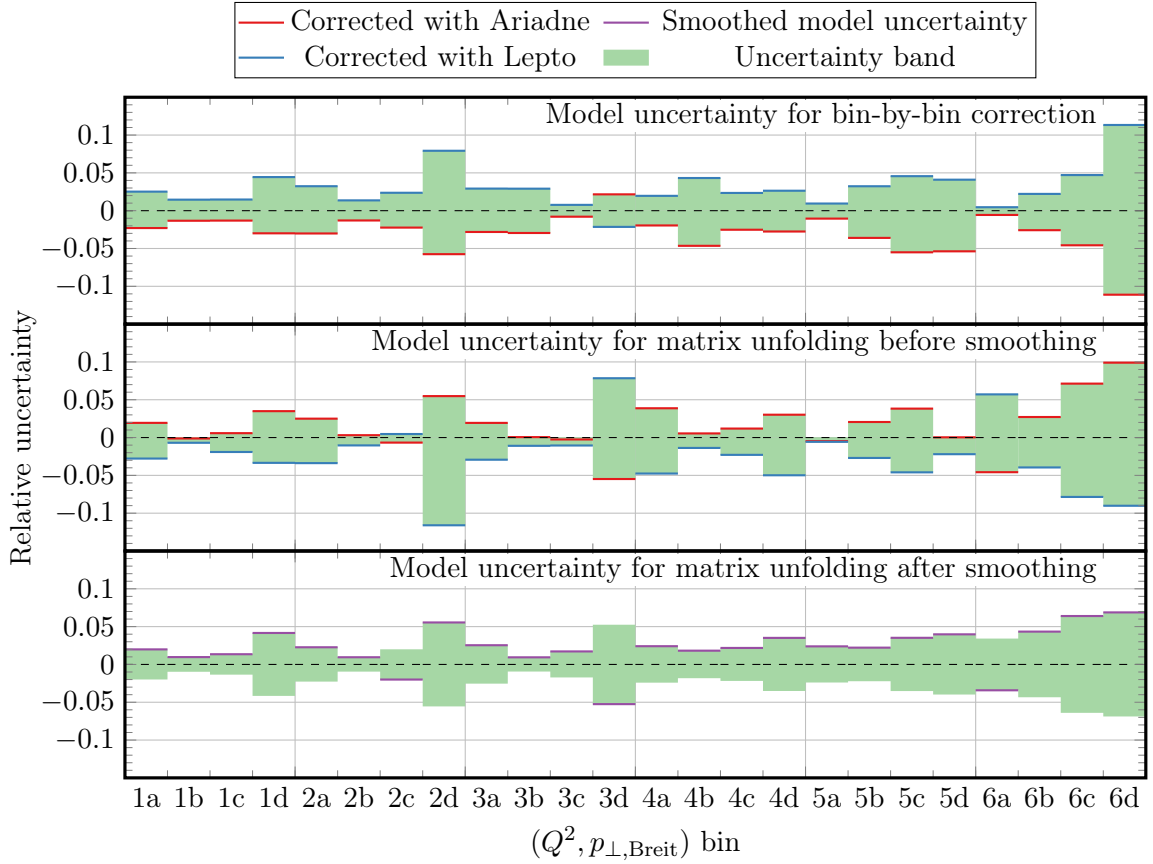
Figure 10.2: Systematic model uncertainty determined for the bin-by-bin correction (top panel) and matrix unfolding (middle and bottom panel). In the first two panels, the individual contributions from each unfolding step are also shown as red and blue lines at the edges of the uncertainty band.

unfolded distribution should agree with the corresponding truth distribution within statistical and model uncertainty. This relation can be used to extract the model uncertainty.

The closure test can be performed in two ways. Either the Ariadne or Lepto MC sample can be used as pseudo-data and unfolded with the other sample. Both of these are performed, and each of them gives the one-sided model uncertainty. The resulting model uncertainty will be approximately symmetric.

The unfolded and truth distributions should agree within statistical and model uncertainties. Therefore, the uncertainty is given by the difference between these two distributions. In each bin and for both ways to perform the closure test, the model uncertainty is given by

$$\Delta_{\text{stat.+model}} := \frac{N_{\text{unfolded}} - N_{\text{truth}}}{N_{\text{unfolded}}},$$

where N_{unfolded} and N_{truth} are the number of jets in the unfolded and truth distributions of the pseudo-data and $\Delta_{\text{stat.+model}}$ is the relative statistical and model uncertainty. This bin-wise approach can only be an approximation, as it neglects correlations of the unfolding uncertainty.

The model uncertainty Δ_{model} is assumed to be a constant fraction of the combined statistical and model uncertainty

$$\Delta_{\text{model}} := f \Delta_{\text{stat.+model}}.$$

The factor f should be between 0 and 1. To determine f, the unfolded distribution is compared to the truth distribution within the correlated statistical and the uncorrelated model uncertainty by computing the χ^2 per degree of freedom.[196] The factor f is adjusted until a χ^2 per degree of freedom of 1 is achieved. This procedure results in values of $f = 0.90$ when using Lepto as pseudo-data and $f = 0.76$ when using Ariadne as pseudo-data. As expected, these factors are only slightly less than 1, indicating that the model uncertainty is larger than the statistical uncertainties of the MC samples.

Afterwards, the model uncertainty is divided by a factor of 2. This is because the model uncertainty is defined to be half the difference between the Ariadne and Lepto samples, analogously to the approach used for the bin-by-bin correction. In other words, each of the two variants of the closure test determines half of the model uncertainty.

The resulting model uncertainty is shown in the second panel of figure 10.2. It can be observed that there are a few bins in which the uncertainty becomes very small. This is because the results from the two closure tests can cross. This would be acceptable for a correlated systematic. As discussed in section 12.1.1, the model uncertainty will be treated as partially uncorrelated. Therefore, bins of zero uncertainty are not acceptable, as they underestimate the uncertainty. Another problem is that this distribution often changes rapidly between neighbouring bins, whereas a physical distribution is expected to be smooth. This is probably due to additional fluctuations that are amplified.

To address both of these issues, a smoothing procedure is applied, which increases the uncertainty near the cross-over points and dampens close-range fluctuations. To simplify this procedure, the distribution is symmetrised. Afterwards, the uncertainty in each bin i is replaced by

$$\Delta_{\text{model,smoothed},i} := \frac{2\Delta_{\text{model},i} + \sum_j \Delta_{\text{model},j}}{2 + \sum_j 1},$$

where j runs over the nearest neighbouring bins of i in Q^2 and $p_{\perp,\text{Breit}}$. There are four nearest neighbours for bins in the central region of the phase space, three for bins at the edges and two for bins in the corners of the phase space. The smoothing procedure will necessarily distort the shape of the uncertainty. To reduce this distortion and preserve more of the original shape, the central bin is given double weight.

The smoothed model uncertainty is given in the third panel of figure 10.2. There are no more bins of zero uncertainty, and fluctuations are significantly reduced. The sign of the uncertainty is added back after smoothing, such that the uncertainty can be treated as partially correlated. Comparing this distribution to the model uncertainty determined for the bin-by-bin correction, one finds that the two have a similar shape and size. As expected, in most bins, the uncertainty is very slightly reduced using matrix unfolding, indicating a reduced dependence of the unfolded cross sections on the MC event generator.

10.3 Jet reweighting

Related to the model uncertainty is the uncertainty on the jet reweighting, which was described in section 8.3.3. This correction aims to make the MC samples more similar to the data,
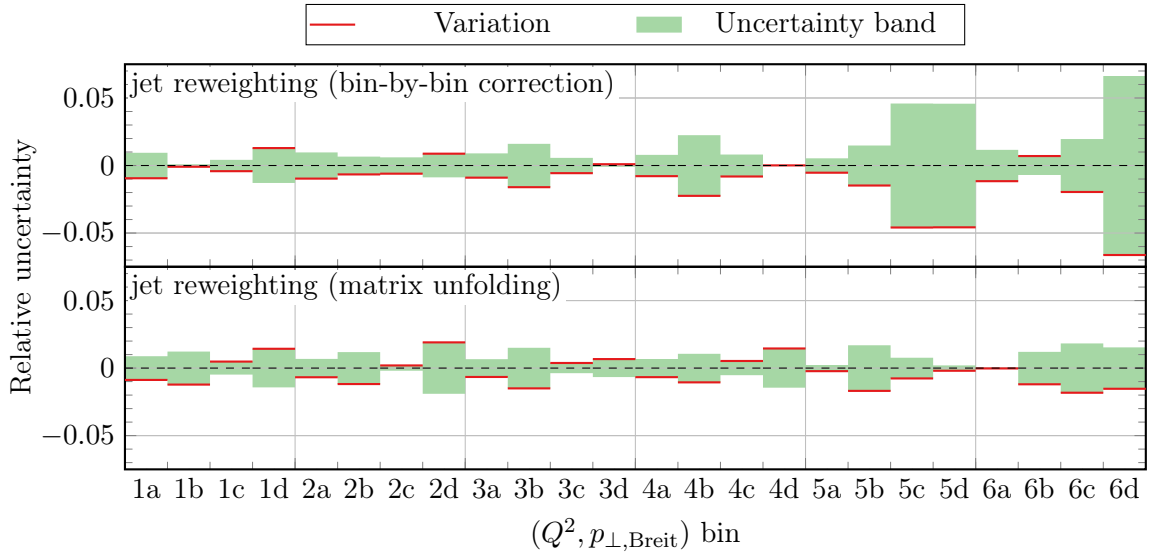
Figure 10.3: Systematic uncertainty associated with the jet reweighting correction of the MC sample using the bin-by-bin correction (top panel) and for matrix unfolding (bottom panel).

thereby reducing the dependence of the unfolded cross sections on the MC model. Also, if the MC models fail to describe the data reasonably, the approach described in the previous section to evaluate this model dependence is no longer applicable.

As described in the corresponding section, two approaches to jet reweighting are considered: multijet reweighting and inclusive jet reweighting. Both methods show similar performance, and therefore, it is justified to use the difference between the two as an estimate for the systematic uncertainty remaining after this correction. Since this uncertainty is also related to the model dependence of the unfolded cross sections, it is expected to be reduced when using matrix unfolding instead of the bin-by-bin correction. Therefore, this systematic uncertainty is derived separately for the two unfolding methods. Since this correction is just a reweighting, no special treatment is needed to reduce fluctuations. The results are shown in figure 10.3. In both cases, the uncertainty is typically below 1.5%, but for the bin-by-bin correction, a significant increase can be observed in the high-Q^2, high-$p_{\perp,\mathrm{Breit}}$ region.

10.4 Electron uncertainties

Electron-energy scale   The electron-energy-scale uncertainty arises for similar reasons as the jet-energy-scale uncertainty but concerning the energy measurement of the DIS electron. In a previous analysis, the uncertainty on the description of the detector response regarding the electron energy has been estimated to be 2%.[80]

This analysis uses almost exclusively the electron energy from the double-angle method, which is approximately independent of any bias in the calorimeter energy measurement. The electron-energy measurement from the calorimeter is used only for the corresponding cut as part of the DIS selection, see section 7.2.2. Consequently, the influence of this uncertainty on the cross sections is minimal, as is depicted in the top panel of figure 10.4.

Electron identification   At ZEUS, two different algorithms are commonly used to identify electrons: EM and Sinistra, see section 6.2. The EM electron finder is used in this analysis
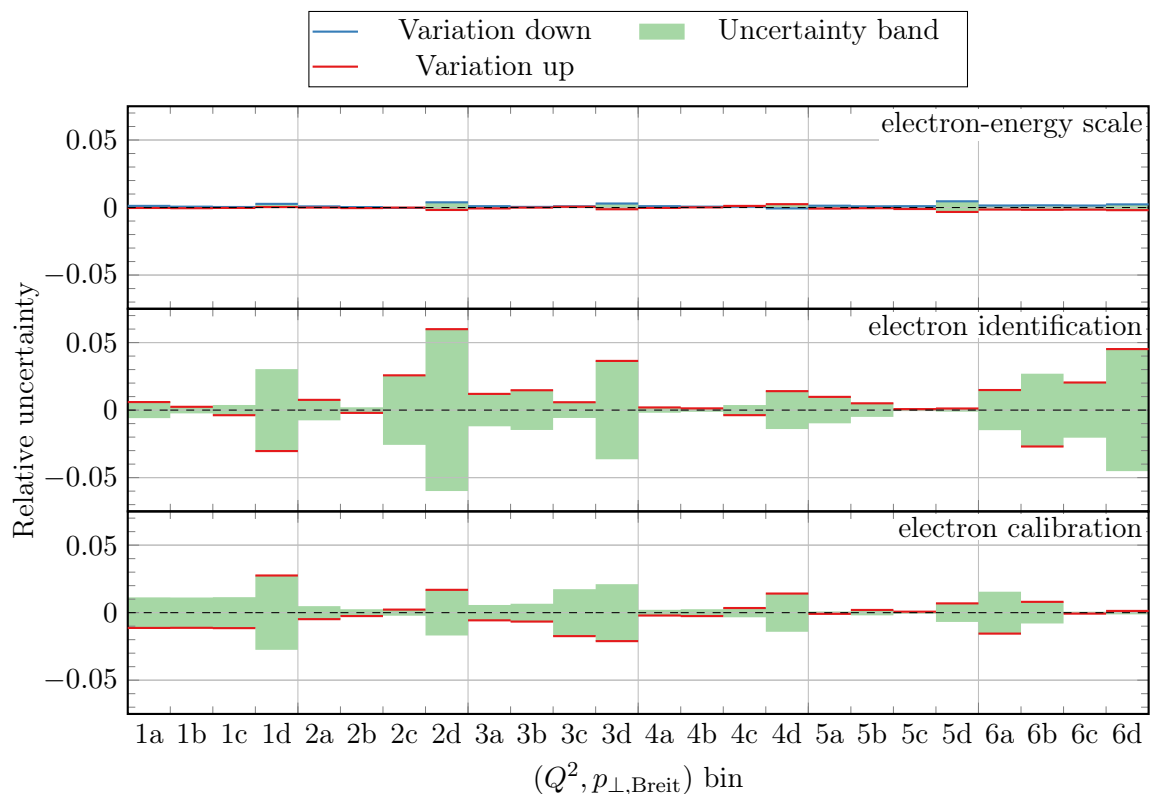
Figure 10.4: Systematic uncertainty associated with the MC electron energy scale (top panel), the electron finding algorithm (middle panel) and the electron calibration (bottom panel).

since it is better optimised for the very high-Q^2 region. The systematic effect of this choice is evaluated by repeating the cross section determination using Sinistra instead of EM, i.e. all quantities involving the DIS electron are determined using Sinistra instead of EM. Correction factors are not rederived for this study. This procedure results in a conservative estimate of the uncertainty, as it is known that Sinistra is less suited for parts of the analysis region. The determined uncertainty is given in the middle panel of figure 10.4. In most bins, the effect is around 1.5%, but the distribution is subject to fluctuations up to 6%. This uncertainty is assumed to cover also the uncertainty associated with choosing the double-angle method for kinematic reconstruction.

Electron-energy calibration In section 8.2.4, a calibration procedure to the electron energy was introduced, which aims to bring the measured energy of the electron in data and MC closer to the true energy. A modified correction procedure is investigated to estimate the systematic uncertainty due to this correction. Instead of applying the correction factors as a two-dimensional function, they are applied only as a one-dimensional function of θ. The effect on the cross section is shown in the bottom panel of figure 10.4. In most bins, the effect is around 1% or less.

10.5 Cut boundaries

To improve the quality of the data sample and its agreement to the MC simulations, a series of cuts is introduced in chapter 7. The exact numerical values of these cuts are, to some extent, arbitrary. To evaluate the systematic effect of these arbitrary choices, the values of the cut boundaries are varied in both data and MC within the uncertainties of the corresponding variables. If the MC samples describe the data well in the regions close to the cut boundaries, these variations are expected to have little impact on the cross sections.

The effect of the variation of a quality cut on the cross sections can be roughly estimated as

(Relative change of number of events in data) − (Rel. change of number of events in MC).

For example, changing the boundary of the electron isolation cut from 0.1 to 0.2 would increase the number of events by 1.3% and 1.2% in data and MC respectively, indicating that there is an ∼ 8% disagreement in this region, see figure 7.4. Thus, the expected change of the cross section is 0.1%, which is negligible compared to the remaining systematic uncertainties. This formula shows that the uncertainty on cuts can be neglected if the fraction of affected events is sufficiently small or if data and MC agree close to the cut boundary. This simple approximation assumes that the affected events are distributed similarly to the remaining events and that the fractional changes in the number of events are small and constant throughout the phase space. A more thorough estimation of the uncertainty is necessary for cuts whose influence is potentially large in some bins.

The following cut variations are identified as potentially large and are investigated in more detail.

Variable		Original value	Variation
Jet momentum	$p_{\perp,\mathrm{lab}}$	$> 3\,\mathrm{GeV}$	$\pm 1\,\mathrm{GeV}$
Electron-track momentum	p_{track}	$> 3\,\mathrm{GeV}$	$\pm 1\,\mathrm{GeV}$
Longitudinal-momentum balance	$E - p_{\mathrm{z}}$	$\in [38, 65]\,\mathrm{GeV}$	$\pm 6\%$
Transverse-momentum balance	$p_{\perp}/\sqrt{E_{\perp}}$	$< 2.5\,\sqrt{\mathrm{GeV}}$	$\pm 0.5\,\sqrt{\mathrm{GeV}}$
Primary-vertex position	$|z_{\mathrm{vertex}}|$	$< 30\,\mathrm{cm}$	$\pm 5\%$
RCAL radius of electron	r_{RCAL}	$< 175\,\mathrm{cm}$	$\pm 2\,\mathrm{cm}$
Electron-track distance	DCA	$< 10\,\mathrm{cm}$	$< 8\,\mathrm{cm}$

For the downward variation of the $E - p_{\mathrm{z}}$ cut, both boundaries are moved downwards simultaneously and similarly for the upwards variation. The electron-energy-scale uncertainty covers the uncertainty on the electron-energy cut E_{EL}.

The resulting uncertainties on the cross section are given in figure 10.5. For most variations and bins, the effect on the cross section is well below 1%. The only notable exception is the high-$p_{\perp,\mathrm{Breit}}$ region of the $E - p_{\mathrm{z}}$ variation, where the effect can reach up to 6%.
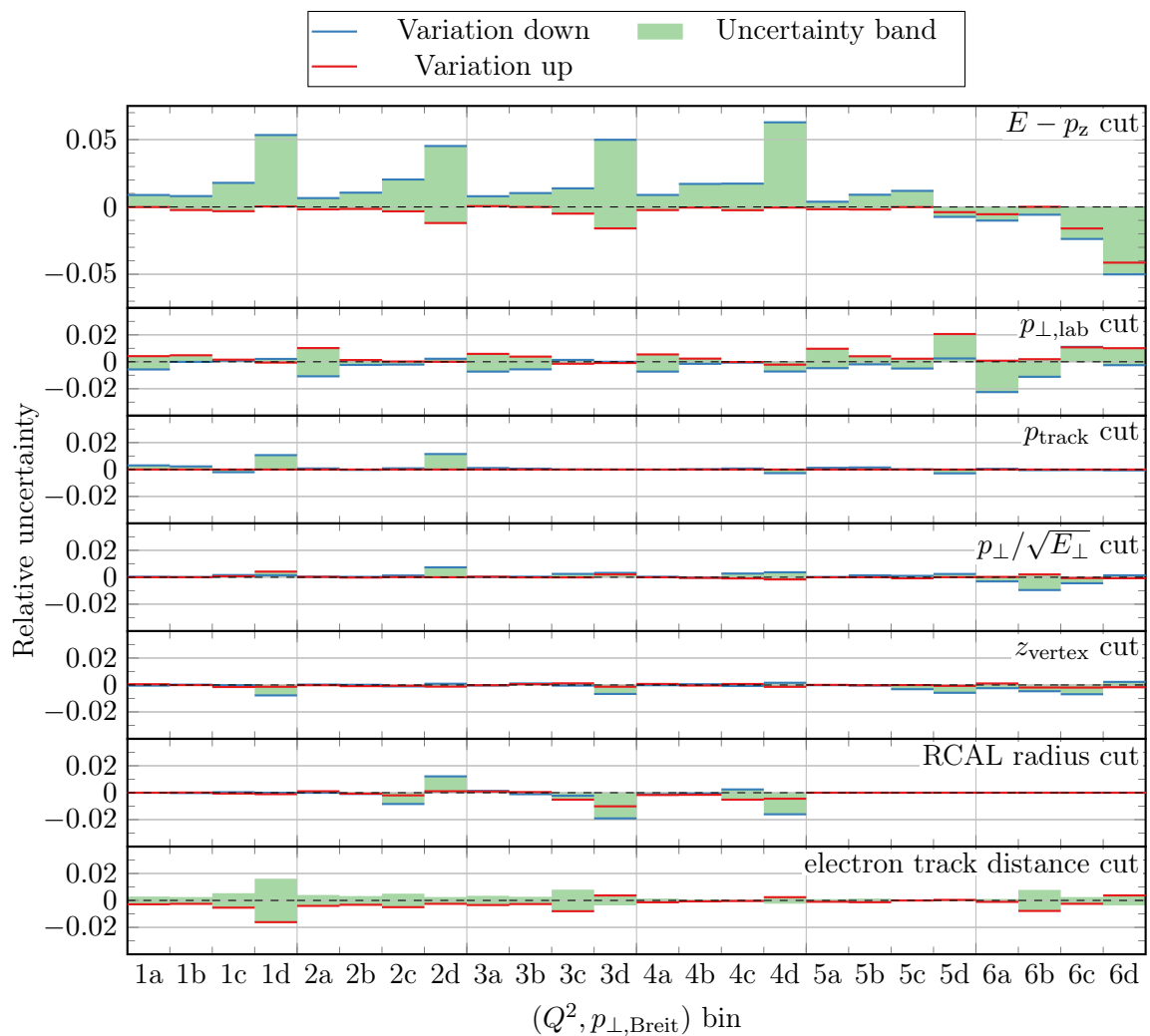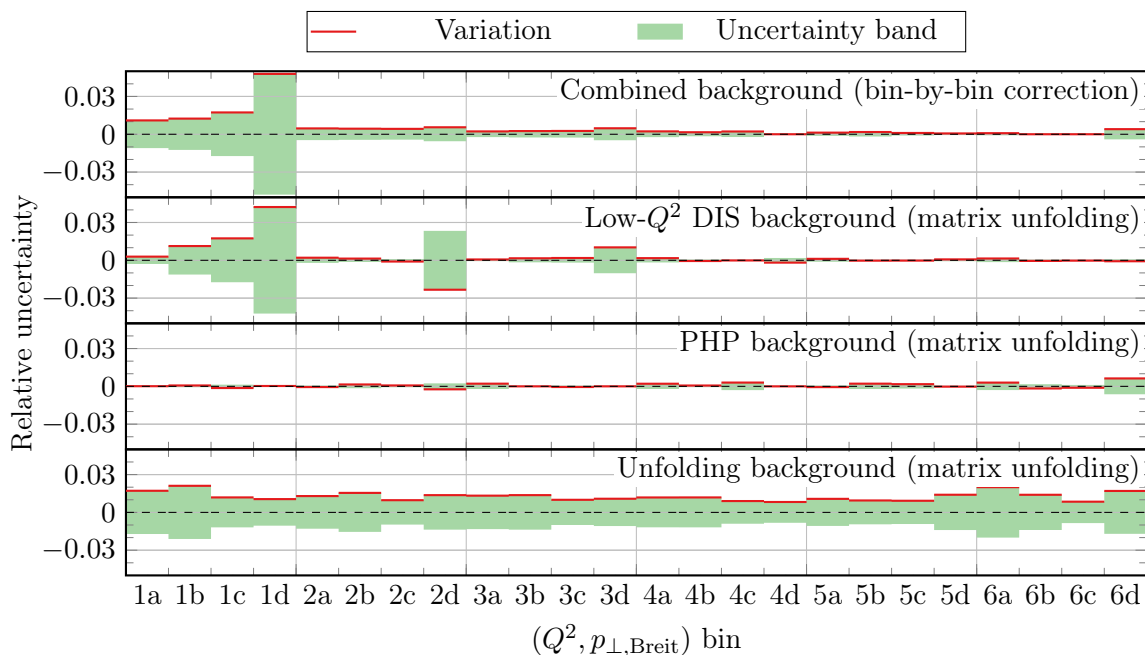
Figure 10.5: Systematic uncertainties associated with the exact numerical value of the quality cuts. The labels 'Variation up/down' refer to the variation of cut value, which does not necessarily change the cross section in the same direction.

10.6 Background contribution

The unfolding procedure assumes that the MC sample describes the data. However, the samples used for the unfolding in this analysis only describe the DIS signal process. During the event selection, several cuts were introduced to reject events from other processes, but a small fraction of these events is expected to remain in the final sample. These background contributions are estimated using additional MC samples. Before unfolding, the estimated background distributions are subtracted from the data distributions.

The MC samples to describe the low-Q^2 DIS and PHP background contributions are introduced in section 5.3.2 and compared to the data and the signal MC distributions in section B.3. The normalisation uncertainty on these samples is conservatively estimated to be 50%. For the bin-by-bin correction, the systematic uncertainty is determined by repeating the unfolding without subtracting the background and taking half of the effect on the cross section as a systematic uncertainty. The uncertainty is shown in figure 10.6.

Figure 10.6: Systematic uncertainties associated with the contribution of background events. The first panel shows the combined uncertainty due to PHP and low-Q^2 DIS events as used for the bin-by-bin correction. The second and third panels show the individual uncertainties due to PHP and low-Q^2 DIS events as used for matrix unfolding. The last panel shows the additional uncertainty due to matrix unfolding, which arises due to jets that are not part of the response matrix.

For matrix unfolding, the uncertainties due to these processes are considered separately and shown in the second and third panels of the figure. The low-Q^2 DIS background uncertainty has a similar shape and size to the combined uncertainty from the bin-by-bin correction. The contribution from PHP events is very small.

When using matrix unfolding, an additional type of background contribution arises. In section 9.4.2, this type of background is referred to as 'unfolding background'. It refers to jets that contribute to the detector-level distribution but cannot be matched to any hadron-level jet in the phase space, i.e. due to misreconstructions or migrations from outside the phase space. This contribution is estimated using the signal DIS MC samples. The normalisation uncertainty on this background is estimated to be 5% by comparing the data and MC distributions and varying the normalisation of the contribution from unmatched jets. The resulting systematic uncertainty is shown in the final panel of figure 10.6. The uncertainty is approximately constant and has a typical size of 1.5%.

10.7 Other sources

Other corrections   In section 8.1 and 8.2, corrections were introduced for the polarisation, vertex-position, track-matching efficiency and track-veto efficiency. The effects of the first two of these corrections on the unfolded cross sections are very small. Therefore, the systematic uncertainty is estimated conservatively by repeating the unfolding without these corrections and taking half of the resulting change of the cross sections as an uncertainty. The track-
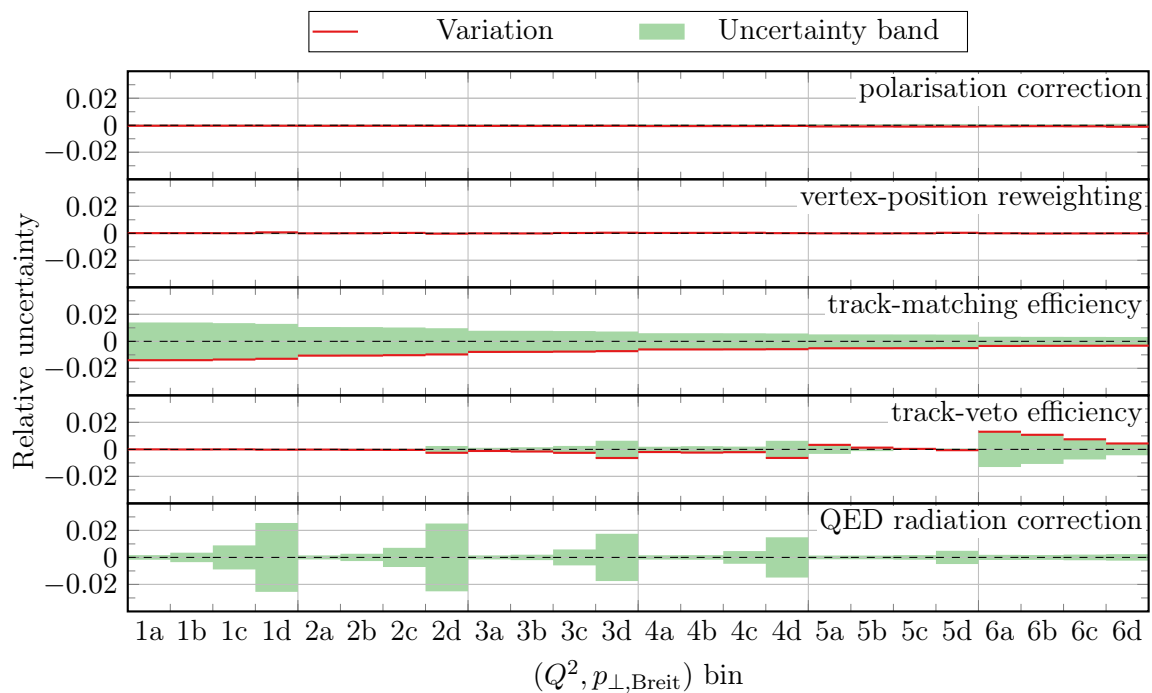
Figure 10.7: Systematic uncertainties associated with the polarisation, vertex position, track-matching efficiency, and track-veto efficiency correction using bin-by-bin correction. The final panel shows the statistical uncertainty on the QED radiation correction.

matching correction is applied halfway. The entire size of this halved correction is used as uncertainty to account for the effect of possibly applying the correction completely or not at all. The systematic uncertainty on the trigger-efficiency correction is estimated by applying the correction as a function of the electron inelasticity y instead of the FLT track multiplicity. The resulting systematic uncertainties are shown in figure 10.7.

As all these corrections rely on reweighting the data or MC samples, they can also be straightforwardly evaluated using matrix unfolding. The determined uncertainties are essentially identical for both methods.

QED radiation correction As described in section 9.5.3, an additional MC sample is generated to correct the unfolded cross sections to QED Born level. The statistical uncertainty on this MC samples is taken as a systematic uncertainty on the data. A sufficient amount of MC events are generated to make this uncertainty subdominant. The resulting uncertainty is also given in figure 10.7. Apart from this statistical uncertainty, no systematic uncertainty is assigned to this correction, since QED effects are well understood.

Luminosity Each run period has a luminosity uncertainty between 1.8% and 2.5% associated with it.[200] The luminosity weighted average of these uncertainties is 1.9%, which is taken as the luminosity uncertainty on this measurement. By ZEUS convention, the luminosity uncertainty is not shown in the cross section figures due to its fully correlated nature.

10.8 Combined uncertainty

The total uncertainty and the contributions from each of the sources listed above are given in figure 10.8 for matrix unfolding and the bin-by-bin correction. The jet energy scale is the dominant systematic uncertainty in most parts of the phase space, especially the low-Q^2 and low-$p_{\perp,\mathrm{Breit}}$ regions. Another significant contribution in the high-Q^2 region comes from the model uncertainty and, in the case of the bin-by-bin correction, from the jet reweighting uncertainty. In the high-$p_{\perp,\mathrm{Breit}}$ region, the data are less well understood due to a lack of statistics. This leads not only to an increase in the statistical uncertainty but also in the electron uncertainties and the variations of the cut boundaries. The combined systematic uncertainty is usually around 5%, except in the high-$p_{\perp,\mathrm{Breit}}$ region, where it increases to about 10%. The numerical values of all systematic uncertainties are listed in appendix C.
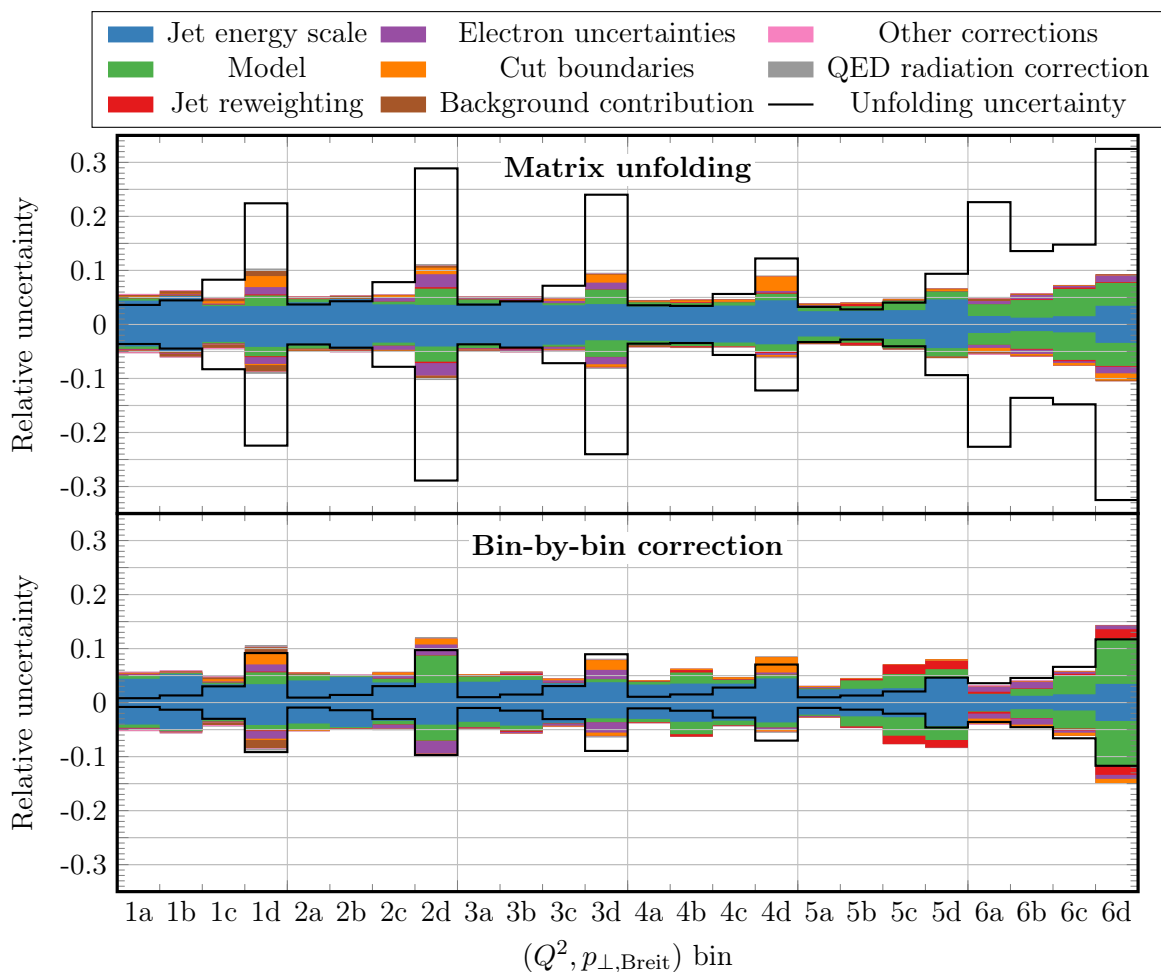
Figure 10.8: Total systematic uncertainty and the individual contributions from each source. The bins are labelled according to table 9.1.

CHAPTER 11

Cross sections

In the previous two chapters, event counts and their uncertainties have been estimated at truth level in the data. By normalising these counts to the luminosity of the measurement, the detector-independent cross sections can be derived, which are the core result of this thesis.

11.1 Cross section definition

The cross section definition is given and motivated in the previous chapters. For convenience, it is repeated here in its entirety.

The cross sections refer to NC DIS processes in unpolarised $e^{\pm}p$ collisions. Protons have an energy of $E_P = 920\,\mathrm{GeV}$ and electrons and positrons of $E_e = 27.5\,\mathrm{GeV}$, leading to a centre-of-mass energy of $\sqrt{s} = 318\,\mathrm{GeV}$. Events are required to belong to the kinematic range $150\,\mathrm{GeV}^2 < Q^2 < 15000\,\mathrm{GeV}^2$ and $0.2 < y < 0.7$. Jets are considered in the phase space region of $7\,\mathrm{GeV} < p_{\perp,\mathrm{Breit}} < 50\,\mathrm{GeV}$ and $-1 < \eta_{\mathrm{lab}} < 2.5$.

The jets are constructed at hadron-level from photons, hadrons and charged leptons with a lifetime of more than $10\,\mathrm{ps}$, excluding neutrinos, see section 7.4. The selected particles are made massless by rescaling their three-momentum to match their energy. The massless four-vectors corresponding to each particle are boosted to the Breit frame of reference. In this frame, jets are constructed using the $k_{\perp}$-algorithm with a distance parameter of $R = 1$. The $p_{\perp}$-weighted scheme is used to combine particles, which leads to massless jets.

The quoted cross sections are defined as the average of e^-p and e^+p cross sections in a ratio of $6:5$. The lepton type affects the cross sections at high Q^2, where Z boson exchange and photon-Z interference become relevant. Correction factors from a definition excluding weak effects are given in table C.4.

Cross sections at QED Born level   Cross sections are provided in two variants. In the first variant, they are defined at QED Born level, meaning they exclude higher-order QED effects, such as initial-state radiation from the electron or quark. In this definition, loops of virtual particles are still considered via the scale dependence of the QED coupling. Table C.4 lists cross sections and the correction factors applied to the data to achieve this definition.

Cross sections including QED radiation   It is desirable to consider the data without the Born level definition. A second variant of cross sections is provided, in which this correction is not applied. Instead, a cut $E_{\mathrm{ISR}} > 8.5\,\mathrm{GeV}$ is applied on the energy of the ISR photon from

the electron line, see section 9.5.3. When considering events including QED radiation from the electron, the kinematic quantities are defined using the electron method, see section 6.5. Table C.7 presents the cross sections in this alternative definition.

11.2 Theoretical predictions

The full potential of an experimental measurement can only be exploited when confronting it with sufficiently precise theoretical predictions. Fixed-order predictions for inclusive jet production in the Breit frame in NC DIS with a photon as exchanged boson are available at NNLO QCD accuracy, as calculated using the NNLOJET program.[159] Like most NNLO calculations, this program is effectively a parton-level event generator and thus gives access to the complete kinematic information of all final-state partons. This allows computing cross sections corresponding precisely to the experimental definition, including the exact jet definition and all phase space cuts.

Cross sections are computed using the HERAPDF2.0Jets NNLO PDF set.[103] The strong coupling constant and the factorisation and renormalisation scales were set to $\alpha_s(M_Z^2) = 0.1155$ and $\mu_f^2 = \mu_r^2 = Q^2 + p_{\perp,\text{Breit}}^2$, as were used in the determination of the PDF set. The calculation is available at QED Born level, so the kinematic quantities are unambiguously defined according to equation (2.1). The final-state partons are considered in the Breit frame, where the same jet clustering algorithm is used that was applied to the data in section 7.3. Due to the relatively small number of partons produced in the fixed-order calculation, jets will often consist of just a single parton.

The NNLO calculation is only available in the zero-mass variable-flavour-number scheme. Massless jets are constructed to minimise inconsistencies with the massless quarks in this scheme. For consistency, massless jets are also constructed in the data, even though the real quarks are massive. The difference between massive and massless quarks is not further considered. It is often assumed that this effect is small, though this has not been proven.[103]

To compare the parton-level jets from the fixed-order calculation, the hadronisation correction described in section 5.2.2 is applied to determine hadron-level cross sections. Furthermore, the calculation only takes photon exchange into account. The weak-boson correction from section 9.5.2 is applied to account for contributions from Z boson exchange and photon-Z interference. After applying this correction, the cross sections correspond to the specific composition of electron-proton and positron-proton data that is present in the data.

The uncertainty of the theory predictions consists mainly of the PDF and scale uncertainty. The PDF uncertainty is computed as described in the corresponding publication.[103] This means that the cross section uncertainty resulting from the fit and model parameters is added in quadrature, separately in the positive and negative directions. The parameterisation uncertainty is computed as the envelope of the cross section values computed with different PDF parameterisations. The scale uncertainty is estimated by performing a six-point variation, in which the factorisation and renormalisation scales are varied up and down by factor two, both independently and simultaneously.[159] Lesser contributions to the uncertainty arise from the statistical and systematic uncertainty of the numerical calculations and the uncertainty of the hadronisation correction.

11.3 Cross section results

Having now discussed all the individual steps of a cross section measurement mentioned in the introduction, cross sections can finally be derived using equation (1.1) from the event counts and the luminosity. Before being comparable to the theory predictions, the QED radiation correction from section 9.5.3 is applied to the data. This changes the cross section definition to QED Born level, which is used in the theoretical predictions.

11.3.1 Cross sections at QED Born level

The derived cross sections are presented in figure 11.1. The cross sections presented here are shown in the matrix unfolded and the bin-by-bin corrected variant. They are compared to the corresponding measurement from the H1 collaboration[121] and the NNLO QCD predictions. All cross sections are well compatible with each other, and some interesting conclusions can be drawn. The numerical information about the ZEUS cross sections, their uncertainties and correlations is given in appendix C.1 for the bin-by-bin correction and in appendix C.2 for matrix unfolding.

The uncertainties of the ZEUS measurement with the bin-by-bin correction appear significantly smaller than the ZEUS measurement using matrix unfolding and the H1 measurement, which also uses matrix unfolding. Especially the unfolding uncertainty is notably smaller. This difference is because the unfolding uncertainty of the matrix unfolded cross sections is anti-correlated between bins, as was discussed in section 9.4.2. Since the figure only shows the diagonal elements of the covariance matrix, the uncertainties appear larger. This effect is difficult to visualise in the figure, but it will decrease the effective size of the uncertainty in subsequent numerical analyses of the cross sections. Furthermore, the uncertainty of the bin-by-bin corrected cross sections is weakly positively correlated, which will usually increase their effective uncertainty. The increased unfolding uncertainty of matrix unfolding is partially compensated by a decreased systematic uncertainty. This is because parts of the systematic model uncertainty are included in the unfolding uncertainty of the matrix unfolding. Combining all these effects, the cross sections derived using the two correction methods possess a similar constraining power, even though their uncertainties appear very different in the figure. This will be further demonstrated in the next chapter, where both sets of cross sections are used in a QCD analysis, and their constraining power is compared.

Due to the parity-violating nature of the weak interaction, the electron-proton and positron-proton cross sections differ at very high-Q^2. The cross sections are defined as the luminosity weighted average of the two lepton beam types. The weak-boson correction that is applied to the theory takes this ratio into account. The same definition was chosen in the H1 measurement, but their electron-positron ratio is slightly different from the one in ZEUS. This effect is considered insignificant for the figure, so it is not corrected. If adjusted to the ZEUS composition, the cross sections from H1 would increase by about 1% in the fourth and fifth Q^2 bin and by about 5% in the highest Q^2 bin.

When comparing the theoretical predictions to the two measurements, one finds that, while compatible, the measurements show similar trends relative to the theory. Overall, the predictions seem to overestimate the cross sections by a few percent. In the high-$p_{\perp,\mathrm{Breit}}$ region at medium to high Q^2, this difference seems to increase to about 20%.
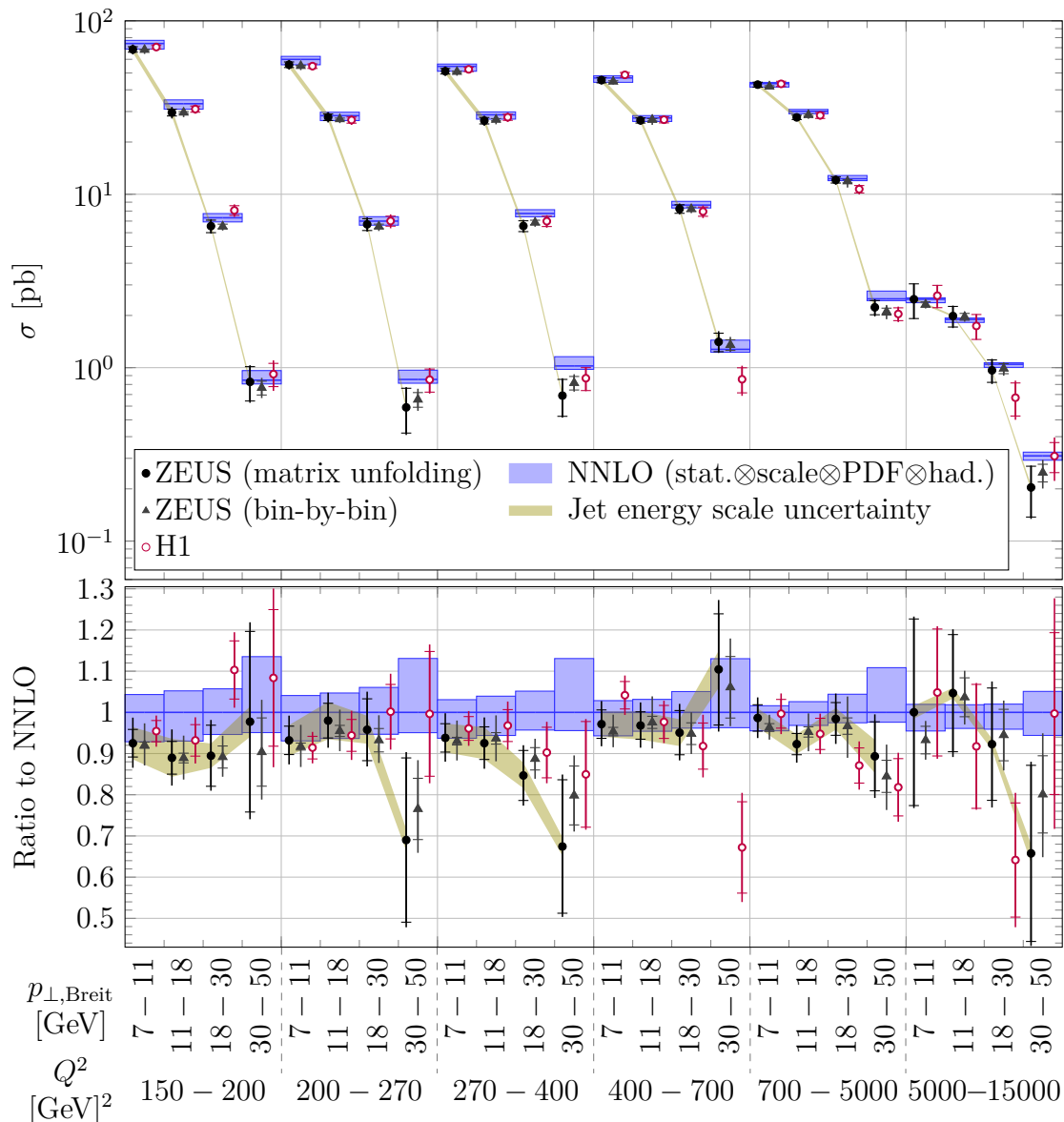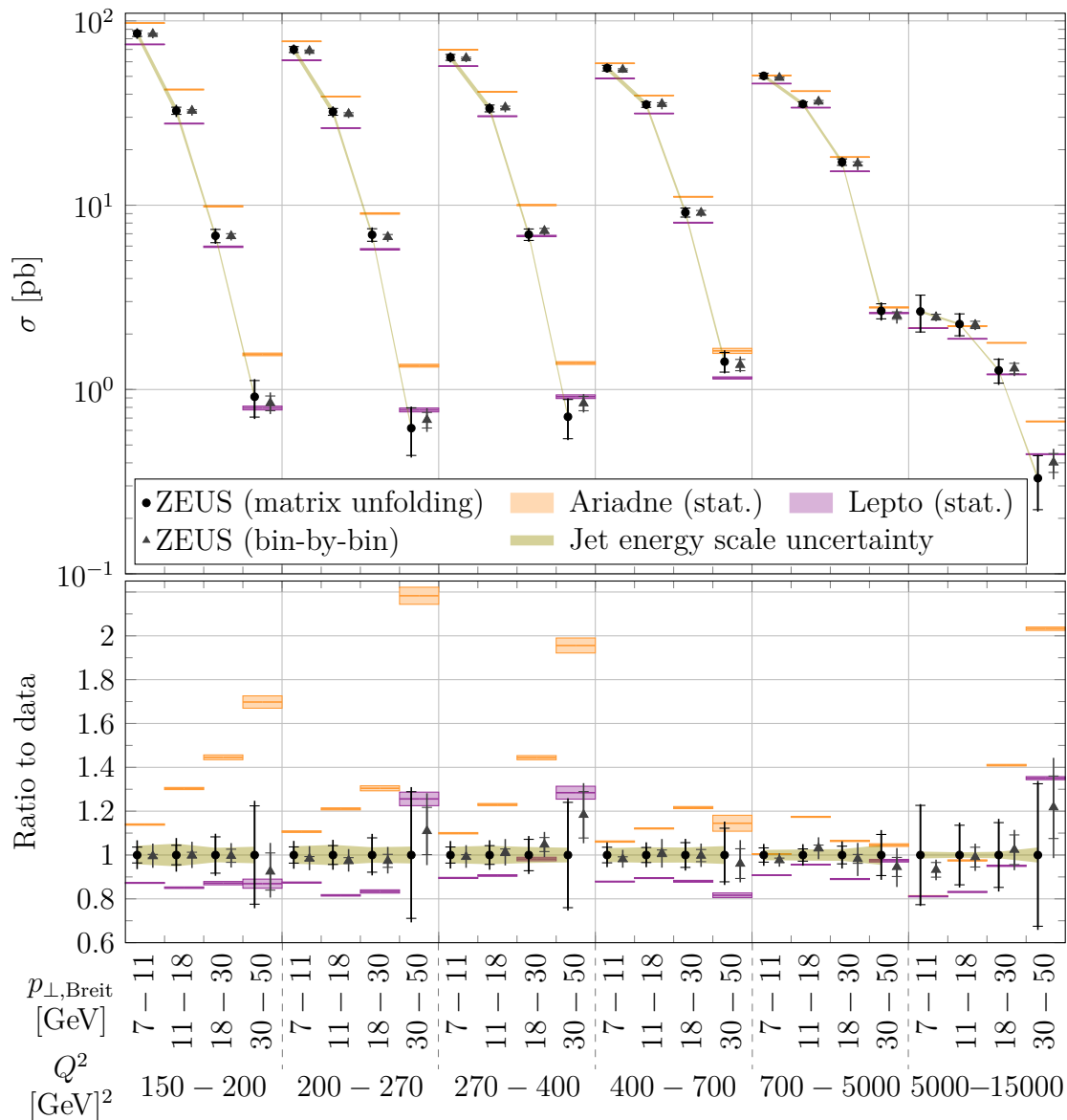
Figure 11.1: The measured double-differential inclusive jet cross sections with $7\,\text{GeV} < p_{\perp,\text{Breit}} < 50\,\text{GeV}$ and $-1 < \eta_{\text{lab}} < 2.5$, in the kinematic range $150\,\text{GeV}^2 < Q^2 < 15000\,\text{GeV}^2$ and $0.2 < y < 0.7$. Shown are the present measurement from ZEUS using matrix unfolding and using the bin-by-bin correction, the corresponding measurement from H1[121] and the NNLO QCD predictions. The inner error bars of the measurements represent the unfolding uncertainty, and the outer error bars the total uncertainty. For the ZEUS measurement with matrix unfolding, the shaded band shows the uncertainty associated with the jet energy scale. The upper panel shows the measured and predicted cross sections, and the lower panel shows the ratio of those cross sections to the NNLO QCD prediction. The data points are shifted horizontally for visual clarity.

11.3.2 Cross sections including QED radiation

Cross sections are also provided in an alternative definition, including leading-order QED radiative effects and a cut on ISR. Currently, no theoretical calculations are available for comparison. In figure 11.2, these cross sections are compared to the uncorrected predictions from the signal MC samples.

The cross sections in this definition are higher than at QED Born level. In the low- and medium-Q^2 region of the measurement, this difference decreases with increasing $p_{\perp,\mathrm{Breit}}$. At low $p_{\perp,\mathrm{Breit}}$, the cross sections, including radiation, are about 25% larger than at QED Born level. At high $p_{\perp,\mathrm{Breit}}$, this difference decreases to about 5%. This trend reverses in the highest Q^2 bin of the measurement. In the highest-Q^2 highest-$p_{\perp,\mathrm{Breit}}$ bin, the difference between the two cross section definitions is about 60%. The difference is shown in figure 9.16.

Without corrections, most importantly without the jet reweighting, the MC samples do not describe the data. The Ariadne sample predicts a more shallow jet spectrum as a function of $p_{\perp,\mathrm{Breit}}$, leading to significant deviations in the high-$p_{\perp,\mathrm{Breit}}$ region. The Lepto sample describes the shape of data much better and only shows a potential systematic deviation in the highest $p_{\perp,\mathrm{Breit}}$ bin. The data usually lies between the predictions from Ariadne and Lepto, which confirms that it is reasonable to use these two samples to estimate the model uncertainty.

Figure 11.2: The measured double-differential inclusive jet cross sections with $7\,\text{GeV} < p_{\perp,\text{Breit}} < 50\,\text{GeV}$ and $-1 < \eta_{\text{lab}} < 2.5$, in the kinematic range $150\,\text{GeV}^2 < Q^2 < 15000\,\text{GeV}^2$, $0.2 < y < 0.7$ and $E_{\text{ISR}} < 8.5\,\text{GeV}$. The present measurement from ZEUS is shown in the matrix unfolded and bin-by-bin corrected variants. The predictions from the uncorrected signal MC samples are also shown, including statistical uncertainties and normalised to the MC luminosity. The inner error bars of the measurement represent the unfolding uncertainty, and the outer error bars are the total uncertainty. The shaded band shows the uncertainty associated with the jet-energy scale for the matrix unfolded variant. The upper panel shows the measured and predicted cross sections, and the lower panel shows the ratio of those cross sections to the matrix unfolded data. The data points are shifted horizontally for visual clarity.

CHAPTER 12

Determination of the strong coupling constant

The main strength of measured cross sections is their ability to constrain theoretical models. In this thesis, the cross sections are considered within the theory of QCD. In the zero-mass scheme of perturbative QCD, this theory has a single free parameter, the strong coupling constant $\alpha_s(M_Z^2)$. To calculate cross sections in lepton-hadron interactions, one additionally requires the PDFs, which parameterise the non-perturbative regime of QCD. The inclusive jet cross sections presented in the previous chapter can constrain $\alpha_s(M_Z^2)$ and, to a lesser extent, the PDFs.

12.1 Global determination of the strong coupling constant

The PDFs $\xi f_i(\xi, \mu_f^2)$ are a set of real scalar functions. Using the DGLAP equations (2.33), the PDFs at one value of the factorisation scale μ_f^2 can be related to those at a different scale. Therefore it is sufficient to describe the PDFs at a single scale, which is referred to as the starting scale $\mu_{f,0}^2$. At this starting scale, the PDFs are parameterised in terms of a finite number of parameters.

Cross sections can then be written as a function of the PDF parameters and of $\alpha_s(M_Z^2)$. To calculate cross sections, the PDFs and $\alpha_s(M_Z^2)$ are evolved to the desired scale, where they are convolved with the matrix element from the fixed-order calculation, as is sketched in equation (2.32). The appropriateness of the chosen parameters may be evaluated by comparing the calculated cross sections against the measured ones. Thus, a non-linear fit may be used to determine the PDF parameters and $\alpha_s(M_Z^2)$ from the measured cross sections. In this thesis, this is accomplished using the XFITTER package.[201,202] This procedure is well established and has been used in many previous analyses.[77,103–105,145]

During the fit, a $\chi^2(\vec{p})$ function is minimised numerically. This function quantifies the difference between calculation and measurement as a function of the PDF parameters and $\alpha_s(M_Z^2)$, collectively referred to as $\vec{p}$. It is defined as

$$\chi^2(\vec{p}) = \min_s \left(\vec{D} \cdot \widetilde{V}^{-1} \cdot \vec{D} + \sum_k s_k^2 + \log\left(\frac{\mathrm{Det}(\widetilde{V})}{\mathrm{Det}(V)}\right) \right), \tag{12.1}$$

where

$$\vec{D} = \left(\mathbb{1} + \sum_k s_k \operatorname{diag}(\vec{\gamma}_k) \cdot \operatorname{diag}(\vec{\beta}) \right) \cdot \vec{\sigma}_{\text{data}} - \vec{\sigma}_{\text{theory}}(\vec{p})$$

$$V = V_{\text{stat}} + V_{\text{uncor}},$$

$$\widetilde{V} = \sqrt{\operatorname{diag}(\vec{\beta})} \cdot V_{\text{stat}} \cdot \sqrt{\operatorname{diag}(\vec{\beta})} + \operatorname{diag}(\vec{\beta}) \cdot V_{\text{uncor}} \cdot \operatorname{diag}(\vec{\beta}),$$

$$\beta_i = \frac{\sigma_{\text{theory},i}(\vec{p})}{\sigma_{\text{data},i}},$$

and $\operatorname{diag}(\vec{x})$ is the diagonal matrix whose entries are the elements of $\vec{x}$. The measured and calculated cross sections are given by $\vec{\sigma}_{\text{data}}$ and $\vec{\sigma}_{\text{theory}}(\vec{p})$. The covariance matrix of the unfolding uncertainty of the data is given by V_{stat}, and the uncorrelated systematic uncertainty is given by the diagonal matrix V_{uncor}. The first term in equation (12.1) is the main term of the expression. It evaluates the difference between measurement and calculation, weighted by the inverse of the unfolding and uncorrelated systematic uncertainty. Correlated systematic uncertainties are technically taken into account as shifts in the data. For each correlated uncertainty k, a vector $\vec{\gamma}_k$ expresses how the measurements change under that uncertainty. Each correlated systematic uncertainty is added to the measurements, scaled by a shift s_k. These shifts are referred to as nuisance parameters, meaning that they are additional parameters that are optimised during the fit but that are not the main point of interest. The second term in (12.1) ensures that the shifts cannot get too large. The minimisation concerning these shifts is performed analytically, such that the numerical stability of the fit is not affected by the presence of a large number of shifts. The β_i in the formula account for the fact that uncertainties should not scale with the measured value of the cross section, but rather with the true one, which is approximated in the fit by $\vec{\sigma}_{\text{theory}}(\vec{p})$. Systematic uncertainties should scale linearly with the calculated cross section, and the unfolding uncertainty, which is dominated by the statistical uncertainty, should scale with the square root of the calculated cross sections.[77,203]

This form of the χ^2 expression is based on the likelihood function

$$\mathcal{L} = \frac{1}{\sqrt{(2\pi)^n \operatorname{Det}(\widetilde{V})}} \, \mathrm{e}^{-\frac{1}{2} \vec{D} \cdot \widetilde{V}^{-1} \cdot \vec{D}} \prod_k \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{1}{2} s_k^2}, \tag{12.2}$$

with n being the number of data points. The first term is a multivariate normal distribution for the data points. The expectation values are the true cross sections, which are approximated by the fitted values $\vec{\sigma}_{\text{theory}}(\vec{p})$. This term includes the scaling of the statistical and uncorrelated systematic uncertainties via $\widetilde{V}$ and the shifts due to correlated systematic uncertainties via $\vec{D}$. The second term in the likelihood function is a normal distribution for each shift with expectation value 0 and standard deviation 1. The χ^2 function is derived from the likelihood as $\chi^2 = -2\operatorname{Log}(\mathcal{L}/\mathcal{L}_0)$, where the normalisation factor $\mathcal{L}_0 = \mathcal{L}|_{\vec{\sigma}_{\text{theory}} = \vec{\sigma}_{\text{data}}, s_k = 0}$ removes constant additive terms. Constant terms are irrelevant for the fit, but their removal ensures that the derived expression is χ^2-distributed and can thus be used as a goodness-of-fit test.[204] Evaluating this expression leads to the appearance of the third term in equation (12.1).

12.1.1 Datasets and uncertainties

The inclusive jet cross sections presented in this thesis alone do not possess enough constraining power to usefully determine the PDFs and $\alpha_{\text{s}}(M_Z^2)$. Therefore, they are analysed together with

the HERA combined inclusive DIS data[77] and previous inclusive jet and dijet measurements from ZEUS during the HERA I and II periods. For details on the combined HERA DIS dataset, see section 4.1. An overview of the ZEUS jet datasets is given in the following table. The last line represents the datasets presented in this thesis. Here, and in the remainder of this chapter, $p_\perp$ will denote $p_{\perp,\mathrm{Breit}}$ when referring to inclusive jet calculations and $\overline{p_{\perp,\mathrm{Breit}}}$ for dijets.

Dataset	Years	L [pb^{-1}]	Points	Q^2	$p_\perp$
HERA I inclusive jets[133]	96–97	38.6 pb^{-1}	30	$> 125\,\mathrm{GeV}^2$	$> 8\,\mathrm{GeV}$
HERA I/II dijets[71]	98–00, 04–07	374 pb^{-1}	16	$> 125\,\mathrm{GeV}^2$	$> 8\,\mathrm{GeV}$
HERA II inclusive jets	04–07	347 pb^{-1}	24	$> 150\,\mathrm{GeV}^2$	$> 7\,\mathrm{GeV}$

Due to a cut on the invariant mass in the dijet measurement, the leading-order ($\mathcal{O}(\alpha_\mathrm{s})$) prediction vanishes in parts of the dijet phase space. Therefore, the NLO calculation ($\mathcal{O}(\alpha_\mathrm{s}^2)$) is effectively only leading order, and the NNLO calculation ($\mathcal{O}(\alpha_\mathrm{s}^3)$) is effectively NLO. This leads to an increased scale uncertainty for the affected points and, more importantly, changes the correlation of the scale uncertainty to the other points. To avoid this issue, the six low-$p_\perp$ dijet points are excluded from the analysis, such that only 16 of the 22 published points are used.

Jet-energy-scale and luminosity uncertainty   In the previous jet datasets, the jet-energy-scale and luminosity uncertainties are treated as fully correlated, while all other uncertainties are treated as uncorrelated. These two sources of uncertainty are also treated as fully correlated in this analysis. For simplicity and since it is approximately symmetric, the jet-energy-scale uncertainty is symmetrised by averaging the absolute values of the relative upward and downward uncertainty. As explained in section 9.3.3, the data analysed in this thesis overlap significantly with those of the dijet measurement. This results in statistical correlations between the two datasets, which are considered via a statistical correlation matrix, as described in the same section. Furthermore, it results in correlations of the systematic uncertainties. In the mentioned section, the overlap of the data used in the two measurements is quantified to be 80%. Correspondingly, the jet-energy-scale and luminosity uncertainties are treated as 80% correlated across both datasets and 20% correlated only within each dataset. Equation (12.1) only allows for fully correlated or fully uncorrelated uncertainties. Technically, this partial correlation is achieved by splitting these uncertainties into two components, one $\sqrt{80\%}$ and one $\sqrt{20\%}$ the size of the original uncertainty. The former component is treated as fully correlated across both datasets. The latter components are treated as fully correlated only within each dataset.

Model uncertainty   For the present dataset, additional correlations are considered. They are summarised in table 12.1. The model uncertainty is treated as half correlated across all points. This is because it is estimated from the smooth distributions of the Ariadne and Lepto MC samples. However, it is not guaranteed that these samples are entirely representative of the model dependence of the cross sections. Similar to the previous paragraph, the half correlation is achieved by splitting the uncertainty into two equal pieces, both $\sqrt{50\%}$ the size of the original. One is treated as fully correlated, and the other as fully uncorrelated.

Uncertainty of the $E - p_\mathrm{z}$ cut   As seen in section 10.5, the upward variation of the uncertainty associated with the $E - p_\mathrm{z}$ cut has little effect on the cross sections, while the

Uncertainty	Fraction uncorrelated	Fraction correlated within inclusive jets	Fraction correlated to dijets
Jet-energy scale	0%	80%	20%
Luminosity	0%	80%	20%
Model	50%	50%	–
$E - p_z$ cut	0%	100%	–
Track-matching efficiency	0%	100%	–
Low-Q^2 DIS	0%	100%	–
Unfolding background	50%	50%	–
Hadronisation	50%	0%	50%*
Theoretical calculations	50%	0%	50%*
All remaining uncertainties	100%	0%	–

Table 12.1: Overview of the systematic uncertainties of the HERA II inclusive jet measurement that are considered correlated in the fit. The '*' in the last column indicates that the uncertainty is treated as correlated not only to the dijets but also to the HERA I inclusive jets.

downward variation affects the high-$p_\perp$ region significantly. This effect is dominated by the variation of the lower boundary of the cut, where the MC does not describe the data perfectly, as seen in figure 7.6. This uncertainty is treated as fully correlated since it quantifies the influence of a single parameter, the $E - p_z$ cut boundary.

Uncertainty due to background normalisation   The uncertainty associated with the normalisation of the low-Q^2 DIS MC samples is also treated as fully correlated since, by definition, it affects all phase space regions simultaneously and in the same way. The unfolding background, discussed in figure 9.5, consists of multiple components, such as migrations in y or η_{lab} and misreconstructions. Therefore, it is treated as half correlated.

Uncertainty due to track-matching efficiency   The track-matching-efficiency correction accounts for mismatched tracks and misidentified photons, see section 8.2.2. It is checked that both of these effects lead to a similar shape of the systematic uncertainty, such that this uncertainty can also be treated as fully correlated.

All remaining experimental systematic uncertainties are either very small or their degree of correlation is difficult to determine. Therefore, they are treated conservatively as uncorrelated. For simplicity, they are symmetrised by averaging the absolute values of their relative upward and downward uncertainties and added in quadrature to form a single uncorrelated systematic uncertainty.

The uncertainty of the hadronisation correction is treated as a systematic uncertainty in the fit. This prescription reduces the influence of the hadronisation uncertainty compared to the previously used offset method.[103] It is assumed to be 50% correlated across all jet datasets and 50% uncorrelated. An additional uncertainty arises due to the statistical and systematic uncertainty of the theoretical calculations. This uncertainty is also treated as a 50% correlated systematic uncertainty.

12.1.2 Fit settings

The strategy of the analysis closely follows that of the previous HERAPDF2.0Jets (N)NLO analyses.[77,103] The general form of the PDF parameterisation is

$$\xi f(\xi, \mu_{\mathrm{f},0}^2) = A_f \xi^{B_f} (1-\xi)^{C_f} (1 + D_f \xi + E_f \xi^2).$$

Relevant parameters for each parton type were identified by observing their effect on the χ^2.[103] The determined optimal parameterisation for all PDFs is

$$\xi g(\xi, \mu_{\mathrm{f},0}^2) = A_g \xi^{B_g} (1-\xi)^{C_g} - A_g' \xi^{B_g'} (1-\xi)^{C_g'}, \tag{12.3a}$$

$$\xi u_{\mathrm{v}}(\xi, \mu_{\mathrm{f},0}^2) = A_{u_{\mathrm{v}}} \xi^{B_{u_{\mathrm{v}}}} (1-\xi)^{C_{u_{\mathrm{v}}}} (1 + E_{u_{\mathrm{v}}} \xi^2), \tag{12.3b}$$

$$\xi d_{\mathrm{v}}(\xi, \mu_{\mathrm{f},0}^2) = A_{d_{\mathrm{v}}} \xi^{B_{d_{\mathrm{v}}}} (1-\xi)^{C_{d_{\mathrm{v}}}}, \tag{12.3c}$$

$$\xi \bar{U}(\xi, \mu_{\mathrm{f},0}^2) = A_{\bar{U}} \xi^{B_{\bar{U}}} (1-\xi)^{C_{\bar{U}}} (1 + D_{\bar{U}} \xi), \tag{12.3d}$$

$$\xi \bar{D}(\xi, \mu_{\mathrm{f},0}^2) = A_{\bar{D}} \xi^{B_{\bar{D}}} (1-\xi)^{C_{\bar{D}}}. \tag{12.3e}$$

Here, g, u_{v} and d_{v} represent the gluon, up-valence-quark and down-valance-quark distributions respectively, $\bar{U} = \bar{u} = u - u_{\mathrm{v}}$ is the up-type sea-quark distribution and $\bar{D} = \bar{d} + \bar{s} = d - d_{\mathrm{v}} + s$ is the down-type sea-quark distribution. The strange sea-quark distribution is approximated as a constant fraction f_s of the down-type sea-quark distribution $s = \bar{s} = f_s \bar{D}$. The distributions of the heavy quarks and the photon are set to zero at the starting scale. The second term of the gluon parameterisation is added to better describe data that is sensitive to the low-x_{Bj} region, such as the inclusive DIS data at low Q^2.[77]

The parameters A_g, $A_{u_{\mathrm{v}}}$ and $A_{d_{\mathrm{v}}}$ are constrained by the sum rules $\int_0^1 u_{\mathrm{v}}(\xi, \mu_{\mathrm{f},0}^2)\mathrm{d}\xi = 2$, $\int_0^1 d_{\mathrm{v}}(\xi, \mu_{\mathrm{f},0}^2)\mathrm{d}\xi = 1$ and $\sum_{i\in\{g,q,\bar{q}\}} \int_0^1 \xi f_i(\xi)\mathrm{d}\xi = 1$. The constraints $A_{\bar{U}} = A_{\bar{D}}(1-f_s)$ and $B_{\bar{U}} = B_{\bar{D}}$ ensure that the anti-up- and anti-down-quark distributions become equal at low ξ. This constraint is motivated by the fact that the valence quarks do not contribute at low ξ. When neglecting the quark masses, isospin symmetry requires $\bar{U}$ and $\bar{D}$ to be equal in this case. This assumption is necessary because the HERA data are not sensitive to the difference between $\bar{U}$ and $\bar{D}$ distributions at low ξ. The parameter C_g' is fixed to 25 to ensure that the corresponding term does not contribute at large ξ.[77,103] This leaves 14 independent parameters to describe the PDFs.

An overview of the settings of the fit is given in table 12.2. The settings of the fit are chosen similarly to the HERAPDF analyses. Low-Q^2 DIS data points are excluded by requiring Q^2 to be greater than $Q_{\min}^2$. The calculations of the inclusive DIS cross sections were performed in a general-mass variable-flavour-number scheme, in which the heavy quark masses m_c and m_b need to be specified.[205] These choices do not affect the jet calculations directly, as they are performed in a massless scheme. The renormalisation and factorisation scales of the inclusive DIS calculations are set to $\mu_{\mathrm{f}}^2 = \mu_{\mathrm{r}}^2 = Q^2$. This is treated as a definition of the PDFs. In the jet calculations, the renormalisation and factorisation scales are chosen to depend also on $p_\perp$ as given in the table.

By definition, the χ^2 function will be approximately parabolic close to its minimum. The experimental/fit uncertainty of the fit is defined as the width of this parabola at the height where it has increased by one with respect to the minimum, i.e. $\chi^2 - \chi_{\min}^2 = 1$. This uncertainty includes the uncertainty of the experimental input, the hadronisation correction and the uncertainty of the theory calculations.

The model uncertainty is evaluated by modifying each of the model parameters listed in the table by its uncertainty and repeating the fit. The changes of the PDFs and $\alpha_{\mathrm{s}}(M_Z^2)$

		NLO	**NNLO**
Model			
Q^2 cut-off	Q^2_{min} [GeV^2]	$3.5^{+1.5}_{-1.0}$	
Strange fraction	f_s	0.4 ± 0.1	
Charm mass	m_c [GeV]	$1.46^{+0.04}_{-\mathrm{symmetrise}}$	$1.41^{+0.04}_{-\mathrm{symmetrise}}$
Bottom mass	m_b [GeV]	4.3 ± 0.10	4.2 ± 0.10
Parameterisation			
Starting scale	$\mu^2_{\mathrm{f},0}$ [GeV^2]	$1.9^{+\mathrm{symmetrise}}_{-0.3}$	
Additional parameters		$D_g, E_g, D_{u_{\mathrm{v}}}, D_{d_{\mathrm{v}}}, E_{d_{\mathrm{v}}}, E_{\bar{U}}, D_{\bar{D}}, E_{\bar{D}}$	
Scales			
Renormalisation scale	μ^2_{r}	$(Q^2 + p^2_\perp)/2$	$Q^2 + p^2_\perp$
Factorisation scale	μ^2_{f}	Q^2	

Table 12.2: Fit settings and associated uncertainties at NLO and NNLO.

are added up in quadrature, separately for the positive and negative directions. For the parameterisation uncertainty, the starting scale is varied similarly. A requirement of the QCD scheme for the inclusive DIS data is that the starting scale is less than the heavy-quark masses. To ensure this during the parameter variations, the starting scale is only varied downwards and the charm mass only upwards. The resulting changes on the PDFs and $\alpha_{\mathrm{s}}(M_Z^2)$ are assumed to be symmetric for the variation in the other direction. Another contribution to the parameterisation uncertainty comes from adding additional parameters. Each of the eight D and E parameters not considered in (12.3) is added to the fit in turn. The envelope of the determined PDFs and $\alpha_{\mathrm{s}}(M_Z^2)$ forms the second contribution to the parameterisation uncertainty.

12.1.3 Scale uncertainty

A scale uncertainty is assigned to $\alpha_{\mathrm{s}}(M_Z^2)$ to account for the uncertainty due to missing higher-order terms in the jet calculations. Thus, it represents the theoretical uncertainty on the $\alpha_{\mathrm{s}}(M_Z^2)$ determination. No scale uncertainty is assigned to the PDFs, as they are determined mainly by the inclusive DIS data, for which the scale choices are treated as part of the definition and thus have no uncertainty associated with them.

The most commonly used method to evaluate the scale uncertainty is to adjust the scale choices and repeat the fit. In this analysis, this is done using a six-point variation with rescaling factors 0.5 and 2, i.e. six additional fits are performed with the renormalisation and factorisation scales set to $(\mu_{\mathrm{r}}, \mu_{\mathrm{f}}) \in \{(0.5, 1), (1, 0.5), (0.5, 0.5), (2, 1), (1, 2), (2, 2)\}(\mu_{\mathrm{r},0}, \mu_{\mathrm{f},0})$, where $\mu_{\mathrm{r},0}$ and $\mu_{\mathrm{f},0}$ are the nominal choices given in table 12.2. The envelope of the resulting $\alpha_{\mathrm{s}}(M_Z^2)$ values is used as the scale uncertainty.[159]

In this method, the scales for calculating all jet points are varied simultaneously and by the same amount. This corresponds to the assumption that the scale dependence of the cross sections is fully correlated. The scale dependence estimates the influence of missing higher orders in the perturbative expansion, which are expected to vary smoothly throughout the phase space. Therefore, this fully correlated assumption is reasonable when considering jet points close in phase space. However, when considering points far away from each other in

phase space or in different final states, such as inclusive jets and dijets, the scale uncertainty of the cross sections cannot be assumed to be fully correlated. It might be uncorrelated or even anti-correlated. Therefore, the fully correlated method described above can give larger uncertainties than necessary, which is a known issue.[103] This effect will contribute especially strongly when fitting jet points across a wide range in phase space. In this analysis, jets are fitted in a relatively small region of phase space, so the effect is expected to be moderate.

Nevertheless, an alternative treatment of the scale uncertainty is investigated. This approach assumes that the scale dependence is half correlated and half uncorrelated across all jet points. While still not ideal, this treatment is expected to be more reasonable than the fully correlated approach. A conceptually similar method has been used in previous analyses.[77,121]

The alternative evaluation of the scale uncertainty proceeds as follows. First, a fit is performed using the nominal scales. Using the PDFs and $\alpha_s(M_Z^2)$ from this fit, cross section scale uncertainties are computed for all jet points using a six-point variation with rescaling factors 0.5 and 2. These cross section scale uncertainties are scaled by $\sqrt{0.5}$ and added to the fit as an uncorrelated systematic uncertainty. When the fit is repeated with this additional uncertainty of the data, the experimental/fit uncertainty of $\alpha_s(M_Z^2)$ will increase. The quadratic difference between the increased and the original experimental/fit uncertainty is used as the uncorrelated contribution to the scale uncertainty of $\alpha_s(M_Z^2)$. Using the fit with the added systematic uncertainty as a starting point, the correlated contribution is computed similarly to the fully correlated case, but with rescaling factors $\sqrt{0.5}$ and $\sqrt{2}$. The total scale uncertainty is finally determined as the quadratic sum of the uncorrelated and correlated contributions.

12.1.4 Results

A fit is performed at NNLO accuracy, using the newly measured cross sections in the matrix unfolded variant. The fit achieves a χ^2 per degree of of freedom of 1419/1200. Appendix B.6.1 provides more details on the fit output. To judge the impact of the new dataset, the result of this fit is compared to a similar fit, excluding the cross sections presented in this thesis, i.e. using only the HERA combined inclusive DIS, ZEUS HERA I inclusive jet and ZEUS HERA I/II dijet datasets. The PDFs are compared in figure 12.1. The PDFs move slightly within their uncertainty. None of the PDFs are significantly constrained by the inclusion of the additional jet dataset. The same behaviour is observed in HERAPDF, see section 4.2. The main effect on the PDFs is due to the updated value of $\alpha_s(M_Z^2)$, which is strongly correlated with the shape of the PDFs, especially that of the gluon PDF, see section 2.3.5. Since the two fits determine similar values of $\alpha_s(M_Z^2)$, the corresponding PDFs are also similar.

The determined values of $\alpha_s(M_Z^2)$ are shown at NLO and NNLO, including all available uncertainties in table 12.3. Values are presented using the dataset presented in this thesis in both the matrix unfolded and bin-by-bin corrected variant. The determined central values are in excellent agreement with each other. The experimental/fit uncertainties are also similar, indicating that both unfolding methods lead to a similar constraining power of the resulting dataset. This verifies the claim made in section 11.3, that even though the uncertainties of the matrix unfolded cross sections appear much larger, they become comparable when including correlations.

The table also shows the fit result, excluding the dataset presented in this thesis. Including the additional dataset reduces the experimental/fit uncertainty of $\alpha_s(M_Z^2)$. This reduction is especially important since the experimental/fit uncertainty is the dominant uncertainty in the current determination. Furthermore, fits have been performed using only the newly
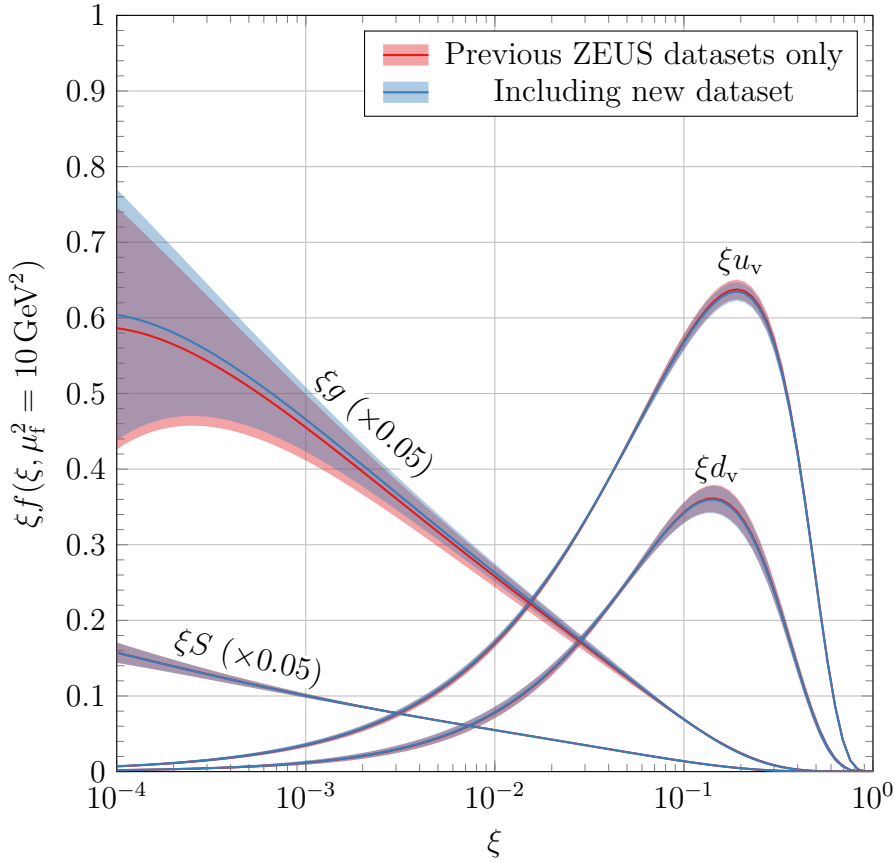
Figure 12.1: The fitted PDFs, before including the inclusive jet cross sections presented in this thesis in the fit and after including them. The sea-quark distribution is defined as $xS = x(\bar{U} + \bar{D})$. Only experimental uncertainties are shown. Note that $\alpha_s(M_Z^2)$ is left free in these fits. Due to the uncertainty of $\alpha_s(M_Z^2)$ and the correlation between $\alpha_s(M_Z^2)$ and the PDFs, the PDF uncertainties obtained here are notably larger than the ones in the corresponding HERAPDF determination, in which $\alpha_s(M_Z^2)$ was fixed, see figure 4.1.

measured cross sections and excluding the previous ZEUS datasets. These fits result in similar experimental/fit uncertainties to those using only the previous ZEUS datasets. This indicates that the new dataset alone has a similar constraining power to both previous datasets. The reduction in uncertainty of the combined fit is suppressed due to the positive correlation between the datasets.

An additional fit is performed at NNLO accuracy and using only the HERA inclusive DIS data and the newly measured cross sections in the bin-by-bin corrected variant. In this fit, the model uncertainty is set to zero. Consequently, the experimental/fit uncertainty decreases compared to the fit including the uncertainty. Taking the quadratic difference between the experimental/fit uncertainties of these two fits yields an estimate of the contribution of the model uncertainty of the cross sections to the uncertainty of $\alpha_s(M_Z^2)$. A value of 0.0005 is determined for this contribution. The central $\alpha_s(M_Z^2)$ value determined in fits with the bin-by-bin corrected cross sections (0.1114) and the matrix unfolded variant (0.1107) are compatible within this model uncertainty. This is further confirmation that the two unfolding

		$\alpha_{\mathrm{s}}(M_Z^2) \pm \Delta_{\mathbf{exp./fit}}$	$\pm \Delta_{\mathbf{mod./par.}}$	$\pm \Delta_{\mathbf{scale}}^{(100\%)}$	$\pm \Delta_{\mathbf{scale}}^{(50\%)}$
NLO	Previous ZEUS datasets only	0.1165 ± 0.0019			
	New dataset only[(m)]	0.1121 ± 0.0019			
	New dataset only[(b)]	0.1121 ± 0.0019			
	All ZEUS datasets[(m)]	0.1160 ± 0.0017	$+0.0007 \atop -0.0009$	$+0.0026 \atop -0.0014$	$+0.0012 \atop -0.0009$
	All ZEUS datasets[(b)]	0.1159 ± 0.0016			
	HERAPDF2.0Jets[77]	0.1183 ± 0.0015	±0.0005		$+0.0037 \atop -0.0033$
NNLO	Previous ZEUS datasets only	0.1149 ± 0.0019			
	New dataset only[(m)]	0.1107 ± 0.0019			
	New dataset only[(b)]	0.1114 ± 0.0016			
	All ZEUS datasets[(m)]	0.1143 ± 0.0017	$+0.0006 \atop -0.0007$	$+0.0012 \atop -0.0005$	$+0.0006 \atop -0.0004$
	All ZEUS datasets[(b)]	0.1141 ± 0.0015			
	HERAPDF2.0Jets[103]	0.1156 ± 0.0011	$+0.0001 \atop -0.0002$	±0.0029	±0.0022

Table 12.3: Fitted values of $\alpha_{\mathrm{s}}(M_Z^2)$ at NLO and NNLO, including all available uncertainties. The total uncertainty is the quadratic sum of the experimental/fit uncertainty $\Delta_{\mathrm{exp./fit}}$, the model/parameterisation uncertainty $\Delta_{\mathrm{mod./par.}}$ and either the scale uncertainty evaluated under the fully correlated ($\Delta_{\mathrm{scale}}^{(100\%)}$) or half correlated ($\Delta_{\mathrm{scale}}^{(50\%)}$) assumption. Shown are the results from fits using, apart from the HERA inclusive DIS data, only previous ZEUS jet datasets, only the dataset presented in this thesis and all of them simultaneously. For fits including the newly measured dataset, the superscript 'm/b' indicates whether the dataset is used in the matrix unfolded and bin-by-bin corrected variant. The corresponding results from HERAPDF at also shown. For the HERAPDF2.0Jets determination at NLO, the listed exp./fit uncertainty includes the hadronisation uncertainty, which is quoted separately in the corresponding publication, but in all the other ones is included in the exp./fit uncertainty by default.

methods are consistent with each other. From now on, only the matrix unfolded cross sections will be considered.

For reference, the table also shows the results from the corresponding HERAPDF analyses, see section 4.2 for details. As expected, the experimental/fit uncertainty of the present analysis is increased since fewer jet datasets are used than in HERAPDF. The model/parameterisation uncertainty is slightly increased compared to HERAPDF. This is potentially also a consequence of the lower number of jet datasets used. Since $\alpha_{\mathrm{s}}(M_Z^2)$ is less constrained, it is more strongly affected by a change in the model parameters. This effect is not investigated further since this is still a subdominant contribution to the overall uncertainty.

Very notably, the scale uncertainty of the present analysis is significantly smaller than that of the HERAPDF analyses. This is a direct consequence of the way this scale uncertainty is evaluated. In the standard treatment, listed as $\Delta_{\mathrm{scale}}^{(100\%)}$ in the table, the scale dependence of the cross sections is assumed to be fully correlated across all jet data points. As discussed in section 12.1.3, this assumption is not justified when fitting points across a wide range in phase space. Varying the scales by the same amount for all jet points can, thus, overestimate the uncertainty associated with the scale choice. Since the HERAPDF analyses also use H1 jet datasets at low-Q^2, this issue affects them much more severely than the current determination.
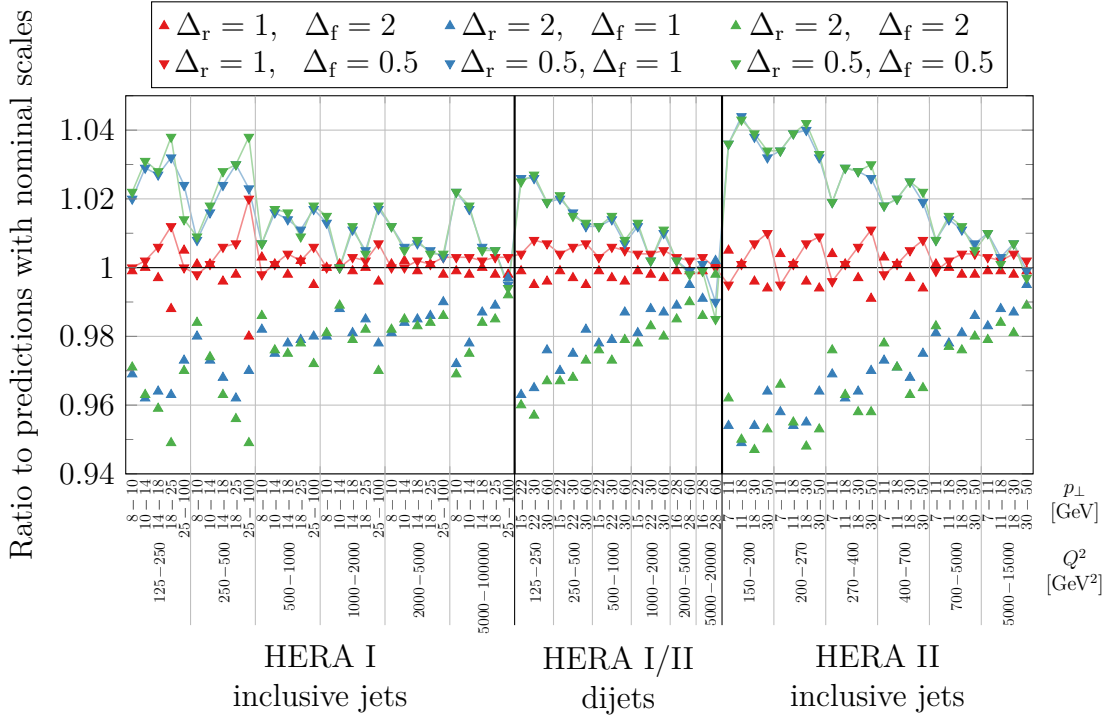
Figure 12.2: Scale dependence of the investigated cross sections. The figure shows the ratio of the cross sections computed with modified scales to the cross sections computed with the nominal scales $\mu_r^2 = \mu_f^2 = Q^2 + p_\perp^2$. In the legend, the symbols $\Delta_{r/f} = \frac{\mu_{r/f}}{\mu_{r/f,0}}$ indicate the factor by which the renormalisation/factorisation scales are modified. Shown are the NNLO QCD predictions corresponding to jet datasets used in the fit presented here. The double-differential points are mapped onto the one-dimensional axis similarly to section 9.4, with the large bins corresponding to Q^2 bins and the points within them corresponding to $p_\perp$ bins. For visual clarity, some points are connected by solid lines.

The half-correlated treatment of the scale uncertainty is designed to partially counter this issue. While still assuming long-range correlations in the cross section scale dependence, this approach effectively allows the fit to adjust the scale choice of each jet point separately. This is achieved by adding an additional uncorrelated uncertainty to the fit that corresponds to the size of the scale dependence, see section 12.1.3. The scale uncertainties determined using this approach are listed in the table as $\Delta_{\text{scale}}^{(50\%)}$. In all cases where the scale uncertainty is evaluated using both approaches, the half-correlated approach leads to a smaller uncertainty.

When fitting points in a very wide range of phase space, the half-correlated treatment might not be sufficient. The scale uncertainty of the cross sections can even be anti-correlated for jet points very far away from each other in phase space. The half-correlated/half-uncorrelated approach does not cover anti-correlations. On the other hand, when fitting points in a very narrow range of phase space, the half-correlated treatment can also be inappropriate. The scale dependence of points close in phase space is likely strongly correlated. In those cases, the half-correlated treatment can underestimate the scale uncertainty.

The scale dependence of cross sections considered in the current fit is shown in figure 12.2. It is dominated by the renormalisation scale, especially in the low-Q^2 regions, which in the figure are shown at the left side of the ranges corresponding to each measurement. In most cases, a

variation of the scales shifts all points in the same direction, indicating a positive correlation of the scale dependence. The most notable exception is the variation of the factorisation scale of the HERA II inclusive jet measurement. Here, the scale dependence is anti-correlated since a scale variation moves points in different directions. The half-correlated treatment is appropriate for the current fit since negative correlations are observed only for a small number of points and for a subdominant scale variation. Even if this were not the case, the half-correlated treatment would still be preferable over the fully correlated one.

Furthermore, the figure shows that especially the high-Q^2 points have a minimal scale dependence. This means they prefer the same $\alpha_s(M_Z^2)$ value, regardless of the chosen scale. Therefore, when performing a fit with changed scales, the determined $\alpha_s(M_Z^2)$ values must be reasonably close to the nominal one. Otherwise, the fit could not describe the high-Q^2 points. This aspect also contributes to the small scale uncertainty of the determined $\alpha_s(M_Z^2)$ value.

Figure 12.3 compares the $\alpha_s(M_Z^2)$ value determined in this analysis at NNLO, from the matrix-unfolded cross sections and using the half-correlated treatment of the scale uncertainty to various other determinations, some of which are explained in more detail in section 4.3 and 4.4. Due to the decreased scale uncertainty, the present measurement is one of the most precise determinations of the strong coupling constant at hadron colliders so far. It achieves the highest precision at HERA, even though some previous determinations use significantly more data from both the ZEUS and H1 experiments. The presented measurement even competes with recent measurements from the LHC. It has the potential to contribute to further reducing the uncertainty of the world average value of $\alpha_s(M_Z^2)$.
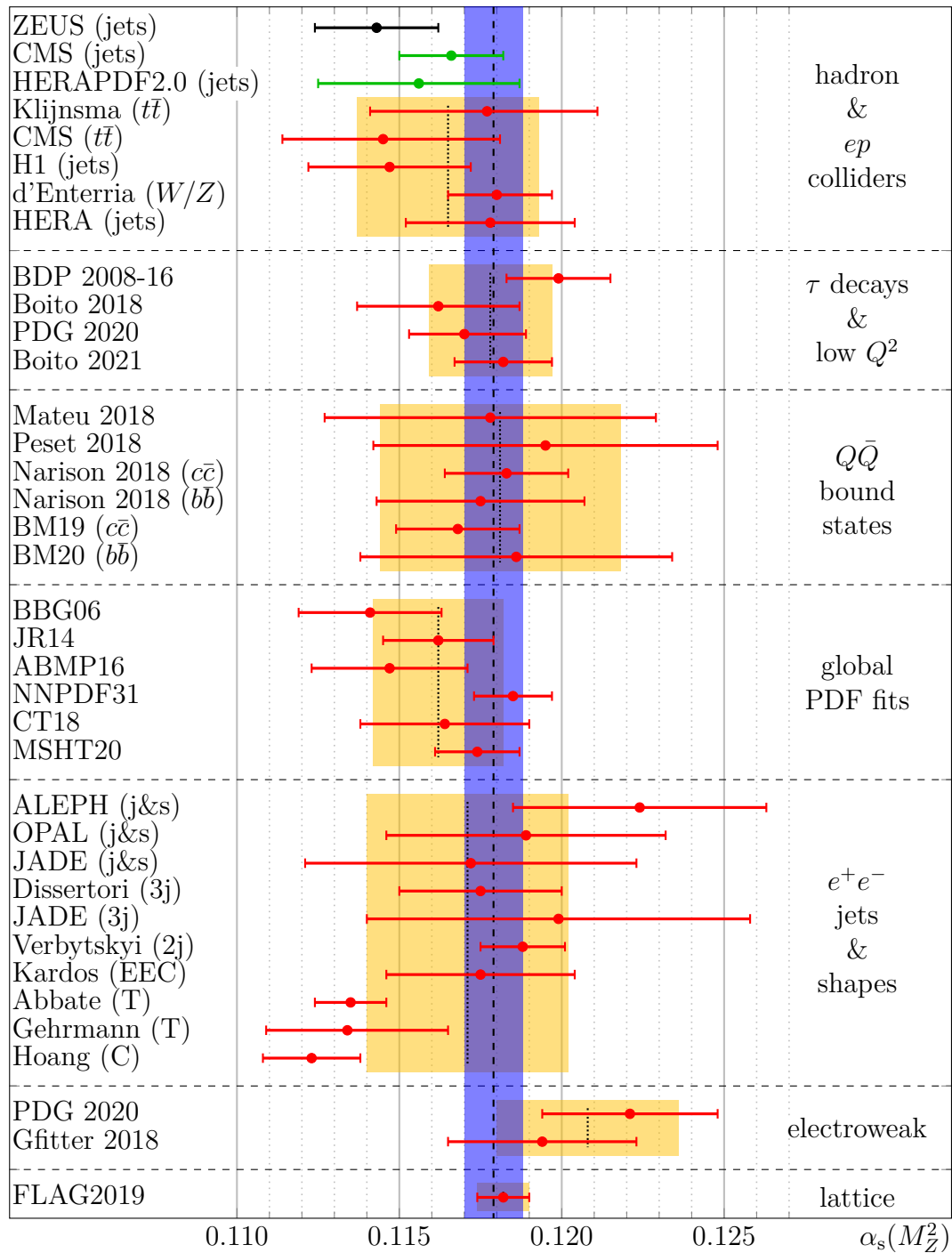
Figure 12.3: Comparison of different determinations of the strong coupling constant. The blue band indicates the Particle Data Group world average, and the red points are the determinations that enter it.[17] The yellow boxes indicate the sub-averages in each of the seven categories. In addition, the latest HERAPDF result and a recent determination from CMS are shown in green.[103,145] The top point represents the value found in this analysis using the half-correlated treatment of the scale uncertainty. All determinations shown are at least at NNLO accuracy in QCD. The error bars indicate the total uncertainty of each determination. Figure adapted from Particle Data Group.

12.2 Determination of the scale dependence of the strong coupling

As discussed in section 2.4.4, the strong coupling α_s is not a constant but depends on the scale at which it is measured. Another valuable test of QCD is to investigate the shape of the $\alpha_s(\mu^2)$ curve. This information can be used to confirm that QCD is the correct theory to describe the strong interaction.

12.2.1 Global and local determinations of α_s

In the previous section, the strong coupling is expressed at the mass of the Z boson, which is a commonly used convention.[17,104,105,145] In xFITTER, $\alpha_s(M_Z^2)$ is the fit parameter that is numerically determined. It should be noted that M_Z is just an arbitrary choice for a reference scale. Since the QCD evolution of $\alpha_s(\mu^2)$ is assumed to hold, adjusting $\alpha_s(M_Z^2)$ affects the entire $\alpha_s(\mu^2)$ curve in a precisely known way. Therefore, one should not think of the fit as fitting $\alpha_s(M_Z^2)$, but rather as fitting the entire $\alpha_s(\mu^2)$ curve, within the constraints of QCD, and quoting the result as $\alpha_s(M_Z^2)$. The fit would have determined the exact same results if the fit parameter was $\alpha_s(\mu_{ref}^2)$ at a different reference scale μ_{ref}. The fact that the fit parameter is $\alpha_s(M_Z^2)$ is just an implementation detail and carries no physical meaning. Using QCD evolution, α_s could also be quoted at any other scale, and the conclusions would be the same. A fit that uses data points at many different scales and thus fits large sections of the $\alpha_s(\mu^2)$ curve is referred to as a global fit.

To study the scale dependence of the strong coupling, $\alpha_s(\mu_i^2)$ is determined at different scales μ_i, and the curve is compared to the QCD prediction. This can be realised as follows. First, each jet data point used in the previous section is assigned a scale μ. Data points are then sorted into groups of points with similar scales μ_i. The strong coupling is then determined using, as far as possible, only the jet data points in one of these groups. The determined value corresponds to $\alpha_s(\mu_i^2)$. Note again that the fit adjusts the entire $\alpha_s(\mu^2)$ curve and quotes the result as $\alpha_s(M_Z^2)$. But since the only inputs to the fit are jet points close to a scale μ_i, this effectively corresponds to fitting $\alpha_s(\mu_i^2)$. The $\alpha_s(M_Z^2)$ value quoted by the fit should be evolved to the scale μ_i to determine the physically meaningful result $\alpha_s(\mu_i^2)$. A fit that only uses data points close to a single scale value will be referred to as a local fit.

12.2.2 Interpretation of the result

Determining $\alpha_s(\mu_i^2)$ exclusively from jet data points at one scale is not feasible. This is because any prediction of jet cross sections necessarily requires PDFs, and those cannot be usefully constrained using just a few jet data points. The next best approach is to use fixed PDFs at different $\alpha_s(M_Z^2)$ values. The concrete procedure to determine $\alpha_s(\mu_i^2)$ in this way is the following: Using the inclusive DIS data only, determine PDF sets at different fixed $\alpha_s(M_Z^2)$ values. Next, for each $\alpha_s(M_Z^2)$ value, use the corresponding PDFs to compute χ^2 according to equation (12.1) for just the jet points. The computed points form a $\chi^2(\alpha_s(M_Z^2))$ curve that will exhibit a minimum from which the preferred $\alpha_s(M_Z^2)$ value of the jet data points can be extracted. This $\alpha_s(M_Z^2)$ value is then evolved to the scale of the jets to obtain $\alpha_s(\mu_i^2)$. This procedure is reasonable because the inclusive DIS data alone do not have a strong preference for an $\alpha_s(M_Z^2)$ value, as discussed in section 4.2.

This procedure introduces another conceptual difficulty. When determining the PDFs, it is assumed that the QCD evolution of the strong coupling holds for the inclusive DIS data.

Therefore, this procedure cannot be considered a measurement of the scale dependence of the strong coupling. If QCD was not the correct theory of the strong interaction and the scale dependence of $\alpha_s(\mu^2)$ would differ, then the determined curve would not reflect the true shape. Instead, it would be a complicated combination of the true shape and the fact that QCD evolution was assumed to hold for the PDFs. Therefore, if the determined curve deviates from the QCD prediction, it would be very difficult to quantify what this means for the scale dependence of $\alpha_s(\mu^2)$. A qualitative statement can, however, be made. If the curves disagree, a potential deviation from QCD would have been found. Any potential deviation from QCD will usually lead to a different curve, which makes this a valuable study.

12.2.3 Procedure

PDFs are determined from the inclusive DIS data only similar to the HERAPDF2.0 NNLO PDF set, but using the updated model parameters and procedures described in the previous section. The experimental/fit, model and parameterisation uncertainties are fully determined for each $\alpha_s(M_Z^2)$ value. Each uncertainty is provided as an alternative PDF set. The experimental/fit uncertainty is provided in the form of so-called PDF eigenvectors.[201] PDF sets are determined for $\alpha_s(M_Z^2)$ from 0.112 to 0.120 in steps of 0.001.

Each jet data point from the datasets used in the previous section is assigned a scale. The jet cross section falls roughly with $1/(Q^2 + p_\perp^2)^2$. Therefore, the barycentre of each data point can be approximated as

$$\frac{1}{\mu^4} = \frac{1}{2} \left(\frac{1}{(Q_{\text{low}}^2 + p_{\perp,\text{low}}^2)^2} + \frac{1}{(Q_{\text{high}}^2 + p_{\perp,\text{high}}^2)^2} \right),$$

where the subscripts 'low' and 'high' represent the lower and upper bin boundaries of each data point. This definition gives a higher weight to the lower bin boundary since the cross section is steeply falling within the bin. The scales assigned to each point are tabulated in the appendix B.6.2.

The jet data points are then split into five groups based on their scale. For this purpose, the jet points are sorted by their scale, and the scale is plotted against the sorted point index. This procedure is shown in figure 12.4. Boundaries between scale bins are placed such that each bin contains a similar number of points and bins are separated from each other, i.e. boundaries are preferably placed in gaps of the scale value. This results in five bins with 11 to 19 jet points each. These assignments are also given in appendix B.6.2. Each group of points is then assigned a representative scale value $\langle \mu \rangle$, as the cross-section-weighted average of the scales of the jet points within that group.

For each group of jet points, a $\chi^2(\alpha_s(M_Z^2))$ curve is then computed using equation (12.1). Similar to the previous section, this curve is approximated as a parabola close to its minimum. The location of this minimum is the central $\alpha_s(M_Z^2)$ value. The uncertainty of the determination is the width of this parabola at the height where it has increased by one unit in χ^2.

12.2.4 Uncertainties

When using the same statistical and systematic uncertainties of the cross sections as in the previous section, the determined uncertainty of $\alpha_s(M_Z^2)$ will correspond to a subset of the experimental/fit uncertainty from the previous section. What is missing to the full experimental/fit uncertainty of $\alpha_s(M_Z^2)$ is the effect of the experimental/fit uncertainty of the used PDFs. This uncertainty can be propagated to $\alpha_s(M_Z^2)$ by adding additional correlated
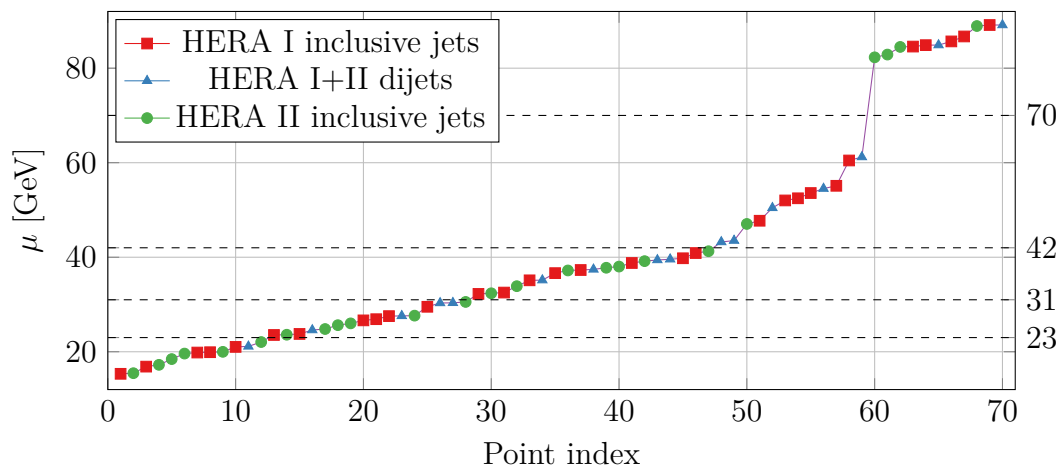
Figure 12.4: Assigned scales of jet points. The scales of the jet data points are drawn after sorting the points by their scale. The symbols indicate from which dataset each point originates. The dashed lines indicate the chosen boundaries of the scale bins.

systematic uncertainties in the fit. These describe the effect of the experimental/fit uncertainties of the PDFs on the jet cross sections. They are determined by calculating cross section predictions using the nominal PDFs and using the PDF eigenvectors and recording the effect on the cross sections. One systematic uncertainty is added for each eigenvector.

The model/parameterisation uncertainty is included similarly. Cross sections are calculated using the nominal PDFs and those corresponding to the model/parameterisation variations. For each variation, the difference between the two is added to the fit as a systematic uncertainty. If the corresponding variation is two-sided, this systematic uncertainty will be asymmetric. When these model/parameterisation uncertainties are added to the fit, the determined uncertainty of $\alpha_s(M_Z^2)$ will increase. The model/parameterisation uncertainty is determined by taking the quadratic difference between this increased uncertainty and the experimental/fit uncertainty determined above. Technically, the procedure of adding additional systematic uncertainties to the fit, corresponding to variations of the PDF set is implemented using the 'Profiling' feature of recent xFITTER versions.[202]

The scale uncertainty is computed similarly to the previous section. Six additional fits are performed, corresponding to a six-point scale variation of the renormalisation and factorisation scales with rescaling factors 0.5 and 2. In each case, a new central $\alpha_s(M_Z^2)$ value is determined. The envelope of the obtained $\alpha_s(M_Z^2)$ values is taken as the scale uncertainty. Since this is a local determination of $\alpha_s(\mu^2)$, it is appropriate to assume the scale dependence of the involved cross sections to be fully correlated.

12.2.5 Results

A cross-check is performed before applying this procedure to the groups of jet points. The procedure is applied to all jet data points simultaneously. Such a single-parameter fit of $\alpha_s(M_Z^2)$, without refitting the PDFs is also a commonly used procedure.[146,206]

The $(\alpha_s(M_Z^2), \chi^2)$ points corresponding to this cross-check are shown in figure 12.5. A parabola is fitted to each set of points. Only the seven points for which $\chi^2 - \chi^2_{\min} < 4$ are considered in these fits because the parabolic approximation is not expected to hold for points far from the minimum. As expected, the width of the curve including model/parameterisation
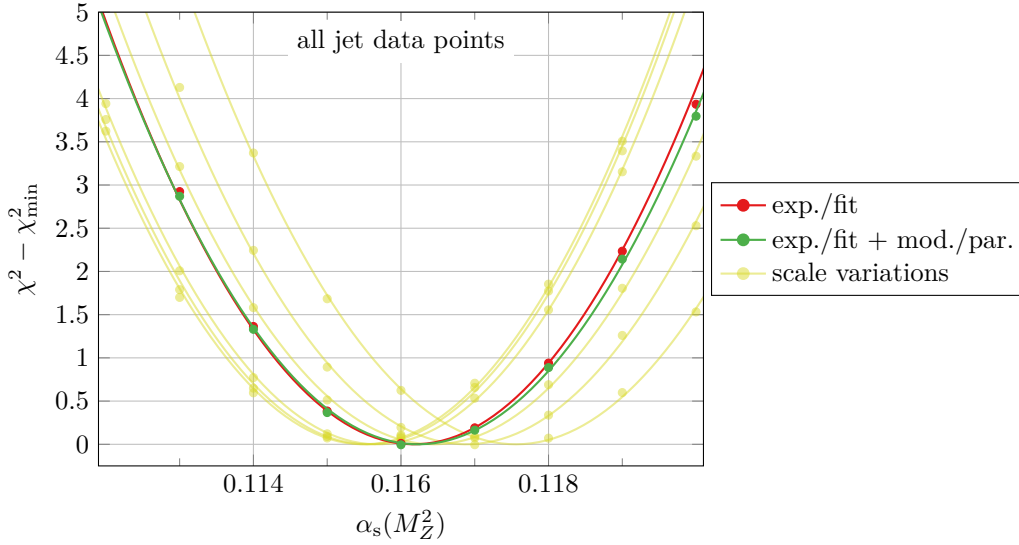
Figure 12.5: The $\chi^2(\alpha_s(M_Z^2))$ curve determined using all jet points at once. The marks indicate the computed $(\alpha_s(M_Z^2), \chi^2)$ points, and the lines represent the fitted parabolas. The red points and line indicate the curve taking only experimental/fit uncertainties into account. For the green points and line, the model/parameterisation uncertainty is additionally included. The yellow points and lines indicate the six curves corresponding to the scale variations.

uncertainties is slightly larger than that of the curve considering only the experimental/fit uncertainty. The central values are well-compatible within the model/parameterisation uncertainty. The curves corresponding to the scale variations do not differ significantly in shape, but the locations of their minima are offset.

The results of this cross-check are given in the last line of table 12.4. The central value agrees with the combined fit presented in the previous section, well within the experimental/fit uncertainty. The experimental/fit uncertainty is slightly increased. This is expected since the single-parameter fit does not take correlations between $\alpha_s(M_Z^2)$ and the PDF parameters into account to the same extent as the combined fit. This effect also affects the central value of the fit. The model/parameterisation and scale uncertainties are similar to the combined fit.

The procedure is then applied to the five groups of jet data points separately. The points and fitted parabolas are shown in figure 12.6. The parabolas are determined from the points that satisfy $\chi^2 - \chi^2_{\min} < 1$. As before, adding the model/parameterisation uncertainty increases the width of the parabolas slightly but does not affect the locations of their minima significantly. The six curves corresponding to the scale variations form three pairs with similar central values. These three pairs correspond to the variations of the renormalisation scale $\mu_r \in \{0.5, 1, 2\}\mu_{r,0}$. It was already seen in figure 12.2 that the scale dependence of the cross sections is dominated by the renormalisation scale, especially for the low-scale data points.

The determined α_s values are listed in table 12.4. The values are quoted as $\alpha_s(M_Z^2)$, including the usual uncertainties. As discussed in section 12.2.1, the fact that the fit quotes the result as $\alpha_s(M_Z^2)$ is an arbitrary choice and does not carry physical meaning. A physically meaningful result is derived by evolving these values to the scale $\langle\mu\rangle$, where the corresponding jet data were actually measured. This value $\alpha_s(\langle\mu\rangle^2)$ is also given in the table.

The determined points are compared to previous analyses and the theoretical prediction in figure 12.7. The points are well-compatible with the prediction. No deviation can be observed

Number of jet points	$\langle\mu\rangle$ [GeV]	$\alpha_s(M_Z^2)$	$\pm\delta_{\text{exp./fit}}$	$\pm\delta_{\text{mod./par.}}$	$\pm\delta_{\text{scale}}$	$\alpha_s(\langle\mu\rangle^2)\pm\delta_{\text{total}}$
12	18	0.1156	±0.0037	±0.0008	$^{+0.0035}_{-0.0025}$	0.1525 ±0.0086
16	26	0.1153	±0.0026	±0.0006	$^{+0.0028}_{-0.0017}$	0.1417 ±0.0054
19	35	0.1167	±0.0024	±0.0003	$^{+0.0018}_{-0.0010}$	0.1363 ±0.0039
12	52	0.1164	±0.0032	±0.0002	$^{+0.0011}_{-0.0003}$	0.1271 ±0.0040
11	84	0.1158	±0.0045	±0.0003	$^{+0.0014}_{-0.0004}$	0.1172 ±0.0047
70	–	0.1161	±0.0019	±0.0004	$^{+0.0014}_{-0.0007}$	–

Table 12.4: Values of the strong coupling determined at different scales. Shown are the number of jet points that enter each determination and the representative scale $\langle\mu\rangle$ of each group. For each determination, the result is quoted as $\alpha_s(M_Z^2)$, including all individual uncertainties and as $\alpha_s(\langle\mu\rangle^2)$, including the total uncertainty. The last line of the table shows the result of the cross-check, for which all jet points are fitted simultaneously.
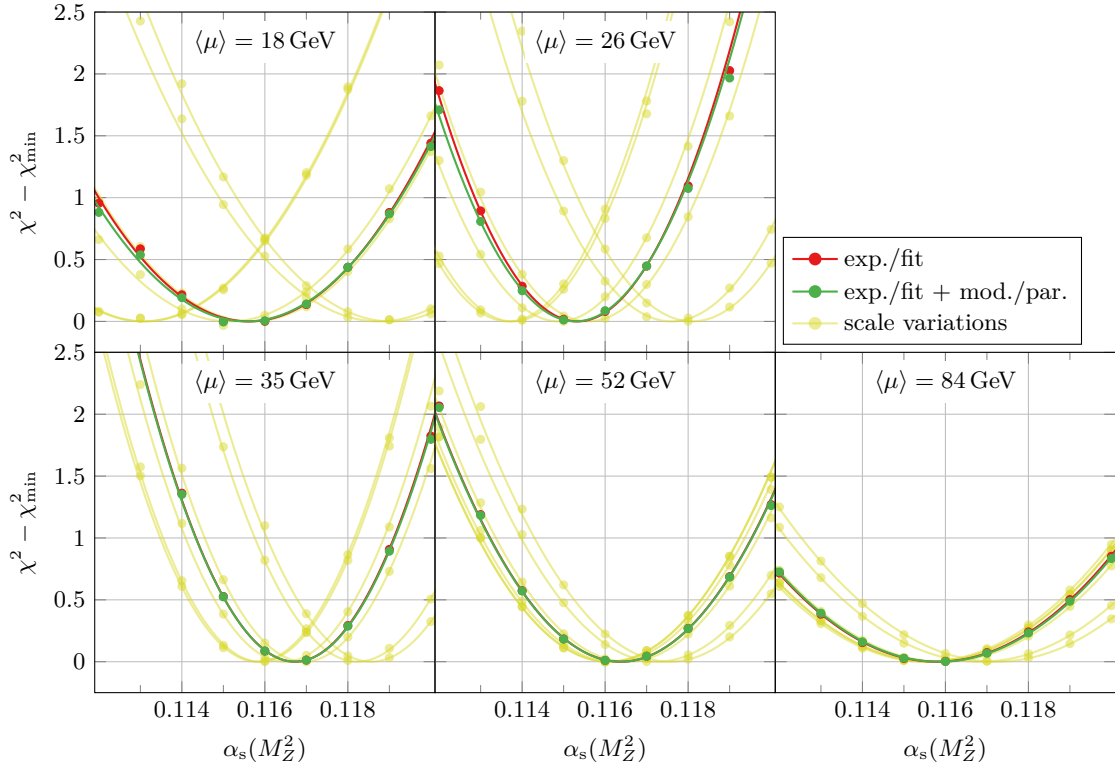
Figure 12.6: The $\chi^2(\alpha_s(M_Z^2))$ curve determined using groups of jet points. See the caption of figure 12.4 for details. Note that the $\alpha_s(M_Z^2)$ axis of the last panel is different from the others.

that would indicate any inconsistency in the employed method or in the assumption that QCD is the correct theory to describe the strong interaction.
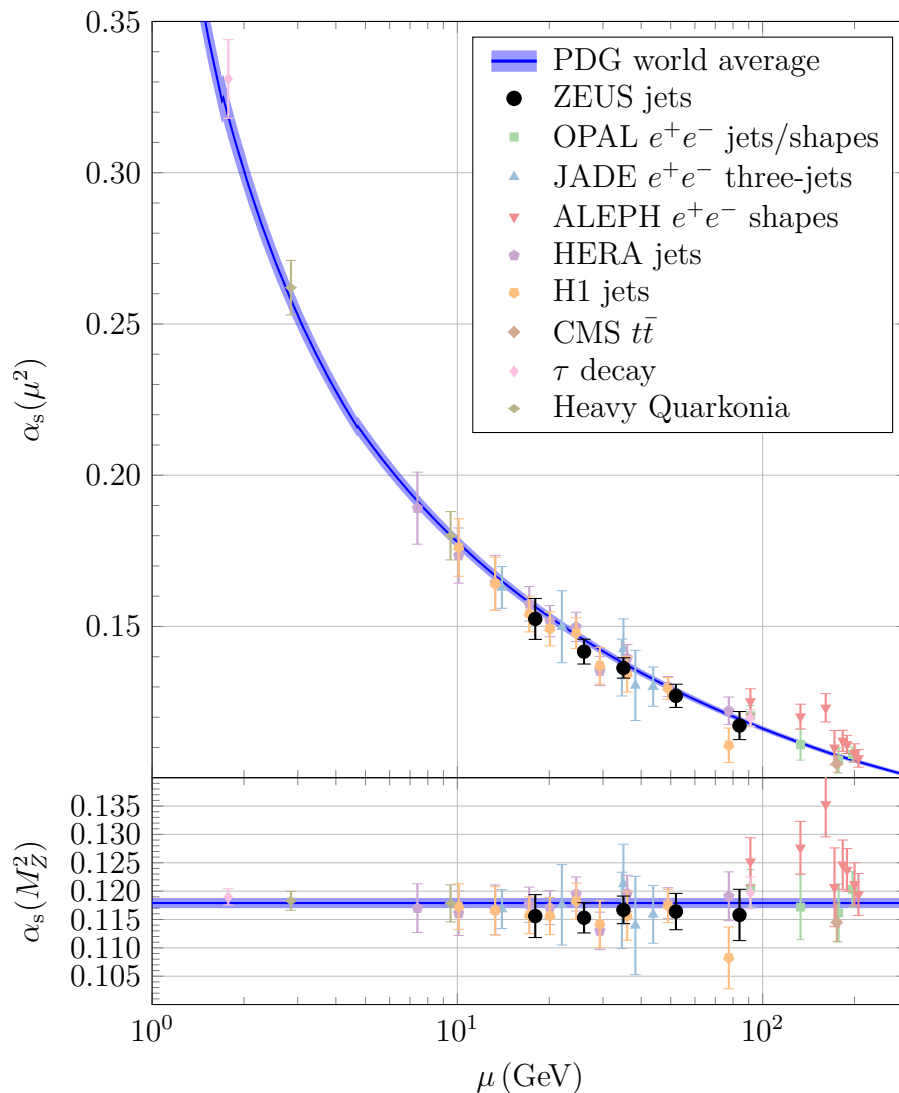
Figure 12.7: Predicted and measured scale dependence of the strong coupling. The blue curve shows the four-loop QCD prediction, assuming the Particle Data Group world average value of $\alpha_\mathrm{s}(M_Z^2) = 0.1179 \pm 0.009$.[17] Also shown are measurements from e^+e^-,[141–143] ep[206,207] and pp[146] collisions, as well as from τ leptons[150] and quarkonium states.[149] The black points represent the values determined in this thesis. The error bars indicate the total uncertainties. All determinations included are at least at NNLO in QCD. The lower panel shows the $\alpha_\mathrm{s}(M_Z^2)$ values computed from each data point.

CHAPTER 13

Conclusions

A measurement of inclusive jet cross sections in NC DIS at the ZEUS detector at the HERA $e^{\pm}p$ collider was performed. The data were taken in the years 2004–2007 and correspond to an integrated luminosity of $347\,\mathrm{pb}^{-1}$. This amounts to about two-thirds of the entire luminosity collected by the ZEUS experiment. Thus, the measured cross sections provide additional insights into jet production in DIS and significant constraining power in subsequent QCD analyses. The cross sections were compared to the corresponding measurement from the H1 collaboration and to NNLO QCD predictions and found to be well compatible.

The cross sections were measured in collisions of protons of energy 920 GeV and electrons or positrons of energy 27.5 GeV, leading to a centre-of-mass energy of 318 GeV. The kinematic range is defined using the virtuality of the exchanged boson $Q^2 \in (150, 15000)\,\mathrm{GeV}^2$ and the inelasticity $y \in (0.2, 0.7)$. Massless jets were reconstructed using the $k_\perp$-algorithm and required have a pseudorapidity $\eta_{\mathrm{lab}} \in (-1, 2.5)$ and a transverse momentum in the Breit frame $p_{\perp,\mathrm{Breit}} \in (7, 50)\,\mathrm{GeV}$, where the Breit frame is the reference frame in which the exchanged boson and the struck parton collide collinearly. In an inclusive jet measurement, each jet that passes these criteria is counted individually. The cross sections were measured double-differentially as a function of Q^2 and $p_{\perp,\mathrm{Breit}}$.

At detector level, the kinematic quantities of the event were reconstructed using the double-angle method, and jets were constructed from calorimeter cells. This reconstruction exploits the highly uniform and well-calibrated response of the high-resolution uranium-scintillator calorimeter of the ZEUS detector.

Several corrections were derived for this analysis and applied to improve the measurement of the detector-level quantities and to correct for deficiencies in the MC simulations. The track-matching- and track-veto-efficiency corrections improved the description of the tracking performance of the ZEUS detector in the MC. The electron calibration removed bias in the measurement of the DIS electron. One of the most critical corrections is the relative jet-energy-scale correction, which ensured that the bias of the jet-energy reconstruction is well described by the MC. After ensuring that it is well described, this bias was corrected by the absolute jet-energy-scale correction. Finally, the signal MC samples were reweighted so their jet distributions agree with those in the data.

After these corrections, the Ariadne and Lepto MC samples describe the data reasonably well in all relevant distributions. These samples were then used to unfold the data to hadron level and derive cross sections. This was done using a multi-dimensional matrix unfolding procedure, the most complex unfolding procedure employed at ZEUS to date. This procedure explicitly considers migrations in Q^2 and $p_{\perp,\mathrm{Breit}}$ within the phase space and across the boundaries of

the analysis phase space. Cross sections were also derived using a more traditional bin-by-bin correction method, and the results were found to be very consistent.

The consistency between the two unfolding methods shows that, when applied correctly, both are similarly valid to determine the cross sections. For the bin-by-bin correction, more elaborate corrections to the MC are required to ensure that it describes the data. For matrix unfolding, agreement between data and MC is less important for the unfolded variables but still crucial in all other variables. Previous measurements from ZEUS were often performed using the bin-by-bin correction. The results presented in this thesis demonstrate that these results are just as valid as more modern measurements with more sophisticated unfolding techniques.

Systematic uncertainties were evaluated by propagating the uncertainties of the employed selection and correction procedures to the cross sections. In most regions of the phase space, the systematic uncertainty is dominated by the uncertainty associated with the description of the jet-energy scale by the MC. The second significant contribution comes from the model uncertainty, which describes the uncertainty due to the parton shower model in the MC samples. Statistical uncertainty plays a subdominant role in most bins.

Available theoretical calculations do not include all physical effects present in the data. Correction factors were derived to make data and theory comparable. The theoretical predictions are corrected for the exchange of a virtual Z boson and photon-Z interference terms. Hadronisation correction factors were derived to make the theoretical predictions of parton level cross sections comparable to the measurement at hardon level. The QED radiation correction is applied to the measured cross sections to remove the influence of higher-order QED effects, most importantly, initial-state radiation of the electron. Cross sections are also provided using an alternative QED correction. These will allow a more rigorous treatment of higher-order QED effects once suitable theoretical predictions become available.

The derived cross sections are compared against the corresponding measurement of the H1 collaboration and to NNLO QCD theory predictions. Since the H1 measurement uses the same binning and cross section definition, the two sets of cross sections are directly comparable. The measurements are found to be in good agreement. Within the combined uncertainty, the measurements agree well with the theoretical predictions. Similar trends are observed between the two measurements and the predictions. Most notably, the predictions slightly overestimate the cross sections, especially at high $p_{\perp,\mathrm{Breit}}$.

This new jet dataset was used together with combined HERA inclusive DIS data and previous jet measurements at ZEUS in a combined QCD fit at NNLO accuracy to simultaneously determine the PDFs of the proton and the strong coupling constant $\alpha_\mathrm{s}(M_Z^2)$. The latter was determined to be

$$\alpha_\mathrm{s}(M_Z^2) = 0.1143 \pm\ 0.0017\ (\mathrm{exp./fit})^{+0.0006}_{-0.0007}\ (\mathrm{model/param.})^{+0.0006}_{-0.0004}\ (\mathrm{scale}).$$

A significantly reduced total uncertainty is observed compared to similar determinations. This is primarily due to the absence of low-Q^2 jet datasets in the determination, which leads to a strongly reduced theoretical uncertainty. Another slight reduction could be achieved by using a more appropriate treatment of this uncertainty. For the first time at ZEUS at NNLO accuracy, the jet datasets were used to determine the scale dependence of the strong coupling. No deviation from the behaviour predicted by QCD is observed.

An earlier version of this analysis was made public as a ZEUS preliminary report.[208] The full publication, including the final results presented in this thesis, was recently made public as a preprint[209] and is currently undergoing peer review for publication.

The measured cross sections are the core result of this thesis. They are a valuable input to future QCD analyses and will become part of the legacy of HERA. It has been demonstrated that this dataset allows highly precise determinations of the strong coupling constant. It will be very interesting to see future analyses combining the ZEUS jet datasets with measurements from other experiments, such as H1 or the LHC experiments. Since the current determination is limited by the experimental/fit uncertainty, the addition of further datasets has the potential to improve the precision of the analysis notably. When adding these datasets to the fit, the treatment of the theoretical uncertainty is likely to become even more relevant. It is highly desirable to develop a further improved treatment of this uncertainty, as this would lead to another significant reduction in the total uncertainty of QCD analyses.

APPENDIX A   Corrections to QCD amplitudes

In this section, the one-loop corrections to the quark and gluon propagators and the quark-gluon vertex are computed. They are required for the renormalisation of QCD, discussed in section 2.4.4. The following six diagrams need to be evaluated:

$$i\Pi^{\mu\nu}(q) = \text{<image>}$$

$$i\Sigma(q) = \text{<image>},$$

$$ig\Gamma^{\mu}(q) = \text{<image>}.$$

The correction to the gluon propagator includes a contribution from a Fadeev-Popov ghost.[39,56] Ghosts are not introduced in the main text, as they are not relevant to the discussion there. Ghosts are additional fields added to the theory to preserve unitarity and never appear as external particles. They are scalar fermions that transform under the adjoint representation of $SU(3)$, i.e. they carry the same colour charge as gluons. Their Feynman rules are

In the following, the renormalised variants of all quantities are considered. The evaluation is performed in dimensional regularisation in $4 - 2\varepsilon$ dimensions and using Feynman parameterisation. Formulae for commonly encountered integrals are available in the literature.[39,49,56] Furthermore, the $SU(3)$ colour factors must be computed.[210] Colour factors are usually written in terms of the Casimir operators $C_{\mathrm{F}} = \frac{N^2-1}{2N}$ and $C_{\mathrm{A}} = N$, to make it clear how the group structure affects the results. For QCD, $N = 3$ is inserted in the end.

Vacuum polarisation

Quark loop   To evaluate a diagram, start by labelling all internal and external vertices and momenta. Then use the Feynman rules from figure 2.7 to assign a term to each element of the diagram. The external propagators do not need to be considered here. In the following,

the symbols $\mu, \nu, \rho, \ldots$ are used to denote spacetime indices, $\alpha, \beta, \gamma, \ldots$ are spinor indices, $i, j, k, \ldots$ are quark-colour indices and $a, b, c, \ldots$ are gluon-colour indices. The expression for the quark loop correction to the vacuum polarisation is

$$
\left(i\Pi^{\mu\nu,ab}(q)\right)^{(\text{quark})} =
$$

$$
= (-1)n_{\text{f}} \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \left(\frac{i(\slashed{k}-\slashed{q})_{\alpha\delta}\delta_{il}}{(k-q)^2}\right) \left(-ig\frac{\lambda^b_{lk}}{2}\gamma^\nu_{\delta\gamma}\right) \left(\frac{i\slashed{k}_{\gamma\beta}\delta_{kj}}{k^2}\right) \left(-ig\frac{\lambda^a_{ji}}{2}\gamma^\mu_{\beta\alpha}\right)
$$

$$
= -g^2 n_{\text{f}} \text{Tr}\left(\frac{\lambda^a}{2}\frac{\lambda^b}{2}\right) \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{\text{Tr}\left(\gamma^\nu\slashed{k}\gamma^\mu(\slashed{k}-\slashed{q})\right)}{k^2(k-q)^2}.
$$

Here, the closed fermion loop introduces an additional factor of -1. The factor n_{f} arises from the summation over quark flavours. The colour factor of this diagram is $\text{Tr}\left(\frac{\lambda^a}{2}\frac{\lambda^b}{2}\right) = \frac{1}{2}\delta^{ab}$. The denominator of the integral is transformed using Feynman parameterisation

$$
\frac{1}{k^2(k-q)^2} = \int_0^1 \mathrm{d}x \frac{1}{x(k-q)^2 + (1-x)k^2} = \int_0^1 \mathrm{d}x \frac{1}{l^2 - C},
$$

with $l = k - xq$ and $C = -x(1-x)q^2$. Since the loop momentum is integrated to infinity, the integration variable can be shifted from k to l. The denominator is even under $l \to -l$. Therefore, terms that are odd in l cancel in the numerator. In particular, $l^\mu l^\nu \to \frac{1}{4}g^{\mu\nu}l^2$. The numerator becomes

$$
\text{Tr}\left(\gamma^\nu\slashed{k}\gamma^\mu(\slashed{k}-\slashed{q})\right) \to -2l^2 g^{\mu\nu} - 4x(1-x)(2q^\mu q^\nu - g^{\mu\nu}q^2).
$$

The resulting integral can be solved using standard formulae. The result is

$$
\left(i\Pi^{\mu\nu,ab}(q)\right)^{(\text{quark})} = i\frac{2}{3}n_{\text{f}}\delta^{ab}f(q)\left(q^\mu q^\nu - g^{\mu\nu}q^2\right),
$$

where

$$
f(q) = \frac{g^2}{16\pi^2}\Gamma(\varepsilon)\left(\frac{-q^2}{4\pi\mu^2}\right)^{-\varepsilon} = \frac{g^2}{16\pi^2}\left(\frac{1}{\varepsilon} - \gamma_{\text{E}} + \log(4\pi) - \log\left(\frac{-q^2}{\mu^2}\right)\right),
$$

and γ_{E} is the Euler-Mascheroni constant.

Gluon loop The diagram involving a gluon loop is computed similarly. It receives an additional combinatoric factor of $\frac{1}{2}$. Its colour factor is $f^{adc}f^{bcd} = -C_{\mathrm{A}}\delta^{ab}$.

$$
\left(i\Pi^{\mu\nu,ab}(q)\right)^{(\text{gluon})} =
$$

$$
= \frac{1}{2}\int \frac{\mathrm{d}^4 k}{(2\pi)^4}\left(\frac{-ig_{\sigma\lambda}\delta_{ed}}{k^2}\right)\left(\frac{-ig_{\rho\kappa}\delta_{cf}}{(k-q)^2}\right)
$$
$$
\cdot\left(-gf^{adc}\big(g^{\mu\lambda}(q+k)^\rho + g^{\lambda\rho}(-k-k+q)^\mu + g^{\rho\mu}(k-q-q)^\lambda\big)\right)
$$
$$
\cdot\left(-gf^{bfe}\big(g^{\nu\kappa}(-q+k-q)^\sigma + g^{\kappa\sigma}(-k+q-k)^\nu + g^{\sigma\nu}(k+q)^\kappa\big)\right)
$$
$$
= -i\frac{C_A}{12}\delta^{ab}f(q)(22q^\mu q^\nu - 19q^2 g^{\mu\nu}).
$$

Ghost loop The ghost loop diagram has a similar structure and the same colour factor as the gluon loop diagram. Since ghosts are fermions, it receives a factor -1.

$$
\left(i\Pi^{\mu\nu,ab}(q)\right)^{(\text{ghost})} =
$$

$$
= (-1)\int \frac{\mathrm{d}^4 k}{(2\pi)^4}\left(-gf^{adc}k^\mu\right)\left(-gf^{bfe}(k-q)^\nu\right)\left(\frac{-i\delta_{ed}}{k^2}\right)\left(\frac{-i\delta_{cf}}{(k-q)^2}\right)
$$
$$
= i\frac{C_A}{12}\delta^{ab}f(q)(2q^\mu q^\nu + q^2 g^{\mu\nu})
$$
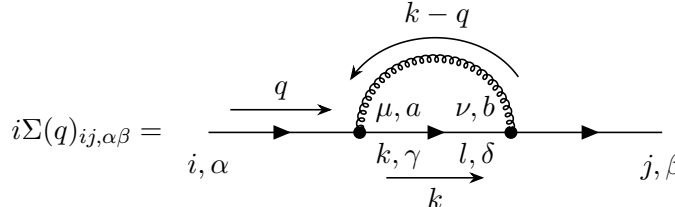
Complete vacuum polarisation Combining these three expressions yields

$$
i\Pi^{\mu\nu,ab}(q) = i\delta^{ab}f(q)\left(\frac{2}{3}n_{\mathrm{f}} - \frac{5}{3}C_A\right)(q^\mu q^\nu - q^2 g^{\mu\nu}),
$$

which is the result quoted in equation (2.25a) in the main text.

Self energy

The quark self-energy diagram has a similar structure to the vacuum polarisation diagrams. It has a colour factor of $\left(\frac{\lambda^a}{2}\frac{\lambda^a}{2}\right)_{ji} = C_F \delta_{ji}$.
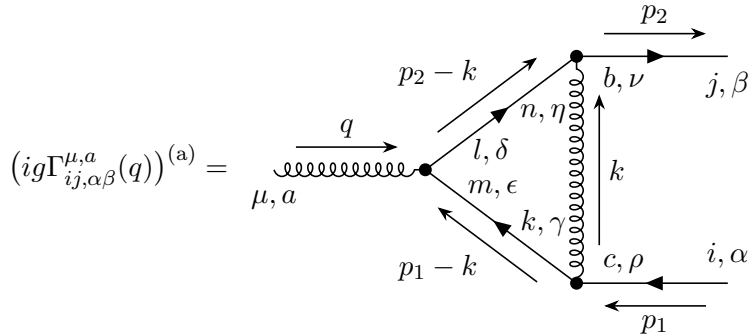
$$
\begin{aligned}
i\Sigma(q)_{ij,\alpha\beta} &= \\
&= \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \left(-ig\frac{\lambda_{jl}^b}{2}\gamma_{\beta\delta}^\nu\right)\left(\frac{i\slashed{k}_{\delta\gamma}\delta_{lk}}{k^2}\right)\left(-ig\frac{\lambda_{ki}^a}{2}\gamma_{\gamma\alpha}^\mu\right)\left(\frac{-ig^{\mu\nu}\delta_{ab}}{(k-q)^2}\right) \\
&= iC_F f(q)\delta_{ji}\slashed{q}_{\beta\alpha}.
\end{aligned}
$$

This is the result quoted in equation (2.25b) in the main text.

Vertex correction

Abelian diagram The Abelian vertex correction diagram is identical to the corresponding diagram in QED, except for the colour factor.

$$
\begin{aligned}
\left(ig\Gamma_{ij,\alpha\beta}^{\mu,a}(q)\right)^{(a)} &= \\
&= \int \frac{\mathrm{d}^4 k}{(2\pi)^4}\left(-ig\frac{\lambda_{jn}^b}{2}\gamma_{\beta\eta}^\nu\right)\left(\frac{i(\slashed{p}_2-\slashed{k})_{\eta\delta}\delta_{nl}}{(p_2-k)^2}\right)\left(-ig\frac{\lambda_{lm}^a}{2}\gamma_{\delta\varepsilon}^\mu\right) \\
&\qquad\cdot\left(\frac{i(\slashed{p}_1-\slashed{k})_{\varepsilon\gamma}\delta_{mk}}{(p_1-k)^2}\right)\left(-ig\frac{\lambda_{ki}^c}{2}\gamma_{\gamma\alpha}^\rho\right)\left(\frac{-ig_{\nu\rho}\delta^{bc}}{k^2}\right) \\
&= -g^3\left(\frac{\lambda^b}{2}\frac{\lambda^a}{2}\frac{\lambda^b}{2}\right)_{ji}\int \frac{\mathrm{d}^4 k}{(2\pi)^4}\frac{\left(\gamma^\nu(\slashed{p}_2-\slashed{k})\gamma^\mu(\slashed{p}_1-\slashed{k})\gamma_\nu\right)_{\beta\alpha}}{k^2(p_1-k)^2(p_2-k)^2}
\end{aligned}
$$

The colour factor is $\left(\frac{\lambda^b}{2}\frac{\lambda^a}{2}\frac{\lambda^b}{2}\right)_{ji} = -\frac{1}{2N}\frac{\lambda^a}{2} = -(\frac{1}{2}C_A - C_F)\frac{\lambda^a}{2}$. The integration requires a different Feynman parameterisation since the loop contains three propagators. Terms involving p_i^2 can be neglected because external particles are taken to be on-shell:

$$\frac{1}{k^2(p_1-k)^2(p_2-k)^2} = 2\int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \int_0^1 \mathrm{d}z \frac{\delta(1-x-y-z)}{(xk^2+y(p_1-k)^2+z(p_2-k)^2)^3}$$
$$= 2\int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \int_0^1 \mathrm{d}z \frac{\delta(1-x-y-z)}{\left(l^2-C\right)^3}.$$
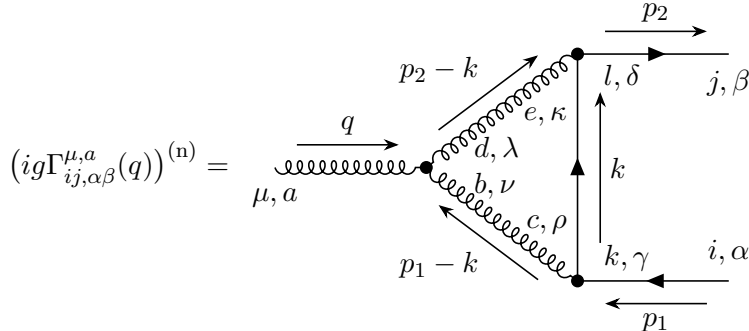
with $l = k - yp_1 - zp_2$ and $C = -yzq^2$. Due to the high mass dimension of the denominator, terms independent of l in the numerator will not diverge at large loop momenta. They are irrelevant to the present discussion and are not evaluated further. Shifting the integration variable leads to

$$\gamma^\nu(\slashed{k}+\slashed{p}_2)\gamma^\mu(\slashed{k}+\slashed{p}_1)\gamma_\nu \rightarrow -\frac{1}{2}l^2\gamma^\mu + \mathcal{O}(1)$$

The resulting integral can be evaluated using standard formulae:

$$\left(ig\Gamma^{\mu,a}_{ij,\alpha\beta}(q)\right)^{(\mathrm{a})} = igf(q)(\frac{1}{2}C_A - C_F)\frac{\lambda^a}{2}\gamma^\mu_{\beta\alpha} + (\text{UV finite})$$

Non-Abelian diagram   The non-Abelian diagram is evaluated analogously to the Abelian one. It has a colour factor of $f^{abc}\left(\frac{\lambda^c}{2}\frac{\lambda^b}{2}\right)_{ji} = -i\frac{C_A}{2}\frac{\lambda^a_{ji}}{2}$.

$$= \int \frac{\mathrm{d}^4k}{(2\pi)^4} \left(-ig\frac{\lambda^c_{ki}}{2}\gamma^\rho_{\gamma\alpha}\right)\left(-ig\frac{\lambda^e_{jl}}{2}\gamma^\kappa_{\beta\delta}\right)\left(\frac{-ig_{\nu\rho}\delta^{bc}}{(p_1-k)^2}\right)\left(\frac{-ig_{\kappa\lambda}\delta^{ed}}{(p_2-k)^2}\right)$$
$$\left(-gf^{abd}\left(g^{\mu\nu}(q-p_1+k)^\lambda + g^{\nu\lambda}(p_1+p_2-2k)^\mu + g^{\lambda\mu}(-p_2+k-q)^\nu\right)\right)$$
$$= -igf(q)\frac{3}{2}C_A\frac{\lambda^a_{ji}}{2}\gamma^\mu_{\beta\alpha} + (\text{UV finite})$$

Combined vertex correction   Combining these expressions leads to the result quoted in equation (2.25c):

$$\left(ig\Gamma^{\mu,a}_{ij,\alpha\beta}(q)\right) = -igf(q)(C_A + C_F)\frac{\lambda^a}{2}\gamma^\mu_{\beta\alpha} + (\text{UV finite}).$$

Additional figures and tables
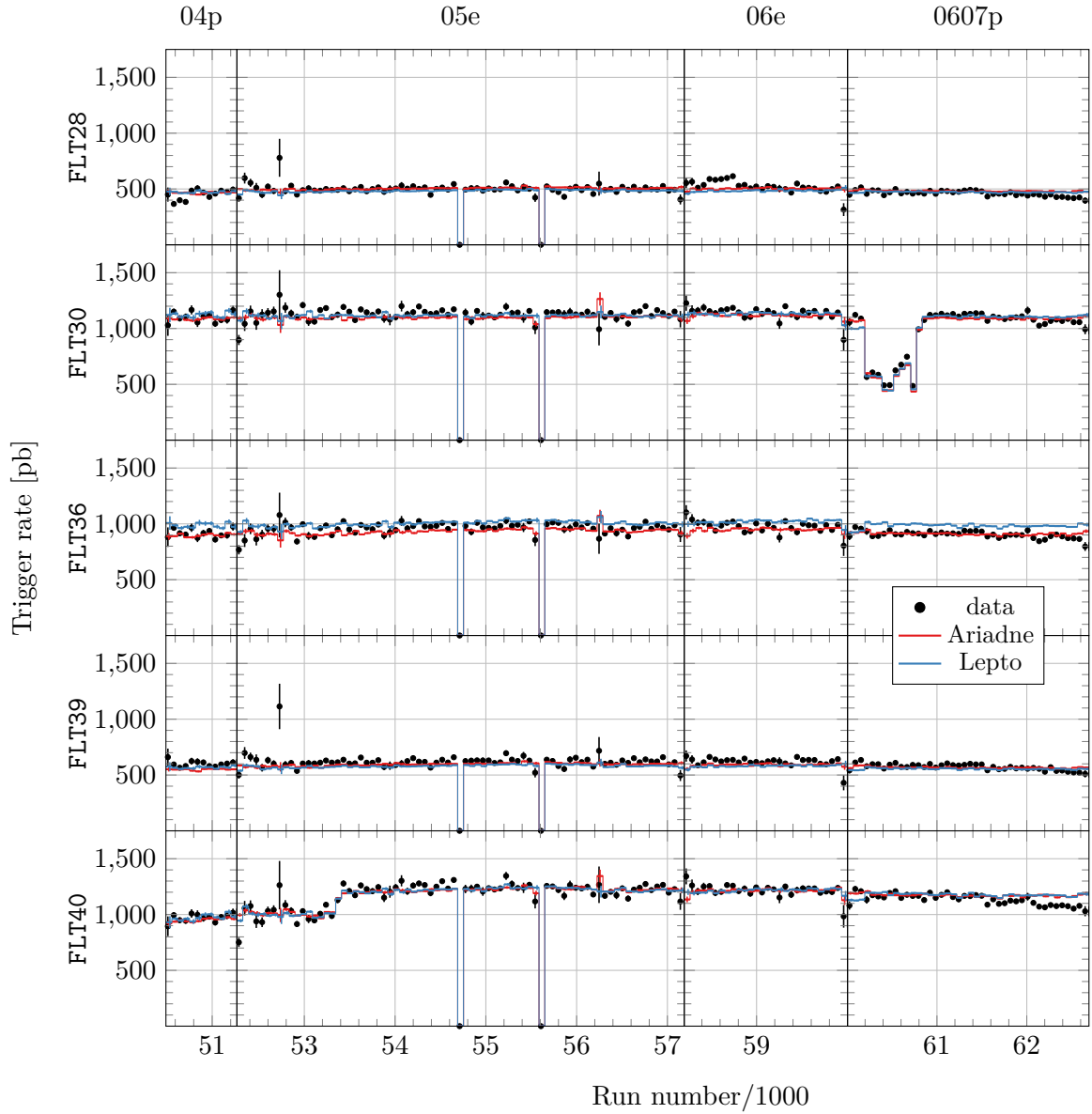
B.1 Trigger rates

Figure B.1: Trigger rates of FLT bits 28, 30, 36, 39 and 40. The rates are shown after the FLT30 efficiency correction, described in section 8.1.2
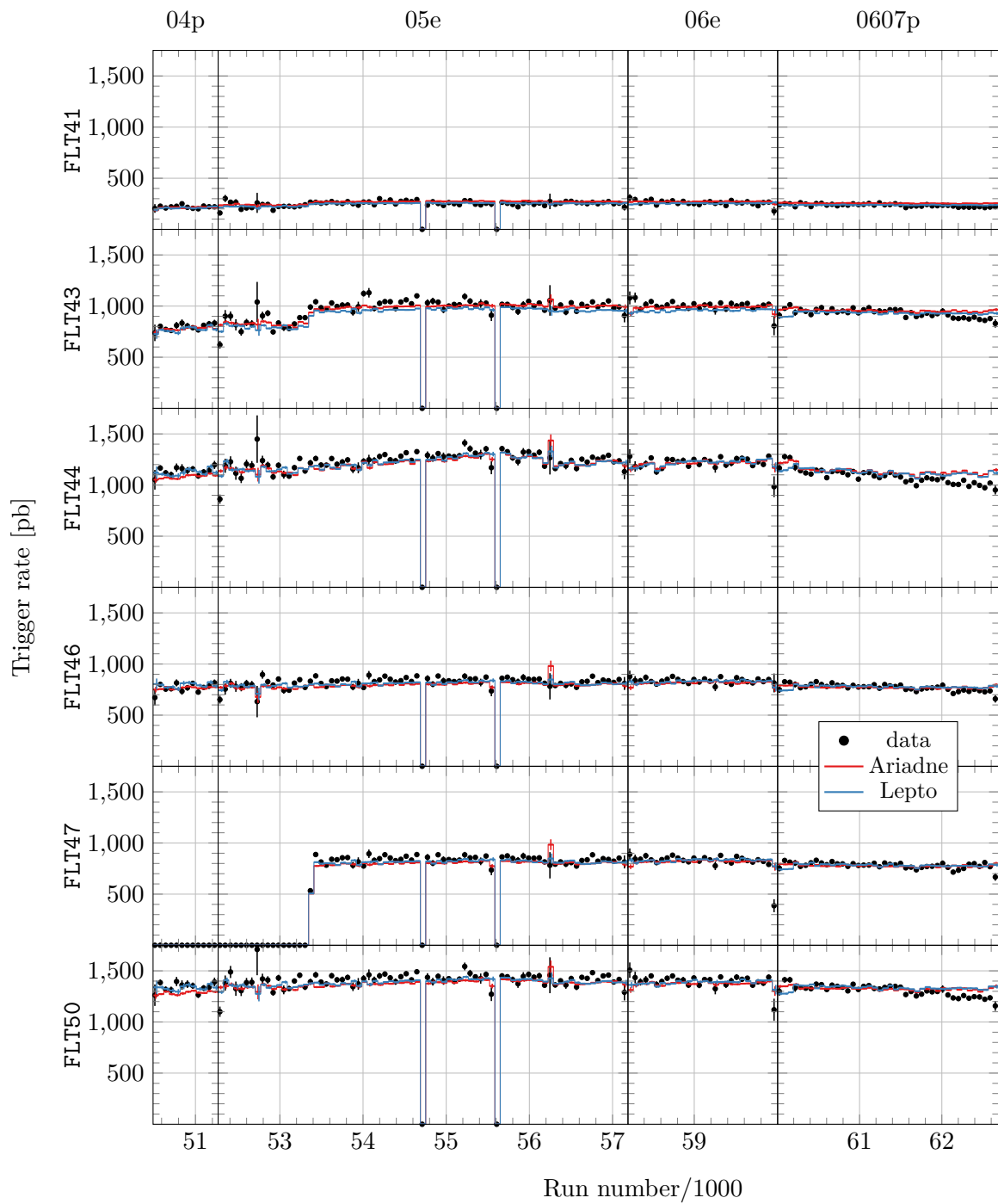
Figure B.2: Trigger rates of FLT bits 41, 43, 44, 46, 47 and 50
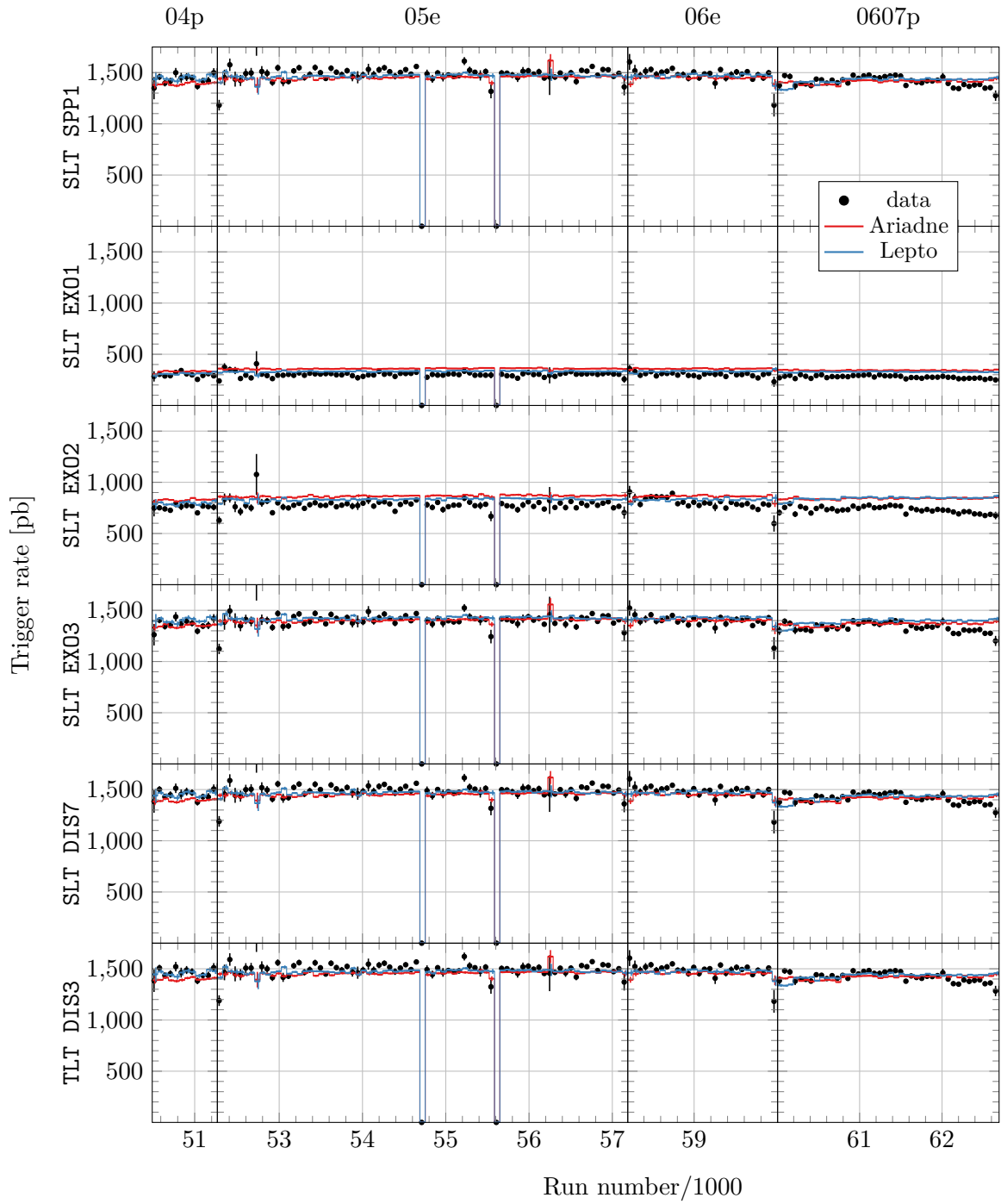
Figure B.3: Trigger rates of SLT and TLT

B.2 Jet distributions
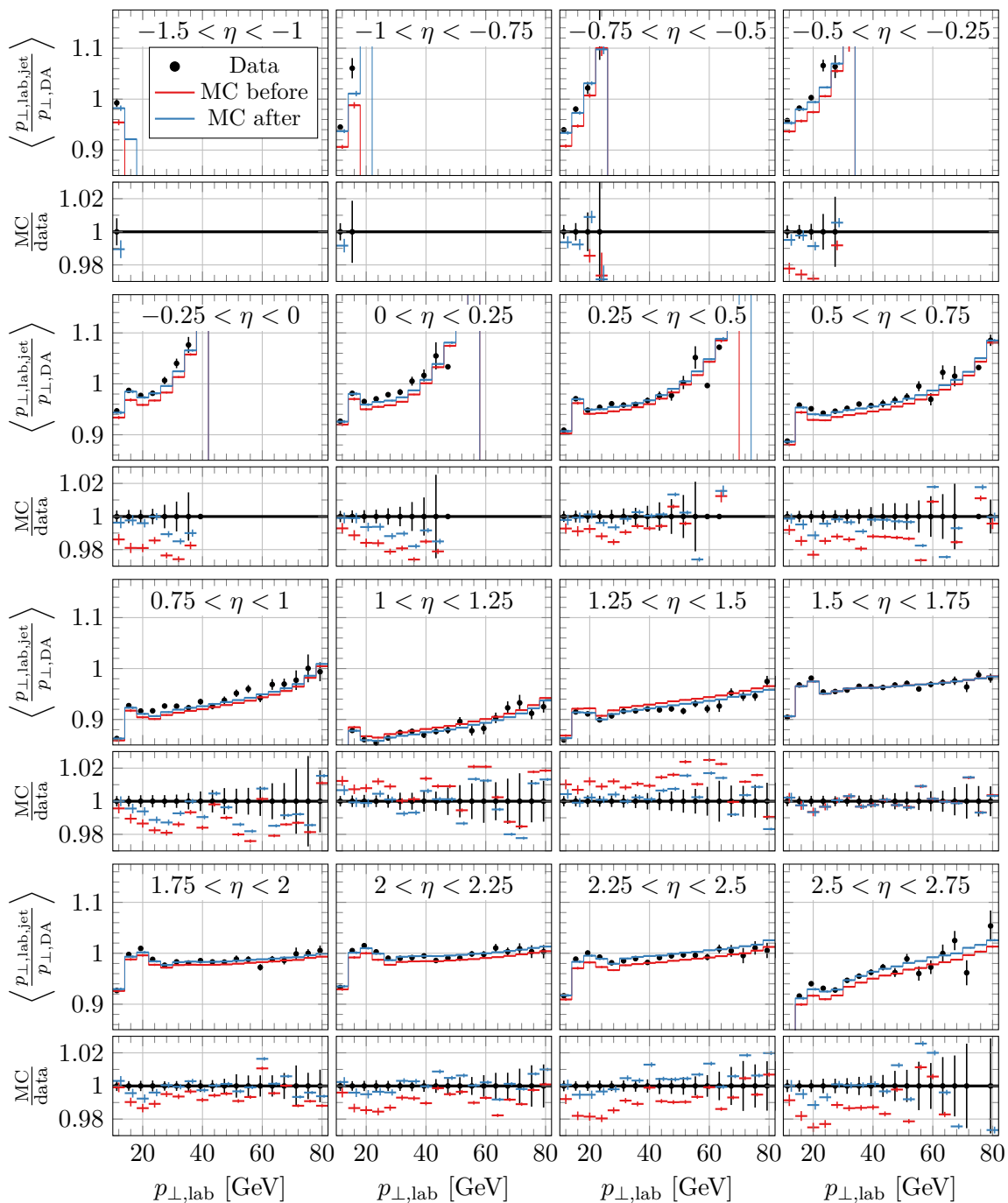
Figure B.4: Average of the ratio $\frac{p_{\perp,\text{lab,jet}}}{p_{\perp,\text{DA}}}$ as a function of $p_{\perp,\text{lab}}$ in different bins of η_{lab}. For detail see section 8.3.1.
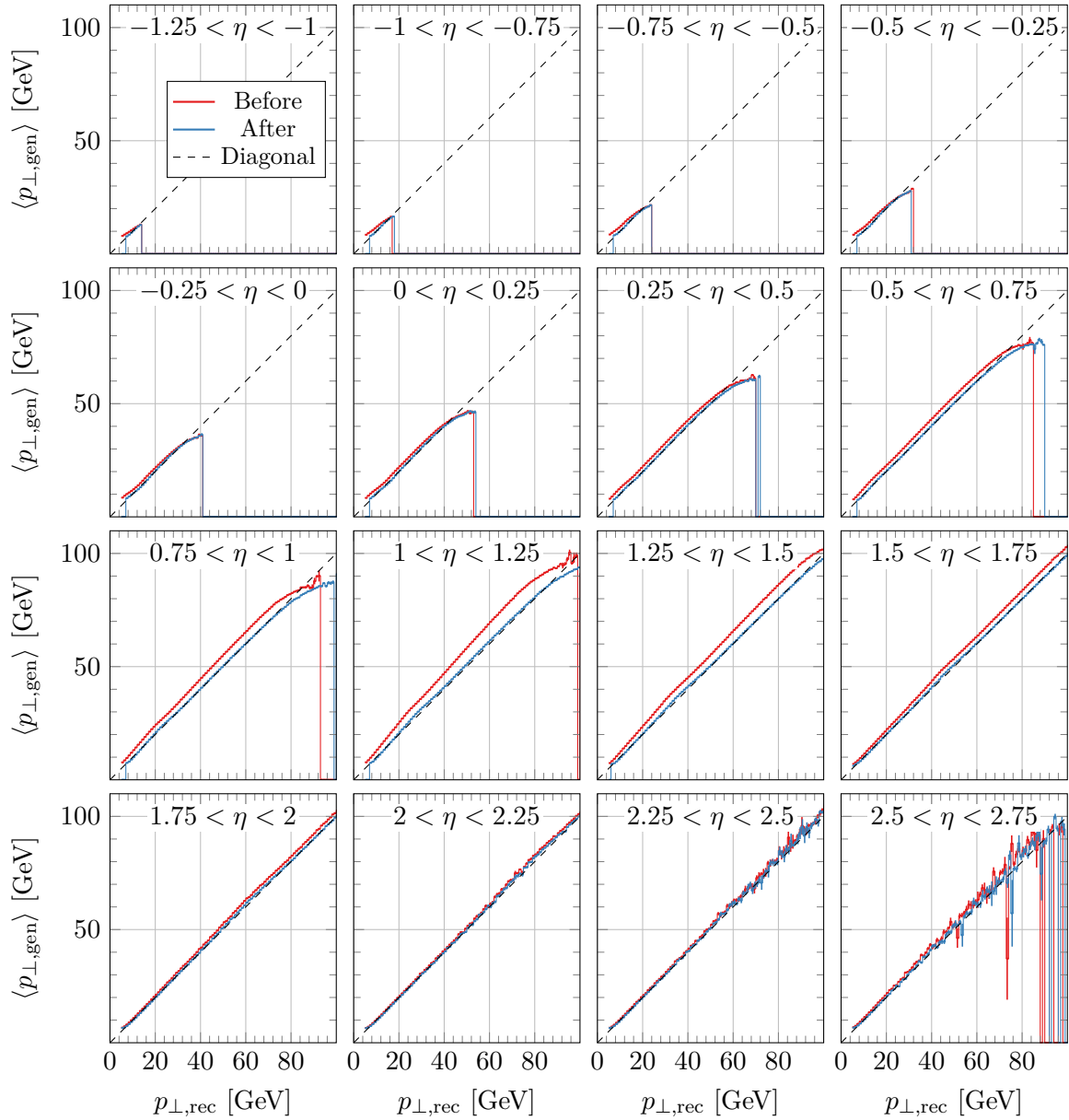
Figure B.5: Average hadron-level jet energy as a function of the detector-level jet energy in different bins of η_{lab}. For details see section 8.3.2.
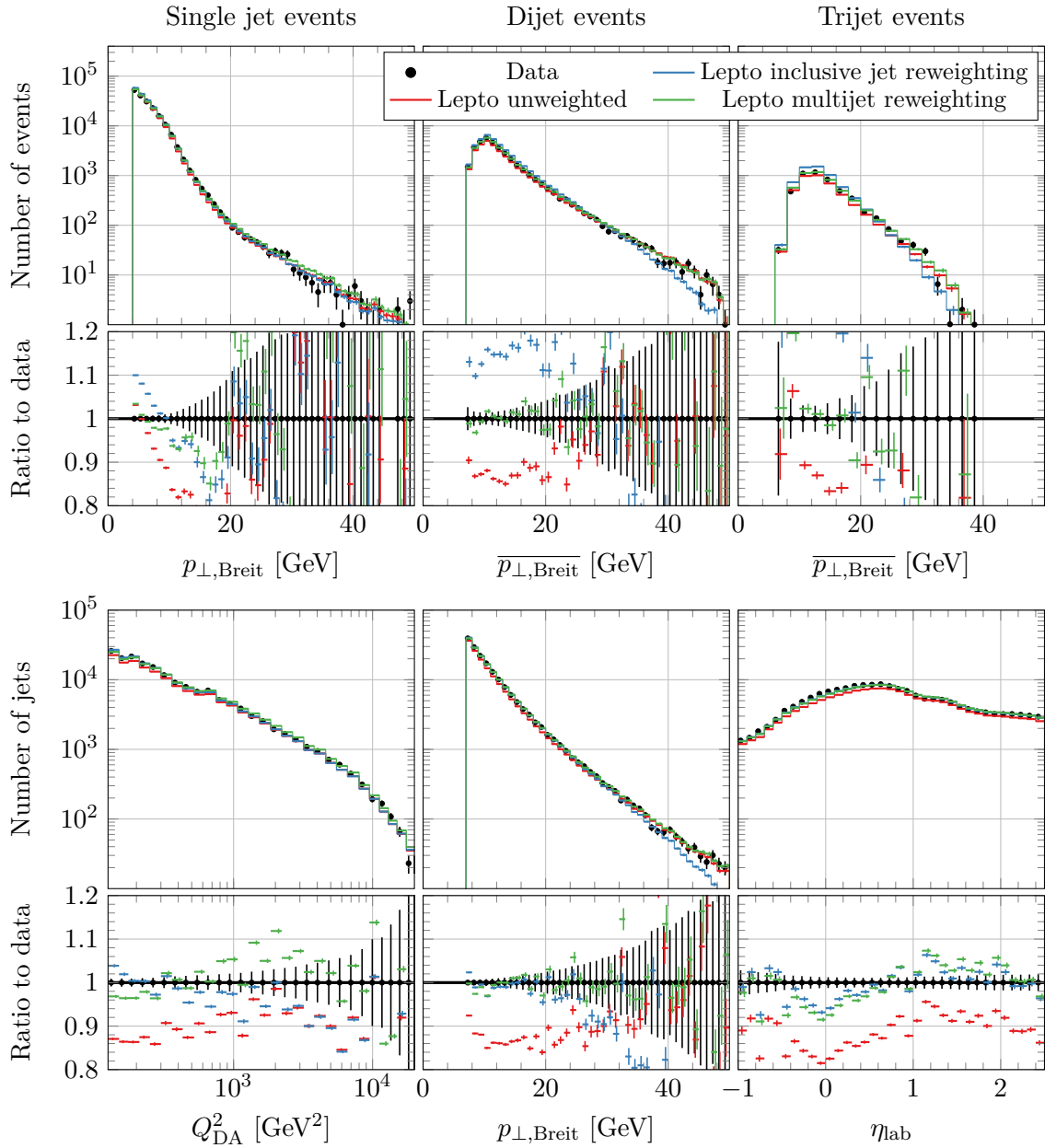
Figure B.6: Multijet (upper row) and inclusive jet (lower row) distributions before and after jet reweighting of the Lepto MC chain. For detail see figure 8.9.

B.3 Corrected data sample
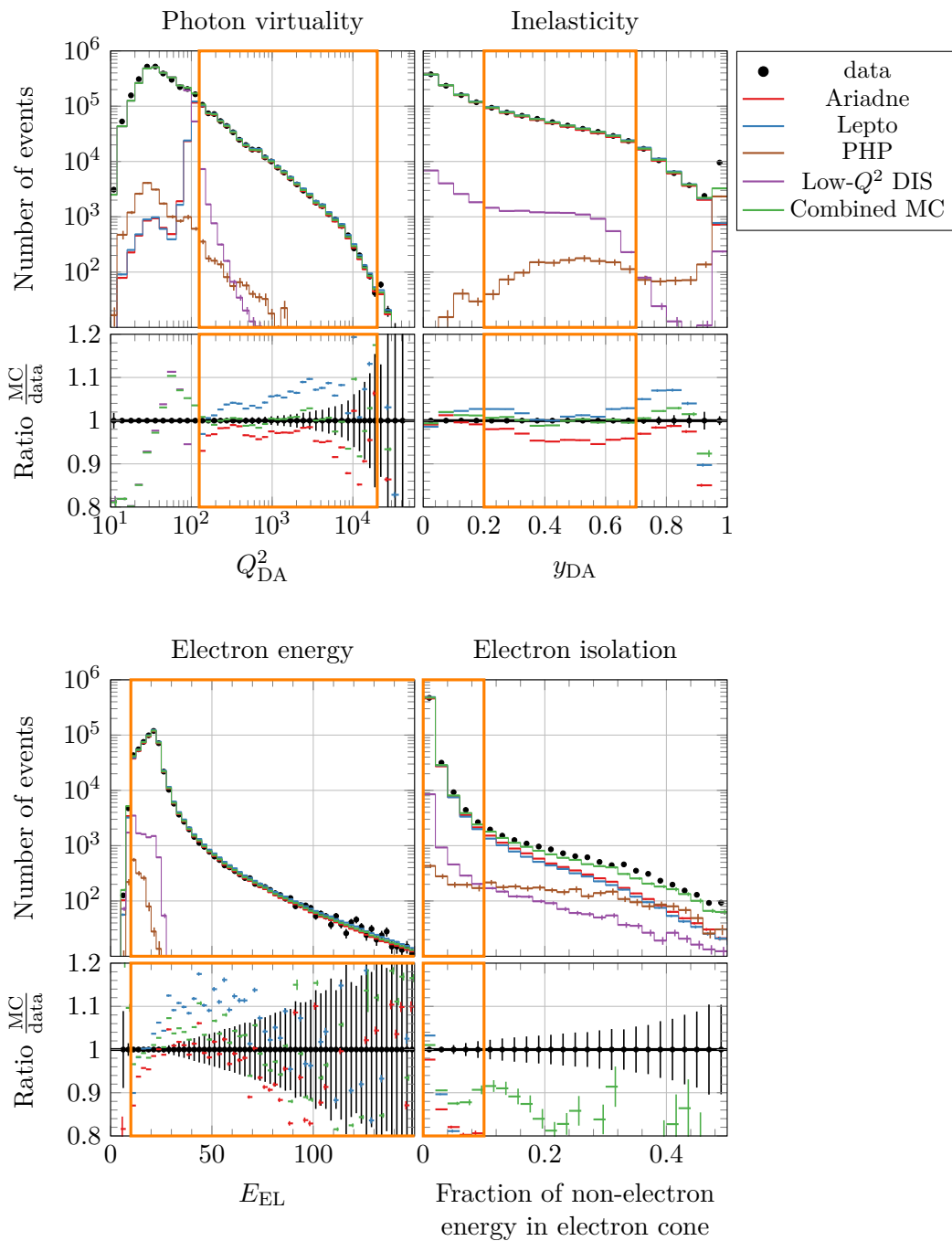
Figure B.7: Distributions of the phase space and lepton quality cuts of the inclusive DIS sample including all corrections. For details on the figure, see caption of figure 7.3. The distributions are shown before corrections in figures 7.3 and 7.4.
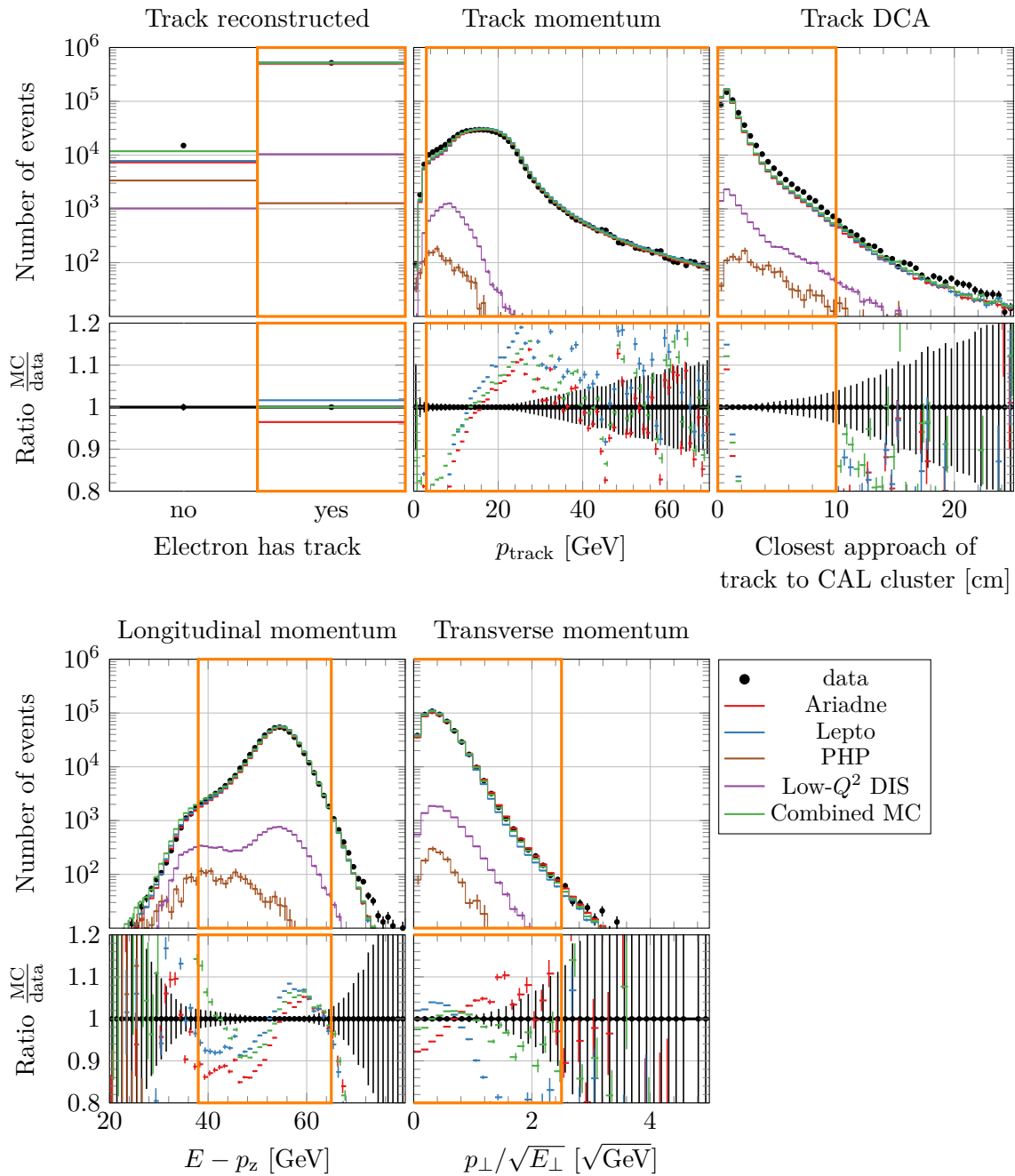
Figure B.8: Distributions of the lepton tracking cuts and DIS selection cuts of the inclusive DIS sample including all corrections. For details on the figure, see caption of figure 7.3. The distributions are shown before corrections in figures 7.5 and 7.6.
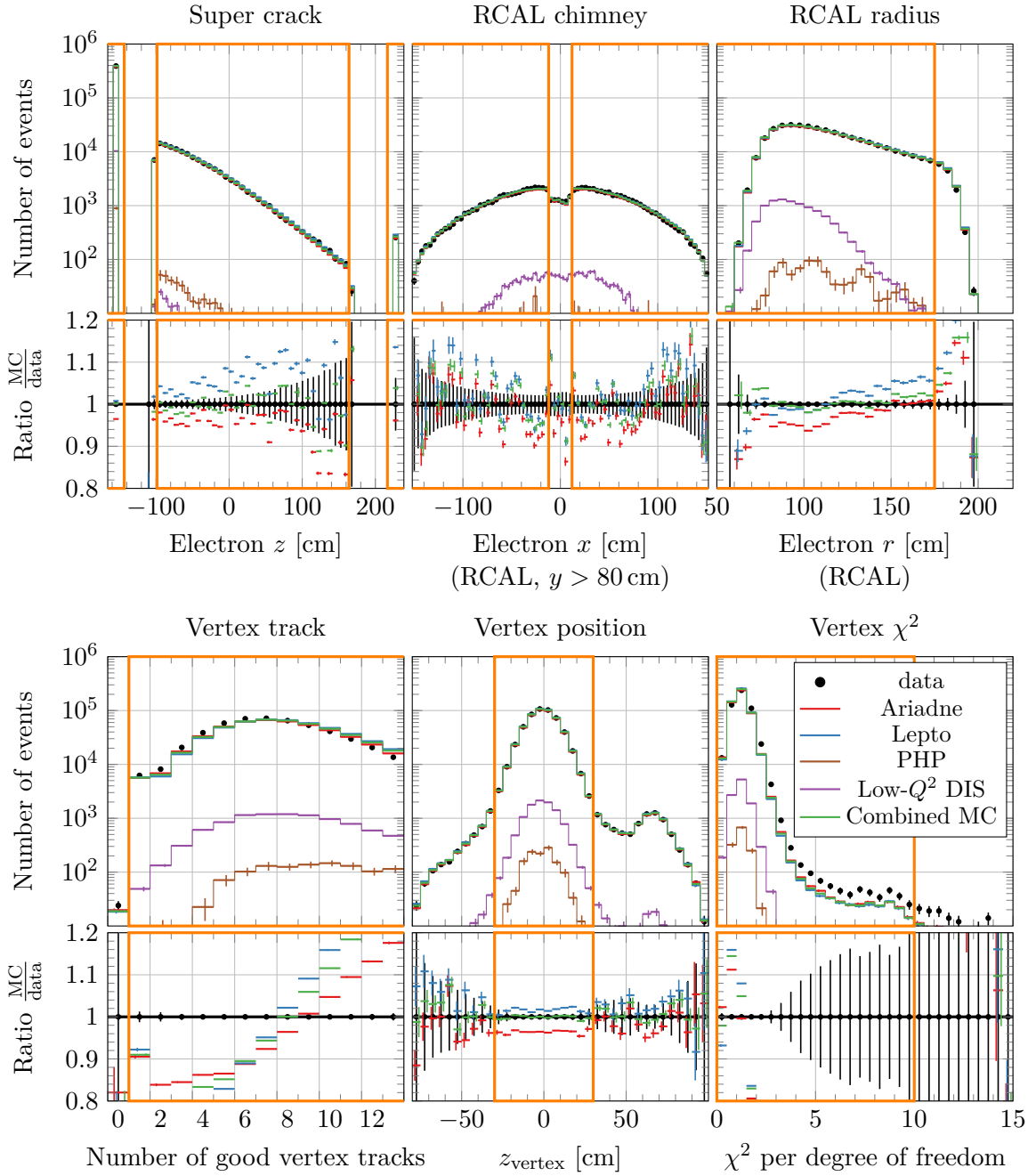
Figure B.9: Distributions of cuts on poorly measured regions of the detector and vertex-related cuts of the inclusive DIS sample including all corrections. For details on the figure, see caption of figure 7.3. The distributions are shown before corrections in figures 7.7 and 7.9.
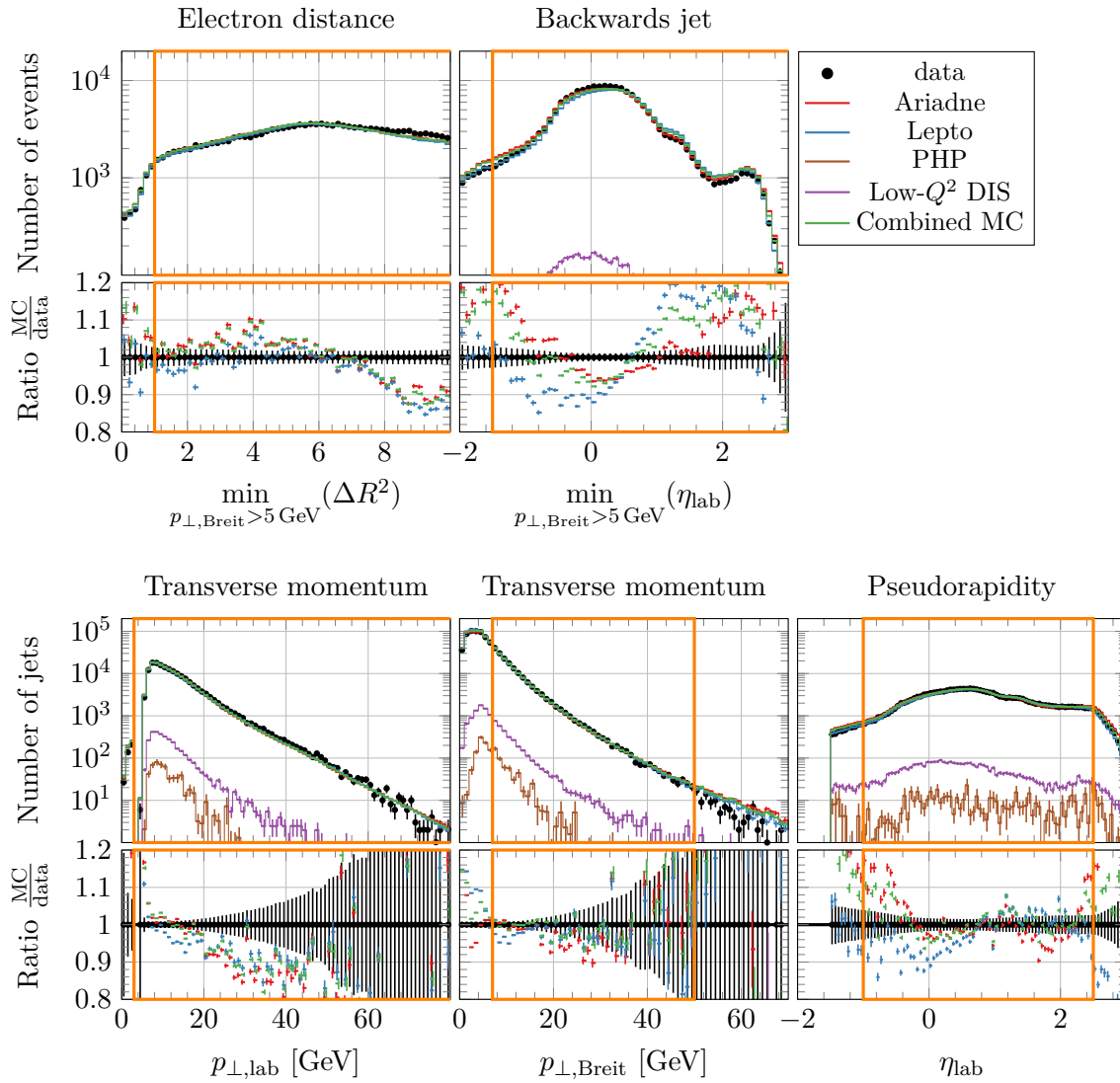
Figure B.10: Distributions of the jet veto cuts and the inclusive jet selection cuts including all corrections. For details, see captions of figures 7.11 and 7.12.
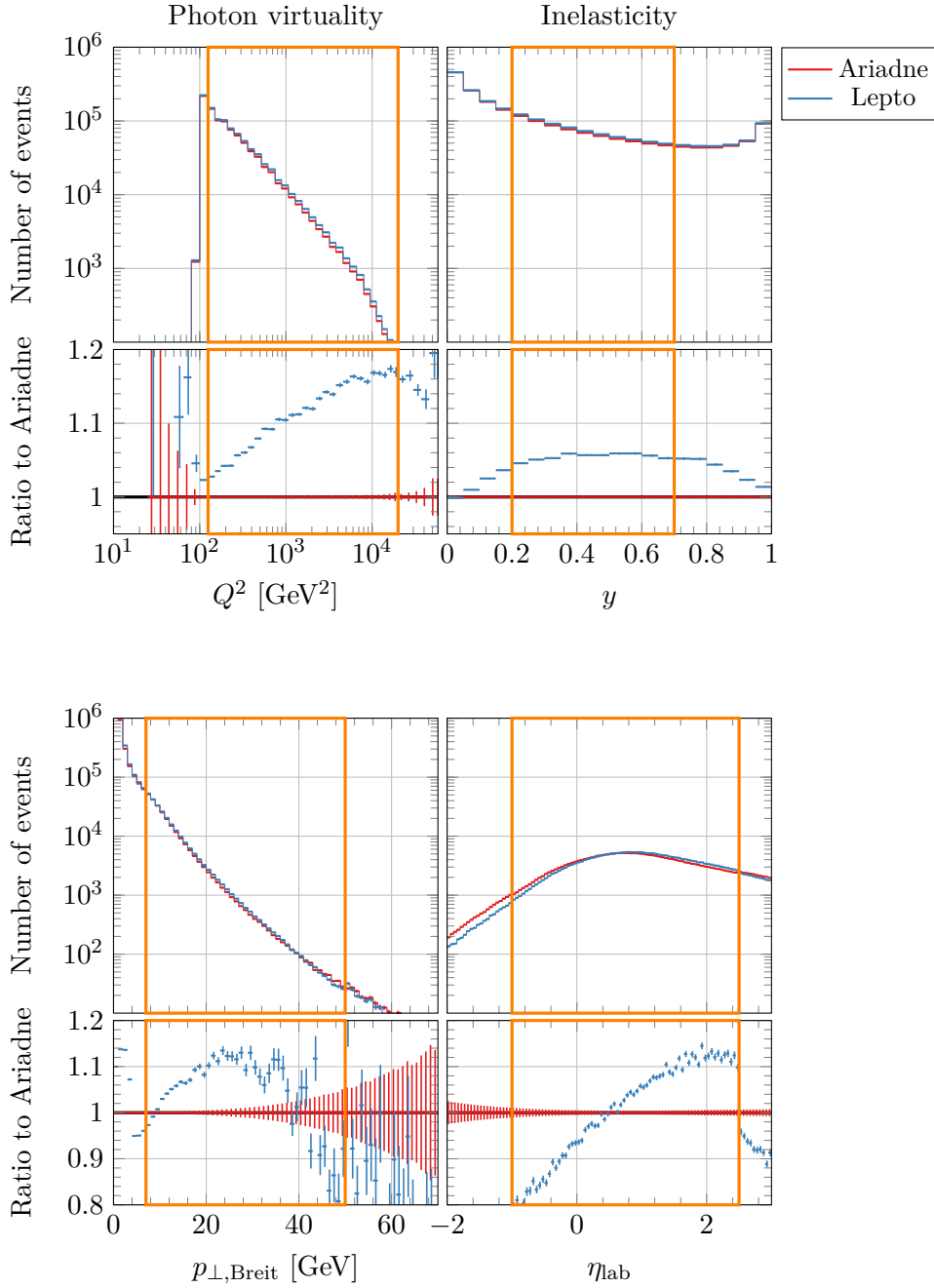
Figure B.11: Distributions of the inclusive DIS and inclusive jet cuts at hadron level including all corrections. For details, see captions of figures 7.13 and 7.14.

B.4 Bin quality
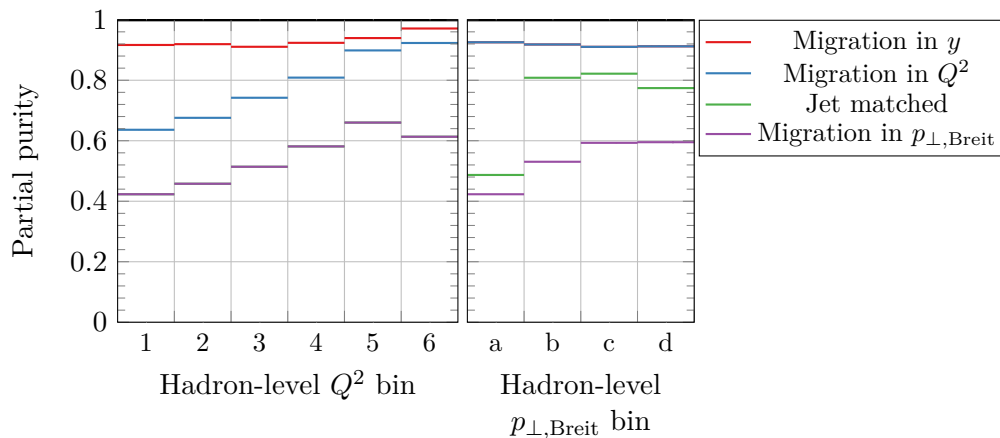
Figure B.12: Partial purity for the single differential distributions in Q^2 and $p_{\perp,\mathrm{Breit}}$ for different stages of the hadron-level reconstruction. The partial purity is the ratio of the number of detector-level jets that pass part of the hadron-level selection to the total number of detector-level jets in a given bin $N_{\mathrm{rec\&partial\ gen}}/N_{\mathrm{rec}}$. For details, see section 9.3.1. The intermediate steps considered are: 1. detector-level jets for which the corresponding hadron-level event belongs to the analysis phase space in y; 2. and the correct region in Q^2; 3. and that could be matched to a hadron-level jet, that is within the analysis phase space in $p_{\perp,\mathrm{Breit}}$ and η_{lab}; 4. and that belongs to the correct region in $p_{\perp,\mathrm{Breit}}$ (except for purity as function of Q^2). The significant contribution to the low purity comes from migrations in Q^2 and y.
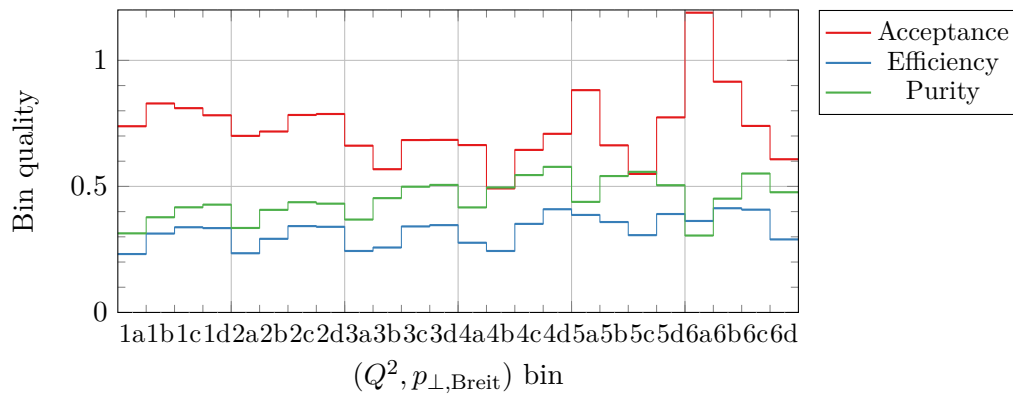
Figure B.13: Double differential bin quality. For details, see the single differential version of this distribution, shown in figure 9.1.
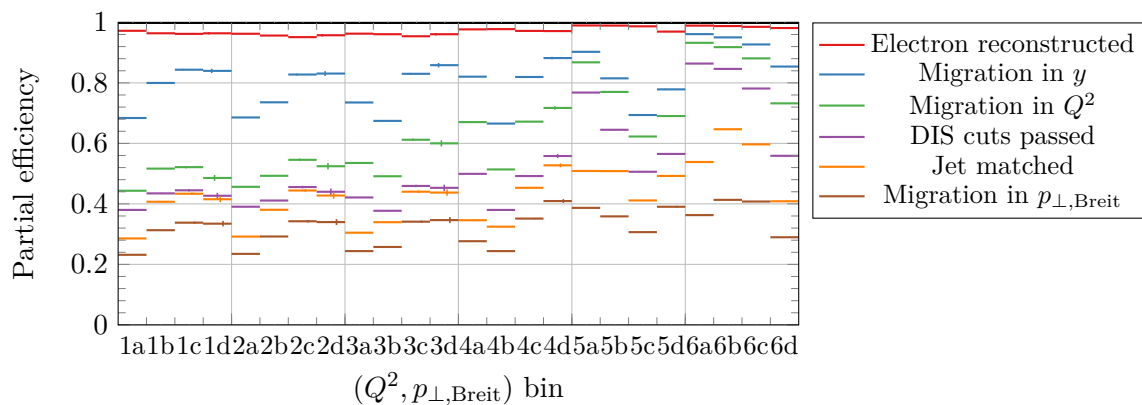
Figure B.14: Double differential partial efficiency. For details, see the single differential version of this distribution, shown in figure 9.2.
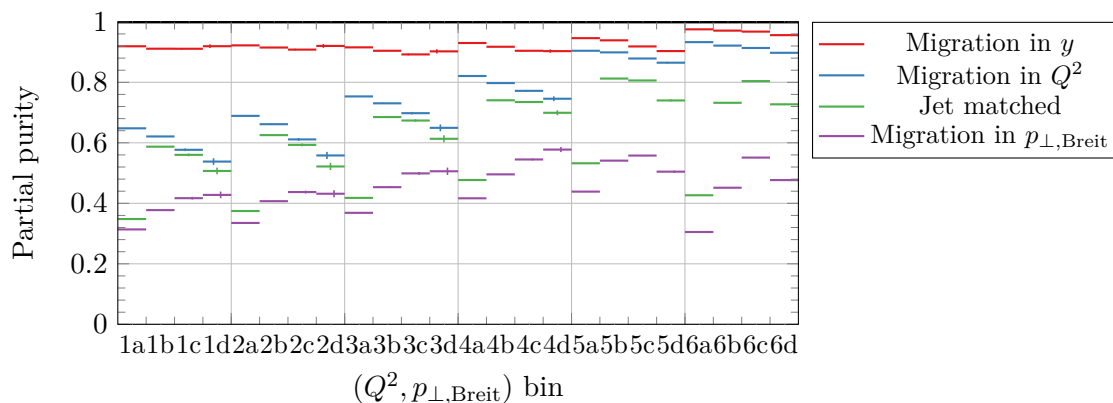
Figure B.15: Double differential partial purity. For details, see the single differential version of this distribution, shown in figure B.12.

B.5 Matrix unfolding templates with Ariadne
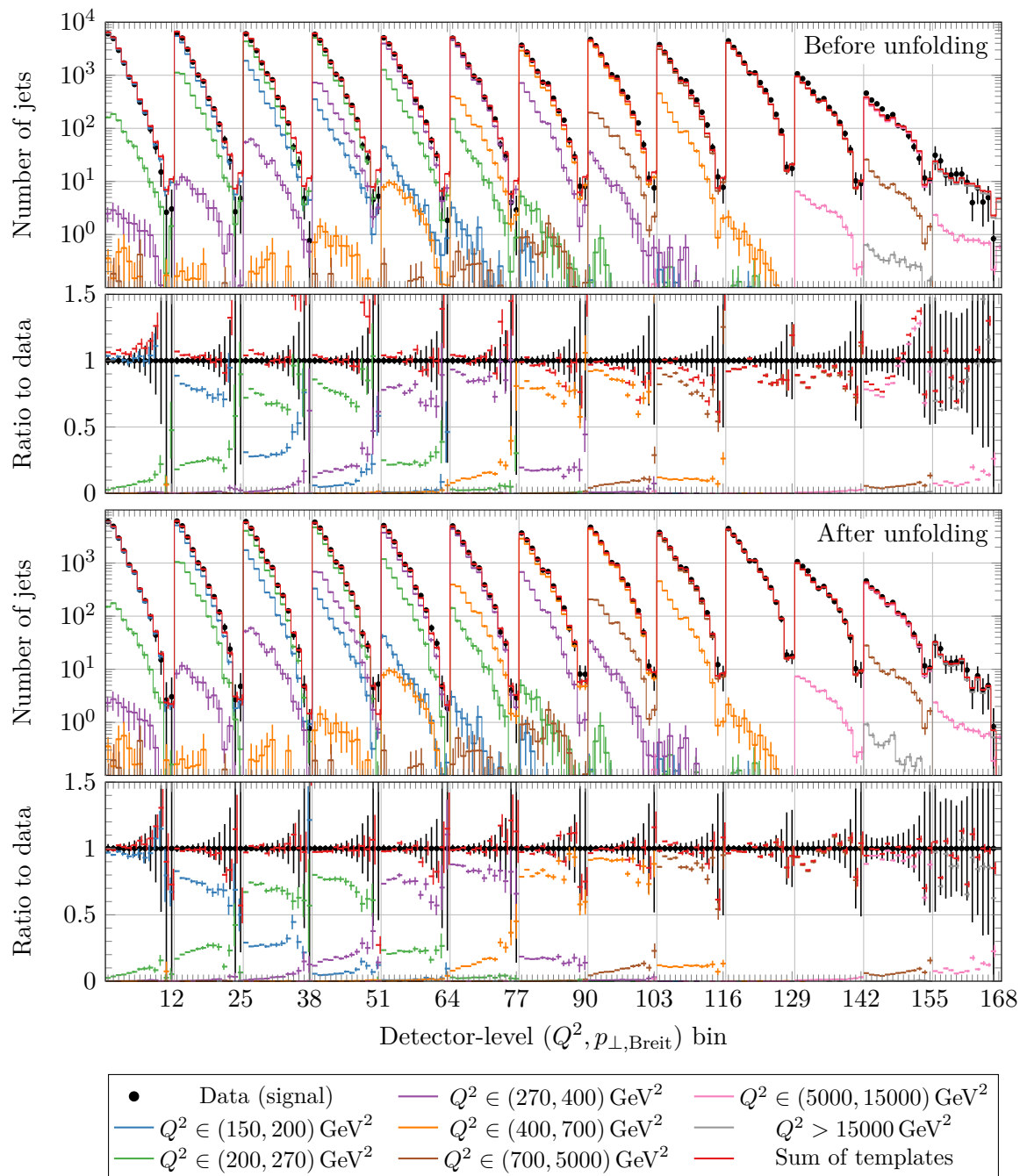
Figure B.16: Matrix unfolding as template fit in Q^2 with Ariadne MC. This figure is similar to 9.12, except that only the Ariadne signal MC sample is used, rather than a combination of Ariadne and Lepto.
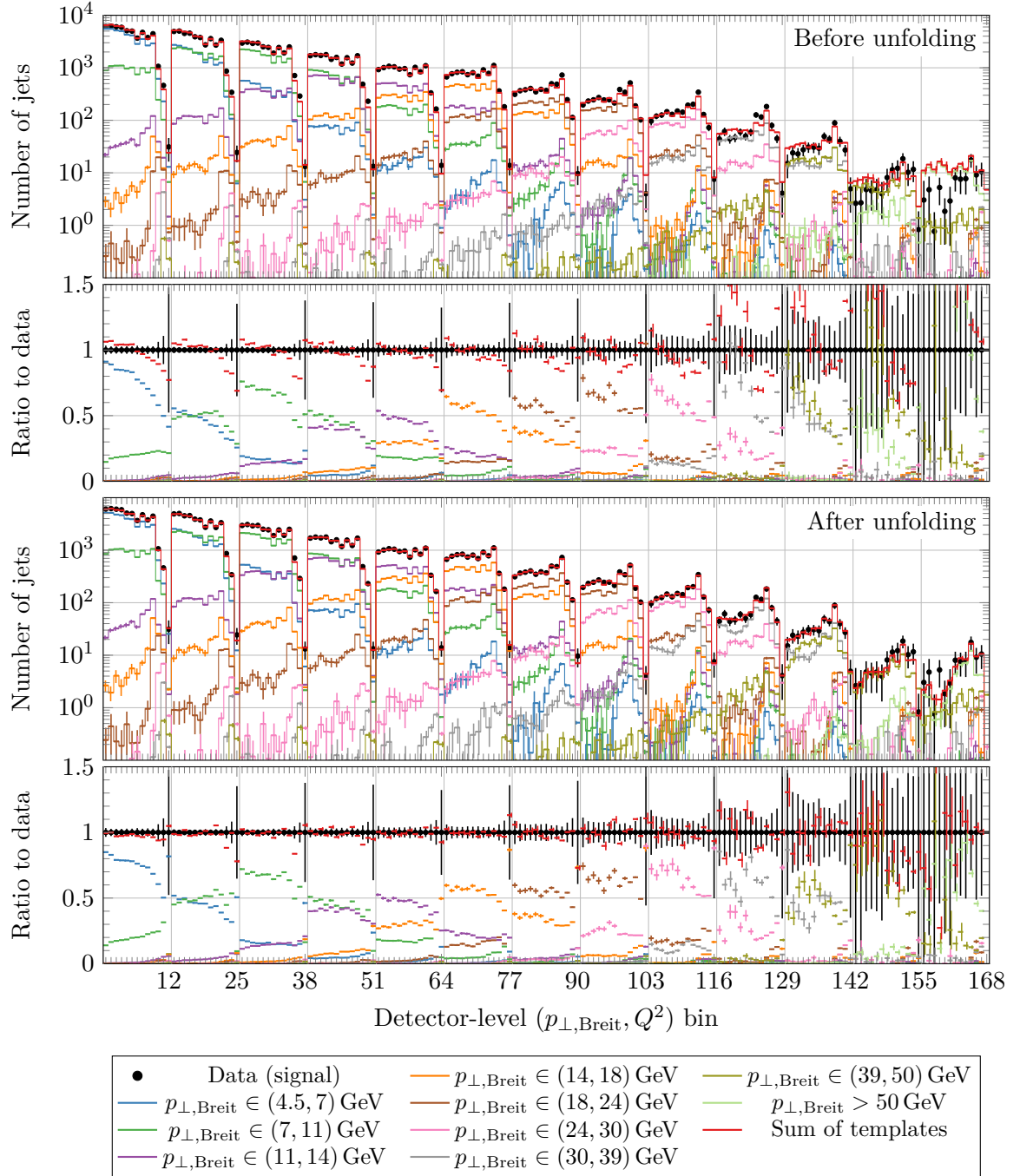
Figure B.17: Matrix unfolding as template fit in $p_{\perp,\text{Breit}}$ with Ariadne MC. The figure is similar to 9.13, except that only the Ariadne signal MC sample is used, rather than a combination of Ariadne and Lepto.

B.6 Aditional information on the α_s analysis

B.6.1 Detailed fit results

Dataset	Partial χ^2/Number of points
HERA NC e^+p DIS, $E_P = 920\,\mathrm{GeV}$	447.65/377
HERA NC e^+p DIS, $E_P = 820\,\mathrm{GeV}$	64.99/70
HERA NC e^+p DIS, $E_P = 575\,\mathrm{GeV}$	219.16/254
HERA NC e^+p DIS, $E_P = 460\,\mathrm{GeV}$	216.58/204
HERA NC e^-p DIS, $E_P = 920\,\mathrm{GeV}$	219.88/159
HERA CC e^+p DIS, $E_P = 920\,\mathrm{GeV}$	47.52/39
HERA CC e^-p DIS, $E_P = 920\,\mathrm{GeV}$	51.73/42
HERA I inclusive jets	26.38/30
HERA I/II dijets	14.65/16
HERA II inclusive jets	14.98/24
Shifts of correlated systematics	96.24
Global χ^2 per degree of freedom	1418.93/1200

Table B.1: The determined χ^2 values from the nominal fit at NNLO.

Parameter	Fitted value
B_g	-0.084941 ± 0.073962
C_g	6.947203 ± 0.747311
A'_g	0.184528 ± 0.179314
B'_g	-0.387773 ± 0.066123
B_{u_v}	0.772017 ± 0.032302
C_{u_v}	4.878890 ± 0.092664
E_{u_v}	10.310995 ± 1.416879
B_{d_v}	0.975858 ± 0.088014
C_{d_v}	4.855352 ± 0.384069
$C_{\bar{U}}$	7.524403 ± 1.491796
$D_{\bar{U}}$	3.584528 ± 2.594193
$A_{\bar{D}}$	0.255332 ± 0.011212
$B_{\bar{D}}$	-0.132296 ± 0.005217
$C_{\bar{D}}$	9.395293 ± 1.834683
$\alpha_\mathrm{s}(M_Z^2)$	0.114259 ± 0.001661

Table B.2: Fitted parameters in the nominal fit at NNLO.

	B_g	C_g	A'_g	B'_g	B_{u_v}	C_{u_v}	E_{u_v}	B_{d_v}	C_{d_v}	$C_{\bar U}$	$D_{\bar U}$	$A_{\bar D}$	$B_{\bar D}$	$C_{\bar D}$	$\alpha_\mathrm{s}(M_Z^2)$
B_g	100	29	−87	−58	9	−4	−1	2	−4	1	7	−12	−20	−6	17
C_g	29	100	−10	5	−26	29	17	−20	−9	27	48	−53	−55	−24	−63
A'_g	−87	−10	100	90	15	−7	−17	7	6	3	1	5	5	−4	−2
B'_g	−58	5	90	100	30	−16	−30	13	8	3	−3	5	0	−8	10
B_{u_v}	9	−26	15	30	100	−27	−72	40	13	33	−5	52	40	7	57
C_{u_v}	−4	29	−7	−16	−27	100	73	−16	−15	46	54	−24	−19	−5	−37
E_{u_v}	−1	17	−17	−30	−72	73	100	−27	−16	12	39	−43	−35	−12	−27
B_{d_v}	2	−20	7	13	40	−16	−27	100	88	42	25	32	27	58	19
C_{d_v}	−4	−9	6	8	13	−15	−16	88	100	23	15	20	18	56	−9
$C_{\bar U}$	1	27	3	3	33	46	12	42	23	100	89	−5	−6	19	−13
$D_{\bar U}$	1	48	1	−3	−5	54	39	25	15	89	100	−36	−35	22	−34
$A_{\bar D}$	−12	−53	5	5	52	−24	−43	32	20	−5	−36	100	96	23	36
$B_{\bar D}$	−20	−55	5	0	40	−19	−35	27	18	−6	−35	96	100	24	26
$C_{\bar D}$	−6	−24	−4	−8	7	−5	−12	58	56	19	22	23	24	100	−16
$\alpha_\mathrm{s}(M_Z^2)$	17	−63	−2	10	57	−37	−27	19	−9	−13	−34	36	26	−16	100

Table B.3: Correlations of the fitted parameters from the nominal fit at NNLO.

Dataset	Uncertainty	Shift
HERA I inclusive jets	Jet-energy scale	-1.2253 ± 0.4671
HERA I/II dijets	Jet-energy scale (20%)	-0.7079 ± 0.8817
HERA II inclusive jets	Jet-energy scale (20%)	0.9345 ± 0.8662
HERA I/II dijets + HERA II inclusive jets	Jet-energy scale (80%)	0.4650 ± 0.6133
HERA I inclusive jets	Luminosity	-0.9849 ± 0.9220
HERA I/II dijets	Luminosity (20%)	-0.3657 ± 0.9459
HERA II inclusive jets	Luminosity (20%)	0.3373 ± 0.9614
HERA I/II dijets + HERA II inclusive jets	Luminosity (80%)	-0.0568 ± 0.8426
HERA II inclusive jets	Model (50%)	0.2668 ± 0.7167
HERA II inclusive jets	Low-Q^2 DIS background	0.3632 ± 0.9447
HERA II inclusive jets	Unfolding background (50%)	0.4532 ± 0.9527
HERA II inclusive jets	$E - p_\mathrm{z}$ cut	0.5566 ± 0.8174
HERA II inclusive jets	Track-matching efficiency	0.5992 ± 0.9476
All jet datasets	Hadronisation (50%)	-0.9323 ± 0.8644
All jet datasets	Theoretical calculation (50%)	-0.5187 ± 0.9436

Table B.4: Shifts of the correlated systematic uncertainties of the jet datasets in the nominal fit at NNLO. For uncertainties that are split into multiple parts, the number in parenthesis indicates the fraction of the original uncertainty, see table 12.1.

B.6.2 Scale assignment for running α_{s} analysis

	Q^2 [GeV^2]	$p_{\perp,\mathrm{Breit}}$ [GeV]	Scale μ [GeV]				
	$125 - 250$	$8 - 10$	15.3				
	$125 - 250$	$10 - 14$	16.9				
	$125 - 250$	$14 - 18$	19.9				
	$125 - 250$	$18 - 25$		23.8			
	$125 - 250$	$25 - 100$			32.5		
	$250 - 500$	$8 - 10$	19.8				
	$250 - 500$	$10 - 14$	21.0				
	$250 - 500$	$14 - 18$		23.6			
	$250 - 500$	$18 - 25$		26.9			
	$250 - 500$	$25 - 100$			35.1		
	$500 - 1000$	$8 - 10$		26.6			
	$500 - 1000$	$10 - 14$		27.5			
	$500 - 1000$	$14 - 18$		29.5			
	$500 - 1000$	$18 - 25$			32.2		
	$500 - 1000$	$25 - 100$			39.8		
	$1000 - 2000$	$8 - 10$			36.6		
	$1000 - 2000$	$10 - 14$			37.3		
	$1000 - 2000$	$14 - 18$			38.8		
	$1000 - 2000$	$18 - 25$			40.9		
	$1000 - 2000$	$25 - 100$				47.7	
	$2000 - 5000$	$8 - 10$				52.0	
	$2000 - 5000$	$10 - 14$				52.5	
	$2000 - 5000$	$14 - 18$				53.6	
	$2000 - 5000$	$18 - 25$				55.1	
	$2000 - 5000$	$25 - 100$				60.5	
	$5000 - 100000$	$8 - 10$					84.6
	$5000 - 100000$	$10 - 14$					84.9
	$5000 - 100000$	$14 - 18$					85.7
	$5000 - 100000$	$18 - 25$					86.7
	$5000 - 100000$	$25 - 100$					89.1

HERA I inclusive jets

Table B.5: Data points from the ZEUS HERA I inclusive jet dataset.[133] For each point, the bin boundaries are given along with the scale that was assigned to this point. The position of the number in the last column indicates to which of the five scale bins the point belongs.

	Q^2 [GeV^2]	$p_\perp$ [GeV]			Scale μ [GeV]		
HERA I/II dijets	$125 - 250$	$15 - 22$	21.1				
	$125 - 250$	$22 - 30$		27.6			
	$125 - 250$	$30 - 60$			37.4		
	$250 - 500$	$15 - 22$		24.6			
	$250 - 500$	$22 - 30$		30.3			
	$250 - 500$	$30 - 60$			39.6		
	$500 - 1000$	$15 - 22$		30.4			
	$500 - 1000$	$22 - 30$			35.1		
	$500 - 1000$	$30 - 60$				43.5	
	$1000 - 2000$	$15 - 22$			39.4		
	$1000 - 2000$	$22 - 30$				43.2	
	$1000 - 2000$	$30 - 60$				50.4	
	$2000 - 5000$	$16 - 28$				54.5	
	$2000 - 5000$	$28 - 60$				61.2	
	$5000 - 20000$	$16 - 28$					84.9
	$5000 - 20000$	$28 - 60$					89.1
HERA II inclusive jets	$150 - 200$	$7 - 11$	15.5				
	$150 - 200$	$11 - 18$	18.4				
	$150 - 200$	$18 - 30$		24.8			
	$150 - 200$	$30 - 50$			37.2		
	$200 - 270$	$7 - 11$	17.2				
	$200 - 270$	$11 - 18$	20.0				
	$200 - 270$	$18 - 30$		26.0			
	$200 - 270$	$30 - 50$			38.0		
	$270 - 400$	$7 - 11$	19.6				
	$270 - 400$	$11 - 18$	22.1				
	$270 - 400$	$18 - 30$		27.6			
	$270 - 400$	$30 - 50$			39.2		
	$400 - 700$	$7 - 11$		23.6			
	$400 - 700$	$11 - 18$		25.6			
	$400 - 700$	$18 - 30$		30.5			
	$400 - 700$	$30 - 50$			41.3		
	$700 - 5000$	$7 - 11$			32.4		
	$700 - 5000$	$11 - 18$			33.9		
	$700 - 5000$	$18 - 30$			37.8		
	$700 - 5000$	$30 - 50$				47.0	
	$5000 - 15000$	$7 - 11$					82.3
	$5000 - 15000$	$11 - 18$					82.9
	$5000 - 15000$	$18 - 30$					84.5
	$5000 - 15000$	$30 - 50$					88.9

Table B.6: Data points from the ZEUS HERA I/II dijet dataset and the dataset presented in this thesis.[71] For details, see caption of table B.6.

Cross section tables

C.1 With bin-by-bin correction

Q^2 [GeV^2]	$p_{\perp,\mathrm{Breit}}$ [GeV]	σ [pb]	δ_{stat} [%]	δ_{uncor} [%]	δ_{JES} [%]	δ_{model} [%]	$\delta_{\mathrm{Bkg.}}$ [%]	δ_{E-p_z} [%]	δ_{TME} [%]	c_{QED}	c_Z	c_{Had}	δ_{Had} [%]
150–200	7–11	68.1	0.8	1.7	$^{+4.4}_{-4.1}$	2.5	1.1	$^{+0.9}_{-0.0}$	1.4	0.686	1.00	0.929	0.8
150–200	11–18	29.6	1.3	1.3	$^{+5.3}_{-5.1}$	1.5	1.2	$^{+0.8}_{-0.2}$	1.4	0.868	1.00	0.973	0.9
150–200	18–30	6.52	3.0	1.7	$^{+3.4}_{-3.2}$	1.5	1.7	$^{+1.8}_{-0.3}$	1.3	0.946	1.00	0.971	0.4
150–200	30–50	0.766	9.2	5.3	$^{+3.4}_{-4.2}$	4.4	4.8	$^{+5.3}_{+0.0}$	1.3	0.901	1.00	0.959	0.3
200–270	7–11	55.0	0.9	1.7	$^{+4.1}_{-3.9}$	3.2	0.5	$^{+0.7}_{-0.2}$	1.1	0.679	1.00	0.936	1.2
200–270	11–18	27.1	1.4	0.9	$^{+4.7}_{-4.5}$	1.4	0.4	$^{+1.1}_{-0.2}$	1.1	0.800	1.00	0.961	0.2
200–270	18–30	6.53	3.0	2.8	$^{+3.7}_{-3.4}$	2.4	0.4	$^{+2.0}_{-0.3}$	1.0	0.953	1.00	0.978	0.2
200–270	30–50	0.655	9.7	6.9	$^{+3.6}_{-4.1}$	7.9	0.6	$^{+4.5}_{-1.2}$	1.0	0.951	1.00	0.985	0.8
270–400	7–11	50.8	1.0	1.8	$^{+3.9}_{-3.6}$	2.9	0.2	$^{+0.8}_{+0.1}$	0.8	0.704	1.00	0.930	1.6
270–400	11–18	26.9	1.5	2.4	$^{+4.2}_{-4.3}$	2.9	0.3	$^{+1.0}_{-0.0}$	0.8	0.690	1.00	0.967	0.1
270–400	18–30	6.87	3.1	2.2	$^{+3.6}_{-3.7}$	0.8	0.3	$^{+1.4}_{-0.5}$	0.8	0.923	1.00	0.973	0.2
270–400	30–50	0.817	9.0	4.9	$^{+3.8}_{-2.9}$	-2.1	0.5	$^{+5.0}_{-1.6}$	0.7	0.956	1.00	0.977	0.4
400–700	7–11	44.7	1.1	1.1	$^{+3.3}_{-3.0}$	2.0	0.2	$^{+0.9}_{-0.2}$	0.6	0.755	1.01	0.916	1.4
400–700	11–18	26.9	1.5	2.3	$^{+3.5}_{-3.6}$	4.3	0.2	$^{+1.7}_{-0.0}$	0.6	0.626	1.01	0.974	0.2
400–700	18–30	8.21	2.8	1.2	$^{+3.5}_{-3.3}$	2.3	0.2	$^{+1.7}_{-0.2}$	0.6	0.854	1.01	0.972	0.3
400–700	30–50	1.35	7.0	2.8	$^{+4.5}_{-3.7}$	2.6	0.0	$^{+6.3}_{-0.1}$	0.6	0.982	1.01	0.973	0.5
700–5000	7–11	41.9	1.0	1.4	$^{+2.5}_{-2.2}$	0.9	0.1	$^{+0.4}_{-0.2}$	0.5	0.826	1.03	0.903	0.6
700–5000	11–18	28.6	1.3	1.6	$^{+2.5}_{-2.5}$	3.2	0.2	$^{+0.9}_{-0.2}$	0.5	0.702	1.03	0.967	0.7
700–5000	18–30	11.9	2.1	4.6	$^{+2.7}_{-2.7}$	4.6	0.1	$^{+1.2}_{-0.0}$	0.5	0.571	1.03	0.974	0.0
700–5000	30–50	2.11	4.6	4.8	$^{+4.6}_{-4.4}$	4.1	0.1	$^{-0.7}_{-0.4}$	0.5	0.714	1.03	0.963	0.4
5000–15000	7–11	2.31	3.6	3.0	$^{+1.6}_{-1.6}$	0.5	0.1	$^{-1.0}_{-0.5}$	0.3	0.930	1.16	0.889	0.6
5000–15000	11–18	1.97	4.6	3.3	$^{+1.3}_{-1.2}$	2.2	0.0	$^{-0.6}_{+0.0}$	0.3	0.867	1.16	0.948	0.3
5000–15000	18–30	0.988	6.6	3.2	$^{+1.5}_{-1.5}$	4.7	0.0	$^{-2.4}_{-1.6}$	0.3	0.722	1.16	0.976	0.1
5000–15000	30–50	0.248	11.7	8.1	$^{+3.5}_{-3.4}$	11.3	0.4	$^{-5.0}_{-4.1}$	0.3	0.433	1.16	0.977	0.9

Table C.1: Double-differential inclusive jet cross sections derived with the bin-by-bin correction, see section 11.1. Listed are the statistical uncertainty δ_{stat}, the sum of the uncorrelated systematic uncertainties δ_{uncor} and the correlated systematic uncertainties associated with the jet energy scale δ_{JES}, MC model δ_{model}, the background contribution $\delta_{\mathrm{Bkg.}}$, the $E - p_z$ cut δ_{E-p_z} and the track-matching efficiency correction δ_{TME}. Not listed explicitly is the luminosity uncertainty of 1.9%, which is fully correlated across all points. The last four columns show the QED Born level correction c_{QED} that has been applied to the data as well as the Z^0 and hadronisation correction and associated uncertainty, c_Z, c_{Had} and δ_{Had}, that need to be applied to the theoretical predictions.

Q^2 [GeV²]	$p_{\perp,\mathrm{Breit}}$ [GeV]	δ_1 [%]	δ_2 [%]	δ_3 [%]	δ_4 [%]	δ_5 [%]	δ_6 [%]	δ_7 [%]	δ_8 [%]	δ_9 [%]	δ_{10} [%]	δ_{12} [%]	δ_{13} [%]	δ_{14} [%]
150–200	7–11	-0.9	$^{-0.0}_{+0.1}$	+0.6	-1.1	$^{+0.4}_{-0.6}$	$^{+0.0}_{+0.3}$	$^{+0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	$^{+0.0}_{-0.0}$	-0.3	-0.0	+0.0	+0.2
150–200	11–18	-0.1	$^{-0.1}_{+0.1}$	+0.2	-1.1	$^{+0.5}_{-0.0}$	$^{+0.0}_{+0.2}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.0}$	$^{+0.0}_{-0.0}$	-0.3	-0.0	-0.0	+0.3
150–200	18–30	-0.4	$^{-0.0}_{+0.0}$	-0.4	-1.1	$^{+0.2}_{+0.1}$	$^{+0.0}_{-0.2}$	$^{+0.1}_{+0.2}$	$^{-0.2}_{-0.0}$	$^{-0.1}_{+0.0}$	-0.5	-0.0	-0.0	+0.9
150–200	30–50	+1.3	$^{+0.0}_{+0.3}$	-3.0	+2.8	$^{-0.1}_{+0.2}$	$^{+0.0}_{+1.1}$	$^{+0.4}_{+0.1}$	$^{-0.1}_{-0.8}$	$^{-0.1}_{+0.0}$	-1.6	-0.0	-0.0	+2.5
200–270	7–11	-1.0	$^{-0.0}_{+0.1}$	+0.8	-0.5	$^{+1.0}_{-1.1}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	-0.4	-0.0	-0.0	+0.1
200–270	11–18	-0.7	$^{-0.0}_{+0.0}$	-0.2	-0.3	$^{+0.1}_{-0.2}$	$^{+0.0}_{-0.0}$	$^{-0.0}_{+0.0}$	$^{-0.1}_{+0.0}$	$^{-0.1}_{-0.1}$	-0.3	-0.0	-0.0	+0.3
200–270	18–30	-0.6	$^{-0.0}_{-0.0}$	+2.6	+0.2	$^{+0.0}_{-0.2}$	$^{+0.0}_{+0.1}$	$^{+0.1}_{+0.1}$	$^{-0.1}_{-0.1}$	$^{-0.2}_{-0.8}$	-0.5	-0.0	-0.0	+0.7
200–270	30–50	+0.9	$^{-0.2}_{+0.4}$	+6.0	+1.7	$^{+0.0}_{+0.2}$	$^{+0.0}_{+1.1}$	$^{+0.0}_{+0.7}$	$^{-0.1}_{+0.1}$	$^{+0.1}_{+1.2}$	-0.3	-0.0	-0.3	+2.5
270–400	7–11	-0.9	$^{-0.1}_{+0.1}$	+1.2	-0.6	$^{+0.6}_{-0.7}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{-0.0}$	$^{+0.1}_{+0.1}$	-0.3	-0.1	-0.1	+0.1
270–400	11–18	-1.6	$^{-0.0}_{+0.0}$	+1.5	-0.7	$^{+0.4}_{-0.6}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{+0.1}_{+0.1}$	$^{+0.0}_{-0.1}$	-0.3	-0.1	-0.2	+0.2
270–400	18–30	-0.6	$^{+0.1}_{+0.1}$	+0.6	-1.7	$^{-0.1}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.2}$	$^{+0.1}_{-0.0}$	$^{-0.5}_{-0.2}$	-0.8	-0.1	-0.3	+0.6
270–400	30–50	+0.1	$^{-0.1}_{+0.3}$	+3.6	-2.1	$^{-0.1}_{+0.0}$	$^{+0.0}_{-0.0}$	$^{+0.2}_{+0.3}$	$^{-0.1}_{-0.7}$	$^{-1.0}_{-1.9}$	+0.4	-0.0	-0.6	+1.8
400–700	7–11	-0.8	$^{-0.0}_{+0.1}$	+0.2	-0.2	$^{+0.5}_{-0.7}$	$^{+0.0}_{-0.0}$	$^{-0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	$^{-0.2}_{-0.1}$	-0.1	-0.1	-0.2	+0.2
400–700	11–18	-2.2	$^{+0.0}_{+0.1}$	+0.1	-0.3	$^{+0.2}_{-0.2}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{-0.0}$	$^{-0.0}_{+0.1}$	$^{-0.2}_{-0.1}$	-0.1	-0.1	-0.2	+0.2
400–700	18–30	-0.8	$^{+0.1}_{+0.1}$	-0.4	+0.3	$^{-0.0}_{-0.1}$	$^{+0.0}_{+0.1}$	$^{-0.1}_{+0.3}$	$^{-0.1}_{-0.1}$	$^{-0.5}_{+0.2}$	-0.0	-0.1	-0.2	+0.5
400–700	30–50	+0.0	$^{+0.2}_{-0.1}$	+1.4	+1.4	$^{-0.2}_{-0.7}$	$^{+0.0}_{-0.3}$	$^{-0.2}_{+0.4}$	$^{-0.1}_{+0.2}$	$^{-0.4}_{-1.6}$	+0.2	-0.1	-0.6	+1.5
700–5000	7–11	-0.5	$^{-0.1}_{+0.1}$	+1.0	-0.1	$^{+1.0}_{-0.5}$	$^{+0.0}_{+0.1}$	$^{-0.0}_{+0.0}$	$^{+0.0}_{-0.0}$	$^{+0.0}_{+0.0}$	-0.1	-0.1	+0.3	+0.1
700–5000	11–18	-1.5	$^{-0.1}_{+0.1}$	+0.5	+0.2	$^{+0.4}_{-0.2}$	$^{+0.0}_{+0.2}$	$^{+0.1}_{+0.0}$	$^{-0.0}_{-0.0}$	$^{+0.0}_{+0.0}$	-0.1	-0.1	+0.1	+0.1
700–5000	18–30	-4.6	$^{-0.1}_{+0.1}$	+0.1	+0.1	$^{+0.2}_{-0.5}$	$^{+0.0}_{+0.0}$	$^{-0.1}_{+0.1}$	$^{-0.0}_{-0.3}$	$^{+0.0}_{+0.0}$	-0.0	-0.1	+0.0	+0.2
700–5000	30–50	-4.6	$^{-0.3}_{+0.5}$	+0.1	+0.7	$^{+2.1}_{+0.2}$	$^{+0.0}_{-0.3}$	$^{+0.0}_{+0.2}$	$^{-0.1}_{-0.6}$	$^{+0.0}_{+0.0}$	+0.0	-0.1	-0.1	+0.5
5000–15000	7–11	-1.2	$^{-0.2}_{+0.2}$	+1.5	-1.6	$^{+0.1}_{-2.2}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{-0.3}$	$^{+0.1}_{-0.2}$	$^{+0.0}_{+0.0}$	-0.1	-0.1	+1.3	+0.2
5000–15000	11–18	+0.7	$^{-0.2}_{+0.2}$	-2.7	+0.8	$^{+0.2}_{-1.1}$	$^{+0.0}_{-0.0}$	$^{+0.2}_{-1.0}$	$^{-0.2}_{-0.5}$	$^{+0.0}_{+0.0}$	-0.8	-0.1	+1.1	+0.2
5000–15000	18–30	-2.0	$^{-0.2}_{+0.1}$	+2.0	-0.1	$^{+1.1}_{+1.1}$	$^{+0.0}_{-0.0}$	$^{-0.1}_{-0.4}$	$^{-0.2}_{-0.7}$	$^{+0.0}_{+0.0}$	-0.3	-0.1	+0.7	+0.2
5000–15000	30–50	-6.6	$^{-0.2}_{+0.2}$	+4.5	+0.1	$^{+1.0}_{-0.3}$	$^{+0.0}_{-0.1}$	$^{-0.1}_{+0.1}$	$^{-0.2}_{+0.2}$	$^{+0.0}_{+0.0}$	+0.4	-0.1	+0.4	+0.2

Table C.2: Breakdown of the uncorrelated uncertainty δ_{uncor} from table C.1. Shown are the uncertainties associated with the reweighting of the MC models (δ_1), the electron-energy scale (δ_2), the electron-finding algorithm (δ_3), the electron calibration (δ_4), the variation of the $p_{\perp,\mathrm{lab}}$ cut of the jets (δ_5), the variation of the p_{track} cut (δ_6), the variation of the $p_\perp/\sqrt{E_\perp}$ cut (δ_7), the variation of the z_{vertex} cut (δ_8), the variation of the r_{RCAL} cut (δ_9), the variation of the electron-track distance cut (δ_{10}), the polarisation correction (δ_{12}), the FLT track veto efficiency correction (δ_{13}) and the correction to QED Born level (δ_{14}). The uncertainty associated with the vertex position reweighting is not shown, as it is zero within the quoted number of digits. For the asymmetric uncertainties, the upper number corresponds to the upward variation of the corresponding parameter and the lower number corresponds to the downward variation. For details on individual uncertainties, see chapter 10.

Q^2 [GeV²]	$p_\perp$ [GeV]	150–200 7–11	150–200 11–18	150–200 18–30	150–200 30–50	200–270 7–11	200–270 11–18	200–270 18–30	200–270 30–50	270–400 7–11	270–400 11–18	270–400 18–30	270–400 30–50	400–700 7–11	400–700 11–18	400–700 18–30	400–700 30–50	700–5000 7–11	700–5000 11–18	700–5000 18–30	700–5000 30–50	5000–15000 7–11	5000–15000 11–18	5000–15000 18–30	5000–15000 30–50
150–200	7–11	100	18	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150–200	11–18	18	100	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150–200	18–30	5	12	100	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150–200	30–50	1	2	9	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200–270	7–11	0	0	0	0	100	19	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200–270	11–18	0	0	0	0	19	100	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200–270	18–30	0	0	0	0	5	12	100	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
200–270	30–50	0	0	0	0	1	1	7	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
270–400	7–11	0	0	0	0	0	0	0	0	100	19	5	1	0	0	0	0	0	0	0	0	0	0	0	0
270–400	11–18	0	0	0	0	0	0	0	0	19	100	13	1	0	0	0	0	0	0	0	0	0	0	0	0
270–400	18–30	0	0	0	0	0	0	0	0	5	13	100	8	0	0	0	0	0	0	0	0	0	0	0	0
270–400	30–50	0	0	0	0	0	0	0	0	1	1	8	100	0	0	0	0	0	0	0	0	0	0	0	0
400–700	7–11	0	0	0	0	0	0	0	0	0	0	0	0	100	20	5	2	0	0	0	0	0	0	0	0
400–700	11–18	0	0	0	0	0	0	0	0	0	0	0	0	20	100	14	2	0	0	0	0	0	0	0	0
400–700	18–30	0	0	0	0	0	0	0	0	0	0	0	0	5	14	100	9	0	0	0	0	0	0	0	0
400–700	30–50	0	0	0	0	0	0	0	0	0	0	0	0	2	2	9	100	0	0	0	0	0	0	0	0
700–5000	7–11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	20	7	2	0	0	0	0
700–5000	11–18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	100	15	3	0	0	0	0
700–5000	18–30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	15	100	9	0	0	0	0
700–5000	30–50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	9	100	0	0	0	0
5000–15000	7–11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	19	10	3
5000–15000	11–18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	100	18	10
5000–15000	18–30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	18	100	13
5000–15000	30–50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	10	13	100
125–250	8–15	15	32	1	0	11	23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125–250	15–22	2	16	27	0	1	13	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125–250	22–30	1	1	27	8	1	1	21	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125–250	30–60	0	1	3	43	0	0	2	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
250–500	8–15	0	0	0	0	5	11	0	0	18	36	1	0	7	14	0	0	0	0	0	0	0	0	0	0
250–500	15–22	0	0	0	0	1	5	8	0	2	20	30	0	1	8	13	0	0	0	0	0	0	0	0	0
250–500	22–30	0	0	0	0	0	0	8	3	1	1	30	4	0	1	14	4	0	0	0	0	0	0	0	0
250–500	30–60	0	0	0	0	0	0	0	10	0	0	4	45	0	0	2	16	0	0	0	0	0	0	0	0
500–1000	8–15	0	0	0	0	0	0	0	0	0	0	0	0	14	26	1	0	11	20	1	0	0	0	0	0
500–1000	15–22	0	0	0	0	0	0	0	0	0	0	0	0	2	16	23	0	2	11	17	0	0	0	0	0
500–1000	22–30	0	0	0	0	0	0	0	0	0	0	0	0	1	1	24	6	1	1	17	4	0	0	0	0
500–1000	30–60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	37	1	0	3	23	0	0	0	0
1000–2000	8–15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	31	1	0	0	0	0	0
1000–2000	15–22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	18	26	0	0	0	0	0
1000–2000	22–30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	28	6	0	0	0	0
1000–2000	30–60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	5	41	0	0	0	0
2000–5000	8–16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	22	2	0	0	0	0	0
2000–5000	16–28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	9	28	3	0	0	0	0
2000–5000	28–60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	30	0	0	0	0
5000–20000	8–16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	40	8	0
5000–20000	16–28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	22	45	8
5000–20000	28–60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	18	53

The first 24 rows correspond to the **Inclusive jet bin**; the remaining rows correspond to the **Dijet bin**. The top-level column header is **Inclusive jet bin**.

Table C.3: Correlation matrix of the unfolding uncertainty within the inclusive jet cross section measurement and between the inclusive jet measurement and the previous dijet measurement.[71] Correlations are given in percent. The transverse momentum $p_\perp$ is $p_{\perp,\mathrm{Breit}}$ for the inclusive jets and $\overline{p_{\perp,\mathrm{Breit}}}$ for the dijets.

C.2 With matrix unfolding

Q^2 [GeV^2]	$p_{\perp,\text{Breit}}$ [GeV]	σ [pb]	δ_{unf} [%]	δ_{uncor} [%]	δ_{JES} [%]	δ_{model} [%]	δ_{Fake} [%]	$\delta_{\text{Low-}Q^2}$ [%]	δ_{E-p_z} [%]	δ_{TME} [%]	c_{QED}	c_Z	c_{Had}	δ_{Had} [%]
150–200	7–11	68.4	3.6	1.7	$^{+4.4}_{-4.1}$	2.0	1.7	0.3	$^{+0.9}_{-0.0}$	1.4	0.686	1.00	0.929	0.8
150–200	11–18	29.6	4.5	1.8	$^{+5.3}_{-5.1}$	1.0	2.1	1.1	$^{+0.8}_{-0.2}$	1.4	0.868	1.00	0.973	0.9
150–200	18–30	6.55	8.3	1.7	$^{+3.4}_{-3.2}$	1.4	1.2	1.7	$^{+1.8}_{-0.3}$	1.3	0.946	1.00	0.971	0.4
150–200	30–50	0.828	22.4	5.3	$^{+3.4}_{-4.2}$	4.2	1.1	4.2	$^{+5.3}_{+0.0}$	1.2	0.901	1.00	0.959	0.3
200–270	7–11	55.9	3.7	1.6	$^{+4.1}_{-3.9}$	2.3	1.3	0.2	$^{+0.7}_{-0.2}$	1.0	0.679	1.00	0.936	1.2
200–270	11–18	27.8	4.3	1.3	$^{+4.7}_{-4.5}$	0.9	1.5	0.1	$^{+1.1}_{-0.2}$	1.0	0.800	1.00	0.961	0.2
200–270	18–30	6.70	7.8	2.8	$^{+3.7}_{-3.4}$	-2.0	1.0	-0.1	$^{+2.0}_{-0.3}$	1.0	0.953	1.00	0.978	0.2
200–270	30–50	0.590	28.9	7.1	$^{+3.6}_{-4.1}$	5.5	1.4	-2.3	$^{+4.5}_{-1.2}$	0.9	0.951	1.00	0.985	0.8
270–400	7–11	51.2	3.7	1.7	$^{+3.9}_{-3.6}$	2.5	1.3	0.1	$^{+0.8}_{+0.1}$	0.7	0.704	1.00	0.930	1.6
270–400	11–18	26.6	4.3	2.3	$^{+4.2}_{-4.3}$	0.9	1.4	0.2	$^{+1.0}_{-0.0}$	0.8	0.690	1.00	0.967	0.1
270–400	18–30	6.56	7.2	2.2	$^{+3.6}_{-3.7}$	1.7	1.0	0.2	$^{+1.4}_{-0.5}$	0.7	0.923	1.00	0.973	0.2
270–400	30–50	0.690	24.0	5.0	$^{+3.8}_{-2.9}$	-5.2	1.1	1.0	$^{+5.0}_{-1.6}$	0.7	0.956	1.00	0.977	0.4
400–700	7–11	45.5	3.5	1.1	$^{+3.3}_{-3.0}$	2.4	1.2	0.2	$^{+0.9}_{-0.2}$	0.6	0.755	1.01	0.916	1.4
400–700	11–18	26.7	3.4	1.2	$^{+3.5}_{-3.6}$	1.8	1.2	-0.1	$^{+1.7}_{-0.0}$	0.6	0.626	1.01	0.974	0.2
400–700	18–30	8.23	5.6	1.0	$^{+3.5}_{-3.3}$	2.2	0.9	-0.0	$^{+1.7}_{-0.2}$	0.6	0.854	1.01	0.972	0.3
400–700	30–50	1.41	12.2	3.2	$^{+4.5}_{-3.7}$	3.5	0.8	-0.2	$^{+6.3}_{-0.1}$	0.6	0.982	1.01	0.973	0.5
700–5000	7–11	42.9	3.3	1.3	$^{+2.5}_{-2.2}$	2.4	1.1	0.1	$^{+0.4}_{-0.2}$	0.5	0.826	1.03	0.903	0.6
700–5000	11–18	27.8	2.8	1.8	$^{+2.5}_{-2.5}$	2.2	0.9	-0.0	$^{+0.9}_{-0.2}$	0.5	0.702	1.03	0.967	0.7
700–5000	18–30	12.1	4.0	0.9	$^{+2.7}_{-2.7}$	3.5	0.9	-0.0	$^{+1.2}_{-0.0}$	0.5	0.571	1.03	0.974	0.0
700–5000	30–50	2.23	9.4	1.6	$^{+4.6}_{-4.4}$	4.0	1.4	0.1	$^{-0.7}_{-0.4}$	0.5	0.714	1.03	0.963	0.4
5000–15000	7–11	2.48	22.6	2.9	$^{+1.6}_{-1.6}$	-3.4	2.0	0.1	$^{-1.0}_{-0.5}$	0.3	0.930	1.16	0.889	0.6
5000–15000	11–18	1.99	13.6	3.5	$^{+1.3}_{-1.2}$	4.3	1.4	-0.1	$^{-0.6}_{+0.0}$	0.3	0.867	1.16	0.948	0.3
5000–15000	18–30	0.965	14.8	3.1	$^{+1.5}_{-1.5}$	6.4	0.8	-0.0	$^{-2.4}_{-1.6}$	0.3	0.722	1.16	0.976	0.1
5000–15000	30–50	0.204	32.5	4.9	$^{+3.5}_{-3.4}$	6.9	1.7	-0.1	$^{-5.0}_{-4.1}$	0.3	0.433	1.16	0.977	0.9

Table C.4: Double-differential inclusive jet cross sections derived with matrix unfolding, see section 11.1. Listed are the unfolding uncertainty δ_{unf}, the sum of the uncorrelated systematic uncertainties δ_{uncor} and the correlated systematic uncertainties associated with the jet energy scale δ_{JES}, MC model δ_{model}, the background contribution from unmatched jets δ_{Fake}, the background contribution from low-Q^2 DIS events $\delta_{\text{Low-}Q^2}$, the $E - p_z$ cut δ_{E-p_z} and the track-matching efficiency correction δ_{TME}. Not listed explicitly is the luminosity uncertainty of 1.9%, which is fully correlated across all points. The last four columns show the QED Born level correction c_{QED} that has been applied to the data as well as the Z^0 and hadronisation correction and associated uncertainty, c_Z, c_{Had} and δ_{Had}, that need to be applied to the theoretical predictions.

Q^2 [GeV²]	$p_{\perp,\text{Breit}}$ [GeV]	δ_1 [%]	δ_2 [%]	δ_3 [%]	δ_4 [%]	δ_5 [%]	δ_6 [%]	δ_7 [%]	δ_8 [%]	δ_9 [%]	δ_{10} [%]	δ_{11} [%]	δ_{12} [%]	δ_{13} [%]	δ_{14} [%]
150–200	7–11	-0.9	$^{-0.0}_{+0.1}$	+0.6	-1.1	$^{+0.4}_{-0.6}$	$^{+0.0}_{+0.3}$	$^{+0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	$^{+0.0}_{-0.0}$	-0.3	+0.0	-0.0	+0.0	+0.2
150–200	11–18	-1.2	$^{-0.1}_{+0.1}$	+0.2	-1.1	$^{+0.5}_{-0.0}$	$^{+0.0}_{+0.2}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.0}$	$^{+0.0}_{-0.0}$	-0.3	+0.1	-0.0	-0.0	+0.3
150–200	18–30	+0.5	$^{-0.0}_{+0.0}$	-0.4	-1.1	$^{+0.2}_{+0.1}$	$^{+0.0}_{-0.2}$	$^{+0.1}_{+0.2}$	$^{-0.2}_{-0.0}$	$^{-0.1}_{+0.0}$	-0.5	-0.1	-0.0	+0.0	+0.9
150–200	30–50	+1.4	$^{+0.0}_{+0.3}$	-3.0	+2.8	$^{-0.1}_{+0.2}$	$^{+0.0}_{+1.1}$	$^{+0.4}_{+0.1}$	$^{-0.1}_{-0.8}$	$^{-0.1}_{+0.0}$	-1.6	+0.0	-0.0	-0.1	+2.5
200–270	7–11	-0.7	$^{-0.0}_{+0.1}$	+0.8	-0.5	$^{+1.0}_{-1.1}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{+0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	-0.4	-0.1	-0.0	+0.0	+0.1
200–270	11–18	-1.2	$^{-0.0}_{+0.0}$	-0.2	-0.3	$^{+0.1}_{-0.2}$	$^{+0.0}_{-0.0}$	$^{-0.0}_{+0.0}$	$^{-0.1}_{+0.0}$	$^{-0.1}_{-0.1}$	-0.3	+0.1	-0.0	-0.0	+0.3
200–270	18–30	+0.2	$^{-0.0}_{-0.0}$	+2.6	+0.2	$^{+0.0}_{-0.2}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.1}$	$^{-0.1}_{-0.1}$	$^{-0.2}_{-0.8}$	-0.5	+0.1	-0.0	-0.0	+0.7
200–270	30–50	+1.9	$^{-0.2}_{+0.4}$	+6.0	+1.7	$^{+0.0}_{+0.2}$	$^{+0.0}_{+1.1}$	$^{+0.0}_{+0.7}$	$^{-0.1}_{+0.1}$	$^{+0.1}_{+1.2}$	-0.3	-0.2	-0.1	+0.0	+2.5
270–400	7–11	-0.7	$^{-0.1}_{+0.1}$	+1.2	-0.6	$^{+0.6}_{-0.7}$	$^{+0.0}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.0}$	$^{+0.1}_{+0.1}$	-0.3	+0.2	-0.1	-0.1	+0.1
270–400	11–18	-1.5	$^{-0.0}_{+0.0}$	+1.5	-0.7	$^{+0.4}_{-0.6}$	$^{+0.0}_{+0.1}$	$^{-0.0}_{+0.0}$	$^{+0.1}_{+0.1}$	$^{+0.0}_{-0.1}$	-0.3	-0.0	-0.1	-0.2	+0.2
270–400	18–30	+0.4	$^{+0.1}_{+0.1}$	+0.6	-1.7	$^{-0.1}_{+0.1}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{+0.2}$	$^{+0.1}_{-0.0}$	$^{-0.5}_{-0.2}$	-0.8	-0.1	-0.1	-0.3	+0.6
270–400	30–50	+0.7	$^{-0.1}_{+0.3}$	+3.6	-2.1	$^{-0.1}_{+0.0}$	$^{+0.0}_{-0.0}$	$^{+0.2}_{+0.3}$	$^{-0.1}_{-0.7}$	$^{-1.0}_{-1.9}$	+0.4	-0.0	-0.1	-0.9	+1.8
400–700	7–11	-0.7	$^{-0.0}_{+0.1}$	+0.2	-0.2	$^{+0.5}_{-0.7}$	$^{+0.0}_{-0.0}$	$^{-0.0}_{+0.0}$	$^{+0.1}_{-0.0}$	$^{-0.2}_{-0.1}$	-0.1	+0.2	-0.1	-0.3	+0.2
400–700	11–18	-1.1	$^{+0.0}_{+0.1}$	+0.1	-0.3	$^{+0.2}_{-0.2}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{-0.0}$	$^{-0.0}_{+0.1}$	$^{-0.2}_{-0.1}$	-0.1	+0.1	-0.1	-0.3	+0.2
400–700	18–30	+0.5	$^{+0.1}_{+0.1}$	-0.4	+0.3	$^{-0.0}_{-0.1}$	$^{+0.0}_{+0.1}$	$^{-0.1}_{+0.3}$	$^{+0.1}_{-0.1}$	$^{-0.5}_{+0.2}$	-0.0	+0.3	-0.1	-0.2	+0.5
400–700	30–50	+1.5	$^{+0.2}_{-0.1}$	+1.4	+1.4	$^{-0.2}_{-0.7}$	$^{+0.0}_{-0.3}$	$^{-0.2}_{+0.4}$	$^{-0.1}_{+0.2}$	$^{-0.4}_{-1.6}$	+0.2	-0.0	-0.0	-0.9	+1.5
700–5000	7–11	-0.2	$^{-0.1}_{+0.1}$	+1.0	-0.1	$^{+1.0}_{-0.5}$	$^{+0.0}_{+0.1}$	$^{-0.0}_{+0.0}$	$^{+0.0}_{-0.0}$	$^{+0.0}_{+0.0}$	-0.1	-0.1	-0.1	+0.3	+0.1
700–5000	11–18	-1.7	$^{-0.1}_{+0.1}$	+0.5	+0.2	$^{+0.4}_{-0.2}$	$^{+0.0}_{+0.2}$	$^{-0.0}_{+0.1}$	$^{-0.0}_{-0.0}$	$^{+0.0}_{+0.0}$	-0.1	+0.2	-0.1	+0.1	+0.1
700–5000	18–30	-0.8	$^{-0.1}_{+0.1}$	+0.1	+0.1	$^{+0.2}_{-0.5}$	$^{+0.0}_{+0.0}$	$^{-0.1}_{-0.0}$	$^{-0.0}_{-0.3}$	$^{+0.0}_{+0.0}$	-0.0	+0.2	-0.1	+0.0	+0.2
700–5000	30–50	-0.2	$^{-0.3}_{+0.5}$	+0.1	+0.7	$^{+2.1}_{+0.2}$	$^{+0.0}_{-0.3}$	$^{+0.0}_{+0.2}$	$^{-0.1}_{-0.6}$	$^{+0.0}_{+0.0}$	+0.0	-0.0	-0.1	-0.1	+0.5
5000–15000	7–11	-0.0	$^{-0.2}_{+0.2}$	+1.5	-1.6	$^{+0.1}_{-2.2}$	$^{+0.0}_{+0.1}$	$^{-0.0}_{-0.3}$	$^{-0.2}_{-0.2}$	$^{+0.0}_{+0.0}$	-0.1	+0.3	-0.1	+1.5	+0.2
5000–15000	11–18	-1.2	$^{-0.2}_{+0.2}$	-2.7	+0.8	$^{+0.2}_{-1.1}$	$^{+0.0}_{-0.0}$	$^{+0.2}_{-1.0}$	$^{-0.2}_{-0.5}$	$^{+0.0}_{+0.0}$	-0.8	-0.2	-0.1	+1.1	+0.2
5000–15000	18–30	-1.8	$^{-0.2}_{+0.1}$	+2.0	-0.1	$^{+1.1}_{+1.1}$	$^{+0.0}_{-0.0}$	$^{-0.1}_{-0.4}$	$^{-0.2}_{-0.7}$	$^{+0.0}_{+0.0}$	-0.3	-0.1	-0.1	+0.7	+0.2
5000–15000	30–50	-1.5	$^{-0.2}_{+0.2}$	+4.5	+0.1	$^{+1.0}_{-0.3}$	$^{+0.0}_{-0.1}$	$^{-0.1}_{+0.1}$	$^{-0.2}_{+0.2}$	$^{+0.0}_{+0.0}$	+0.4	+0.6	+0.2	+0.1	+0.2

Table C.5: Breakdown of the uncorrelated uncertainty δ_{uncor} from table C.4. Shown are the uncertainties associated with the reweighting of the MC models (δ_1), the electron-energy scale (δ_2), the electron-finding algorithm (δ_3), the electron calibration (δ_4), the variation of the $p_{\perp,\text{lab}}$ cut of the jets (δ_5), the variation of the p_{track} cut (δ_6), the variation of the $p_\perp/\sqrt{E_\perp}$ cut (δ_7), the variation of the z_{vertex} cut (δ_8), the variation of the r_{RCAL} cut (δ_9), the variation of the electron-track distance cut (δ_{10}), the relative normalisation of the background from photoproduction events (δ_{11}), the polarisation correction (δ_{12}), the FLT track veto efficiency correction (δ_{13}) and the correction to QED Born level (δ_{14}). The uncertainty associated with the vertex position reweighting is not shown, as it is zero within the quoted number of digits. For the asymmetric uncertainties, the upper number corresponds to the upward variation of the corresponding parameter and the lower number corresponds to the downward variation. For details on individual uncertainties, see chapter 10.

	Q^2 [GeV²]	$p_\perp$ [GeV]	150–200 7–11	150–200 11–18	150–200 18–30	150–200 30–50	200–270 7–11	200–270 11–18	200–270 18–30	200–270 30–50	270–400 7–11	270–400 11–18	270–400 18–30	270–400 30–50	400–700 7–11	400–700 11–18	400–700 18–30	400–700 30–50	700–5000 7–11	700–5000 11–18	700–5000 18–30	700–5000 30–50	5000–15000 7–11	5000–15000 11–18	5000–15000 18–30	5000–15000 30–50
Inclusive jet bin	150–200	7–11	100	−38	−7	0	−13	6	2	0	13	−5	−1	−0	9	−3	−1	−0	9	−4	−2	−1	2	−1	−1	−1
	150–200	11–18	−38	100	4	4	5	−26	−0	−1	−4	7	1	1	−3	1	0	−0	−3	2	1	0	−1	0	0	0
	150–200	18–30	−7	4	100	−1	0	3	−36	1	−0	−0	9	−0	−0	0	−1	0	−0	0	0	0	−0	0	0	0
	150–200	30–50	0	4	−1	100	0	−0	2	−47	0	0	−1	14	0	−0	0	−3	0	0	0	0	0	−0	−0	−0
	200–270	7–11	−13	5	0	0	100	−41	−8	−1	−10	4	2	0	9	−3	−1	−0	7	−3	−1	−1	2	−1	−1	−0
	200–270	11–18	6	−26	3	−0	−41	100	−4	−1	3	−17	1	0	−3	4	0	−0	−3	1	0	0	−1	0	0	0
	200–270	18–30	2	−0	−36	2	−8	−4	100	−5	−1	2	−26	2	−1	−0	5	−0	−1	0	−0	0	−0	0	0	0
	200–270	30–50	0	−1	1	−47	−1	−1	−5	100	−0	0	3	−39	−0	0	−1	10	−0	0	0	−1	−0	0	0	0
	270–400	7–11	13	−4	−0	0	−10	3	−1	−0	100	−38	−8	−1	−3	1	1	0	7	−3	−1	−1	2	−1	−0	−0
	270–400	11–18	−5	7	−0	0	4	−17	2	0	−38	100	−4	−1	0	−9	1	0	−2	2	0	0	−1	0	0	0
	270–400	18–30	−1	1	9	−1	2	1	−26	3	−8	−4	100	−3	−0	2	−16	1	−0	0	1	−0	−0	0	0	0
	270–400	30–50	−0	1	−0	14	0	0	2	−39	−1	−1	−3	100	−0	0	2	−25	−0	0	−0	3	−0	0	−0	−0
	400–700	7–11	9	−3	−0	0	9	−3	−1	−0	−3	0	−0	−0	100	−39	−9	−1	1	−0	−0	−0	2	−1	−0	−0
	400–700	11–18	−3	1	0	−0	−3	4	−0	0	1	−9	2	0	−39	100	−5	−1	−1	−4	1	0	−1	0	0	0
	400–700	18–30	−1	0	−1	0	−1	0	5	−1	1	1	−16	2	−9	−5	100	−5	−0	1	−7	1	−0	0	0	0
	400–700	30–50	−0	−0	0	−3	−0	−0	−0	10	0	0	1	−25	−1	−1	−5	100	−0	0	1	−10	−0	0	0	0
	700–5000	7–11	9	−3	−0	0	7	−3	−1	−0	7	−2	−0	−0	1	−1	−0	−0	100	−50	−10	−2	1	−0	−0	−0
	700–5000	11–18	−4	2	0	0	−3	1	0	0	−3	2	0	0	−0	−4	1	0	−50	100	−10	−2	−0	−1	0	0
	700–5000	18–30	−2	1	0	0	−1	0	−0	0	−1	0	1	−0	−0	1	−7	1	−10	−10	100	−7	−0	0	−1	−0
	700–5000	30–50	−1	0	0	0	−1	0	0	−1	−1	0	−0	3	−0	0	1	−10	−2	−2	−7	100	−0	−0	−0	−1
	5000–15000	7–11	2	−1	−0	0	2	−1	−0	−0	2	−1	−0	−0	2	−1	−0	−0	1	−0	−0	−0	100	−67	−15	−3
	5000–15000	11–18	−1	0	0	−0	−1	0	0	0	−1	0	0	0	−1	0	0	0	−0	−1	0	−0	−67	100	−12	1
	5000–15000	18–30	−1	0	0	−0	−1	0	0	0	−0	0	0	−0	−0	0	0	0	−0	0	−1	−0	−15	−12	100	−17
	5000–15000	30–50	−1	0	0	−0	−0	0	0	0	−0	0	0	−0	−0	0	0	0	−0	0	−0	−1	−3	1	−17	100
Dijet bin	125–250	8–15	2	26	5	3	−1	11	−0	−1	−7	2	1	1	−6	3	1	−0	−6	3	1	1	−2	1	0	0
	125–250	15–22	2	13	18	2	1	4	7	−1	−0	−1	−1	0	−0	0	0	−0	−0	0	0	0	−0	0	0	0
	125–250	22–30	−0	4	20	5	−1	−0	10	−1	−0	0	−2	0	−0	0	0	−0	−0	0	0	0	−0	0	0	0
	125–250	30–60	1	4	2	36	−0	−0	−1	10	−0	0	0	−2	−0	0	0	0	−0	0	0	0	−0	0	0	0
	250–500	8–15	−9	2	1	0	−5	6	1	0	4	19	0	0	−4	7	0	0	−7	3	1	1	−2	1	0	0
	250–500	15–22	−1	0	−1	0	−1	2	2	0	1	9	15	0	−1	2	5	0	−1	0	−0	0	−0	0	0	0
	250–500	22–30	−1	1	−1	0	−1	0	3	−0	−1	0	21	0	−1	0	6	1	−1	0	−1	0	−0	0	0	0
	250–500	30–60	−0	1	0	−1	−0	0	−0	2	−1	0	1	33	−0	0	0	10	−0	0	0	−1	−0	0	0	0
	500–1000	8–15	−8	3	0	−0	−6	3	1	0	−7	−1	1	0	4	19	−0	0	1	13	1	1	−2	1	0	0
	500–1000	15–22	−2	1	−0	0	−1	1	1	−0	−2	−1	−4	0	1	10	18	1	−1	6	9	−0	−0	0	−0	−0
	500–1000	22–30	−1	1	−0	−0	−1	0	1	0	−1	0	−4	−1	−1	−0	22	2	−1	−0	14	−0	−0	0	−0	0
	500–1000	30–60	−1	0	0	−1	−1	0	−0	4	−1	0	0	−10	−1	0	−1	37	−1	0	−1	22	−0	0	−0	−0
	1000–2000	8–15	−6	2	0	−0	−5	2	0	0	−4	2	0	0	−5	−0	1	0	6	24	−1	0	−1	0	0	0
	1000–2000	15–22	−2	1	0	0	−1	0	−0	0	−1	1	1	−0	−2	−0	−2	0	1	13	17	0	−0	−0	−0	−0
	1000–2000	22–30	−1	0	0	0	−1	0	−0	0	−1	0	0	0	−1	0	−2	−1	−1	−0	25	1	−0	0	−1	−0
	1000–2000	30–60	−1	1	0	0	−1	0	0	−1	−1	0	0	1	−1	0	0	−5	−0	1	−1	42	−0	0	−0	−1
	2000–5000	8–16	−4	1	0	−0	−3	1	0	0	−3	1	0	−0	−4	−0	0	0	3	20	−1	−0	−1	1	0	0
	2000–5000	16–28	−2	1	0	0	−1	0	−0	−0	−1	0	1	0	−1	0	−3	−0	0	4	28	0	−0	0	0	−0
	2000–5000	28–60	−1	0	0	0	−1	0	0	−0	−1	0	0	1	−1	0	−0	−3	−1	1	3	25	−0	0	−0	0
	5000–20000	8–16	−2	1	0	−0	−2	1	0	0	−2	1	0	0	−2	1	0	0	−2	1	0	0	8	24	−3	0
	5000–20000	16–28	−1	0	0	−0	−1	0	0	0	−1	0	0	0	−1	0	0	0	−1	0	−0	0	3	9	31	−1
	5000–20000	28–60	−1	0	0	−0	−1	0	0	0	−1	0	0	−0	−1	0	0	0	−1	0	0	−0	0	1	7	39

Table C.6: Correlation matrix of the unfolding uncertainty within the inclusive jet cross section measurement and between the inclusive jet measurement and the previous dijet measurement.[71] Correlations are given in per cent. The transverse momentum $p_\perp$ is $p_{\perp,\mathrm{Breit}}$ for the inclusive jets and $\overline{p_{\perp,\mathrm{Breit}}}$ for the dijets.

Q^2 [GeV^2]	$p_{\perp,\text{Breit}}$ [GeV]	σ [pb]	δ_{uncor} [%]	c'_{QED}	δ'_{QED} [%]
150–200	7–11	85.4	1.7	0.856	0.2
150–200	11–18	32.6	1.8	0.954	0.4
150–200	18–30	6.82	1.8	0.986	1.0
150–200	30–50	0.913	5.6	0.994	3.0
200–270	7–11	70.0	1.6	0.851	0.2
200–270	11–18	32.1	1.3	0.924	0.3
200–270	18–30	6.91	2.8	0.983	0.8
200–270	30–50	0.617	7.3	0.994	3.0
270–400	7–11	63.4	1.7	0.871	0.2
270–400	11–18	33.5	2.3	0.870	0.3
270–400	18–30	6.93	2.2	0.976	0.6
270–400	30–50	0.712	5.0	0.986	1.8
400–700	7–11	55.4	1.1	0.919	0.2
400–700	11–18	35.2	1.2	0.825	0.2
400–700	18–30	9.13	1.1	0.947	0.5
400–700	30–50	1.42	3.3	0.988	1.5
700–5000	7–11	50.3	1.3	0.970	0.1
700–5000	11–18	35.4	1.8	0.896	0.1
700–5000	18–30	17.1	0.9	0.809	0.2
700–5000	30–50	2.67	1.5	0.855	0.5
5000–15000	7–11	2.65	2.9	0.994	0.2
5000–15000	11–18	2.27	3.5	0.989	0.2
5000–15000	18–30	1.270	3.1	0.950	0.2
5000–15000	30–50	0.330	4.9	0.702	0.2

Table C.7: Cross sections using the alternative QED correction factors. These numbers correspond to a cross section definition including QED radiation on the electron side and including a cut $E_{\text{ISR}} < 8.5\,\text{GeV}$ on the energy of the ISR photon. Also given are the corresponding QED correction factors that have been applied and their uncertainties. All numbers and definitions that are not listed are identical to those from table C.4. The column δ_{uncor} represents the sum of uncorrelated systematic uncertainties from table C.5, but using the uncertainties of the alternative QED correction, rather than the nominal one. See section 9.5.3 for more details.

Bibliography

[1] X. Fan et al. *Measurement of the Electron Magnetic Moment.* In: *Physical Review Letters* 130.7 (2023). DOI: 10.1103/physrevlett.130.071801. arXiv: 2209.13084 (cit. on p. 1).

[2] N. Bezginov et al. *A measurement of the atomic hydrogen Lamb shift and the proton charge radius.* In: *Science* 365.6457 (2019), pp. 1007–1012. DOI: 10.1126/science.aau7807 (cit. on p. 1).

[3] R. Bouchendira et al. *New Determination of the Fine Structure Constant and Test of the Quantum Electrodynamics.* In: *Physical Review Letters* 106 (2011), p. 080801. DOI: 10.1103/PhysRevLett.106.080801 (cit. on p. 1).

[4] D. Goldberg. *The Standard Model in a Nutshell.* Princeton University Press (2017). ISBN: 978-0-691-16759-6 (cit. on pp. 1, 5, 7).

[5] S.L. Glashow. *The renormalizability of vector meson interactions.* In: *Nuclear Physics* 10 (1959), pp. 107–117. DOI: 10.1016/0029-5582(59)90196-8 (cit. on pp. 1, 7).

[6] A. Salam and J.C. Ward. *Electromagnetic and weak interactions.* In: *Physics Letters* 13.2 (1964), pp. 168–171. DOI: 10.1016/0031-9163(64)90711-5 (cit. on pp. 1, 7).

[7] S. Weinberg. *A Model of Leptons.* In: *Physical Review Letters* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (cit. on pp. 1, 7).

[8] P.W. Anderson. *Plasmons, Gauge Invariance, and Mass.* In: *Physical Review* 130 (1963), pp. 439–442. DOI: 10.1103/PhysRev.130.439 (cit. on pp. 1, 7).

[9] F. Englert and R. Brout. *Broken Symmetry and the Mass of Gauge Vector Mesons.* In: *Physical Review Letters* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on pp. 1, 7).

[10] P.W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons.* In: *Physical Review Letters* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on pp. 1, 7).

[11] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. *Global Conservation Laws and Massless Particles.* In: *Physical Review Letters* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on pp. 1, 7).

[12] K. Nishijima. *BRS invariance, asymptotic freedom and color confinement (a review).* In: *Czechoslovak Journal of Physics* 46.1 (1996), pp. 1–40. DOI: 10.1007/BF01692238 (cit. on pp. 1, 16, 20).

[13] UA1 Collaboration. *Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider.* In: *Physics Letters B* 126.5 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0 (cit. on p. 1).

[14] UA2 Collaboration. *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider.* In: *Physics Letters B* 122.5 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2 (cit. on p. 1).

[15] LHCb Collaboration. *List of hadrons observed at the LHC.* LHCB-FIGURE-2021-001. 2021. URL: http://cds.cern.ch/record/2749030 (cit. on p. 1).

[16] T. Aoyama et al. *Tenth-order electron anomalous magnetic moment: Contribution of diagrams without closed lepton loops.* In: *Physical Review D* 91 (2015), p. 033006. DOI: 10.1103/PhysRevD.91.033006 (cit. on pp. 1, 45).

[17] Particle Data Group. *Review of Particle Physics.* In: *Progress of Theoretical and Experimental Physics* 2022.8 (2022). DOI: 10.1093/ptep/ptac097 (cit. on pp. 1, 5, 6, 12, 21, 42, 43, 152, 153, 158).

[18] N. Jones. *General Relativity and the Standard Model: Why evidence for one does not disconfirm the other.* In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 40.2 (2009), pp. 124–132. DOI: 10.1016/j.shpsb.2008.10.004 (cit. on p. 1).

[19] J.G. de Swart, G. Bertone, and J. van Dongen. *How dark matter came to matter.* In: *Nature Astronomy* 1.3 (2017), p. 0059. DOI: 10.1038/s41550-017-0059 (cit. on p. 1).

[20] KATRIN Collaboration. *Direct neutrino-mass measurement with sub-electronvolt sensitivity.* In: *Nature Physics* 18.2 (2022), pp. 160–166. DOI: 10.1038/s41567-021-01463-1 (cit. on p. 1).

[21] L. Canetti, M. Drewes, and M. Shaposhnikov. *Matter and antimatter in the universe.* In: *New Journal of Physics* 14.9 (2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012 (cit. on p. 1).

[22] R.D. Peccei and H.R. Quinn. *CP Conservation in the Presence of Pseudoparticles.* In: *Physical Review Letters* 38 (1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440 (cit. on p. 1).

[23] A. Crivellin. *Anomalies in Particle Physics.* 2023. arXiv: 2304.01694 (cit. on p. 1).

[24] O. Brüning, A. Seryi, and S. Verdú-Andrés. *Electron-Hadron Colliders: EIC, LHeC and FCC-eh.* In: *Front. Phys.* 10 (2022), p. 886473. DOI: 10.3389/fphy.2022.886473 (cit. on p. 1).

[25] F. Grotelüschen et al. *50 Years of DESY.* 2019. URL: https://pr.desy.de/sites/sites_desygroups/sites_extern/site_pr/content/e104098/e104109/DESY50_anniversabrochure_E_eng.pdf. (Cit. on pp. 1, 2, 29).

[26] ZEUS Collaboration. *Inclusive-jet production in NC DIS with HERA II.* ZEUS-prel-10-002. 2010. URL: https://inspirehep.net/literature/1217265 (cit. on p. 2).

[27] D. Lontkovskyi. *Measurement of Jet Production with the ZEUS Detector.* PhD thesis. University Hamburg (2016). URL: https://inspirehep.net/literature/1785225 (cit. on p. 2).

[28] B. Pullman. *The Atom in the History of Human Thought.* Oxford University Press (1998). ISBN: 0195114477 (cit. on p. 5).

[29] D. Bernoulli. *Hydrodynamica.* Sumptibus J.R. Dulseckeri (1739). DOI: 10.3931/e-rara-3911 (cit. on p. 5).

[30] E. Rutherford. *The scattering of α and β particles by matter and the structure of the atom.* In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 21.125 (1911), pp. 669–688. DOI: 10.1080/14786435.2011.617037 (cit. on pp. 5, 7).

[31] O. Masson. *XXIV. The constitution of atoms.* 1921. DOI: 10.1080/14786442108636219 (cit. on p. 5).

[32] M. Fierz. *Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin*. In: *Helvetica Physica Acta* 12 (1939), pp. 3–37. DOI: `10.5169/seals-110930` (cit. on p. 5).

[33] W. Pauli. *The Connection Between Spin and Statistics*. In: *Physical Review* 58 (1940), pp. 716–722. DOI: `10.1103/PhysRev.58.716` (cit. on p. 5).

[34] W. Pauli. *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*. In: *Zeitschrift für Physik* 31.1 (1925), pp. 765–783. DOI: `10.1007/BF02980631` (cit. on p. 5).

[35] T.D. Lee and C.N. Yang. *Question of Parity Conservation in Weak Interactions*. In: *Physical Review* 104 (1956), pp. 254–258. DOI: `10.1103/PhysRev.104.254` (cit. on p. 6).

[36] R. Devenish and A. Cooper-Sarkar. *Deep Inelastic Scattering*. Oxford University Press (2003). ISBN: 9780198506713. DOI: `10.1093/acprof:oso/9780198506713.001.0001` (cit. on pp. 8–10, 12, 15, 16, 22, 23, 25).

[37] H1 Collaboration. *Comparison of Deep Inelastic Scattering with Photoproduction Interactions at HERA*. In: *Physics Letters B* 358.3-4 (1995), pp. 412–422. DOI: `10.1016/0370-2693(95)01054-t`. arXiv: `hep-ex/9508013v2` (cit. on p. 8).

[38] R.P. Feynman. *QED: the strange theory of light and matter*. Princeton University Press (2006). ISBN: 9780691125756 (cit. on pp. 9, 14, 15).

[39] W. Greiner, S. Schramm, and E. Stein. *Quantum chromodynamics*. Springer (2007). ISBN: 9783540485346 (cit. on pp. 9, 13, 19, 48, 163).

[40] R.W. McAllister and R.Hofstadter. *Elastic Scattering of 188-MeV Electrons from the Proton and the Alpha Particle*. In: *Physical Review* 102 (1956), pp. 851–856. DOI: `10.1103/PhysRev.102.851` (cit. on p. 9).

[41] E.D. Bloom et al. *High-Energy Inelastic $e - p$ Scattering at 6°and 10°*. In: *Physical Review Letters* 23 (1969), pp. 930–934. DOI: `10.1103/PhysRevLett.23.930` (cit. on p. 10).

[42] G. Zweig. *An $SU(3)$ model for strong interaction symmetry and its breaking. Version 2*. In: *Developments in the Quark Theory of Hadrons. Volume 1* (1964), pp. 22–101. DOI: `10.17181/CERN-TH-412` (cit. on p. 11).

[43] R.P. Feynman. *The Behaviour of Hadron Collisions at Extreme Energies*. In: *3rd International Conference on High Energy Collisions* 690905 (1969), pp. 237–258. URL: `https://www.osti.gov/biblio/4153889` (cit. on p. 11).

[44] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. *Advantages of the color octet gluon picture*. In: *Physics Letters B* 47.4 (1973), pp. 365–368. DOI: `10.1016/0370-2693(73)90625-4` (cit. on p. 11).

[45] TASSO Collaboration. *Evidence for planar events in e^+e^- annihilation at high energies*. In: *Physics Letters B* 86.2 (1979), pp. 243–249. DOI: `10.1016/0370-2693(79)90830-X` (cit. on p. 13).

[46] MARK-J Collaboration. *Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA*. In: *Physical Review Letters* 43 (1979), pp. 830–833. DOI: `10.1103/PhysRevLett.43.830` (cit. on p. 13).

[47] PLUTO Collaboration. *Evidence for gluon bremsstrahlung in e^+e^- annihilations at high energies*. In: *Physics Letters B* 86.3 (1979), pp. 418–425. DOI: `10.1016/0370-2693(79)90869-4` (cit. on p. 13).

[48] JADE Collaboration. *Observation of planar three-jet events in e^+e^- annihilation and evidence for gluon bremsstrahlung.* In: *Physics Letters B* 91.1 (1980), pp. 142–147. DOI: 10.1016/0370-2693(80)90680-2 (cit. on p. 13).

[49] R.D. Field. *Applications of Perturbative QCD.* Addison-Wesley Publishing Company (1989). ISBN: 9780201483628 (cit. on pp. 13, 19, 163).

[50] E. Schrödinger. *Quantisierung als Eigenwertproblem.* In: *Annalen der Physik* 384.4 (1926), pp. 361–376. DOI: 10.1002/andp.19263840404 (cit. on p. 14).

[51] E. Schrödinger. *An Undulatory Theory of the Mechanics of Atoms and Molecules.* In: *Physical Review* 28 (1926), pp. 1049–1070. DOI: 10.1103/PhysRev.28.1049 (cit. on p. 14).

[52] O. Klein. *Quantentheorie und fünfdimensionale Relativitätstheorie.* In: *Zeitschrift für Physik* 37.12 (1926), pp. 895–906. DOI: 10.1007/BF01397481 (cit. on p. 14).

[53] W. Gordon. *Der Comptoneffekt nach der Schrödingerschen Theorie.* In: *Zeitschrift für Physik* 40.1 (1926), pp. 117–133. DOI: 10.1007/BF01390840 (cit. on p. 14).

[54] P.A.M. Dirac and R.H. Fowler. *The quantum theory of the electron.* In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023 (cit. on p. 14).

[55] Nobel Prize Outreach AB 2023. *The Nobel Prize in Physics 2004.* 2004. URL: https://www.nobelprize.org/prizes/physics/2004/summary/. (Cit. on p. 16).

[56] M.E. Peskin and D.V. Schroeder. *An Introduction To Quantum Field Theory.* Avalon Publishing (1995). ISBN: 9780813345437 (cit. on pp. 17, 18, 163).

[57] M.L. Mangano. *Introduction to QCD.* In: (2000). DOI: 10.5170/CERN-1999-004.53 (cit. on p. 18).

[58] A. Deur, S. J. Brodsky, and G. F. de Téramond. *The QCD running coupling.* In: *Progress in Particle and Nuclear Physics* 90 (2016), pp. 1–74. DOI: 10.1016/j.ppnp.2016.04.003. arXiv: 1604.08082 (cit. on p. 21).

[59] R.S. Thorne and R.G. Roberts. *Ordered analysis of heavy flavor production in deep-inelastic scattering.* In: *Physical Review D* 57.11 (1998), pp. 6871–6898. DOI: 10.1103/physrevd.57.6871. arXiv: hep-ph/9709442 (cit. on p. 21).

[60] K.G. Chetyrkin, J.H. Kühn, and M. Steinhauser. *RunDec: a Mathematica package for running and decoupling of the strong coupling and quark masses.* In: *Computer Physics Communications* 133.1 (2000), pp. 43–65. DOI: 10.1016/s0010-4655(00)00155-7. arXiv: hep-ph/0004189 (cit. on p. 21).

[61] J.C. Collins, D.E. Soper, and G. Sterman. *Factorization of Hard Processes in QCD.* In: *Perturbative QCD* (2004), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 (cit. on p. 22).

[62] F. Manigrasso. *Direct calculation of Parton Distribution Functions (PDFs) on the lattice.* PhD thesis. Humboldt University Berlin (2022). DOI: 10.18452/24645 (cit. on p. 22).

[63] T. Matsuura. *Higher order corrections to the Drell-Yan process.* PhD thesis. University Leiden (1989). URL: https://inspirehep.net/literature/278862 (cit. on p. 22).

[64] M. Cacciari, G.P. Salam, and G. Soyez. *The anti-k_t jet clustering algorithm.* In: *Journal of High Energy Physics* 2008.4 (2008), p. 63. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 (cit. on pp. 24, 25).

[65] G.P. Salam and G. Soyez. *A practical seedless infrared-safe cone jet algorithm.* In: *Journal of High Energy Physics* 2007.5 (2007), p. 86. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 (cit. on p. 23).

[66] R. Atkin. *Review of jet reconstruction algorithms.* In: *Journal of Physics: Conference Series* 645.1 (2015), p. 012008. DOI: 10.1088/1742-6596/645/1/012008 (cit. on p. 24).

[67] S. Catani et al. *Longitudinally-invariant $k_\perp$-clustering algorithms for hadron-hadron collisions.* In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M (cit. on p. 24).

[68] C. Matteo and G.P. Salam. *Dispelling the N^3 myth for the k_t jet-finder.* In: *Physics Letters B* 641.1 (2006), pp. 57–61. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210 (cit. on p. 25).

[69] ZEUS Collaboration. *Inclusive-jet cross sections in NC DIS at HERA and a comparison of the k_T, anti-k_T and SIScone jet algorithms.* In: *Physics Letters B* 691.3 (2010), pp. 127–137. DOI: 10.1016/j.physletb.2010.06.015. arXiv: 1003.2923 (cit. on p. 25).

[70] D. Bardin et al. *Complete $\mathcal{O}(\alpha)$ QED corrections to the process $ep \rightarrow eX$ in mixed variables.* In: *Physics Letters B* 357.3 (1995), pp. 456–463. DOI: 10.1016/0370-2693(95)00932-b. arXiv: hep-ph/9504423 (cit. on p. 25).

[71] ZEUS Collaboration. *Inclusive dijet Cross Sections in Neutral Current Deep Inelastic Scattering at HERA.* In: *The European Physical Journal C* 70.4 (2010), p. 965. arXiv: 1010.6167 (cit. on pp. 25, 40, 64, 70, 71, 73, 77, 97, 103, 115, 143, 187, 191, 194).

[72] I. Makarenko. *Multijet Production with the ZEUS Detector at HERA.* PhD thesis. University Hamburg (2017). DOI: 10.3204/PUBDB-2018-01866 (cit. on pp. 26, 40, 61, 70, 72, 73, 76, 84, 86, 99).

[73] H. Alvensleben et al. *Validity of Quantum Electrodynamics at Extremely Small Distances.* In: *Physical Review Letters* 21 (1968), pp. 1501–1503. DOI: 10.1103/PhysRevLett.21.1501 (cit. on p. 29).

[74] ARGUS Collaboration. *Physics with ARGUS.* In: *Physics Reports* 276.5 (1996), pp. 224–405. DOI: 10.1016/S0370-1573(96)00008-7 (cit. on p. 29).

[75] P. Söding. *On the discovery of the gluon.* In: *The European Physical Journal H* 35.1 (2010), pp. 3–28. DOI: 10.1140/epjh/e2010-00002-5 (cit. on p. 29).

[76] H1 and ZEUS Collaborations. *Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA.* In: *The European Physical Journal C* 78.6 (2018), p. 473. DOI: 10.1140/epjc/s10052-018-5848-3. arXiv: 1804.01019 (cit. on pp. 29, 38).

[77] H1 and ZEUS Collaborations. *Combinations of Measurement of inclusive deep inelastic $e^{\pm}p$ Scattering Cross Sections and QCD Analysis of HERA Data.* In: *The European Physical Journal C* 75 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 (cit. on pp. 29, 37–40, 141–143, 145, 147, 149).

[78] H1 and ZEUS Collaborations. *Combined inclusive diffractive cross sections measured with forward proton spectrometers in deep inelastic ep scattering at HERA*. In: *The European Physical Journal C* 72.10 (2012). DOI: 10.1140/epjc/s10052-012-2175-y. arXiv: 1207.4864 (cit. on p. 29).

[79] ZEUS Collaboration. *Measurement of Charged and Neutral Current e^-p Deep Inelastic Scattering Cross Sections at High Q^2*. In: *Physical Review Letters* 75 (1995), pp. 1006–1011. DOI: 10.1103/PhysRevLett.75.1006 (cit. on p. 29).

[80] J. Behr. *Jets at high Q^2 at HERA and test beam measurement with the EUDET pixel telescope*. PhD thesis. University Hamburg (2010). DOI: 10.3204/DESY-THESIS-2010-038 (cit. on pp. 30, 58, 61, 69, 70, 72, 73, 76, 84, 86, 89, 91, 97, 99, 103, 124, 128).

[81] M. Seidel. *Luminosity Upgrade of HERA*. In: *18th Biennial Particle Accelerator Conference* (1999). URL: https://cds.cern.ch/record/553192 (cit. on p. 30).

[82] A.A. Sokolov and I.M. Ternov. *On Polarization and Spin Effects in the Theory of Synchrotron Radiation*. In: *Soviet Physics Doklady* 8 (1964), p. 1203 (cit. on p. 30).

[83] M. Beckmann et al. *The Longitudinal Polarimeter at HERA*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 479 (2002), p. 334. DOI: 10.1016/S0168-9002(01)00901-9. arXiv: physics/0009047 (cit. on p. 31).

[84] D.P. Barber et al. *The HERA polarimeter and the first observation of electron spin polarization at HERA*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 329 (1993), p. 79. DOI: 10.1016/0168-9002(93)90924-7 (cit. on p. 31).

[85] ZEUS Collaboration. *The ZEUS Detector*. 1993. URL: http://www-zeus.desy.de/bluebook/bluebook.html (cit. on pp. 31–36).

[86] R. Shehzadi. *Measurement of beauty production in deep inelastic scattering at HERA using decays into electrons*. PhD thesis. University Bonn (2011). URL: https://inspirehep.net/literature/1180236 (cit. on pp. 31, 32).

[87] A. Polini et al. *The design and performance of the ZEUS microvertex detector*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 581 (2007), p. 656. DOI: 10.1016/j.nima.2007.08.167. arXiv: 0708.3011 (cit. on pp. 31, 33).

[88] B. Foster et al. *The Design and Construction of the ZEUS Central Tracking Detector*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 338 (1994), p. 254. DOI: 10.1016/0168-9002(94)91313-7 (cit. on pp. 31, 34).

[89] S. Fourletov. *Straw tube tracking detector (STT) for ZEUS*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 535 (2004), p. 191. DOI: 10.1016/j.nima.2004.07.212 (cit. on p. 32).

[90] A. Bamberger et al. *The Small Angle Rear Tracking Detector of ZEUS*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 401 (1997), p. 63. DOI: 10.1016/S0168-9002(97)01029-2 (cit. on p. 32).

[91] M. Derrick et al. *Design and Construction of the ZEUS Barrel Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 309 (1991), p. 77. DOI: `10.1016/0168-9002(91)90094-7` (cit. on pp. 32, 34).

[92] A. Andresen et al. *Construction and Beam Test of the ZEUS Forward and Rear Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 309 (1991), p. 101. DOI: `10.1016/0168-9002(91)90095-8` (cit. on pp. 32, 34).

[93] I. Kudla et al. *Test of a Prototype of the ZEUS Backing Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 300 (1991), p. 480. DOI: `10.1016/0168-9002(91)90382-Z` (cit. on p. 32).

[94] N. Harnew et al. *Vertex Triggering Using Time Difference Measurements in the ZEUS Central Tracking Detector*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 279 (1989), p. 290. DOI: `10.1016/0168-9002(89)91096-6` (cit. on p. 33).

[95] M. Livan and R. Wigmans. *Misconceptions about Calorimetry*. 2017. arXiv: `1704.00661` (cit. on pp. 34, 35).

[96] A. Dwuraźny et al. *Experimental Study of Electron–Hadron Separation in Calorimeters Using Silicon Diodes*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 277 (1989), p. 176. DOI: `10.1016/0168-9002(89)90550-0` (cit. on p. 34).

[97] A. Caldwell et al. *Design and Implementation of a High-Precision Readout System for the ZEUS Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 321 (1992), p. 356. DOI: `10.1016/0168-9002(92)90413-X` (cit. on p. 35).

[98] A. Bernstein et al. *Beam Tests of the ZEUS Barrel Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 336 (1993), p. 23. DOI: `10.1016/0168-9002(93)91078-2` (cit. on p. 35).

[99] M. Helbich et al. *The Spectrometer System for Measuring ZEUS Luminosity at HERA*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 565 (2006), p. 572. DOI: `10.1016/j.nima.2006.06.049`. arXiv: `physics/0512153` (cit. on p. 35).

[100] L. Adamczyk et al. *Measurement of the luminosity in the ZEUS experiment at HERA II*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 744 (2014), pp. 80–90. DOI: `10.1016/j.nima.2014.01.053`. arXiv: `1306.1391` (cit. on p. 35).

[101] W.H. Smith, K. Tokushuku, and L.W. Wiggers. *The ZEUS trigger system*. In: *Conference on Computing in High-Energy Physics* (1992). DOI: `10.5170/CERN-1992-007.222` (cit. on p. 36).

[102] L. Barzè et al. *Neutral-current Drell-Yan with combined QCD and electroweak corrections in the POWHEG BOX*. In: *The European Physical Journal C* 73.6 (2013). DOI: `10.1140/epjc/s10052-013-2474-y`. arXiv: `1302.4606` (cit. on p. 37).

[103] H1 and ZEUS Collaborations. *Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions*. In: *The European Physical Journal C* 82.3 (2022), p. 243. DOI: 10.1140/epjc/s10052-022-10083-9. arXiv: 2112.01120 (cit. on pp. 37–41, 136, 141, 144, 145, 147, 149, 152).

[104] CTEQ collaboration. *Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions*. In: *The European Physical Journal C* 12 (2000), p. 375. DOI: 10.1007/s100529900196. arXiv: hep-ph/9903282 (cit. on pp. 37, 52, 141, 153).

[105] S. Alekhin et al. *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II*. In: *Physical Review D* 96.1 (2017). DOI: 10.1103/physrevd.96.014011. arXiv: 1701.05838 (cit. on pp. 37, 42, 141, 153).

[106] T.-J. Hou et al. *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*. In: *Physical Review D* 103.1 (2021). DOI: 10.1103/physrevd.103.014013. arXiv: 1912.10053 (cit. on pp. 37, 42).

[107] R.D. Ball et al. *Parton distributions from high-precision collider data*. In: *The European Physical Journal C* 77.10 (2017). DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 (cit. on p. 37).

[108] S. Bailey et al. *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*. In: *The European Physical Journal C* 81.4 (2021). DOI: 10.1140/epjc/s10052-021-09057-0. arXiv: 2012.04684 (cit. on p. 37).

[109] D.M. South and M. Turcato. *Review of searches for rare processes and physics beyond the Standard Model at HERA*. In: *The European Physical Journal C* 76.6 (2016). DOI: 10.1140/epjc/s10052-016-4152-3. arXiv: 1605.03459 (cit. on p. 38).

[110] ZEUS Collaboration. *An NLO QCD analysis of inclusive cross-section and jet-production data from the ZEUS experiment*. In: *The European Physical Journal C* 42.1 (2005), pp. 1–16. DOI: 10.1140/epjc/s2005-02293-x. arXiv: hep-ph/0503274 (cit. on p. 39).

[111] H1 Collaboration. *Measurement of inclusive jet cross sections in photoproduction at HERA*. In: *The European Physical Journal C* 29.4 (2003), pp. 497–513. DOI: 10.1140/epjc/s2003-01262-9. arXiv: hep-ex/0302034 (cit. on p. 40).

[112] H1 Collaboration. *Measurement of the inclusive di-jet cross section in photoproduction and determination of an effective parton distribution in the photon*. In: *The European Physical Journal C* 1.1-2 (1998), pp. 97–107. DOI: 10.1007/bf01245800. arXiv: hep-ex/9709004 (cit. on p. 40).

[113] H1 Collaboration. *Measurement of dijet cross sections in photoproduction at HERA*. In: *The European Physical Journal C* 25.1 (2002), pp. 13–23. DOI: 10.1007/s10052-002-1004-0. arXiv: hep-ex/0201006 (cit. on p. 40).

[114] H1 Collaboration. *Measurement of di-jet cross-sections in photoproduction and photon structure*. In: *Physics Letters B* 483.1-3 (2000), pp. 36–48. DOI: 10.1016/s0370-2693(00)00576-1. arXiv: hep-ex/0003011 (cit. on p. 40).

[115] H1 Collaboration. *Photoproduction of dijets with high transverse momenta at HERA*. In: *Physics Letters B* 639.1 (2006), pp. 21–31. DOI: 10.1016/j.physletb.2006.05.060. arXiv: hep-ex/0603014 (cit. on p. 40).

[116] H1 Collaboration. *Measurement of inclusive jet cross-sections in deep-inelastic ep scattering at HERA*. In: *Physics Letters B* 542.3-4 (2002), pp. 193–206. DOI: 10.1016/s0370-2693(02)02375-4. arXiv: hep-ex/0206029 (cit. on p. 40).

[117] H1 Collaboration. *Jet production in ep collisions at low Q^2 and determination of α_s*. In: *The European Physical Journal C* 67.1-2 (2010), pp. 1–24. DOI: 10.1140/epjc/s10052-010-1282-x. arXiv: 0911.5678 (cit. on p. 40).

[118] H1 Collaboration. *Measurement of jet production cross sections in deep-inelastic ep scattering at HERA*. In: *The European Physical Journal C* 77.4 (2017), p. 215. DOI: 10.1140/epjc/s10052-017-4717-9. arXiv: 1611.03421 (cit. on p. 40).

[119] H1 Collaboration. *Measurement and QCD analysis of jet cross sections in deep-inelastic positron-proton collisions at $\sqrt{s} = 300\,GeV$*. In: *The European Physical Journal C* 19.2 (2001), pp. 289–311. DOI: 10.1007/s100520100621 (cit. on p. 40).

[120] H1 Collaboration. *Measurement of inclusive jet production in deep-inelastic scattering at high Q^2 and determination of the strong coupling*. In: *Physics Letters B* 653.2-4 (2007), pp. 134–144. DOI: 10.1016/j.physletb.2007.07.050. arXiv: 0706.3722 (cit. on p. 40).

[121] H1 Collaboration. *Measurement of Multijet Production in ep Collisions at High Q^2 and Determination of the Strong Coupling α_s*. In: *The European Physical Journal C* 65 (2015), p. 2. arXiv: 1406.4709 (cit. on pp. 40, 42, 64, 69, 96, 137, 138, 147).

[122] ZEUS Collaboration. *Inclusive jet differential cross sections in photoproduction at HERA*. In: *Physics Letters B* 342.1 (1995), pp. 417–432. DOI: 10.1016/0370-2693(94)01510-J (cit. on p. 40).

[123] ZEUS Collaboration. *High-E_T inclusive jet cross sections in photoproduction at HERA*. In: *The European Physical Journal C* 4.4 (1998), pp. 591–606. DOI: 10.1007/s100529800916. arXiv: hep-ex/9802012 (cit. on p. 40).

[124] ZEUS Collaboration. *Inclusive-jet photoproduction at HERA and determination of α_s*. In: *Nuclear Physics B* 864.1 (2012), pp. 1–37. DOI: 10.1016/j.nuclphysb.2012.06.006. arXiv: 1205.6153 (cit. on p. 40).

[125] ZEUS Collaboration. *Dijet cross sections in photoproduction at HERA*. In: *Physics Letters B* 348.3 (1995), pp. 665–680. DOI: 10.1016/0370-2693(95)00275-P (cit. on p. 40).

[126] ZEUS Collaboration. *Dijet cross sections in photoproduction at HERA*. In: *The European Physical Journal C* 1.1-2 (1998), pp. 109–122. DOI: 10.1007/bf01245801. arXiv: hep-ex/9710018 (cit. on p. 40).

[127] ZEUS Collaboration. *Measurement of dijet photoproduction at high transverse energies at HERA*. In: *The European Physical Journal C* 11.1 (1999), pp. 35–50. DOI: 10.1007/s100529900166. arXiv: hep-ex/9905046 (cit. on p. 40).

[128] ZEUS Collaboration. *High-mass dijet cross sections in photoproduction at HERA*. In: *Physics Letters B* 531.1-2 (2002), pp. 9–27. DOI: 10.1016/s0370-2693(02)01327-8. arXiv: hep-ex/0112030 (cit. on p. 40).

[129] ZEUS Collaboration. *Dijet photoproduction at HERA and the structure of the photon*. In: *The European Physical Journal C* 23.4 (2002), pp. 615–631. DOI: 10.1007/s100520200936. arXiv: hep-ex/0112029 (cit. on p. 40).

[130] ZEUS Collaboration. *High-E_T dijet photoproduction at HERA*. In: *Physical Review D* 76.7 (2007). DOI: 10.1103/physrevd.76.072011. arXiv: 0706.3809 (cit. on p. 40).

[131] ZEUS Collaboration. *Measurement of three-jet distributions in photoproduction at HERA*. In: *Physics Letters B* 443.1-4 (1998), pp. 394–408. DOI: `10.1016/s0370-2693(98)01360-4`. arXiv: `hep-ex/9810046` (cit. on p. 40).

[132] ZEUS Collaboration. *Multijet production at low x_{Bj} in deep inelastic scattering at HERA*. In: *Nuclear Physics B* 786.1 (2007), pp. 152–180. DOI: `10.1016/j.nuclphysb.2007.05.027`. arXiv: `0705.1931` (cit. on p. 40).

[133] ZEUS Collaboration. *Inclusive Jet Cross Sections in the Breit Frame in Neutral Current Deep Inelastic Scattering at HERA and Determination of α_s*. In: *Physics Letters* B 547 (2002), p. 164. DOI: `10.1016/S0370-2693(02)02763-6`. arXiv: `hep-ex/0208037` (cit. on pp. 40, 64, 97, 143, 186).

[134] ZEUS Collaboration. *Inclusive-jet and dijet cross sections in deep inelastic scattering at HERA*. In: *Nuclear Physics B* 765.1-2 (2007), pp. 1–30. DOI: `10.1016/j.nuclphysb.2006.09.018`. arXiv: `hep-ex/0608048` (cit. on p. 40).

[135] ZEUS Collaboration. *Measurement of dijet production in neutral current deep inelastic scattering at high Q^2 and determination of α_s*. In: *Physics Letters B* 507.1 (2001), pp. 70–88. DOI: `10.1016/S0370-2693(01)00421-X`. arXiv: `hep-ex/0102042` (cit. on p. 40).

[136] ZEUS Collaboration. *Trijet production in deep inelastic scattering at HERA*. ZEUS-prel-14-008. URL: `https://inspirehep.net/literature/1337742` (cit. on p. 40).

[137] U. Klein. *ZEUS luminosity calculation*. 2023. URL: `https://zeusdp.desy.de/physics/lumi/index.html`. (Cit. on p. 40).

[138] L.W. Whitlow et al. *A precise extraction of $R = \sigma_L/\sigma_T$ from a global analysis of the SLAC deep inelastic e-p and e-d scattering cross sections*. In: *Physics Letters B* 250.1 (1990), pp. 193–198. DOI: `10.1016/0370-2693(90)91176-C` (cit. on p. 41).

[139] BCDMS Collaboration. *A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2*. In: *Physics Letters B* 223.3 (1989), pp. 485–489. DOI: `https://doi.org/10.1016/0370-2693(89)91637-7` (cit. on p. 41).

[140] NuTeV Collaboration. *Precise measurement of dimuon production cross sections in $\nu_\mu Fe$ and $\bar{\nu}_\mu Fe$ deep inelastic scattering at the Tevatron*. In: *Physical Review D* 64.11 (2001). DOI: `10.1103/physrevd.64.112006`. arXiv: `hep-ex/0102049` (cit. on p. 41).

[141] OPAL collaboration. *Determination of α_s using OPAL hadronic event shapes at $\sqrt{s} = 91 - 209\,GeV$ and resummed NNLO calculations*. In: *The European Physical Journal C* 71.9 (2011). DOI: `10.1140/epjc/s10052-011-1733-z`. arXiv: `1101.1470` (cit. on pp. 41, 158).

[142] G. Dissertori et al. *First determination of the strong coupling constant using NNLO predictions for hadronic event shapes in e^+e^- annihilations*. In: *Journal of High Energy Physics* 2008.02 (2008), pp. 040–040. DOI: `10.1088/1126-6708/2008/02/040`. arXiv: `0712.0327` (cit. on pp. 41, 158).

[143] JADE collaboration. *Measurement of the strong coupling α S from the three-jet rate in $e + e−$-annihilation using JADE data*. In: *The European Physical Journal C* 73.3 (2013). DOI: `10.1140/epjc/s10052-013-2332-y`. arXiv: `1205.3714` (cit. on pp. 41, 42, 158).

[144] T. Klijnsma et al. *Determination of the strong coupling constant* $\alpha_s(m_Z)$ *from measurements of the total cross section for top-antitop-quark production*. In: *The European Physical Journal C* 77.11 (2017). DOI: 10.1140/epjc/s10052-017-5340-5. arXiv: 1708.07495 (cit. on p. 41).

[145] CMS Collaboration. *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at* $\sqrt{s}=13$ *TeV*. In: *Journal of High Energy Physics* 2022.2 (2022), p. 142. DOI: 10.1007/JHEP02(2022)142. arXiv: 2111.10431 (cit. on pp. 41, 42, 106, 123, 141, 152, 153).

[146] CMS Collaboration. *Measurement of the* $t\bar{t}$ *production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at* $\sqrt{s} = 13$ *TeV*. In: *The European Physical Journal C* 79.5 (2019). DOI: 10.1140/epjc/s10052-019-6863-8. arXiv: 1812.10505 (cit. on pp. 41, 42, 155, 158).

[147] D. d'Enterria and A. Poldaru. *Extraction of the strong coupling* $\alpha_s(m_Z)$ *from a combined NNLO analysis of inclusive electroweak boson cross sections at hadron colliders*. In: *Journal of High Energy Physics* 2020.6 (2020). DOI: 10.1007/jhep06(2020)016 (cit. on p. 41).

[148] V. Mateu and P.G. Ortega. *Bottom and charm mass determinations from global fits to* $Q\bar{Q}$ *bound states at* N^3LO. In: *Journal of High Energy Physics* 2018.1 (2018). DOI: 10.1007/jhep01(2018)122. arXiv: 1711.05755 (cit. on p. 42).

[149] S. Narison. *QCD parameter correlations from heavy quarkonia*. In: *International Journal of Modern Physics A* 33.10 (2018), p. 1850045. DOI: 10.1142/s0217751x18500458. arXiv: 1801.00592 (cit. on pp. 42, 158).

[150] A. Pich. *Precision tau physics*. In: *Progress in Particle and Nuclear Physics* 75 (2014), pp. 41–85. DOI: 10.1016/j.ppnp.2013.11.002. arXiv: 1310.7922 (cit. on pp. 42, 158).

[151] J. Haller et al. *Update of the global electroweak fit and constraints on two-Higgs-doublet models*. In: *The European Physical Journal C* 78.8 (2018). DOI: 10.1140/epjc/s10052-018-6131-3. arXiv: 1803.01853 (cit. on p. 42).

[152] S. Forte and Z. Kassabov. *Why* α_s *cannot be determined from hadronic processes without simultaneously determining the parton distributions*. In: *The European Physical Journal C* 80.3 (2020). DOI: 10.1140/epjc/s10052-020-7748-6. arXiv: 2001.04986 (cit. on p. 42).

[153] CMS Collaboration. *Measurement of* $t\bar{t}$ *normalised multi-differential cross sections in pp collisions at* $\sqrt{s} = 13$ *TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions*. In: 80.7 (2020). DOI: 10.1140/epjc/s10052-020-7917-7. eprint: 1904.05237 (cit. on p. 42).

[154] R.D. Ball et al. *Precision determination of the strong coupling constant within a global PDF analysis*. In: *The European Physical Journal C* 78.5 (2018). DOI: 10.1140/epjc/s10052-018-5897-7. arXiv: 1802.03398 (cit. on p. 42).

[155] T. Cridge et al. *An investigation of the* α_s *and heavy quark mass dependence in the MSHT20 global PDF analysis*. In: *The European Physical Journal C* 81.8 (2021). DOI: 10.1140/epjc/s10052-021-09533-7. arXiv: 2106.10289 (cit. on p. 42).

[156] S. Aoki et al. *FLAG Review 2019*. In: *The European Physical Journal C* 80.2 (2020). DOI: 10.1140/epjc/s10052-019-7354-7. arXiv: 1902.08191 (cit. on p. 42).

[157] A. Bazavov et al. *Nonperturbative QCD simulations with 2+1 flavors of improved staggered quarks.* In: *Reviews of Modern Physics* 82.2 (2010), pp. 1349–1417. DOI: 10.1103/revmodphys.82.1349. arXiv: 0903.3598 (cit. on p. 43).

[158] S. Dürr et al. *Ab Initio Determination of Light Hadron Masses.* In: *Science* 322.5905 (2008), pp. 1224–1227. DOI: 10.1126/science.1163233. arXiv: 0906.3599 (cit. on p. 43).

[159] J. Currie et al. *NNLO QCD corrections to jet production in deep inelastic scattering.* In: *Journal of High Energy Physics* 2017.18 (2017). [Erratum: JHEP **12**, 042 (2020)]. DOI: 10.1007/jhep07(2017)018. arXiv: 1703.05977 (cit. on pp. 45, 46, 48, 64, 120, 136, 146).

[160] X. Chen et al. *NNLO QCD corrections to Higgs boson production at large transverse momentum.* In: *Journal of High Energy Physics* 2016.10 (2016). DOI: 10.1007/jhep10(2016)066. arXiv: 1607.08817 (cit. on pp. 46, 47).

[161] J.M. Campbell, M.A. Cullen, and E.W.N. Glover. *Four jet event shapes in electron-positron annihilation.* In: *The European Physical Journal C* 9.2 (1999), pp. 245–265. DOI: 10.1007/s100529900034 (cit. on p. 47).

[162] Z. Kunszt and D.E. Soper. *Calculation of jet cross sections in hadron collisions at order α_s^3.* In: *Physical Review D* 46 (1992), pp. 192–221. DOI: 10.1103/PhysRevD.46.192 (cit. on p. 47).

[163] A. Gehrmann-De Ridder, T. Gehrmann, and E.W.N. Glover. *Antenna subtraction at NNLO.* In: *Journal of High Energy Physics* 2005.09 (2005), pp. 056–056. DOI: 10.1088/1126-6708/2005/09/056. arXiv: hep-ph/0505111 (cit. on p. 48).

[164] P. Connor. *Precision measurement of the inclusive b jet production in proton proton collisions with the CMS experiment at the LHC at $\sqrt{s} = 13\,TeV$.* Dissertation, Universität Hamburg, 2018. PhD thesis. University Hamburg (2018). DOI: 10.3204/PUBDB-2018-02244 (cit. on pp. 48, 108, 112).

[165] G. Ingelman, A. Edin, and J. Rathsman. *Lepto 6.5: A Monte Carlo Generator for Deep Inelastic Lepton–Nucleon Scattering.* In: *Computer Physics Communications* 101 (1997), p. 108. DOI: 10.1016/S0010-4655(96)00157-9. arXiv: hep-ph/9605286 (cit. on pp. 48, 50, 53).

[166] H. Jung. *Hard diffractive scattering in high energy ep collisions and the Monte Carlo Generator RAPGAP.* In: *Computer Physics Communications* 86.1 (1995), pp. 147–161. DOI: 10.1016/0010-4655(94)00150-Z (cit. on pp. 48, 51).

[167] L. Lönnblad. *Ariadne Version 4 – A Program for Simulation of QCD Cascades Implementing the Colour Dipole Model.* In: *Computer Physics Communications* 71 (1992), p. 15. DOI: 10.1016/0010-4655(92)90068-A (cit. on pp. 48, 51).

[168] G.S. Bali and K. Schilling. *Static quark-antiquark potential: Scaling behavior and finite-size effects in SU(3) lattice gauge theory.* In: *Physical Review D* 46 (1992), pp. 2636–2646. DOI: 10.1103/PhysRevD.46.2636 (cit. on p. 49).

[169] B. Andersson et al. *Parton fragmentation and string dynamics.* In: *Physics Reports* 97.2 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7 (cit. on p. 49).

[170] T. Sjöstrand. *High-Energy Physics Event Generation with Pythia 5.7 and Jetset 7.4.* In: *Computer Physics Communications* 82 (1994), p. 74. DOI: 10.1016/0010-4655(94)90132-5. arXiv: hep-ph/9508391 (cit. on pp. 49, 51).

[171] A. Metz and A. Vossen. *Parton fragmentation functions*. In: *Progress in Particle and Nuclear Physics* 91 (2016), pp. 136–202. DOI: 10.1016/j.ppnp.2016.08.003. arXiv: 1607.02521 (cit. on p. 49).

[172] B. Webber. *Parton shower Monte Carlo event generators*. In: *Scholarpedia* 6.12 (2011), p. 10662. DOI: 10.4249/scholarpedia.10662 (cit. on p. 50).

[173] G. Corcella et al. *Herwig 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*. In: *Journal of High Energy Physics* 1 (2001), p. 10. DOI: 10.1088/1126-6708/2001/01/010. arXiv: hep-ph/0011363 (cit. on p. 51).

[174] A. Kwiatkowski, H. Spiesberger, and H.-J. Möhring. *Heracles: An Event Generator for ep Interactions at HERA Energies Including Radiative Processes*. In: *Computer Physics Communications* 69 (1992), p. 155. DOI: 10.1016/0010-4655(92)90136-M (cit. on pp. 51, 53).

[175] R. Brun et al. *GEANT3 User's Guide*. 1987. URL: https://cds.cern.ch/record/1119728 (cit. on p. 51).

[176] J. Malka and K. Wichmann. *The ZEUS data preservation project*. In: *Journal of Physics: Conference Series* 396.2 (2012), p. 022033. DOI: 10.1088/1742-6596/396/2/022033 (cit. on p. 51).

[177] ZEUS Collaboration. *ZECoNt, the ZEUS Common Ntuple project*. 2023. URL: https://www-zeus.desy.de/ZEUS_ONLY/analysis/comntp/comntp_v08.php. (Cit. on p. 52).

[178] H. Spiesberger et al. *Radiative corrections at HERA*. In: (1992). URL: http://cds.cern.ch/record/237380 (cit. on p. 53).

[179] ZEUS Collaboration. *Measurement of High-Q^2 Neutral Current e^+p Deep Inelastic Scattering Cross Sections at HERA*. In: *The European Physical Journal C* C 11 (1999), p. 427. DOI: 10.1007/s100520050645. arXiv: hep-ex/9905032 (cit. on pp. 56, 97).

[180] H. Abramowicz, A. Caldwell, and R. Sinkus. *Neural Network Based Electron Identification in the ZEUS Calorimeter*. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* A 365 (1995), p. 508. DOI: 10.1016/0168-9002(95)00612-5 (cit. on p. 56).

[181] A.L-D. Viani and S. Schlenstedt. *Electron finder efficiencies and impurities. A comparison between SINISTRA95, EM and EMNET*. ZEUS internal note. ZEUS note 99-077. 2000 (cit. on p. 56).

[182] S. Bentvelsen, J. Engelen, and P. Kooijman. *Reconstruction of (x, Q^2) and extraction of structure functions in neutral current scattering at HERA*. In: *Proceedings of the Workshop on Physics at HERA*. NIKHEF-H-92-02. (1992). URL: https://inspirehep.net/literature/332514 (cit. on pp. 58, 63).

[183] K.C. Hoeger. *Measurement of x, y, Q^2 in Neutral Current Events*. In: *Proceedings of the Workshop on Physics at HERA*. Vol. 1. (1992), 43ff (cit. on p. 58).

[184] H. Raach. *Inclusive jet cross-sections in neutral current deep inelastic scattering in the Breit frame at ZEUS*. PhD thesis. University Freiburg (2001). URL: https://inspirehep.net/literature/565494 (cit. on p. 61).

[185] T. Theedt. *Measurement of Dijet Cross Sections in Deep Inelastic ep Scattering at HERA*. PhD thesis. University Hamburg (2009). DOI: `10.3204/DESY-THESIS-2009-046` (cit. on p. 61).

[186] M. Jimenez. *Tests of color dynamics and precision measurements of α_s using jet production in DIS with the ZEUS detector at HERA*. PhD thesis. Autonoma University Madrid (2008). URL: `https://inspirehep.net/literature/890589` (cit. on p. 61).

[187] H1 Collaboration. *Measurement of the inclusive ep scattering cross section at low Q^2 and x at HERA*. In: *The European Physical Journal C* 63.4 (2009), pp. 625–678. DOI: `10.1140/epjc/s10052-009-1128-6`. arXiv: `0904.0929` (cit. on p. 63).

[188] ZEUS Collaboration. *ZeVis – ZEUS event visualisation*. 2023. URL: `https://www-zeus.desy.de/~zevis/`. (Cit. on p. 66).

[189] F. Januschek. *Measurements of e^+p neutral current deep inelastic scattering with a longitudinally polarised positron beam and x-ray radiation damage for silicon sensors*. PhD thesis. University Hamburg (2011). URL: `https://inspirehep.net/literature/1116568` (cit. on pp. 67, 70, 72, 73, 76, 81, 82, 84–86, 99).

[190] J. Grosse-Knetter. *Corrections for the Hadronic Final State*. ZEUS internal note. ZEUS note 98-031. 1998 (cit. on pp. 68, 71).

[191] ZEUS Collaboration. *Measurement of high-Q^2 neutral current deep inelastic e^+p scattering cross sections with a longitudinally polarized positron beam at HERA*. In: *Physical Review D* 87 (2013), p. 052014. DOI: `10.1103/PhysRevD.87.052014`. arXiv: `1208.6138` (cit. on pp. 70, 71, 73, 125).

[192] J. Edmonds, J. Grosse-Knetter, and A. Quadt. *Neutral Current Cross Sections at High Q^2*. ZEUS internal note. ZEUS note 99-016. 1999 (cit. on p. 81).

[193] S. Schmitt. *Data Unfolding Methods in High Energy Physics*. In: *EPJ Web of Conferences* 137 (2017), p. 11008. DOI: `10.1051/epjconf/201713711008`. arXiv: `abs/1611.01927` (cit. on pp. 96, 105).

[194] S. Schmitt. *TUnfold, an algorithm for correcting migration effects in high energy physics*. In: *Journal of Instrumentation* 7.10 (2012), T10003–T10003. DOI: `10.1088/1748-0221/7/10/t10003` (cit. on pp. 96, 106).

[195] T.P. Stewart. *Measurement of high-Q^2 neutral current cross-sections with longitudinally polarised positrons with the ZEUS detector*. PhD thesis. Toronto University (2012). URL: `https://inspirehep.net/literature/1126165` (cit. on p. 99).

[196] D.A. Britzger. *Regularized Unfolding of Jet Cross Sections in Deep-Inelastic ep Scattering at HERA and Determination of the Strong Coupling Constant*. PhD thesis. University Hamburg (2013). DOI: `10.3204/DESY-THESIS-2013-045` (cit. on pp. 107, 127).

[197] A. Arbuzov et al. *Hector 1.00: A Program for the Calculation of QED, QCD and Electroweak Corrections to ep and $\ell^{\pm}N$ Deep Inelastic Neutral and Charged Current Scattering*. In: *Computer Physics Communications* 94 (1996), p. 128. DOI: `10.1016/0010-4655(96)00005-7`. arXiv: `hep-ph/9511434` (cit. on p. 121).

[198] ZEUS Collaboration. *Measurement of high-Q^2 e^-p neutral current cross sections at HERA and the extraction of xF_3*. In: *The European Physical Journal C* 28 (2003), p. 175. DOI: `10.1140/epjc/s2003-01163-y`. arXiv: `hep-ex/0208040` (cit. on p. 125).

[199] ZEUS Collaboration. *Measurement of high-Q^2 neutral current deep inelastic e^-p scattering cross sections with a longitudinally polarised electron beam at HERA*. In: *The European Physical Journal C* 62.4 (2009), pp. 625–658. DOI: 10.1140/epjc/s10052-009-1055-6. arXiv: 0901.2385 (cit. on p. 125).

[200] ZEUS Collaboration. *Introduction to ZEUS analysis*. 2023. URL: https://zeusdp.desy.de/ZEUS_ONLY/analysis/primer/. (Cit. on p. 133).

[201] S. Alekhin et al. *HERAFitter, Open Source QCD Fit Project*. In: *P* 75.7 (2015), p. 304. DOI: 10.1140/epjc/s10052-015-3480-z. arXiv: 1410.4412 (cit. on pp. 141, 154).

[202] xFitter developers. *xFitter – A PDF fit program from HERA*. 2023. URL: https://gitlab.cern.ch/fitters/xfitter. (Cit. on pp. 141, 155).

[203] H1 Collaboration. *Inclusive deep inelastic scattering at high Q^2 with longitudinally polarised lepton beams at HERA*. In: *Journal of High Energy Physics* 2012.9 (2012), p. 61. DOI: 10.1007/JHEP09(2012)061. arXiv: 1206.7007 (cit. on p. 142).

[204] R.D. Cousins. *On Goodness-of-Fit Tests*. 2016. URL: http://cousins.web.cern.ch/cousins/ongoodness6march2016.pdf (cit. on p. 142).

[205] R.S. Thorne. *Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections*. In: *Physical Review D* 86.7 (2012). DOI: 10.1103/physrevd.86.074017. arXiv: 1201.6180 (cit. on p. 145).

[206] H1 Collaboration. *Determination of the strong coupling constant $\alpha_s(M_Z)$ in next-to-next-to-leading order QCD using H1 jet cross section measurements*. In: *The European Physical Journal C* 77.11 (2017). [Erratum: EPJC 81.8, 738 (2021)], p. 791. DOI: 10.1140/epjc/s10052-017-5314-7. arXiv: 1709.07251 (cit. on pp. 155, 158).

[207] D. Britzger et al. *Calculations for deep inelastic scattering using fast interpolation grid techniques at NNLO in QCD and the extraction of α_s from HERA data*. In: *The European Physical Journal C* 79.10 (2019). DOI: 10.1140/epjc/s10052-019-7351-x. arXiv: 1906.05303 (cit. on p. 158).

[208] ZEUS Collaboration. *Measurement and QCD analysis of inclusive jet production in deep inelastic scattering at HERA*. ZEUS-prel-22-001. 2022 (cit. on p. 160).

[209] ZEUS Collaboration. *Measurement of jet production in deep inelastic scattering and NNLO determination of the strong coupling at ZEUS*. 2023. arXiv: 2309.02889 (cit. on p. 160).

[210] H. Haber. *Useful relations among the generators in the defining and adjoint representations of SU(N)*. In: *SciPost Physics Lecture Notes* (2021). DOI: 10.21468/scipostphyslectnotes.21. arXiv: 1912.13302 (cit. on p. 163).

List of Figures