Regularized Unfolding of Jet Cross Sections in Deep-Inelastic ep Scattering at HERA and Determination of the Strong Coupling Constant

Dissertation

zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Daniel Andreas Britzger

aus Marktoberdorf

Hamburg
2013

Gutachter der Dissertation:	Dr. Günter Grindhammer
	Dr. Alexandre Glazov
Gutachter der Disputation:	Prof. Dr. Peter Schleper
	Dr. Alexandre Glazov
Datum der Disputation:	14. Juni 2013
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Peter Hauschildt
Dekan der MIN Fakultät:	Prof. Dr. Daniela Pfannkuche
Leiter des Department Physik:	Prof. Dr. Heinrich Graener

Abstract

In this thesis double-differential cross sections for jet production in neutral current deep-inelastic $e^{\pm}p$ scattering (DIS) are presented at the center-of-mass energy of $\sqrt{s} = 319\,\text{GeV}$, and in the kinematic range of the squared four-momentum transfer $150 < Q^2 < 15\,000\,\text{GeV}^2$ and the inelasticity $0.2 < y < 0.7$. Jets are reconstructed in the Breit frame of reference using the k_T-algorithm and are constrained to the pseudorapidity range $-1.0 < \eta^\text{jet}_\text{lab} < -2.5$ in the laboratory rest frame. Inclusive jet measurements are performed for jets with transverse momenta of $7 < p^\text{jet}_\text{T} < 50\,\text{GeV}$. Dijet and trijet observables are measured for events with at least two or three jets with transverse momenta of $5 < p^\text{jet}_\text{T} < 50\,\text{GeV}$, where the invariant mass of the two jets with the highest transverse momenta is required to be greater than 16 GeV.

The data are corrected for detector effects, such as limited acceptance and migrations due to limited detector resolutions, using a multidimensional regularized unfolding procedure which features a correct propagation of the statistical uncertainty. The matrix based unfolding corrects the neutral current DIS, the inclusive jet, the dijet and the trijet measurements simultaneously, and it considers migrations in up to eight variables. The simultaneous unfolding enables to constrain contributions from jet multiplicities differing between detector and hadron level using the neutral current DIS kinematics of such events. Furthermore, jet cross sections normalized to the inclusive neutral current DIS cross section and ratios of jet cross sections are obtained, since the statistical correlations between these observables are known.

The jet cross sections are used to determine the strong coupling constant $\alpha_s(M_Z)$ at the scale of the mass of the Z^0 boson in the framework of perturbative quantum chromodynamics in next-to-leading order. Values are derived separately for the absolute and normalized jet cross section measurements. A higher sensitivity to $\alpha_s(M_Z)$ is obtained in a simultaneous least-square-minimization procedure to the three jet cross sections, taking the statistical correlations and correlations due to other experimental uncertainties into account. The most precise value is obtained from all normalized jet cross sections, yielding $\alpha_s(M_Z) = 0.1165 \pm 0.0008$, which benefits from the high statistical precision of the inclusive jet measurement, the increased sensitivity to $\alpha_s(M_Z)$ of the trijet cross section, and from the cancellation of normalization uncertainties. However, the value of the strong coupling constant is currently only determinable from this measurement with a precision of 3 to 4 % due to the limited precision of the theoretical predictions at next-to-leading order.

Kurzfassung

In dieser Arbeit werden doppelt-differenziell gemessene Wirkungsquerschnitte für Jetproduktion in tiefunelastischer $e^{\pm}p$-Streuung (DIS) bei einer Schwerpunktsenergie von $\sqrt{s} = 319\,\mathrm{GeV}$ im kinematischen Bereich des quadratischen Viererimpulsübertrags $150 < Q^2 < 15\,000\,\mathrm{GeV}^2$ und einer Inelastizität von $0{,}2 < y < 0{,}7$ vorgestellt. Jets werden im Breit-Bezugssystem mithilfe des k_T-Algorithmus rekonstruiert und sind auf den Pseudorapiditätsbereich $-1{,}0 < \eta^\mathrm{jet}_\mathrm{lab} < 2{,}5$ im Laborsystem beschränkt. Eine inklusive Jetmessung wird für Jets mit einem Transversalimpuls von $7 < p^\mathrm{jet}_\mathrm{T} < 50\,\mathrm{GeV}$ durchgeführt. Dijet- und Trijet-Observablen werden für Ereignisse gemessen welche mindestens zwei bzw. drei Jets mit einem Transversalimpuls von $5 < p^\mathrm{jet}_\mathrm{T} < 50\,\mathrm{GeV}$ aufweisen und deren invariante Masse der beiden Jets mit dem höchsten Impuls mindestens 16 GeV beträgt.

Bei diesen Messungen werden Detektoreffekte, wie beispielweise die begrenzte Akzeptanz oder kinematische Migrationen aufgrund der endlichen Detektorauflösung, mithilfe der Methode der mehrdimensionalen regularisierten Entfaltung korrigiert, welche sich insbesondere durch eine vollständige Fehlerfortpflanzung auszeichnet. Die auf einer Matrix basierende Entfaltungsmethode korrigiert die inklusive Messung der Streuung durch neutralen Strom, die inklusive Jetmessung, die Dijet- sowie die Trijetmessung gleichzeitig, wobei Migrationen in bis zu acht Variablen berücksichtigt werden. Diese simultane Entfaltung ermöglicht es auch die Beiträge aufgrund verschiedener Jetmultiplizitäten auf Detektor- und Hadronlevel durch die Verwendung der Kinematik des neutralen Stroms dieser Ereignisse zu berücksichtigen. Zudem können Jetwirkungsquerschnitte normiert auf die Wirkungsquerschnitte des neutralen Stroms und Verhältnisse von Jetwirkungsquerschnitten bestimmt werden, da die statistischen Korrelationen zwischen den einzelnen Observablen bekannt sind.

Die Jetwirkungsquerschnitte werden angewendet um die Starke Wechselwirkungskonstante $\alpha_s(M_Z)$ bei der Energieskala der Masse des Z^0-Bosons im Rahmen der perturbativen Quantenchromodynamik in nachführender Ordnung zu bestimmen. Werte für $\alpha_s(M_Z)$ werden durch eine Minimierung der kleinsten Quadrate zuerst separat von den absoluten sowie von den normierten Messungen bestimmt. Eine höhere Sensitivität auf $\alpha_s(M_Z)$ wird bei der gleichzeitigen Berücksichtigung aller drei Jetmessungen erzielt, wobei die statistischen Korrelationen und die Korrelationen weiterer experimenteller Unsicherheiten berücksichtigt werden. Der präziseste Wert für $\alpha_s(M_Z)$ von 0.1165 mit einer experimentellen Unsicherheit von ± 0.0008 wird bei gleichzeitiger Berücksichtigung aller drei normierten Jetwirkungsquerschnitte erzielt. Seine Genauigkeit profitiert von der hohen Statistik der inklusiven Jetmessung, der erhöhten Sensitivität der Trijetmessung auf $\alpha_s(M_Z)$ sowie der Aufhebung der Normierungsunsicherheiten. Dennoch kann die Kopplungskonstante der starken Wechselwirkung derzeit von diesen Messgrößen lediglich mit einer Genauigkeit von 3 bis 4 % wegen der begrenzten Vorhersagekraft der theoretischen Rechnungen in nachführender Ordnung bestimmt werden.

Contents

Chapter 1

Introduction

Jet production in deep-inelastic $e^{\pm}p$ scattering (DIS) is an important process to study quantum chromodynamics, the theory of the strong interaction. In DIS, the measurement of observables related to jets can be performed in the so-called Breit frame of reference, in which the presence of jets is directly related to a strong vertex. In the present thesis, cross sections of inclusive jet production, where every single jet is counted, and cross sections of dijet and trijet production, where events with at least two or three jets are counted, are measured. Also jet cross sections normalized to the neutral current DIS cross sections are obtained. The neutral current DIS process of this measurement is in the kinematic range of the squared four-momentum transfer $150 < Q^2 < 15\,000\,\mathrm{GeV}^2$ and the inelasticity $0.2 < y < 0.7$. The theoretical concepts for jet production in DIS are outlined in chapter 2 and provide also the theoretical background for the determination of the strong coupling constant. In the subsequent chapter the experimental setup of the HERA collider and the H1 detector is summarized. The event recording, reconstruction and the determination of the observables are given in chapter 4 and the data is compared to Monte Carlo simulations. The focus of this work lies particularly on a regularized unfolding method for the correction of detector effects. The regularized unfolding is mathematically introduced in chapter 5 and the details of the method employed, which corrects the neutral current DIS, the inclusive jet, the dijet and the trijet measurements simultaneously, is outlined in chapter 6. The jet cross sections obtained are of the final precision as achieved by H1 and are presented in chapter 7 and are compared to theoretical predictions. The multidimensional unfolding method is studied in detail in chapter 9 and is compared to the widely used bin-by-bin method in chapter 8. The measured data are utilized for the determination of the strong coupling constant in chapter 10.

Chapter 2

Theory

Physical studies are always an interplay between experimental measurements and theoretical predictions. A fundamental input to any kind of research are precise experimental observations of nature, while theoretical concepts try to describe these measurements, and therefore contribute to the structured understanding of natures phenomena. Moreover theoretical understanding is required to design reasonable experiments and to define valuable quantities which are ascertainable in experiment.
In this thesis theoretical concepts are used when employing Monte Carlo event generators for the correction of detector effects. Moreover, precise theoretical predictions of the strong interaction are compared to the measured cross sections, and furthermore the measured cross sections are used to extract the strong coupling constant. Since this fundamental constant is not directly observable in experiment but can only be determined within the scope of quantum chromodynamics as the theory of the strong interaction, a precise understanding of the underlying theoretical concepts is mandatory, especially for the meaningful interpretation of the results. This chapter will outline the relevant theoretical concepts, with a particular focus on the determination of the strong coupling constant.

2.1 Basic theoretical concepts

Most of the theoretical concepts of particle physics are embed in the framework of the standard model of particle physics. In this work, the theory of the strong interaction, quantum chromodynamics, is of main interest and is discussed in more details.

2.1.1 The standard model of particle physics

The *standard model of particle physics* (SM) is a collection of theoretical concepts which describe fundamental particles and their interactions. This model is based on relativistic quantum field theory, where the interaction between quanta is mediated by quantized gauge fields.
The fundamental particles are assumed to be massive fermions of spin 1/2, which come in three families. The fermions carry quantum numbers of electric charge, (weak) isospin T, and color, which determines their coupling to the force-mediating gauge fields. The weak hypercharge can be calculated from the third component of the the weak isospin and the electric charge, through $Y_W = 2(Q - T_3)$.
The electromagnetic and the weak force are described by a U(1)×SU(2) symmetry. It is spontaneously broken through the coupling to a scalar-field, the Higgs-field. After spon-

taneous symmetry breaking, the force-mediating gauge bosons are a mixture of the the original gauge fields and result in the massive $W^{\pm}$ and the Z^0 bosons as well as the mass-less exchange particle of the electromagnetic force, the photon. The weak force only couples to left-handed particles. The fermions interact with the gauge bosons in the weak eigen-states, which are given by rotating the mass-eigenstates by the CKM matrix [1, 2]. The corresponding unitary matrix in the lepton sector is the PMNS matrix [3–6], which gives the strength of neutrino oscillations.
The strong force is described by a SU(3)-symmetry and is assumed to be an unbroken symmetry. The structure of the gauge group predicts eight mass-less bosons, the gluons, which carry two color charges each. Only quarks, but no leptons carry color charges. As a consequence of the self-coupling of the gluons, quarks (besides the top quark) are only observable in color-less bound states of two or three quarks, the mesons or baryons, respectively.
The field quanta of the Higgs-field, which causes electroweak symmetry breaking, have probably been detected recently at the LHC [7,8]. As the Higgs-mechanism is the principle of mass-generation of the $W^{\pm}$ and Z^0 bosons, it is expected that a similar mechanism might cause the creation of the fermion masses, e.g. through a Yukawa-coupling of the fermions to the Higgs-field with coupling strengths of the order of the fermion masses.
The standard model obeys the rules of special relativity, but does not involve the metric and rules of general relativity. It is therefore not capable of describing gravitational effects.

2.1.2 Quantum chromodynamics

Quantum chromodynamics [9,10] (QCD)[1] is the theory of the strong force. It fully describes the interaction between quarks[2] and gluons[3], the fundamental building blocks of hadrons, within the standard model of particle physics.
QCD is a quantum field theory with the special unitary group SU(3) as gauge group. The corresponding quantum number is called color, in correspondence to the three degrees of freedom of the fundamental representation of the SU(3). The gauge invariant Lagrangian is given by [10–12]

$$\mathcal{L}_{\text{classical}} = \sum_f \bar{\psi}_{f,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu \lambda^C_{ab} \mathcal{A}^C_\mu - m_f \delta_{ab})\psi_{f,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu} \ , \tag{2.1}$$

where $\psi_{f,a}$ are the quark-field spinors for the quark flavor f and the color-index a that runs from $a = 1$ to $N_c = 3$. This implies that each quark with mass m_q, comes in three colors. The γ^μ symbols denote the Dirac γ-matrices and λ^C_{ab} correspond to the eight 3×3 Gell-Mann matrices which are one possible representation of the infinitesimal generators of the SU(3) group. For a proper quantization procedure the *gauge-fixing* term and the *Faddeev-Popov ghost* terms have to be added [13]. The full QCD Lagrangian then reads [14]

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{classical}} + \mathcal{L}_{\text{gauge-fixing}} + \mathcal{L}_{\text{ghost}} \ . \tag{2.2}$$

[1]The name *quantum chronodynamics* is derived from the greek word $X\rho\omega\mu\alpha$ ("chroma"), which means color. *Chromodynamcis* is a made-up word in analogy to the name of the theory of the electric charge, electrodynamics.

[2]The name *quark* was proposed by M. Gell-Mann and refers to a sentence in the book "Finnegans Wake" by J. Joyce: "Three quarks for Muster Mark!". G. Zweig proposed the name *aces*, while R. Feynman named the constituents of the hadron *parton* (which is nowadays used for gluons and partons).

[3]M. Gell-Mann derived the name *gluon* from the word "glue" in order to emphasize the fact that these quanta *glue* the hadron constituents together, although the more stringent name would have been *chromon* as proposed by H. Fritzsch, following the terminus of greek names for fundamental particles.

The field-strength tensor of QCD is derived from the gluon fields $\mathcal{A}_\alpha^A$ and is given by

$$F_j^{\mu\nu} = \partial^\nu \mathcal{A}_j^\mu - \partial^\mu \mathcal{A}_j^\nu + g_s f_{jkl} \mathcal{A}_k^\mu \mathcal{A}_l^\nu \ , \tag{2.3}$$

where f_{jkl} are the structure constants of the SU(3) Lie algebra. The third summand of the field-strength tensor is a direct consequence of the non-Abelian structure of the SU(3) gauge group, where the commutators of the generators are non-vanishing, but obey the commutation relation

$$[\lambda_i, \lambda_j] = i f^{ijk} \lambda_k \ . \tag{2.4}$$

These non-trivial structure functions are one of the main differences of QCD to the well-established quantum electrodynamics (QED) and results in a self-coupling of the gauge-field. This term can be interpreted in the way the qunanta of the gauge field, the gluons, are carrying two color charges. From the Lagrangian the Feynman rules can derived, which also help as graphical illustration of the theory. The Feynman diagrams of the interaction vertices are depicted in figure 2.1.

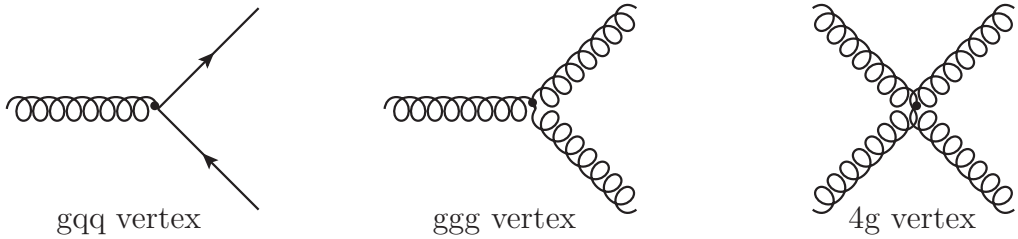


Figure 2.1: The three fundamental vertices of QCD as derived from the QCD-Lagrangian. The solid lines denote quark propagators and the curly lines denote gluon propagators. The gluon radiation vertex qgg is in analogy to QED. The ggg-vertex, which is proportional to the coupling g, and the 4-gluon vertex which is proportional to (g_s^2). The last two indicate the gluon self-coupling as a consequence of the non-Abelian structure of the gauge group.

The big success of QCD is owed the fact that QCD is able to describe two quite opposing features of the strong force. *Asymptotic freedom* [15, 16] refers to the weakness of the short-distance interaction, while the *confinement* of quarks follows from its strength at long distances [17]. Quarks or any other building blocks of hadrons have never been observed as free particles[4] but only confined states like mesons and baryons were observed in experiments. The success of the quark model and the eightfold way, which explains the hadron spectra by assuming that these objects are bound states, gave rise to the assumption of a substructure of hadrons. Since such hypothetical constituents only show up in bound states, the force between them have to be strong. However, the parton model (see section 2.5.3.1) was able to describe scattering data very well. Paradoxically, it assumes that these constituents do not interact (strongly) at all. The parton model further predicted that the constituents must be spin 1/2-particles. Both effects are intrinsic features of QCD and are accommodated through the structure of the force between two quarks as function of their distance, which is expressed through an effective coupling (see section 2.1.4.2). With the discovery of the gluon at the PETRA collider [18–20], QCD became the standard theory of the strong interaction.

[4]Quarks never exist on their own for a time longer than $\sim 1/\Lambda_{\mathrm{QCD}}$. However, the 1995 discovered top-quark cannot form a bound state due to its short lifetime and its large mass.

2.1.3 Perturbative QCD

Further predictions and consequences of QCD are complicated to derive, although various techniques have been developed, like e.g. lattice QCD or effective theories. To derive prediction for scattering processes from the QCD Lagrangian, perturbative methods are employed. However, the appearance of infrared and ultra-violet divergences set direct and indirect limitations on the application of perturbative methods and it becomes obvious that fixed-order perturbation theory captures only a small part for the prescription of QCD effects.

2.1.3.1 Perturbative expansion of an observable

A physical observable [5] σ is expressed in *perturbative QCD* (pQCD) by a perturbative series in powers of the strong coupling strength α_s according to

$$\sigma = k_0 + \alpha_s^1 k_1 + \alpha_s^2 k_2 + \ldots = \sum_{i=0}^{n} \alpha_s^i k_i \ , \tag{2.5}$$

where k are so-called perturbative coefficients and can be calculated from Feynman diagrams in order-by-order. If the coupling parameter is sufficiently small, i.e. if $\alpha_s \ll 1$, the series will converge sufficiently quickly. In this case, it can provide a realistic prediction of the observable even if only a limited number of perturbative orders are known. The contributions k_i from each order have to be derived from appropriate Feynman diagrams of QCD, related to the corresponding order in α_s. Perturbative calculations in leading order (LO) are defined as lowest order of the perturbative expansion which still gives contributions to the observable, i.e. in LO it is not necessarily $n = 0$. Therefore, only tree-level Feynman diagrams without internal loops can contribute to the leading order calculation.

2.1.3.2 Ultra-violet divergences

At higher orders of the perturbative series the calculated observable receives contributions from diagrams including loop corrections. In these graphs the integrals have to be performed over all (internal) loop momenta. As a consequence, these integrals are ultra-violet (UV) divergent for integration-momenta going to infinity. In order to obtain a finite (physical) quantity these divergences have to be made finite by a regularization procedure, for instance by introducing a momentum cut-off scale or by dimensional regularization [21]. This procedure introduces an arbitrary new scale μ_r, the *renormalization scale*, and the UV divergences are absorbed in the definition of the coupling strength. This modification of the coupling exactly compensates the contributions from self-energies to all orders [22] through the vertex correction. The most commonly used regularization scheme is the *modified minimal-subtraction scheme* ($\overline{\text{MS}}$ scheme) [23], which is also employed for all calculations consistently in this thesis. The consequences of the renormalization procedure leads to the 'running' of the strong coupling and is discussed in more detail in section 2.1.4.

2.1.3.3 Soft and collinear divergences

Already tree-level amplitudes are typically divergent for most observables. These divergences arise from singularities at the phase-space boundary and also from vanishing gluon

[5] The letter σ could stand for any physical observable which could be of interest, but not only cross sections as suggested by 'σ'.

(and quark) masses, and two distinct kinematical configurations of infrared divergences for mass-less calculations can be identified.

- **Soft** If the gluon or quark energy is tending to the soft limit the amplitude becomes singular;
- **Collinear** The collinear limit occurs if the 4-momentum of the gluon is parallel to that either of the incoming or outgoing parton and their angular separation vanishes, i.e. $\theta_{qg} \to 0$ (or π)[6].

The consequences of soft and collinear divergences are on the one hand, that observables have to be defined *infrared safe* [24], which means that they should be defined as inclusive as possible and should not depend on the long-distance behavior of the theory[7]. In order to resolve multiple partons[8] the definition of infrared-safe observables with *jet-algorithms* becomes necessary and is discussed in more detail in section 2.4. On the other hand, the divergent long-distance dependence of the scattering amplitudes has to be seperated from the short-distance dependence. This will lead to the *factorization theorem*, which is discussed in more detail with its consequences in section 2.1.5. Furthermore, Monte Carlo event generators model the effects from soft and collinear splittings using the technique of *parton showers*, and is discussed in more detail in section 2.3.

2.1.4 Renormalization - The running of the strong coupling

As soon as higher order corrections, i.e. internal loops in self-energies or vertex corrections, are included in Feynman diagrams, the appearing ultra-violet divergences have to be renormalized[9]. This prescription introduces an additional, but arbitrary, energy scale μ_r, the *renormalization scale*. As a consequence, the coupling parameter $\alpha_s(\mu_r)$ and any physical quantity which is expressed as a series of the coupling parameter become functions of μ_r.

2.1.4.1 The renormalization group equation

In order to preserve the predictive power of the theory, any physical quantity $\mathcal{R}$, like e.g. inclusive cross sections or jet-production rates, should not depend on a particular arbitrary choice of the parameter μ_r. This postulate can be expressed mathematically by the renormalization group equation (RGE)

$$\mu_r^2 \frac{\mathrm{d}}{\mathrm{d}\mu_r^2}\mathcal{R}(Q^2/\mu_r^2, \alpha_s) = \mu_r^2 \frac{\partial \mathcal{R}}{\partial \mu_r^2} + \mu_r^2 \frac{\partial \alpha_s}{\partial \mu_r^2}\frac{\partial \mathcal{R}}{\partial \alpha_s} \stackrel{!}{=} 0 \ , \tag{2.6}$$

where $\mathcal{R}$ is a dimensionless quantity $\mathcal{R} = \mathcal{R}(Q^2/\mu_r^2, \alpha_s)$ depending on a single energy scale Q^2 and is calculated in the perturbative limit $n \to \infty$ in a power series of the renormalized coupling $\alpha_s = \alpha_s(\mu_r)$.

[6] This so-called mass singularity would be absent if either gluon or the quark had mass.

[7] Infact, the infrared divergence in the phase space integrals is canceled by loop integrals, for inclusive enough observables according to the Bloch-Nordsieck theorem [25] and Kinoshita-Lee-Nauenberg theorem [26, 27].

[8] In experiment, besides the limited resolution of any detector, a soft gluon-quark system cannot even be resolved principally as a consequence of confinement and the observed hadrons will only be detected as a 1-particle configuration. In order to resolve a 2-parton final state configuration, the gluon should have an energy E_g and opening angle θ_{qg} at least bigger than the detector resolution. However, in the calculation the unresolved gluon splitting will contribute to the sum of the amplitudes.

[9] The renormalizibility of a non-abelian gauge theory like QCD was demonstrated by 't Hooft and Veltman [28].

The RGE implies that any dependence of $\mathcal{R}$ on μ_r must be canceled. This behavior is enforced by an appropriate μ_r-dependence of the coupling. From equation 2.6 follows that the coupling has to satisfy the following differential equation[10]:

$$\mu_r^2 \frac{\mathrm{d}\alpha_s}{\mathrm{d}\mu_r^2} = \beta(\alpha_s) = -\beta_0 \alpha_s^2 - \beta_1 \alpha_s^3 - \beta_2 \alpha_s^4 - \dots \ . \tag{2.7}$$

Assuming $\alpha_s = \alpha_s(\mu_r)/(4\pi)$ the β_i coefficients are:

$$\begin{aligned} \beta_0 &= 11 - \frac{2}{3} n_f \ , \\ \beta_1 &= 102 - \frac{38}{3} n_f \ , \\ \beta_2 &= \frac{2857}{2} - \frac{5033}{18} n_f + \frac{325}{54} n_f^2 \ , \end{aligned}$$

where n_f is the number of flavors and the SU(3) color-factors have been applied[11] The 1-loop [15, 16] and 2-loop [30] β-coefficents, β_0 and β_1 respectively, are independent on the employed renormalization procedure, whereas the 3-loop coefficient β_2 [31, 32] is scheme dependent and is given here in the $\overline{\text{MS}}$-scheme. The β-coefficients in the $\overline{\text{MS}}$-scheme are currently known up to 4-loops [33]. The fact that the $(-\beta_i)$-coefficients are negative[12] is the origin of asymptotic freedom.

2.1.4.2 The running of the strong coupling

The energy dependence of the renormalized coupling $\alpha_s(\mu_r)$ can be precisely determined by integrating equation 2.7. An exact analytic solution exists only in 1-loop approximation by

$$\alpha_s(\mu_r) = \frac{1}{\beta_0 \ln(\mu_r^2/\Lambda^2)} \ , \tag{2.8}$$

where Λ is a constant of integration (the so-called *asymptotic scale parameter*), which can be interpreted as the non-perturbative scale of QCD where the perturbatively defined coupling would diverge.
One possible approximate analytic solution of the RGE up to 4-loop is given by [35]

$$\begin{aligned} \alpha_s(\mu_r) \simeq \frac{1}{\beta_0 t} \Big(1 - \frac{\beta_1}{\beta_0^2} + \frac{\beta_1^2(\ln^2 t - \ln t - 1) + \beta_0 \beta_2}{\beta_0^4 t^2} \\ - \frac{\beta_1^3(\ln^3 t - \frac{5}{2}\ln^2 t - 2\ln t + \frac{1}{2}) + 3\beta_0\beta_1\beta_2 \ln t - \frac{1}{2}\beta_0^2\beta_3}{\beta_0^6 t^3} \Big) \ , \end{aligned} \tag{2.9}$$

where $t = \ln(\mu_r^2/\Lambda^2)$. Using equation 2.9 the value of $\alpha_s(\mu_r)$ can be determined at any scale μ_r if either Λ or, alternatively, the value of $\alpha_s(\mu_0)$ at an arbitrary scale μ_0 is known.

[10] In some articles equation 2.7 is labeled 'renormalization group equation' instead of equation 2.6.

[11] The splitting factor for a gluon to split into a $q\bar{q}$-pair is set to the $T_F n_f = 1/2 n_f$, the gluon radiation color factor to $C_F = 4/3$ and the color-factor associated triple gluon vertex the number of colors to $C_A \equiv N_C = 3$. The experimental values for the color factors are $C_A = 2.89 \pm 0.21$ and $C_F = 1.30 \pm 0.09$ [29] and are in good agreement with the SU(3)-values.

[12] If the number of active flavors would be $n_f > 33/2$ there would be no asymptotic freedom and following quarks could not be regarded as free within a bound state, e.g. the proton. There are no experimental hints for more than six quark flavors up to date. Even under the assumption of a symmetry of the number of flavors in the quark and the lepton flavor, no hints for a fourth family are observed [34].

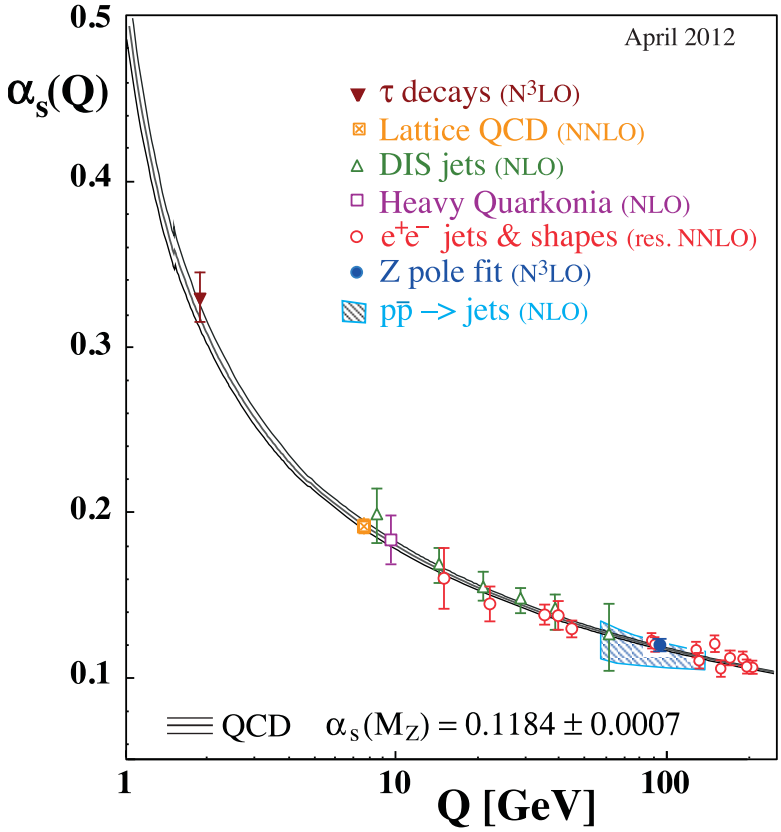


Figure 2.2: Illustration of the running coupling $\alpha_s(\mu_r)$ as predicted by QCD over two orders of magnitude as a function of the scale $Q \equiv \mu_r$ (taken from [12]). Values of $\alpha_s(\mu_r)$ for several classes of determinations of $\alpha_s(M_Z)$ are depicted at the scales of their determination. Besides the pre-averaged values (see section 2.2), also an averaged value of HERA inclusive jet-measurements at high-Q^2 [36] and from inclusive jets in $p\bar{p}$-collisions at the Tevatron collider are shown.

As a free parameter of the theory, this value has to be determined from experimental data. For convenience it has become common practice to quote the value of $\alpha_s(M_Z)$ at the well-known Z^0 boson mass as reference scale μ_0 and to determine the corresponding value of Λ by iteratively solving equation 2.9. The running coupling $\alpha_s(\mu_r)$ is shown in figure 2.2 together with experimental data at the scales of their determination.

A general analytic form of the renormalized coupling (i.e. independent of the renormalization scheme) as expression of $\alpha_s(\mu_0) = \alpha_s(M_Z)$ instead of Λ, in 1-loop approximation can be written as

$$\alpha_s(\mu_r, \alpha_s(M_Z)) = \frac{\alpha_s(M_Z)}{1 + \alpha_s(M_Z)\beta_0 \ln(\frac{\mu_r^2}{M_Z^2})} \,, \tag{2.10}$$

where $\alpha_s = \alpha_s/(4\pi)$. For a given value of $\alpha_s(M_Z)$ equation 2.10 determines the coupling strength at a vertex for any given momentum transfer. This dependence of the strong coupling which has now become the *effective coupling* strength as a function of the relevant scale at the vertex is often called the *running of the strong coupling*. Contributions to the running of the strong coupling beyond the 1-loop approximation are still significant (as equation 2.9 implies), but the impact of more higher orders beyond 3-loop approximation become very small at relevant scales.

2.1.4.3 The scale dependence of the perturbative series

The principal independence of any perturbatively calculated quantity σ on the renormalization scale μ_r is expressed by the RGE (equation 2.6). However, this behavior can only be enforced in infinite order, but as a consequence of the truncation of the finite series the quantity σ remains explicitly dependent on μ_r. This can be seen, when inserting the perturbative expansion of σ (equation 2.5) in the RGE (equation 2.6)

$$\begin{aligned} 0 \stackrel{!}{=} \mu_r^2 \frac{\partial k_0}{\partial \mu_r^2} + \alpha_s(\mu_r)\mu_r^2 \frac{\partial k_1}{\partial \mu_r^2} + \alpha_s^2(\mu_r)\left(\mu_r^2 \frac{\partial k_2}{\partial \mu_r^2} - k_1\beta_0\right) \\ + \alpha_s^3(\mu_r)\left(\mu_r^2 \frac{\partial k_3}{\partial \mu_r^2} - (k_1\beta_1 + 2k_2\beta_0)\right) + \mathcal{O}(\alpha_s^4) \ . \end{aligned} \tag{2.11}$$

In order to solve this equation, the coefficients of $\alpha_s^n(\mu_r)$ have to vanish separately for each order n. The perturbative coefficients k can thus be determined by integration to

$$\begin{aligned} k_0 &= \text{const.} \ , \\ k_1 &= \text{const.} \ , \\ k_2 &= k_2\left(\frac{Q^2}{\mu_r^2}\right) = k_2(1) - \beta_0 k_1 \ln\left(\frac{Q^2}{\mu_r^2}\right) \ , \\ k_3 &= k_3\left(\frac{Q^2}{\mu_r^2}\right) = k_3(1) - (\beta_1 k_1 + 2\beta_0 k_2) \ln\left(\frac{Q^2}{\mu_r^2}\right) \ . \end{aligned} \tag{2.12}$$

In infinite order the scale dependencies of $\alpha_s(\mu_r)$ and of the coefficients $k(Q^2/\mu_r^2)$ cancel and the quantity σ becomes invariant under the choice of μ_r. This cancellation is not granted for a finite truncated series and all predictions of pQCD include an explicit dependence on the renormalization scale. This dependence is called *(renormalization) scale-dependence*[13]. As the equations 2.12 imply, the scale dependence is distinct for LO predictions, since k_1 does not depend on μ_r and thus no cancellation of the scale dependence of $\alpha_s(\mu_r)$ can occur[14]. Regarding higher orders, the logarithmic energy dependence of the coupling $\alpha_s(\mu_r)$ is partly canceled by the scale dependence of the perturbative coefficients $k_i(\mu_r)$. The scale dependence of a perturbatively calculated quantity can be expressed by differentiating the finite series in equation 2.11 of order $n = N$ by μ_r [37]

$$\mu_r^2 \frac{\mathrm{d}}{\mathrm{d}\mu_r^2}\sigma_N = \mu_r^2 \frac{\mathrm{d}}{\mathrm{d}\mu_r^2} \sum_{i=0}^{N} k_i\left(\frac{Q^2}{\mu_r^2}\right) \alpha_s^i(\mu_r) \propto \mathcal{O}(\alpha_s^{N+1}(\mu_r)) \ . \tag{2.13}$$

With $\alpha_s(\mu_r) < 1$, in order that perturbative QCD is applicable, the calculated observable σ shows less dependence on the arbitrary choice of μ_r the more terms are included in the perturbative series and vanishes in the perturbative limit $N \to \infty$.
It can be seen from equations 2.11 and 2.12 that each (higher) order is always composed of a renormalization scale dependent and a scale independent part. In order to account for a convergence of the perturbative series, the scale independent part is supposed to decrease with increasing order. The scale dependent part, however, can be estimated

[13] In section 2.1.5 a second scale is introduced, the factorization scale. Throughout this thesis, 'scale dependence' refers to the dependence of any quantity on both the renormalization and factorization scale, the renormalization and the factorization scale.

[14] The LO contribution is calculated solely from tree-level diagrams where no virtual loops and thus no μ_r-dependence is present.

from preceding contributions. Hence, it has became standard to estimate the contribution from (missing) higher orders by a variation of the renormalization scale. As given by equation 2.13 such a variation is proportional to the relevant contribution of the next higher order. By convention, to estimate these missing higher orders the cross sections are calculated with the varied scale μ_r of

$$\mu_r = 2^c \mu_{r,0} \ , \tag{2.14}$$

where the *scale factor* c has to take values between

$$-1 \leq c \leq 1 \ . \tag{2.15}$$

This refers to a scale variation of a factor of two 'up' and 'down'. We assume here that the dependence of the cross section on the applied renormalization scheme is also been accounted for by the variation of the renormalization scale and no other renormalization schemes other than the $\overline{\text{MS}}$-scheme will be investigated.

2.1.4.4 The renormalization scale dependence of jet cross sections

In the following we study the dependence of jet cross sections on various scale choices of the form $\mu_r = 2^c \mu_{r,0}$, where $\mu_{r,0}$ is calculated from a simple formula of Q^2 and p_T and the scale factor c which is used to vary the scale. The study is performed for bins of the inclusive jet measurement and the trijet measurement. The phase space of these measurements is defined in chapter 4 and summarized in table A.1.
We will consider only scales $\mu_{r,0}$ which are composed of the virtuality Q^2 (c.f. equation 2.35) and of the transverse momentum of the individual jet p_T or the average jet transverse momentum $\langle p_\text{T} \rangle$ for inclusive jets or multijet events, respectively. As *nominal scale*, we choose the quadratic average

$$\tilde{\mu}_{r,0} = \sqrt{\frac{Q^2 + p_\text{T}^2}{2}} \ . \tag{2.16}$$

The ratio of cross sections $\sigma(\mu_r)$ calculated for various scale choices w.r.t. this nominal cross section $\sigma(\tilde{\mu}_{r,0})$ is shown as a function of the scale factor c in figure 2.4 for some selected bins of the inclusive jet measurement. The LO prediction follows the scale-dependence of the effective coupling $\alpha_s(\mu_r)$, since the perturbative coefficients in LO are scale independent due to the absence of loop-corrections. The NLO correction is proportional to $\alpha_s(\mu_r)^2 \ln(Q^2/\mu_r^2)$ and partly compensates the scale dependence arising from $\sigma_\text{LO} \propto \alpha_s(\mu_r)k$ and the NLO cross section is therefore less scale dependent (as predicted by equation 2.13). The shape of the scale dependence at NLO follows always the same shape: for small values of the scale factor c, the cross section increases up to some maximum, and then decreases again. At very small scale factors c the cancellation between the $\alpha_s(\mu_r)$ and k_2 becomes worse, and the cross section decreases rapidly due to the $\ln(\mu_r^2)$ dependence. Such a renormalization scale can be considered as very un-natural.
The scale dependence shows two distinct values which could be chosen as scales according to basic principles:

- The principle of minimum sensitivity (PMS) [38]. The scale (factor) is chosen where $\partial\sigma/\partial\mu_r = 0$, to account for the minimal dependence of the cross section on the 'unphysical' parameter. As it can be seen in figure 2.3 this can result in a very large cross section prediction at very small scales, e.g. for the lowest Q^2 and p_T bin of the inclusive jet cross section, and to values, which are distant from a 'natural' scale.

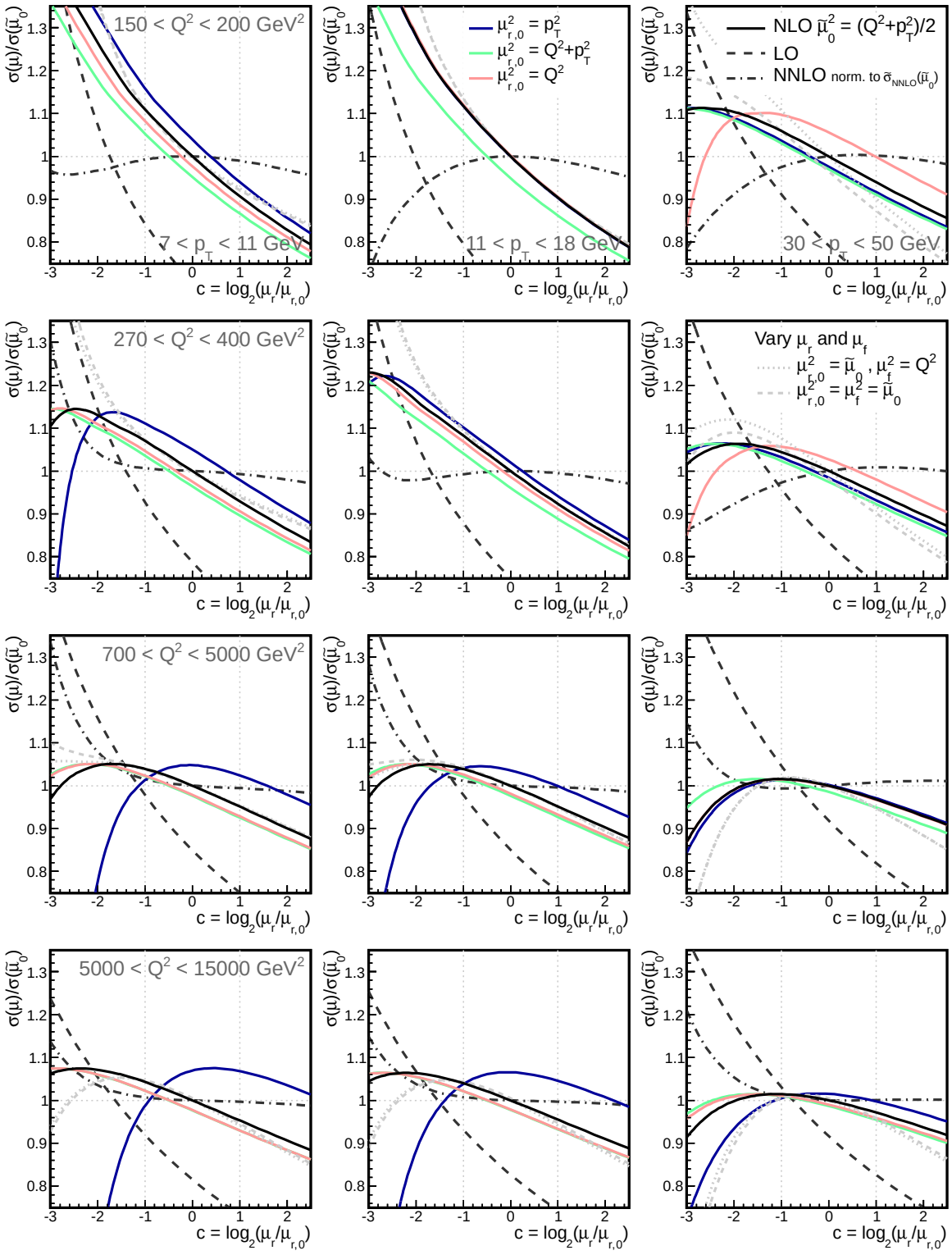


Figure 2.3: Dependence of selected double-differential inclusive jet cross section bins on a variation of the renormalization scale for various choices of the renormalization scale. The cross sections ratio is always built w.r.t. the cross section calculated with $\mu_{r,0} = 1 \cdot \sqrt{(Q^2 + p_T^2)/2}$. The factorization scale is always set to $\mu_f^2 = Q^2$. The renormalization scale is typically varied by a factor of 2, i.e. for $\mu_r = 2^c \mu_{r,0}$ with $-1 < c < 1$, in order to determine an estimate for missing higher orders (called *theory uncertainty*). The gray curves show the relative cross section, when the factorization and the renormalization scales are varied together by the same factor. As comparison, also the LO prediction is shown, where the perturbative coefficients are scale independent and the scale dependence only arises from the running of the coupling constant $\alpha_s(\mu_r)$. The dashed-dotted line shows an estimated NNLO cross section prediction. The scale dependence decreases with increasing order perturbation theory.

- The principle of fastest apparent convergence (FAC) or also called the method of effective charges (MEC) [39], where an observable is regarded similar to a the coupling strength. The knowledge of β-parameters up to k-th order enables to estimate the observable up to order α_s^{k+1} in $(k+1)$-loop approximation. This can be re-interpreted as the scale-factor, where the LO cross section is equal to the NLO order cross section. This however, can result in very small scales and very large cross section predictions.

There are several more methods, which propose to estimate the scale from basic principles, like the BLM procedure [40] and several extensions (like se-BLM or x-BLM) or the principal of maximum conformality [41]. A good overview about common methods is e.g. given in [42] and a recent proposal including some extensive discussion is given in [43]. However, there is no common agreement on how to optimize the choice of the scale. Therefore, we follow the classical approach to identify the scale with typical observables of each process.

For comparison the approximate scale dependence of NNLO cross sections is also shown in figure 2.3. The NNLO perturbative coefficients $k_{3,a}^{\mathrm{NNLO}}$ for each PDF flavor a is calculated according to equation 2.12 from the LO and the NLO coefficients. There are however no full NNLO calculations available since the scale independent coefficients $k_3(1)$ are still unknown up to date[15]. This calculation is normalized to its nominal prediction $\sigma_{\mathrm{NNLO}}(2^c\tilde{\mu_r})/\sigma_{\mathrm{NNLO}}(\tilde{\mu_r})$ in figure 2.3, in order that the unknown contributions cancel (the scale dependence through of $\alpha_s^3(\mu_r)k_3$ is therefore being neglected). The benefit of NNLO can be clearly seen by the reduced scale dependence of the cross sections within some reasonable variation of the renormalization scale[16].

The scale dependence of the dijet cross section shows a very similar behavior for the several scale choices to the inclusive jet cross sections, when the transverse momentum p_{T} is replaced by the average transverse momentum $\langle p_{\mathrm{T}} \rangle$.

The scale dependence of the trijet cross section is shown in figure 2.4 for selected bins of the double-differential measurement[17]. Compared to the inclusive-jet cross sections, the scale dependence is much more distinct for the LO calculation, since this is already proportional to $\alpha_s^2(\mu_r)$. The NLO calculation shows the typical shape, where the PMS value is mostly at $c < 0$. For a choice of $\mu_r = \langle p_{\mathrm{T}} \rangle$ the calculation in some bins is entirely in the regime where the cancellation of the renormalized coupling and the perturbative coefficients is worse. For that reason, we do not consider such a scale as a reasonable choice.

2.1.4.5 The choice of the renormalization scale

The renormalization scale has to be chosen when predicting cross sections in finite order calculations. It contributes as a parameter in the effective coupling $\alpha_s(\mu_r)$, and also in the perturbative coefficients $k(\mu_r)$ of higher order corrections with terms of $\ln(Q_0^2/\mu_r^2)$, and only in the perturbative limit these two dependencies cancel.

The renormalization scale can be interpreted as a cut-off of ultraviolet divergences and is therefore reasonably be identified with some *natural scale* of that process, especially in the

[15] Only very recently, the first calculations of 2-loop amplitudes for exclusive quantities, like inclusive jet production, in pure gluonic interactions were demonstrated [44], which provides the first step towards full NNLO calculations and demonstrates the principal possibility to arrange the appropriate cancellations of the singularities in NNLO.

[16] This is only a reasonable approximation if $k_3(1)$ is small and the term $\alpha_s(\mu_r)^3 k_3(1)/\sigma_0$ can still be neglected in the displayed ratio.

[17] The phase space of the trijet measurement is summarized in table A.1.

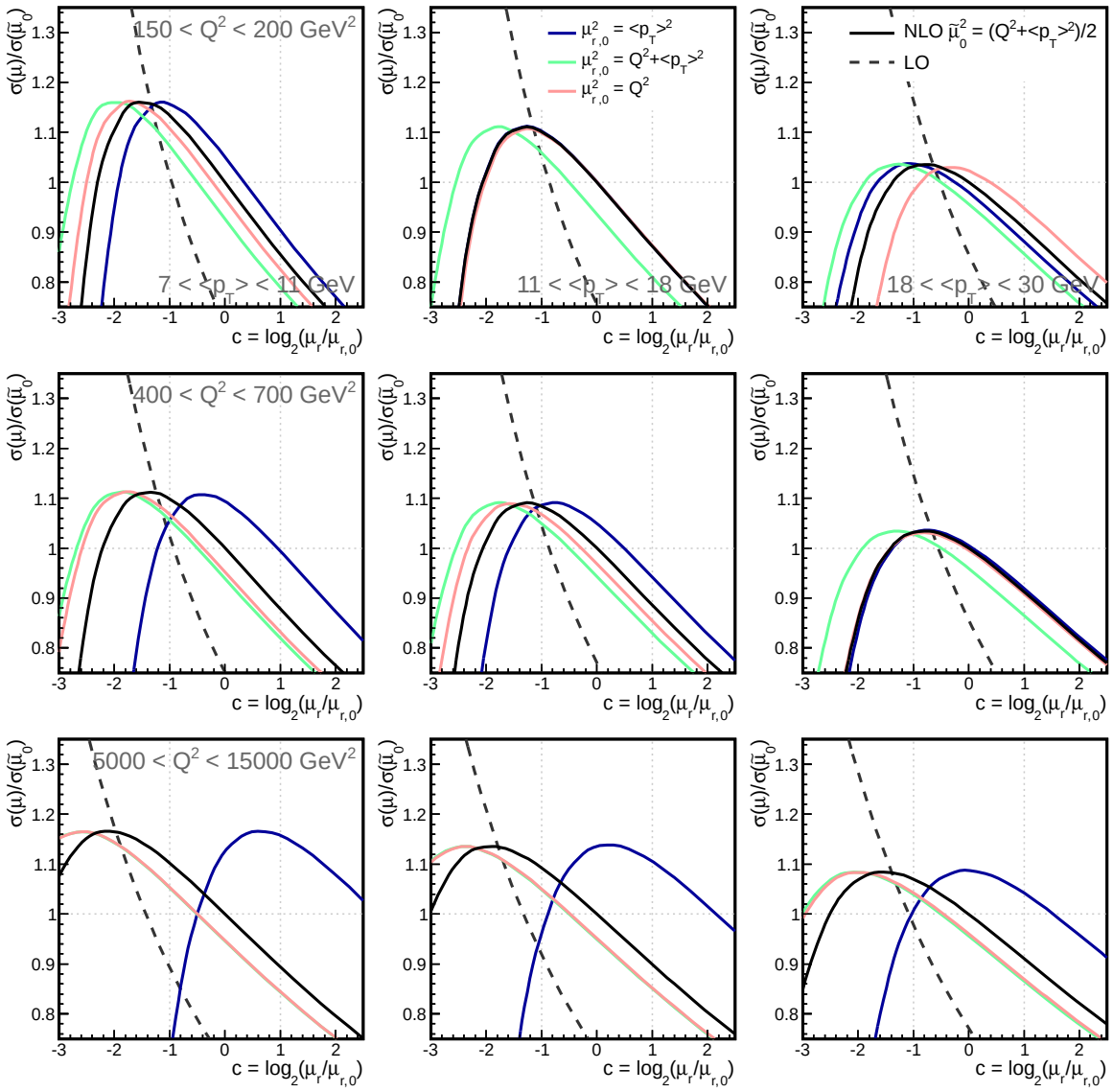


Figure 2.4: Illustration of the renormalization scale dependence in selected bins of the double-differential trijet measurement for various choices of μ_r. The LO cross section is already of $\mathcal{O}(\alpha_s^2)$ and hence the scale dependence is more distinct than for inclusive jet cross sections. This will increase the sensitivity to $\alpha_s(M_Z)$ of the measurement, but on the other hand also increases the theoretical uncertainty.

calculation of $k(\mu_r)$[18]. In DIS the scale of the process would naturally be identified with Q^2. On the other hand, it would be also be natural to identify the renormalization scale with the specific scale at each strong vertex in order to evaluate the coupling strength $\alpha_s(\mu_r)$ at an appropriate value and it becomes reasonable to select μ_r proportional to $p_{\mathrm{T}}^{\mathrm{jet}}$. As a compromise it has became standard to calculate the renormalization scale from a composed formula involving Q^2 and p_{T} and we select the average of the quadratic values as the nominal renormalization scale[19]

$$\mu_r = \sqrt{\frac{Q^2 + p_{\mathrm{T}}^2}{2}} \,, \tag{2.17}$$

As studied in figures 2.3 and 2.4, this choice of the scale also features a reasonable behavior under scale variations. Any resulting difference on the cross section from a different scale choice is accounted for through the scale variation by a factor of 2.

Besides those two obvious scales Q^2 and p_{T}, also other observables could be input for the renormalization scale. For example, the quantity [45]

$$k_{\mathrm{T,i}} := E_i \sqrt{2(1 - \cos\theta_{i\mathrm{P}})} \,, \tag{2.18}$$

where for a parton or jet i the angle in the Breit frame[20] towards the incident proton $\theta_{i\mathrm{P}}$ becomes relevant. The proposed scale would be $\mu_{\mathrm{MZ}} := \mu_r = \sum_i k_{\mathrm{T,i}}$ for dijet cross sections in DIS. It always holds that $\mu_{\mathrm{MZ}}^2 > Q^2$, but also $\mu_{\mathrm{MZ}} > \sum_i p_{\mathrm{T,i}}$, and thus this scale results in quite large values of μ_r. The advantage of the quantity μ_{MZ} would be a smooth interpolation between the expected natural scales for the two asymptotic limits. It approaches Q in the parton limit, i.e. where the p_{T} vanishes, and it corresponds to twice[21] the jet transverse momentum in the photoproduction limit [45]. For other processes like e.g. pp and p$\bar{\mathrm{p}}$ collisions, where only jet quantities like p_{T} are accessible, also scales involving the separation of the two jets in a dijet system were proposed [46].

2.1.5 Factorization and parton density functions

Perturbative methods are only applicable if the power series is converging sufficiently fast, i.e. $\alpha_s(\mu_r) \ll 1$. However, the effective coupling becomes large if the renormalization scale reaches the asymptotic scale $\mu_r \to \Lambda_{\mathrm{QCD}} \approx 200\,\mathrm{MeV}$ and perturbative QCD breaks down. For instance, at these small scales (corresponding to long distances) the color-recombination and formation of hadrons take place.
The concept of *factorization* separates the short-distance interactions (hard processes of the partons) from long-distance interactions at an additional arbitrary scale, the *factorization scale* μ_f. The calculation of processes which involve hadrons, i.e. also in DIS, can be split (factorized) into the hard process, which can be calculated by pQCD, and the soft process, i.e. the distribution of partons inside the hadrons.

[18] The perturbative coefficients $k(\mu_r)$ are evaluated at a fixed scale Q_0^2, which is chosen to be Q^2 in the calculation and hence, the scale dependence formally cancels through $\log(Q_0^2/\mu_r^2)$. However, since the scale Q_0^2 was chosen freely, the problem persists and is only hidden in $k(\mu_r)$.

[19] Throughout this thesis, p_{T} denotes $p_{\mathrm{T}}^{\mathrm{jet}}$ in case of the inclusive jet measurement and $\langle p_{\mathrm{T}} \rangle_2$ and $\langle p_{\mathrm{T}} \rangle_3$ in case of the dijet and trijet measurement, respectively. See chapter 4 for the precise definition of the observables.

[20] See section 2.6 for the definition of the Breit frame.

[21] In [45] it is stated that the scale μ_{MZ} corresponds to the jet transverse momentum when the photon virtuality becomes negligible. However, from the formula given in the article for the LO approximation it follows, that the scale μ_{MZ} approaches twice the transverse momentum of the dijet system.

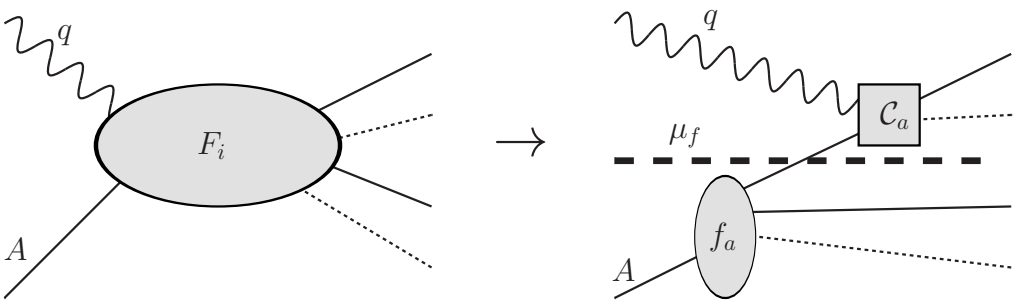


Figure 2.5: Graphical illustration of the factorization theorem in DIS for one incoming hadron A. The factorization scale sub-divides at a certain factorization scale μ_f the cross section calculation into two parts: a hard contribution which is calculable in pQCD and a soft contribution which has to be parametrized. The hard contribution is process dependent, while the soft contribution is universal for each hadron in any high energy process. As example, an hadronic structure function F_i (which could be a jet-cross section as well) becomes a convolution of the parton density function f_a and the perturbative coefficients $\mathcal{C}_a$ (for the definition of this convolution see equation 2.19) (illustration adapted from [48]).

2.1.5.1 The factorization theorem

By introducing a separation criterion between long and short distance effects, any observable involving initial state partons can be expressed by a convolution of a hard partonic cross section and the *parton distribution functions* (PDF). Let $W^{\mu\nu}$ being the hadronic tensor which enters the feynman calculus, this *factorization theorem* can be expressed in the Bjorken limit where the momentum transfer Q^2 is large and x is fixed by [47]

$$W(q,p) = f_a \otimes \mathcal{C}_a = \sum_a \int_x^1 \frac{d\xi}{\xi} f_a(\xi,\mu_f)\mathcal{C}_a(q,\xi p,\mu_r,\mu_f,\alpha_s(\mu_r)) \ , \tag{2.19}$$

where f_a is the parton density function of a parton of type a, and $\mathcal{C}_a$ is the hard scattering coefficient of the process. The quantity $f_a d\xi$ can be interpreted as the probability to find a parton of type a, e.g. a gluon or any quark, carrying the hadron's longitudinal momentum fraction ξ to $\xi + \mathrm{d}\xi$, at an effective scale μ_f. A graphical representation of equation 2.19 is shown in figure 2.5. The factorization theorem ensures that the dependence of $W^{\mu\nu}$ on large distance effects below the factorization scale μ_f is entirely taken into account by the PDFs f_a, and the calculation can be separated into two parts: a hard process which can be calculated perturbatively, and a soft process, which is parametrized by the PDFs and which have to be determined experimentally. In this procedure, μ_f defines the borderline between the short distance and long distance effects, while the employed factorization scheme defines the method of reshuffling the finite pieces[22].
The appearance of a factorization scale μ_f is not an intrinsic feature of the theory but an artifact of the limited power of the calculation. Hence, any physical observable does not depend on the arbitrary choice of μ_f which is true for pQCD only in the perturbative limit[23]. In a finite series, a variation of the scale parameter cannot be fully compensated, and as a consequence, the calculated observable obtains some *(factorization) scale dependence*, which is of the order $n+1$.

[22]The parton density functions and the perturbative coefficients must always be determined consistently using the identical factorization scheme. The most commonly applied factorization scheme is the modified minimal subtraction scheme ($\overline{\mathrm{MS}}$) which is also used for all calculations in this thesis.

[23]In the perturbative limit, a shift of μ_f merely results in a reshuffling between the theoretical factors $f_a(\mu_f)$ and $\mathcal{C}_a(\mu_f)$, but the integral in equation 2.19 remains invariant [48].

2.1.5.2 The QCD evolution equations

As a consequence of the factorization theorem, it becomes necessary to determine the parton distribution function of the involved hadrons $f_a(x, \mu_f)$. The dependence of the PDF on μ_f can be written as

$$\mu_f \frac{\mathrm{d}}{\mathrm{d}\mu_f} f_a(x, \mu_f) = \sum_b \int_x^1 \frac{\mathrm{d}\xi}{\xi} \mathcal{P}_{ab}(\frac{x}{\xi}, \alpha_s(\mu_f)) f_b(\xi, \mu_f) \ , \tag{2.20}$$

where f_a is a parton distribution function of flavor a which can be either a quark or a gluon, and $\mathcal{P}_{ab}$ are the *splitting functions*[24] (or also known as the Altarelli-Parisi kernels). These so-called *QCD evolution equations* are in LO also called Dokshitzer-Gribov-Lipatov-Altarelli-Parisi equations [49–51] (DGLAP) and have different structures between the gluon and the quark PDFs. Using the notation of a Mellin convolution [25] they read for the quark densities $q_i(x, \mu_f)$ of flavor[26] i

$$\mu_f \frac{\mathrm{d}}{\mathrm{d}\mu_f} q_i(x, \mu_f) = [q_i \otimes \mathcal{P}_{qq}] + [g \otimes \mathcal{P}_{qg}] \tag{2.22}$$

and for the gluon density $g(x, \mu_f)$

$$\mu_f \frac{\mathrm{d}}{\mathrm{d}\mu_f} g(x, \mu_f) = [\sum_i^{n_f} q_i \otimes \mathcal{P}_{gq}] + [g \otimes \mathcal{P}_{gg}] \ , \tag{2.23}$$

where the sum i has to be performed over all active quark and anti-quark flavors n_f. In order to emphasize the fact that the gluon and quark densities are coupled due to the splitting functions the DGLAP equations are often written in a closed matrix notation by

$$\mu_f \frac{\mathrm{d}}{\mathrm{d}\mu_f} \begin{pmatrix} \boldsymbol{q}(x, \mu_f) \\ g(x, \mu_f) \end{pmatrix} = \sum_j^{2n_f} \begin{pmatrix} \boldsymbol{\mathcal{P}}_{qq_j} & \boldsymbol{\mathcal{P}}_{qg} \\ \mathcal{P}_{gq_j} & \mathcal{P}_{gg} \end{pmatrix} \otimes \begin{pmatrix} \boldsymbol{q} \\ g \end{pmatrix} \ , \tag{2.24}$$

where $\mathbf{q}$ is a $2n_f$ dimensional vector of the quark and anti-quark densities q_i and all flavors have to be respected in the general form of the spitting functions $\mathcal{P}_{q_i q_j}$.
The splitting functions $\mathcal{P}_{ab}$ can be interpreted (in LO) as the probability to emit a parton of flavor a with momentum fraction x from an incident parton of flavor b with momentum fraction ξ. The splitting functions can be expanded in powers of the coupling α_s at the scale[27] μ_f by

$$\mathcal{P}_{ab}(\frac{x}{\xi}, \alpha_s(\mu_f)) \simeq \frac{\alpha_s(\mu_f)}{4\pi} \mathcal{P}_{ab}^{(0)}(\frac{x}{\xi}) + \left(\frac{\alpha_s(\mu_f)}{4\pi}\right)^2 \mathcal{P}_{ab}^{(1)}(\frac{x}{\xi}) + \dots \ . \tag{2.25}$$

[24] The k_t ordered splitting functions refer to vertices of the four processes of 1) an incident quark radiating a gluon $\mathcal{P}_{qg}$, 2) an incident quark suffering a soft gluon radiation $\mathcal{P}_{qq}$, 3) an incident gluon splitting into a quark-pair $\mathcal{P}_{gq}$ and 4) an incident gluon radiating a gluon $\mathcal{P}_{gg}$.

[25] The Mellin convolution in the momentum variable is defined as

$$[f \otimes \mathcal{P}] = [\mathcal{P} \otimes f] := \int_x^1 \frac{\mathrm{d}\xi}{\xi} f(\xi, \mu_f) \mathcal{P}(\frac{x}{\xi}) \tag{2.21}$$

[26] Altough pure QCD does not distinguish flavors (besides different masses and the number of 'active' flavors), this separation becomes relevant for other SM processes when determining PDFs.

[27] Notable is that the factorization scale μ_f takes the role of the renormalization scale to evaluate the renormalized effective coupling $\alpha_s(\mu_f)$ since it is the only applicable scale.

and can be directly derived from QCD at any order, which makes them an intrinsic part of the theory [50]. The splitting functions were first calculated to NLO in α_s [52, 53] and are currently known up to NNLO [54, 55], which are mandatory ingredients for precise calculations at hadron colliders.
An important feature of the DGLAP evolutions is that the lower limit of the convolution integral is not zero but equal to x. This allows for definite theoretical predictions without the knowledge of the parton densities at x values lower than in the process under consideration.

2.1.5.3 Parton density functions of the proton

The DGLAP equations provide a prescription of the evolution of the parton density functions from a starting scale Q_0 to any arbitrary larger scales μ_f. However, for the determination of PDFs at the starting scale Q_0 and at each x-value perturbative methods are not applicable because of IR singularities[28], and the PDFs have to be obtained from experimental data.
In order to have universal PDFs, i.e. applicable for all hadron induced standard model processes, it is standard to distinguish 13 different PDFs for the twelve quark flavors and the gluon ($\bar{t}, \bar{b}, \bar{c}, \bar{s}, \bar{d}, \bar{u}, g, u, d, s, c, b$ and t)[29]. The top PDFs t and $\bar{t}$ are set equal to zero, because of the large rest-mass of the top quark. The heavy flavor PDFs $\bar{b}, b, \bar{c}$ and c have non-zero density at scales $\mu_f \gg m_b$. The treatment of heavy flavors for scales $\mu_f \sim m_{\text{heavy}}$ is complicated and is currently under active discussion [56].
The PDFs are determined in (so-called 'global') fits to a large number of experimental datasets which have sensitivity to the different quark flavors. In these fits the parton distributions are parametrized by reasonable flexible functions of x at a fixed starting scale Q_0 and are evolved by applying the evolution equations 2.24 to higher scales. The entirety of all PDFs have to fulfill *sum-rules* (see e.g. [57] and references therein) and the free parameters of the parameterizations are determined in an iterative χ^2-minimization[30].
Sets of parton density functions are currently available from various groups, which mostly employ similar fitting-techniques and have to rely on the same measurements. Recent PDF sets are e.g. released by the CTEQ-TEA collaboration [59], the MSTW group [60, 61] and the ABM group [62] which provide PDF sets for LO, NLO and NNLO fits and also have published simultaneous PDF and $\alpha_s(M_Z)$ determinations. The HERAPDF group constrain their input data mostly to HERA data sets [58], i.e. the inclusion of structure function measurements, but shows a promising effort for opening their fitting framework to data from other experiments for global fits[31]. The NNPDF group applies a neural network

[28] Also lattice QCD methods are not yet applicable due to limitations in the present calculations.

[29] The color-degree of freedom of the quarks and the gluon is integrated, because the QCD coupling is equal to all colors and the quantum number color cannot be resolved experimentally, but the flavor (and electric charges) can be distinguished in experiments.

[30] The most valuable input to these fits come from DIS-structure function measurements [58]. These measurements can determine the gluon-density from scaling violations, since this process corresponds to an inclusive way to measure radiation. Also the direct measurement of F_L is directly connected to the gluon density. In order to have higher constraining power to the gluon density, further measurements like jet-measurements, which have direct sensitivity to the gluon density, can be included in PDF fits. However, since the gluon-density g has to be determined through an expression of the form $\sigma \propto g \cdot \alpha_s(\mu_r)$ (which involves a strong correlation of the gluon and the strong coupling constant, i.e. a harder gluon-density can always be compensate by a smaller strong coupling constant $\alpha_s(M_Z)$) the simultaneous determination of PDFs and $\alpha_s(M_Z)$ is most promising, if data with additional constraints on the strong coupling itself are included into PDF fits.

[31] See the HERAFitter website for more details: `http://www.herafitter.org`

approach to determine PDFs with unbiased paramterization assumptions, i.e. by applying 256 free parameters to the 7 independent PDFs [63–65].

2.1.5.4 The choice of the factorization scale

The factorization scale is an artificial parameter of the theory and has to be chosen when calculating cross sections. In a finite perturbative calculation the observable retains a *factorizatization scale dependence* of $\mathcal{O}(\alpha_s^{N+1})$ since the hard process cannot compensate the change of μ_f due to the missing of higher orders. For its choice we address two dependencies of μ_f in the cross section calculation.
The factorization method accounts for the elimination of soft divergences and it becomes therefore necessary to choose a reasonably large factorization scale which is safely above these singularities, i.e. larger than the mass of the incident hadrons $\mu_f \gg m_P$. Such a scale is also very well above $\mu_f \gg \Lambda_{\mathrm{QCD}}$, in order that perturbative methods are applicable. The typical scale of the hard process is therefore considered to be a reasonable choice for the factorization scale. In DIS, such a scale would naturally be identified with the virtuality Q^2, or p_{T} in case of jet-production[32].
A more mathematically motivated constraint on the choice of μ_f is suggested by the QCD evolution equations, which have to be applied to determine the PDFs $f_a(x, \mu_f)$ in the full (x, μ_f)-space for all parton flavors from the limited data which is available. For the solution of the DGLAP equations the splitting functions 2.25 are employed, which are expressed as functions of $\alpha_s(\mu_f)$. Whenever, the factorization scale is different from the renormalization scale, these functions have to be rewritten explicitly in terms of $\alpha_s(\mu_r)$. A Taylor expansion on a logarithmic scale around μ_r yields

$$\alpha_s(\mu_f) = \alpha_s(\mu_r) + \alpha_s'(\mu_r) L_R + \frac{1}{2}\alpha_s''(\mu_r) L_R^2 + \dots \ , \tag{2.26}$$

where $L_R = \ln(\mu_f^2/\mu_r^2)$ and the derivatives w.r.t. μ_r can be calculated using equation 2.7. This results in a replacement of the $\alpha_s(\mu_f)$-expressions in the splitting functions (equation 2.25) by [66]

$$\begin{aligned}
\alpha_s(\mu_f) =& \alpha_s(\mu_r) - \beta_0 L_R \alpha_s^2(\mu_r) - (\beta_1 L_R - \beta_2^2 L_R^2)\alpha_s^3(\mu_r) + \mathcal{O}(\alpha_s^4) \\
\alpha_s^2(\mu_f) =& \alpha_s^2(\mu_r) - 2\beta_0 L_R \alpha_s^3(\mu_r) + \mathcal{O}(\alpha_s^4) \\
\alpha_s^3(\mu_f) =& \alpha_s^3(\mu_r) + \mathcal{O}(\alpha_s^4) \ .
\end{aligned} \tag{2.27}$$

As a consequence, the factorization scale is strongly connected to the renormalization scale through the logarithms L_R, although both scales are principally independent. In order to avoid that these logarithms become too large, it would be natural to choose $\mu_r = \mu_f$.
Henceforth, we will apply a factorization scale of

$$\mu_f^2 = Q^2, \tag{2.28}$$

since the argument of consistent scales for various processes seems to be a stronger argument than the mathematically motivated one. Furthermore, it seems to be reasonable to choose a consistent scale for various processes. For inclusive DIS[33], the only applicable

[32] In contradiction to the considerations we performed for the choice of the renormalization scale, the specific scale of each vertex is not supposed to be relevant.

[33] In hadron-hadron collisions the only applicable scale would be proportional to the jet transverse momenta.

scale is by default Q^2 and is usually chosen [12] for such processes, which are also the main ingredients for the determination of the PDFs.
Nevertheless, the contribution from missing higher orders will be estimated by a variation of the factorization scale, which also accounts for different choices of μ_f, and will be considered as an uncertainty on the results.

2.1.5.5 The scale dependence of the inclusive jet cross section

The factorization scale dependence as well as the renormalization scale dependence of the inclusive jet cross section is shown in figure 2.6 for some selected bins[34], where each color defines a change of the cross section of 5% w.r.t. the nominal cross section. Following the arguments given in the previous section the nominal factorization scale μ_f is set to $\mu_f^2 = Q^2$, while the renormalization scale μ_r is set to $\mu_r^2 = (Q^2 + p_{\mathrm{T}}^2)/2$. To study the relation between the renormalization and the factorization scale, we select statistically median events in Q^2 and p_{T} in each bin and refer to it as *median event*. For these events, the scale factor which connects the two scales is depicted as the yellow line for each bin. The white lines indicate the scale factors for $|\log_2(\mu_r/\mu_f)| = 1$. Relative cross section predictions for several choices of $\mu_r = \mu_f$ are shown in color coding. These values are displayed at scale factors which have to be applied to the nominal scale choice for a median event. It is observed that although the factorization scale is different from the renormalization, their absolute logarithmic ratio $|\log_2(\mu_r/\mu_f)|$ is always smaller than 2 for median events in every bin. In most regions of the phase space the factorization scale dependence is smaller compared to the renormalization scale dependence. The factorization scale dependence is only relevant for the phase space region of large values of $p_{\mathrm{T}}^{\mathrm{jet}}$, where the cross section decreases up to 5% if μ_f is decreased by a factor of 2 (i.e. $\log_2(\mu_f/\mu_{f,0}) = -1$).

2.2 The world average value of the strong coupling constant $\alpha_s(M_Z)$

The only free parameters in the QCD Lagrangian are the quark masses and the coupling constant. If the quark masses are fixed, the only remaining free parameter of this comprehensive theory is the effective coupling which is typically quoted at the scale of the Z^0-mass (see section 2.1.4.2). The value of $\alpha_s(M_Z)$ can be determined from many experimental observables.
A recent world average value from many experimental data is determined by S. Bethke and presented in [12] and [67]. The pre-averaged input values are shown in figure 2.7. The input is constrained to data, where at least NNLO calculations are available. This unfortunately excludes jet-data from DIS as well as from hadron-hadron collisions. Pre-averaged values of $\alpha_s(M_Z)$ are determined from four classes of processes. The most stringent input values come from various lattice calculations. A further class of $\alpha_s(M_Z)$ determination are from hadronic τ-decay widths based on N^{3}LO calculations and τ-lifetime measurements. Another class is from e^+e^--annihilation data, where event shape observables and jet-rates are measured. These are mostly dominated by theoretical uncertainties, e.g. arising from hadronization corrections uncertainties. Further determinations are from DIS structure functions (PDF fits), where the very high accuracy of HERA data plays an important role and especially the higher order contributions to the NC and CC inclusive DIS cross sections

[34] See chapter 7 for the precise phase space definition of the inclusive jet cross section bins.

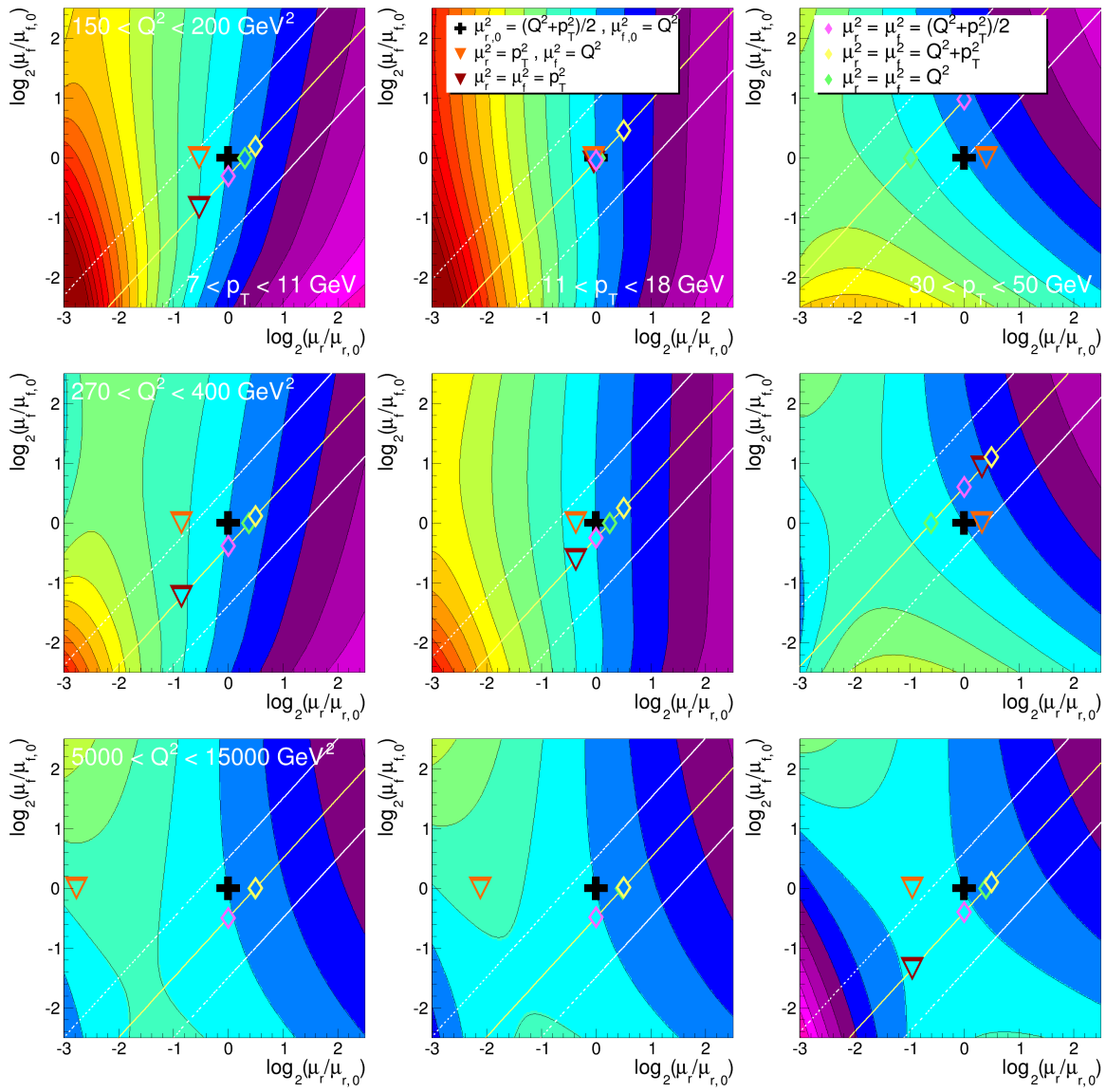


Figure 2.6: Relative dependence of the inclusive jet cross section on the variation of the renormalization and the factorization scale. The color coding illustrates the relative cross section w.r.t. to the cross section calculated for the scales $\mu_{r,0}^2 = (Q^2 + p_{\mathrm{T}}^2)/2$ and $\mu_{f,0}^2 = Q^2$. The cross section decreases towards blueish colors and increases for reddish colors and the black equipotential lines indicate each a change of 5% of the cross section.
For a median event of a bin, the yellow line indicates the scale factors, where the renormalization scale is equal to the factorization scale and the logarithms $\log(\mu_f/\mu_r)$ in the perturbative calculation vanish. The white lines shows scale factors where $|\log(\mu_f/\mu_r)| = 1$. The colored diamonds and triangles indicate scale factors for a median event for various other choices of the scales. The filled color is the relative cross section predicted by such a scale choice w.r.t. to the standard choice, which is displayed with a black cross.

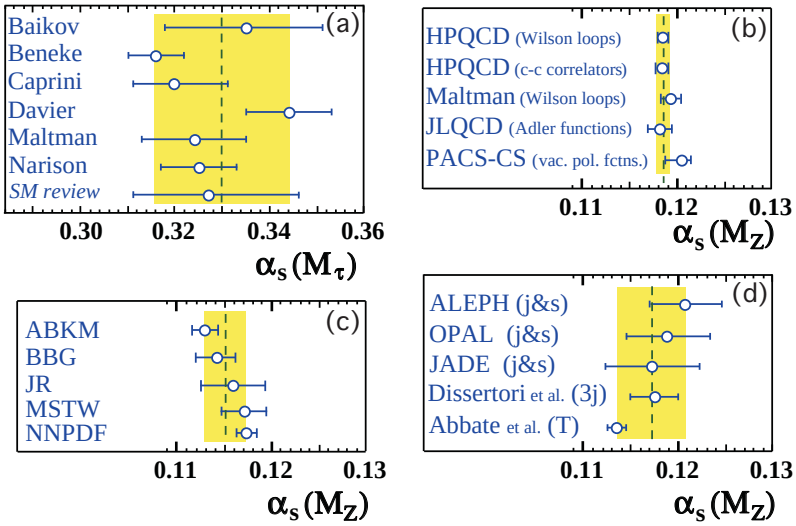


Figure 2.7: The summary of α_s-determinations for a) hadronic τ-decays, b) lattice calculations, c) DIS structure functions and d) e^+e^--annihilation data. All individual determinations are at least based on NNLO calculations. The yellow band indicates the assigned error to each sub-class of processes as explained in the full report [12].

give stringent constraints on the strong coupling. The world average value is determined from four pre-averaged values to be

$$\alpha_s(M_Z) = 0.1184 \pm 0.0007, \tag{2.29}$$

where the value and its uncertainty is mostly dominated by lattice QCD calculations.

2.3 Hadronization and parton shower

Perturbatively calculated cross sections cannot be compared directly to data, because two additional soft QCD effects have to be considered. Although these phenomena are partly integrated by the applied jet-algorithm (c.f. section 2.4), they must be accounted for in data-theory comparisons.

2.3.1 Multi-parton production

The pQCD calculations fail to predict multi-particle final states, as observed in the experiment, since at most $(n+1)$-leg matrix elements are considered in the NLO calculation. The creation of further partons can be estimated in event generators (see section 3.5.1) by attaching the *parton shower* prescription to LO matrix elements or by applying the *(color) dipole showering*. In principle, parton showers can be attached to any fixed order calculation, and predictions with NLO matrix elements are already available for many processes [68]. These are for instance also jet-production in hadron-hadron collisions [69], but no such predictions for jet-production in DIS are available up to date.

2.3.1.1 Parton showers

Starting from the final state partons of the LO matrix elements, each parton undergoes subsequent emission of additional partons. This showering is similar to QED radiation effects, but the radiated gluons themselves carry color charge and can therefore emit further

partons. These splittings factorize if they are approximated in the collinear or the soft gluon emission limit. The leading-order splitting functions are identical to the DGLAP splitting functions. The splitting or not-splitting probability is subsequently applied to any parton, by using Monte Carlo values for the splitting variables.
Hence, a parton shower develops starting from each (colored) parton of the hard process. The splitting cannot take place with all partons being on their mass-shell. Starting with an off-shell parton from the hard scattering, the splittings are strongly ordered in the *evolution variable*, e.g. the virtuality, and the partons become increasingly on mass-shell if they undergo splittings. The shower starts with an upper limit on the evolution variable which is derived from the hard process, and the shower terminates when the evolution variable has fallen below some hadronization scale Q_0. In an event generator, at this scale the final configuration of all partons momenta can be passed to a hadronization model. A subsequent momentum shuffling to the showering ensures momentum conservation. The parton shower method is basically an approximation to the higher-order correction to the hard process and could be principally attached to any fixed order calculations. If attached to matrix elements beyond LO, e.g. used by NLO generators or generators based on multi-let tree level diagrams, it must be taken care that no double-counting of contributions are included.

2.3.1.2 Color dipole model

An alternative approach to simulate multi-parton final states is the so-called *color dipole model*. In this approach, a gluon splitting is described by a dipole radiation pattern of a pair of partons. In addition the process $g \to q\bar{q}$ is included to account for additional quarks. In this model, each quark or antiquark is connected to a color partner, and each gluon is connected to two color partners. A pair of color partners then form a dipole, which becomes two dipoles after the gluon radiation. The splitting cascade continues until a stopping criterion is reached, which can be a minimum transverse momentum or the dipole mass reaches some minimum scale. Since dipole radiation refers to a $2 \to 3$ process, the momentum conservation can be satisfied with all partons be on their mass-shell.

2.3.2 Hadronization

Color charged partons cannot be observed as free particles, but have to undergo *hadronization* in order to form color-less bound states. This is a direct consequence of confinement (c.f. section 2.1.2). The characteristic scale of hadronization effects is of the order of the hadrons mass. This scale is typically set to $Q_0 = 1\,\mathrm{GeV}$ in event generators, where the parton cascades stops, and the hadronization model is applied. There are two relevant hadronization models: the *cluster fragmentation model* and the *Lund string fragmentation model*. Both models are depicted in figure 2.8.

2.3.2.1 The Lund String model

The *Lund string fragmentation model* [71] treats the quark-antiquark color attraction through a string whose energy follows a classical potential. The energy, which is assigned to each string follows a Coulomb potential $V(r) \propto a/r + br$ as a function of the distance r of the two quarks. The potential is motivated by observations taken from lattice QCD, that at large separations of the two quarks the potential raises linearly. Gluons are assumed to cause kinks in theses strings of the order of their four-momenta. The string energy density

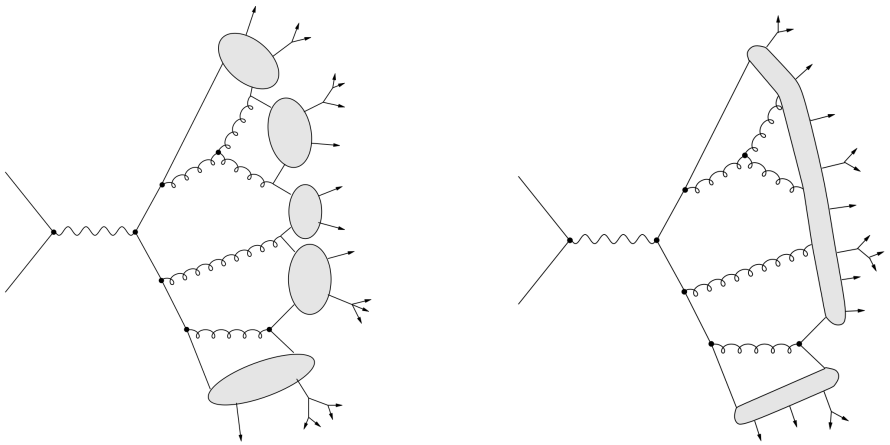


Figure 2.8: Illustration of the cluster fragmentation model (left) and the Lund string fragmentation model (right) in e^+e^--collisions (taken from [70]). The cluster model first splits all gluons into light $q\bar{q}$-pairs and then "proto-hadrons" are formed, which then decay into stable hadrons. The string model spans classical strings between the quark with the string energy being calculated from the gluon-field energy plus contributions from intermediate gluons. Strings can split up in order to form the hadronic final state. In DIS, also the partons from the proton remnant are input to the hadronization models.

is of the order of $\mathcal{O}(1\,\mathrm{GeV/fm})$. If the string energy exceeds a certain threshold of the order of the hadron masses, the string breaks up with some probability and creates a new quark-antiquark pair. The new quark and anti-quarks are reconnected with a new string to the primary antiquark or quark, respectively. When all the available energy is used up, the remaining quarks are combined into mesons and baryons. In hadron induced processes, like DIS, also the hadron remnant and also the initial state radiated partons take place in this prescripton. Multiple quark-antiquark pairs are then treated in the large N-limit, where each parton in the system has its unique color partner.

2.3.2.2 The cluster fragmentation model

The *cluster fragmentation model* [72] is based on the preconfinement [73] property of QCD. In a first step, this algorithm splits all gluons from the shower in the planar approximation into (light) quark-antiquark pairs. All quarks are then grouped together with their color partners into *clusters*. The mass spectrum of these color singlet pairs is asymptotically independent of the energy or the production mechanism, and peak at the low mass scale $\sim Q_0$, where the parton shower terminated. Most of the clusters are then being decayed in an isotropic quasi-two-body phase space model and are identified as (stable) hadrons corresponding to their flavor content. Clusters which are too light to decay are identified as the lightest hadron of this flavor.

2.4 Jet algorithms

In collider experiments single partons cannot be observed, but instead an appearance of a spray of collimated hadrons within some region of the detector is measured. If all the deposited energy, or better all measured four-momenta of the single hadrons, within some certain geometrical region is reasonably 'bundled', we call it a *jet*.
The combination of single components to form a jet is called *jet-algorithm* (or *jet-finder*) and can be performed in various different ways. A jet-algorithm must be identically defined for input parameters coming from the experiment (i.e. particle candidates), from an event

generator (i.e. hadrons or partons), or from fixed-order pQCD calculations (i.e. partons). The input parameters to a jet finder are consistently called 'particles'. Any reasonable jet-algorithm has to fulfill the demands of infrared and collinear safety (*IR-safe*) and has to factorize in hadron collisions[35] which is in particular demanding to have predictive power to all orders in pQCD and should result in a good one-to-one correspondence of pQCD partons to jets.
There are two classes of jet-finders which are most commonly used in high-energy physics: cone-type algorithms and sequential recombination algorithms. The first one defines a cone of a certain radius, and all particles within this cone are recombined in order to form the jet four-momentum. The cone axis can be either chosen from the input particles, through an iterative procedure or by maximizing the energy flow within the cone[36]. The other type of algorithm, a sequential recombination algorithm, is employed in this work. These kind of algorithms are infrared and collinear safe to all orders in pQCD. A set of rules are subsequently applied to a list of all particles. In the prescription of the generalized k_{T} jet algorithm, in a first step a distance-measure between all particles are calculated by

$$d_{ij} = \min(k_{\mathrm{T},i}^{2a}, k_{\mathrm{T},j}^{2a}) \frac{\Delta_{ij}^2}{R_0^2} \ , \tag{2.30}$$

$$d_i = k_{\mathrm{T},i}^{2a} \ , \tag{2.31}$$

where $\Delta_{ij}^2 = (\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2$ denotes the distance between two particles i and j (ϕ being the azimuthal angle and η being the pseudorapidity). The dimensionless parameter R_0 is an angular radius and can be interpreted as the jet-size in correspondence to cone-algorithms. Then the iterative algorithm follows the prescription:

1. Find smallest value of all the d_{ij} and d_j and label it $d_{\min}$.

2a. If $d_{\min}$ is d_{ij} then recombine these two particles and replace the two jets by a new pseudo-particle in the list.

2b. If $d_{\min}$ is d_i then call i a jet and remove it from the list of particles.

3. Recalculate the distances d_{ij} and repeat from step 1 until no particles but only jets are left.

The recombination of two particles in step 2a can be calculated according to several recombination schemes. We employ the $\mathrm{P_T}$-recombination scheme, where the new pseudo-particle is calculated according to

$$\begin{aligned} k_{\mathrm{T},ij} &= k_{\mathrm{T},i} + k_{\mathrm{T},j} \ , \\ \phi_{ij} &= (\phi_i k_{\mathrm{T},i} + \phi_j k_{\mathrm{T},j})/k_{\mathrm{T},ij} \ , \\ \eta_{ij} &= (\eta_i k_{\mathrm{T},i} + \eta_j k_{\mathrm{T},j})/k_{\mathrm{T},ij} \ . \end{aligned} \tag{2.32}$$

The $\mathrm{P_T}$-recombination scheme assumes mass-less particles as inputs to the jet algorithm, in order to be invariant under longitudinal boosts. This is a relevant criterion for this analysis in order to enable the boost from the laboratory frame of reference to the Breit

[35] This means that the outcome of the algorithm should be invariant under the radiation of a soft or collinear parton, which cannot be resolved experimentally.

[36] It turned out that seeded cone algorithms are not infrared safe to all orders in pQCD. The algorithm of a seedless infrared safe cone (SISCone) fulfills all requirement and has further comparably low computing costs, which is also a relevant criterion for the practical use in collider experiments.

frame, where the jet-finding is performed. In this thesis, the native distance parameter of $R_0 = 1$ is chosen (following the studies in [74]).
The parameter a in equation 2.31 can be chosen freely and denotes the distinction of three kinds of algorithms.

- $\mathbf{a = 1}$: The inclusive k_T algorithm [75,76] which recombines first the softest particles as it is applied in this analysis.
- $\mathbf{a = 0}$: The Cambridge/Aachen algorithm, [77,78] which recombines first the closest particles. It was originally developed for e^+e^- colliders and was then adopted to DIS experiments [79].
- $\mathbf{a = -1}$: The anti-k_T algorithm [80], which recombines first the hardest particles, and is now preferably used at the LHC.

2.5 Deep-inelastic lepton-proton scattering

In *deep-inelastic scattering* processes $\ell A \to \ell' X$ a lepton ℓ scatters off a hadron A via the exchange of a vector boson, i.e. a photon γ, a Z^0 ('neutral current') or $W^\pm$ boson ('charged current'), where the final state X and the outgoing lepton ℓ' persist. At HERA, the process of an electron or positron (both are denoted *electron* throughout this thesis) scattering off a proton was investigated. In this thesis, only neutral current (NC) interactions are studied, where the outgoing lepton is equal to the incident charged lepton, and the contribution from Z^0 exchange is taken into account by correction factors in calculations.

2.5.1 The kinematics of DIS

The kinematics of a DIS process can be described by only a few invariant scalars, using the incoming lepton momentum k, the hadron momentum p and the outgoing lepton momentum k' (where k, k' and p denote four-momenta k^μ, k'^μ and p^μ, respectively). A diagram for the NC scattering process is depicted in figure 2.9. The momentum transfer carried from the exchanged boson is

$$q = k - k' \ . \tag{2.33}$$

The centre-of-mass energy $\sqrt{s}$ is the given by the Mandelstam variable

$$\sqrt{s} = \sqrt{(k+p)^2} \ . \tag{2.34}$$

In the kinematic region of HERA, the centre-of-mass energy can be approximated conveniently in the massless limit, using the energy of the incoming electron E_e and proton E_p, by $\sqrt{s} \simeq 4E_eE_p$. The virtuality Q^2 of the process is given by

$$Q^2 = -q^2 \ , \tag{2.35}$$

where the minus sign is used by convention to have the variable positive. The dimensionless Bjorken variable[37] is given by

$$x_\mathrm{Bj} = \frac{Q^2}{2p \cdot q} \ . \tag{2.36}$$

[37] In LO approximation the Bjorken variable can be interpreted as the fractional momentum of the struck quark with respect to the longitudinal proton momentum. In higher order the Bjorken variable is essentially different from the proton momentum fraction ξ which appears in the Mellin convolutions of the parton densities with the Wilson coefficients and is defined only within a certain factorization scheme as a theoretical variable.

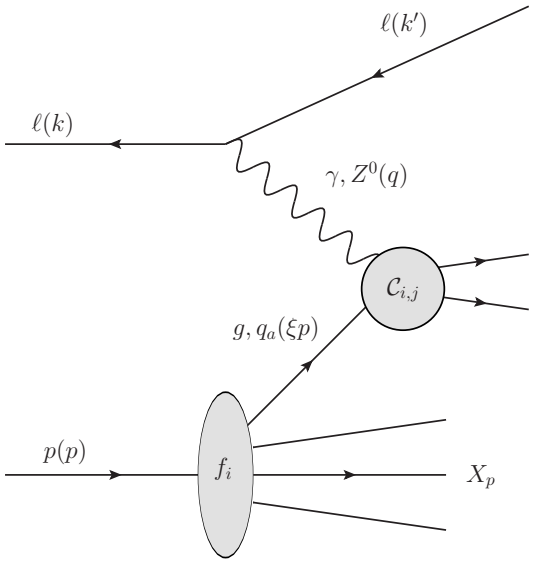


Figure 2.9: The diagram for (neutral current) deep-inelastic scattering: ℓ denotes an incoming charged lepton or anti-lepton and p denotes the incoming proton with parton densities f_i. Some relevant kinematic variables are given in brackets. The exchanged boson can be a photon, a γZ^0-interference, or a Z^0 term. The hard scattering coefficient $\mathcal{C}_{i,j}$ is calculated in pQCD, where the incoming parton is either a gluon or an (anti-)quark.

The process is called *deep-inelastic* if the momentum transfer Q^2 is large compared to the hadron's rest mass, i.e. $Q^2 \gg m_p^2 \approx 1\,\mathrm{GeV}^2$.
Another important dimensionless quantity is the *inelasticity* variable

$$y = \frac{p \cdot q}{p \cdot k} \ , \tag{2.37}$$

which corresponds to the energy loss of the lepton in the rest frame of the hadron. Both, x and y, are limited to values between 0 and 1.

2.5.2 The inclusive DIS cross section

The cross section of unpolarized inclusive deep-inelastic scattering is given by the electroweak theory, which describes how the vector boson couples to leptons, while higher order corrections are dominantly given by QCD[38]. The differential (electroweak) Born cross section can be written as [57, 81]

$$\mathrm{d}\sigma = \frac{2\pi y \alpha_{em}^2}{Q^4} \sum_i \eta_i(Q^2) L_i^{\mu\nu}(k,q) W_{i,\mu\nu}(p,q) \ , \tag{2.38}$$

where definitions of equations 2.35, 2.36 and 2.37 are used, α_{em} denotes the electromagnetic coupling, and i denotes the four electroweak current interactions, namely $i = |\gamma|^2, |\gamma Z|, |Z^2|$ for neutral current and $i = |W^\pm|^2$ for charged current interactions. We are only interested in neutral current photon exchange, where the factor η becomes $\eta_{|\gamma|^2}(Q^2) = 1$ and the leptonic tensor reads [12]

$$L_{\mu\nu}^{|\gamma|^2} = 2(k_\mu k'_\nu + k'_\mu k_\nu - k \cdot k' g_{\mu\nu} - i\lambda \varepsilon_{\mu\nu\alpha\beta} k^\alpha k'^\beta) \ , \tag{2.39}$$

[38] In ep scattering, for inelasticities $y \to 0$ and $y \to 1$ the QED radiative corrections due to ISR and FSR radiation off the electron are larger than the QCD corrections. However, these corrections are well-understood theoretically.

where $\lambda = \pm 1$ denotes the helicity of the incoming lepton. The hadronic tensor $W^{\mu\nu}$ depends on p and q and describes the coupling of the vector boson to the hadronic system. Although the inclusive cross section in equation 2.38 is only in LO electroweak theory, it involves the dominant higher order corrections, which are in powers of the strong coupling α_s, expressed through the hadronic tensor. By taken the Lorentz-invariance of $W^{\mu\nu}$ and the current conservation into account, the hadronic tensor can be written by three tensor structures which multiply to a structure function F_1, F_2 and F_3, by [12, 57]

$$\begin{aligned} W_{\mu\nu} = &- \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) F_1(x_{\mathrm{Bj}}, Q^2) \\ &+ \left(p_\mu - q_\mu \frac{p\cdot q}{q^2}\right)\left(p_\nu - q_\nu \frac{p\cdot q}{q^2}\right)\frac{1}{p\cdot q} F_2(x_{\mathrm{Bj}}, Q^2) \\ &- i\varepsilon_{\mu\nu\lambda\sigma}\frac{p^\lambda q^\sigma}{2p\cdot q} F_3(x_{\mathrm{Bj}}, Q^2)\ . \end{aligned} \tag{2.40}$$

Since the structure functions $F_{1,2,3}$ are Lorentz-scalars they can only depend on the invariants x_{Bj} and Q^2.

2.5.3 The neutral current DIS cross section calculations

By using equations 2.38, 2.39 and 2.40, the double differential cross section for unpolarized deep-inelastic scattering as functions of x_{Bj} and Q^2 is generically written in terms of structure functions by

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{Bj}}\mathrm{d}Q^2} = \frac{4\pi\alpha_{em}^2}{2x_{\mathrm{Bj}}Q^4}\left(Y_+ F_2 - y^2 F_L \mp Y_- x F_3\right)\cdot(1+\Delta^{\mathrm{weak}}), \tag{2.41}$$

where $Y_\pm = 1 \pm (1-y)^2$ and the structure functions are determined from experiment. The *longitudinal structure function* F_L is defined as $F_L = F_2 - 2xF_1$. The factor Δ^{weak} involves corrections in higher orders of the electroweak theory of $\mathcal{O}(\alpha_{em}^3)$.
The mass-less structure functions can be written in a perturbative expansion in powers of $\alpha_s(\mu_r)/(4\pi)$ and according to the factorization theorem in forms of Mellin convolutions of the parton density functions f_j and the perturbative coefficients $\mathcal{C}_{i,j}^{(n)}$ by

$$\mathcal{F}_i(x_{\mathrm{Bj}}, Q^2) = x_{\mathrm{Bj}} \sum_{n=0}^{\infty} \frac{\alpha_s(\mu_r)^n}{(4\pi)^n} \sum_{j=q,\bar{q},g} f_j(\xi,\mu_f) \otimes \mathcal{C}_{i,j}^{(n)}\left(\frac{x_{\mathrm{Bj}}}{\xi}, Q^2, \mu_r, \mu_f\right) + \mathcal{O}\left(\frac{m}{Q^2}\right), \tag{2.42}$$

where j denotes the parton flavor and $\mathcal{F}_i$ labels any structure function (e.g. F_2, $x_{\mathrm{Bj}}F_3$, F_L). The coefficient functions $\mathcal{C}_{i,j}^{(n)}$ can be calculated from Feynman diagrams. As example, a LO and one of the NLO diagrams are illustrated in figure 2.10. In this work, the calculations are performed using the NLO coefficient functions [82] (i.e. $\mathcal{O}(\alpha_s)$). The $\mathcal{C}_{i,j}^{(n)}$ are currently known up to NNLO [83–86][39] (i.e. $\mathcal{O}(\alpha_s^2)$)[40].

[39] The full analytic coefficients in NNLO are about $\mathcal{O}(100)$ pages in normalsize fonts.

[40] Here, it must be taken care for the counting of higher orders. For the longitudinal structure function F_L the LO (i.e. the first non-vanishing order) is of order $\mathcal{O}(\alpha_s)$ and thus the NLO is of $\mathcal{O}(\alpha_s^2)$. For the structure functions F_1 and F_2 the LO is of zeroth order in α_s, namely just $\mathcal{O}(\alpha_{em}^2)$, while the NLO is of the order $\mathcal{O}(\alpha_{em}^2\alpha_s)$.

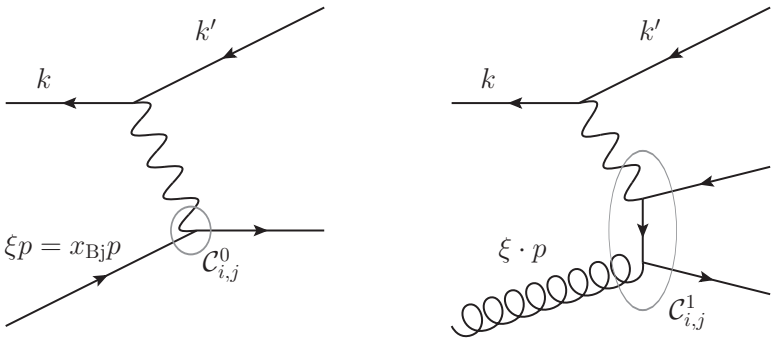


Figure 2.10: Some Feynman diagrams for the hard scattering part of the DIS cross section. The left one is the only contributing LO diagram in DIS ($\mathcal{O}({\alpha_s}^0)$) and corresponds to the quark parton model approximation. Only quarks from the proton PDFs contribute. The right diagram is one of the higher order contribution, where already gluon contributions are factorized in the PDF.

2.5.3.1 The quark parton model

In zeroth order in α_s it is assumed that the vector boson couples directly to the quarks inside the proton. The coefficients $\mathcal{C}_{i,j}^{(0)}$ become factorization scheme independent (i.e. no dependence on μ_f) and since they have no strong vertex included simplify to a delta function

$$\mathcal{C}_{1,j}^{(0)}(x_{\mathrm{Bj}}, Q^2) = \frac{1}{2} e_j^2 \delta(\frac{x_{\mathrm{Bj}}}{\xi} - 1) \tag{2.43}$$

$$\mathcal{C}_{2,j}^{(0)}(x_{\mathrm{Bj}}, Q^2) = x_{\mathrm{Bj}} e_j^2 \delta(\frac{x_{\mathrm{Bj}}}{\xi} - 1), \tag{2.44}$$

where e is the electric charge of the interacting quark of flavor i and the factor $1/2$ arises from the quark spin. In this approximation, the integration variable ξ is identical to the Bjorken scaling variable $\xi = x_{\mathrm{Bj}}$ and the structure functions read

$$F_1(x_{\mathrm{Bj}}, Q^2) \simeq \frac{1}{2} \sum_{j=q,\bar{q}} e_j^2 f_j^{(0)}(x_{\mathrm{Bj}}) + \mathcal{O}(\alpha_s) + \mathcal{O}(m/Q^2) \tag{2.45}$$

$$F_2(x_{\mathrm{Bj}}, Q^2) \simeq x_{\mathrm{Bj}} \sum_{j=q,\bar{q}} e_j^2 f_j^{(0)}(x_{\mathrm{Bj}}) + \mathcal{O}(\alpha_s) + \mathcal{O}(m/Q^2) \tag{2.46}$$

$$F_L(x_{\mathrm{Bj}}, Q^2) = 0, \tag{2.47}$$

and they fulfill the Callan-Gross relation [87]

$$F_2(x_{\mathrm{Bj}}) = 2x_{\mathrm{Bj}} F_1(x_{\mathrm{Bj}}). \tag{2.48}$$

This zeroth order approximation in α_s corresponds to the *quark parton model* (QPM) [88, 89]. The QPM was successful to explain the scaling behavior of F_2 (the absence of Q^2 dependence [90]) detected in early DIS experiments [91] and proved the proton consists of spin-$\frac{1}{2}$ particles, the quarks. In this approximation, the longitudinal structure function F_L is zero.

2.6 The Breit frame of reference

Lowest order DIS is not directly sensitive to the strong coupling, since the corresponding QPM matrix element is only of zeroth order in the strong coupling. The transverse momentum of the scattered lepton balances exactly the transverse momentum of the outgoing parton (jet) in the laboratory reference frame. Therefore, a meaningful jet-cross section has to be defined at least in first order in α_s. This can be achieved, by performing the analysis in the Breit frame of reference [89, 92], which is defined by

$$2x_{\mathrm{Bj}}\vec{p} + \vec{q} \stackrel{!}{=} 0, \tag{2.49}$$

where $\vec{p}$ and $\vec{q}$ are the momenta of the incoming proton and the exchanged boson, respectively, and the longitudinal (z) direction is given by the incoming boson. The full coordinate system is defined such that the scattering occurs in the (x, z)-plane, and thus having vanishing p_y momentum. The boson becomes purely space-like in the longitudinal direction, i.e. $q = (0, 0, 0, -Q)$. A graphical illustration is given in figure 2.11.

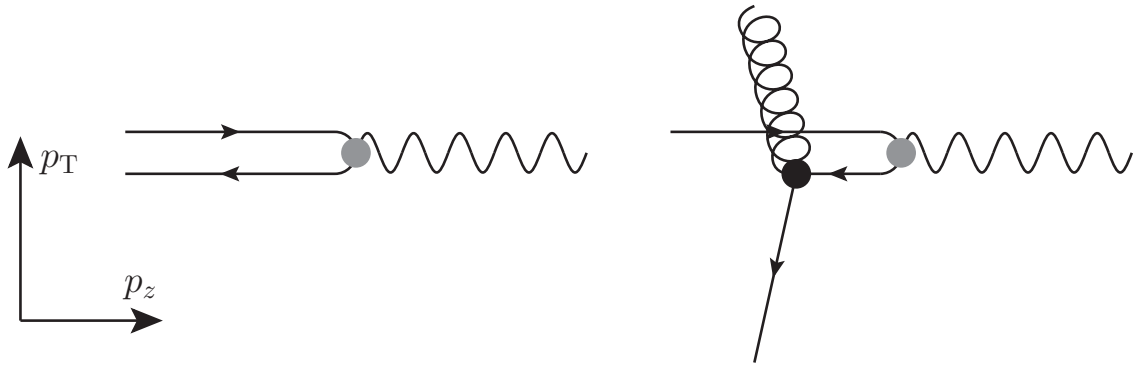


Figure 2.11: Schematic illustration of the Breit reference frame in the (p_z, p_{T})-plane. In the quark-parton model approximation, the incoming boson and quark, as well as the outgoing quark have no transverse momentum. If a strong vertex is present, as depicted by the black dot, the radiated gluon as well as the outgoing quark receive transverse momentum.

In the quark-parton model approximation, the Breit frame corresponds to the reference frame in which the exchanged boson and the struck quark collide head-on. Then, the incident quark has only transverse momentum of $p_z = Q/2$ and the outgoing quark is scattered back longitudinally with $p' = (0, 0, 0, -Q/2)$.
This reference frame provides the advantage that events occuring at high scales can be clearly distinguished from born-level DIS processes experimentally, by requiring some sufficiently high transverse momentum of the jets. A consequence of this requirement is that in lowest order always two outgoing partons are involved, since the Born-level cross section is of the order $\mathcal{O}(\alpha_{em}^2 \alpha_s)$.

2.7 Calculation of next-to-leading order jet cross sections

Perturbative calculations at higher orders in α_s are very challenging due to the large number of contributing Feynman diagrams and due to the large number of (soft) singularities (see section 2.1.3.3). Especially, since jet-measurements involve diagrams with many external

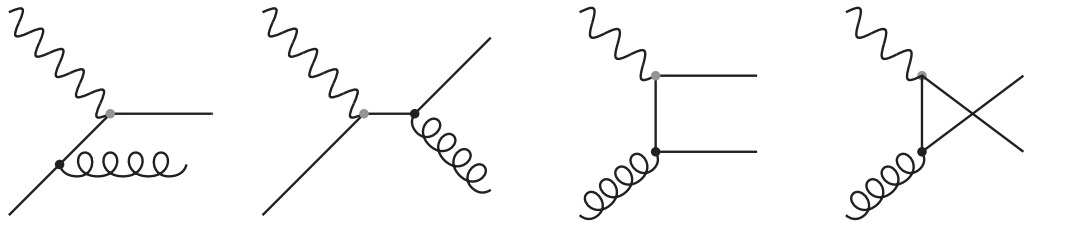


Figure 2.12: The LO Feynman diagrams for jet-production in neutral current DIS. QCD vertices are depicted with a black dot, and electroweak vertices are depicted by a gray dot. QCD compton graphs refer to diagrams, where a gluon is radiated from a quark propagator as higher order corrections to QPM diagrams. Boson gluon fusion diagrams refer to processes, in which the electroweak boson interacts with a gluon inside the proton through a quark line.

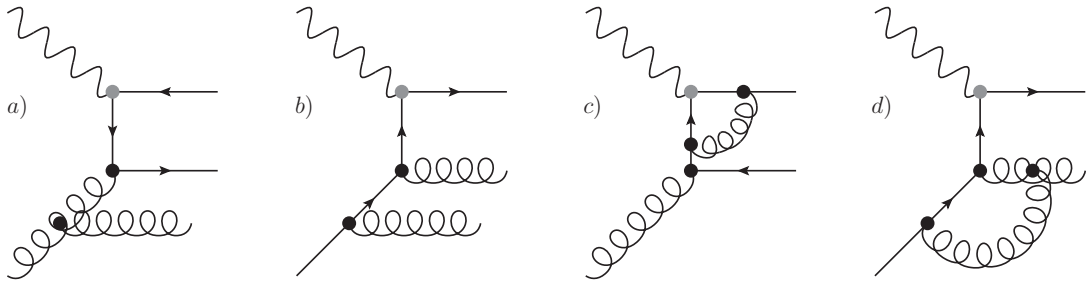


Figure 2.13: Some NLO corrections to the 2-jet Feynman diagrams. The diagrams a) and b) show real corrections, and diagram c) and d) depict some of the virtual loop corrections. Diagrams a) and b) further are some of the LO contributions to the three-jet cross section.

legs. The four LO diagrams for 2-jet production are shown in figure 2.12 and in LO there is always a one-to-one correspondence between partons and jets (c.f. section for the definition of a jet 2.4). Higher-order corrections consist of diagrams involving virtual loop-corrections or real corrections, e.g. by gluon radiation as indicated in figure 2.13. The NLO real corrections to 2-jet diagrams are identical to the 3-jet LO contributions.
Fixed order calculations in pQCD of an n-jet observable can be performed separately, i.e. order by order, by calculating the *partonic cross section*

$$\sigma_a = \sigma_a^{\mathrm{LO}} + \sigma_a^{\mathrm{NLO}} + \mathcal{O}(\mathrm{NNLO}) \simeq \int_n \mathrm{d}\sigma_a^B + \left[\int_{n+1} \mathrm{d}\sigma^R + \int_n \mathrm{d}\sigma^V \right], \tag{2.50}$$

where in NC DIS jet-production the born cross section σ_a^{LO} is of order $\mathcal{O}(\alpha_s^{n-1})$. Here, the calculations are performed using the program NLOJET++ [93, 94], which has implemented the *(dipole) subtraction method*[41] [96] for the separation of the divergences.
The Feynman diagrams of the NLO correction σ_a^{NLO} consist of real emission diagrams (R) which are contributions of $n+1$ parton configurations, and n-leg graphs involving virtual corrections (V) of (one) internal loop (compare figure 2.13).
Both two contributions are singular if the dimensional regularization parameter vanishes, $\epsilon \to 0$, due to the existence of $1/\epsilon$ and $1/\epsilon^2$ poles. These poles can be separated in the calculation and cancel exactly between the real $1\gamma \to n+1$ and virtual $1\gamma \to n$ diagrams in order-by-order pQCD, if the physical measurement is infrared safe, according to the

[41] The dipole subtraction method is preferred over the *slicing-method*, which is an alternative method for the separation of the divergences, since no approximations are made (further advantages are outlined in [95]).

Kinoshita-Lee-Nauenberg theorem [26, 27]. This requirement, that the emission of a soft or collinear parton must not influence the result of the measurement, is mathematically expressed by a *measurement function* $\mathcal{S}_n$ (or also called *jet function*)[42]. The real and virtual contribution to the partonic cross section can then be written in terms of finite local counterterms, by [96]

$$\sigma_a^{\mathrm{NLO}}(p_a, \mu_f) = \int_{n+1} \left(\mathrm{d}\sigma^R(p_a) - \mathrm{d}\sigma^A(p_a)\right) + \left[\int_{n+1} \mathrm{d}\sigma^A(p_a) + \int_m \mathrm{d}\sigma^V(p_a) + \int_m \mathrm{d}\sigma^C(p_a, \mu_f)\right]. \tag{2.51}$$

The partonic cross sections therein are made infrared safe by applying the jet functions to the divergent matrix elements

$$\mathrm{d}\sigma_n \simeq \mathrm{d}\Phi |\mathcal{M}_n(p_1, \ldots, p_n)|^2 \mathcal{S}(p_1, \ldots, p_n), \tag{2.52}$$

where $\mathrm{d}\Phi$ is the n-particle phase space and $\mathcal{M}_n$ are the relevant matrix elements. The final (differential) cross section for jet-observables can then be calculated by a convolution of the partonic cross sections with the PDFs according to the factorization theorem (see equation 2.19):

$$\sigma = \sum_{a=q,\bar{q},g} \left[\sigma_a^{\mathrm{LO}} + \sigma_a^{\mathrm{NLO}}\right] \otimes f_a. \tag{2.53}$$

Perturbative methods can only be applied to predict cross sections for partonic quantities, i.e. where the color-recombination process of hadronization is not considered. In order to predict jet cross sections on hadron level, which then can be compared to the measured cross sections, the perturbatively calculated parton level cross section has to be further corrected for hadronization effects (c.f. section 2.3.2).
The definition of jet cross sections, and thus their practical calculation typically involve cuts on various observables. These are for example restrictions on the phase space from limited detector acceptance in the jet pseudorapdity or cuts on various kinematic jet observables like the jet transverse momentum or the invariant mass of multiple jets. These cuts can be expressed by a *cut function* $\Theta(\mathrm{cuts})$ and the phase space integration becomes $\mathrm{d}\Phi \rightarrow \mathrm{d}\Phi\Theta(\mathrm{cuts})$. Such phase space constraints make the analytic calculation of the integrals in the jet cross sections according to equations 2.50 and 2.53 practically impossible and therefore Monte Carlo methods are employed. There, it must be taken care of that the subtraction terms $\mathrm{d}\sigma^A$ from the real and virtual contributions cancel accurately. This can become challenging, since the corresponding terms from real and virtual contributions have slightly different parton configurations and one of the two contributions could be inside the accepted phase space where the counter term is not. Typically, such contributions have particular large phase space weight according to their infrared origin. As a consequence, the calculation of jet cross sections by applying Monte Carlo methods can demand a large amount of computing power in order to ensure the accurate cancellation of the subtraction terms. This becomes even more evident if a multi-differential measurement is performed or the measurement comprises very small phase space bins (i.e. $\int \mathrm{d}\Phi\Theta(\mathrm{cuts})$ is getting small). For a fast repeated calculation of jet-cross sections we apply the fastNLO framework [98,99]. This tool enables to calculate perturbative coefficients with nlojet++ with high statistical accuracy, and these are stored such that the coupling strength and the PDF can be changed

[42] The requirements on the measurement function in DIS, where also the initial state parton must be considered are outline in [96, 97].

without a recalculation of the coefficients. We have improved the fastNLO concept towards a scale independent prescription [100], where also the renormalization and factorization scales can be changed without a recalculation of the matrix elements. The fastNLO method is outlined in appendix C.

2.8 $\alpha_s(M_Z)$ determinations from jet-measurements in DIS

Values of the strong coupling have been determined from many measurements at the HERA collider. Here we briefly summarize recent determinations from jet-measurements which are performed by the H1 and ZEUS experiments. All mentioned values are displayed in figure 2.14.
The H1 experiment has determined values of $\alpha_s(M_Z)$ from double-differential inclusive jet, dijet and trijet cross sections in NC DIS at low Q^2 between $5 < Q^2 < 100\,\mathrm{GeV}^2$ with an integrated luminosity of $43.5\,\mathrm{pb}^{-1}$ [101]. Jets were defined using the k_T cluster algorithm in the Breit frame with a transverse momentum $p_\mathrm{T}^\mathrm{jet} > 5\,\mathrm{GeV}$ and a pseudorapidity of $-1 < \eta_\mathrm{lab}^\mathrm{jet} < 2.5$. Furthermore, a value of $\alpha_s(M_Z)$ was derived from the double-differential three-jet over two-jet ratio. In NC DIS at high Q^2, with values of $150 < Q^2 < 15\,000\,\mathrm{GeV}^2$, H1 has determined $\alpha_s(M_Z)$ from normalized jet measurements [102] with an integrated luminosity of $395\,\mathrm{pb}^{-1}$ from HERA-I and HERA-II data. Jets were searched using the k_T algorithm in the Breit frame with a minimum $p_\mathrm{T}^\mathrm{jet} > 5\,\mathrm{GeV}$. The acceptance was restricted to a pseudorapidity range of $-0.8 < \eta_\mathrm{lab}^\mathrm{jet} < 2$ and inclusive jet and dijet cross sections were measured double-differentially and being normalized to the NC DIS cross section. Furthermore, a value for $\alpha_s(M_Z)$ from single-differential normalized trijet cross sections was obtained from this data. A value of $\alpha_s(M_Z)$ from inclusive jets in the high Q^2 region by H1 is derived from the double-differential inclusive jet-cross sections using HERA-I data at a center-of-mass energy of $\sqrt{s} = 319\,\mathrm{GeV}$ with an integrated luminosity of $65.4\,\mathrm{pb}^{-1}$ [103]. The phase space of this publication is identical to the one of the analysis presented here.

The ZEUS experiment has determined values of $\alpha_s(M_Z)$ from double-differential inclusive jet-cross sections at very large Q^2 values ($Q^2 > 500\,\mathrm{GeV}$) for different jet algorithms [104, 105]. Jets were searched using the k_T, the anti-k_T and the SIScone jet algorithms with a jet-size parameter of $R = 1$ and with an $E_T^\mathrm{jet} > 8\,\mathrm{GeV}$ in a pseudorapidity range in the Breit frame of $-2 < \eta_\mathrm{Breit}^\mathrm{jet} < 1.5$. The data was taken in the HERA-I period with an integrated luminosity of $\mathcal{L} = 81.7\,\mathrm{pb}^{-1}$. The values of $\alpha_s(M_Z)$ of the three jet-algorithms are consistent within errors. A preliminary updated value for k_T-jets was derived from the HERA-II data set with $\mathcal{L} = 299\,\mathrm{pb}^{-1}$ [106]. A value of $\alpha_s(M_Z)$ from a dijet measurement was derived from the dijet fraction R_{2+1} from HERA-I data with $\mathcal{L} = 38.4\,\mathrm{pb}^{-1}$ [107]. A more recent determination of $\alpha_s(M_Z)$ from ZEUS dijet data, taking the full HERA-II data, was performed in a private study [108]. It takes into account double-differential inclusive dijet cross sections with an integrated luminosity of $374\,\mathrm{pb}^{-1}$ for high Q^2 values of $125 < Q^2 < 20\,000\,\mathrm{GeV}^2$ for jets with an $E_\mathrm{T} > 8\,\mathrm{GeV}$ [109]. Furthermore, the ZEUS experiment derived values of $\alpha_s(M_Z)$ from single-differential inclusive jets in photoproduction [110]. Jets were searched using the k_T algorithm and jets with high E_T ($21 < E_T^\mathrm{jet} < 71\,\mathrm{GeV}$) were input to the α_s-determination. ZEUS has also determined a value of $\alpha_s(M_Z)$ from the single differential trijet over dijet ratio from HERA-I data of an integrated luminosity of $82.2\,\mathrm{pb}^{-1}$ [111]. Trijet and dijet events are searched

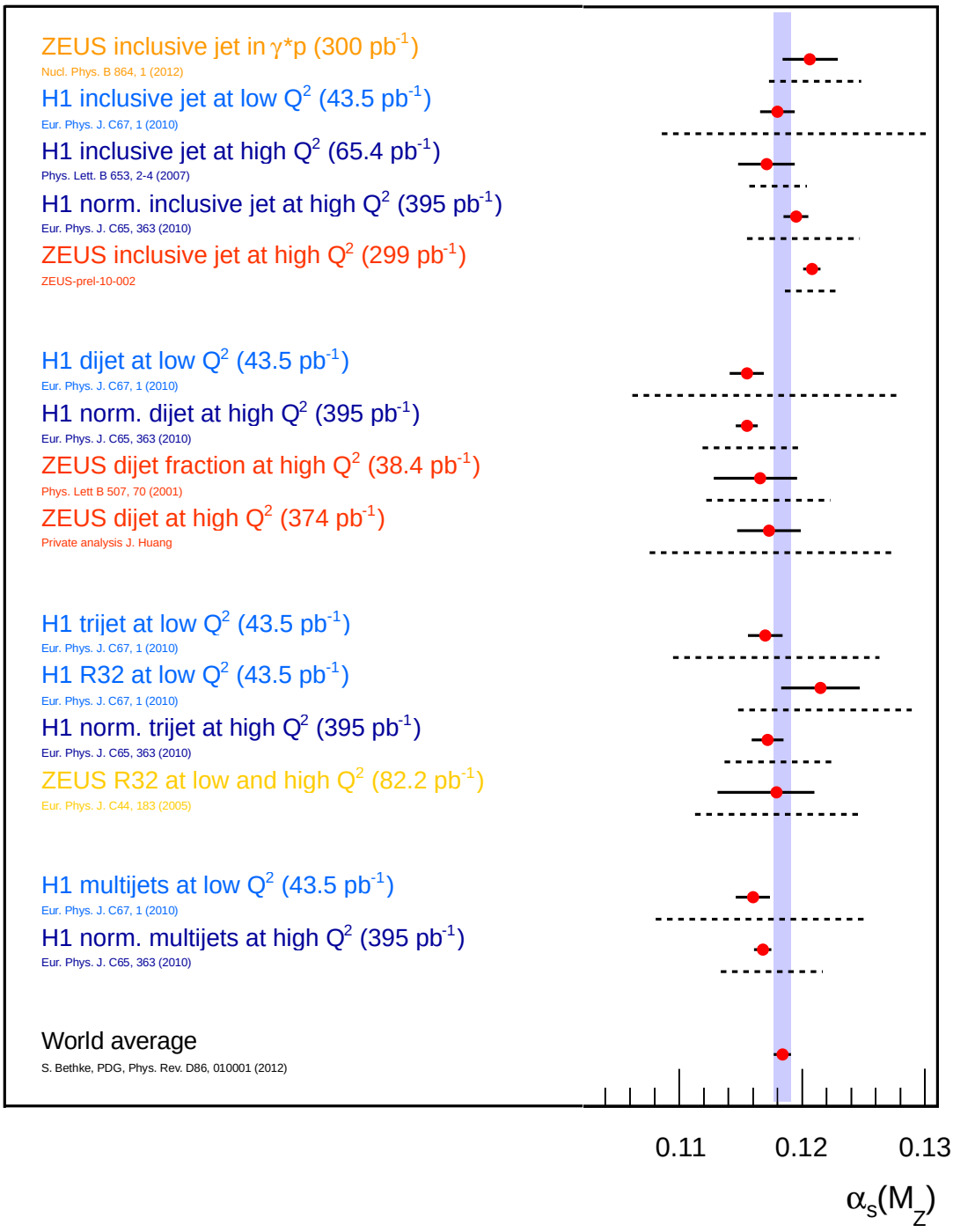


Figure 2.14: Values of $\alpha_s(M_Z)$ from jet-measurements at HERA in comparison to the 2012 world average value. Values are shown for inclusive jet, dijet and trijet measurements in photoproduction and for low Q^2 and high Q^2 DIS. The black error bars indicate the experimental uncertainty and the dashed error bars show the estimated theoretical uncertainty. Jet data at low Q^2 has typically smaller statistical uncertainties, but the resulting $\alpha_s(M_Z)$ has a larger theoretical uncertainties due to the small scale of the process. The value derived from photoproduction data further has to take large uncertainties from the poor knowledge of the photon PDF into account.

using the k_{T}-algorithm in a jet-acceptance of $-1 < \eta^{jet}_{\mathrm{Lab}} < 2.5$ with a transverse energy of $E^{\mathrm{jet}}_{\mathrm{T,B}} > 5\,\mathrm{GeV}$ in the Breit frame. The measurement is performed single differentially in Q^2 of values between $10 < Q^2 < 5000\,\mathrm{GeV}^2$.

The dominant uncertainty of all values of $\alpha_s(M_Z)$ from HERA jets is the theoretical uncertainty, since only NLO calculations are available. Especially values derived from low Q^2 measurements suffer from large uncertainties arising from scale variations in the NLO calculations. Normalized cross sections typically have smaller errors since experimental uncertainties cancel partly or fully. Similar to the cancellation of uncertainties also the sensitivity on the strong coupling is reduced for ratios. For this reason, no value of $\alpha_s(M_Z)$ for the trijet over dijet ratio was derived for high Q^2 data by H1 [74], since the trijet over NC DIS ratio has a larger sensitivity to α_s.
It is remarkable that values of $\alpha_s(M_Z)$ derived from dijet cross sections show typically lower values, than those from inclusive jet data. However, within the theoretical uncertainty, all values of $\alpha_s(M_Z)$ are well compatible with the world average, although none of the data points are input to this average.

Chapter 3

Experimental setup

In this chapter the experimental setup of the H1 experiment at the HERA collider is described. The relevant detector components for this analysis and the calibration of the hadronic final state are recounted.

3.1 The HERA collider

The HERA collider (from german **H**adron **E**lektron **R**ing **A**nlage) is an electron proton collider at DESY in Hamburg. The collider is situated in a circular tunnel of 6.3 km length with protons and electrons provided by the pre-accelerators PETRA and DESY. A schematic view on the accelerator complex is shown in figure 3.1. The HERA collider was operational from 1992 to 2007, with the exception of an interruption in the years 2000 to 2002 where a luminosity upgrade was performed, by introducing more focused beams in the interaction region. The two phases of operation are denoted HERA-I (pre-upgrade) and HERA-II (post-upgrade) phase. This analysis makes use of data taken during the HERA-II phase.

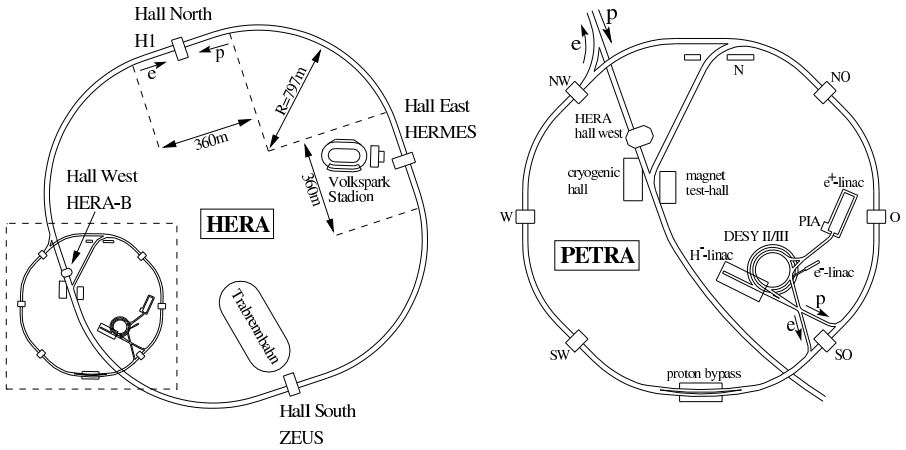


Figure 3.1: Schematic geographical view of the HERA collider, the experiments and the pre-accelerator complex (left) and an enlargement of the pre-accelerator complex (right) [112]. The H1 experiment is situated at the interaction point of the electron and the proton beam in the Hall North. The HERA tunnel consists of straight sections of 360 m and of arches with a radius of 797 m. The arches are equipped with dipole and quadrupole magnets, while the straight sections are used for beam acceleration and beam quality devices as well as for setting the spin orientation of the lepton beams for the two collider experiments.

HERA-II parameters

Run period	e^+p 2003/04	e^-p 2005	e^-p 2006	e^+p 2006/07
Lepton type	e^+	e^-	e^-	e^+
Lepton E_ℓ (proton E_p) energy	27.6 GeV/c (920 GeV/c)			
Center of mass energy	319 GeV/c			
Number of lepton bunches	184	142	180	180
Number of proton bunches	184	140	180	180
Lepton (proton) bunch length	10 mm (191 mm)			
Beam sizes $\sigma_x \times \sigma_y$	112 µm × 30 µm			
Average (maximum) lepton beam current [mA]	25 (52)	22 (42)	22 (62)	25 (45)
Average (maximum) proton beam current [mA]	80 (122)	78 (102)	79 (108)	89 (112)
Delivered integrated luminosity [pb^{-1}]	92.8	218	89.7	184
H1 recorded integrated luminosity (good+medium) [pb^{-1}] [113]	60.3	168	69.7	151
H1 maximum integrated run luminosity [nb^{-1}] [113]	117	260	153	164
Peak luminosity [µb^{-1}/s]	38.0	48.0	46.2	40.8

Analysis specific parameters

Considered run ranges	357160 - 392213	399629 - 436893	444312 - 466997	468530 - 500611
Integrated luminosity of selected runs (BH analysis) [pb^{-1}]	53.2	101.9	57.7	139.7
QED Compton correction to integrated luminosity [114]	0.999	0.991	1.009	1.039
Total selected integrated luminosity from QED Compton analysis	$357.6 \pm 8.9\,\text{pb}^{-1}$			

Table 3.1: Table of average and maximally achieved beam parameters of the HERA lepton and proton beam during the HERA-II running period, which is subdivided into four running periods [113]. Shown are also delivered and the recorded integrated luminosities by H1 from the online luminosity system [113]. In the four lower rows, the selected run ranges and their integrated luminosities are given.

The HERA collider was operated with electrons or positrons[1] of an energy of $E_e = 27.6\,\text{GeV}$ and a proton energy of $E_p = 920\,\text{GeV}$ yielding ep-collisions with a center-of-mass energy of

$$s \simeq \sqrt{4E_pE_e} = 319\,\text{GeV}. \tag{3.1}$$

An overview of the achieved running conditions and provided luminosities is given in table 3.1.

Four experiments are located at the HERA collider. Two experiments were designed as 4π multipurpose experiments, H1 and ZEUS. The other two experiments were fixed target experiments, HERMES[2] and HERA-B[3].

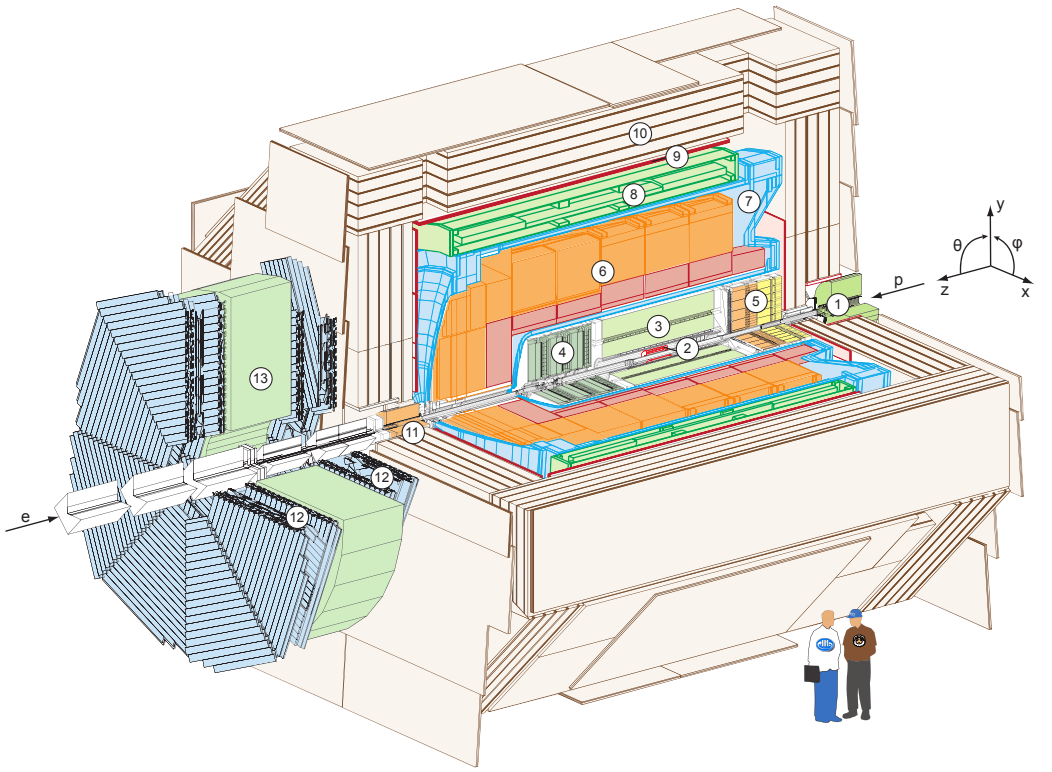
[1]The beam currents of luminosity fills were initially around $I_p \simeq 110\,\text{mA}$ for the proton and $I_e \simeq 50\,\text{mA}$ for the electron beam. The currents decreased in a typical 12 hour run by around 60 %. The beams were structured into 220 bunches with typically around 174 were filled for physics collisions, which results in a collision frequency of 10.4 MHz or every 96 ns. Some unpaired bunches of protons and electrons were used for beam monitoring.

[2]The HERMES expreiment is dedicated to measuring polarized electron-nucleon scattering, using the lepton beam on a gas target.

[3]The HERA-B experiment was intended to measure production cross sections of B-mesons using protons and CP violation in the B-meson system, by directing the proton beam on a fixed wire target.

3.2 The H1 detector

The H1 detector is one the two multi-purpose detectors located at the HERA collider. Its dimensions are about $12 \times 10 \times 15\,\text{m}$ and it has an approximate weight of 2800 tons. The H1 detector consists of several detector components which are arranged in an asymmetric onion-shell design. A schematic drawing of the H1 detector is shown in figure 3.2.



1. Beam pipe and beam magnets
2. Silicon tracking detector
3. Central tracking detector
4. Forward tracking detector
5. Spacal calorimeter (em and had)
6. Liquid Argon calorimeter (em and had)
7. Liquid Argon cryostat
8. Superconductiong coil
9. Muon chambers
10. Instrumented iron (streamer tube detectors)
11. Plug calorimeter
12. Forward muon detector
13. Muon toroid magnet

Figure 3.2: A schematic drawing of the H1 detector based on the GEANT-based detector simulation. The asymmetric layout in the z-direction is typical for the two $e^{\pm}p$-detectors. The proton beam enters from the right, which also defines the positive z-axis of the right-handed H1 coordinate system together with the x-axis pointing to the HERA-ring center. The relevant components are described in the text. The detector components at large z, which are located outside of the experimental hall in the HERA tunnel are not shown. These are the components of the luminosity system: electron taggers (ET) at $z = -5.4\,\text{m}$ and at $z = -33.4\,\text{m}$ and a photon calorimeter (PD) at $z = -102.9\,\text{m}$. Other components in the forward direction are a proton spectrometer (FPS) at $z = 81\,\text{m}$ and $z = 90\,\text{m}$ as well as a forward neutron spectrometer at $z = 110\,\text{m}$, which are used e.g. for measuring diffractive processes. At a distance of $z = 220\,\text{m}$ a further proton spectrometer is installed.

3.2.1 Subsystems of the H1 detector

The tracking system is subdivided into the *central tracking detector* (CTD) and the *forward tracking detector* (FTD) and the *backward proportional chambers* (BPC). The relevant calorimetry systems are the *Liquid Argon calorimeter* (LAr) in the central and forward direction and the Spaghetti Calorimeter (*SpaCal*) in the backward direction. The tracker and the LAr are surrounded by a superconducting solenoid magnet, which provides a uniform axial-symmetric magnetic field of 1.15 T. Outside of the solenoid, an iron return yoke is installed, which is equipped with streamer tube detectors for muon measurements. A full description of the H1 detector can be found in refs. [115, 116]. In the following, only the relevant parts for this analysis will be discussed.

3.2.2 The H1 coordinate system

The H1 reference system is defined as a right-handed coordinate system, where the x-axis is pointing to the HERA-ring center and the z-axis is given by the proton beam direction. The origin of the coordinate system is set to the nominal interaction point. The spherical coordinate system is defined by the azimuthal angle $0 < \phi < 2\pi$, the angle w.r.t. the x-axis in the (x, y)-plane, and the polar angle $0 < \theta < \pi$, where $\theta = \pi$ defines the electron beam direction.
It is convenient to use the *pseudorapidity* η instead of the polar angle θ which is defined for ultra-relativistic particles by

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \ . \tag{3.2}$$

The advantage of the pseudorapidity over the polar angle is its linear tranformation under Lorentz-boosts along the z-axis. Colloquially, the sign of the pseudorapidity subdivides the detector into the *forward direction* $\theta < \pi/2$ and the *backward direction* $\theta > \pi/2$, which are expressed by positive or negative pseudorapidities, respectively.

3.2.3 The tracking detectors

The H1 tracking system consists of the CTD, the FTD and the BPC, where drift-chambers, multi-wire proportional chambers and silicon strip detectors are employed. It was operated inside a 1.16 T solenoidal magnetic field. A side view of the full tracking system and an azimuthal view of the CTD is shown in figure 3.3.
The innermost tracking device is the silicon track detector [118, 119] which directly surrounds the interaction region. It is subdivided into the two-layer barrel like *Central Silicon Track detector* (CST), and the disk like endcaps in backward (BST) [120] and forward (FST) [121] direction. The employed silicon strip pads achieve a spacial resolution in the r-ϕ-plane of 12 µm and of 22 µm in z-direction in the CTD [118] and enable a precise vertex determination. The most important components of the tracking system are the *Central Jet Chambers* (CJC1 and CJC2) covering an angular range of $20\,° < \theta < 165\,°$. The CJC1 and CJC2 consist of 720 and 1920 sense wires, respectively, which are spanned in parallel to the beam line. The spacial resolution reaches 170 µm in the r-ϕ-plane. The CJC is complemented by the *Central Outer Z chamber* (COZ), which is situated between the CJC1 and CJC2. It consists of 24 drift chambers with radially oriented wires and reaches a z-position resolution of 200 to 260 µm. The *Central Inner* [122] and *Outer Proportional Chambers*

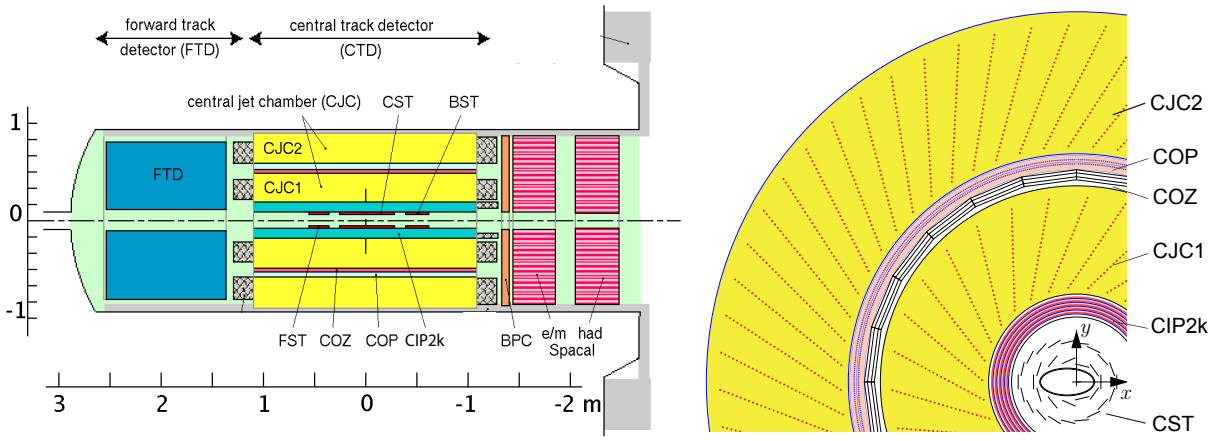


Figure 3.3: Side view (left) and azimuthal schematic drawing (right) of the H1 tracking system [117]. The most important parts for the track finding are the Central Jet Chambers (CJC1 and CJC2) with improved z-resolution from the COZ as well as the silicon trackers (FST, CST and BST). The Central Inner Proportional chamber (CIP) was optimized for triggering purposes.

(CIP2k and COP) are mounted inside the CJC1 and CJC2, respectively, and were optimized for triggering. The CIP2k determined the z-position of the event vertex after 2.3 µs with a z-resolution of 16 cm [123]. The trajectories of charged particles are measured with a transverse momentum resolution of $\sigma(p_T)/p_T \approx 0.2\,\% p_T/\text{GeV} \oplus 1.5\,\%$ [124].
The *Forward Track Detector* (FTD) [125] is built from three supermodules, containing four to five drift chambers of different wire geometries. The typical one hit resolution is 210 µm in the r-ϕ plane. However, the performance of the FTD suffers from the CTD in front of it and results in a track finding efficiency of approximately 70 % in each supermodule.
The tracking system is completed by the *Backward Proportional Chamber* (BPC) in front of the SpaCal. It consists of six wire layers and helped to discriminate charged and neutral particles entering the SpaCal.

3.2.4 The calorimetry system

The H1 calorimetry system is subdivided into the LAr [126] and the SpaCal [116], with both systems consisting of an electromagnetic (em) and an hadronic (had) part. The aim of the calorimetric system is to provide a precise energy measurement of single particles and of jets as well as to generate trigger signals. This is achieved by stopping the incident particles and converting the deposited energy into an electronic signal.

3.2.4.1 The Liquid Argon Calorimeter

The Liquid Argon Calorimeter (LAr) covers the polar angular range of $4\,^\circ < \theta < 154\,^\circ$ and the full azimuthal range (see figure 3.4). The entire calorimeter is located in a vessel filled with liquid Argon at a temperature of 90 K and is surrounded by a superconducting solenoid. The LAr was designed as a sampling calorimeter, consisting of passive layers made of a heavy high Z material for the generation and absorption of particle cascades (showers), and active layers are for the signal generation through ionization and charge collection. The LAr consists of an electromagnetic part of approximately 20 to 30 radiation lengths, using lead as absorber plates, and an hadronic part using stainless steel as absorber plates. The LAr is structured in seven layers f_0 to f_6 and subdivided in about 45 000 individual calorimeter cells, which were grouped into 256 big towers. The big towers

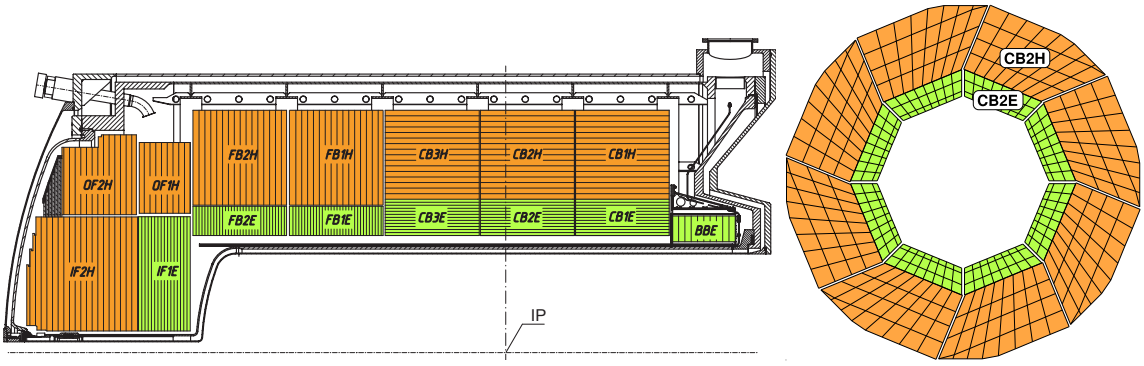


Figure 3.4: Side view (left) and azimuthal schematic drawing (right) at the interaction region of the H1 Liquid Argon Calorimeter. The LAr is segmented into eight wheels, where the five forward wheels (IF, OF, FB1, FB2) and the backward barrel (BBE) had the absorber planes perpendicular to the beam direction, while the central barrel wheels (CB1, CB2, CB3) had absorber plates parallel to the beam axis. Each wheel consists of eight sections in ϕ-directions.

delivered level one trigger signals, if the deposited online energy was above 4.8 GeV for big towers of the BBE or CB, or above 6.2 GeV or 25.6 GeV for big towers located in the FB or the IF, respectively. This enabled to deliver trigger signals from the scattered lepton as well as from the hadronic final state.

The LAr is a non-compensating calorimeter, i.e. with a different response to electromagnetically induced showers compared to hadronic showers, and therefore these two shower types have to be classified and calibrated individually, where a hadronic shower can have also a fractional electromagnetic shower.
Neighboring cells with energy deposits, which are not classified as noise, are grouped together into clusters [127]. The clusters undergo a software compensation procedure [126, 128, 129] to account for the different response of hadronic and electromagnetic clusters. For electromagnetic showers, the reconstructed energy is directly proportional to the collected charge and the calibration is expressed by a the charge-energy conversion factor for each calorimeter wheel. The intrinsic resolution of the electromagnetic part of the LAr (EMC) was determined from electron test beam measurements for an FB module to $\sigma(E_{\mathrm{em}})/E_{\mathrm{em}} = 11\,\%/\sqrt{E} \oplus 0.15\,\mathrm{GeV}/E \oplus 0.6\,\%$ [123].
The energy scale of hadronic showers is obtained by applying a software compensation and calibration procedure. The reconstructed energy of a cell is weighted with respect to the cell energy at the electromagnetic scale by a constant and a cell-energy-density dependent exponential term. The intrinsic resolution achieved for charged pions was determined from test-beam measurements to $\sigma(E_{\mathrm{had}})/E_{\mathrm{had}} = 55\%/\sqrt{E} \oplus E \oplus 1.6\,\%$ [126].

3.2.4.2 Determination of the electromagnetic fraction in showers in the LAr

The determination of the energy of hadronic showers is further improved, by a determination of the electromagnetic fraction of each cluster. A neural network is employed [130], which takes various variables about the shower shape into account. These variables are derived from the geometrical information and the measured energy of the individual cells. The most valuable discriminants are the cluster energy fraction in the third layer f_3 and the covariance of the longitudinal and radial moments c_{LR} of the the cluster. Also the energy of each cluster is valuable input to the neural net. Neural networks are trained for each calorimeter wheel and for individual running periods individually. For the trainig it

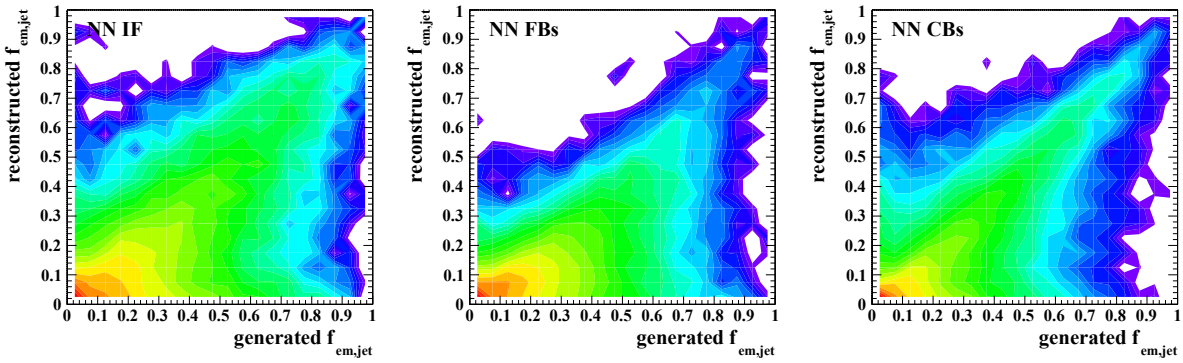


Figure 3.5: Classification of the electromagnetic fraction of jets using the neural networks compared to some comparable generated electromagnetic fraction in three wheels of the LAr [130]. The color code shows the very good linear classification capability also for complex objects, like jets, of this neural network approach.

is assumed, that single photons and neutral pions are incident particles inducing a pure electromagnetic shower, and charged pions are particles inducing hadronic showers. Three separate networks for different energy regimes are combined for the final networks.
The neural network approach enables to determine the electromagnetic fraction f_{em} for each cluster on an event-by-event basis and shows an excellent performance for determining pure electromagnetic clusters [130]. The classification capability of clusters belonging to individual jets is studied by defining the generated electromagnetic fraction of jets from its (electromagnetic) constituents by calculating

$$f_{\mathrm{em}}^{\mathrm{gen}} = \frac{1}{E_{\mathrm{jet}}^{\mathrm{gen}}} \sum_{\{e,\gamma\}} E_{\mathrm{em}}^{i} \ . \tag{3.3}$$

This fraction specifies all energies deposited by photons or electrons/positrons within a jet as electromagnetic. The reconstructed electromagnetic energy fraction $f_{\mathrm{em}}^{\mathrm{rec}}$ is defined by

$$f_{\mathrm{em}}^{\mathrm{rec}} = \frac{1}{E_{\mathrm{jet}}^{\mathrm{rec}}} \left(\sum_{\mathrm{em\ cand}} E_{\mathrm{cand}} + \sum_{\mathrm{cls}} P_{\mathrm{em}}^{i} E_{\mathrm{cls}}^{i} \right) , \tag{3.4}$$

where the first sum runs over all particles found by the electron finder[4] and the second one over all clusters which are associated to this jet. The probability P_{em}^{i} of each cluster is determined from the neural nets. The neural network approach shows an almost perfect linear classification capability, when comparing these reconstructed and generated variables (compare figure 3.5). The output of the neural network is input for the determination of precise calibration constants for the LAr. These constants can be determined directly from data due to the over-constrained kinematics of NC events. The calibration data sample of single jet neutral current events is statistically independent of the data sample of the jet measurements in this analysis.

[4]The electron finder removes the scattered electron and its clusters from the HFS.

3.2.4.3 The SpaCal

The so-called *Spaghetti Calorimeter* (SpaCal) [116] is a lead/scintillating-fiber calorimeter in the backward region of H1, where it covers the angular range $153\,° < \theta < 177.5\,°$. It consists of a separate electromagnetic and hadronic section (see figure 3.3) and was mainly designed for a precise measurement of the energy and direction of the scattered electron for low-Q^2-events and was optimized for a good electron/pion separation power [131].
The electromagnetic section is built from scintillating fibers with a diameter of 0.5 mm embedded into a lead matrix and corresponds to a total of 28 radiation lengths. Incident particles are initiating particle showers and the shower particles caused scintillations in the plastic fibers. 2340 fibers were grouped together in $4 \times 4\,\mathrm{cm}^2$ cells and conduct the light from the shower to photo-multiplier tubes (PMT). The small fiber diameters and the high granularity of the read-out cells result in an excellent electromagnetic energy resolution of [132] $\sigma(E)/E = 7\,\%/\sqrt{E/\mathrm{GeV}} \oplus 1\%$.

3.3 The luminosity measurement

The integrated luminosity $\mathcal{L}$ is the proportionality between the (experimental) event rate and the cross section σ for a given process:

$$N = \mathcal{L}\sigma \ . \tag{3.5}$$

It was determined for present running conditions from a precise measurement of a theoretically well known process. The H1 luminosity was determined by measuring the Bethe-Heitler-process $ep \to ep\gamma$ [133], where the photon is radiated under a very small angle w.r.t. the electron beam axis $\lesssim 0.45\,\mathrm{mrad}$. The photon is measured using the Photon Detector (PD), a Čherenkov sampling calorimeter with tungsten absorbers [116], which is located at a distance of $z = -102.9\,\mathrm{m}$. The electron is determined for calibration and cross checks using the Electron Tagger (ET), placed at $z = -5.4\,\mathrm{m}$, adjacent to the beam-pipe.
A more precise determination of the luminosity was performed offline, where the Bethe-Heitler event-rate is corrected for background from bremsstrahlung events in beam-gas interactions $eA \to eA\gamma$. Also the efficiency and acceptance of the PD and the correction for pile-up events were taken into account. The beam-gas interaction rate is estimated from pilot-bunches.

The overall normalization for the HERA-II running period is determined using a precision measurement of the QED Compton (QEDC) process [114]. This process enables an independent determination of the integrated luminosity, but the small cross section compared to the Bethe-Heitler process provides only little time-dependent information with the sufficient precision. QED Compton events are defined by two deposits in the electromagnetic section of the SpaCal which are back-to-back in the azimuthal angle ϕ. The position information is determined using the CIP chambers. The result of the QEDC analysis yields an uncertainty of 2.3 % on the integrated luminosity for the HERA-II period and a small positive correction w.r.t. the Bethe-Heitler measurement. The precision is limited by systematic uncertainties of the SpaCal energy scale and resolution, the non-elastic QEDC background and theory uncertainties from missing higher orders [114].

3.4 The trigger system

The bunch crossing rate of 10.4 MHz at HERA by far exceeds the maximum reachable storage rate of H1 of $\sim$ 10 Hz. Hence, a four level trigger system is used to downscale the frequency by selecting only relevant events and by downscaling the frequency of abundant event types. A schematical illustration of the H1 trigger system is shown in figure 3.6.

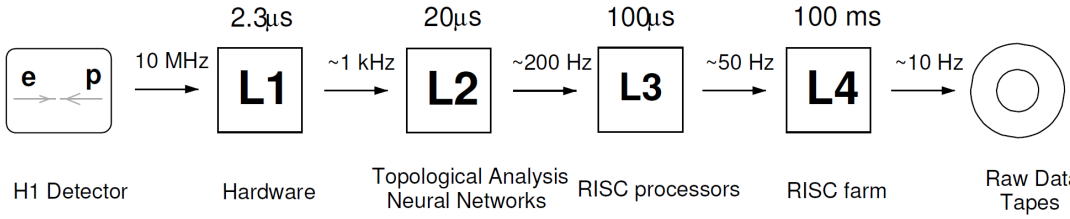


Figure 3.6: Schematic illustration of the data pipeline in the H1 trigger system. The collision rate of 10.4 MHz is reduced successively to the data storage rate of 10 Hz. The typical rates and the dead-times of each subtrigger system is shown.

The first level trigger *L1* is based on the information of 256 trigger elements of several detector components. These are processed practically without dead-time to 128 raw subtriggers. If one of the raw subtriggers is activated, the pipelines from the subdetector are read out and a dead-time of $\sim$ 2.3 µs accumulates. Some of the L1 raw subtriggers with very high repetition rate are prescaled by a factor of n and generate only a positive trigger signal in every n-th occurrence.
The second level trigger *L2* consists of two independent subtriggers. A topological trigger (L2TT) founds a decision based on topological event signatures out of some subdetector signals. A neural network trigger (L2NN) [134] is running on a massively parallel processor architecture. The networks were trained with samples from charged current, neutral current and other special events, using information from all L1 subtrigger systems (CTD, LAr, SpaCal). The L2 subtriggers provide high efficiency and good background suppression of non-*ep* background by reducing the rate to $\sim$ 200 Hz. If the L2 subtrigger give a positive signal, roughly 2700 channels of the H1 detector are read out with a dead-time of around 20 µs.
The third level trigger *L3* was commissioned in 2005. This Fast Track Trigger (FTT) [135, 136] finds a decision based on tracks and was optimized for finding heavy quark decays. The FTT reduces the output rate to $\sim$ 50 Hz with a latency time of 100 µs.
The fourth level trigger (L4/L5) performs a complete event reconstruction. All events are classified and non-*ep* events are rejected, apart from 1 % for monitoring purposes. Some high rate categories are further downscaled to achieve a 5-10 Hz storage rate, where raw data from all detector components are written to the Production Output Tapes (POT). Also the reconstructed quantities are written to the Data Summary Tapes (DST). The dead-time ends after $\sim$ 100 ms and the pipelines are restarted.

3.5 Event generators and detector simulation

The method of using Monte Carlo events together with a detector simulation is widely used to control physics measurements. It is applied for the precise understanding of the response of complicated experimental setups, like e.g. the resolution of the various detector components, and for the estimation of systematic uncertainties. In cross section measurements it is applied to determine the acceptance, purities and the stability of the measurement as well as the background contribution from other processes.
In this analysis MC event generators are used together with the simulation of the H1 detector for the determination and optimization of the migration matrix in the unfolding procedure, as well as for the estimation of background contamination. Furthermore, the corrections for higher-order electroweak effects and from the hadronization process are determined with MC generators.

3.5.1 Monte Carlo event generators

Monte Carlo (MC) event generators are computer programs to simulate high energy physics collisions. They create particle collision events based on various subprocesses and phenomenological models. Most Monte Carlo event generators are based on perturbatively calculated matrix elements in leading-order for the hard scattering process. Softer contributions to the events are simulated by parton showers, evolved from the final as well as the initial state partons (see section 2.3). The generation of hadrons is performed by applying hadronization models, like e.g. the Lund String model or the cluster fragmentation model (see section 2.3). The simulated events consist of a list of four vectors of partons and particles and can be processed by the detector simulation, which also simulates the decay of long-lived hadrons.

3.5.1.1 RAPGAP

The Rapgap event generator [137] is used for simulating NC DIS signal events. It is based on LO matrix elements for γ^* and Z^0 exchange for NC DIS or order $\mathcal{O}(\alpha_{em}^2)$ and $(\mathcal{O}(\alpha_{em}^2\alpha_s))$ (c.f. figure 2.12) matched with leading-order DGLAP parton showers. The final state partons are passed to the Lund string fragmentation model as implemented in the package JETSET [138]. Higher order QED radiative effects including electroweak effects, like radiation and virtual corrections, are modeled using the HERACLES program [139].

3.5.1.2 DJANGOH

The Djangoh event generator [140] is also used for simulating NC DIS events. It is based on LO matrix elements from the LEPTO generator [141] and is interfaced to the color dipole model (CDM) [142] for the generation of parton showers (see section 2.3.1.2), as implemented in ARIADNE [143, 144]. QED effects are implemented through the HERACLES program, and the hadronization is performed by the JETSET program.

3.5.1.3 LEPTO

The program LEPTO [141] is an event generator based on LO matrix elements and DGLAP parton shower. QED effects are simulated using the HERACLES program and the hadronization is simulated using the JETSET program. The LEPTO program includes all electroweak processes in leading-order DIS. It is used to determine the contribution to the

cross section arising from Z^0 exchange. This has to be corrected for, since NLO calculations do not include such effects.

3.5.1.4 SHERPA

The multi-purpose event generator SHERPA [145] is based on multi-leg tree-level matrix elements. In this analysis it is applied to determine the uncertainty of the hadronization correction, which is defined by the difference between the correction factor derived by applying the Lund string model or the cluster fragmentation model to the parton level event (more details in section 7.2.5.1).

3.5.1.5 PYTHIA

The multi-purpose event generator PYTHIA [146, 147] is used to determine contributions to the selected events which arise from photo-production processes (i.e. $Q^2 \to 0\,\mathrm{GeV}^2$). These may contribute to the cross section due to a wrongly identified scattered lepton and have to be subtracted as background contribution in this analysis.

3.5.1.6 COMPTON

The COMPTON generator [148, 149] is used to simulate the background contribution arising from the quasi-real QED Compton processes $ep \to e\gamma X$. These background contributions are discriminated by requiring two or more tracks or hadronic activity in the found jets, where however a small fraction contributes still from wrongly identified jets to the event selection with around 0.3 %.

3.5.1.7 GRAPE

The contribution from lepton-pair production $ep \to e\ell^+\ell^- X$ is estimated using the GRAPE program [150]. The contribution is found to be negligible with a probability of less than one predicted event in the full sample.

3.5.2 Simulation of the experimental setup

The events generated by the Monte Carlo event generators are processed through a simulation of the H1 detector, which is shown in figure 3.2. The GEANT3 [151] based H1Sim [152] package simulates the interaction of particles with matter and simulates the detector response. For an efficient simulation, electromagnetic subshowers in the LAr are generated, using the Gflash module [153, 154], where showers, including their longitudinal and lateral fluctuations and correlations, are parametrized. The actual running conditions during the data are taking into account in the simulation. The data format of simulated events is identical to real data. The events created by H1Sim are treated the same way as data, using the same analysis chain and the same reconstruction methods.

Chapter 4

Event reconstruction and data selection

This chapter summarizes the event recording, reconstruction and the determination of observables. This work benefits from extensive studies using H1 data on reconstruction methods and optimizations in the same or very similar phase space for jet measurements (e.g. [74, 79]) and NC DIS measurements (e.g. [155–157]) which were performed in many previous works. This analysis is based on work described in ref. [130], where also the improved calibration of the LAr is outlined. A significant extension to [130] is the consideration of an extended phase space in order to better take into account migrations arising from limited resolution of the detector. The final measurement phase space, however, is identical to the one from the previous analysis. The chapter will conclude with control distributions for the relevant observables of this analysis.

4.1 Event reconstruction

4.1.1 The trigger

In this analysis the trigger S67 is used. It combines logically the three L1 trigger conditions of the LAr calorimeter t_{LAr} (see section 3.2.4.1), a timing requirement t_{T0} and a veto condition t_{VET}:

$$t_{\mathrm{S67}} = t_{\mathrm{LAr}} \wedge t_{\mathrm{T0}} \wedge t_{\mathrm{VET}}. \tag{4.1}$$

The LAr subtrigger t_{LAr} is designed such that it can be triggered by both, the scattered electron and the HFS, which results in an almost perfect efficiency. Only some small localized regions have reduced efficiencies due to malfunctioning hardware or calorimeter clusters with high noise levels. These regions are removed by time-dependent fiducial volume cuts in the ϕ-z plane [156] from the analysis. For this analysis, the efficiency of the t_{LAr} was determined to be flat with values close to 100 % [156] and larger than 99.5 % [130] for electron energies above $E_e > 13\,\mathrm{GeV}$. To avoid a low efficiency of the LAr trigger, an electron energy of above $E_e > 11\,\mathrm{GeV}$ is required. In the region of $11 < E_e < 13\,\mathrm{GeV}$, where the efficiency after the fiducial volume cut is around 99 %, the MC events are reweighted for this effect [130].
The timing information t_{T0} is provided by either the CIP, or the LAr, or by both, with an efficiency above 99.5 % [130, 156].
The trigger condition t_{S67} includes also a veto condition t_{VET} to suppress non-ep background from beam halo or beam-gas interactions. It takes information from the ToF counters, the

CIP, the Tail Catcher and since 2006 also from the muon chambers into account. The efficiency of the veto condition was proved to be stable over time and a total veto efficiency of 99.3 % was determined for the HERA-II period [130,156] and the MC events are corrected for this.
The efficiency of the S67 subtrigger t_{S67} is determined by the product of the efficiencies of the trigger components to be 99.0 % and is found to be flat in Q^2 and y. After corrections for the inefficiencies, it is well described by the MC, and a systematic uncertainty of 1.2 % is assigned [130].

4.1.2 Track and vertex finding

The individual hits in the tracking detectors have to be combined to form tracks, which represent the trajectories of charged particles, in order to find the primary interaction vertex, assist in the electron identification, and determine the momenta of the charged particles, which are also input for the definition of jets.

4.1.2.1 Alignment

The alignment parameters of the individual sub-detectors are determined for each run [158] using the millipede algorithm [159]. The relative alignment of the LAr w.r.t. the tracking detectors are described by determining ten position parameters for the calorimeter [160], which are determined from a neutral current event sample using scattered electron tracks, with COZ and CST information, extrapolated to the LAr surface [157].

4.1.2.2 Track reconstruction

The track finding combines the hit-information of the tracking detectors in order to find the trajectories of charged particles, which ideally follow a helix description for a homogeneous magnetic field in z-direction. The H1 track finding algorithm [115, 161] starts from approximate track parameters using hits from the CJC, FST and BST, with the algorithm adding hits successively to determine the five helix parameters in a fast helix fit. The hits which are associated with a track[1] are input for a broken line fit [162]. The broken line fit is repeated[2] with hits from the CIP, COZ and CST taken into account additionally [157]. The broken line fits accounts for deviations from the ideal helix trajectory arising from magnetic field inhomogeneity, continuous energy loss along the trajectory and multiple scattering effects.

4.1.2.3 Vertex position

The reconstructed tracks are used to determine the position of the primary event vertex and the event timing T_0. The vertex is determined from a simultaneous constrained fit [161] to all primary tracks, where tracks with large χ^2 contributions are removed. Iteratively, the vertex position is adjusted by adding CST hits to the tracks. The primary vertex position is then determined in a simultaneous constrained fit of all CJC and CST hits[3] [161]. The event timing T_0 is determined as the weighted mean of the accepted primary tracks [161].

[1] Hits which are not associated with a track in the fast helix fit, are assumed to be noise and are rejected.
[2] The resulting tracks are labeled DTNV, according to the name of their storage bank and the fact, that they are *non-vertex-fitted tracks*.
[3] The fits are repeated with the knowledge of the vertex position into account. The resulting tracks are denoted DTRA tracks or named *vertex-fitted tracks*.

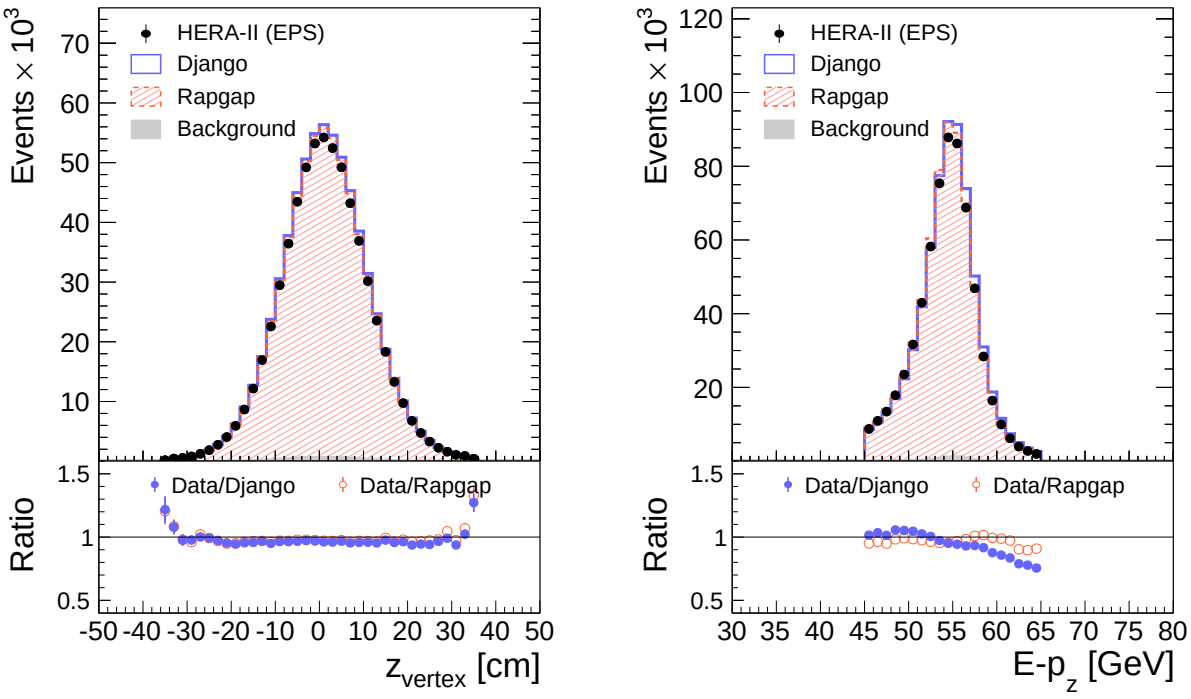


Figure 4.1: Distribution of the vertex z-position in the selected range (left) and the longitudinal momentum balance $E - p_z$ (left) for the selected NC DIS sample. The z_{vertex}-distribution for the MC simulation is estimated by a superposition of Gaussian distribution, and is further slightly reweighted for an accurate agreement with data.

The found vertex position is checked to be compatible with the electron track, since a high track multiplicity could also result from nuclear interaction of the hadronic final state with the beam pipe and can thus mimic a wrong primary vertex position. The so-called NC optimal-vertex procedure [157] improves the efficiency to find the correct vertex from 90 % [155] to almost 100 % [157].
The vertex x and y position are determined very precisely for each run, due to the intrinsically small beam sizes particularly in y. The z-position of the interaction point varies due to the large longitudinal beam sizes: $\sigma_z^p \approx 13\,\text{cm}$ for the proton beam and $\sigma_z^e \approx 2\,\text{cm}$ for the electron beam. For events to be accepted, the vertex z-position is therefore required to be within approximately three standard deviations, which corresponds to $-35 < z_{\text{vertex}} < 35\,\text{cm}$. The measured and simulated vertex z-positions are shown in figure 4.1.
In the simulation, the vertex z-position is estimated by a superposition of Gaussian distributions with parameters, which are determined from data for each run separately. For a very accurate agreement between data and simulated events the vertex z-position of MC events is further slightly reweighted [157].

4.1.3 Electron identification

The neutral current DIS process is identified by finding the scattered electron in the final state. In the desired phase space of $Q^2 \geq 150\,\text{GeV}^2$, the electron is mostly scattered into the acceptance of the LAr. However, at this lower phase space boundary the acceptance decreases dependent on the vertex z-position. Insensitive regions of the calorimeter are excluded from the analysis. These are the azimuthal cracks in between LAr octants and radial cracks at the z-position between the CB2 and CB3 wheels.

The scattered lepton in this analysis is found in a two step procedure. First, clusters of electron candidates in the LAr are identified by the QESCAT algorithm[4] [163, 164]. The efficiency of the algorithm was determined from data to be better than 99.5 % using an independent track-based electron finder [130] and is largely independent of the electron energy and the polar angle, and it is well described by the Monte Carlo. The uncertainty of the electron finding efficiency is assigned to be 0.5 % in the central direction and 2 % in the forward region, since the track-based algorithm is not applicable there, and the efficiency can only be determined from MC simulations [165].
In a second step a track has to be assigned to the electron candidate cluster. A vertex fitted track (DTRA track) is associated to the electron cluster, if the extrapolated track trajectory into the LAr has a distance of closest approach $d_{\mathrm{ca}} < 8\,\mathrm{cm}$ w.r.t. the center of gravity of the cluster. If no vertex fitted track is found, a non-vertex fitted track (DTNV) may be used instead, if it is within the appropriate distance. These requirement reduce the amount of wrongly identified scattered electrons to 0.3 % [130].
The track-cluster-link procedure for the electron identification is closely related to the vertex finding efficiency, due to the usage of vertex-fitted tracks. The efficiency of finding a primary vertex and a track-associated cluster is found to be about 95 %. The Monte Carlo has to be corrected for 0.5-2 %, using θ-dependent correction factors, which are derived for each run-period separately [130]. A systematic uncertainty of 1 % is assigned to the vertex and electron-track requirements.

4.1.4 Reconstruction of the hadronic final state

Tracks and noise suppressed calorimetry information are input for the reconstruction of the four-vectors of the particles of the hadronic final state (HFS). The so-called Hadroo2 algorithm (Hadronic reconstruction in H1OO 2) [166] is employed[5]. Within the acceptance of the CJC, the transverse momentum of a jet[6] is on average composed of 45-50 % from calorimetric information. In the forward region, where the efficiency of the FTD is poor,

[4]The algorithm gathers energy deposits in a cone with a radius of 7.5 ° around prominent seed clusters in the electromagnetic part of the LAr. These clusters have to fulfill a number of estimator criteria constructed from the size, the shape and the energy deposit. The energy is required to be above 5 GeV in at least four cluster cells. The clusters must have an energy fraction in the em part of the LAr $f_{\mathrm{EMC}} > 0.94 + 0.05\cos(2\theta)$, and the energy deposit has to be localized, by demanding the energy fraction in the hottest neighboring cells to be above 0.8 (0.4) in the backward (forward) region (There are four cells summed up in the BBE part of the LAr, eight cells in the CB3 and FB, and twelve cells in the IF.). A compact cluster shape is demanded by requiring the cluster radius to be smaller than some certain θ-dependent value. An isolation criteria is further introduced, by demanding the energy deposits in an η-ϕ-cone with a radius of $R_{\eta\phi} = 0.25$ to be larger than 0.98 times the cluster energy, in order to have a distinct separation of the HFS from the scattered electron. The fraction is only 0.95, if less than 300 MeV are recorded inside the isolation cone in the hadronic part of the LAr. The cluster is flagged as isolated, if the energy deposit inside of a further enlarged cone of $R_{\eta\phi} = 0.5$ is smaller than 1.05 times the cluster energy. In case of multiple electron candidates, an isolated cluster with the highest p_{T} is selected as the scattered electron [155].

[5]The Hadroo2 algorithm composes the four-vector of an HFS particle of the information of either the track or the calorimetric clusters, depending on which of the two measurements yields a better estimate of the resolution. The measurement of the other system is discarded. Hence, at lower particle momenta, where the track bending radius is small, and thus the track information gives a higher accuracy compared to the calorimetric measurement, while the latter provides better resolution at larger energies. Both systems' resolutions meet at an energy of approximately 25 GeV. Calorimetric clusters, which could not be associated to a track, are also turned into HFS particles and are assumed to originate e.g. from neutral particles.

[6]See section 4.1.6 for the definition of the jet transverse momentum

the calorimetric fraction of the jet transverse momentum is up to 90 % [130].
The LAr clusters within an HFS object are corrected by calibration constants, which are determined from the neural net (compare section 3.2.4.2) and for a certain class of events from the momentum balance of the scattered lepton and the jet. These constants are determined from data for the four HERA-II periods and are dependent on the cluster energy and the cluster pseudorapidity and are applied separately to the electromagnetic and hadronic fraction of a cluster. Clusters, which are associated to a jet, receive a further correction dependent on the jet transverse momentum and the jet pseudorapidity [130].
By using tracks and the calibrated LAr clusters, the Hadroo2 energy flow algorithm achieves an overall jet energy scale uncertainty of better than 1 % and a resolution of the jet transverse momentum of 9-11 % [130].

4.1.5 Determination of the DIS kinematic variables

In deep-inelastic scattering the event kinematics is overconstrained, if the incoming beam energies (E_e^0 and E_p^0) and hence the center-of-mass energy are known, and only two kinematic variables out of Q^2, y, and x_Bj have to be determined. In this analysis the *electron-sigma* method (eΣ) [167] is employed for the determination of the kinematics, which was shown to be the most precise method in the phase space of this analysis [74]. It uses the electron method [168] to compute the virtuality Q^2, whereas the scaling variable x_Bj is determined from the sigma method [169].
The scattered electron four-vector is expressed in terms of the electron energy E_e, the azimuthal angle θ_e and the polar angle ϕ_e, using the calorimetric energy for E_e and the position measurement for θ_e and ϕ_e. The azimuthal angle θ_e is determined from the vertex-fitted track associated with the electron, otherwise from the position of the calorimeter cluster. The virtuality Q^2 of the process is then given by

$$Q^2 := Q_\mathrm{e}^2 = 4E_\mathrm{e}E_\mathrm{e}^0 \cos^2\left(\frac{\theta_\mathrm{e}}{2}\right). \tag{4.2}$$

By defining the hadronic longitudinal momentum balance of the HFS $\Sigma = \sum_h E_\mathrm{h} - p_{z,\mathrm{h}}$, where the sum runs over all objects except the scattered electron, a temporary inelasticity and virtuality of the process are measured through

$$y_\Sigma = \frac{\Sigma}{E - p_z} \quad \text{and} \quad Q_\Sigma^2 = \frac{E_\mathrm{e}^2 \sin^2\theta_\mathrm{e}}{1 - y_\Sigma}\,.$$

The quantity $E - p_z = \Sigma + E_\mathrm{e} - p_{z,\mathrm{e}} = 2E_\mathrm{e}^0$ is an alternative expression of the electron beam energy. The Bjorken scaling variable x_Bj is then expressed by

$$x_\mathrm{Bj} := x_\Sigma = \frac{Q_\Sigma^2}{s y_\Sigma}\,. \tag{4.3}$$

The inelasticity of the process is calculated by a mixture of the electron and sigma reconstruction methods by

$$y := y_{\mathrm{e}\Sigma} = \frac{Q_\mathrm{e}^2}{s x_\Sigma}\,. \tag{4.4}$$

The electron-sigma method benefits from the good electron resolution for the reconstruction of Q^2 and a low sensitivity of x_Bj to initial state radiation of the incident electron.

4.1.6 Construction of jet four-vectors

The four-vectors of all HFS objects are input to the jet-finding algorithm (see section 2.4), which is performed in the Breit frame of reference. Therefore, all HFS objects are boosted using a boost vector determined from the exchange boson. For reconstructing the boost variables, the electron-sigma method is better suited than the electron method, because of the lower sensitivity to initial state radiation of the electron [79]. A cut on $E - p_z$ and the upper phase space cut on y of 0.7 improves further the stability. Overall, the β-value of the boost achieves a resolution of around 2 % and the distribution is excellently described by the two employed MC models [130].
For the jet finding, the inclusive longitudinally invariant k_T-algorithm is applied with a jet size parameter of $R_{k_\mathrm{T}} = 1.0$, and the p_T-recombination scheme for combining objects is used. The algorithm determines the four-vectors of each jet found, and the transverse momentum of a jet in the Breit frame is denoted as $p_\mathrm{T}^\mathrm{jet}$ in the following.
To account for the acceptance of the LAr, the jet four-vectors are boosted back into the laboratory frame of reference using the inverse of the previously applied boost. The jets are restricted to a laboratory frame pseudorapidity of $-1.5 < \eta_\mathrm{lab} < 2.75$. The lower cut is motivated by the angular coverage of the LAr and the upper cut rejects jets with clusters close to the beam pipe. Jets are further required to have a significant transverse momentum in the laboratory frame $p_\mathrm{T}^\mathrm{lab} > 2.5\,\mathrm{GeV}$ to be measured accurately. Jets consisting only of one cluster are removed, since they may result from radiated photons and are often in close proximity to the scattered electron [130].

4.1.7 Longitudinal energy loss

The H1 detector cannot cover the entire 4π sphere, but has no acceptance in the region of the beam pipe. The longitudinal momentum balance $E - p_z$, where all HFS objects and the scattered electron are taken into account, gives a measure of the longitudinal energy loss in an event. This quantity should peak at twice the electron beam energy for completely and perfectly reconstructed events. The measured distribution for the NC event sample is shown in figure 4.1.
Here we require $E - p_z$ to be within

$$45 < E - p_z < 65\,\mathrm{GeV}\ , \tag{4.5}$$

which suppresses events from photoproduction processes, where the outgoing electron leaves under a very small polar angle $\theta_\mathrm{e} \approx 0\,^\circ$ and an HFS object was wrongly identified as the scattered electron. This constraint further suppresses events with hard initial state QED radiation, where e.g. a photon with an energy $E_\gamma > 5\,\mathrm{GeV}$ is radiated collinear to the beam.

4.2 Background rejection

The selected data consist of events of the desired process (so-called *signal events*), but also events from other *ep*-scattering processes (*ep background*) and from events which do not result from *ep*-scattering (*non-ep background*), e.g. from cosmic radiation or beam gas interactions. These two distinct background sources have to be suppressed by applying suitable cuts, and in case of the *ep*-background, the residual background contribution has to be estimated by MC and subtracted from data.

4.2.1 Non-*ep* background rejection

The requirement of a timing condition already at trigger level reduces the non-*ep* background significantly. A residual contribution from beam gas collisions, beam halo muons and muons from cosmic showers are rejected with dedicated background finders [170]. An event is rejected if the two commonly used beam-halo finders, ibg0 and ibg1, or if the cosmic muon finders ibg5, ibg6 and ibg7 are fired in a pairwise logical conjunction. A beam-gas event is assumed if the ibgam0 finder responds and a large particle multiplicity $N_{\text{part}} > 50$ is found. A study of the non-*ep* background finders was performed in [155] and established the remaining data to be essentially background free. An additional optimization and a detailed list of all finders is given in [74].

4.2.2 Background rejection from other processes

The main contributions in the selected data of other than neutral current DIS within the selected data are from QED Compton processes (QEDC), where the electron is wrongly identified. These events are simulated using the COMPTON generator. The contribution of QEDC events in the selected events are reduced by $\approx 50\,\%$ [130] by requiring $\cos(|\pi - \Delta\phi|) < 0.95$, where $|\pi - \Delta\phi|$ is the acoplanarity, with $\Delta\phi$ the azimuthal angle between the scattered electron and the photon. The remaining QED Compton events are estimated to contribute with $1.0\,\%$ to events in the measurement phase space. Other minor background contributions are NC DIS processes with low Q^2 and photoproduction events with a relative contribution of approximately $0.1\,\%$. Negligible amounts of background of below $0.01\,\%$ are found for lepton pair production, charged current processes and deeply virtual Compton scattering.

4.3 Reweighting of the MC models

A reliable correction of detector effects must be based on MC models which precisely describe the measured data. This can be ensured by a reweighting of the model predictions, using reweighting constants which are derived on detector level quantities from the comparison to data and are applied as function of generator level observables. Thus, by changing distributions on generator level, the reweighting changes the underlying model in order to improve the description of data. After each reweighting step a large number of distributions is checked to ensure that the simulation does not show unexpected behaviors and the overall description of data is improved. Compared to [130], the reweighing functions are determined from distributions in the extended NC DIS phase space of $100 < Q^2 < 30\,000\,\text{GeV}$ and $0.08 < y < 0.94$ and for jets in the range $-1.5 < \eta^{\text{jet}}_{\text{lab}} < 2.75$ and $3 < p^{\text{jet}}_{\text{T}} < 60\,\text{GeV}$.
A good description is particularly important for the bin-by-bin correction, since migrations in and out of the measurement phase space fully rely on the predictions from the simulation. Reweighting constants for observables which are directly considered in the unfolding process cancel according to equation 5.12 in the regularized unfolding. The employed MC models Django and Rapgap provide a reasonable prescription of neutral current DIS observables within an agreement of approximately $10\,\%$. However, the dynamics of jets coincide with the data only within approximately 10 to $30\,\%$, where especially variables related to high jet-multiplicities are demanding and can show even larger deviations from the data, since such event topologies are only generated through the parton shower.

Multiplicative reweighting constants are derived for Django and Rapgap separately in an iterative procedure. These are polynomials of degree 2 as function of $\log Q^2$ and determined individually for events with different jet multiplicites N_{jet}. Further reweighing functions are derived from the pseudorapidity and transverse momentum of the jet with the smallest and highest pseudorapidity in the event. Additional small corrections are applied for three-jet events as function of the sum of the pseudorapidities of the two leading jets and the pt of the third-leading jet [171, 172]. All these constants are determined as two-dimensional polynomials of degree 2 with either $\log p_{\text{T}}^{\text{Fwd}}$, $\log p_{\text{T}}^{\text{Bwd}}$ or $\log Q^2$ as the second observable. Events with exactly one jet further receive corrections dependent on $\eta_{\text{lab}}^{\text{Breit}}$ and trijet events as function of $\langle p_{\text{T}} \rangle_3$. All reweighting constants are applied to the weight of the event.
After applying the reweighting constants, particularly events with high jet multiplicities show an improved description of data. All control distributions displayed in this chapter are then obtained from the reweighted MC models.

4.4 Selection and phase space definitions

4.4.1 Run selection

This analysis considers the entire data which were recorded by H1 during the HERA-II running period. All runs[7] which are labeled as *good* or *medium* quality run, based on detector performance criteria as determined during the data taking, are considered. Furthermore, the relevant detector components LAr, CJC, CIP and ToF must be active. Runs with an insignificant integrated luminosity below $0.2\,\text{nb}^{-1}$ are rejected. The selected run ranges, separated for the four individual run periods, are given in table 3.1.

4.4.2 Phase space considerations

This analysis aims to determine cross sections in the identical phase space as in a previous work [130, 173], but introduces a regularized unfolding procedure for the correction of detector effects. Also other comparable analyses by H1 have the same phase space in Q^2 [102, 103] or in the jet momenta [101, 102].
The unfolding procedure, as explained in the next chapter, takes into account migrations into the phase space of the measurement, denoted *measurement phase space* (MPS). For the estimation of these migrations the phase space is extended in the variables which are subject to the unfolding procedure. This enlarged phase space is labeled *extended phase space* (EPS). It is partly associated with detector regions with poor acceptance or where detector components overlap. Therfore, the measured cross sections are considered to be reliable only in the reduced MPS. The MPS is a full subspace of the EPS and is constrained through more stringent requirements of the detector acceptance and by considerations on the resolution and reconstruction quality. Events in the EPS but not in the MPS may contribute to the measured cross sections through migrations, e.g. resulting from resolution effects, or in a minor way through the normalization constraints in the unfolding procedure. The MPS also respects constraints derived from phase space regions for which pQCD calculations are safe.
The event selection is performed in the EPS on detector level, whereas on hadron level the MPS is of interest for the finally quoted cross sections. However, for an accurate unfolding,

[7] A run is defined by a fill of lepton an proton bunches, which typically lasted up to several hours until the beam currents decreased significantly, and the beams were dumped.

also on hadron level the EPS is defined identically to the detector level EPS.
The inclusive neutral current event sample is selected such that it covers the largest accessible phase space as allowed by detector performance constraints.
The jet samples are then sub-samples of the neutral current DIS sample by applying cuts on jet quantities, e.g. for the dijet and trijet event sample, or by defining the inclusive jet sample by counting each jet.

4.4.3 Neutral current phase space

The neutral current events which are input to the unfolding procedure are selected in the EPS, which is defined by

$$100 < Q^2 < 40\,000\,\mathrm{GeV}^2$$
$$0.08 < y < 0.7$$

and no direct constraint on x_{Bj} is applied. The lower limit on Q^2 is due to the acceptance of the LAr and the requirement of the scattered electron to be identified in this detector component. The upper limit on y is motivated by the strong suppression of events with larger values of the inelasticity due to the requirement on the scattered electron energy $E_{\mathrm{e}} > 11\,\mathrm{GeV}$ for efficient triggering. The lower bound of the inelasticity is chosen such that scattering of the HFS into the very forward direction is suppressed, which improves the reconstruction of the boost vector. The control distributions of the NC DIS sample in the EPS are shown in figure 4.2.
The MPS of the NC DIS measurement has stronger constraints and is restricted to $150 < Q^2 < 15\,000\,\mathrm{GeV}^2$ and $0.2 < y < 0.7$, in agreement with previous measurements [102], which is preferred due to the electron acceptance in the LAr and the acceptance of the HFS in the forward direction.

4.4.4 Inclusive jet phase space

In the inclusive jet measurement, every single jet contributes to the measurement. Jets are searched in the Breit frame on detector level in NC DIS events in the EPS and every jet with a transverse momentum of

$$p_{\mathrm{T}}^{\mathrm{jet}} > 3\,\mathrm{GeV} \tag{4.6}$$

is counted. The jet four-momenta are boosted 'back' to the laboratory reference frame where the pseudorapidity of the jets are restricted to the region of

$$-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75\ . \tag{4.7}$$

The found jet-multiplicities of the NC DIS event sample in the EPS and MPS are shown in figure 4.3 for $p_{\mathrm{T}}^{\mathrm{jet}} > 3\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\mathrm{jet}} > 5\,\mathrm{GeV}$ respectively.
The MPS of the inclusive jet measurement is restricted to $7 < p_{\mathrm{T}}^{\mathrm{jet}} < 50\,\mathrm{GeV}$ in the pseudorapidity region of $-1.0 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.5$. The $p_{\mathrm{T}}^{\mathrm{jet}}$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ distributions for the EPS and the MPS are shown in figure 4.4.

4.4.5 Jet shape distributions

A valuable criterion for the quality of the Monte Carlo generator employed is obtained from the study of the jet substructure. The internal structure of a jet can be quantified by

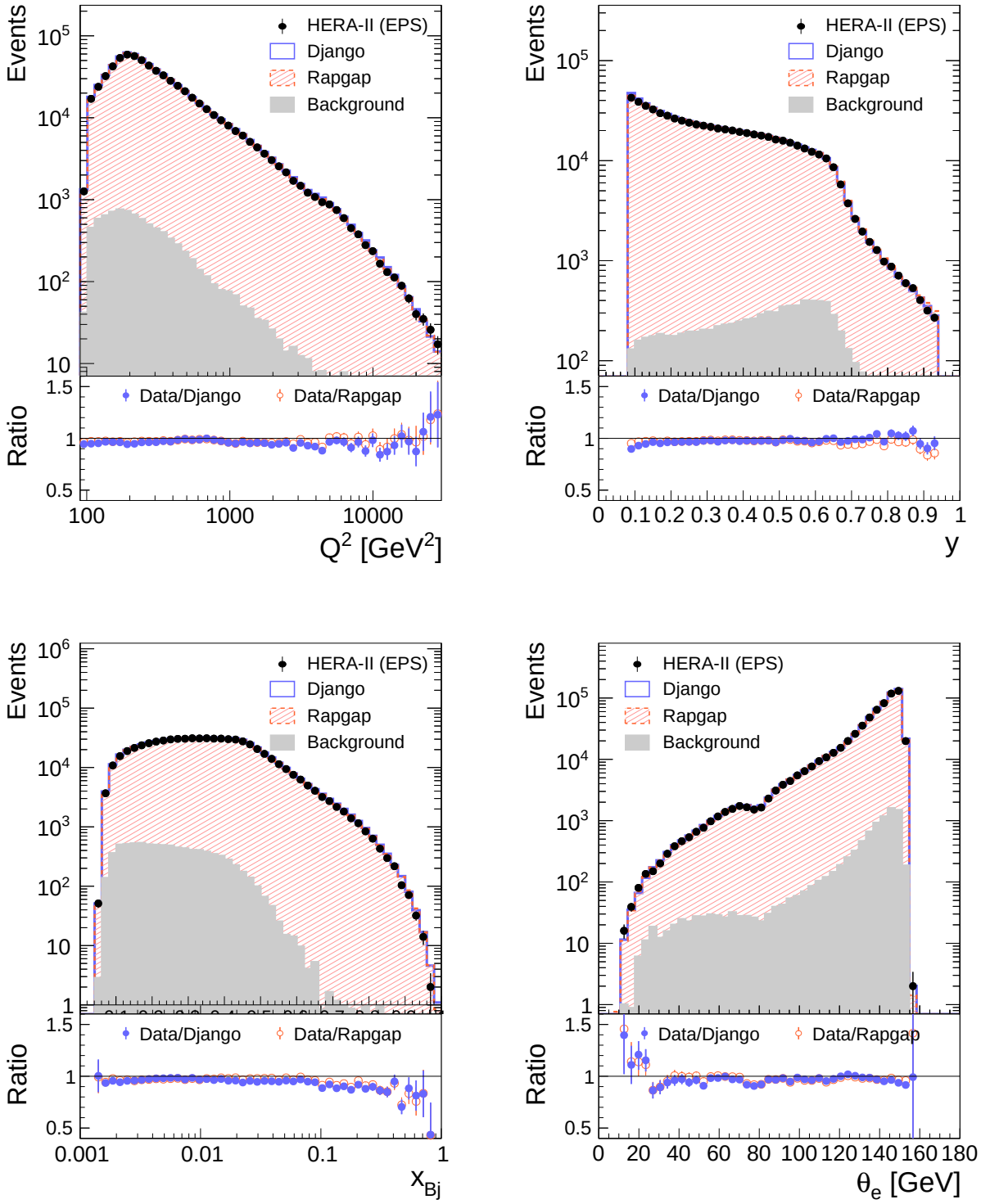


Figure 4.2: Control Plots of the NC DIS event sample in the extended phase space (EPS). The data are compared to the Monte Carlo predictions from Rapgap and Django, and also the sum of all *ep*-background contributions are shown. Due to the electron energy requirement of $E_e > 11\,\text{GeV}$ for a high trigger efficiency, the y distribution falls off steeply at $y \gtrsim 0.7$, and therefore these data are not used in the MPS.

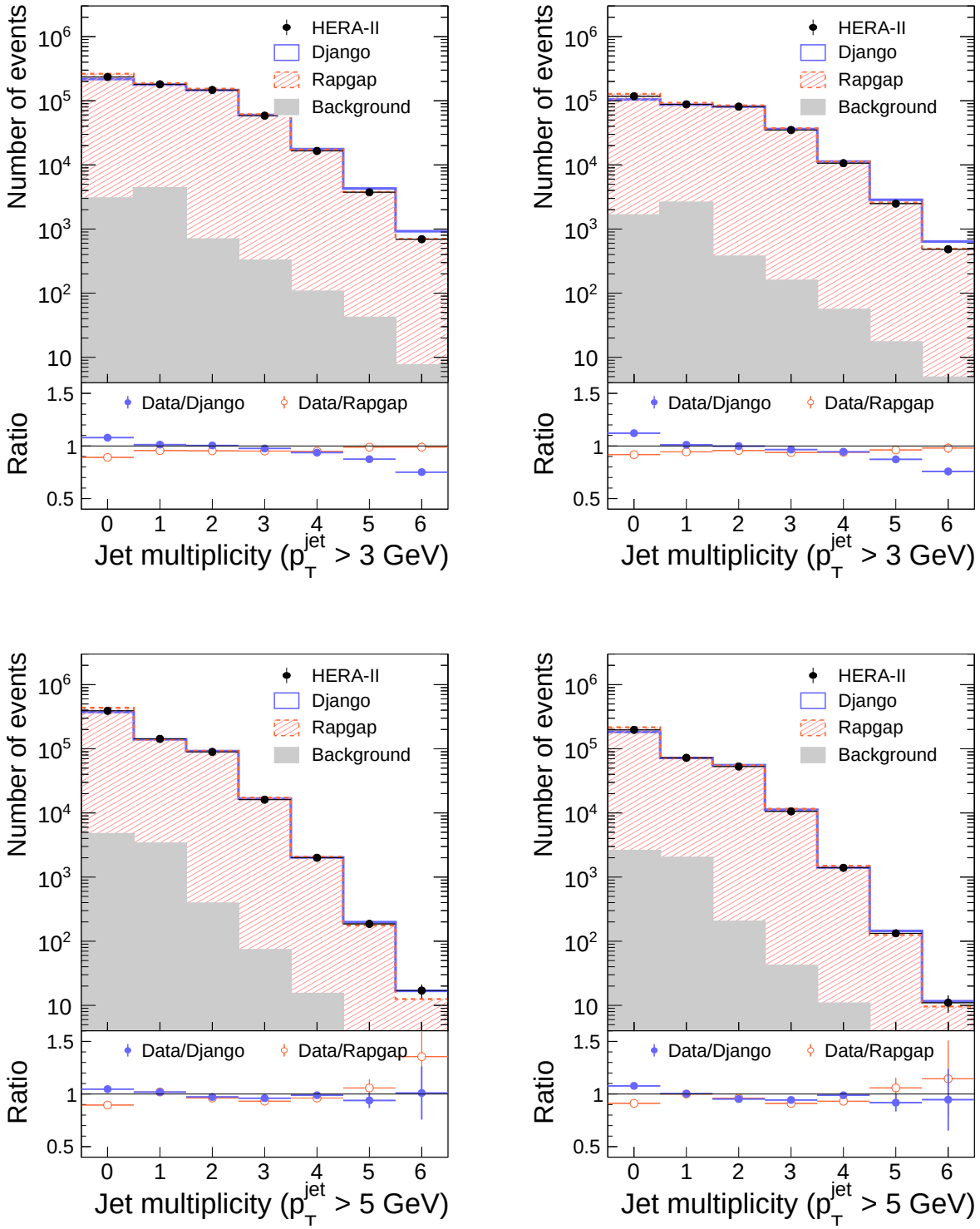


Figure 4.3: Jet multiplicities in NC DIS events in the EPS (left) and the MPS (right). The upper figures show the jet-multiplicities for minimum transverse jet momenta of $p_T^{jet} > 3\,\text{GeV}$, which refers to the EPS of the inclusive jet, dijet and trijet measurement. The lower figures show jet-multiplicities for $p_T^{jet} > 5\,\text{GeV}$, which refers to the MPS of the dijet and trijet measurements. The extension of the phase space increases the amount of dijet events by a factor of 4 and for trijet events by a factor of 8. This increases the stability of the measurement, due to the inclusion of migrations in the unfolding procedure.

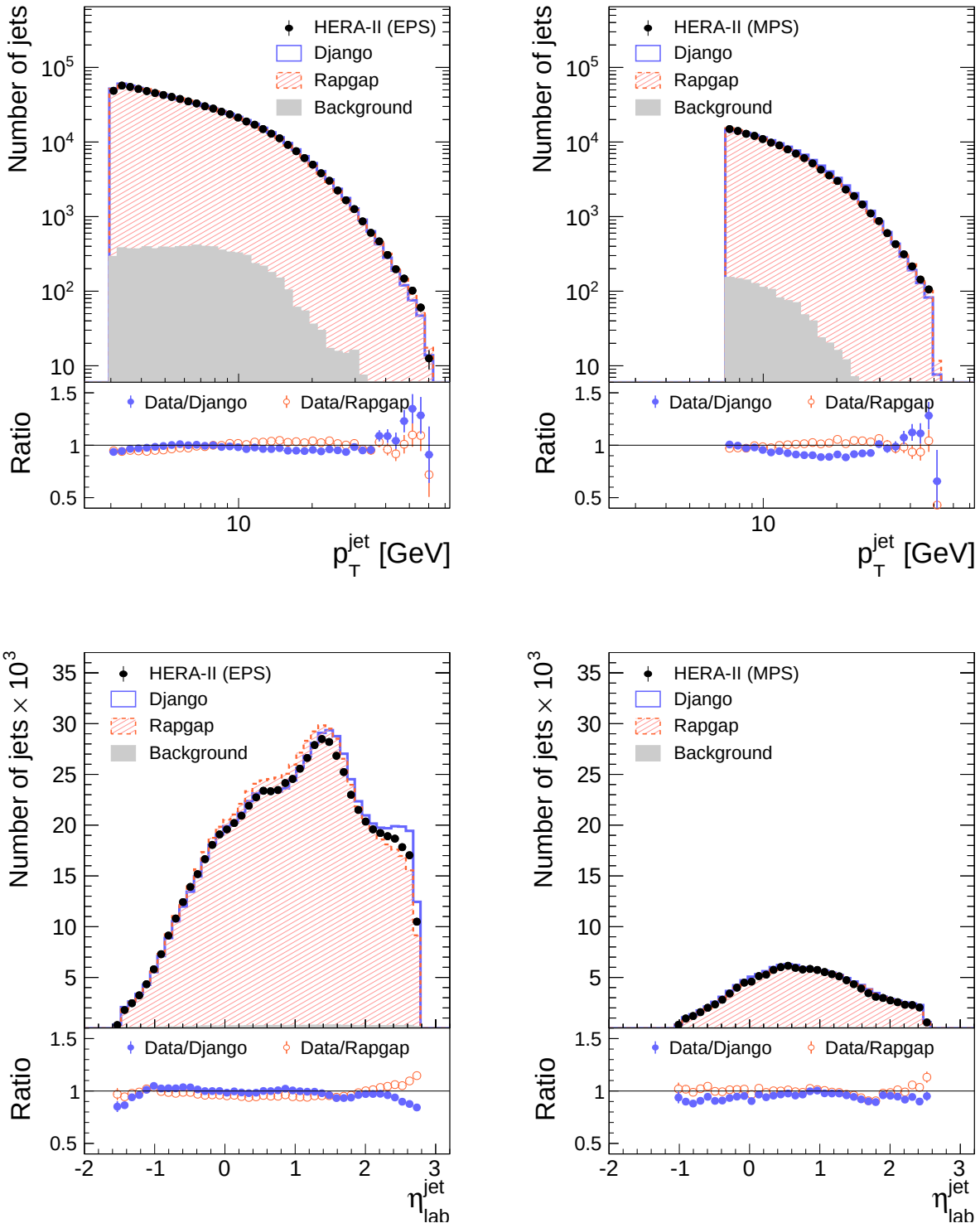


Figure 4.4: Control distributions of the inclusive jet measurement in the EPS (left) and the MPS (right). The upper figures show the jet transverse momentum in the Breit frame $p_{\mathrm{T}}^{\mathrm{jet}}$ and the lower plots the pseudorapidity in the laboratory frame $\eta_{\mathrm{lab}}^{\mathrm{jet}}$. The EPS in increased w.r.t. the MPS in $p_{\mathrm{T}}^{\mathrm{jet}}$ and in the NC DIS observables Q^2 and y. The steeply falling spectra of Q^2 and $p_{\mathrm{T}}^{\mathrm{jet}}$ lead to a strongly decreased number of jets in the MPS compared to the EPS, which becomes quite visible for $\eta_{\mathrm{lab}}^{\mathrm{jet}}$.

the *differential jet shape* $\rho(r)$, which characterizes the broadening of a jet. The observable[8] $\rho(r)$ is defined as the fraction of the jet transverse momentum, deposited within two cones[9] of radii r and $r + \Delta r$ around the jet axis relative to the total transverse momentum of the jet [176]

$$\rho(r) = \frac{\sum_i \hat{p}_{\mathrm{T},i}}{p_{\mathrm{T}}^{\mathrm{jet}}} \quad \text{with} \quad \hat{p}_{\mathrm{T},i} = \begin{cases} p_{\mathrm{T},i} & \text{if } r_i \in (r, r + \Delta r) \\ 0 & \text{else} \end{cases} . \tag{4.9}$$

The sum runs over all objects i contributing to the jet [175], with their distance from the jet axis defined in the Breit frame by $r_i = \sqrt{\Delta\eta^2_{\mathrm{Breit},i} + \Delta\phi_i^2}$, where $\Delta\eta_{\mathrm{Breit},i}$ and $\Delta\phi_i$ are the distance to the jet axis in pseudorapdity and azimuthal angle, respectively. The *mean differential jet shape* is defined as

$$\langle\rho(r)\rangle = \frac{1}{N_{\mathrm{jets}}} \sum_{\mathrm{jets}} \rho(r) \ , \tag{4.10}$$

where the sum runs over all jets of the inclusive jet selection. The differential jet shape is sensitive to hard as well as soft processes. Close to the jet axis, the jet shape is dominated by soft collinear gluon emission, whereas at large distances from the jet axis, the jet shape is sensitive to large angle gluon emission which can be well calculated perturbatively by a leading-logarithm parton shower Monte Carlo simulation [177].

The jet substructure is potentially sensitive to the parton originating from the hard interaction, and gluon induced jets are expected to be broader and contain more particles [14, 178][10].

The mean differential jet shape for jets with transverse momenta of $7 < p_{\mathrm{T}}^{\mathrm{jet}} < 11\,\mathrm{GeV}$ and $18 < p_{\mathrm{T}}^{\mathrm{jet}} < 30\,\mathrm{GeV}$ are displayed in figure 4.5. Since in our case the k_{T}-algorithm is employed for the definition of the jet, which in general does not result in a circular shape in (η,ϕ) [180], also objects with distances larger than the jet radius parameter $r_i > R_0$ ($R_0 = 1$) can contribute to the jet. We find that the two MC generators employed describe the average jet substructure very well. The softer jets between $7 < p_{\mathrm{T}}^{\mathrm{jet}} < 11\,\mathrm{GeV}$ show a broader distribution $\rho(r)$. Narrower distributions are expected for jets with lower momentum, and more for gluonic induced jets than for quark induced jets. The jets predicted by Django with transverse momenta of $18 < p_{\mathrm{T}}^{\mathrm{jet}} < 30\,\mathrm{GeV}$ are slightly narrower than the jets from data.

[8] It is also common to define the *integrated jet shape* by [174]

$$\psi(r) = \frac{\sum_{i, r_i < r} p_{\mathrm{T},i}}{p_{\mathrm{T}}^{\mathrm{jet}}} \ , \tag{4.8}$$

where all HFS objects forming the jet within the subcone of radius r are considered in the sum. Since the integrated jet shape is calculated by integrating the differential jet shape and hence does not resolve particle production at large angles from the jet axis but is dominated from the soft collinear contributions at small values of r, this observable is not explicitly studied here.

[9] Due to the fixed cones, the definition of the jet shape observable is infrared unsafe in a perturbative expansion, but is infrared safe in a resummed calculation [175].

[10] The distinction between gluon or quark induced jets on an event by event basis is not trivial, but could be measured statistically for instance in e^+e^- collisions [179], where it was also found that b-quark induced jets are broader than light-quark induced jets.

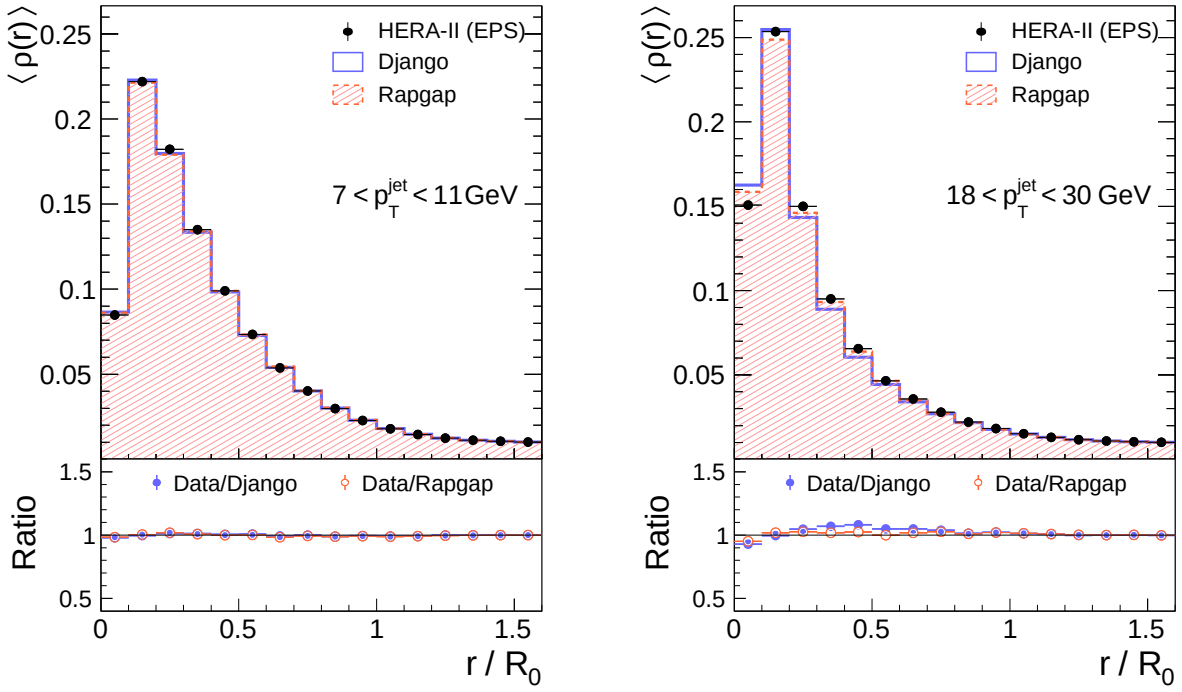


Figure 4.5: Mean differential jet shape $\rho(r)$ for inclusive jets with transverse momenta of $7 < p_{\mathrm{T}}^{\mathrm{jet}} < 11\,\mathrm{GeV}$ (left) and $18 < p_{\mathrm{T}}^{\mathrm{jet}} < 30\,\mathrm{GeV}$ (right).

4.4.6 Dijet phase space

4.4.6.1 The dijet extended phase space

A dijet event selection is performed, based on the selected neutral current DIS event sample. An event is labeled as *dijet event* if two or more jets are measured with a transverse momenta

$$3 < p_{\mathrm{T}}^{\mathrm{jet}} < 50\,\mathrm{GeV} \tag{4.11}$$

in the pseudorapidity region

$$-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75\ . \tag{4.12}$$

The neutral current kinematic variables of the dijet sample are shown in figure 4.6.
The jets are ordered according to their transverse momentum and the two jets with the largest $p_{\mathrm{T}}^{\mathrm{jet}}$ are labeled *jet1* and *jet2*, respectively, and are also referred to as *leading jet* and *sub-leading jets*, respectively. The dijet events are measured as a function of the *average transverse momentum* of the two leading jets

$$\langle p_{\mathrm{T}} \rangle_2 := \frac{1}{2}\left(p_{\mathrm{T}}^{\mathrm{jet1}} + p_{\mathrm{T}}^{\mathrm{jet2}}\right)\ . \tag{4.13}$$

The invariant mass of a dijet event is calculated using the four-vectors $p^{\mathrm{jet}i}$ of the two leading jets

$$M_{12}^2 = \left(p^{\mathrm{jet1}} + p^{\mathrm{jet2}}\right)^2\ . \tag{4.14}$$

These two distributions of the dijet observables are shown in figure 4.7 for dijet events.
Another characteristic variable of the dijet system is the observable ξ_2 defined as

$$\xi_2 = x_{\mathrm{Bj}}\left(1 + \frac{M_{12}^2}{Q^2}\right) = x_{\mathrm{Bj}} + \frac{M_{12}^2}{ys}\ , \tag{4.15}$$

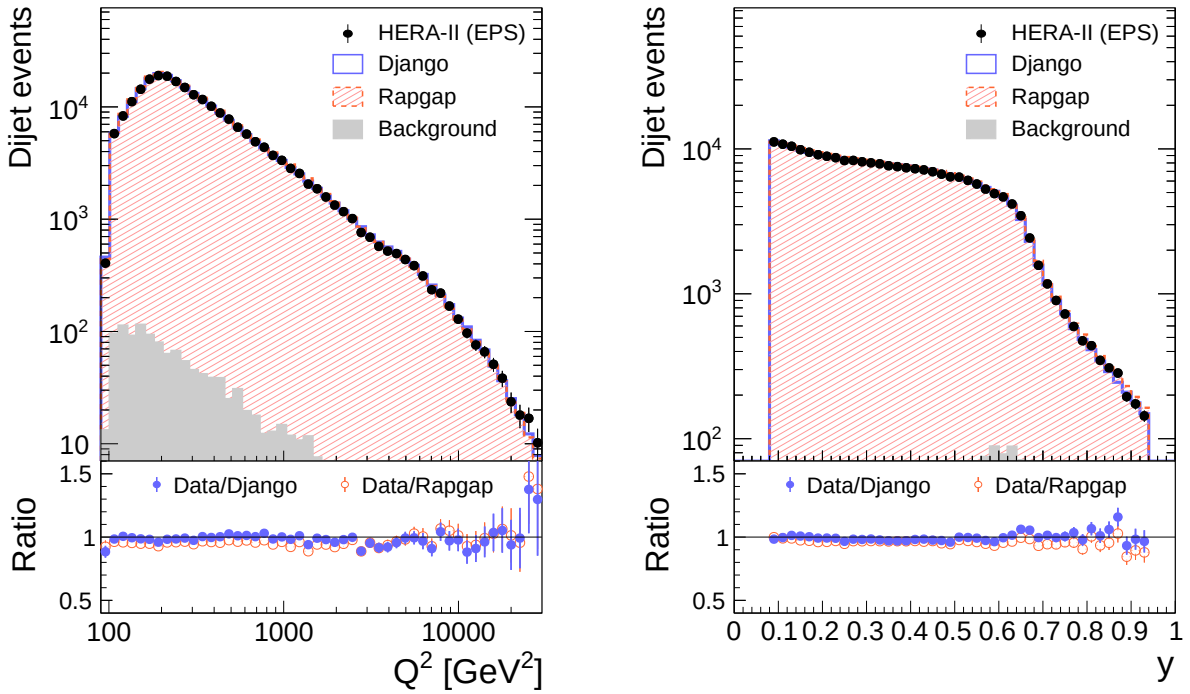


Figure 4.6: Control distributions of the NC DIS kinematic variables Q^2 and y of the dijet event sample in the EPS.

which corresponds to the momentum fraction of the struck parton w.r.t. the incoming proton in the leading order approximation. The ξ_2 distribution is shown in figure 4.8.

4.4.6.2 The dijet measurement phase space

The dijet MPS will be restricted to the MPS of the NC DIS measurement for DIS quantities, and the phase space of the jets is constrained to the more reliable acceptance regions $-1.0 < \eta^{\text{jet}}_{\text{lab}} < 2.5$ and to jet transverse momenta $p^{\text{jet}}_{\text{T}} > 5\,\text{GeV}$. The lower cut on $p^{\text{jet}}_{\text{T}}$ is motivated by the resolution and the jet-energy scale. The upper cut on $p^{\text{jet}}_{\text{T}}$ is for consistency with previous measurements [102, 173] and helps to improve the predictive power of pQCD calculations, since it constrains the renormalization scale accordingly. It also preserves a consistency with the inclusive jet sample. In order to ensure the convergence of the NLO calculations, the MPS will be further constrained to $M_{12} > 16\,\text{GeV}$. The average leading jet momenta are measured for dijet events in the MPS above $\langle p_{\text{T}} \rangle_2 > 7\,\text{GeV}$.

4.4.7 Trijet phase space

4.4.7.1 The trijet extended phase space

A three jet event selection is performed based on the selected neutral current DIS event sample. This sub-sample is named *trijet* sample and an event is labeled *trijet event* if three or more jets with transverse momenta

$$3 < p^{\text{jet}}_{\text{T}} < 50\,\text{GeV} \tag{4.16}$$

in the pseudorapidity region

$$-1.5 < \eta^{\text{jet}}_{\text{lab}} < 2.75 \tag{4.17}$$

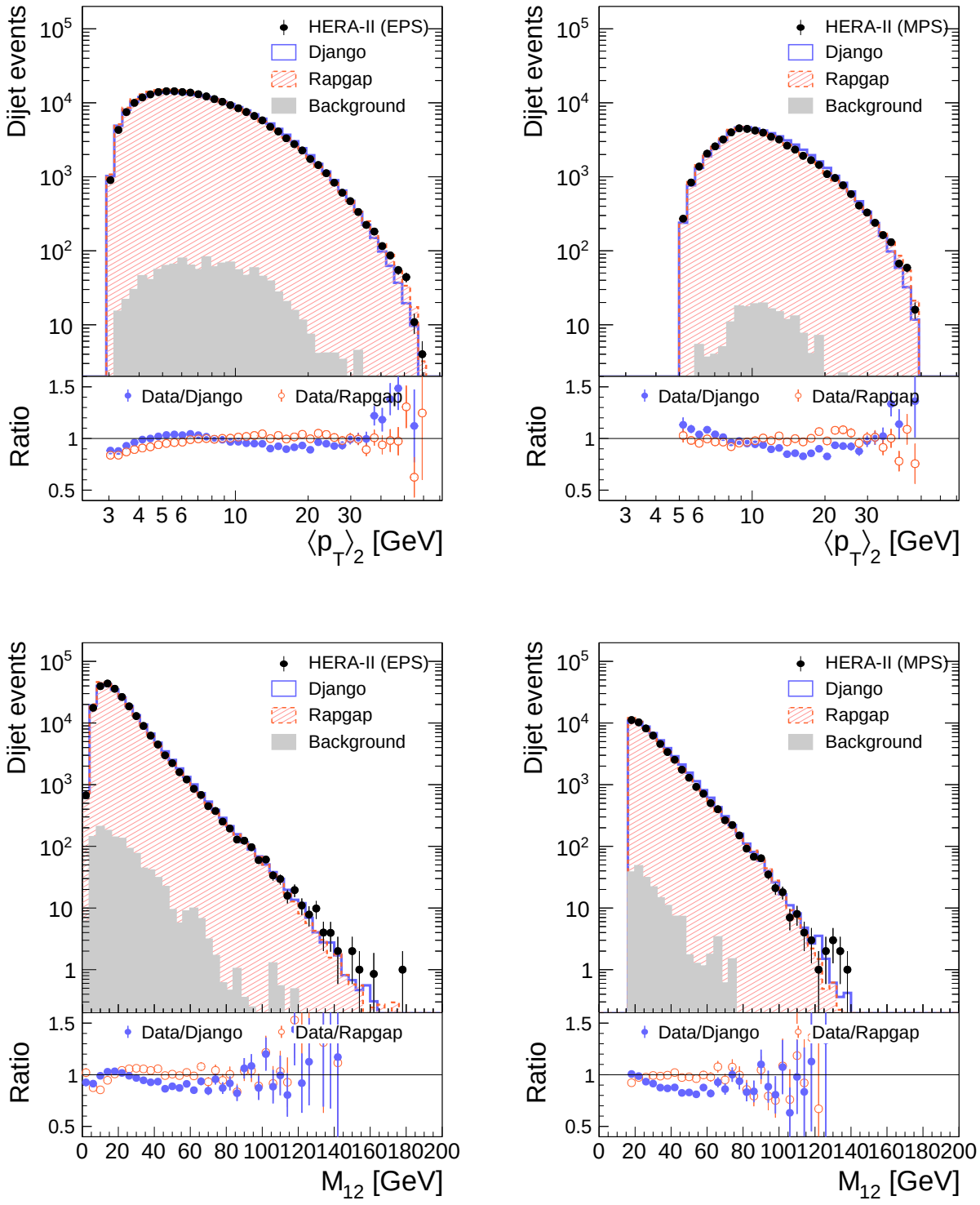


Figure 4.7: Control distributions of dijet observables $\langle p_{\mathrm{T}} \rangle_2$ (top) and M_{12} (bottom) in the EPS (left) and MPS (right). The MPS of the dijet measurement is constrained by $M_{12} > 16\,\mathrm{GeV}$ and by $p_{\mathrm{T}}^{\mathrm{jet}} > 5\,\mathrm{GeV}$ and $-1.0 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.5$. The dijet measurement is published for values of $\langle p_{\mathrm{T}} \rangle_2 > 7\,\mathrm{GeV}$.

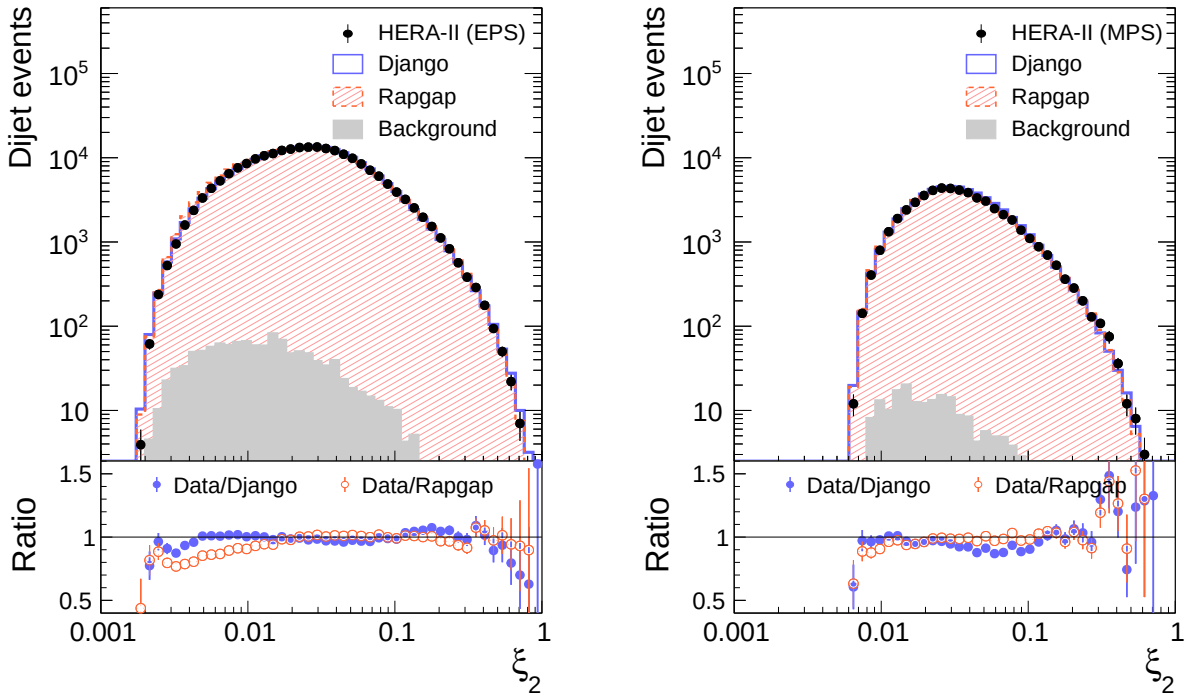


Figure 4.8: Distribution of ξ_2 of dijet events in the EPS (left) and MPS (right). The EPS incorporates dijet events with $p_{\mathrm{T}}^{\mathrm{jet}} > 3\,\mathrm{GeV}$ and no cut on M_{12}, and thus reaches lower values of ξ_2.

are measured. The neutral current kinematic variables Q^2 and y for the trijet event sample are shown in figure 4.9.
Similar considerations as for the dijet sample are applied to constrain the trijet phase space. The average transverse momentum of the jets of a trijet event is calculated by

$$\langle p_{\mathrm{T}} \rangle_3 := \frac{1}{2}\left(p_{\mathrm{T}}^{\mathrm{jet1}} + p_{\mathrm{T}}^{\mathrm{jet2}} + p_{\mathrm{T}}^{\mathrm{jet3}}\right) \ . \tag{4.18}$$

The distribution of $\langle p_{\mathrm{T}} \rangle_3$ for trijet events are shown in figure 4.10.
The invariant mass of a three jet topology is calculated using the three leading jets four-momenta in the respected $p_{\mathrm{T}}^{\mathrm{jet}}$-range

$$M_{123}^2 = \left(p^{\mathrm{jet1}} + p^{\mathrm{jet2}} + p^{\mathrm{jet3}}\right)^2 \ , \tag{4.19}$$

and the observable ξ_3 is calculated for a trijet event by

$$\xi_3 = x_{\mathrm{Bj}}\left(1 + \frac{M_{123}^2}{Q^2}\right) = x_{\mathrm{Bj}} + \frac{M_{123}^2}{ys} \ . \tag{4.20}$$

Whenever it is convenient, we drop the explicit subscript and use the shorthand notation $\langle p_{\mathrm{T}} \rangle$ for the two observables $\langle p_{\mathrm{T}} \rangle_2$ and $\langle p_{\mathrm{T}} \rangle_3$, and similarly for ξ_2 and ξ_3, if it is clear from the context which one of the observables is meant.

4.4.7.2 The trijet measurement phase space

The MPS for the trijet events is based on the dijet MPS, such, that the trijet sample is a full sub-sample of the dijet sample, but at least three jets are required in each event. Therefore, the invariant mass is required to be identical to the dijet measurement and calculated only from the two leading jets: $M_{12} > 16\,\mathrm{GeV}$. The statistics in the measurement phase space is increased in comparison to previous analyses [102] due to the comparably larger phase space in $\eta_{\mathrm{lab}}^{\mathrm{jet}}$.

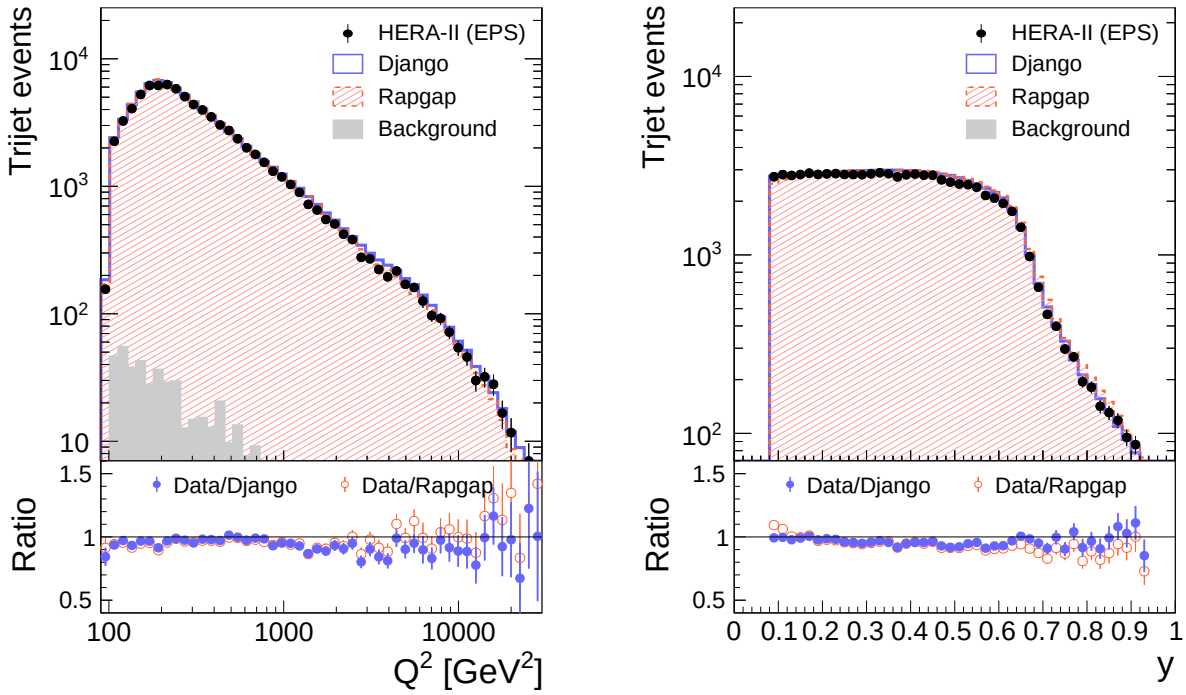


Figure 4.9: Control distributions of the NC DIS kinematic variables Q^2 and y of the trijet event sample in the EPS.

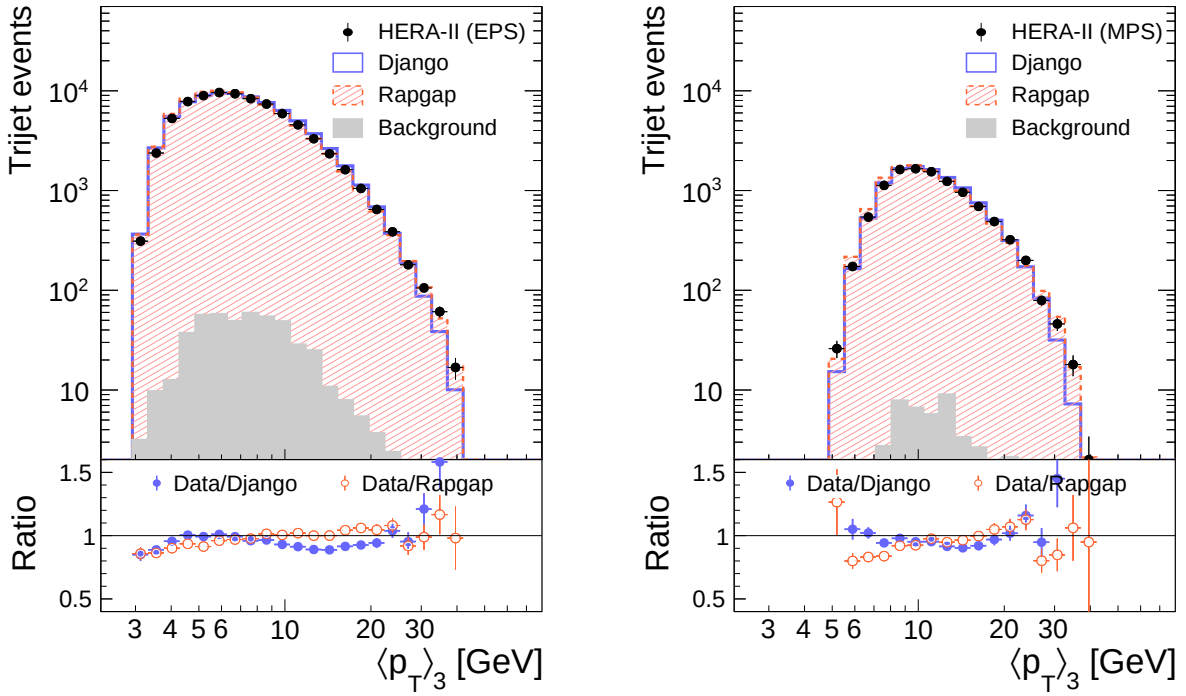


Figure 4.10: Control distributions of $\langle p_\mathrm{T} \rangle_3$ for the trijet measurement in the EPS (left) and the MPS (right). The extended phase space in $\eta^\mathrm{jet}_\mathrm{lab}$ and $p^\mathrm{jet}_\mathrm{T}$ significantly increases the amount of trijet events in the EPS.

Chapter 5

Regularized unfolding

In this chapter the method of regularized unfolding for correction of the detector acceptance and other detector effects is explained. The detailed description of the migration matrix employed is given in the next chapter. The resulting cross sections are presented in chapter 7. A comparison of the unfolding and a bin-by-bin method is performed in chapter 8. A study of the unfolding procedure and a comparison of several different definitions of the migration matrix is given in chapter 9.

5.1 Detector effects

The aim of any measurement is to determine or improve the knowledge about some physical quantity. In this analysis, these are jet-cross sections and the strong coupling constant. There are, however, often experimental and physical effects, which influence the measured data, such that the desired quantities cannot be determined directly. Therefore, data have to be corrected for these effects, such that the measured quantities can be compared to the theoretical predictions, and optimally no biases from the correction method should be introduced.
Corrections have to be applied for effects, which are included in the measured data and which are well-known, but which are by themselves of no interest and which are not included in the theoretical calculation. These are contributions from other *ep*-processes, and in this analysis for instance also higher order QED effects, which can be determined to very high accuracy.
Experimental corrections become necessary since the experimental apparatus cannot measure the quantity with the necessary precision. There may be two distinct detector effects present:

- Acceptance and efficiency effects
- Migrations and limited resolution

These effects are corrected for in the unfolding procedure.
The acceptance of the experimental setup in collider experiments cannot cover 4π due to beam holes, and insensitive detector regions from support structures or where detector components are put together. Also malfunctioning hardware reduces the acceptance. A single particle cannot be measured, if it passes an insensitive region. However, in such cases, secondary particles or parts of particle cascades may enter sensitive parts of the detector and only partial information about properties of a the primary particle are recorded.

A measuring apparatus always has a finite resolution, i.e. it is only capable of determining a quantity with a limited precision. This leads to migrations, which means that quantities are counted in a different bin on detector level than on generator level. If parts of complex objects, like particles in jets, pass insensitive regions, the measured quantity can be even heavily distorted.
In order to correct for these effect, the detector and its response needs to be well modeled in the simulation and also the MC generators employed need to provide a reliable event structure. The quality of the simulation of the H1 detector has been confirmed in many previous analyses. The control distributions for this analysis (see chapter 4) show a reasonable agreement of data and the simulation for relevant quantities.

5.2 Method of regularized unfolding

The method of *regularized unfolding* as implemented in the TUnfold package [181] is applied in this analysis to correct for detector effects. In this section, the method is explained briefly, and all relevant formulae for this analysis are given and discussed. The error treatment through the unfolding process is further outlined. More details about TUnfold may be found in [181]; and about regularized unfolding in general in [182–187].

5.2.1 The unfolding problem

In a general prescription, the detector is expressed by a response function, or kernel, $A(x, y)$, which acts on the true distribution $f(x)$. The detector level distribution is then calculated by a Fredholm integral of first kind

$$g(y) = \int \mathrm{d}x A(x, y) f(x) \ . \tag{5.1}$$

If the efficiency correction is included in the response function, it is normalized to $\int \mathrm{d}y A(x, y) = 1$, for any value of x.

In a counting experiment, equation 5.1 has to be approximated and expressed through the discrete equation

$$y = \boldsymbol{A}x, \tag{5.2}$$

where y is the measured detector level and is expressed in terms of a vector of dimension n, while x is the true distribution of dimension m, and $\mathbf{A}$ is an $n \times m$ matrix, which is referred to as *detector response* or *migration matrix*. The response matrix $\mathbf{A}$ is determined from Monte Carlo events and a simulation of the experimental setup. The necessary approximation in the transition to a discrete problem has to be performed in an optimal way, in order to have an accurate description of the experimental device.
The determination of the true distribution x could mathematically be performed by a direct matrix inversion of $\boldsymbol{A}$ by

$$x = \boldsymbol{A}^{-1}y, \tag{5.3}$$

if $n = m$ and $\det \boldsymbol{A} \neq 0$. However, this is in general not possible, since the unfolding problem is an ill-posed problem [188], or the matrix is not directly invertible, when $m \neq n$. The vector x may show large fluctuations, as it is for instance demonstrated in [189]. These oscillations are interpreted as being unphysical.

5.2.2 The regularized unfolding as implemented in TUnfold

A solution to find x in equation 5.2 is alternatively given by a least square minimization approach, which is expressed by finding the minimum χ^2 as function of x in

$$\chi^2 = \chi^2_{\mathrm{A}} + \chi^2_{\mathrm{L}} \tag{5.4}$$

using

$$\chi^2_{\mathrm{A}} = (y - \boldsymbol{A}x)^{\mathrm{T}} \boldsymbol{V}_y^{-1} (y - \boldsymbol{A}x) \tag{5.5}$$

and

$$\chi^2_{\mathrm{L}} = \tau^2 (x - x_0)^{\mathrm{T}} (\mathbf{L}^{\mathrm{T}}\mathbf{L})(x - x_0) \ , \tag{5.6}$$

where $\boldsymbol{V}_y$ denotes the covariance matrix on detector level as determined by the measurement. The expression χ^2_{L} is a Tikhonov type regularization term [190], with τ being a small but free regularization parameter, x_0 being the true distribution of the model, and $\mathbf{L}$ is the regularization condition. The regularization suppresses large fluctuations on x due to statistical fluctuations in y. The regularization matrix $\mathbf{L}$ can also introduce smoothness conditions between bins of the type of an n-th discrete derivative. The regularization condition and regularization strength is studied in more detail in section 9.3.
The true vector is found by determining the stationary point of χ^2 by setting the first derivative to zero and solving the equation

$$\frac{\partial \chi^2}{\partial x} = 0 \tag{5.7}$$

as function of x. The true vector x is given by [181]

$$x = \left(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}_y^{-1} \boldsymbol{A} + \tau^2 \boldsymbol{L}^2\right)^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}_y^{-1} y =: \boldsymbol{B} y \ . \tag{5.8}$$

This calculus avoids the direct inversion of the response matrix $\boldsymbol{A}$, but only the quantity in brackets including the regularization term has to be inverted. It has to be ensured, that the covariance matrix $\boldsymbol{V}_y$ is invertible. This is typically the case for a well-defined measurement, i.e. where two bins are not fully correlated and $\boldsymbol{V}_y$ has full rank. The precise definitions of the input quantities for equation 5.8 are given for y in equation 5.19, for $\boldsymbol{A}$ in equation 5.12 and for $\boldsymbol{V}_y$ in equation 5.20.

5.3 Determination of the migration matrix

The migration matrix $\boldsymbol{A}$ is determined using Monte Carlo event generators and simulation of the detector. Each MC event is assigned a generator level weight w_{gen} and a reconstruction level weight w_{rec}, where $w_{\mathrm{rec}} \leq w_{\mathrm{gen}}$ and the difference accounts for effects like for instance trigger inefficiencies. A two-dimensional *migration histogram*

$$\boldsymbol{N} = (N_{ij}) = \left(\sum_k w^{\mathrm{rec}} |_{\mathrm{rec}(j) \wedge \mathrm{gen}(i)} \right) \tag{5.9}$$

is determined, using the unfolding bin grid, which counts the detector level weights w_{rec} of each quantity of every event k, reconstructed in bin j of the detector level bin grid and generated in bin i of the generator level bin grid. These entries account for the mapping between the generator level and the detector level. The differences $\epsilon = w_{\mathrm{gen}} - w_{\mathrm{rec}}$ of the

events generated in bin i are counted in the generator level bin grid, and are technically included as an overflow bin of bin of the histogram $\boldsymbol{N}$. These bins are denoted as bins for *efficiency corrections* and are shown in figure 6.1 as the rightmost column. Monte Carlo events (or jet-observables), which are present only on generator level in bin i, but have zero weight on detector level, are counted in the efficiency correction bin i of the generator level bin grid.
The histogram $\boldsymbol{N}$, including the efficiency correction, fulfills the two projection equalities

$$y_{\mathrm{MC}} = \mathrm{proj}_x(\boldsymbol{N}) = (\sum_i^m N_{ij}) \ , \tag{5.10}$$

$$x_{\mathrm{MC}} = \mathrm{proj}_y(\boldsymbol{N}) = (\sum_j^{n+1} N_{ij}) \ . \tag{5.11}$$

The migration matrix $\boldsymbol{A}$ is then determined by normalizing each row of the migration histogram

$$\boldsymbol{A} = (A_{ij}) = \left(\frac{N_{ij}}{\sum_j^{n+1} N_{ij}} \right) \ . \tag{5.12}$$

The element of the migration matrix A_{ij} can therefore be interpreted as the probability of an event generated in bin i being reconstructed in bin j. The migration matrix satisfies the equation[1]

$$y_{\mathrm{Model}} = \boldsymbol{A} x_{\mathrm{Model}} \tag{5.13}$$

for the model vectors.

In this analysis, the migration histogram is determined using the two different Monte Carlo models Django (Dj) and Rapgap (Rg) (see section 3.5), by $\boldsymbol{N} = \boldsymbol{N}_{\mathrm{Dj}} + \boldsymbol{N}_{\mathrm{Dj}}$, which is equivalent to averaging the two migration matrices

$$\boldsymbol{A} = \frac{\boldsymbol{A}_{\mathrm{Dj}} + \boldsymbol{A}_{\mathrm{Rg}}}{2} \ . \tag{5.14}$$

5.4 Error propagation

Several sources of uncertainties have to be identified and propagated through the unfolding process. These are statistical uncertainties of the data, correlated systematic uncertainties like the jet energy scale, the model uncertainty, the background normalization uncertainty or the uncertainty from limited MC statistics.

5.4.1 Statistical uncertainty

The matrix $\boldsymbol{B}$ is the effective inverse response matrix of the detector and is employed for the error propagation. The *covariance matrix* (or *error matrix*) provides the statistical uncertainties of the unfolded result x, which is then calculated by

$$\boldsymbol{V}_x = \boldsymbol{B} \boldsymbol{V}_y \boldsymbol{B}^{\mathrm{T}} \ . \tag{5.15}$$

This true level covariance matrix includes all statistical uncertainties and their correlations from the measurements y on detector level ($\boldsymbol{V}_y$) and propagated to the true level x.

[1] Equation 5.13 is also denoted as *folding equation*.

The *correlation matrix* is defined through

$$\boldsymbol{P} = (\rho_{ij}) = \left(\frac{V_{ij}}{\sqrt{V_{ii} V_{jj}}} \right) \tag{5.16}$$

and expresses the strength of the correlations between different bins. Two bins i and j are denoted as *anti-correlated* if $\rho_{ij} < 0$ and as *(positively) correlated* if $\rho_{ij} > 0$.
The square root of the diagonal elements of $\boldsymbol{V}_x$ are usually quoted as the statistical uncertainty in a particular bin. This, however, neglects the correlations between different bins, which are usually not shown graphically. The correlations are important, since even if the diagonal elements of the covariance matrix are large, the measurement can still be very precise, if the correlations are largely negative, i.e. $\rho_{ij} \to -1$. On the other hand, positive correlations can also be interpreted as a correlated systematic shift.

5.4.2 Systematic uncertainties

Systematic uncertainties are estimated by simulating the size of the effect using the detector simulation. For each source of systematic uncertainty, the effect is simulated twice: for an upwards and a downwards shift. These "shifted" detector simulations are employed to determine an alternative migration matrix $\boldsymbol{A}^k = \boldsymbol{A} + (\Delta \boldsymbol{A})^k$ for each shift k of the various sources of systematic uncertainties[2]. The effect of a shifted migration matrix on the true level distribution is determined by error propagation, by the replacements

$$\begin{aligned} x &\to x + \Delta x \\ \boldsymbol{A} &\to \boldsymbol{A} + (\Delta \boldsymbol{A}) \end{aligned}$$

in equation 5.8 (the index k has been dropped for better readability) and solving the equation as function of Δx. By linearizing the problem, i.e. neglecting quadratic terms like $(\Delta \boldsymbol{A})^2$ and $\Delta \boldsymbol{A} \Delta x$, the systematic shift of each source of correlated systematic uncertainty is given by [181]

$$\Delta x = (\Delta \boldsymbol{B})(y - \boldsymbol{A}x) - \boldsymbol{B}(\Delta \boldsymbol{A})x \ , \tag{5.17}$$

using the shorthand notation $(\Delta \boldsymbol{B}) \equiv \left(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}_y^{-1} \boldsymbol{A} + \tau^2 \boldsymbol{L}^2 \right)^{-1} (\Delta \boldsymbol{A})^{\mathrm{T}} \boldsymbol{V}_y^{-1}$. The systematic uncertainties of one source are expected to be correlated. Hence, the systematic uncertainties may be expressed in terms of a correlated error matrix

$$\boldsymbol{V}_{\mathrm{sys}}^{\mathrm{corr}} = (V_{ij}^{\mathrm{corr}}) = (\Delta x)(\Delta x)^{\mathrm{T}} = ((\Delta x)_i (\Delta x)_j) \ . \tag{5.18}$$

The propagation of the systematic uncertainties through the unfolding process in order to determine $x + (\Delta x)^k$ is equivalent to an unfolded result $\tilde{x}_k$, which is determined using the migration matrix A^k in the unfolding procedure. It has been confirmed for all sources of systematic uncertainties of this analysis that the difference between $\tilde{x}_k$ and $x + (\Delta x)^k$ is negligible.

[2] Throughout this thesis, we will use 'Δx' for specifying an absolute uncertainty on the value x (e.g. $\Delta x = 0.3\,\mathrm{pb}$), and the notation '$\delta x$' whenever relative uncertainties are meant (e.g. $\delta x = 10\,\%$). These notations are connected through $\Delta x = x \cdot \delta x$.

5.4.3 Model uncertainty

Although the response matrix $\boldsymbol{A}$ is expected to be universal for each experimental device, i.e. independent of the physics of the MC model which is employed for its determination, due to the approximation of equation 5.1 through the discrete form 5.2, there might be shortcomings in the completeness of the description of the detector response. This results in small residual differences in A_{ij} and thus on the unfolded result x, if different models are employed for the determination of the response matrix.
In this analysis, the two MC models Rapgap (Rg) and Django (Dj) are used to determine the migration matrix. A model uncertainty Δx is defined in analogy to the systematic uncertainties, by determining an up- and down variation using the alternative migration matrices $\boldsymbol{A}^{\mathrm{Dj}}$ and $\boldsymbol{A}^{\mathrm{Rg}}$.

5.4.4 Uncertainties from limited Monte Carlo statistics

The migration matrix $\boldsymbol{A}$ is determined from a limited sample of MC events which is approximately 25 times the luminosity of the data statistics for each of the two MC models. This kind of uncertainty on $\boldsymbol{A}$ may also be propagated through the unfolding process [181] in order to determine a covariance matrix $\boldsymbol{V}_x^{\boldsymbol{A},\mathrm{stat}}$ from this source of uncertainty. In this analysis it has been verified, that the statistics of the MC event sample is large enough, such that it is negligible compared to $\boldsymbol{V}_x$ for all bins of the measurement phase space.

5.4.5 Background uncertainties

The vector of simulated background events b from other *ep*-scattering processes is subtracted on detector level from the data vector y_0 prior to the unfolding for each background source l. The vector, which is input to the unfolding is therefore calculated by

$$y = y_0 - \sum_l b_l \ . \tag{5.19}$$

The statistical errors of the MC generated background events ($\Delta^{\mathrm{MC}}_{\mathrm{stat}} b_l$) have to be added in quadrature to the statistical uncertainty of the data. Furthermore, each background source is assigned a relative normalization uncertainty $\delta_{\mathrm{Norm}} b_l$, e.g. 30 % for QED Compton events (c.f. 4.2.2). The absolute normalization uncertainty is then given by $\Delta_{\mathrm{Norm}} b_l = \delta_{\mathrm{Norm}} b_l b_l$, which has to be added to the statistical uncertainty of $\boldsymbol{V}_y^0$ of the data. The covariance matrix, which is employed for the unfolding procedure is then given by [181]

$$\boldsymbol{V}_y = (V_y^0)_{ij} + \sum_l \left(\delta_{ij} (\Delta^{\mathrm{MC}}_{\mathrm{stat}} b_l)_i^2 + (\Delta_{\mathrm{Norm}} b_l)(\Delta_{\mathrm{Norm}} b_l)^{\mathrm{T}} \right) \ , \tag{5.20}$$

where the sum l runs over all relevant background processes. Consequently, the covariance matrix after the unfolding $\boldsymbol{V}_x$ also incorporates these correlated background normalization uncertainties, which cannot be disentangled from the statistical uncertainties any more[3].

[3]The covariance matrix of only the statistical uncertainties on true level, i.e. without the uncertainty from the *ep*-backgrounds, could principally be propagated through $\boldsymbol{V}_x^{\mathrm{stat}} = \boldsymbol{B} \boldsymbol{V}_y^0 \boldsymbol{B}^{\mathrm{T}}$.

5.5 Inclusion of bin-wise correction factors in the migration matrix

The migration matrix represents the effect of the experimental device on the hadron level distribution. However, it might be necessary to correct the data in addition with bin-wise correction factors. For instance, the higher-order QED corrections cannot be represented through a migration matrix, since the 'radiative' and the 'non-radiative' calculations are performed individually.
Here, we have developed a method to include bin-by-bin correction factors in the unfolding procedure. The bin-wise correction factors cannot be applied to each bin (or row) in the migration histogram, since each generator level bin is normalized, when determining the migration matrix. A bin-wise correction factor c_i can instead be included in the migration histogram, by applying it to the content of the efficiency correction bin $\epsilon_{0,i}$, which determines the normalization, and subtracting in addition its impact on the reconstructed contributions $(x_{0,i} - \epsilon_{0,i})$ also from the efficiency bin. Hence, the efficiency bin has to be adapted following

$$\epsilon_{0,i} \to c_i \epsilon_{0,i} - (1 - c_i)(x_{0,i} - \epsilon_{0,i}) \ , \tag{5.21}$$

where $x_{0,i}$ is the generator level prediction in bin i of the model which is calculated accordingly equation 5.11. This modification is identical to an additive contribution γ_i to each efficiency bin in the migration histogram of

$$\epsilon_{0,i} \to \epsilon_{0,i} + \gamma_i =: \epsilon_{0,i} + [(c_i - 1)\, x_{0,i}] \ . \tag{5.22}$$

If the correction factors c_i are of the same order in each bin (e.g. around unity), the inclusion of bin-wise correction factors in the unfolding histogram is equivalent to a bin-by-bin correction of the unfolded result.
An advantage of the inclusion of correction factors into the migration matrix is the finer generator level bin grid of the unfolding scheme compared to the final measurement binning. There are e.g. two bins on generator level in each measurement bin in $p_{\mathrm{T}}^{\mathrm{jet}}$ for the inclusive jet measurement. This enables a more accurate application of the correction. For the case of QED corrections, another advantage is the consistent definition of the non-radiative generator level of the MC generators. In the study in section 9 this leads to a reduced model uncertainty.

Chapter 6

Definition of the migration matrix

The definition of the migration matrix is the most sensible part when performing a regularized unfolding. The matrix $\boldsymbol{A}$ should be capable of describing accurately all detector effects and should be the best approximation of the general differential detector response function $A(x, y)$. Furthermore, the applied corrections should be independent of the employed model and not introducing a model dependence.
The structure of the matrix, i.e. the structure of the detector level and generator level vector is called *migration scheme*. A migration scheme is defined through the included observables for the description of the migrations, by their binning on detector and generator level, by the number of side-bins, etc. The migration scheme as used in this work is the outcome of many studies. It is naturally not possible to define the migration scheme from first principles. The choice of the finally employed migration scheme is motivated by the comparison of different schemes, where some tendencies could be derived concerning the usability of various variables.

6.1 Considerations concerning the definition of the migration matrix

There is a wide literature about unfolding, where many suggestions and hints for an optimal unfolding scheme are made. On the other hand, there is also a wide disagreement about the correct method and proper implementation of methods. Most of the physically motivated studies are based on toy models, incorporating artificial physical spectra and artificial detector models. There are a few remarks to be made on toy based unfolding studies:

- **Detector response**
  It is often assumed in toy models that the physical spectra are distorted by assuming a finite resolution in exactly that variable which is to be measured differentially. This, however, is not appropriate for a complicated detector device, where resolution and acceptance effects take place primarily in other variables. If the MC model, which is employed for the determination of the migration matrix, does not perfectly describe these other variables, which are integrated out, the unfolding may be biased toward the Monte Carlo prediction.

- **Physical spectra: resolution and statistics**
  In several toy studies, a resonant peak for example is distorted by finite resolution effects. This, however, does not correspond to the unfolding problem as faced in

this work. In many physical problems, a steeply falling cross section is measured and has to be unfolded. This implies several peculiarities: firstly, the spectrum vary by many orders of magnitude, e.g. more than ten orders of magnitudes for LHC jet cross sections. This may result in large numerical inaccuracies, for instance when for a matrix inversion small eigenvalues are dropped although these may still be relevant in phase space regions with smaller statistics. Secondly, there may be phase space regions with only few entries, one entry, or no even entry. Since normal distributed errors are assumed in the least square ansatz, this may introduce a bias which has to be corrected for [191]. Furthermore, an issue are the conflicting wishes on the resolution, the cross section and the theoretical accuracy, which are closely related. In the phase space region of low transverse momentum jets, where the jet-production cross sections are large, a fine binning on the basis of the statistics is likely to be chosen. On the other hand, at low transverse momentum the resolution is poor and therefore migrations are large. However, from the theory perspective this region is of special interest, since it is particular sensitive to the QCD coupling strength due to confinement, or may constrain the region where pQCD is applicable. These theoretical studies can be performed with very high experimental accuracy, since the statistical precision of the measurements are large.

- **Boundaries of the measured phase space**
  The usable phase space of any measurement is or has to be restricted. At large scales it is mostly constrained by physical properties, arising for example from the fixed center-of-mass energy of the collider, which results close to that boundary in vanishing cross sections, and migrations become negligible w.r.t. the bin size.
  The restriction at the "opposite" boundary of the phase space at low scales is usually more complicated. In many measurements, the resolution may become poor, since e.g. the electronic noise dominates over the expected deposited energy. In these phase space regions, where the measurements are no longer reliable, the phase space boundary has to be defined. This implies that especially at this boundary, the migrations are getting large and are important to be considered.
  To account for these migrations, the phase space has to be extended to an even less reliable phase space region to provide an estimate of these migrations. However, also this extended phase space region has to be estimated again from data, which is even more distorted by detector effects and, furthermore, the cross section typically increases drastically because of the physical spectrum. All contributions from the outside phase space region, which migrate into the (extended or measurement) phase space, have to be estimated from Monte Carlo events. Furthermore, the MC predictions may be poor at these soft scales, and are even further distorted by the increasing size of detector effects at low energies.

Following the above arguments, the setup of the migration scheme cannot rely solely on recommendations, which are based on toy model studies. Instead, the structure of the migration scheme in this work is iteratively optimized to obtain a model dependence as small as possible (see chapter 9).
Only few parameters can be chosen, when defining the set-up of the migration matrix. These are:

- The choice of the observables to describe the migrations
- The number of observables

- The bin grid of the observables
- The extended and measurement phase space

Basically, any observable can be considered for the description of the migrations, and hence a the first bullet constitutes a large freedom for a multidimensional[1] unfolding. Also the opportunity to include migrations in quantities which are subject to the phase space definition, e.g. the invariant mass cut for the dijet measurement, gives additional freedom.

The 'perfect' unfolding scheme would include as many observables as possible with very fine binning on detector and generator level. This, however, is not possible, since data has only limited statistics, and the statistical uncertainty in each bin should optimally satisfy a normal distribution, in order to apply the unfolding method. There are also constraints from the computing power[2], since the covariance matrix $\boldsymbol{V}_y$ and the matrix expression inside of $\boldsymbol{B}$ have to be inverted directly within a reasonable time.

6.2 Structure of the migration matrix

In this work a simultaneous unfolding of the neutral current DIS, the inclusive jet, the dijet and the trijet measurements is performed. These measurements together with this procedure also allows to determine normalized cross sections, such as jet cross sections normalized to the NC DIS measurement and the trijet over dijet ratio. The combined unfolding has the further advantage to estimate the amount of detector-level-only contributions. This feature is explained and studied in more detail in section 6.5.

The migration matrix is composed of a 4×4 structure of submatrices and is shown[3] schematically in figure 6.1. The four diagonal elements $\boldsymbol{E}$, $\boldsymbol{J}_1$, $\boldsymbol{J}_2$ and $\boldsymbol{J}_3$ are the individual migration matrices of the four single measurements. The unfolding schemes of these four matrices are different for each measurement and are explained in detail in the following sections.

[1] The phrase 'multidimensional' is used to describe an unfolding scheme, which accounts for migrations in multiple variables at the same time.

[2] Unfolding schemes, which take many observables in a multidimensional setup into account have to deal with large vectors and thus large matrices. The migration matrix, which is employed in this analysis has an overall size of 3166028 entries. Although many phase space regions are empty and already some optimization has been performed, this matrix allocates more than 25 MB in memory, while the covariance matrix $\boldsymbol{V}_y$ allocates 167 MB in memory, when determined in double precision. The memory demand increases rapidly, when also statistical uncertainties for each entry and migration matrices for systematic uncertainties have to be stored, and multiple models are used. State-of-the-art operating systems, like scientific linux 5 (SL5), are 32-bit operating systems and thus can only handle a maximum of approximately 3 GB in memory at the same time.

[3] In the common mathematical notation, the diagonal elements of a matrix are written from the upper left to the lower right corner. In the computer code which is used here, the matrix is stored as a histogram. The common histogram plotting tools draw the diagonal elements from the lower left to the upper right. Hence, graphical illustrations of matrices are shown in the histogram style, while numeric values are shown in the correct mathematical manner in this work. An element of a matrix may therefore also be denoted as a 'bin' in correspondence to the technical implementation. In graphical illustrations, we show the detector level always on the x-axis and the generator level on the y-axis. Since the x-dimension is typically larger than the y-dimension, this convention simplifies the illustration in landscape formatted illustrations.

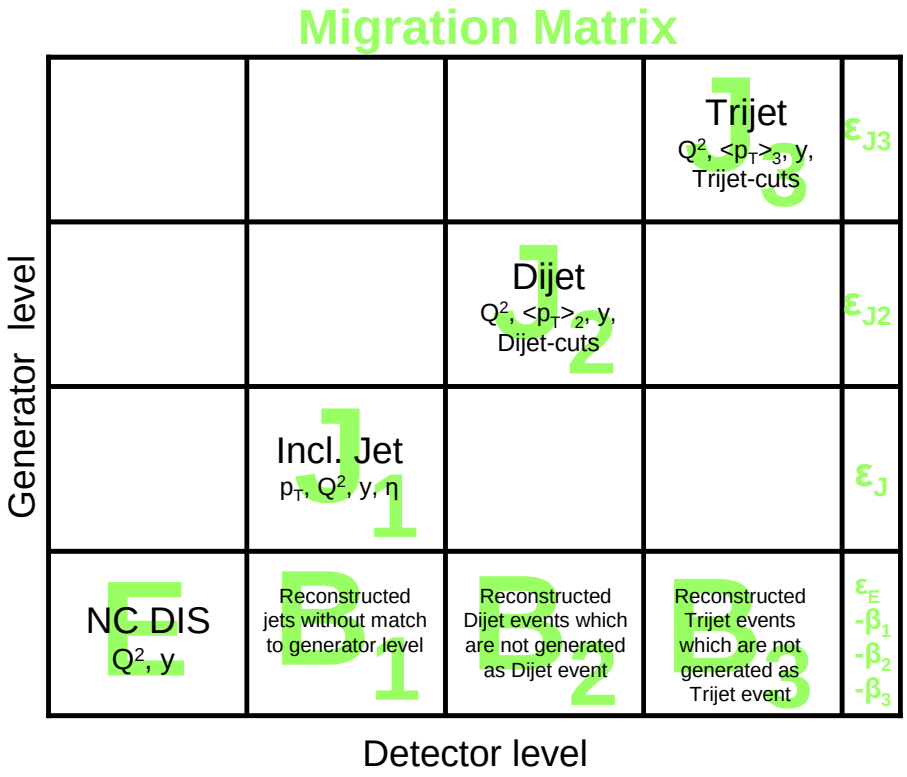


Figure 6.1: Schematic illustration of the migration matrix as used for the unfolding in this analysis to correct for detector effects. The matrix incorporates four submatrices, which are the migration matrices for the individual measurements. The matrices $\boldsymbol{B}_i$, which connect the jet measurements with the NC DIS generator level, are used to estimate the number of jets (or events) which have no connection to the generator level in the relevant jet quantity, but, however, the DIS event kinematics (Q^2,y) is available. The column for the normalization is shown at the rightmost.

The size of the matrix is given by the size of the submatrices and is $m \times n$, where the detector level vector y has the size m and the generator level has the size n. Figure 6.1 further shows the additional column, which is employed for the normalization (see section 5.3). The three matrices $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ connect the detector level measurement of the jet data with the generator level of the NC DIS measurement. These matrices are used in this work to determine the number of jets (or events) which are present in Monte Carlo events on detector level, but are not connected to any bin on the generator level. Such entries are only present for the jet measurements. The NC DIS measurements has an extended phase space on generator level, such, that all NC DIS events on detector-level can be connected to the generator level.

The eight submatrices, which connect a measurement on detector level to a different measurement on generator level are left empty. The contributions from the inclusive jet detector level to the dijet and trijet observables are ambivalent, since every jet is being counted compared the an event based measurement. The submatrices, which connect the djiet and the trijet measurement would have only little contributions of detector-level and generator-level-only contributions, where, however, also the observable $\langle p_{\mathrm{T}} \rangle$ is defined differently. This 4×4 structure of the migration matrix $\boldsymbol{A}$ is used throughout this work. The unfolding schemes which are employed for the submatrices of the individual measurements are investigated in more detail in section 9.2.

6.3 Statistical correlations on detector level

The four measurements are performed on the identical NC DIS event selection, where multiple observables are counted simultaneously for events containing jets, and are therefore statistically correlated. Furthermore, the inclusive jet measurement has positive statistical correlations on detector level, since multiple jets in one single event are statistically correlated. The unfolding procedure takes these correlations through the covariance matrix $\boldsymbol{V}_y$ into account.
For the determination of the covariance matrix, in each event the measured vector y^{event} is determined. For example, let's assume a two jet event where the two jets have identical $p_{\mathrm{T}}^{\mathrm{jet}}$ the measured vector is of the form[4] $y^{\mathrm{event}} = (\ldots 1 \ldots 2 \ldots 1 \ldots 0 \ldots)$. Different events are statistically independent, but all measured quantities of one single event are fully correlated, and hence, a covariance matrix is determined for each event by

$$\boldsymbol{V}_y^{\mathrm{event}} = (V_{ij}) = \left(y_i^{\mathrm{event}} \cdot y_j^{\mathrm{event}} \right) \ . \tag{6.1}$$

The covariance matrix $\boldsymbol{V}_y$ of the measurement is then determined by error propagation $\boldsymbol{V}_y = \sum_{\mathrm{events}} \boldsymbol{V}_y^{\mathrm{event}}$. This matrix incorporates all statistical correlations between the single measurements and also the statistical correlations of the inclusive jet data. The matrix $\boldsymbol{V}_y$ is typically invertible[5].

[4] Regarding the overall structure of the migration matrix in figure 6.1, this event is counted in one bin of the NC DIS measurement, where all other values of y^{event} are zero. The two jets are counted in the inclusive jet measurement in the same bin, and hence with a bin content of '2'. All other bins give no contribution to the y^{event} ('...'). The event is further counted in one bin of the dijet measurement ('1'), but not in the trijet measurement ('0').

[5] In the case of an uncorrelated measurement, the covariance matrix $\boldsymbol{V}_y$ is always invertible since it is only a diagonal matrix. In the case of a correlated measurement this may not necessarily be the case, if e.g. two bins are fully correlated. This would correspond to a measurement, where two times the same

Neutral current DIS measurement

Migration scheme	Two-dimensional unfolding in Q^2 and y
Generator level	
Measurement phase space	$150 < Q^2 < 15\,000\,\text{GeV}^2$ $0.2 < y < 0.7$
Q^2 number of bins (measurement)	8 (6)
Q^2 bin grid [GeV^2]	60 – 150 – 200 – 270 – 400 – 700 – 5000 – 15 000 – 40 000
Q^2 bin grid measurement [GeV]	150 – 200 – 270 – 400 – 700 – 5000 – 15 000
y number of bins (measurement)	2 (1)
y bin grid	0 – 0.2 – 0.7 (– 1.0) migrations from $0.7 < y < 1.0$ are treated as background
y bin grid measurement	0.2 – 0.7
Detector level	
Q^2 number of bins	14
Q^2 bin grid [GeV^2]	100 – 135 – 150 – 165 – 200 –230 – 270 – 320 – 400 – 500 – 700 – 1200 – 5000 – 15 000 – 40 000
y number of bins	3
y bin grid	0.08 – 0.15 – 0.35 – 0.7

Table 6.1: Unfolding scheme for the neutral current DIS measurement.

6.4 Migration scheme for the neutral current DIS measurement

The unfolding of the neutral current DIS measurement is performed in a two-dimensional unfolding scheme, employing the observables Q^2 and y. The applied bin grids on detector and generator level are depicted in table 6.1. We refer to the *measurement phase space* (MPS) for NC DIS kinematics as the phase space region defined by $150 < Q^2 < 15\,000\,\text{GeV}^2$ and $0.2 < y < 0.7$. In this phase space detector acceptance and efficiencies are large and jet cross sections will be published. The measurement phase space is a sub-space of the *extended phase space* (EPS), which enlarges the MPS through side-bins for an accurate determination of migrations into and out of the MPS[6].
The generator level binning is chosen to be the measurement binning of the six bins in Q^2 between $150 < Q^2 < 15\,000\,\text{GeV}$ plus two side-bins in order to account for migrations into and out of the phase space. The histogram of the Q^2 migrations only is shown in figure 6.2, which also illustrates the chosen bin grids. The y measurement phase space is represented by one bin of $0.2 < y < 0.7$ and is enlarged through two side-bins for the EPS. All generated events can therefore be assigned a dedicated bin number.

quantity would have been measured. It can also occur, when two quantities of only one single event are measured (or repeatedly such a constellation). In this analysis, the statistics are large enough, and the observables of the four measurements are sufficiently distinct, and therefore the covariance matrix has always full rank and is invertible.

[6]The NC DIS extended phase space (EPS) is defined differently on detector and on generator level. On generator level it is defined as $60 < Q^2 < 40\,000\,\text{GeV}^2$ and $0.0 < y < 1.0$, whereas on detector level it is defined through $100 < Q^2 < 40\,000\,\text{GeV}^2$ and $0.08 < y < 0.7$.

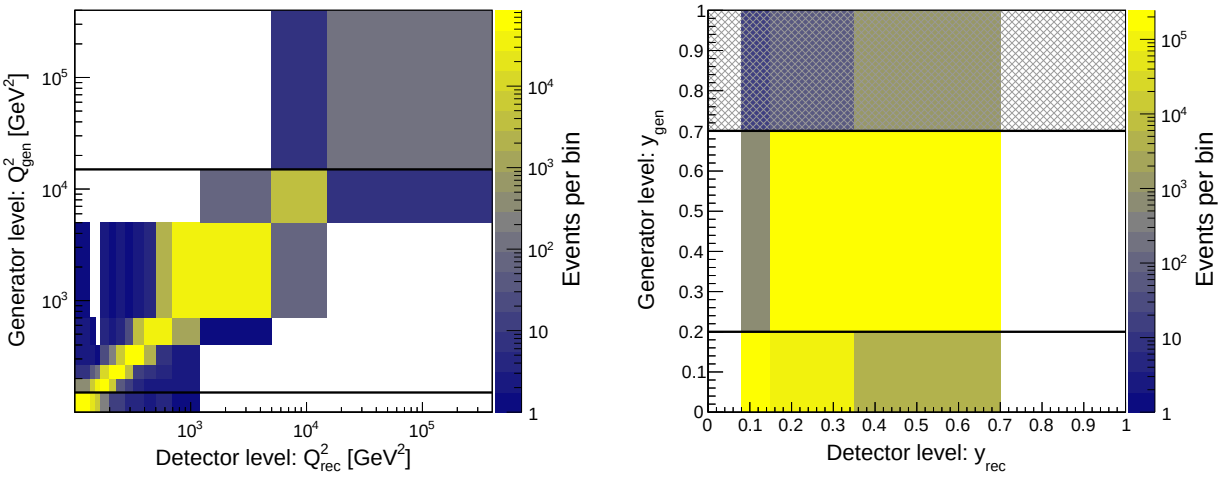


Figure 6.2: Binning scheme for Q^2 (left) and y (right) of the NC DIS migration histogram. The unfolding scheme used in the NC DIS measurement is a two-dimensional unfolding, built up from these two bin grids, where the Q^2 bin grid is contained into the y bin grid. The color coding shows the number of NC DIS events per bin, where the 'other' dimension is integrated out. The black lines indicate the phase space of the measurement (MPS). No data are taken for $y_{\text{rec}} > 0.7$, due to the requirement on the electron energy of $E_e > 11\,\text{GeV}$. Therefore, the grey shaded area at $y_{\text{gen}} > 0.7$ is subtracted prior to the unfolding, since this phase space cannot be reasonably determined from data. Identical bin grids are also employed for the unfolding of the jet measurements.

The bin grid in Q^2 on detector level is chosen such, that each generator level bin is subdivided into two bins. The phase space at the lower bound is extended by two side-bins down to $Q^2 = 100\,\text{GeV}^2$, where the electron acceptance due to the S67 trigger requirement falls off to 20 %. These two side-bins between 100 to $150\,\text{GeV}^2$ are used to determine the events of the generator level side-bin from 60 - $150\,\text{GeV}^2$. The two generator level bins between 5000 to $40\,000\,\text{GeV}^2$ are represented on detector level through the same bin boundaries, since the statistics is getting poor, and due to the large bin size the migrations are small.

6.5 Detector-level-only entries

When the migration histogram is filled from MC events, each contribution has to be assigned a generator level and a detector level bin. However, there are orphaned contributions which are present solely on generator or detector level. Generator-level-only entries are treated as inefficiencies, whereas detector-level-only contribution have to be treated carefully, since these contribute to the measured statistics.

6.5.1 Appearance of detector-level-only contributions

Whenever the generator-level phase space is constrained and at the same time migrations are present, contributions in MC events are present which are measured in a well-defined detector level bin, but cannot be assigned to a generator-level bin. Such contributions are for example present in the region of low $p_{\text{T}}^{\text{jet}}$, where the generator level phase space is constrained due to the necessary requirement of a minimum p_{T} for the jet finding algorithm and for the applicability of perturbative QCD predictions. These contributions (i.e. jets or events) are referred to as *detector-level-only* entries and may arise from two effects:

a) **Generator level phase space restriction**
The generator level phase space of the migration matrix may not cover the phase space to include all generated events. For instance, there are cuts to restrict the phase space of the measurement, like the inelasticity y or the invariant mass of the two leading jets M_{12} (or the observable in which the measurement is binned in, like e.g. Q^2, or the jet transverse momentum $p_{\mathrm{T}}^{\mathrm{jet}}$). If migrations in such a quantity occur, such an event may appear in the detector level phase space, but cannot be assigned to a generator level bin.

b) **Measurements of multiplicities**
Measurements, which are sensitive to the multiplicity of a quantity, can have different multiplicities on generator level than on detector level, or the quantity cannot be matched together between the two levels. This is the case for the inclusive jet measurement, where each single jet is measured. The jet-multiplicity is defined on detector level and generator level individually through the jet-finding algorithm and different jet-multiplicities can be present.

In order to fulfill equation 5.10, detector-level-only contributions have to be included in the migration histogram.

In the NC DIS measurement, detector-level-only contributions only of type a) can be present, but only if the generator level phase space is restricted. The inclusive jet measurement, however, has different jet-multiplicities on detector and generator level and, hence, detector-level-only jets of type b) are present. Furthermore, the inclusive jet measurement has detector-level-only contributions of type a) due to the NC DIS phase space, but also contributions of type a) due to the jet phase space restrictions from jet-quantities. Such contributions occur, for instance, if a jet was generated with a transverse momentum of 3.5 GeV, but was reconstructed with only 2.8 GeV and since jets only above 3 GeV are constructed by the jet algorithm, the reconstructed jet becomes orphan.

The dijet and trijet measurements exhibit detector-level-only contributions of type a) and of type b) because of their sensitivity to the jet-multiplicity[7].

6.5.2 Simple inclusion of detector-level-only entries into the unfolding

Two simple solutions for the inclusion of detector-level-only contributions are feasible.
The detector-level-only entries can be included into the migration matrix in one additional generator level bin (i.e. a row in the matrix), which is not related to a certain phase space region. When such a migration matrix is employed for the unfolding, this will result in a large (statistical) uncertainty of that generator level bin after the unfolding process, since that bin is not well constrained by the data [192]. As a consequence, that

[7]For the dijet and trijet measurements, certain event topologies may be measured on detector level, which can be interpreted either as type a) or as type b). Assuming a detector-level dijet event, with two jets of transverse momentum of 3.5 GeV and thus $\langle p_{\mathrm{T}} \rangle = 3.5$ GeV. The jets are generated, however, with a transverse momentum of 2.5 and 3.3 GeV a hence $\langle p_{\mathrm{T}} \rangle = 2.9$ GeV. When the sub-leading jet is rejected due to a minimum $p_{\mathrm{T}}^{\mathrm{jet}}$requirement, this detector-level-only event is of type a), but if this jet is allowed to be used for the calculation of $\langle p_{\mathrm{T}} \rangle$, this event is of type b), if the generator level phase space is constrained to $\langle p_{\mathrm{T}} \rangle > 3.0$ GeV.

bin may have large (negative) correlations w.r.t. the bins of the phase space. This will result in positive correlations of the generator level phase space bins and is similar to a normalization uncertainty of the unfolded result.
Another possibility to include detector-level-only entries in an unfolding process, is the subtraction of these contributions prior to the unfolding similar to a background contamination. Since this contribution has to be determined solely from Monte Carlo generators, it has to be assigned a normalization uncertainty which can be large for instance for multiplicity measurements, since differing multiplicites on generator and detector level cannot be measured in a dedicated analysis. If treated correctly, this will result in the same size of correlated statistical uncertainties of the phase space bins, as if this row would be included in the unfolding process, as an unconstrained generator level bin.

6.5.3 Improved inclusion of detector-level-only entries into the migration matrix

In order to include the detector-level-only contributions in the migration scheme, the full information on such kind of events is exploited. In a first step, the generator level phase space of the NC DIS event selection is extended to the entire generated phase phase. In observables like x_{Bj} or y, or at the upper phase space boundary in Q^2 all events are then included in the migration scheme on generator-level. At the lower phase space boundary in Q^2, the phase space is chosen such, that all generated events are included. In this analysis, the signal Monte Carlo events are generated with $Q^2 > 60\,\mathrm{GeV}$. Events with lower generated values of Q^2 may still migrate into the detector level phase space due to wrongly reconstructed quantities. These events are treated as background and are subtracted prior to the unfolding process, including a normalization uncertainty of this background contribution of 10 %. This ensures that detector-level-only entries of type a) are not present for the NC DIS measurement.

The three jet-measurements have detector-level-only contributions of type a) and type b). This may be either single jets in case of the inclusive jet measurement, or an event in case of the dijet or the trijet measurement. Although, the jet-quantities cannot be assigned a corresponding generator-level quantity, there is always the event information of the underlying NC DIS event available. Hence, this NC DIS information is used to estimate the amount of detector-level-only contributions in a three step procedure:

1. The detector-level-only contributions are binned on detector-level according to the measured quantities (Q^2,y,p_{T}). On generator level, these contributions are binned in the bin rid of the NC DIS measurement, using the NC DIS observables, since no jet-quantity is available.

2. The detector-level-only contributions are included, using the detector-level weight of that MC event in the unfolding scheme (see figure 6.1) in the sub-matrix, which connects the jet-measurement on detector-level to the NC DIS measurement on generator level. In figure 6.1, these matrices are denoted $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ for the inclusive jet, dijet and trijet measurement, respectively.

3. In order to preserve the normalization of the NC DIS measurement (c.f. equation 5.11), for each detector-level-only contribution, the detector level weight has to be subtracted from the normalization bin of the generator level bin of the NC DIS

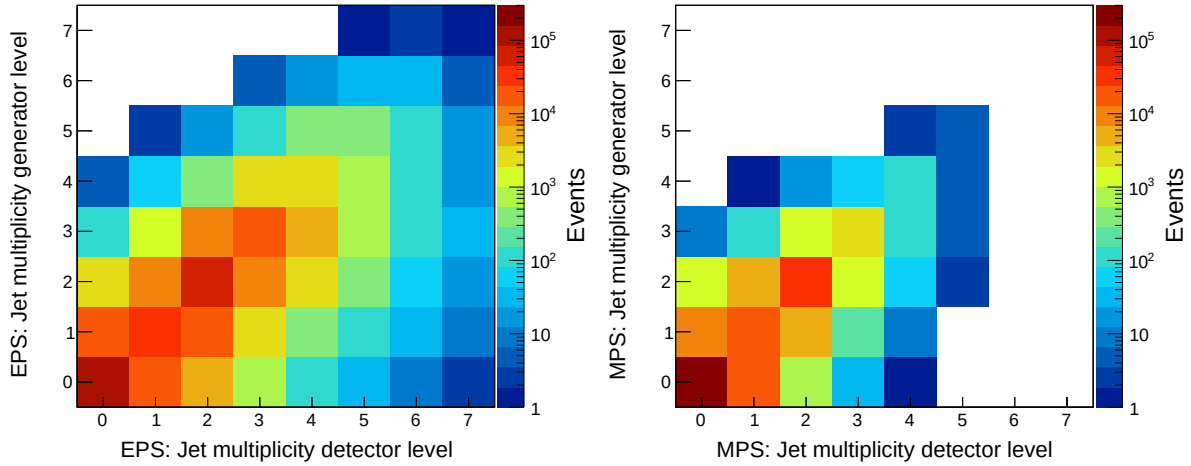


Figure 6.3: Jet multiplicities on generator level versus detector level for the extended phase space (EPS) (left) and the measurement phase space (MPS) (right) for NC DIS events. The color coding indicates the number of events. The reduced requirement on the individual $p_{\mathrm{T}}^{\mathrm{jet}} > 3\,\mathrm{GeV}$ for the EPS compared to the MPS, where it is $p_{\mathrm{T}}^{\mathrm{jet}} > 7\,\mathrm{GeV}$, largely increases the number of found jets. The difference in the jet-multiplicities on the two levels leads to orphan jets, which cannot be connected to the opposite level. These are denoted as *detector-level-only* or *generator-level-only* jets. Such contributions are, however, eve possible when $N_{\mathrm{jet}}^{\mathrm{rec}} = N_{\mathrm{jet}}^{\mathrm{gen}}$.

measurement. The sum of these negative contributions are denoted $-\beta_1$, $-\beta_2$ and $-\beta_3$ for the inclusive-jet, dijet and trijet measurement, respectively, in figure 6.1.

This procedure enables to unfold the detector-level-only entries, using the information of the NC DIS observables. Since the NC DIS measurement is entirely determined from data in the unfolding process, the amount of detector-level-only contributions is therefore also determined from data instead of MC event generators. The influence of detector-level-only contributions of the jet-measurements onto the NC DIS measurement is negligible, since the NC DIS statistics is three to four orders of magnitude larger than the detector-level-only contributions of the jet-measurements in each generator level bin.

6.6 Unfolding of the inclusive jet measurement

For the determination of the migration matrix for the inclusive jet measurement, a jet-matching has to be performed, which connects jets on detector level with jets on generator level.

6.6.1 Jet multiplicities and jet-matching

The jet algorithm defines 'jets' through clustering of HFS objects on detector level and final state particles on generator level. The jet multiplicities on generator and detector level are shown in figure 6.3 for the EPS and the MPS. The extended phase space has an increased number of jets and closely related to that, also higher jet multiplicities. The jet multiplicities can differ on generator and detector level for individual events. In order to determine the jet resolution and the kinematic migrations, jets of the two levels have to be matched together.

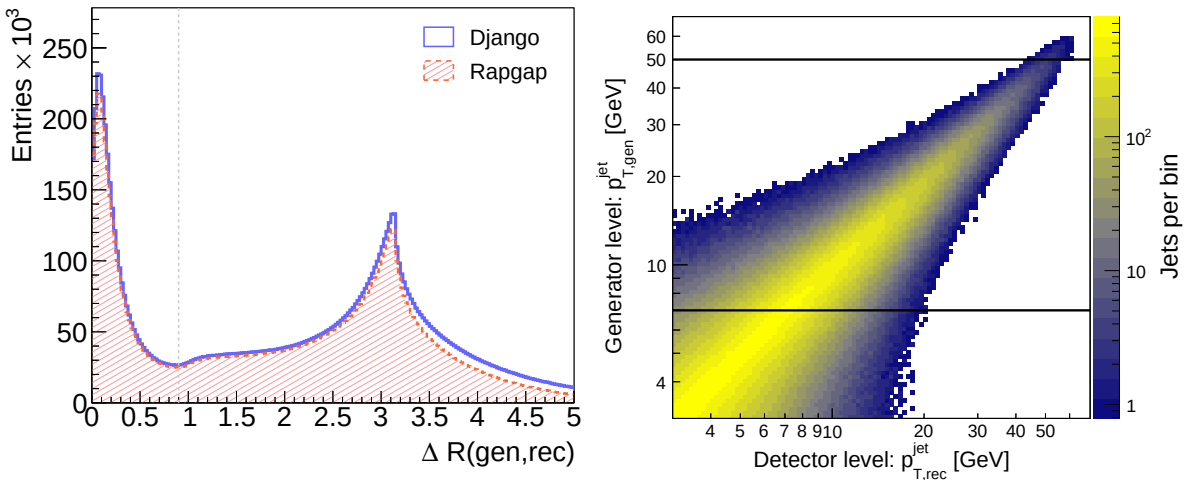


Figure 6.4: Distance of each detector level jet to each generator level jet for Django and Rapgap events (left). The grey dashed line shows the maximum distance for the jet-matching criterion. The right plot shows the relation of the jet transverse momentum between generator and the detector level for the matched jets. This logarithmic plot illustrates the linear response after calibration and software reweighting of the H1 detector.

Here, a closest pair algorithm with a distance measure

$$(\Delta R)_{ij} = \sqrt{(\Delta\eta^{\text{jet}}_{\text{lab}})^2_{ij} + (\Delta\phi)^2_{ij}} \tag{6.2}$$

is applied, where $(\Delta\eta^{\text{jet}}_{\text{lab}})_{ij} = \eta^{\text{lab}}_{i,\text{rec}} - \eta^{\text{lab}}_{j,\text{gen}}$ and $(\Delta\phi)_{ij} = \phi_i - \phi_j$ are the distance of a detector level jet i to a generator level jet j in the laboratory frame pseudorapidity and the azimuthal angle, respectively. The distance measure must fulfill

$$(\Delta R)_{ij} < 0.9 \tag{6.3}$$

so that no geometrical opposite jets are matched. The distance $(\Delta R)_{ij}$ of generator level jets to detector level jets is depicted in figure 6.4. The closest pair algorithm calculates an ordered list of all distances $(\Delta R)_{ij}$ between generator level and detector level jets. Iteratively, the pair with the smallest distance is matched together and the remaining distances to these two jets are removed from the list, until no pairs are left. The connected jets are used for determining the migrations in the migration matrix. The values of the transverse momentum $p^{\text{jet}}_{\text{T}}$ for the matched jets are shown as $p^{\text{gen}}_{\text{T}}$ versus $p^{\text{rec}}_{\text{T}}$ in figure 6.4. The jets, which could not be matched, are treated as detector-level-only or generator-level-only entries, following section 6.5, when filling the migration histogram. The advantage of the extended phase space becomes obvious when regarding the amount of detector-level-only jets within all detector-level-jets, for matched jets in the MPS or in the EPS. At the lower MPS boundary at $p^{\text{jet}}_{\text{T}} = 7\,\text{GeV}$ the amount of detector-level-only jets is reduced from 65 % to only 7 % and to only 1 % for $p^{\text{jet}}_{\text{T}} > 11\,\text{GeV}$, as is can be seen in figure 6.5[8].
The distances $(\Delta R)_{ij}$ are calculated, using laboratory frame observables [79], i.e. $\eta^{\text{jet}}_{\text{lab}}$ and $\phi^{\text{jet}}_{\text{lab}}$. These observables are preferred over quantities in the Breit frame of reference, since they are not affected by resolution effects, which enter calculating the variables for the boost to the Breit frame. Hence, the four-vectors of the jets, which were found in

[8] In this figure, the inclusive jet 'EPS' is only extended in jet quantities w.r.t. the MPS, i.e. in $\eta^{\text{jet}}_{\text{lab}}$ and $p^{\text{jet}}_{\text{T}}$. The phase space is not enlarged in Q^2 and y. On detector-level, the phase space is only extended in $p^{\text{jet}}_{\text{T}}$.

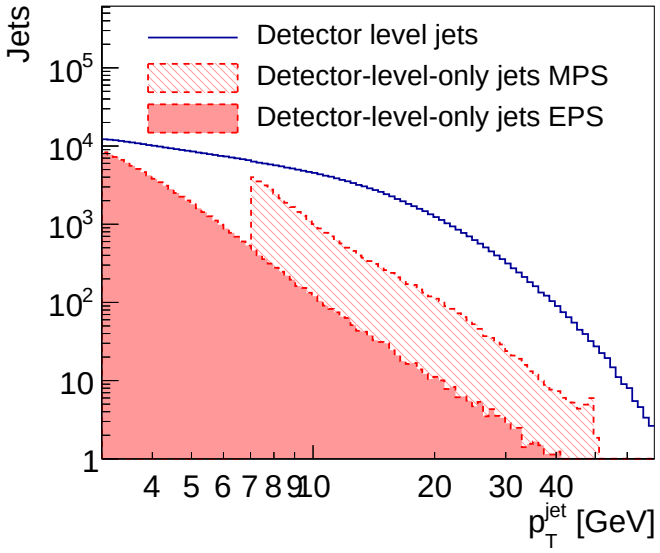


Figure 6.5: Number of detector-level-only jets compared to the total amount of all detector level jets. At the lowest p_T^{jet}, the amount of detector-level only jets in MPS is around 65 % of the measured jet sample due to the finite resolution of the detector. The extension of the phase space reduces the amount of detector-level-only entries within MPS to only 1-6 %, and thus increases significantly the stability of the measurement.

the Breit frame of reference, are boosted back to the laboratory frame of reference, using the inverse of the boost vector which was used to determine the Breit frame four-vectors. The resolution of the pseudorapity in the laboratory frame is significantly better than its resolution in the Breit frame, as shown in figure 6.6. However, the impact on the distance measure when calculated in the Breit or the laboratory frame of reference, was found to be insignificant (c.f. section 9.2.2).
The distance measure only includes geometrical distances and no kinematic variables. This ensures that no biases in relevant observables are introduced through the jet-matching procedure. However, if the jet-matching is performed using a kinematic distance measure, like Δp_T^{jet}, no difference on the unfolded result is visible (c.f. section 9.2.2).

6.6.2 Migration scheme for the inclusive jet measurement

The unfolding of the inclusive jet measurement is performed in a four-dimensional unfolding scheme in the variables y, Q^2, η_{lab}^{jet} and p_T^{jet}. The unfolding scheme is depicted in table 6.2. The four-dimensional unfolding scheme is technically set up by inserting a Q^2 bin grid into each bin of the p_T^{jet} bin grid, which is inserted in each bin of the η_{lab}^{jet} bin grid, which is inserted in each bin of the y bin grid. The bin grid and the DIS phase space definitions for the MPS and the EPS are identical to the one for the NC DIS unfolding matrix. The measurement phase space (MPS) of jets is defined by requiring $-1.0 < \eta_{lab}^{jet} < 2.5$ and $7 < p_T^{jet} < 50\,\text{GeV}$. The EPS is defined by extending p_T^{jet} in order to account for migrations into and out of the phase space down to $p_T^{jet} = 3\,\text{GeV}$. Migrations in the jet pseudorapidity η_{lab}^{jet} are taken into account through two side-bins up to $\eta_{lab}^{jet} = 2.75$ and down to $\eta_{lab}^{jet} = -1.5$. The two bins are differential in Q^2, y and p_T^{jet}, while the y and p_T^{jet} bin grids are identical to the bin grids in the MPS. Since the statistics in these side-bins is too small for a fully differential Q^2 bin grid, only three bins in Q^2 are employed. The cause of migrations in η_{lab}^{jet} are detector effects in the reconstruction of the HFS, which are mostly independent of the scattered electron kinematics and are therefore largely independent of Q^2. The bin grids in

Inclusive jet measurement

Migration scheme	Four-dimensional unfolding in Q^2 y, $p_{\mathrm{T}}^{\mathrm{jet}}$, $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$side-bins differential in Q^2, y and $p_{\mathrm{T}}^{\mathrm{jet}}$
Generator level	
NC DIS measurement phase space	$150 < Q^2 < 15\,000\,\mathrm{GeV}^2$
	$0.2 < y < 0.7$
Inclusive jet MPS	$7 < p_{\mathrm{T}}^{\mathrm{jet}} < 50\,\mathrm{GeV}$
	$-1.0 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.5$
Inclusive jet EPS	$3 < p_{\mathrm{T}}^{\mathrm{jet}} < 100\,\mathrm{GeV}$
	$-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75$
Q^2 and y binning	c.f. NC DIS measurement
$p_{\mathrm{T}}^{\mathrm{jet}}$ number of bins (measurement)	10 (4)
$p_{\mathrm{T}}^{\mathrm{jet}}$ bin grid [GeV]	3 – 5 – 7 – 9 – 11 – 14 – 18 – 22 – 30 – 50 – 100
$p_{\mathrm{T}}^{\mathrm{jet}}$ measurement bin grid [GeV]	7 – 11 – 18 – 30 – 50
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ number of bins (measurement)	2 (1)
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ bin grid	-1.0 – 0.75 – 2.5
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bins	2 $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bins, differential in Q^2, y and $p_{\mathrm{T}}^{\mathrm{jet}}$ with $-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < -1.0$ and $2.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75$
differential binning of $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bin	$p_{\mathrm{T}}^{\mathrm{jet}}$and y binning as in EPS times three Q^2 bins: $60 – 230 – 700 – 40\,000\,\mathrm{GeV}^2$
Detector level	
Q^2 and y bin grid	c.f. NC DIS measurement
$p_{\mathrm{T}}^{\mathrm{jet}}$ number of bins	16
$p_{\mathrm{T}}^{\mathrm{jet}}$ bin grid [GeV]	3 – 4.4 – 6.0 – 7 – 8 – 9 – 10 – 11.3 – 14 – 15.8 – 18 – 19.8 – 22 – 28 – 34 – 50 – 100
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ number of bins	3
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ bin grid	-1.0 – 0.0 – 1.2 – 2.5
$\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bins	2 $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bins, differential in Q^2, y and $p_{\mathrm{T}}^{\mathrm{jet}}$ $-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < -1.0$ and $2.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75$
differential binning of $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ side-bin	$p_{\mathrm{T}}^{\mathrm{jet}}$and y binning as the detector level bin grids times two Q^2 bins: $60 – 400 – 40\,000\,\mathrm{GeV}^2$

Table 6.2: Unfolding scheme for the inclusive jet measurement.

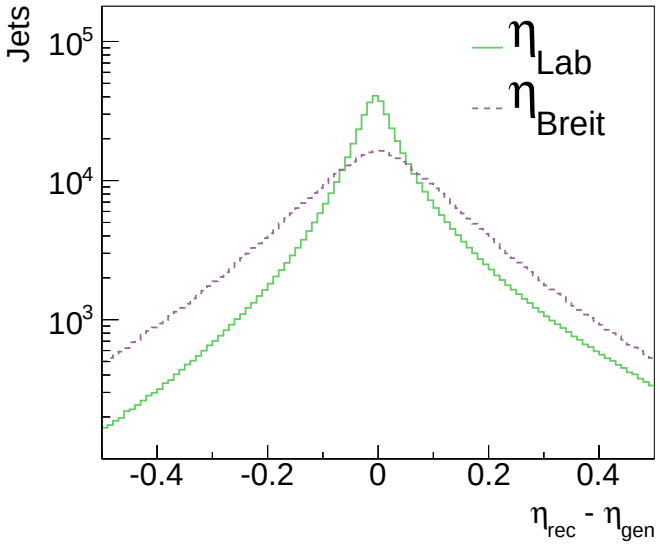


Figure 6.6: Resolution of the jet pseudorapidity η in the laboratory and the Breit frame of reference for inclusive jets which are matched to the opposite level.

Q^2 and y are shown in 6.7, illustrating the size of the migrations. In order to account for the sizeable migrations in $p_{\mathrm{T}}^{\mathrm{jet}}$, the number of bins on detector level is increased, which leads to very small bin sizes, down to 1 GeV. On generator level each bin of the measurement phase space is subdivided into two bins. These two bins are summed after the unfolding process to determine the cross section in the measurement bin grid[9]. Since migrations in the low-$p_{\mathrm{T}}^{\mathrm{jet}}$ region are large compared to the small bin size, and also the jet production cross section increases significantly, the low-$p_{\mathrm{T}}^{\mathrm{jet}}$ region is extended using two $p_{\mathrm{T}}^{\mathrm{jet}}$ side-bins on generator level and three side-bins on detector level, each between 3 to 7 GeV.

6.7 Migration scheme for the dijet measurement

The unfolding of the dijet measurement is performed using a three-dimensional unfolding scheme in the kinematic observables Q^2, y and $\langle p_{\mathrm{T}} \rangle$. The details are given in table 6.3. Technically, the $\langle p_{\mathrm{T}} \rangle$ bin grid is inserted in each Q^2 bin, and the Q^2 bin grid is inserted in each y bin. The Q^2 bin grid is extended with additional side-bins in quantities, which define the dijet phase space.
The migrations in Q^2 and y are given by the reconstruction of the NC DIS kinematics, and thus the same bin grid for these two observables are used as for the NC DIS measurement (c.f. table 6.1).
It was found that migrations in $\langle p_{\mathrm{T}} \rangle$ are dominant and cause the largest difference between the two employed models. Therefore, the binning on detector level was chosen to be very fine with a bin width of 1.0 GeV in the low $\langle p_{\mathrm{T}} \rangle$-region, and the bins on generator level are subdivided into multiple bins and are recombined for the measurement binning. The size of the migrations in $\langle p_{\mathrm{T}} \rangle$ is shown in the detector and generator level bin grids in figure 6.8.

The migrations of the dijet measurement are related to the migrations of individual jets, but no jet-matching has to be performed. Instead, a dijet event is given by a certain event

[9]The measurement is performed in four bins from $\langle p_{\mathrm{T}} \rangle = 7$ to 50 GeV, which is referred to as the 'bin grid of the measurement'. The bin grid on generator level in the MPS incorporates a finer binning of eight bins and should not be mistaken with the first.

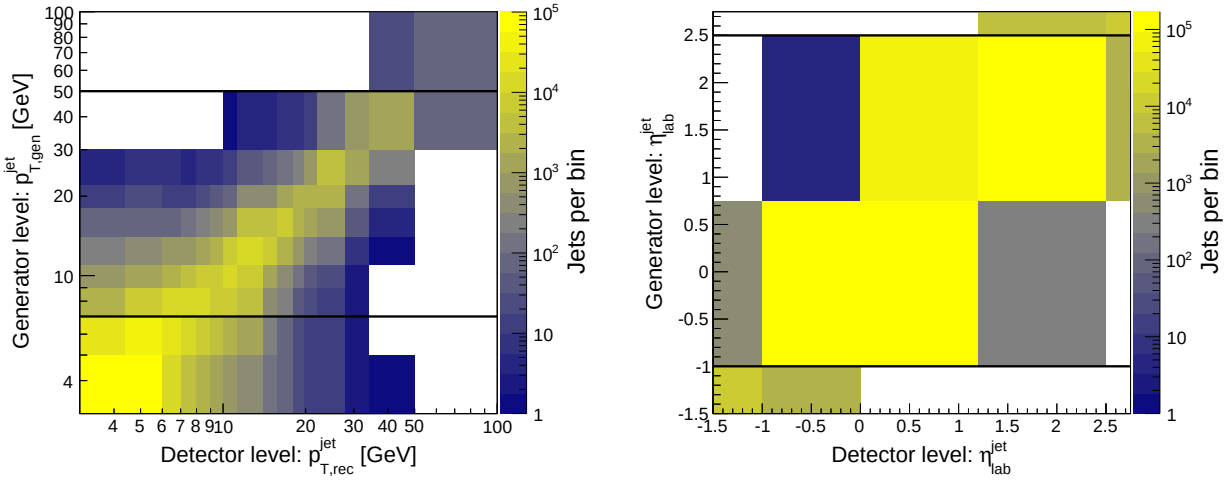


Figure 6.7: Binning scheme of the migration histogram for the inclusive jet measurement in $p_{\mathrm{T}}^{\mathrm{jet}}$ (left) and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ (right) in the Breit frame. The unfolding scheme of the inclusive jet measurement is four-dimensional in the variables Q^2, y, $p_{\mathrm{T}}^{\mathrm{jet}}$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$. The black lines indicate the phase space of the measurement, where the $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ bins and some $p_{\mathrm{T}}^{\mathrm{jet}}$ bins are added for the final measurement. The side-bins in $\eta_{\mathrm{lab}}^{\mathrm{jet}}$, i.e. where $\eta_{\mathrm{lab}}^{\mathrm{jet}} < -1.0$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}} > 2.5$, are not fully differential in Q^2, because of the low statistics and the Q^2 bin grid is reduced to a 3×2 bin grid instead for this phase space region. The color coding shows the number of matched jets from the MC predictions after integration over the other variables.

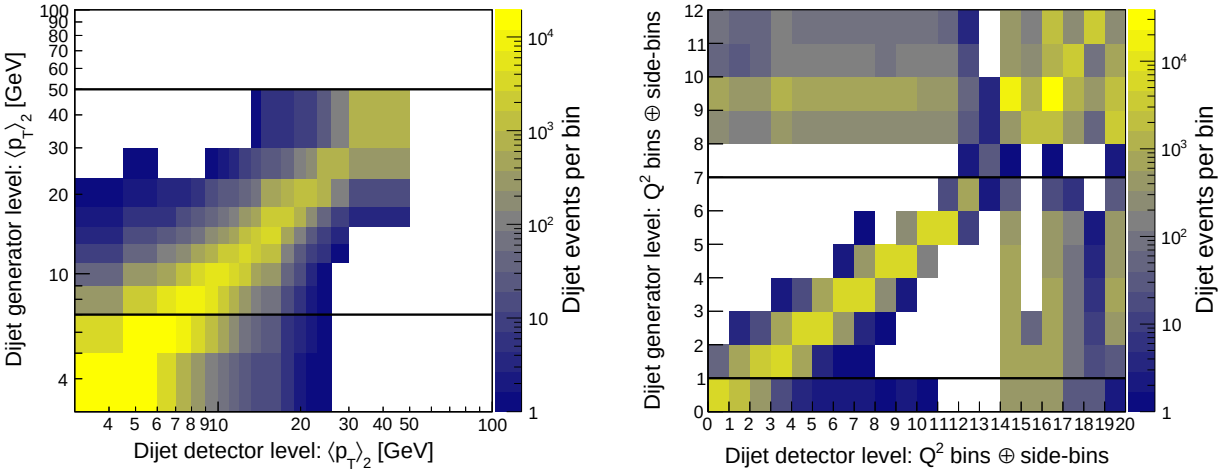


Figure 6.8: Binning scheme of the dijet measurement in $\langle p_{\mathrm{T}} \rangle$ (left). The unfolding of the dijet measurement is based on a three-dimensional unfolding scheme in Q^2, y and $\langle p_{\mathrm{T}} \rangle$. The phase space of the dijet measurement is extended through additional bins in Q^2 and side-bins as shown in the right plot. The bins 1-14 (1-8) on detector (generator) level refer to the Q^2 migrations and six (four) additional bins for dijet events in the extended phase space in M_{12}, $p_{\mathrm{T}}^{\mathrm{jet}}$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$. See the table 6.4 for the exact definition of these side-bins. By including these side-bins in the Q^2 bin grid, these bins are differential in $\langle p_{\mathrm{T}} \rangle$ (and y). The black lines indicate the phase space of the measurement.

Dijet measurement

Migration scheme	Three-dimensional unfolding in Q^2, y and $\langle p_\mathrm{T} \rangle$ and side-bins in M_{12}, $p_\mathrm{T}^\mathrm{jet2}$ and $\eta_\mathrm{lab}^\mathrm{jet}$
Generator level	
NC DIS measurement phase space	$150 < Q^2 < 15\,000\,\mathrm{GeV}^2$
	$0.2 < y < 0.7$
Dijet MPS	$7 < \langle p_\mathrm{T} \rangle < 50\,\mathrm{GeV}$
	$-1.0 < \eta_\mathrm{lab}^\mathrm{jet} < 2.5$
	$M_{12} > 16\,\mathrm{GeV}$
	$p_\mathrm{T}^\mathrm{jet} < 50\,\mathrm{GeV}$
	$p_\mathrm{T}^\mathrm{jet2} > 5\,\mathrm{GeV}$
Q^2 and y binning	c.f. NC DIS measurement
$\langle p_\mathrm{T} \rangle$ number of bins (measurement)	11 (4)
$\langle p_\mathrm{T} \rangle$ bin grid [GeV]	3 – 5 – 7 – 9 – 11 – 13 – 15 – 18 – 23 – 30 – 50 – 100
$\langle p_\mathrm{T} \rangle$ measurement bin grid [GeV]	7 – 11 – 18 – 30 – 50
Side-bins	Four side-bins for $M_{12} < 16\,\mathrm{GeV}$, $p_\mathrm{T}^\mathrm{jet2} < 5\,\mathrm{GeV}$, $\eta_\mathrm{lab}^\mathrm{jet} < -1.0$ and $\eta_\mathrm{lab}^\mathrm{jet} > 2.5$, following table 6.4, where each bin includes the $\langle p_\mathrm{T} \rangle$ and y bin grid
Detector level	
Q^2 and y bin grid	c.f. NC DIS measurement
$\langle p_\mathrm{T} \rangle$ number of bins	18
$\langle p_\mathrm{T} \rangle$ bin grid [GeV]	3 – 4.5 – 6 – 7 – 8 – 9 – 10 – 11 – 12 – 13.2 – 14.4 – 17 – 19 – 21 – 23 – 26 – 30 – 50 – 100
Side-bins	Six side-bins for migrations in M_{12}, $p_\mathrm{T}^\mathrm{jet2}$ and $\eta_\mathrm{lab}^\mathrm{jet}$ following the definition of table 6.4

Table 6.3: Unfolding scheme for the dijet measurement.

Side-bins of dijet unfolding

Detector level bin	Generator level bin	Necessary distinction to dijet MPS
15	10	$M_{12} < 16\,\text{GeV}$
17	10	$3 < p_{\text{T}}^{\text{jet2}} < 5\,\text{GeV}$ and $M_{12} < 16\,\text{GeV}$
20	9	$3 < p_{\text{T}}^{\text{jet2}} < 5\,\text{GeV}$ and $Q^2 > 200\,\text{GeV}^2$
16	9	$3 < p_{\text{T}}^{\text{jet2}} < 5\,\text{GeV}$ and $Q^2 < 200\,\text{GeV}^2$
18	11	$-1.0 < \eta_{\text{lab}}^{\text{jet}} < 2.75$
19	12	$-1.5 < \eta_{\text{lab}}^{\text{jet}} < 2.5$

Table 6.4: Side-bins of the dijet unfolding scheme for quantities which define the dijet phase space. The given bin numbers refer to the generator level bin numbering of the exemplary matrix in figure 6.8. The two bins 18 and 19, where the phase space is extended in $\eta_{\text{lab}}^{\text{jet}}$, the dijet events are searched in the EPS. All bins incorporate a full $\langle p_{\text{T}} \rangle$ and y matrix. Side-bins of the trijet unfolding are defined analogously using $p_{\text{T}}^{\text{jet3}}$ and M_{12}.

topology, with requirements on individual jets (e.g. $p_{\text{T}}^{\text{jet}}$ or $\eta_{\text{lab}}^{\text{jet}}$), combined jet quantities (e.g. M_{12}) or event quantities (e.g. two jets are required). Hence, migrations between the bins involving these quantities defining a dijet event are present. The migrations are caused by resolution effects of individual jets or arise from differing jet multiplicities on detector and generator level.

Migrations into and out of the dijet phase space are determined in an enlarged phase space (EPS) through the extension of the MPS in the variables $\langle p_{\text{T}} \rangle$, M_{12}, $p_{\text{T}}^{\text{jet2}}$ and $\eta_{\text{lab}}^{\text{jet}}$. Six additional bins represent events with at least two jets, where the dijet phase space is extended in one or two observables compared to the MPS. The definition of these six bins is outlined in table 6.4. One bin extends the EPS with dijet events where the $p_{\text{T}}^{\text{jet}}$ of the sub-leading jet is below 5 GeV, one bin where the invariant mass of the two leading jets fulfills $M_{12} < 16\,\text{GeV}$ and one bin where these two requirements of the MPS are not fulfilled[10]. Since migrations in $p_{\text{T}}^{\text{jet2}}$ are large, because of the decreased resolution and the increased jet-production cross section, the bin with $p_{\text{T}}^{\text{jet2}} < 5\,\text{GeV}$ is subdivided into two bins in Q^2 with $Q^2 < 200\,\text{GeV}^2$ and $Q^2 > 200\,\text{GeV}^2$. Two bins with $\eta_{\text{lab}}^{\text{jet}} > 2.5$ and $\eta_{\text{lab}}^{\text{jet}} < -1.0$ represent migrations of individual jets in $\eta_{\text{lab}}^{\text{jet}}$[11]. These bins contribute to the measurement, if for instance one out of two jets in an event is generated at $\eta_{\text{lab}}^{\text{jet}} < 2.5$ and reconstructed at $\eta_{\text{lab}}^{\text{jet}} > 2.5$.

The migrations in M_{12} and $p_{\text{T}}^{\text{jet2}}$ are closely related the jet transverse momenta, but can be assumed to be independent of the electron kinematics (i.e. of Q^2). Therefore, these side-bins are included in the migration scheme into the Q^2 bin grid (the structure of the Q^2 dimension is shown as a binned histogram in figure 6.8) and each of these side-bins

[10] This definition of the extension of the MPS is essentially a 2×2 matrix, where one bin is the MPS. Since all bins incorporate a full $\langle p_{\text{T}} \rangle$-migration scheme, the $\langle p_{\text{T}} \rangle$ bin grid must span the phase space of $3 < \langle p_{\text{T}} \rangle < 50\,\text{GeV}$.

[11] In order to have a properly defined EPS (i.e. no multivariate definition), the events for the enlarged $\eta_{\text{lab}}^{\text{jet}}$-phase space are searched in the EPS with the following cuts: $3 < p_{\text{T}}^{\text{jet}} < 50\,\text{GeV}$, no cut on M_{12}, $\langle p_{\text{T}} \rangle > 3\,\text{GeV}$ and two jets in the given pseudorapidity region. However, these events are not allowed to have two jets in the $\eta_{\text{lab}}^{\text{jet}}$-region of $-1.0 < \eta_{\text{lab}}^{\text{jet}} < 2.5$.

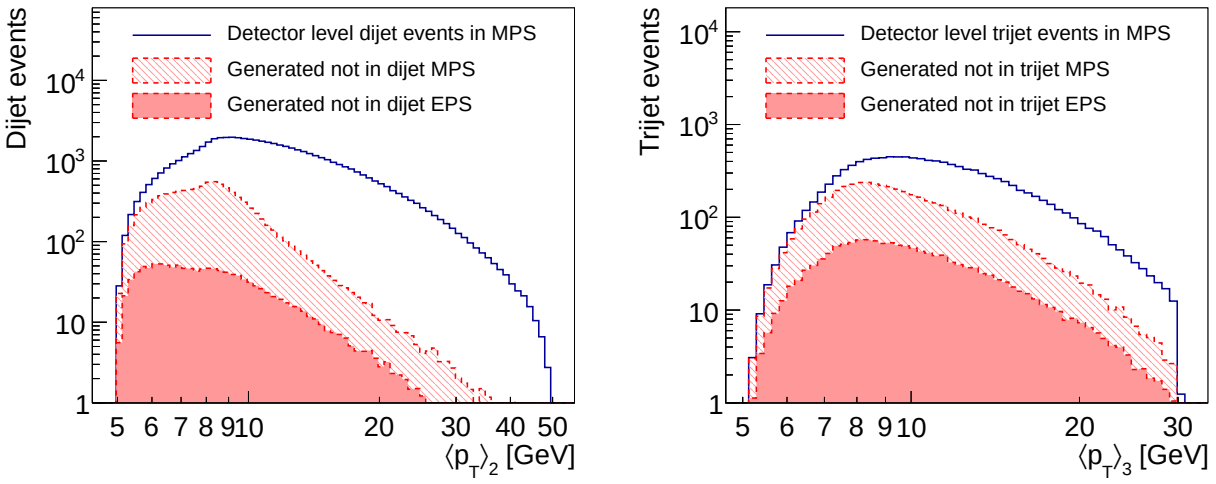


Figure 6.9: The amount of detector-level-only entries for the dijet (left) and the trijet (right) measurement, compared to the amount of all detector-level events in the MPS. Shown are events which are on detector level in the MPS, but are not in the MPS or EPS on generator level. The increased phase space on generator level, in the quantities M_{12}, p_T^{jet} and η_{lab}^{jet}, enable, to determine the migrations and thus improves the stability of the measurement.

include a full $\langle p_T \rangle$ bin grid.
The extension of the MPS reduces significantly the amount of detector-level-only dijet events. The detector-level-only contributions in bins of the MPS are almost negligible as it can be seen in figure 6.9, which increases substantially the stability of the measurement, especially at low values of $\langle p_T \rangle$, where migrations among the jet-algorithm threshold of $p_T^{jet} > 3\,\text{GeV}$ become relevant. The few amount of remaining detector-level-only dijet event are estimated by NC DIS events (see section 6.5.3).
The phase space of dijet events is further constrained by $p_T^{jet} < 50\,\text{GeV}$. Migrations near this upper limit are negligible because of the small cross section and the good resolution, and they are not taken into account.
Migrations of one-jet events to dijet events are not included in the migration scheme, because such monojet events contribute only insignificantly to the generator level MPS. These contributions could be at most of the size of the remaining detector-level-only contributions as shown in figure 6.9. However, a brief study of such contributions is performed in section 9.2.3.

6.8 Migration scheme for the trijet measurement

The unfolding of the trijet measurement is performed in a three-dimensional unfolding scheme in the variables y, Q^2 and $\langle p_T \rangle$, with additional side-bins for quantities, which define the trijet phase space. A detailed description is found in table 6.5. The technical implementation of the three dimensional bin grid is identical to the dijet unfolding scheme, where the dijet-$\langle p_T \rangle_2$ is replaced by the trijet-$\langle p_T \rangle_3$. Due to the decreased statistic of the trijet measurement compared to the dijet measurement, the $\langle p_T \rangle$ bin grid has a reduced number of bins compared to the dijet bin grid on detector and on generator level. The generator level bin grid is identical to the MPS and is extended with side-bins. The increased stability of the trijet measurement through the EPS is illustrated in figure 6.9 (right). The number of bins in $\langle p_T \rangle$ on detector level is chosen to be 1.5 times the number of $\langle p_T \rangle$-bins on generator level. The side-bins of the trijet phase space in M_{12}, η_{lab}^{jet} and

Trijet measurement

Migration scheme	Three-dimensional unfolding in Q^2, y and $\langle p_T \rangle$ and side-bins in M_{12}, p_T^{jet2} and η_{lab}^{jet}
Generator level	
NC DIS measurement phase space	$150 < Q^2 < 15\,000\,\text{GeV}^2$
	$0.2 < y < 0.7$
Trijet MPS	$7 < \langle p_T \rangle < 30\,\text{GeV}$
	$-1.0 < \eta_{lab}^{jet} < 2.5$
	$M_{12} > 16\,\text{GeV}$
	$p_T^{jet} < 50\,\text{GeV}$
	$p_T^{jet3} > 5\,\text{GeV}$
Q^2 and y binning	c.f. NC DIS measurement
$\langle p_T \rangle$ number of bins (measurement)	6 (3)
$\langle p_T \rangle$ bin grid [GeV]	3 – 5 – 7 – 11 – 18 – 30 – 100
$\langle p_T \rangle$ bin grid measurement [GeV]	7 – 11 – 18 – 30
Side-bins	Six side-bins for M_{12}, p_T^{jet3} and η_{lab}^{jet} following table 6.4 including the $\langle p_T \rangle$ and y generator level bin grid
Detector level	
Q^2 and y bin grid	c.f. NC DIS measurement
$\langle p_T \rangle$ number of bins	9
$\langle p_T \rangle$ bin grid [GeV]	3 – 5 – 6.7 – 8.4 – 10.5 – 13.5 – 17 – 22 – 30 – 100
Side-bins	Four side-bins for M_{12}, p_T^{jet3} and η_{lab}^{jet} following table 6.4 including the $\langle p_T \rangle$ and y detector level bin grid

Table 6.5: Unfolding scheme for the trijet measurement.

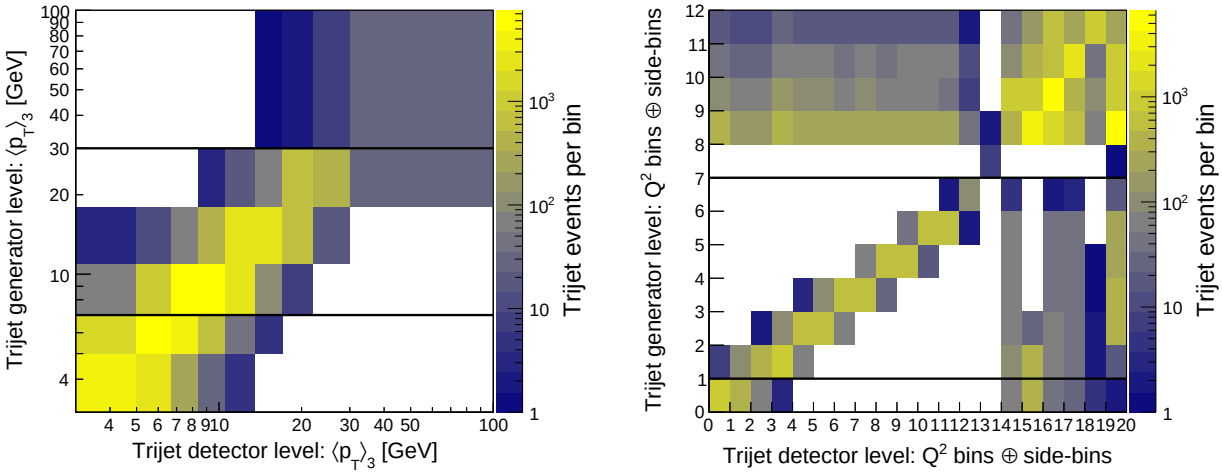


Figure 6.10: Binning scheme of the trijet measurement in $\langle p_{\mathrm{T}} \rangle$ (left). The Q^2 bin grid is extended with additional side-bins for migrations into and out of the trijet MPS. The right figure shows the binned migration histogram in Q^2 together with these side-bins in bins 15-20 and 9-12 on detector and generator level, respectively. The black lines indicate the phase space of the measurement.

$p_{\mathrm{T}}^{\mathrm{jet3}}$ are defined in analogy to the dijet unfolding scheme. The $\langle p_{\mathrm{T}} \rangle$ bin grid and the the size of the migrations are shown in figure 6.10, where the bin grid in Q^2 and the side-bins are shown.

6.9 Radiative corrections

The measured data will be compared to pQCD calculations which do not include higher order QED effects. Therefore, the data have to be corrected for these effects. In this analysis, the radiative corrections are included in the unfolding histogram.
Higher order QED effects are included in the two employed MC event generators Rapgap and Django using the HERACLES program [139] at next-to-leading order precision in α_{em}. These are QED corrections such as real radiation of a photon off the incoming or the outgoing lepton (initial and final state radiation), virtual QED corrections at the lepton vertex and self-energy corrections of the propagators, i.e. fermion loops of the photon propagator and electroweak corrections.
The influence on the cross section due to higher order QED effects may arise from radiated photons, which are reconstructed as jets or which lead to wrongly reconstructed kinematics, since the radiated photon is not associated with the scattered lepton. These effects also affect the reconstruction of the boost variables and therefore also the jet four-vectors are influenced by QED radiation.
The size of higher order QED corrections are quantified for each MC generator by generating an event sample using QED radiative corrections and compared to an event sample where no higher-order QED effects are included[12]. A correction factor c_i^{QED} for each bin i to correct the measurement for higher order QED effects is defined by

$$c_i^{\mathrm{QED}} = \frac{\sigma_i^{\mathrm{NoRad}}}{\sigma_i^{\mathrm{Rad}}}, \tag{6.4}$$

where σ_i^{Rad} and $\sigma_i^{\mathrm{NoRad}}$ are the cross section predictions in bin i with and without the

[12]Since the QED correction factors are determined from two different physics simulations, no migration matrix but only bin-by-bin correction factors for this correction can be derived.

simulation of higher-order QED effects, respectively. The simulation of the non-radiative events includes the running of α_{em} and the renormalized electron mass.
The higher-order QED correction is included in the migration matrix following section 5.5. The correction factor $\gamma_i^{\mathrm{QED}} = \epsilon_{0,i}(1 - c_i^{\mathrm{QED}})x_{0,i}$ is filled into the efficiency bins of the migration histogram prior to the unfolding process. Consequently, the unfolding process corrects the data for detector and QED effects, simultaneously.

6.10 Applying the unfolding

The unfolding histogram $\boldsymbol{N}$, the covariance matrix $\boldsymbol{V}_y$ and the measurement vector y are determined using the unfolding scheme as outlined in the previous sections. The vector y and the covariance matrix $\boldsymbol{V}_y$ are determined from the data (see section 6.3). The migration matrix is determined with the signal Monte Carlo event generators (see section 5.3). The background subtraction is performed following section 5.4.

6.10.1 High-y background subtraction

For the unfolding procedure, the phase space of the inelasticity y is extended with side-bins in order to estimate migrations in y. The low-y side-bin on generator level with $y_{\mathrm{gen}} < 0.2$ is determined from the unfolding, mainly from data of $0.08 < y_{\mathrm{rec}} < 0.2$.
The side-bin on generator level ($y_{\mathrm{rec}} > 0.7$), would have to be determined dominantly from detector level events with $y_{\mathrm{rec}} > 0.7$. In this phase space region, however, the requirement of $E_e > 11\,\mathrm{GeV}$ to insure a high precision measurement of the scattered electron, suppresses significantly the amount of reconstructed events. Therefore, the generator level side-bin $y_{\mathrm{gen}} > 0.7$ cannot be w constrained by data.
In order to estimate the amount of events which are reconstructed with $y_{\mathrm{rec}} < 0.7$ and generated with $y_{\mathrm{gen}} > 0.7$, and hence migrate out of the phase space, these events are subtracted prior to the unfolding. This *high-y contribution* is estimated using the signal event generators Django and Rapgap. The high-y background is treated identically to other background sources, and a normalization uncertainty of 5 % is applied, which is the difference between these two models. The amount of high-y background is around 0.5 % in the MPS. The distribution is exemplarily shown for dijet events in figure 6.11.
Contributions, which are generated in the MPS (i.e. $y < 0.7$), but reconstructed at $y_{\mathrm{rec}} > 0.7$ are treated as inefficencies and are included in the overflow bin.

6.10.2 Unfolding procedure

The unfolded generator level is found by calculating the minimum of χ^2 in equation 5.4. Some regions of the phase space in the multidimensional unfolding scheme are physically not accessible and several bins of the data histogram and of the generator level prediction of the model are empty. These bins are removed from the unfolding. A further number of 20 generator level bins with small but non-vanishing entries are not constrained through the predicted migrations from detector level entries of data. These rows are also removed from the migration histogram. The response matrix which is used by TUnfold has a size of 690×3071 non-vanishing rows and columns.

For the regularization condition the *size* is used, which is expressed by the unity matrix for $\boldsymbol{L}^2$: $\boldsymbol{L}^2 = \mathbb{1}$. The regularization parameter τ is determined by performing the automated

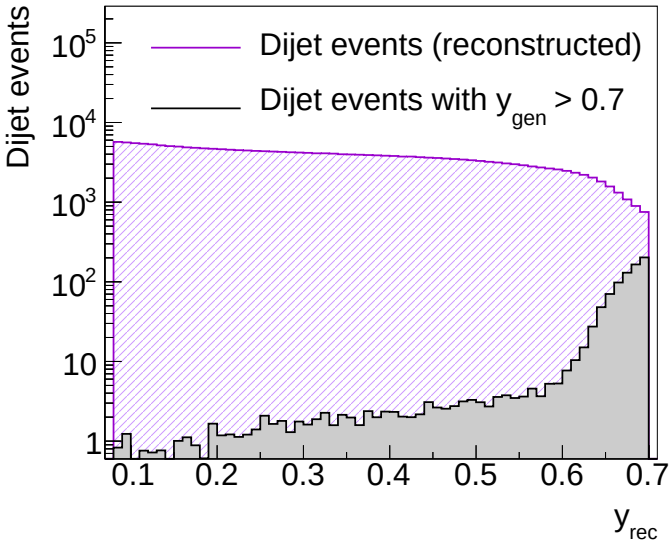


Figure 6.11: Distribution of y of dijet events on detector level from MC prediction. The grey shaded area refers to dijet events, which are reconstructed inside the detector-level phase space $0.08 < y_{\text{rec}} < 0.7$ but are initially generated at $y_{\text{gen}} > 0.7$. This contribution is referred to as 'high-y'-background and is subtracted prior to the unfolding from data.

L-curve scan as implemented in TUnfold [181] using 500 iterations. It is described in section 9.3.2. The L-curve scan finds a regularization parameter of

$$\log_{10} \tau = -4.231. \tag{6.5}$$

This value corresponds to

$$\chi^2_{\text{A}} = 3563.86 \tag{6.6}$$

and

$$\chi^2_{\text{L}} = 1322.55 \tag{6.7}$$

of the L-curve. A detailed study of the regularization parameter and the regularization condition is performed in section 9.3.

The value of

$$\frac{\chi^2_{\text{A}}}{n_{\text{rec}}} = 1.160 \tag{6.8}$$

implies that the found true level x can be folded with the employed migration matrix to describe the detector level data within the statistical precision. The degrees of freedom of the unfolding problem can be estimated by $n_{\text{ndf}} \simeq n_{\text{rec}} - n_{\text{gen}} = 2381$. This expression for the number of degrees of freedom does, however, not incorporate additional degrees of freedom from the regularization condition. Moreover, it ignores statistical correlations of the data, e.g. the trijet events are a subsample of the dijet events. For the χ^2 over the number of degrees of freedom a value of

$$\frac{\chi^2_{\text{A}}}{n_{\text{ndf}}} = 1.497 \tag{6.9}$$

is obtained. This rather large value is expected to be due to the unclear determination of the correct number of degrees of freedom and the systematic uncertainties are not considered in the calculation of χ^2. The value is compatible with values obtained in other unfolding problems (e.g. [193]).

Chapter 7

Cross section results

This chapter summarizes the measured jet cross sections and shows the comparison with theoretical predictions. Normalized jet cross sections, i.e. normalized w.r.t. the inclusive NC DIS cross section, are also presented. We will refer to *absolute cross sections* if jet cross sections are meant literally, i.e. if no normalization is performed. In this analysis an emphasis is put on the simultaneous measurement of the inclusive jet, dijet and trijt cross sections and therefore the results are always displayed together. The numerical values of the cross sections, the uncertainties and the correlation coefficients are given in the appendix A.

7.1 Definition of the cross sections

The measurement is based on data taken by the H1 detector during the HERA-II running period of the HERA collider, corrsponding to an integrated luminosity of

$$\mathcal{L} = 357.60 \pm 8.9\,\mathrm{pb}^{-1} \ .$$

The cross sections for jet-production in DIS are measured in the NC DIS phase space defined by

$$150 < Q^2 < 15\,000\,\mathrm{GeV}^2 \quad \text{and} \quad 0.2 < y < 0.7 \ .$$

Jets are reconstructed in the Breit frame using the k_T jet algorithm with $R_0 = 1$. The definitions of the jet phase space and the bin grids of the three jet measurements can be found in tables 6.2, 6.3 and 6.5 for the inclusive jet, the dijet and the trijet measurement, respectively. Background from other ep processes are subtracted from the data (c.f. section 4.2.2).
The cross sections are quoted on the 'non-radiative hadron' level. The data are corrected for higher order QED effects, except for the running of $\alpha_{em}(Q^2)$, and for acceptance and detector effects, using the previously described regularized unfolding procedure (c.f. chapter 6). Some bins of the measurement bin grid are subdivided for the unfolding process and have to be re-combined. The summation algorithm is implemented in TUnfold and performs also an error propagation for the statistical covariance matrix.

The quoted cross sections σ_i in each bin i are defined through

$$\sigma_i = \frac{N_i^{\mathrm{unfolded}}}{\mathcal{L}}, \tag{7.1}$$

Bin labels Q^2

Bin number q	Q^2 range in GeV^2
1	$150 \leq Q^2 < 200$
2	$200 \leq Q^2 < 270$
3	$270 \leq Q^2 < 400$
4	$400 \leq Q^2 < 700$
5	$700 \leq Q^2 < 5000$
6	$5000 \leq Q^2 < 15\,000$

Bin labels p_T

Bin number p	Label	p_T range in GeV
1	α	$7 \leq p_T < 11$
2	β	$11 \leq p_T < 18$
3	γ	$18 \leq p_T < 30$
4	δ	$30 \leq p_T < 50$

Table 7.1: Bin numbering scheme for Q^2 and p_T-bins. Bins of the double differential measurement are referred to for instance through 3γ or bin number 11, following formula 7.2.

where N_i^{unfolded} is the value of each bin after the unfolding. The variable N_i^{unfolded} includes all corrections above mentioned and corresponds to the measured number of events[1] on non-radiative hadron level in bin i, where no QED effects and other ep processes were present.
The measurements are made differentially in Q^2 and in p_T^{jet} for the inclusive jet or $\langle p_T \rangle_2$ or $\langle p_T \rangle_3$ for the dijet and trijet measurement, respectively[2]. Single bins in Q^2 and p_T are referred to using the bin labels as defined in table 7.1. Bins of the double-differential measurements are labeled using the two single-differential bin labels, e.g. 3γ. It is also convenient to define increasing integer bin numbers b for the double-differential measurements by

$$b = (q-1)n_{p_T} + p \ , \tag{7.2}$$

where $n_{p_T} = 4$ in case of the inclusive jet and dijet measurement, and $n_{p_T} = 3$ for the trijet measurement, and q and p are the bin numbers from table 7.1. Thus, bin $b = 11$ is then identical to bin 3γ.
Two bins of the trijet measurement are not quoted in the final tables. The bin 6α is removed from the measurement due to the small purity and in agreement with previous measurements, and bin 6γ of the trijet measurement is removed due to its very large statistical uncertainty after the unfolding of 89 %.

7.2 Calculation of theoretical predictions

The data are compared to parton level theoretical predictions in next-to-leading order in pQCD. Hadronization effects and effects of Z-exchange at high values of Q^2 bins, which are not present in the QCD predictions, are included through bin-wise correction factors.

7.2.1 NLO calculations

In this chapter, the measured cross sections are compared to NLO calculations. The predictions σ_i^{pQCD} are calculated using the fastNLO framework [98–100] using perturbative

[1] The number of events on generator level is not an integer number, due to the estimated background subtraction and the unfolding procedure.

[2] Throughout this thesis 'p_T' denotes any of the three measured observables p_T^{jet}, $\langle p_T \rangle_2$ and $\langle p_T \rangle_3$. The expression $\langle p_T \rangle$ is used for the observables of the dijet and trijet measurements. The explicit notation, $\langle p_T \rangle_2$ and $\langle p_T \rangle_3$, is only used if the distinction between the dijet or trijet observable is necessary.

coefficients, which have been obtained with the NLOJET++ program [93, 94] (c.f. 2.7). The PDFs are accessed via the LHAPDF routines [194]. The α_s-evolution is performed using the evolution routines as provided together with the PDF sets in LHAPDF or using the fastNLO version of the GRV calculation [195], which gives identical results to the CRunDec-evolution code [35, 196]. The running of the electromagnetic coupling is taken from the code by H. Spiesberger [197, 198].

For the data-theory comparison the following choices for the renormalization μ_r and factorization scale μ_f are made (c.f. 2.1.4.5 and 2.1.5.4):

$$\mu_r^2 = (Q^2 + p_{\mathrm{T}}^2)/2 \tag{7.3}$$
$$\mu_f^2 = Q^2, \tag{7.4}$$

where p_{T} is identified with each $p_{\mathrm{T}}^{\mathrm{jet}}$ fulfilling the requirements in case of inclusive jets, and with $\langle p_{\mathrm{T}} \rangle$ in case of the dijet an trijet measurement.
The calculation of the NC DIS cross section for the prediction (c.f. 2.5.3) of the normalized jet cross section is performed using the QCDNUM program [66].

7.2.2 The theoretical cross section

Corrections to pQCD calculations are necessary for electroweak and hadronization effects. These effects are taken into account by multiplicative bin-wise correction factors c^{Z^0} and c^{had} to the pQCD calculation[3]. The theory prediction σ_{Theory} in bin i is then given by

$$\sigma_i^{\mathrm{Theory}} = c_i^{\mathrm{EW}} c_i^{\mathrm{had}} \sigma_i^{\mathrm{pQCD}}. \tag{7.5}$$

The corrections to the pQCD prediction are only dependent on theoretical models and their parameters, i.e. no experimental details are included. Therefore, theoretical correction factors are quoted separately and improvements in their accuracy are expected in the future, e.g. when NLO event generators with parton showers and adapted hadronization models become available.

7.2.3 Hadronization correction

The *hadronization correction* considers long-range effects in the cross section calculation such as the fragmentation of partons into hadrons (c.f. 2.3.2). For each bin i of the measurement, the hadronization correction factor $c_{i,k}^{\mathrm{had}}$ is given by

$$c_{i,k}^{\mathrm{had}} = \frac{\sigma_{i,k}^{\mathrm{hadron}}}{\sigma_{i,k}^{\mathrm{parton}}}, \tag{7.6}$$

where the cross sections σ^{parton} and σ^{hadron} are calculated using the event generator k ($k \in \{\mathrm{Dj}, \mathrm{Rg}\}$) after and before the application of a certain fragmentation model. In the former case, the colored partons are input to the jet algorithm. In the latter case, stable particles with a lifetime $\tau > 3 \cdot 10^{-10}\,\mathrm{s}$ are input to the jet algorithm.

[3] The theoretical predictions are colloquially denoted *NLO calculations*, and often it is not distinguished if corrections are applied or not. Here, we refer to *pQCD calculations*, if pure NLO calculations without corrections are meant.

The two MC generators employed, Django and Rapgap, are both interfaced to the Lund string fragmentation model (c.f. 2.3.2.1) in order to simulate hadronization effects. However, due to their different approach to generate parton showers (CDM vs. ME+PS), the correction factors $c^{\text{had}}_{i,\text{Dj}}$ and $c^{\text{had}}_{i,\text{Rg}}$ differ by 1 to 3 %. The hadronization correction factor is determined from the average of the factors obtained from the reweighted Django and Rapgap predictions:

$$c^{\text{had}}_i = \frac{c^{\text{had}}_{i,\text{Dj}} + c^{\text{had}}_{i,\text{Rg}}}{2}. \tag{7.7}$$

The hadronization correction is generally below unity and decreases for larger transverse jet momenta. The small effect of 2 to 8 % is an advantage of the inclusive k_T jet algorithm. Together with the anti-k_T jet algorithm it has the lowest sensitivity to hadronization effects compared to other jet algorithms [199], and is therefore preferred [78]. It was also demonstrated that the hadronization corrections are significantly reduced for virtualities of $Q^2 \gtrsim 200\,\text{GeV}^2$ and transverse jet momenta of $p^{\text{jet}}_\text{T} \gtrsim 8\,\text{GeV}$, which holds also for this analysis [78].
The hadronization corrections and their uncertainties (section 7.2.5.1) are shown in figure 7.1 for all bins of the three measurements.

7.2.4 Electroweak correction

The fixed order pQCD calculations for multi-jet production include only leading-order electromagnetic interactions at the lepton vertex. The influence of electroweak effects, like the γZ-interference terms and Z^0-exchange, become relevant when $Q^2 \to M^2_Z$, i.e. $Q^2 \gtrsim 1000\,\text{GeV}^2$. These effects are estimated using the LEPTO event generator, where cross sections are calculated including these effects ($\sigma^{(\gamma,\gamma Z,Z^0)}$) and excluding these effects, i.e. keeping only electromagnetic terms (denoted σ^γ). The correction factor c^{EW} is defined for each bin i by

$$c^{\text{ew}}_i = \frac{\sigma^{(\gamma,\gamma Z,Z^0)}_i}{\sigma^\gamma_i}. \tag{7.8}$$

The electroweak correction is dependent on the charge of the incoming lepton. Therefore, the $\sigma^{(\gamma,\gamma Z,Z^0)}_i$ calculation is weighted by the luminosities of e^+p and e^-p run-periods (see table 3.1). The electroweak correction factor is displayed in figure 7.1. For the second highest Q^2 bin it is 2 % and for the highest about 10 %. The electroweak correction has some p_T-dependence[4], but due to the mixture of e^+p and e^-p run-periods, the correction factors become flat in p_T. In case of normalized cross sections or cross section ratios, the electroweak correction cancels.
The running of the electromagnetic coupling $\alpha_{em}(Q^2)$ is taken into account in the NLO calculations and no correction has to be applied. Effects of real photon radiation from the incoming or out-going lepton are corrected for in the unfolding procedure.

7.2.5 Uncertainties of the theory prediction

7.2.5.1 Uncertainty of the hadronization corrections

The uncertainty of the hadronization correction is estimated using the Sherpa event generator [145] (c.f. section 3.5.1.4). Events with up to $2 \to 5$ parton topology are generated in the matrix element on tree level, and interfaced to the Lund string fragmentation model

[4] The acceptance of the DIS phase space (Q^2,y) has a small dependence on p^{jet}_T.

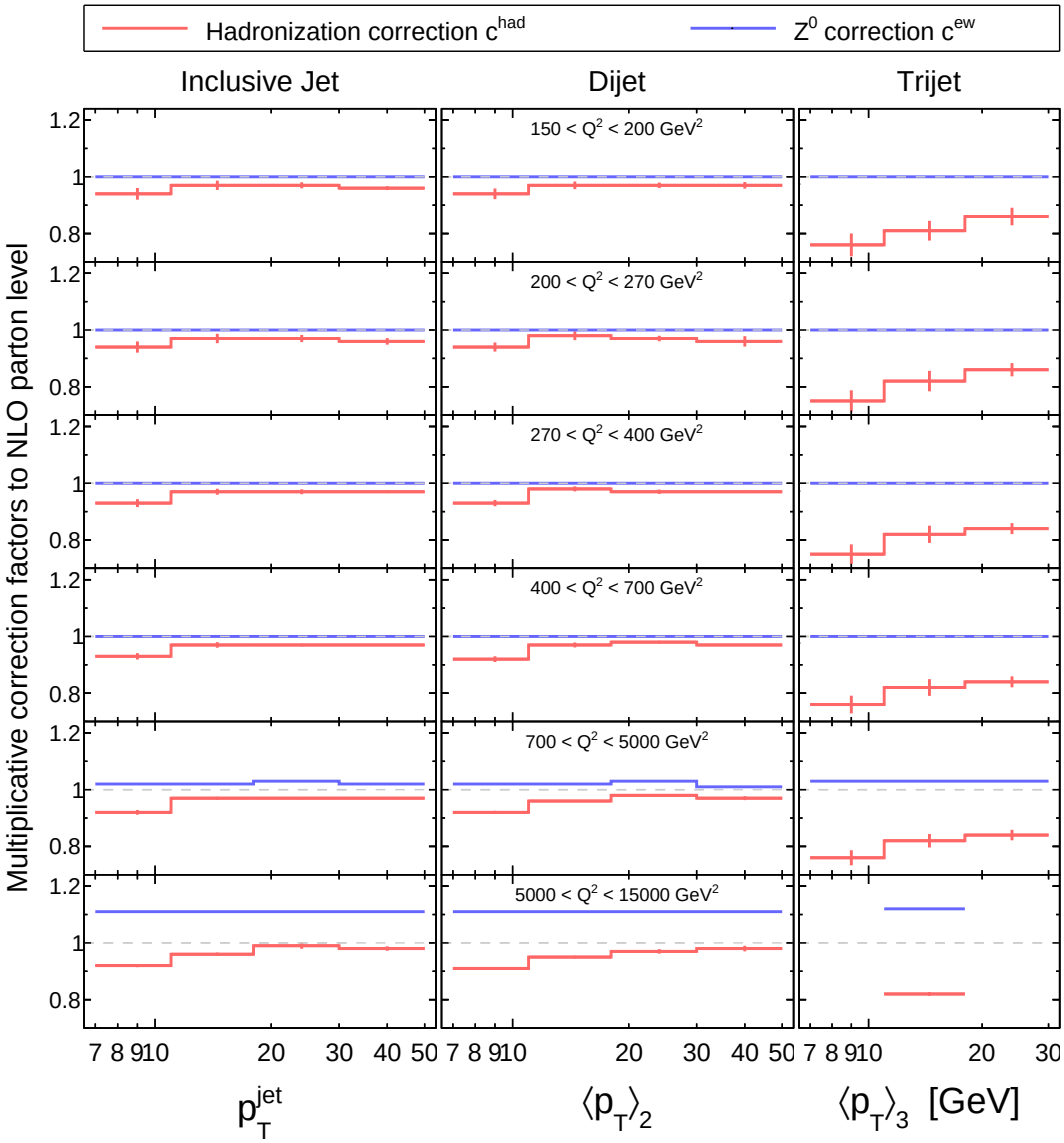


Figure 7.1: Bin-wise multiplicative correction factors applied to the pQCD calculations. The hadronization correction is indicated with its error as derived from two different hadronization models. The uncertainty on the Z^0 correction is negligible.

and to the cluster fragmentation model (c.f. section 2.3.2.1). Half the difference between the two correction factors is quoted as uncertainty on the hadronization correction Δc_{had}. The uncertainty is between 1 to 2 % for the inclusive jet and dijet measurements and between 0.5 to 5 % for the trijet measurement. The size of these uncertainties is slightly larger than the full difference derived from Rapgap and Django. This estimate for the uncertainty is expected to include the difference of various fragmentation models and the uncertainty on their parameters. However, it is expected, that the precision of hadronization corrections improve significantly as soon as event generators based on NLO calculations become available.

7.2.5.2 Uncertainty of the electroweak theory

The correction for electroweak effects c_{ew} is precisely determinable, since the masses and couplings of the electroweak theory are well known and therefore no uncertainty on c_{ew} is assigned. Additionally, the contribution from Z^0 and γZ interference terms to the cross sections become only relevant in the region $Q^2 \gtrsim 1000\,\text{GeV}^2$, where the statistical uncertainty of the data is the dominant uncertainty.
A difference of 1.5 % on the predictions of jet and NC DIS cross sections was observed when using different evolution codes for the electromagnetic coupling $\alpha_{em}(\mu)$. However, the very recent code [198] includes the latest data from hadronic contributions to $\alpha_{em}(\mu)$ [200] and is assumed to be reliable and no uncertainty is assigned.

7.2.5.3 Uncertainty of the pQCD calculation

The uncertainty on the perturbative calculation is determined by varying the renormalization and factorization scales simultaneously by a conventional factor of 0.5 and 2[5]. This *theory uncertainty* is assumed to reflect the effect from missing higher orders in the truncated perturbative series. The scale dependence is smaller for bins, where the renormalization scale is large, because of the proportionality of $\sigma \propto \alpha_s(\mu_r)$. The error of the theory uncertainty may be asymmetric, the scale variations may even predict smaller cross sections for both, the up and the down variation of the scale. This is for instance the case for the highest $\langle p_{\text{T}} \rangle$ bin in the highest Q^2 bin. The theory uncertainty is the dominant uncertainty of the measurement and is between 5 and 15 %.

7.2.5.4 Uncertainty of the PDFs

The uncertainty of the NLO calculation due to the limited knowledge of the PDFs is determined from the MSTW2008 eigenvectors following the formula for asymmetric PDF uncertainties [60] and represents the uncertainty at 68 % CL.. The PDF uncertainty is almost symmetric and around 1 % for all bins. The predictions using other PDFs do not deviate typically by more than $2\,\sigma$ of the PDF uncertainty.

[5] See also figures 2.3, 2.4 and 2.6 for the scale dependence on the cross sections

7.3 Jet cross section results

In this section, the measured cross section results are presented and compared to NLO calculations.

7.3.1 Double-differential cross sections for jet-production in DIS

An overview of the measured cross sections for the three jet-measurements is shown in figure 7.2. In order to imitate the double differential manner of the measurement the cross sections σ_i are divided by the p_{T} and Q^2 bin-widths on the plots (indicated by "[GeV^{-3}]") and scaled for better visibility by 10^i, with i increasing with decreasing Q^2 (the values of i are given in the legend of the figure) . The inner error bars show the statistical uncertainties and the outer error bars the statistical and all systematic uncertainties added in quadrature. However, these uncertainties are typically smaller than the marker size and are not visible in this figure. The cross sections show a steeply falling behavior as function of p_{T} and Q^2, and they span over five orders of magnitude.
The cross sections are shown together with NLO predictions employing the MSTW2008 parton density functions and a value of $\alpha_s(M_Z) = 0.118$. The NLO calculations are corrected for non-perturbative hadronization effects and electroweak contributions. The width of the blue band of the NLO predictions indicates the uncertainty obtained from varying the factorization and renormalization scale by a factor of 2 and 0.5. The data show overall a good agreement with the NLO predictions.

7.3.2 Data-theory comparison

The measured cross sections are compared to NLO calculations using the MSTW2008 PDF set in figure 7.3. Predictions from other PDF sets, which have all been determined consistently with $\alpha_s(M_Z) = 0.118$ and at NLO precision, are compared to the predictions from the MSTW2008 PDF set in figure 7.4.
Within the statistical and systematic uncertainties, the theory shows an overall good description of the measurement. On average, there are more data points below the theory predictions than above. This points towards a smaller value of $\alpha_s(M_Z)$, which decreases the predicted cross sections. Some bins of the inclusive jet and dijet measurement show an increased deviation from the NLO predictions of up to $2\,\sigma$ and these bins are always below the calculations. This, however, could also be caused by statistical fluctuations. The inclusive jet and dijet data in the lowest p_{T}-bin are always tend to be below the theory predictions, as well as most of the highest p_{T}-bins.

The measurement is dominated by systematic uncertainties, which are mostly larger than the statistical uncertainties. Due to the unfolding, there are large negative correlations from migrations between neighboring bins of an individual measurement. Therefore, the data are more precise as perceived visually from these plots and the correct statistical precision is only obtained together with the bin-to-bin correlations (see figure 7.5). The systematic uncertainties are dominated by the model uncertainty, the JES and RCES uncertainty and the uncertainty on the luminosity measurement, as described in section 7.3.3.2.

The largest uncertainty in this comparison is the theory uncertainty. It is between 5 and 15 % and is larger at smaller scales, where in contrast the statistical precision is high. Up to date, it is unknown, if higher orders will increase or decrease the predicted cross section.

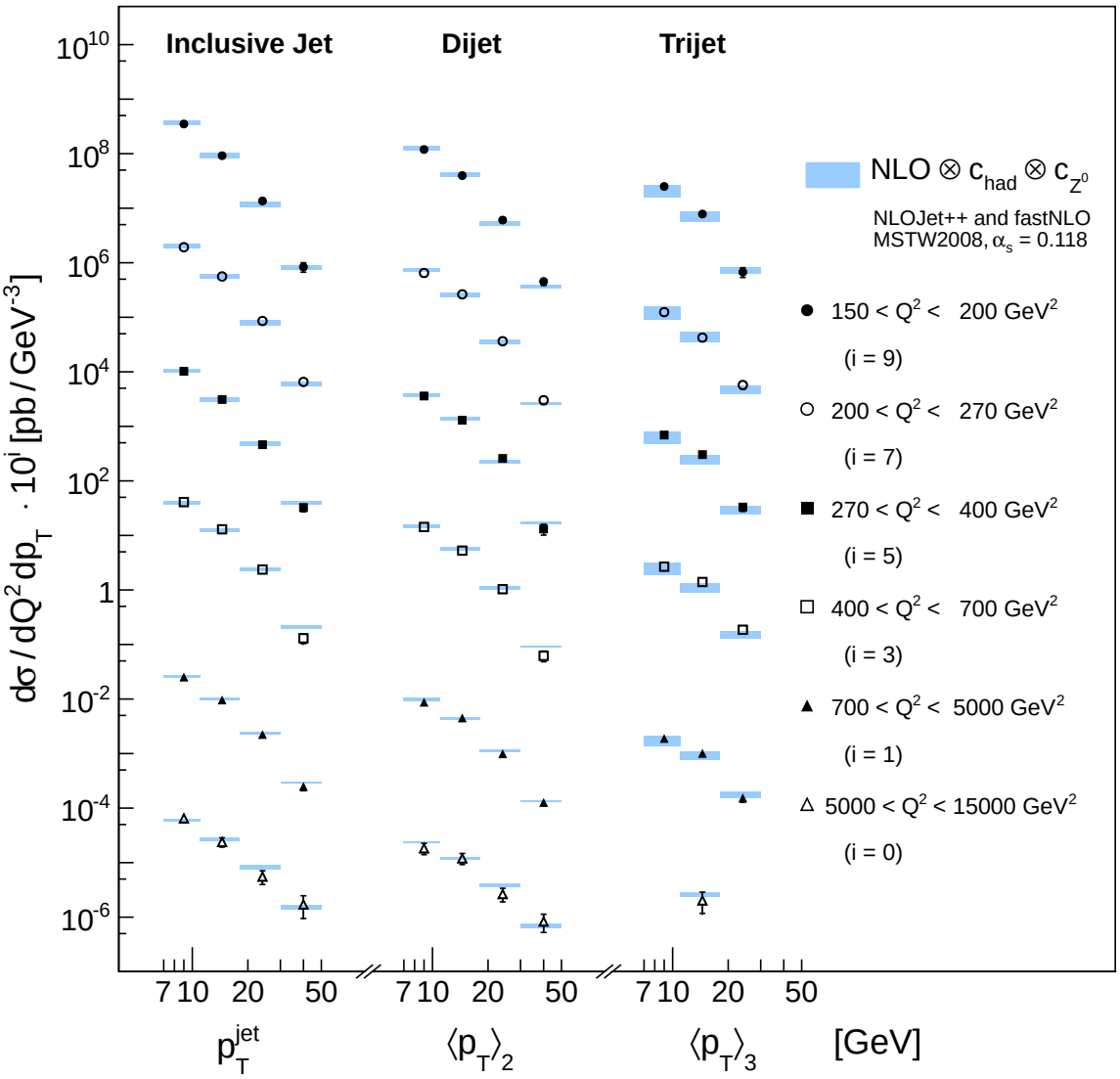


Figure 7.2: Measured double differential cross sections for jet production in DIS as function of Q^2 and p_{T}. The inner and outer error bars indicate the statistical uncertainties and the statistical and systematic uncertainties added in quadrature. They are typically smaller than the marker size. The pQCD predictions in next-to-leading order are derived using NLOJET++ and employ the MSTW2008 parton density functions and a value of $\alpha_s(M_Z) = 0.118$. The theory predictions include non-perturbative corrections for hadronization and for electroweak effects. The width of the NLO band indicate the uncertainty from missing higher orders as determined from scale variations. The cross sections for individual Q^2 bins are multiplied by a factor of 10^i for better readability.

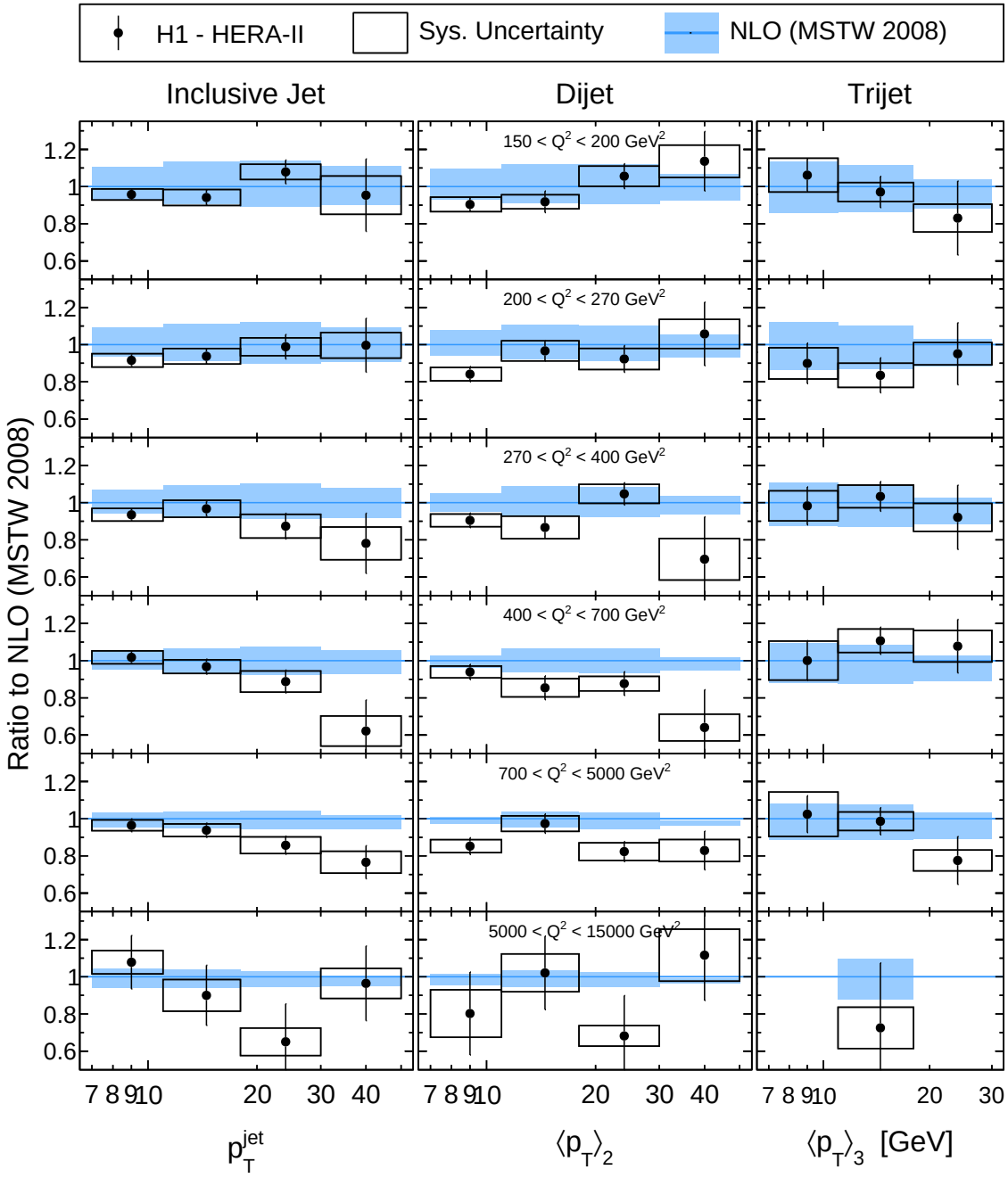


Figure 7.3: Ratio of jet cross sections to NLO predictions using the MSTW2008 PDF set. The error bars on the data indicate the statistical uncertainty. The statistical uncertainties seem to be large due to the (negative) correlations from the unfolding, which are most distinct between neighboring bins in p_{T} and Q^2. The systematic uncertainty box shows all systematic uncertainties added in quadrature, where the model uncertainty, the JES and RCES uncertainty and the luminosity uncertainty dominate. The blue band shows the theory uncertainty determined from simultaneous variations of the renormalization and factorization scale by a factor of 0.5 and 2. The NLO calculations are corrected for non-perturbative hadronization effects and electroweak contributions. Predictions from other PDF sets are displayed in figure 7.4.

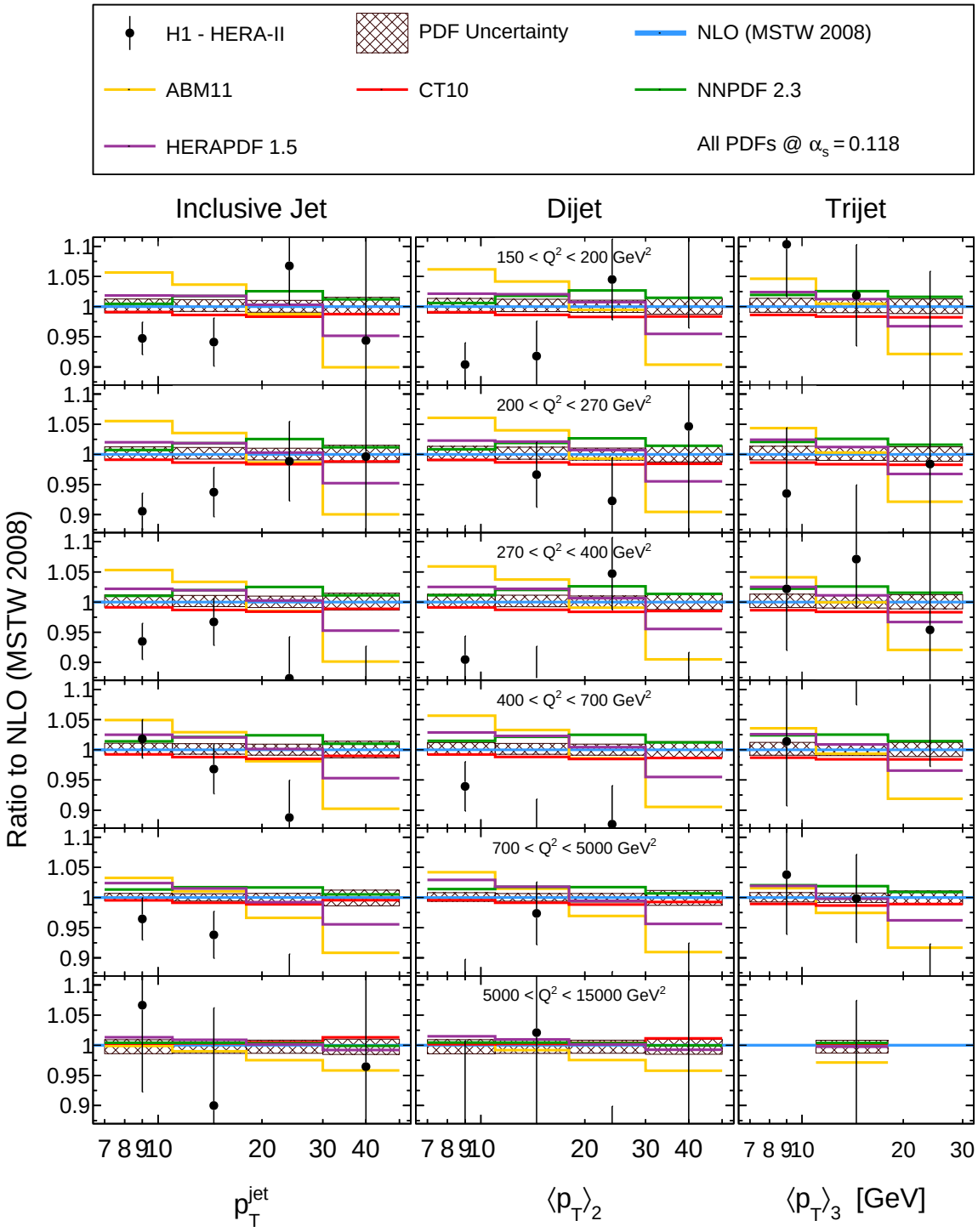


Figure 7.4: Ratio of NLO predictions from various PDF sets to predictions using the MSTW2008 PDF set. For comparison, also the data points are displayed together with its statistical uncertainty, which however are often outside of the displayed range in this enlarged presentation. All PDFs used are determined at NLO and with a value of $\alpha_s(M_Z) = 0.118$. The hatched area shows the PDF uncertainty on the NLO calculations from the MSTW2008 eigenvector set at a confidence level of 68 %, which is typically around 1 %.

Although the NNLO perturbative coefficients are exptected to increase the contribution from the hard matrix element to the cross section [44], the unknown higher orders may also have lower predictions, since the PDFs determined in NNLO precision, and especially $\alpha_s(M_Z)$ determined from NNLO calculations, compensate the effect from the hard process (see e.g. [61, 62]).

The CT10 PDF prediction is between 1 to 2 % below, and the NNPDF 2.3 prediction is around 2 % above the MSTW2008 prediction. The HERAPDF1.5 prediction is 2 % above the MSTW2008 calculation for low p_{T}-bins but in the highest p_{T}-bins the HERAPDF1.5 prediction is around 5 % below the MSTW2008 prediction (3 % for the highest trijet $\langle p_{\mathrm{T}} \rangle$-bins). This is a result from the softer valence quark density at high-x values of the HERAPDF1.5 PDF set compared to other PDF sets. All PDFs are evolved with the α_s-evolution code, which is provided together with the PDF grids and may therefore differ. The effect of the α_s evolution code has been tested by using the same evolution for all calculations and no visible effect on the cross section could be observed.

7.3.3 Experimental uncertainties

The uncertainties on the measurement are discussed separately for the correlated statistical uncertainties and the systematic uncertainties.

7.3.3.1 Statistical uncertainties and correlations

The regularized unfolding procedure applied features a full linear error propagation, which takes all kinematic migrations into account (see equation 5.15). The correlation matrix of the measured bins is illustrated in figure 7.5. The graphical illustration of the correlation matrix also shows the bins of the NC DIS measurement after the unfolding, which is used for the calculation of the normalized jet cross sections.
The correlations are essential to interpret the unfolded cross section results. On graphical illustrations and tables only the square root of the diagonal elements of the covariance matrix are shown. These can be large, although the statistical precision is high. This is the case when negative correlations which result from the unfolding are present. Neighboring bins in p_{T} are largely anti-correlated especially in the low p_{T}-region of the dijet and trijet measurements, where $\varrho_{\alpha\beta} \approx -0.4$. In the low p_{T}-region of the inclusive jet measurement, the correlations are around $\varrho_{\alpha\beta} \approx -0.2$ and are not as large as for the dijet measurement. This is because the inclusive jets have positive statistical correlations of around $\varrho_{\alpha\beta} \approx +0.2$ between the p_{T}-bins α and β and around $\varrho_{\gamma\delta} \approx +0.1$ between the two highest p_{T}-bins prior to the unfolding.
The correlations between p_{T}-bins are almost identical for different Q^2-bins and vice-versa. This is expected, since the detector response on jet-quantities is largely independent from migrations of quantitites related to the scattered electron.
The correlations between the single measurements are dominated by positive statistical correlations, but after the unfolding also effects from migrations play a role. Whenever bins with large positive migrations are used together in a fit, the statistical precision of the fit results do not increase significantly. For instance, if $\varrho_{ij} = 1.0$ no benefit in precision is gained[6], if the two bins i and j are both used together, compared to only one bin being used.

[6] In the case of $\varrho_{ij} = 1$, the two bins i and j are fully correlated, i.e. they would refer to the same quantity. As a consequence, the covariance matrix $\boldsymbol{V}$ has not full rank and therefore cannot be inverted.

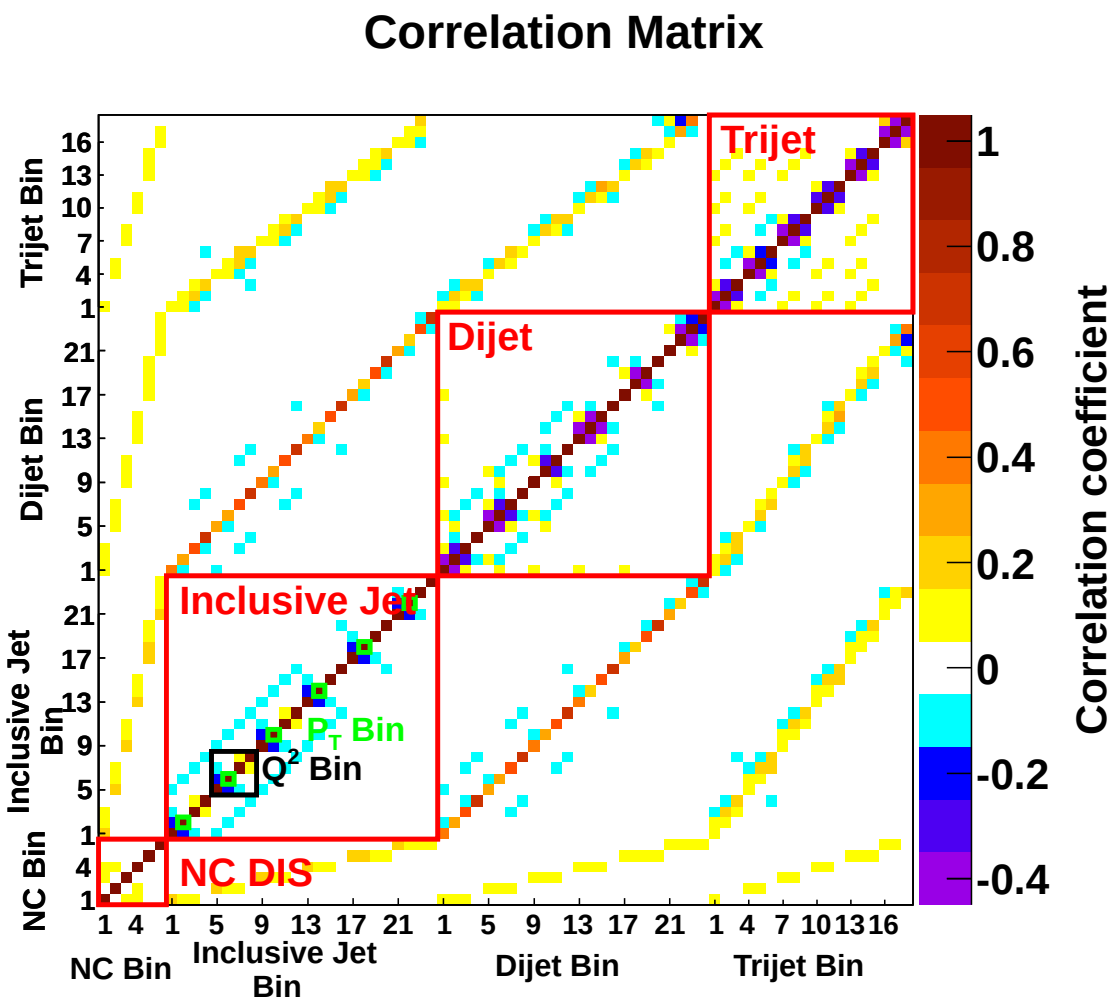


Figure 7.5: Correlation matrix of the three jet cross section measurements and the NC DIS measurement. Exemplary for the inclusive jet measurement, the green boxes indicate the identical p_{T}-bins in several Q^2-bins, and the black box indicates one Q^2 bin with different p_{T}-bins (see the bin numbering in equation 7.2).
The symmetric matrix shows negative correlations between bins, where migrations are large. Positive correlations arise from statistical correlations between the measurements. The initially measured positive correlations of the inclusive jet measurement is now dominated by negative correlations from the unfolding. Correlations between neighboring p_{T}-bins are larger than between neighboring Q^2 bins. The correlations between the jet measurements and the NC DIS measurement are being used in the error propagation for the normalized jet cross sections.

7.3.3.2 Systematic uncertainties

Various sources of systematic uncertainties influence the experimental precision of the measured cross sections. These can be attributed to uncertainties in the reconstruction of the electron or the HFS, as well as to normalization uncertainties, like the uncertainty of the trigger efficiency and of the luminosity measurement.
The size of each systematic uncertainty is determined by simulating the size and sign of the effect in the detector simulation and determining two alternative migration histograms (c.f. section 5.4.2). The systematic uncertainty is then determined on hadron level using equation 5.17.
A detailed comparison of the dominant uncertainties is shown in figures 7.6 and 7.7, where the statistical uncertainty (only the uncorrelated uncertainty of each data point can be displayed) and the model uncertainty is shown.
The uncertainty due to the HFS measurement, consisting of the jet energy scale uncertainty ($\delta_{\mathrm{JES}}\sigma$) and the remaining cluster energy scale uncertainty ($\delta_{\mathrm{RCES}}\sigma$), are shown separately. The systematic uncertainties due to the lepton reconstruction are displayed in figure 7.8.

Uncertainties of the hadronic final state measurement The uncertainty of the hadronic final state is subdivided in two components: these are uncertainties on clusters which are part of detector-level jets and receive calibration constants from the jet calibration procedure (see section 4.1.4) in addition to the neural network calibration [130], and on the other and these are clusters outside of detector-level jets, which receive calibrations constants only from the neural network calibration. Since the calibration constants for jets are only determined for detector-level jets down to $p_{\mathrm{T,lab}}^{\mathrm{jet}} > 7\,\mathrm{GeV}$ a considerable amount of clusters contribute particularly to jets with low $p_{\mathrm{T}}^{\mathrm{jet}}$. The uncertainties on the hadronic final state measurement are displayed in figure 7.7.
Calibration constants for HFS objects, which are part of a jet with $p_{\mathrm{T,lab}}^{\mathrm{jet}} > 7\,\mathrm{GeV}$, are derived from well contained DIS events from the transverse momentum balance of the HFS and the scattered electron in the laboratory frame of reference. The uncertainty due to the *jet energy scale (JES)* on the cross sections is determined by varying the energy of all HFS objects by $\pm 1\%$ [130]. This yields an uncertainty on the cross sections between 2 to 6 %, with the larger values for higher jet p_{T}. The JES uncertainty of the cross section measurements decreases slightly with increasing Q^2 values. The JES uncertainty is the dominant experimental uncertainty, as is usually the case for jet measurements.
The energy of HFS objects which are not part of a jet[7], and hence do not receive jet calibration constants, allowing for a larger uncertainty than the 1 %, and are varied by 2 % to obtain the so-called *remaining cluster energy scale uncertainty* (RCES). This systematic uncertainty is taken to be independent of the JES uncertainty. Clusters which are not part of a laboratory frame jet are mostly singular clusters with small energy deposits, or these are clusters within jets with a lower momentum than $p_{\mathrm{T,lab}}^{\mathrm{jet}} < 7\,\mathrm{GeV}$ for which no jet calibrations constants have been determined. The RCES uncertainty on the jet cross section is therefore large in the low-p_{T} region. It is around 1 to 2 % of the inclusive jet and the dijet cross section, and is up to 7 % for the trijet cross section, which is more sensitive to low-p_{T} jets compared to the dijet measurement.
Several noise suppression algorithms for the LAr electronics are applied. An uncertainty

[7] For the determination of the jet calibration constants, the jets are defined in the laboratory frame and are therefore differently from the jets found in the Breit frame.

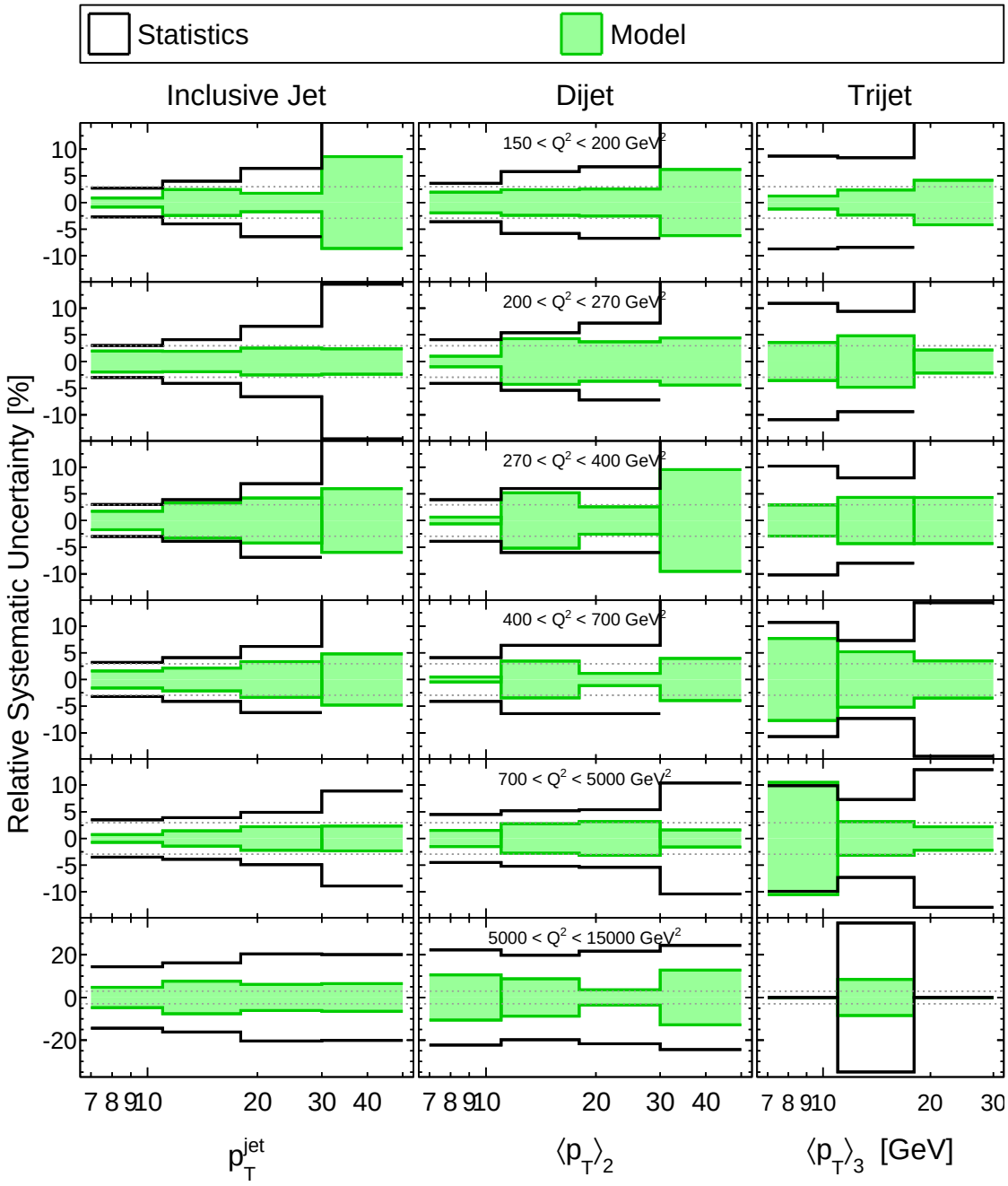


Figure 7.6: Comparison of the dominant uncertainties of the measured cross sections. These are the statistical uncertainty, shown as black lines, and the model uncertainty shown as green lines. The gray dashed line shows the overall normalization uncertainty of 2.9 %, where the uncertainty of the luminosity, the track-cluster-link uncertainty and the trigger uncertainty are added in quadrature. The statistical uncertainty is reduced, as soon as multiple bins with negative correlations are used together, for instance in a fit. In the highest p_{T} bins, the statistical uncertainty is often larger than the displayed range. The numerical values are given in the appendix A.

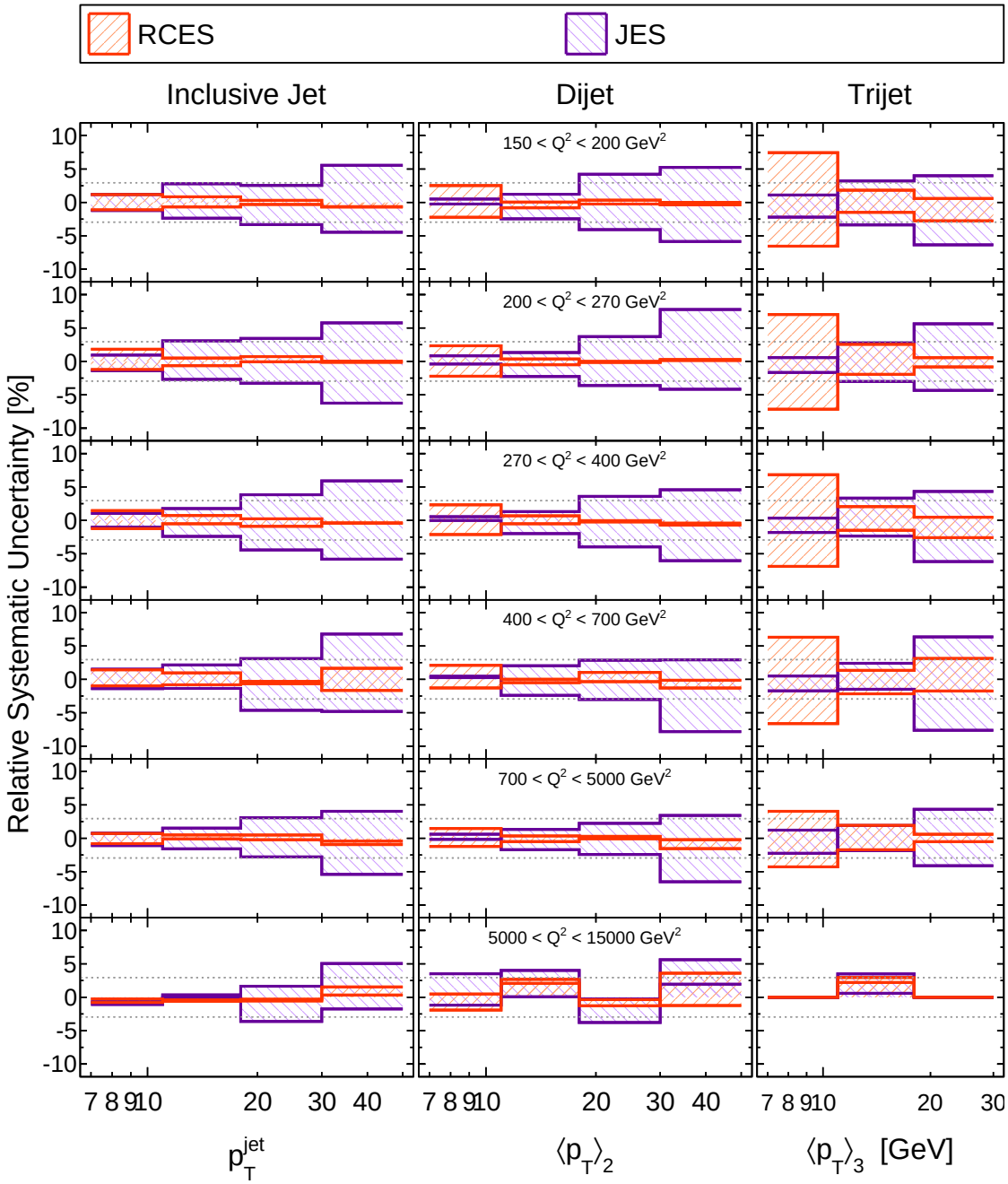


Figure 7.7: Comparison of the uncertainties of HFS measurement. The violet line show the uncertainty due to the jet energy scale uncertainty (JES). The uncertainty due to energy scale of the remaining clusters (RCES), i.e. objects outside of calibrated jets, is depicted as red line. The uncertainty of the LAr noise suppression algorithm accounts for another 0.5 to 0.9 % uncertainty of the HFS measurement and is not displayed here. The gray dashed line shows the overall normalization uncertainty of 2.9 %. Mind the different scale in the highest Q^2 bin.

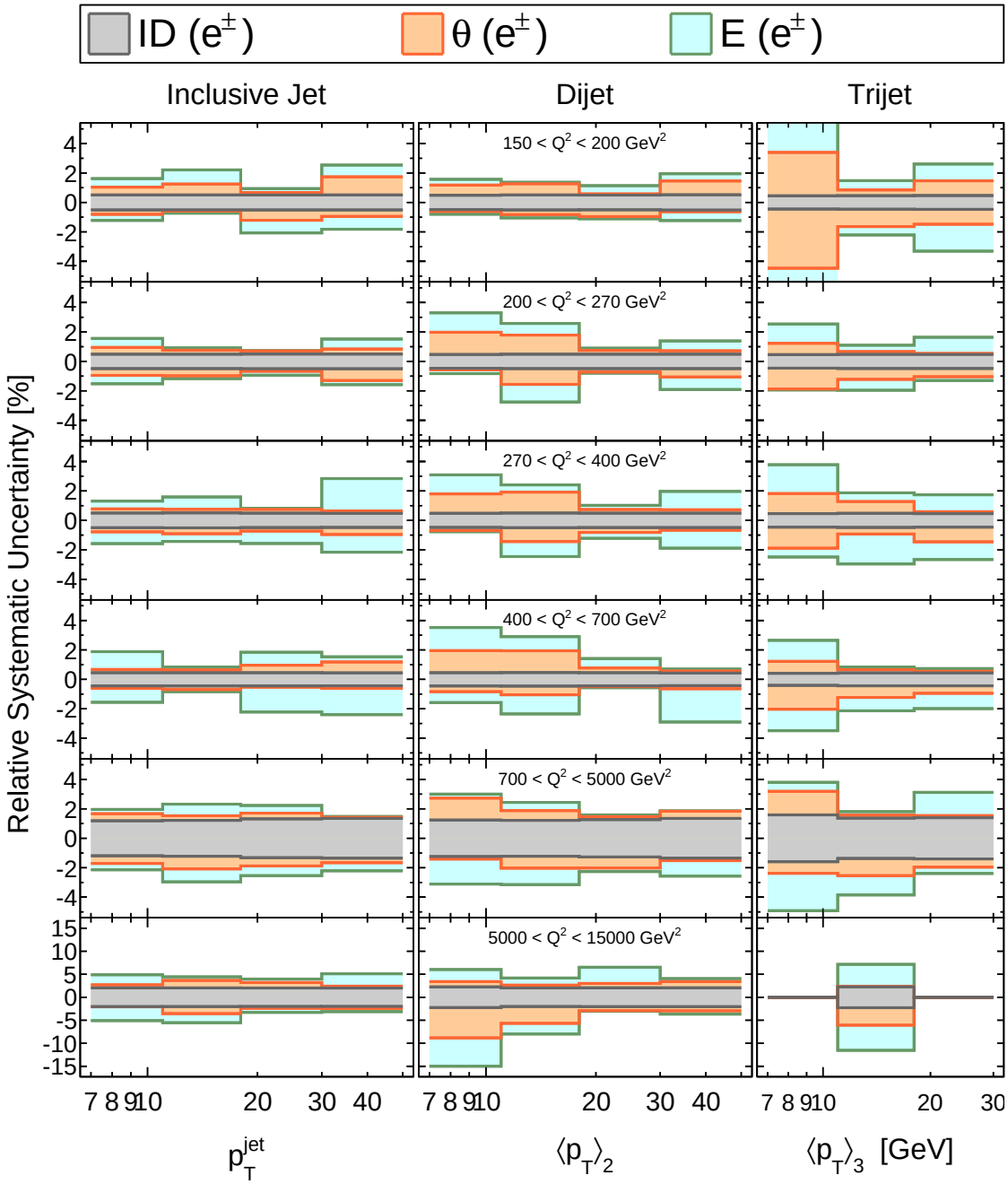


Figure 7.8: Comparison of the relative uncertainties of the measured cross sections due to the measurements of the electron polar angle ($\theta(e^{\pm})$), the electron energy ($\mathrm{E}(e^{\pm})$), and the electron identification ($\mathrm{ID}(e^{\pm})$). These three uncertainties are added linearly on this plot for better visibility. The uncertainties are usually smaller than the normalization uncertainty of 2.9 %. The partly large fluctuations of the size of the uncertainties between adjacent bins are caused by the unfolding procedure. Mind the enlarged scale for the highest Q^2 bin.

of the noise contribution to the cross section is determined by adding randomly 20 % of all rejected noise clusters to the list of signal clusters [171]. An increase of the cross section of 0.5 %, 0.6 % and 0.9 % are observed for the inclusive jet, the dijet and trijet measurement, respectively.

Uncertainties of the lepton reconstruction Three sources of uncertainties are considered and are displayed in figure 7.8. In the measured phase space, the electron is always measured using the LAr calorimeter. The position of the LAr w.r.t. the CTD is aligned with a precision of 1 mrad, which is interpreted as the uncertainty of the electron polar angle measurement θ_e. The resulting size of the uncertainty on the cross sections is around $\delta_{\theta_e}\sigma \simeq 0.5\,\%$, except for the highest Q^2 bin, where it is about 1.5 %.
The electron finding efficiency is 0.5 % in the central direction and 2 % in the forward direction where $z_{\text{impact}} > 0$. This leads to a Q^2 dependent uncertainty for the electron identification (ID(e)) of around $\delta_{\text{ID(e)}}\sigma \simeq 0.5\,\%$ for the lower Q^2 bins and up to 2 % for the highest Q^2 bin.
The energy of the electron is measured with a precision of 0.5 % in the central and backward region ($z_{\text{impact}} < 100\,\text{cm}$) and with 1 % precision in the forward region of the detector. The simulation of this effect determines the uncertainty $\delta_{E_e}\sigma$ on the cross sections between 0.5 and 2 %, where the larger uncertainties are mostly observed in high p_{T} or high Q^2 bins.

Normalization uncertainties The uncertainty of the luminosity measurement is determined to be 2.5 % for the HERA-II run period [114].
The uncertainty of the trigger efficiency takes the efficiency of the LAr trigger element, the veto and the timing condition into account and is determined to be 1.2 % [130].
The efficiency of the requirement of a link between the primary vertex, the electron track and the electron cluster is well described by the MC simulations within 1 % [130], which is assigned to the *track-cluster-link* uncertainty.
The trigger and the track-cluster-link uncertainty are determined to be independent of the kinematics and are therefore also regarded as a normalization uncertainty. The squared sum of the three normalization uncertainties is 2.95 %, where the dominant uncertainty is the luminosity uncertainty. All normalization uncertainties cancel for normalized cross sections. Uncertainties of the electron reconstruction cancel largely for normalized cross sections.

Model uncertainty The model uncertainty is determined from the difference of the migration histograms of Django and Rapgap compared to the averaged migration histogram (c.f. section 5.4.3). The model uncertainty $\delta_{\text{model}}\sigma$ is between 2 and 7 % and is thus the dominant systematic uncertainty (compare figure 7.6). The size of the model uncertainty and its effect on the unfolding is studied in more detail in section 9.1.

7.4 Normalized jet cross sections

Normalized jet cross sections are determined by the ratio of the absolute jet cross section w.r.t. the NC DIS cross section as obtained from the unfolding procedure. Jet cross section ratios are defined as one jet measurement w.r.t. to another, like the trijet over dijet ratio $R_{3/2}$.

7.4.1 Definition of normalized jet cross sections

The normalized jet cross section is defined through

$$\sigma^{\text{jet/NC}} = \frac{\sigma^{\text{jet}}}{\sigma^{\text{NC}}} \ . \tag{7.9}$$

Using equation 7.1, the double-differential normalized jet cross section in the Q^2-bin q and the p_{T}-bin p is determined by

$$\sigma^{\text{jet/NC}}_{(q,p)} = \frac{N^{\text{unfolded}}_{(q,p),\text{jet}}}{N^{\text{unfolded}}_{q,\text{NC}}} \ , \tag{7.10}$$

where $N^{\text{unfolded}}_{(q,p),\text{jet}}$ is the number of events in bin (p,q) on hadron level of the unfolded jet measurement and $N^{\text{unfolded}}_{q,\text{NC}}$ is the number of events of the unfolded NC DIS measurement in bin q. The luminosity cancels and thus also the uncertainty on the luminosity and the remaining normalization uncertainties cancel for normalized cross sections.
For the theory predictions, the effect from electroweak interactions cancels in the ratio. The jet cross sections have to be corrected for hadronization effects, while the inclusive NC DIS cross sections are insensitive to such effects. The normalized jet cross sections are thus calculated through

$$\sigma^{\text{NLO}}_{\text{jet/NC}} = \frac{\sigma^{\text{NLO}}_{\text{jet}} c_{\text{had}}}{\sigma^{\text{NLO}}_{\text{NC}}} \ . \tag{7.11}$$

7.4.2 Error propagation for normalized cross sections

The statistical and experimental uncertainties on the normalized cross sections are determined by linear error propagation.

7.4.2.1 Statistical error propagation

The statistical uncertainty is determined through linear error propagation by calculating the covariance matrix $\boldsymbol{V}^{\text{jet/NC}}$ through a first order Taylor expansion which yields [12]

$$\boldsymbol{V}^{\text{jet/NC}} \simeq \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^{\text{T}} \ , \tag{7.12}$$

where the elements A_{ij} are given by

$$A_{ij} = \left.\frac{\partial \sigma^{\text{jet/NC}}_i}{\partial N_j}\right|_{N^{\text{unfolded}}} \ , \tag{7.13}$$

where $\boldsymbol{V}$ denotes the covariance matrix after the unfolding of the NC DIS measurement and the jet measurements (figure 7.5 shows the correlation matrix $\boldsymbol{P}$ which is calculated from $\boldsymbol{V}$), j are the indices w.r.t. the bin numbering in $\boldsymbol{V}$ and run over all bins of the NC DIS and jet measurements, and i denotes a bin (q,p) of the normalized measurement (c.f.

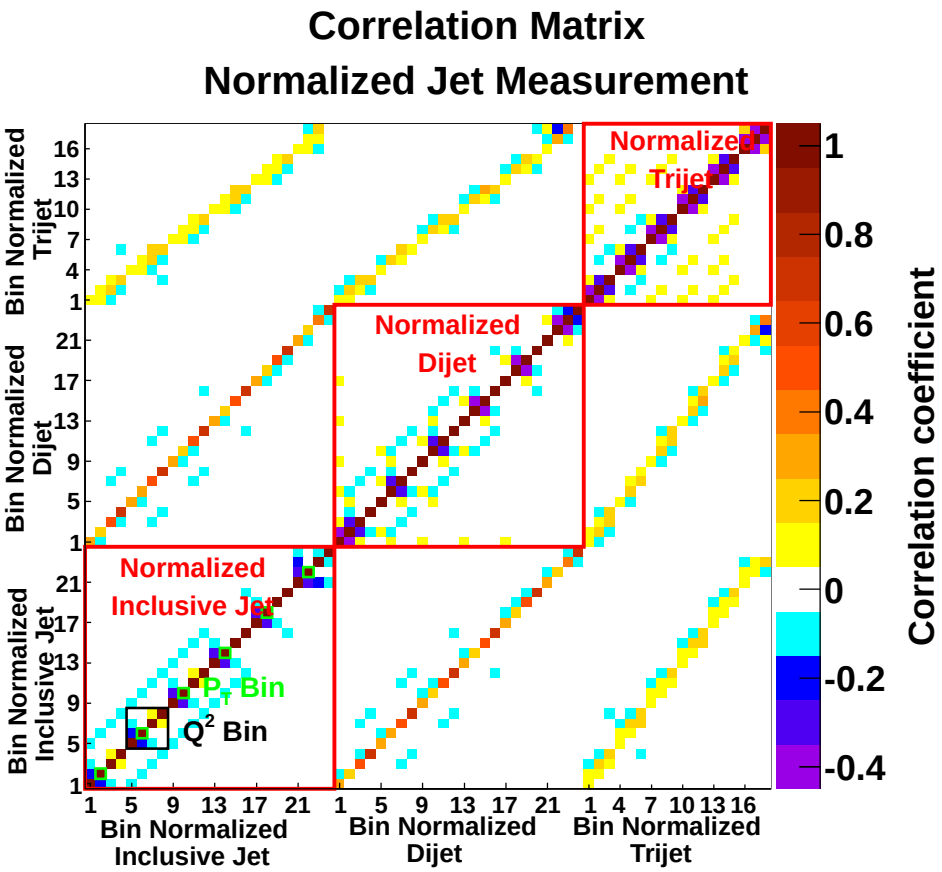


Figure 7.9: Correlation matrix of the three normalized jet cross section measurements. The correlation matrix is calculated from the covariance matrix of the normalized measurements, which is determined by linear error propagation from the covariance matrix of the NC DIS and the respective jet measurement. The correlations of the normalized jet measurements are almost identical to the correlations of the absolute jet measurements in figure 7.5.

equation 7.10).
This method is identical to a direct measurement of the normalized jet cross sections, since all migrations are respected and a full error propagation is performed. The correlation matrix $\boldsymbol{P}$ (see equation 5.16 for its defintion) for the three normalized jet measurements in shown in figure 7.9. The correlations of the normalized jet measurement are almost identical to the correlations of the absolute jet measurements.

7.4.2.2 Example of error propagation for normalized cross sections

In order to illustrate the effect of equation 7.12 for cross section ratios, we want to study the uncertainty of one data point of a cross section ratio $R_{A/B} = A/B$, which is defined as the division of a cross section A by another cross section B. The absolute uncertainties on the numbers are labeled ΔR, ΔA and ΔB, respectively, and are represented by the square root of the variances. The covariance matrix of the bin A and B may be written as

$$V_{AB} = \begin{pmatrix} \Delta A^2 & \varrho_{AB}\Delta A\Delta B \\ \varrho_{AB}\Delta A\Delta B & \Delta B^2 \end{pmatrix}, \tag{7.14}$$

where ϱ_{AB} denotes the correlation coefficient between the two data points.
The quadratic relative uncertainty on $R_{A/B}$ is then given through equation 7.12 by

$$\left(\frac{\Delta R}{R}\right)^2 = \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2 - 2\frac{\varrho_{AB}\Delta A\Delta B}{AB}. \tag{7.15}$$

The first two summands are the terms of an uncorrelated error propagation. In case of a cross section ratio, the third summand has negative sign, and thus decreases the relative statistical uncertainty of the cross section ratio, as soon as the positive correlations[8] are taken into account. We can conclude, that in case of normalized jet cross sections and cross section ratios, the statistical uncertainty always increases compared to the absolute jet cross sections[9].

A cross section ratio is typically defined such that the numerator events A are an entire subsample of the denominator events in bin B. For instance, this is the case for the normalized jet cross sections, where the jet event sample is a subsample of the NC DIS event selection, and the binning observable Q^2 is identical for both measurements[10]. In a counting experiment, where N_k denotes the number of events in k ($k = \{A, B\}$), the covariance matrix then reads

$$V_{AB} = \begin{pmatrix} N_A & N_B \\ N_B & N_B \end{pmatrix}, \tag{7.16}$$

since $\varrho_{AB}\Delta A\Delta B = N_B$ [11]. The squared relative uncertainty of the cross section ratio $R_{A/B}$ has then the simple form

$$\left(\frac{\Delta R}{R}\right)^2 = \frac{1}{N_A} - \frac{1}{N_B}, \tag{7.17}$$

which also holds for cross sections, since the luminosity cancels out in the calculation of $R_{A/B}$. Equation 7.17 is of course no longer exactly valid, as soon as background subtraction and/or an unfolding procedure have been performed.

7.4.2.3 Error propagation of systematic uncertainties

The systematic uncertainties are determined from the absolute cross sections by error propagation according to equation 7.12. The covariance matrix of the up (or down) shift of each source of systematic uncertainties is assumed to be fully correlated and is calculated following equation 5.18. The resulting shift from a systematic up (or down) variation is then given by the square root of the variances. The sign of the first bin of the shift is chosen to be identical to the sign of the shift of the first bin in the numerator. The signs of all other bins are then given by the sign of the covariances w.r.t. that bin.

[8] In this case of statistical error propagation, the correlation coefficient ϱ_{AB} is always positive. If R would be defined as $R = AB$, the third summand would have positive sign.

[9] If $2\frac{\Delta A}{A}\varrho_{AB} > \frac{\Delta B}{B}$, i.e. when $N_A > 4N_B$ in a trivial counting experiment, the relative uncertainty would decrease. Since the normalization is typically performed w.r.t. to a higher statistics sample, this is never the case.

[10] In the case of the jet cross section $R_{3/2}$ (see section 7.6) as it is defined in this work and previous analyses, this is only approximately given due to the different definition of the observables for the numerator and for the denominator events.

[11] This is only the case for uncorrelated measurements, where only one observable is counted in each event. That means, that this equation is not fulfilled exactly for the inclusive jet measurement.

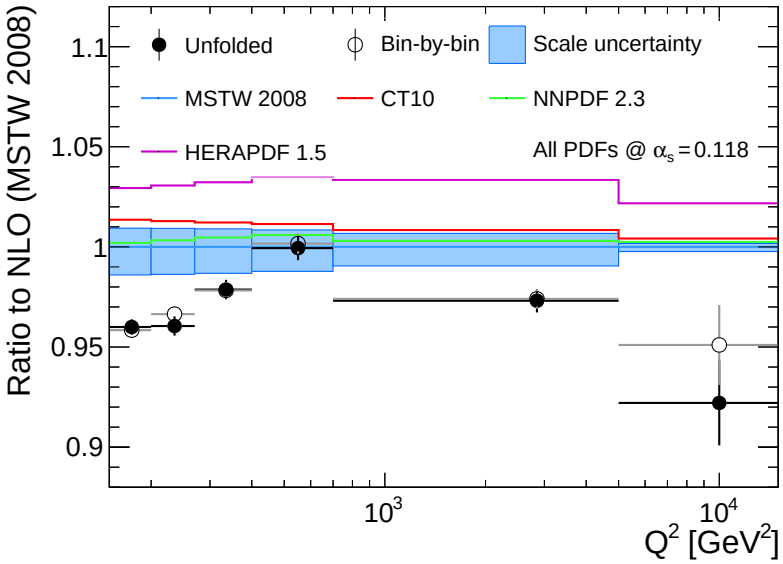


Figure 7.10: The ratio of the inclusive neutral current DIS cross section in the jet phase space to NLO predictions. The theory is calculated using QCDNUM at NLO precision and the MSTW2008 PDF set with $\alpha_s(M_Z) = 0.118$, including Z^0 and γZ interference terms [197, 198]. The data are corrected for higher order QED effects. The NC DIS cross section is used for the determination of the normalized jet cross sections. Further shown are the cross sections, which are derived using the bin-by-bin correction method divided by the theory predictions. They are consistent with the unfolded result within the statistical uncertainties.

For the determination of the model uncertainty, all bins are assumed to be uncorrelated in the error propagation. Therefore, the model uncertainties of the numerator and denominator bin are effectively summed in a quadratic summation for the cross section ratio. Hence, the relative model uncertainty increases in cross section ratios compared to absolute jet cross sections. As a consequence of the uncorrelated treatment the resulting shift from the model uncertainty is always in the same direction for all bins, since there are no correlations present and no sign can be determined. This is a quite conservative approach and accounts to the observation of large, but supposely unphysical, fluctuations of that uncertainty after the unfolding.

7.4.3 The inclusive neutral current DIS cross section

The inclusive NC DIS cross section $\sigma_{\text{NC}}(Q^2)$ are predicted by calculating the unpolarized NC process $e^{\pm}p \to e^{\pm}X$ using the QCDNUM program [66] at NLO precision[12]. The renormalization and factorization scales are set to $\mu_f^2 = Q^2$ and the same PDF set as for the jet calculations is used. The value of the strong coupling constant is set to $\alpha_s(M_Z) = 0.118$. For the comparison with data, the calculations of the inclusive NC DIS cross sections σ_{NC} include electroweak effects[13]. The unfolded NC DIS cross sections are displayed in figure 7.10 and compared to NLO predictions. The unfolded cross sections are also compared to the bin-by-bin corrected results, which have been verified by an independent NC DIS analysis [130, 157, 201]. The unfolded cross sections are consistent with the bin-by-bin

[12]The inclusive NC DIS cross section calculation in next-to-leading order is of order $\mathcal{O}(\alpha_{em}^2\alpha_s)$, while the jet cross section calculation in NLO is of $\mathcal{O}(\alpha_{em}^2\alpha_s^2)$. This is not in contradiction, since both calculations are performed in 1-loop precision.

[13]For normalized jet cross sections the electroweak effects cancel and are therefore not included for these predictions.

corrected cross sections within the statistical precision. The influence of the detector-level-only entries of the jet measurements on the NC DIS cross section is within the statistical uncertainty (see section 9.6). The data are described by the NLO predictions within an experimental normalization uncertainty of around 2 to 4 % and confirms the observation [201], that probably the luminosity is overestimated by around this factor for the HERA-II period.

7.4.4 Uncertainties of normalized jet cross sections

The main motivation for the measurement of normalized jet cross sections is the cancellation of experimental uncertainties. All normalization uncertainties, like the uncertainty on the luminosity, the trigger and the track-cluster-link uncertainty cancels completely (2.9 %). The uncertainties on the electron reconstruction cancels partly, in particular when the the uncertainty is of the same size for the numerator and denominator bin of the cross section ratio. This is the case for the uncertainty of the electron polar angle and the electron identification. The uncertainty of the electron energy has an opposite effect on the NC DIS cross section than on the jet cross sections. Therefore, this uncertainty increases slightly compared to absolute jet cross sections.
The uncertainties, which are predominantly due to the measurement of the hadronic final state, such as the JES uncertainty and the LAr noise uncertainty, are of the same relative size for the normalized cross sections as for the absolute jet cross sections. An overview of the relevant systematic uncertainties is displayed in figure 7.11. The dominant systematic uncertainty of the normalized jet cross sections arises from the JES uncertainty and the model uncertainty, which are of similar size as the statistical uncertainty. The contribution of all other systematic uncertainties is comparably small for most bins, when added in quadrature.

7.4.5 Normalized jet cross section results

The result of the normalized jet cross sections are displayed together with NLO predictions in figure 7.12 for the normalized inclusive jet, the normalized dijet and the normalized trijet measurement. The statistical correlation matrix of the three measurements was already shown in figure 7.9. The normalized jet cross sections are overall well described by the NLO predictions. The level of agreement can be better seen in figure 7.13, which shows the ratio of the normalized jet cross sections w.r.t. the NLO calculations. Most bins of the three normalized jet measurements are well described by the theoretical predictions. However, the predictions have the tendency to be above the normalized dijet measurement, which points towards a lower value of $\alpha_s(M_Z)$. Also, the NLO calculations are above the inclusive jet and dijet measurement for the highest p_{T}-bins in the Q^2 region of $270 < Q^2 < 5000\,\mathrm{GeV}^2$.

The predictions of various PDF sets differ by about 2 to 8 %, although the factorization scale is chosen consistently as $\mu_f^2 = Q^2$ for the calculation of the NC DIS cross sections $\sigma_{\mathrm{NC}}^{\mathrm{NLO}}$ and the jet cross sections $\sigma_{\mathrm{jet}}^{\mathrm{NLO}}$.
The PDF dependence arises from the different sensitivity of σ_{jet} and σ_{NC} to the PDF flavors and from contributions from different x-regions. The jet cross sections in every bin σ_{jet} are sensitive only to a small x-region (see e.g. figure 10.3), whereas partons from a large x-range contribute to the inclusive cross section σ_{NC}, since the hadronic final state is integrated out. Therefore σ_{NC} is sensitive to different x-regions than σ_{jet}. Furthermore, the different flavor sensitivity of σ_{NC} and σ_{jet} causes a non-vanishing PDF uncertainty, since the PDF

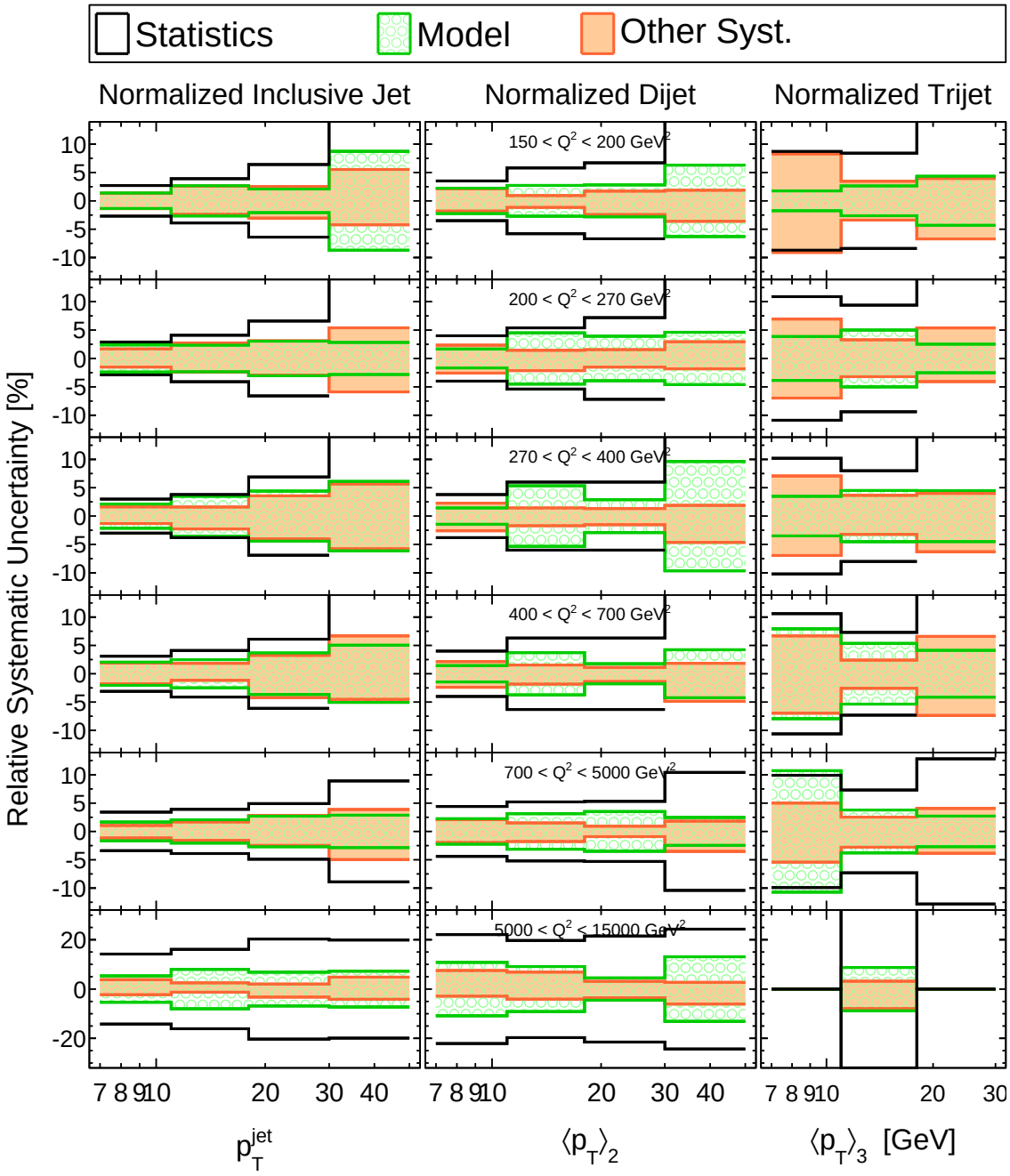


Figure 7.11: Relative uncertainties on the normalized jet cross sections. The dominant uncertainties are the statistical uncertainty (appears larger due to the correlations) and the model uncertainty. The size of the systematic uncertainties are similar to the uncertainty of the absolute jet cross sections, whereas all normalization uncertainties cancel. The remaining systematic uncertainties from the JES and RCES and LAr noise, as well as of the electron reconstruction ($\delta_{\mathrm{ID(e)}}\sigma$, $\delta_{\theta_e}\sigma$ and $\delta_{E_e}\sigma$) are added in quadrature on this plot. In some bins, the statistical uncertainties are larger than displayed range. The numerical values are given in appendix A.

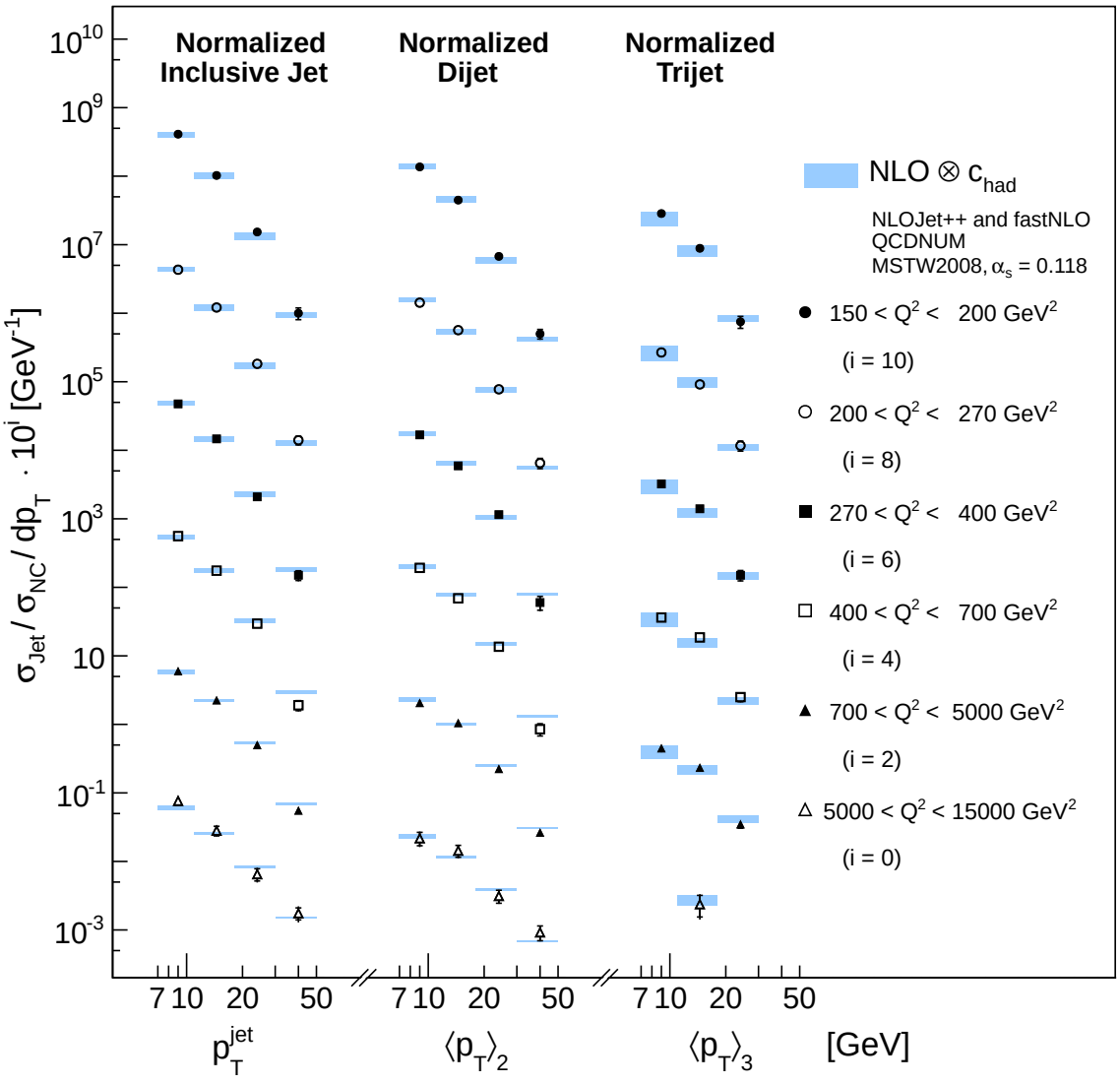


Figure 7.12: Double differential normalized inclusive jet, dijet and trijet jet cross sections. The NLO predictions are calculated using NLOJET++ together with fastNLO for the jet cross sections and QCDNUM for the inclusive NC DIS cross sections for the normalization, employing the MSTW2008 PDF set and $\alpha_s(M_Z) = 0.118$. The uncertainties are typically smaller than the markers. The uncertainties on the NLO calculations, shown as the blue bans, are determined by varying the renormalization scale by a factor of 0.5 and 2 for the NC DIS and the jet cross section calculations simultaneously.

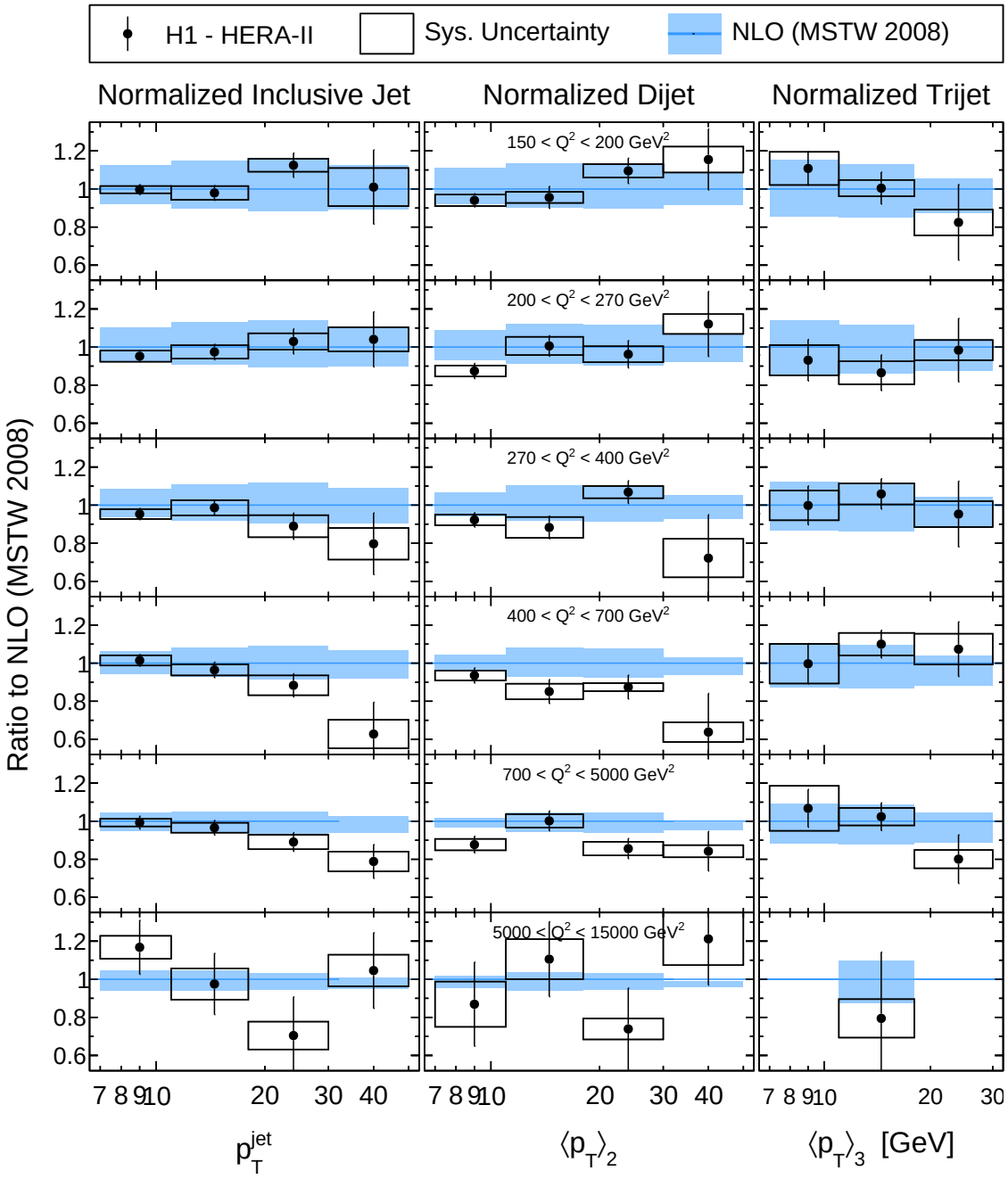


Figure 7.13: Ratio of the normalized jet cross sections w.r.t. NLO predictions using the MSTW2008 PDF. The uncertainty on the theory predictions are obtained by varying the renormalization and factorization scale simultaneously by a factor 0.5 and 2 for both, the jet cross section and the NC DIS calculation. The normalized jet cross sections have a non-vanishing PDF dependence due to the different x-dependence and sensitivity on different parton flavors of the NC DIS and the jet cross sections calculations. The systematic uncertainties are reduced for normalized cross sections compared to absolute jet cross section, since all normalization uncertainties cancel fully and uncertainties on the electron reconstruction cancels partly.

Bin labels ξ dijet

Bin number p_ξ	Label	ξ range
1	a	$0.006 \leq \xi < 0.02$
2	b	$0.02 \leq \xi < 0.04$
3	c	$0.04 \leq \xi < 0.08$
4	d	$0.08 \leq \xi < 0.316$

Bin labels ξ trijet

Bin number p_ξ	Label	ξ range
1	A	$0.01 \leq \xi < 0.04$
2	B	$0.04 \leq \xi < 0.08$
3	C	$0.08 \leq \xi < 0.5$

Table 7.2: Bin numbering scheme for ξ-bins for dijet and trijet events. Bins of the double-differential measurement are referred to by the Q^2-bin number as given in table 7.1 and the respective bin label, for instance 'bin 5c' corresponds to $700 < Q^2 < 5000\,\text{GeV}$ and $0.04 < \xi < 0.08$.

eigenvectors represent shifts for all flavors simultaneously. For instance, when an eigenvector has a large effect on the gluon density, due to the sum-rules, the quark densities have to compensate this effect. Such an eigenvector has an opposite effect on $\sigma_{\text{NC}}^{\text{NLO}}$ than on $\sigma_{\text{jet}}^{\text{NLO}}$.

The relative theoretical uncertainty from scale variations on the normalized jet cross sections are reduced compared to the absolute jet cross sections, since a variation of the renormalization scale has the same sign for both cross section calculations and is performed simultaneously for $\sigma_{\text{jet}}^{\text{NLO}}$ and $\sigma_{\text{NC}}^{\text{NLO}}$. Nevertheless, due to the smaller sensitivity of the NC DIS cross sections to $\alpha_s(\mu_r)$, which is implicitly the cause for the renormalization scale dependence of cross sections, this effect is rather small.

7.5 Jet cross sections differential in ξ

Cross sections for dijet and trijet production are also measured double-differentially as function of Q^2 and ξ. The variable ξ is defined for the dijet and trijet events[14] in equations 4.15 and 4.20 and is proportional to the momentum fraction of the incident parton w.r.t. the incident proton momentum. It takes values $\xi < 1$ and becomes large for processes at large scales, i.e. events at large Q^2 or $\langle p_{\text{T}} \rangle$.
The dijet and trijet extended as well as the measurement phase space (on generator and detector level) are defined identically to the $\langle p_{\text{T}} \rangle$-measurement (given in table 6.3 and 6.5)[15]. Dijet (trijet) events are measured in four (three) bins of ξ as given in table 7.2 and six bins in Q^2 (see table 7.1). The ξ phase space of the dijet (trijet) measurement is therefore restricted to $0.006 < \xi < 0.316$ [16] ($0.08 < \xi < 0.5$), whereas the extended phase space for the unfolding is not restricted in ξ. The bins 5a, 6a, 6b and 6c of the dijet measurement and 5A, 6A and 6B of the trijet measurement are further on excluded from the measurement due to vanishing cross sections.

The regularized unfolding is performed simultaneously for the inclusive jet measurement (as function of Q^2 and $p_{\text{T}}^{\text{jet}}$), the dijet measurement and trijet measurement, both as function of Q^2 and ξ. The data are taken in the extended dijet and trijet phase space. The

[14] The variable 'ξ' is used generically for the dijet and trijet definitions, ξ_2 and ξ_3, respectively, unless otherwise stated.

[15] The phase space of the measurement is defined in the NC DIS phase space $150 < Q^2 < 15\,000\,\text{GeV}^2$ and $0.2 < y < 0.7$. Dijet (trijet) events are defined by finding two (three) jets in the Breit frame with $7 < p_{\text{T}}^{\text{jet}} < 50\,\text{GeV}$ in the laboratory frame pseudorapidity $-1.0 < \eta_{\text{lab}}^{\text{jet}} < 2.5$. The invariant mass of the two leading jets have to fulfill $M_{12} > 16\,\text{GeV}$.

[16] $0.316 \approx 10^{-0.5}$

unfolding is initially derived from the definition of the $\langle p_T \rangle$-measurement[17], where the $\langle p_T \rangle$-dimension is replaced by a ξ-bin grid, and is then improved with a fourth dimension for M_{12} and M_{123}. The invariant mass of the two (three) leading jets directly enters the calculation of ξ, but the phase space is restricted to $M_{12} > 16\,\text{GeV}$, and therefore these migrations become relevant. The unfolding of the ξ-measurement is described and studied in more detail section 9.4.

The cross sections for dijet and trijet production as function of Q^2 and ξ are presented in figure 7.14. The ratio of the measured cross sections to the NLO predictions are shown in figure 7.15. We observe an overall good agreement between the measurement and the theory predictions within the uncertainties. However, the dijet data are typically below the calculations, similarly to the $\langle p_T \rangle$ measurement. The trijet measurement suffers from large systematic uncertainties, which are mostly dominated by the model uncertainty.
A detailed split-up of the dominant systematic uncertainties is shown in figure 7.16. The dijet measurement is dominated by the (correlated) statistical uncertainty. However, also the model uncertainty can reach up to 10 %. It shows large fluctuations after the unfolding in the lowest Q^2 bin. Such large fluctuations between adjacent bins are expected to be unphysical and should be interpolated, if used in a phenomenological study. The trijet measurement is limited by the model uncertainty which reaches values up to 40 % on the cross section value. Further systematic uncertainties are the normalization uncertainty from the luminosity uncertainty, the trigger uncertainty the track-cluster-link uncertainty with an overall normalization uncertainty of 2.9 %, and the LAr noise uncertainty of 0.6 % and 0.9 % for the dijet and trijet measurement, respectively. The uncertainties due to the lepton-reconstruction are found to be small with values of below 1 % for the dijet cross sections and below 2 % for the trijet cross sections.

Similarly, as for the $\langle p_T \rangle$-measurement, also the ξ-measurement can be normalized w.r.t. the inclusive NC DIS measurement, using the information of the covariance matrix after the unfolding. The normalized cross sections for dijet and trijet production as function of Q^2 and ξ are shown in figure 7.17. The normalized cross sections benefit from the full cancellation of all normalization uncertainties. However, since the measurements are dominated by statistical and model uncertainties, as well as uncertainties due to the reconstruction of the hadronic final state, the advantage w.r.t. the absolute cross sections turns out to be small.

[17] Since the unfoldings are performed separately for $\langle p_T \rangle$ and ξ, no covariance matrix between the '$\langle p_T \rangle$' and 'ξ'-measurement was derived.

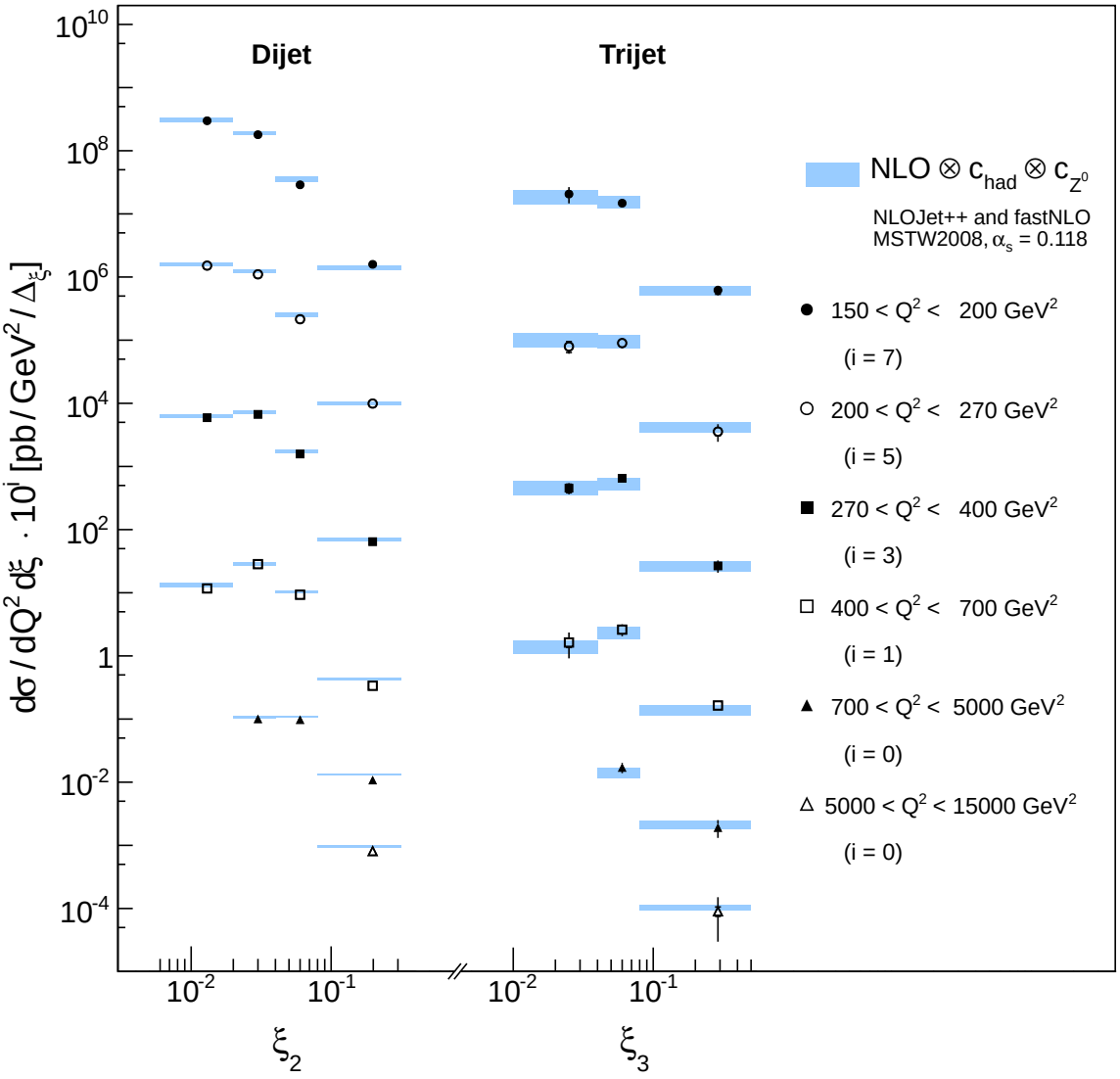


Figure 7.14: Cross sections for dijet and trijet production as function of Q^2 and ξ on the non-radiative hadron level. The cross sections have been unfolded simultaneously with the inclusive jet measurement (not shown here), and the full covariance matrix of the three measurements is available. The inner error bars (mostly smaller than the marker size) indicate the statistical precision, and the outer error bars give the systematic uncertainties added in quadrature. The data are compared to pQCD calculations in NLO. The latter are corrected for hadronization effects and Z^0 and γZ interference enhancement. The blue bands indicate the NLO calculations within the scale variation by a conventional factor of 2 up and down.

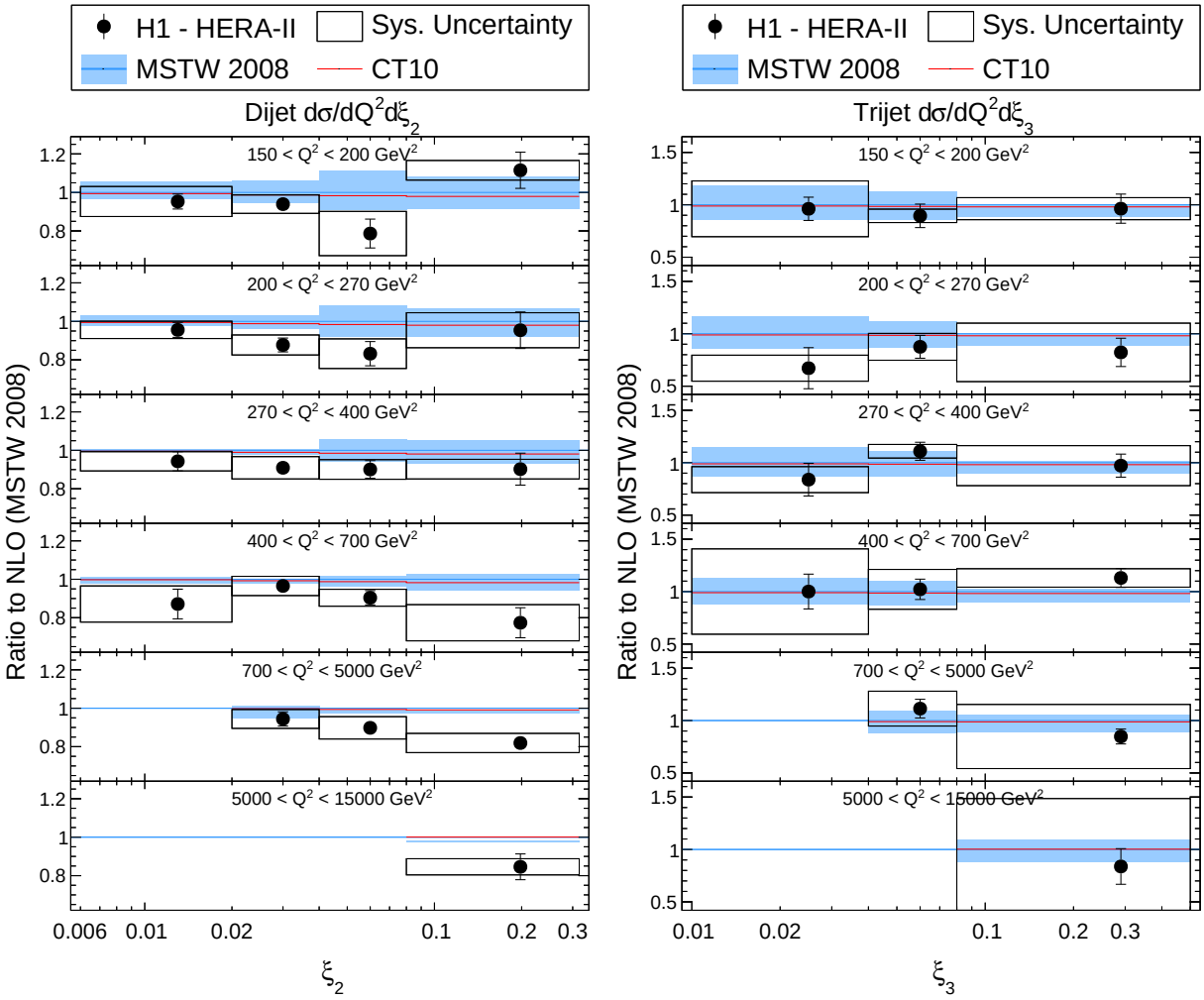


Figure 7.15: Ratio of the double differential (left) dijet and trijet cross sections (right) as function of Q^2 and ξ w.r.t. NLO predictions using the MSTW2008 PDFs (please note the different scales). The PDFs have been determined at $\alpha_s(M_Z) = 0.118$. The systematic uncertainties indicate the sum of all systematic uncertainties added in quadrature and are dominated by the model uncertainty.

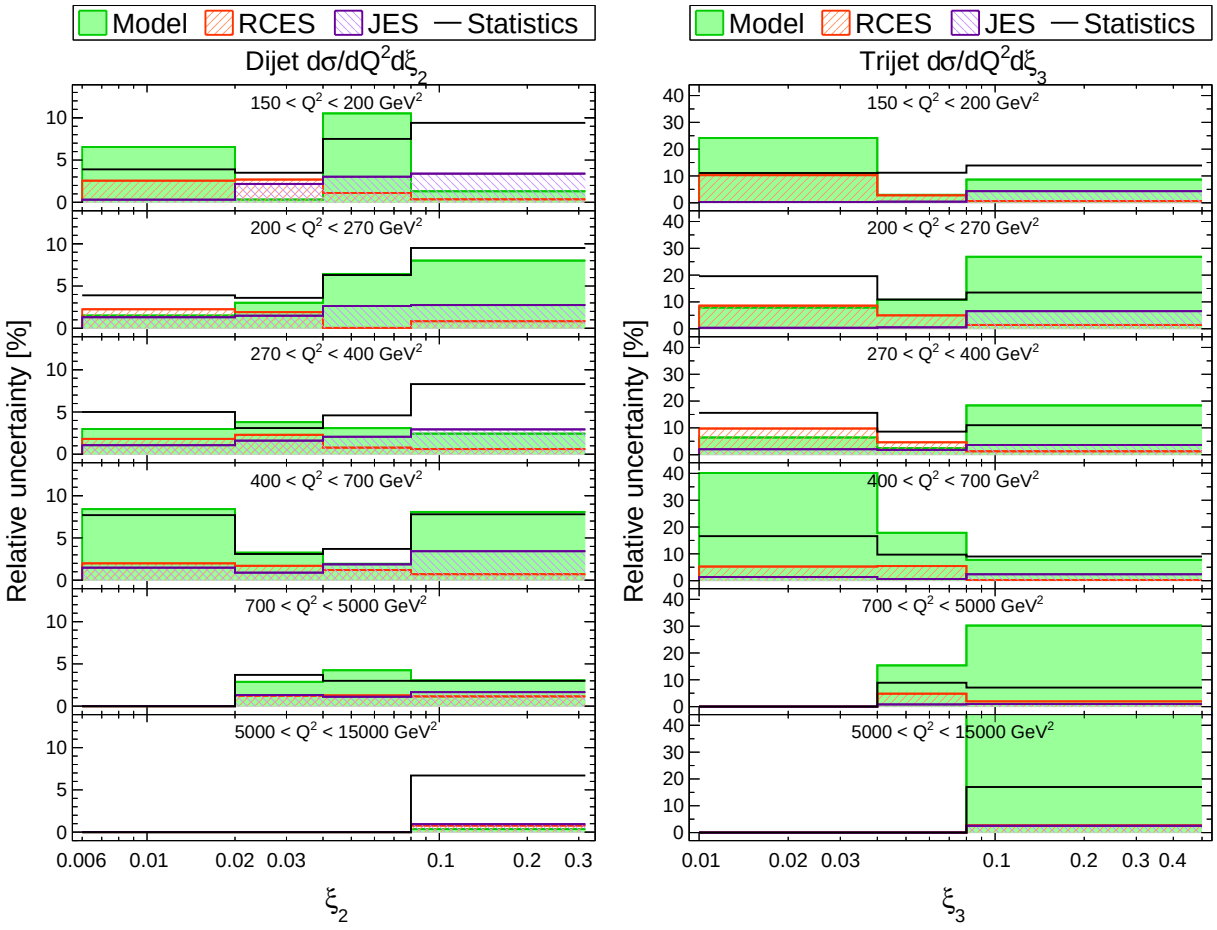


Figure 7.16: The dominant relative uncertainties of the dijet (left) and trijet measurement (right) as function of Q^2 and ξ, where the (linear) average of the 'up' and 'down'-uncertainties is displayed. The dijet measurement is dominated by the (correlated) statistical uncertainty and the model uncertainty. The trijet measurement is limited by the model uncertainty. The overall normalization uncertainty is 2.9 % and the uncertainties due to the scattered lepton reconstruction are below 2 %.

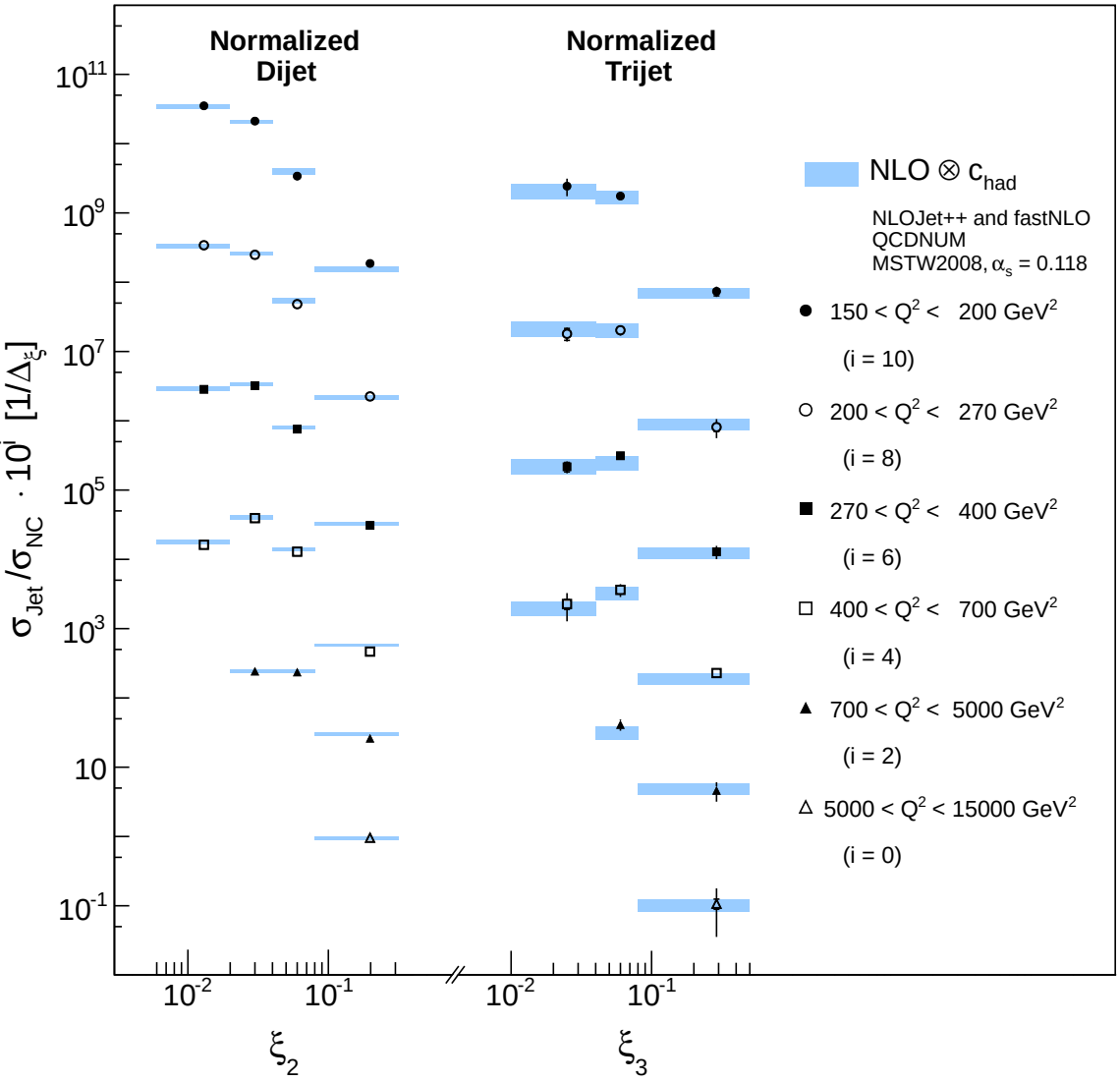


Figure 7.17: Normalized cross sections for dijet and trijet production as function of Q^2 and ξ on the non-radiative hadron level. The normalization is performed w.r.t. the inclusive NC DIS cross section as function of Q^2. An analytic error propagation using the full covariance matrix between the measurements is performed. More details are given in figure 7.14 and figure 7.12.

7.6 Cross section ratio $R_{3/2}$

The cross section ratio $R_{3/2}$ of the trijet to dijet cross sections is presented in this section. This ratio is measured for the first time double-differentially in ep. It is defined by

$$R_{3/2,i} = \frac{\sigma_i^{\text{trijet}}}{\sigma_i^{\text{dijet}}}. \tag{7.18}$$

The advantages offered by of the jet-cross section ratio, are the full cancellation of normalization uncertainties as well as the partial cancellations of other systematic uncertainties. Further advantages of $R_{3/2}$ are also the partial cancellation of some theoretical uncertainties. The NLO prediction is calculated according to

$$R_{3/2}^{\text{NLO}} = \frac{\sigma_{\text{trijet}}^{\text{NLO}}}{\sigma_{\text{dijet}}^{\text{NLO}}}. \tag{7.19}$$

In this expression, the hadronization corrections are determined seperately for the trijet and the dijet cross sections. The dependence of the electromagnetic interaction and the electroweak corrections cancel. Also the contribution from some matrix elements would cancel directly in this expression in pQCD, like e.g. the real correction to the dijet leading-order diagrams, which is of leading-order for σ_{trijet} (see figures 2.13a and b). Due to the necessary cancellations of the singularities in the numeric calculations, the quantity $R_{3/2}^{\text{NLO}}$ can not be calculated directly, but only as the ratio of $\sigma_{\text{trijet}}^{\text{NLO}}$ to $\sigma_{\text{dijet}}^{\text{NLO}}$. The dependence on the PDF cancels partly in the ratio, and thus the PDF uncertainty decreases compared to the one for the absolute cross sections. A residual PDF dependence persists, however, because of the different x-dependence of the truncated $\sigma_{\text{trijet}}^{\text{NLO}}$ and $\sigma_{\text{dijet}}^{\text{NLO}}$ calculations, with similar implications as already discussed for the normalized jet cross sections.
The partial cancellation of theoretical uncertainties also leads to a decrease of the $\alpha_s(M_Z)$ dependence of $R_{3/2}$, as was shown already in [74]. This reduces the sensitivity of $R_{3/2}$ in a determination of $\alpha_s(M_Z)$ compared to the sensitivity of the absolute trijet measurement. Furthermore, the statistical precision is less compared to the one of the dijet cross sections, since the trijet measurement is a subsample of the dijet measurement (see section 7.4.2.1 on the error propagation), and therefore $R_{3/2}$ is not used for the determination of $\alpha_s(M_Z)$ in this work.
It must be further mentioned that the definition of $R_{3/2}$ is not strictly given by any convention. Here, we define the double-differential ratio as function of two different observables; namely $\langle p_{\text{T}} \rangle_3$ over $\langle p_{\text{T}} \rangle_2$, in correspondence to a previous single-differential analysis [74,102] and since these observables are also measured separately. Thus, trijet events with a given $\langle p_{\text{T}} \rangle_3$ may be measured in different $\langle p_{\text{T}} \rangle_2$-bins as the dijet event. For the benefit of the cancellation of experimental uncertainties, a dedicated measurement of '$R_{3/2}$' would be necessary, where also the observables are defined more carefully. One could also put a focus on the cancellation of theoretical uncertainties, as it is e.g. done for $R_{\Delta R}$ in [202] or for the newly proposed observable $R_{\Delta\phi}$ [203].
The double-differential measurement of $R_{3/2}$ is presented in figure 7.18. The shape of the data is well described by the NLO predictions. Some single bins have relatively large deviations from the NLO predictions, but also the statistical and systematic uncertainties are $\geq 10\,\%$ in most of the bins.
The measurements are compared to predictions using different PDF sets, where it can be observed that the PDF dependence has been canceled to a large degree.

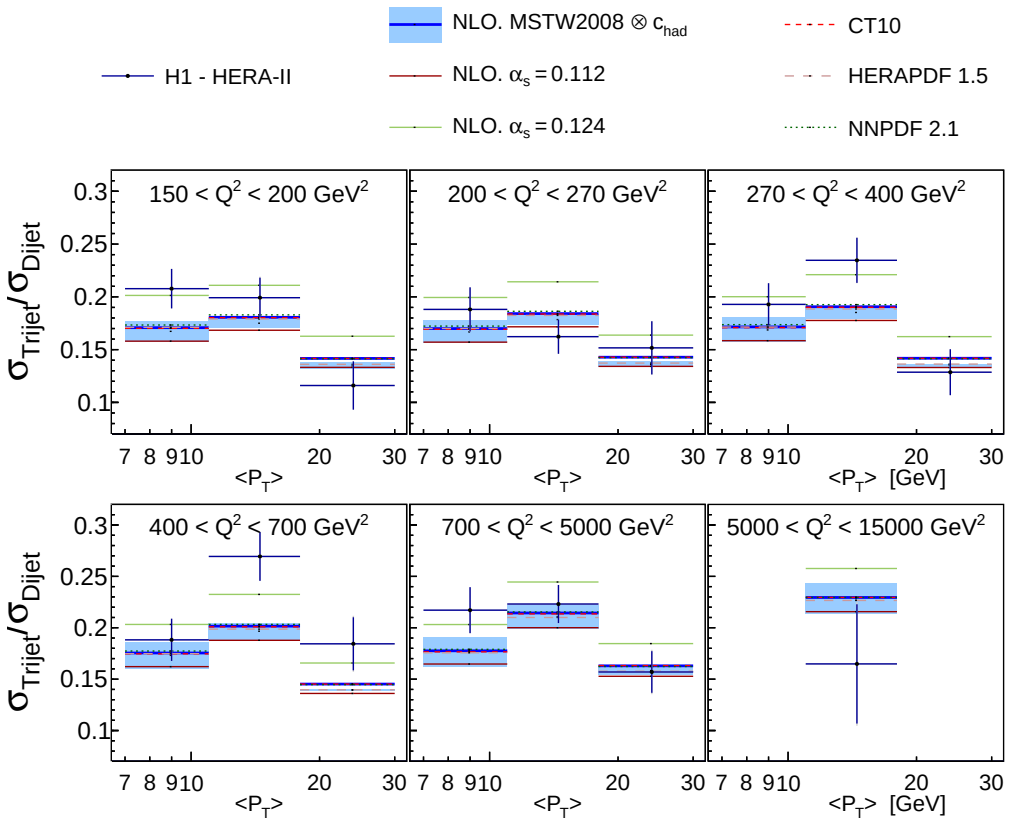


Figure 7.18: The double differential ratio of trijet to dijet cross sections $R_{3/2}$. The measurements are compared to NLO predictions using different PDF sets, where it can be seen that the residual PDF dependence largely cancels in the cross section ratio. Predictions for different values of $\alpha_s(M_Z)$ are also depicted in the plot.

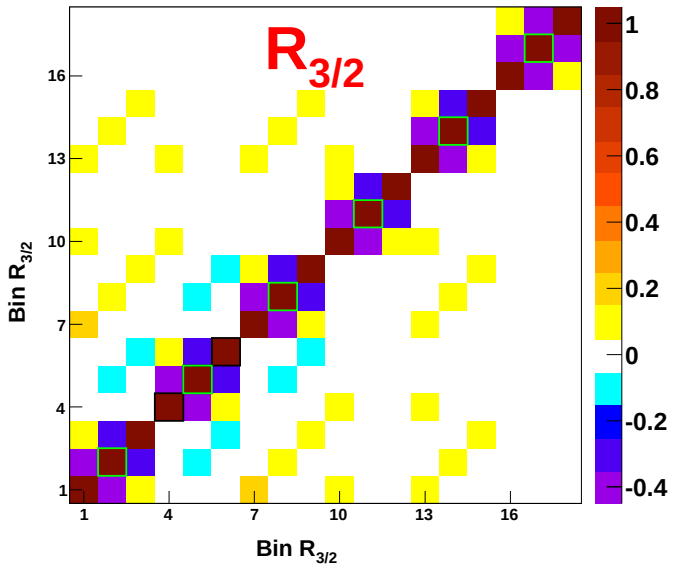


Figure 7.19: Correlation matrix of the double-differential measurement of $R_{3/2}$. The bin numbers are given by equation 7.2 with $n_{p_T} = 3$.

The statistical correlation matrix, which is determined by error propagation of the statistical covariance matrix after the unfolding process is shown in figure 7.19. The statistical correlations in neighboring $\langle p_{\mathrm{T}} \rangle$-bins are around -0.2 to -0.4 and therefore, the statistical uncertainties appear to be larger in the plots than they actually are.

Chapter 8

Comparison of regularized unfolding to a bin-by-bin correction

In this chapter the applied regularized unfolding method is compared to the bin-by-bin correction method. The bin-by-bin correction method is first introduced, and the error determination applied is explained. The correction factors derived and the cross section results are compared, as well as the statistical uncertainty, the systematic uncertainties and the model uncertainty.

8.1 The method of bin-by-bin correction

The *bin-by-bin correction* method (BBB) or *correction factor method* (CFM) is an alternative method to the regularized unfolding to correct the measurement for detector effects, such as limited acceptance and migration effects. Due to the simplicity of the method it is historically widely used in many measurements and is often also denoted as "unfolding".
The bin-by-bin method employs an identical bin grid on generator and on detector level. The detector correction is performed by a multiplicative correction factor c_i^{det} which is applied to each bin i and which is defined as

$$c_i^{\text{det}} = \frac{N_i^{\text{gen,MC}}}{N_i^{\text{rec,MC}}} \ , \tag{8.1}$$

where $N_{\text{gen}}^{\text{MC}}$ and $N_{\text{rec}}^{\text{MC}}$ denote the number of generated events and reconstructed events, respectively, in a bin.
The correction factor for higher-order QED effects c_i^{rad} is defined for each bin i as the ratio of events generated without higher-order QED effects over the number of events including these effects (see section 6.9).
The bin-by-bin correction factors c_i^{det} and c_i^{rad} are determined for the two Monte Carlo event generators employed and an *experimental correction* factor is defined as

$$c_i^{\text{exp}} = c_i^{\text{det}} \cdot c_i^{\text{rad}} \ . \tag{8.2}$$

In this analysis, the total experimental correction factor is determined from the arithmetic mean of the correction factors from Django $c_{\text{Dj}}^{\text{exp}}$ and Rapgap $c_{\text{Rg}}^{\text{exp}}$, by

$$c_i^{\text{exp}} = \frac{c_{\text{Dj,i}}^{\text{exp}} + c_{\text{Rg,i}}^{\text{exp}}}{2} \ . \tag{8.3}$$

The measured cross section is then defined, using the background subtracted data and the luminosity, through

$$\sigma_i = \frac{N_i^{\text{data}} - N_i^{\text{bkg}}}{\mathcal{L}} c_i^{\text{exp}} \ . \tag{8.4}$$

The uncertainty on the experimental correction is denoted as model uncertainty $\delta_{\text{model}}\sigma$ and is chosen to be the half difference between the corrections from Rapgap and Django:

$$\delta_{\text{Model}}\sigma_i = \frac{c_{\text{Dj},i}^{\text{det}} - c_{i,\text{Rg}}^{\text{det}}}{2} \ . \tag{8.5}$$

Due to a "crossing" of the predicted correction factors of the single models, the model uncertainty can be negligibly small in some bins. In order to quote a reasonable value, the model uncertainty undergoes a single exponential smoothing algorithm [204], which interpolates the size of the uncertainty w.r.t. to the neighboring bins in two dimensions (see [130] for details).

The bin-by-bin correction method relies on an accurate description of the measured data and the migrations by the MC event generator than in contrast to the regularized unfolding procedure, because in the former method the detector correction is based on the absolute values $N_i^{\text{gen,MC}}$ and $N_i^{\text{rec,MC}}$. The regularized unfolding has the advantage of representing the behavior of the experimental device by a normalized migration matrix and thus is dependent only on relative numbers. The migration matrix of the regularized unfolding enables furthermore to estimate the migrations between different bins in the detector more accurately. Another advantage of the regularized unfolding is the usage of a finer bin grid, compared to the bin-by-bin correction. This enables a more detailed representation of the detector effects[1].

In the following, we want to investigate the difference in the size of the correction between the two correction methods, their difference in the statistical uncertainty and their dependence on the model predictions.

8.2 Purity and acceptance

In order to quantify the detector effects and the size of migrations which have to be corrected for, it is useful to study the migrations into and out of a bin. For each bin i three numbers are determined. The number of events[2] which migrate out of the bin, migrate into the bin, or stay in the bin are determined from the MC event generator's hadron level. It is helpful to distinguish the two distinct regions into which and out of which migrations take place: from other bins j than bin i of the measurement, or from an "outside of the phase space" region, which also includes migrations resulting from different jet multiplicities. Therefore, the five uncorrelated event numbers are defined:

[1] This is no general limitation of the bin-by-bin correction, which could principally also be applied on a finer bin grid.

[2] Here, "event" is synonymously used for any quantity which is measured. In case of inclusive jets for instance this is a jet instead of the actual event.

No migrations

$N_i^{\text{stay}} = N_{\text{gen}(i)\wedge\text{rec}(i)}$	Number of events which are generated and reconstruced in the same bin.

Migrations into the detector level bin

$N_i^{\text{in}} = N_{\text{gen}(j)\wedge\text{rec}(i)}$	Number of events which are reconstructed in bin i, but are generated in another bin j of the measurement phase space.
$N_i^{\text{gain}} = N_{\neg(\text{gen}(j)\vee\text{gen}(i))\wedge\text{rec}(i)}$	Number of events which are reconstructed in bin i, but are generated not within a phase space bin, or jets which result from different jet-multiplicities on generator than on detector level.

Migrations out of the generator level bin

$N_i^{\text{out}} = N_{\text{gen}(i)\wedge\text{rec}(j)}$	Number of events which are generated in bin i, but are reconstructed in another bin j.
$N_i^{\text{lost}} = N_{\text{gen}(i)\wedge\neg(\text{rec}(j)\vee\text{rec}(i))}$	Number of events which are generated in bin i, but are not reconstructed in any phase space bin. These are for instance events, which are lost due to the limited acceptance of the detector or due to trigger inefficiencies.

The total number of generated events N_i^{gen} and reconstructed events N_i^{rec} can be expressed through

$$N_i^{\text{gen}} = N_i^{\text{stay}} + N_i^{\text{out}} + N_i^{\text{lost}} \tag{8.6}$$

$$N_i^{\text{rec}} = N_i^{\text{stay}} + N_i^{\text{in}} + N_i^{\text{gain}} \ . \tag{8.7}$$

The *acceptance* of the bin i of the experimental device is defined by

$$A_i = \frac{N_i^{\text{rec}}}{N_i^{\text{gen}}}. \tag{8.8}$$

Hence, the (bin-by-bin) detector correction factor can also be expressed through $c_i^{\text{det}} = 1/A_i$.
A characteristic number to quantify the amount of migrations in a measurement is the *Purity* P_i. The purity is defined as [205]

$$P_i = \frac{N_i^{\text{stay}}}{N_i^{\text{rec}}} \tag{8.9}$$

and represents the fraction of number of events which are generated as well as reconstructed in a bin over the total number of reconstructed events in that bin. It expresses the fraction of events which indeed belong to the considered bin i.
The values for the purity and acceptance are displayed in figure 8.1 for the three double-differential jet measurements. The purity and acceptance are characteristic quantities for any experimental device and measurement. Since the unfolding method employed

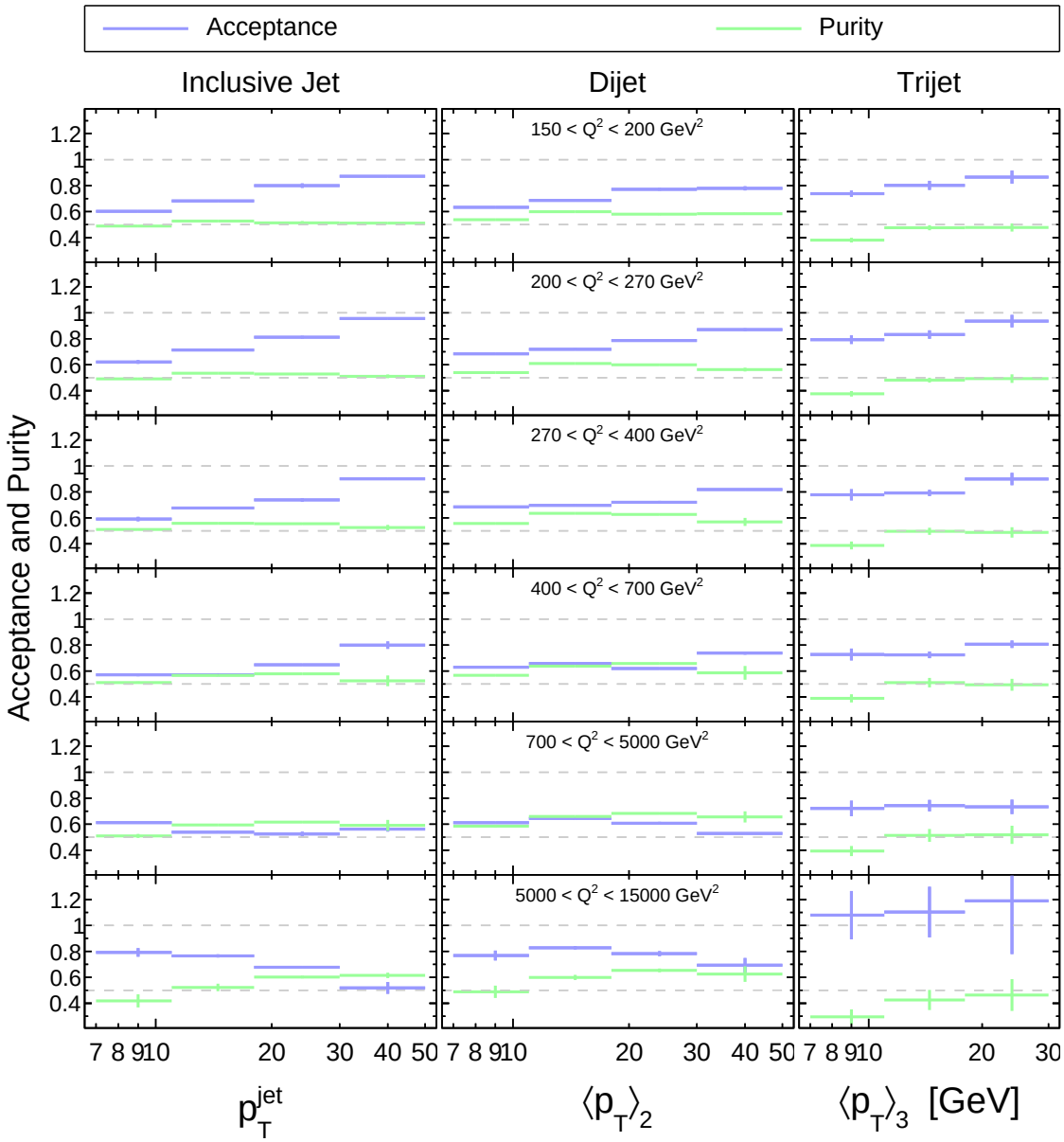


Figure 8.1: The purity P and acceptance A for the three jet measurements. The values are calculated as the average from Django and Rapgap and the uncertainty bars indicate the half difference between the two models.

incorporates a different bin grid on detector level than on generator level, these values are specific for the bin-by-bin correction method.
Another frequently used number is the *stability* S_i which is defined as

$$S_i = \frac{N_i^{\text{stay}}}{N_i^{\text{gen}} - N_i^{\text{lost}}} \,. \tag{8.10}$$

It expresses the fraction of "stay" events w.r.t. the number of generated events, which stay within the phase space of the measurement. This definition of the stability is of course not characteristic for a bin, since it is dependent on the phase space definition of the full measurement.

8.3 Discussion of the bin-by-bin method

In this section we want to illuminate the biases of the bin-by-bin method[3], which are largely not existent for the regularized unfolding.

8.3.1 The central value

The bin-by-bin method gives only an unbiased result if no migrations are present. This, however, is typically not the case. Let's regard the effect of the bin-by-bin method in the presence of migrations. Here, it is firstly irrelevant if migrations take place towards other bins of the phase space or towards the outside of the phase space region. Therefore, we introduce the shorthand notations $N_i^+ = N_i^{\text{in}} + N_i^{\text{gain}}$ and $N_i^- = N_i^{\text{out}} + N_i^{\text{lost}}$. Neglecting the radiative correction factor and background contributions from other processes the corrected hadron level number of events N_i^{bbb} reads

$$N_i^{\text{bbb}} = N_i^{\text{data}} c_i^{\text{det}} = N_i^{\text{gen}} \frac{N_i^{\text{data}}}{N_i^{\text{stay}} + N_i^+}. \tag{8.11}$$

The migrations have to be estimated, using the Monte Carlo prediction and thus, the expected number of events can be inserted and above equation reads [205]

$$N_i^{\text{bbb}} = N_i^{\text{gen}} \frac{N_i^{\text{staytrue}} + N_i^+}{N_i^{\text{stay}} + N_i^+}, \tag{8.12}$$

where N_i^{staytrue} is the true number of measured events which are in bin i on hadron and detector level. If the migrations are now getting large, i.e. $N_i^+ \to N_i^{\text{rec}}$, the corrected number of hadron level events is identical to the Monte Carlo prediction $N_i^{\text{bbb}} \to N_i^{\text{gen}}$, since the migrations out of the bin and into the bin are fixed by Monte Carlo assumptions [205]. The data will thus follow the MC that gave the correction factors [189, 206]. The size of the bias in the estimate is hard to calculate because of the nonlinearity of the procedure [189, 207]. The bias arises from the fact, that the finite number of events contains information only about the visible component of the function being measured, which is only the projection on the subspace of the detector response [208].

[3] Some people even denote the bin-by-bin method as "a disaster" [206] due to these biases.

8.3.2 The statistical uncertainty

The statistical uncertainty for the bin-by-bin correction is calculated through the application of the correction factor to the statistical uncertainty on the detector level uncertainty

$$\Delta N_i^{\mathrm{bbb}} = c_i^{\mathrm{exp}} \Delta N_i^{\mathrm{rec}}. \tag{8.13}$$

We want to show the biases which are introduced in the bin-by-bin correction through an example toy model.

Assuming two bins I and X are measured. The bin I is a phase space bin, and the bin X can either represent another phase space bin or the outside of the measurement phase space region. The number of generator level events are chosen to be $N_I^{\mathrm{gen}} = 20$ and $N_X^{\mathrm{gen}} = 30$. A reasonable detector response matrix $\boldsymbol{A}$ is assumed such that the folding equation $\boldsymbol{A}x = y$ reads

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \begin{pmatrix} 20 \\ 30 \end{pmatrix} = \begin{pmatrix} 28 \\ 22 \end{pmatrix}. \tag{8.14}$$

The bin-by-bin correction is compared to a direct matrix inversion, which is possible for this simple problem. The toy model is constructed such, that the corrected generator level is identical for both correction methods (please keep in mind the comments in the previous section on the biases which may be introduced in the bin-by-bin correction method). Given the detector response matrix $\boldsymbol{A}$, the Monte Carlo generator level prediction and the measured data, the two corrections can be compared. The results of the two correction methods are outlined in table 8.1.

For the constructed toy model, the corrected hadron level results are identical for the two compared correction methods. However, the statistical uncertainties differ. The bin-by-bin method underestimates the statistical uncertainty on the phase space bin I (and X), since the migrations from other bins are not propagated correctly to the hadron level. The bin-by-bin method determines the uncertainty on bin I to $\Delta N_I^{\mathrm{gen,BBB}} \approx 4.26$, while the correct statistical uncertainty would be $\Delta N_I^{\mathrm{gen,matrix}} \approx 6.50$. When bin X is interpreted as the outside of the phase space region and only bin I is measured, the derived statistical uncertainty is simply wrong, since the knowledge of X is discarded.

In this toy model we can also regard bin X as a second bin within the phase space. We can then combine the bins I and X and determine the "total cross section" T. Of course, both correction methods predict again the same bin content $N_T = 50$. When now propagating the statistical uncertainty from the hadron level covariance matrix, the bin-by-bin correction predicts a statistical uncertainty of $\Delta N_T^{\mathrm{BBB}} \approx 7.09 > \sqrt{50}$, while the matrix inversion determines the correct statistical uncertainty of $\Delta N_T^{\mathrm{matrix}} = \sqrt{50}$.

It is further remarkable, that the absolute uncertainty of bin X is larger than the uncertainty of the combined bin T for the unfolding method. Such peculiarities are only possible, if the correlations are considered[4]. This means, that the precision of bin X is in fact increased, if the information of bin I is further considered. In case of the bin-by-bin

[4]The proof for the uncorrelated case is trivial: $(\Delta N_T)^2 = \sum_i^{\mathrm{bins}} (\Delta N_i)^2 > (\Delta N_j)^2$, for every bin j, since $(\Delta N_i)^2 > 0$ for real numbers.

Toy model "unfolding"

	Bin-by-bin	**Matrix inversion**
Monte Carlo input		
Detector response ("folding")	$\boldsymbol{A} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$	
Monte Carlo hadron level	$N_I^{\text{gen,MC}} = 20$ and $N_X^{\text{gen,MC}} = 30$	
Monte Carlo detector level	$N_I^{\text{rec,MC}} = 22$ and $N_X^{\text{rec,MC}} = 28$	
Correction method		
Correction equation	$x_i = c_i^{\text{det}} \cdot N_i$	$\vec{x} = \boldsymbol{A}^{-1}\vec{y}$
Correction	$c_I^{\text{det}} = \frac{20}{22}$ $c_X^{\text{det}} = \frac{30}{28}$	$\boldsymbol{A}^{-1} = \frac{1}{3}\begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}$
Error "propagation" to hadron level	$\Delta N_i^{\text{gen}} = c_i^{\text{det}} \cdot \Delta N_i$	$\boldsymbol{V}_x = \boldsymbol{A}^{-1}\boldsymbol{V}_y\left(\boldsymbol{A}^{-1}\right)^{\text{T}}$
Conclusion of error propagation	$\boldsymbol{V}_y = \begin{pmatrix} (\delta N_I^{\text{gen}})^2 & 0 \\ 0 & (\delta N_X^{\text{gen}})^2 \end{pmatrix}$	$\delta N_i^{\text{gen}} = \sqrt{V_{ii}}$
Detector level		
Measured bin I with uncertainty	$N_I = 22 \pm \sqrt{22}$	
Measured bin X with uncertainty	$N_X = 28 \pm \sqrt{28}$	
Covariance matrix $\boldsymbol{V}_y$	$\boldsymbol{V}_y = \begin{pmatrix} 22 & 0 \\ 0 & 28 \end{pmatrix}$	
Hadron level		
Corrected result bin I	$N_I^{\text{gen}} = 20$	
Corrected result bin X	$N_X^{\text{gen}} = 30$	
Uncertainty on bin I	$\Delta N_I^{\text{gen}} = \frac{20}{\sqrt{22}} \approx 4.26$	$\Delta N_I^{\text{gen}} = \frac{\sqrt{380}}{3} \approx 6.50$
Uncertainty on bin X	$\Delta N_X^{\text{gen}} = \frac{30}{\sqrt{28}} \approx 5.67$	$\Delta N_X^{\text{gen}} = \frac{\sqrt{470}}{3} \approx 7.23$
Covariance matrix	$\boldsymbol{V}_x = \begin{pmatrix} \frac{400}{22} & 0 \\ 0 & \frac{900}{28} \end{pmatrix}$	$\boldsymbol{V}_x = \frac{1}{9}\begin{pmatrix} 380 & -200 \\ -200 & 470 \end{pmatrix}$
Combined hadron level bin $\boldsymbol{T := I + X}$		
Combined bin content $N_T = N_I + N_X$	$N_T^{\text{gen}} = 50$	
Error propagation to T	$(\Delta N_T)^2 = V_{II} + V_{XX} + 2V_{IX}$	
Uncertainty on combined bin ΔN_T	$\Delta N_T = \sqrt{\frac{3875}{77}} \approx 7.09$	$\Delta N_T = \sqrt{50} \approx 7.07$

Table 8.1: Results for a toy model detector correction. Compared are the bin-by-bin correction method and a direct matrix inversion. The toy model is constructed such that the generator level results are identical for both correction methods. One can see that the predicted statistical uncertainties of the bin-by-bin method are not reliable.

method in contrast, the precision of a bin is only dependent on its own information.

We conclude that the bin-by-bin correction method is not capable of a reliable error esti-

mation as soon as migrations are present due to the non-linearity of the procedure [207]. The statistical uncertainties are underestimated, since the correlations to other bins and the outside phase space region are not propagated. The matrix based method on the other hand enables the correct error propagation, since the effect of the migrations are propagated correctly through the unfolding, and the overall statistical precision is preserved.

8.4 Monte Carlo study

The difference between the two correction methods can also be studied using Monte Carlo event generators. If a pseudo-data set is generated with one Monte Carlo generator, i.e. Django or Rapgap in this work, then the unfolding matrix or the correction factors can be obtained from the prediction of the other event generator. Since the underlying "true" distribution is known, the comparison of the corrected distribution to the generated hadron level distribution is an estimate about the biases of a correction method.

Monte Carlo pseudo-data[5] x_A^{true} are generated with model A with the same statistical precision as the real data. The migration histogram $\boldsymbol{A}_B$ is determined using the full Monte Carlo statistics of the other model $B \neq A$. In case of the bin-by-bin method, the correction factors c_B are obtained from model B. The notation for the two correction methods can be synchronized by defining

$$\boldsymbol{A}^{\text{bin-by-bin}} = \begin{pmatrix} \frac{1}{c_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{c_2} & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \frac{1}{c_n} \end{pmatrix} . \tag{8.15}$$

The corrected hadron level pseudo-data x_A is then determined from the reconstructed pseudo-data[6] y_A, which is corrected using a different model through

$$x_A = \boldsymbol{B}_B^m y_A \ , \tag{8.16}$$

where $\boldsymbol{B}_B^m$ denotes the detector correction using the correction method $m \in \{\text{unfold, bin-by-bin}\}$. For the regularized unfolding $\boldsymbol{B}^{\text{unfold}}$ is determined using equation 5.8, or in case of the bin-by-bin method it is a diagonal matrix and defined as $\boldsymbol{B}^{\text{bin-by-bin}} = (\boldsymbol{A}^{\text{bin-by-bin}})^{-1}$ with the elements $(B_{ii}) = c_i^{\text{exp}}$. The corrected pseudo-data x_A are then compared to their initial hadron level x_A^{true} by calculating the bin-wise *pull* values p_i, which are defined through

$$p_i = \frac{x_{A,i} - x_{A,i}^{\text{true}}}{\Delta x_{A,i}} \ . \tag{8.17}$$

The statistical uncertainty $\Delta x_{A,i}$ is determined through error propagation of the detector level statistics $\boldsymbol{V}_x = \boldsymbol{B}_m \boldsymbol{V}_y \boldsymbol{B}_m^{\text{T}}$ and hence $\Delta x_i = \sqrt{V_{ii}}$, where $\boldsymbol{B}_m$ differs for the two correction methods.

[5] For better readability, the hadron level and detector level bin grids are not labeled explicitly. It becomes clear from the context that in case of the bin-by-bin correction, the bin grid of x and y are the measurement bin grid, while for the regularized unfolding, the bin grid of x and y have been defined in chapter 5.

[6] Formally, the reconstructed pseudo data may be written as $y_A = \boldsymbol{A}_A^m x_A$. This, however, holds only for the central value, since the statistical uncertainties are obtained on detector level.

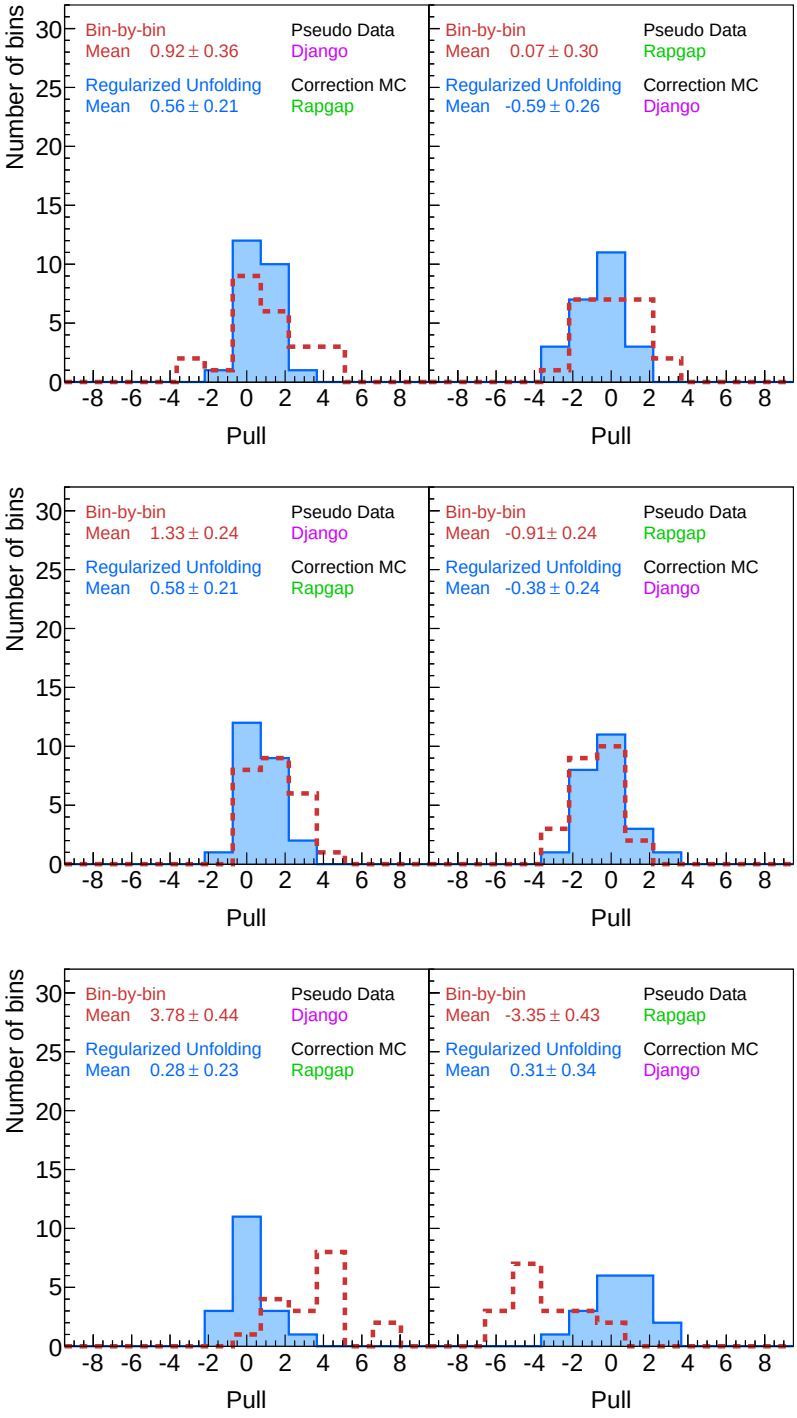


Figure 8.2: The pull values between the "true" and the unfolded hadron level for the two correction methods. Shown are all bins of the inclusive jet (top), the dijet (middle) and trijet pseudo-data (bottom). The left pads show the Django pseudo data corrected using the Rapgap model, while the right pads show Rapgap pseudo-data corrected with the Django model. The arithmetic mean values show that the bin-by-bin correction is biased, while the unfolded data are better compatible with their true distribution. This is particularly distinct for the trijet pseudo-data, where the two models predict significantly different corrections factors (compare uncertainty on bin-by-bin correction factor in figure 8.3).

The pull values for the bins of the inclusive jet, the dijet and the trijet measurements are illustrated in figure 8.2. The *mean* values are calculated using the arithmetic mean from all phase space bins. One observes, that the mean values for the bin-by-bin correction are often not compatible with zero. This means, that the corrected data is not compatible with its true distribution and a bias is present. The unfolded pseudo-data, however, is better compatible with the true level, although small shifts can be observed for the inclusive jet and dijet measurement. The pull values of the trijet pseudo-data differs significantly from 0. Although both models give a good description of data on detector level (see section 4.4.7), they differ significantly on generator level, which causes the large differences of the correction factors, and hence yields large pull values. These differences between the generator and detector level of the two models, can only be explained by significantly different migrations into the phase space, which is for instance indicated by the purity (see uncertainty of the purity in figure 8.1). Such absolute corrections are not relevant for the unfolding, since it is based on relative numbers.
It must be mentioned, that the calculation of the pull values also include the statistical uncertainty. As it was outlined in the previous section, these uncertainties are smaller for the bin-by-bin method than for the regularized unfolding. The difference in the pull values, is therefore also an effect of the difference of the statistical uncertainties between the two correction methods.

8.5 Comparison of the correction methods

In this section we want to compare the cross sections, which are derived using the bin-by-bin correction method, to the ones obtained with the regularized unfolding method.

8.5.1 Comparison of (detector) correction factors

The detector correction factors for the regularized unfolding and the bin-by-bin method are displayed in figure 8.3. In this figure, the correction factor for the regularized unfolding is calculated through $c_i^{\text{unfold}} = N_i^{\text{unfolded}}/(N_i^{\text{Data}} - N_i^{\text{Bkg}})$, in correspondence to the bin-by-bin correction factor. The value N_i^{unfolded} is the number of events or jets at hadron-level after the unfolding.

One observes that the overall size of the correction is comparable between the two methods. However, the multiplicative correction factors can show some significant differences, which are up to a factor of 0.4. Especially in bins with high statistics, the correction factors differ often by about 0.1, which leads to a difference in the measured cross section of around 10 %. The trijet correction factors are quite similar for the two methods. For the inclusive jet and dijet correction factors one observes that the regularized unfolding gives a smaller correction than the bin-by-bin method. In the bins 6c of the inclusive jet and dijet measurements the effective correction factors differ significantly. It seems that the regularized unfolding has large fluctuations of the correction factor. This may arise from the limited statistics of the input data, but maybe also from the limited MC statistics in this phase space region.

The difference in the size of the correction is not surprising. The measurements are influenced by migrations from outside of the phase space region. In the regularized unfolding, these migrations are largely taken into account by the extended phase space in p_{T} and Q^2

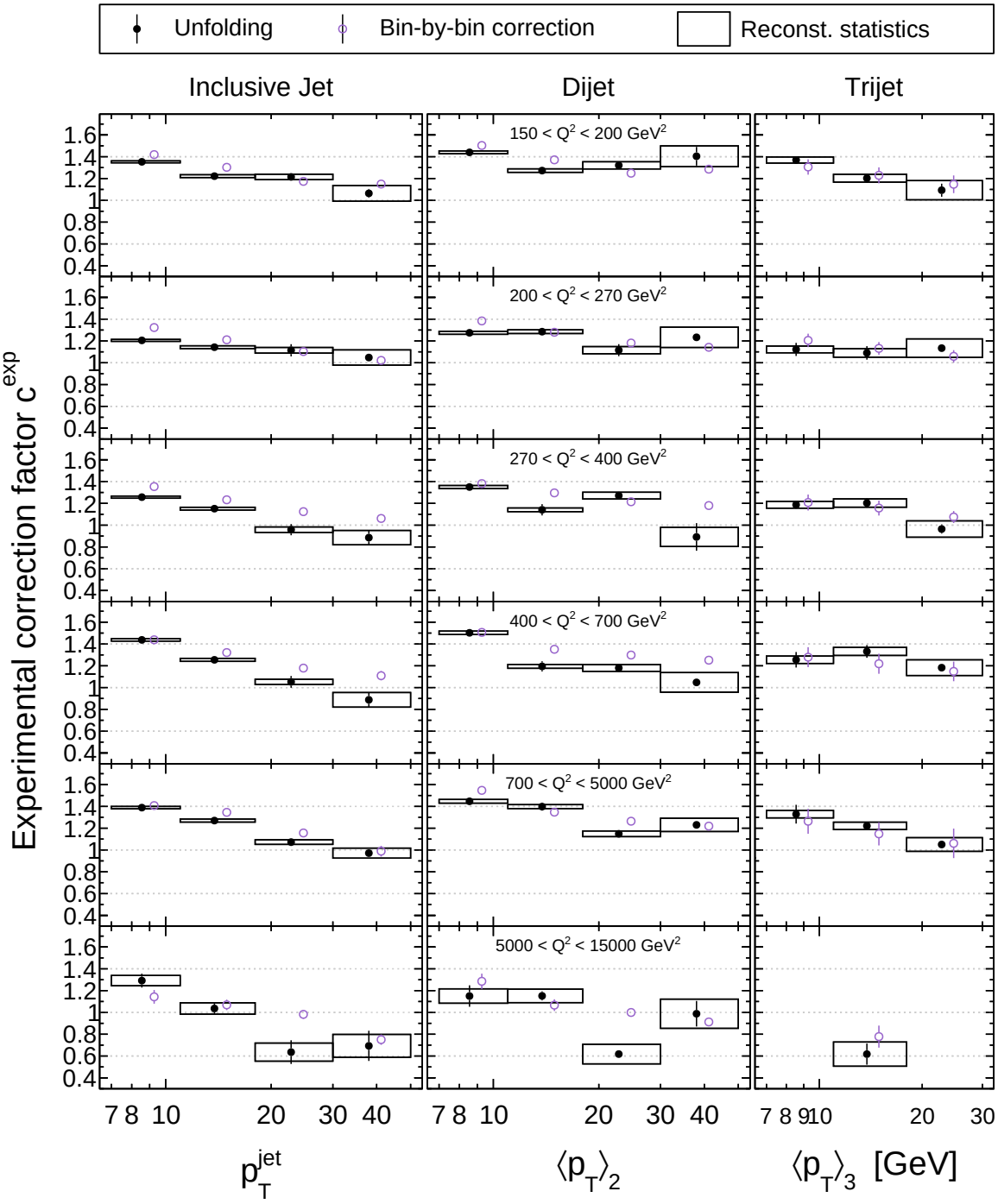


Figure 8.3: Experimental correction factors derived with the regularized unfolding method and with the bin-by-bin method. The error bars on the factors indicate the model uncertainty as derived for the corresponding method. The black boxes indicate the statistical uncertainty on detector level. This statistical precision is input to the regularized unfolding (the positive correlations between the single measurements and within the inclusive jets are also respected in the unfolding procedure). For better visibility the markers are slightly displaced horizontally w.r.t. each other.

(see chapter 6). Furthermore, the migrations between the phase space bins are modeled correctly.
In figure 8.3, we also show the estimated model uncertainties. These are interpreted as the uncertainty from the nescience of the true hadron level. One can recognize that this uncertainty may not be misinterpreted as the uncertainty of the correction method, since the two models are not compatible with each others within the model uncertainty (see error bars in figure 8.3).
Also, the statistical uncertainty on the detector level for each bin are shown in figure 8.3. Since the regularized unfolding method is a data driven method, and the covariance matrix enters the least square minimization, the correction factors cannot be determined with a higher precision than the statistical precision. The determined model uncertainty is therefore also partly caused by the limited statistics and fluctuations of the data.

8.5.2 Comparison of the statistical and model uncertainties

The size of the relative statistical uncertainties and the model uncertainties are compared for the cross sections obtained with the regularized unfolding and the bin-by-bin method in figure 8.4. The statistical uncertainties of the bin-by-bin corrected cross sections $\delta\sigma^{\mathrm{BBB}}$ are around a factor 2 smaller than for the unfolded cross sections $\delta\sigma^{\mathrm{unfold}}$. As argued in the previous section, the bin-by-bin uncertainties are not reliable, since this method does not treat migrations correctly and thus underestimates the statistical uncertainty in each single bin. On the other hand, the statistical uncertainties of the unfolded cross sections suggest a smaller statistical precision than they actually have. Due to the negative correlations, which are caused by migrations and which are mostly present between neighboring bins, the unfolded data has a larger precision when all bins of a single measurement are used together, i.e. in a fit. For this purpose, the correlation matrix of the unfolded data was shown in figure 7.5. The bin-by-bin corrected cross sections have a priori no correlations between bins.
The estimated model uncertainty is roughly comparable for the two correction methods. For the inclusive jet measurement, the model uncertainty of the unfolding is larger by about 3 to 7 % in around half of the number of bins, which are mostly high p_{T} and high Q^2 bins. For the trijet measurement, the model dependence is reduced for the unfolded cross sections in most of the bins. For the dijet measurement, the difference of the model dependence between the correction methods is negligible in most of the bins. Overall, no systematic change of the model dependence in single phase space regions is observed.

8.5.3 Data shift

The effect on the cross section from the different correction methods can be estimated from figure 8.3. The pull values, as defined in equation 8.17, give a more quantitative comparison of the results. The pull values for all bins of the three measurements are shown in figure 8.5. Two definitions for the pulls are conceivable: the true distribution is assumed to be given by the unfolded data or by the bin-by-bin corrected data. The difference is then in the sign of the pull value p_i, but also in the value of the statistical uncertainty Δx_i. The obvious difference in the size of the pull values is explained by the underestimated statistical uncertainty in the bin-by-bin correction, which increases the size of the corresponding pull values (right pad in figure 8.5). The arithmetic means of the pull values of the bin-by-bin corrected result vs. the unfolded result $\langle p_{\mathrm{BBB}} \rangle$ are 1.15 to 2.05 for

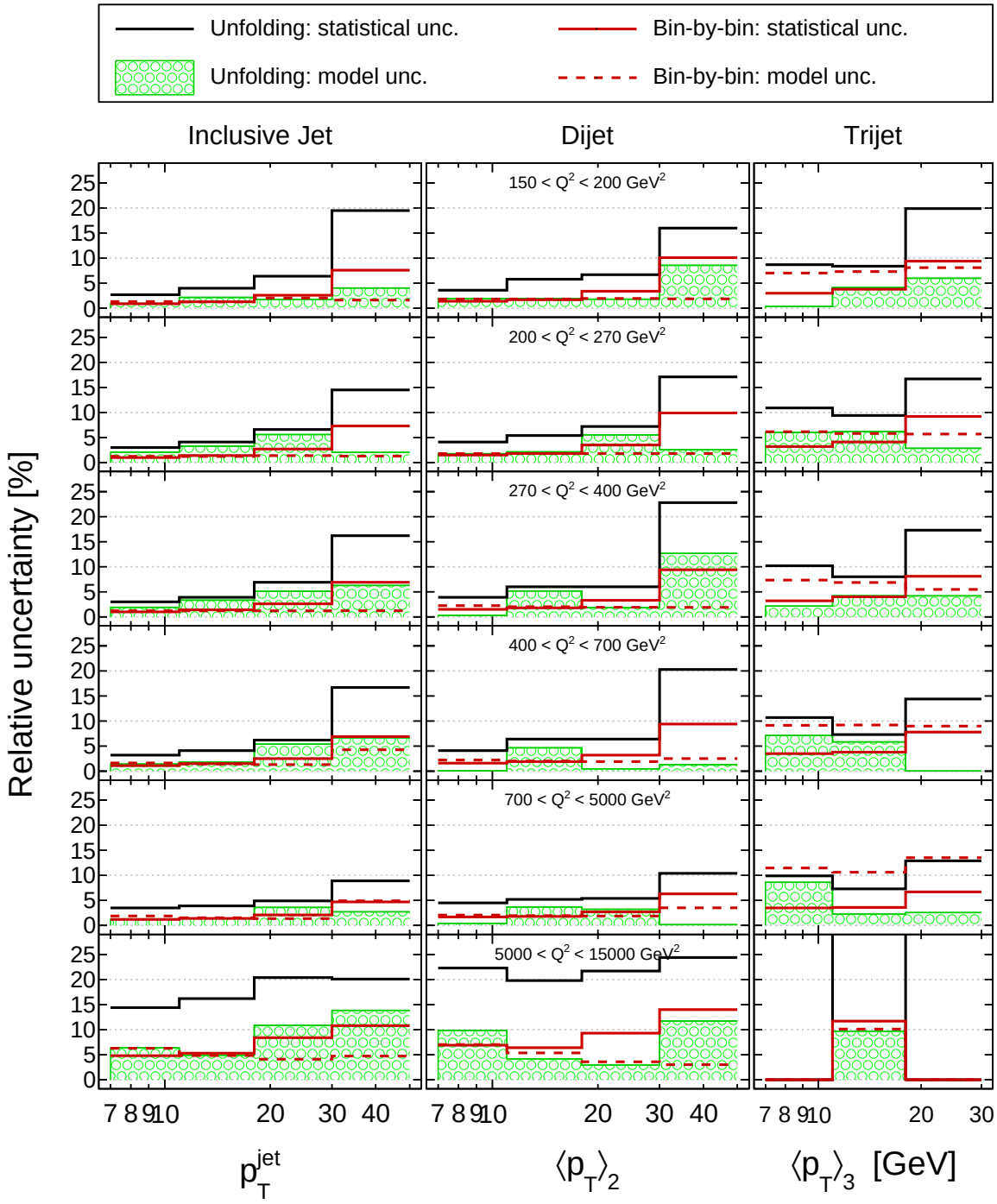


Figure 8.4: The statistical uncertainty and the model uncertainty for the unfolded and the bin-by-bin corrected cross sections. The statistical uncertainty of the bin-by-bin corrected cross sections is around a factor of 2 smaller than the unfolded precision. However, the bin-by-bin corrected results are biased, and the statistical uncertainties are not reliable (c.f. section 8.3.2). Furthermore, the statistical uncertainty of the unfolded cross sections have large negative correlations (see correlation matrix in figure 7.5), and therefore the statistical precision of a single measurement is better than suggested by the statistical uncertainty only.

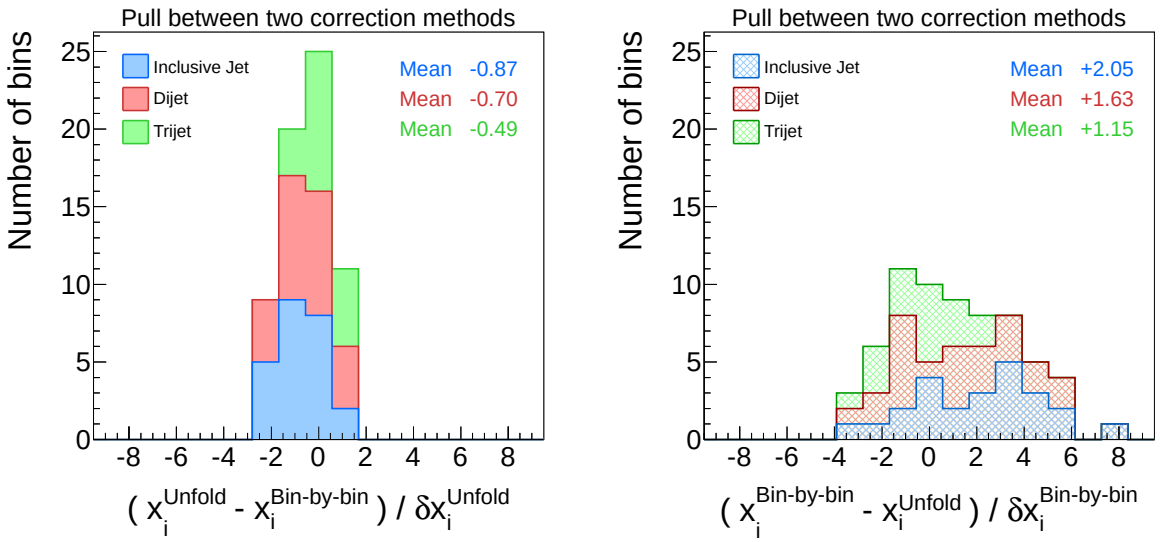


Figure 8.5: Pull values between the results of the two correction methods for all bins of the three measurements. The pull values are color coded to distinguish bins in different measurements. The left plot assumes the bin-by-bin corrected data as the 'true' distribution. The right plot assumes the unfolded data as the 'true' distribution. The large difference between the two pull distributions arises from the different determination of the statistical uncertainty Δx.

the three different measurements, while $\langle p_{\text{unfold}} \rangle$ is only -0.87, -0.70 and -0.49 for the inclusive jet, dijet and trijet bins, respectively.
When assuming the unfolded result is the true hadron level distribution of the data, the right pad in figure 8.5 shows the size of the bias of the bin-by-bin method on data.
We observe that the unfolded cross sections are on average smaller for the regularized unfolding than for the bin-by-bin method. The difference of the statistical uncertainties can roughly be estimated by $\delta(\Delta x) \approx -\langle p_{\text{BBB}} \rangle / \langle p_{\text{Unfold}} \rangle$. Hence, the bin-by-bin method underestimates the statistical uncertainty by more than a factor of 2 for the present measurements.

8.5.4 The total cross section and its uncertainty

A measure of the overall precision on the double-differential measurement can be indicated by the *total cross section* σ_{tot} and its uncertainty. The total cross section is defined as the sum of all double-differential phase space bins

$$\sigma_{\text{tot}} = \sum_{i}^{\text{bins}} \sigma_i. \tag{8.18}$$

The error propagation is performed, following equation 7.12. This procedure is fully correct for the unfolded cross sections $\sigma_{\text{tot}}^{\text{unfold}}$, where all (negative) correlations are taken into account. However, in the philosophy of correction factors, the total cross section $\sigma_{\text{tot}}^{\text{BBB,0-dim}}$ would be determined from one single bin with one correction factor $c_{\text{tot}}^{\text{exp}}$. The total cross section following equation 8.18 will be labeled $\sigma_{\text{tot}}^{\text{BBB,2-dim}}$, for comparison. In table 8.2 the three cross section values and their statistical uncertainties are compared. The total cross section of the bin-by-bin method determines a larger cross section, which was already noted in the previous sections. The statistical uncertainties of the two bin-by-bin cross sections differ slightly. The bin-by-bin corrected cross sections suggest a significantly better

Total cross section

	Inclusive Jet	**Dijet**	**Trijet**
Unfolding $\sigma_{\text{tot}}^{\text{unfold}}$ [pb]	$455.48 \pm 0.82\%$	$194.34 \pm 1.04\%$	$33.13 \pm 3.51\%$
Propagated bin-by-bin $\sigma_{\text{tot}}^{\text{BBB,2-dim}}$ [pb]	$480.54 \pm 0.39\%$	$182.19 \pm 0.48\%$	$33.82 \pm 1.06\%$
Genuine bin-by-bin $\sigma_{\text{tot}}^{\text{BBB,0-dim}}$ [pb]	$481.17 \pm 0.39\%$	$207.85 \pm 0.53\%$	$36.02 \pm 1.08\%$

Table 8.2: Total cross section of the inclusive jet, dijet and trijet measurement. The total cross section of the bin-by-bin method is shown for comparison as genuine total cross section and compared to the double-differentially summed cross section. The bin-by-bin method suggests a largely higher precision than actually given in these measurements.

statistical precision, by more than a factor of 2 compared to the unfolded results. The statistical uncertainty of the trijet measurement is even underestimated by a factor of 3.5. This is because, the regularized unfolding method includes the migrations to the outside of the phase space region correctly. If also this phase space region could be measured, the statistical precision would increase significantly, because of the negative correlations to these bins (c.f. the example in section 8.3.2).

Chapter 9

Study of the unfolding scheme

The unfolding scheme employed in this work was outlined in chapter 6. It could not be developed from first principles, but is the result of iterative testing, trials and improvements. The degrees of freedom for this study are the bin grid, the observables, and the definitions of side-bins. The unfolding scheme was optimized using Monte Carlo studies and was judged on the basis of a χ^2-method.
In the second part of this chapter, the dependence of the unfolded result on the regularization strength and the regularization method is studied.

9.1 The χ^2 method

The task of the unfolding is to find the unknown hadron level distribution of the data. The unfolded result shows typically some dependence on the Monte Carlo generator employed. Due to the limited data statistics and the increase in computing time for additional multi-dimensional observables, the unfolding can only be performed in a few observables at a time. We will further observe in this study that variables which are integrated over the measurement bin grid do not contribute to the model dependence.

In this study we want to optimize the unfolding such that it is as independent as possible of the model used. We investigate the model dependence of an unfolding scheme by calculating χ^2/n_{dof} using the phase space bins and Monte Carlo pseudo-data. We define the χ^2-value through

$$\chi^2_{d(m)}/n_{\mathrm{dof}} = (x_{\mathrm{unfold}} - x_{\mathrm{true}})^{\mathrm{T}} \boldsymbol{V}_x^{-1} (x_{\mathrm{unfold}} - x_{\mathrm{true}})/n_{\mathrm{dof}} \ , \tag{9.1}$$

where x_{unfold} denotes the vector of the unfolded bin values in the measurement phase space (MPS) and x_{true} is the true hadron level vector of the Monte Carlo pseudo-data. The indices d and m denote the Monte Carlo generator used for the pseudo-data at detector level and the MC generator used to determine the migration histogram, respectively. The covariance matrix $\boldsymbol{V}_x$ includes the propagated statistical uncertainty including correlations. The χ^2-value is normalized to the number of bins n_{dof}. A small value of $\chi^2_{d(m)}/n_{\mathrm{dof}} \approx 1$ indicates that the unfolded result is compatible with its true hadron level within the statistical precision.

The χ^2/n_{dof} value can be determined individually for each measurement of an observable, i.e. the inclusive jet $p_{\mathrm{T}}^{\mathrm{jet}}$, dijet or trijet measurement of $\langle p_{\mathrm{T}} \rangle$, and for all three jet measurements at once. If calculated for multiple observables, the positive correlations between

them are taken into account in the covariance matrix. The advantage of the χ^2 calculus over pull values (c.f. equation 8.17), are the inclusion of correlations, i.e. the bin-wise pull values become small, if the statistical uncertainty is large. In the unfolding procedure, the covariances are only becoming large, if the variances are largely negative.

Using the two Monte Carlo generators Django and Rapgap, we can determine four χ^2/n_{dof} values:

- **Django - Django:** $\chi^2_{\mathrm{Dj(Dj)}}/n_{\mathrm{dof}}$
  The pseudo-data consist of 10 % of the Django event sample, and the migration histogram is determined with the remaining, statistically independent, 90 % of the Django event sample. The covariance matrix $\boldsymbol{V}_y$ is calculated from the correlation matrix $\boldsymbol{P}_y^{\mathrm{Data}}$ using real data, but with the statistical uncertainty of the Django pseudo data.

- **Rapgap - Rapgap:** $\chi^2_{\mathrm{Rg(Rg)}}/n_{\mathrm{dof}}$
  The same procedure as for 'Django - Django', but the pseudo-data and the migration histogram are determined from the Rapgap event sample.

- **Django - Rapgap:** $\chi^2_{\mathrm{Dj(Rg)}}/n_{\mathrm{dof}}$
  The pseudo data consist of 10 % of the Django event sample, but the statistical precision is rescaled to the real data statistics. The migration histogram is determined from the full Rapgap event sample.

- **Rapgap - Django:** $\chi^2_{\mathrm{Rg(Dj)}}/n_{\mathrm{dof}}$
  The same procedure as for 'Django-Rapgap', but with the role of the two generators exchanged.

All calculations are performed to obtain the on non-radiative hadron level just as for real data. Therefore, the correction factor c_{rad} from the model, which is used for the determination of the migration matrix is applied, whenever necessary.
The χ^2/n_{dof}-values for pseudo-data which are unfolded with the same model, i.e. $d = m$, have to fulfill $\chi^2/n_{\mathrm{dof}} \lesssim 1$. The values are expected to be smaller than unity, since the two vectors x_{unfold} and x_{true} are from the same data set and thus are not statistically independent. However, the value must deviate from 0, since the pseudo-data are statistically independent of the migration histogram.

The values of $\chi^2_{d(m)}/n_{\mathrm{dof}}$ for the employed unfolding after the optimization of the unfolding scheme are shown in table 9.1. The χ^2/n_{dof}-values, when a different model is employed for the migration histogram than for the pseudo-data, are between 1.2 and 2.1. These values indicate a deviation from the true hadron level which is larger than the statistical fluctuations, and hence a small model dependence is still present (see figure 8.4 for the size of the model uncertainty).
For the study of finding the optimal unfolding scheme, we define a characteristic χ^2/n_{dof}-value for each unfolding scheme through

$$\chi^2/n_{\mathrm{dof}} = \frac{\chi^2_{\mathrm{Dj(Rg)}} + \chi^2_{\mathrm{Rg(Dj)}}}{2n_{\mathrm{dof}}} \ . \tag{9.2}$$

Since the two input values may differ, we quote the half-difference between $\chi^2_{\mathrm{Rg(Dj)}}/n_{\mathrm{dof}}$ and $\chi^2_{\mathrm{Dj(Rg)}}/n_{\mathrm{dof}}$ as uncertainty, which we regard as statistical fluctuation. Whenever,

χ^2/n_{dof} **values**

	Inclusive jet $\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}}$	**Dijet** $\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}\langle p_{\mathrm{T}}\rangle_2}$	**Trijet** $\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}\langle p_{\mathrm{T}}\rangle_3}$	**All**	**Dijet** $\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}\xi}$	**Trijet** $\frac{\mathrm{d}\sigma}{\mathrm{d}Q^2\mathrm{d}\xi}$
n_{dof}	24	24	18	66	21	14
$\chi^2_{\text{Dj(Dj)}}/n_{\text{dof}}$	0.445	0.499	1.009	0.668	0.429	0.818
$\chi^2_{\text{Rg(Rg)}}/n_{\text{dof}}$	1.007	0.855	0.630	0.857	0.208	0.811
$\chi^2_{\text{Dj(Rg)}}/n_{\text{dof}}$	1.698	1.978	1.214	1.487	2.564	2.883
$\chi^2_{\text{Rg(Dj)}}/n_{\text{dof}}$	2.130	2.139	1.822	1.817	2.339	0.746

Table 9.1: Values for χ^2/n_{dof} for the final unfolding scheme employed, using statistically independent Monte Carlo pseudo-data. The calculation of the χ^2/n_{dof}-values includes all correlations. The values are calculated for the double-differential inclusive jet, dijet and trijet cross sections, as well as for all three measurements together (denoted "All"). When pseudo-data are unfolded with the migration histogram of the same model, the value is expected to be smaller than unity, since the true distribution and the unfolded distribution are not statistically independent. However, the value must be larger than zero, since the unfolded distribution is determined from an statistically independent event sample.

this uncertainty is getting large, we consider the unfolding as *not stable*. This may for example result from poor statistics in some phase space regions in a highly multidimensional unfolding.

9.2 Study of the bin grid definition

In this section the bin grid definitions of the inclusive jet, dijet and trijet unfolding histograms are studied in detail. The basic structure of the combined unfolding of all measurements has been introduced in chapter 6. A schematic illustration is shown in figure 6.1. The three jet measurements have always connections through detector-level-only entries with the NC DIS measurement. Therefore, the generator level bin grid for the NC DIS observervables y and Q^2 are identically defined for all four measurements. For simplicity and technical reasons, the same holds for the detector level bin grids for y and Q^2.

9.2.1 Technical concepts for a multidimensional unfolding

Migration matrices, which represent migrations in several kinematic observables, can be realized by two technical concepts.

- **Multidimensional unfolding**
  The *multidimensional* unfolding is a straight forward realization of an unfolding in multiple variables. In each observable, where migrations are considered, a generator level and detector level bin grid is defined. In order to have an over-constrained problem, the number of bins on generator level n_{gen} and on detector level n_{rec} should follow the rule of thumb
  $$n_{\text{rec}} \gtrsim 1.5\, n_{\text{gen}} \tag{9.3}$$
  for every individual variable. The multidimensional unfolding is then defined by a migration scheme where in each of the $n_{\mathcal{I}}^{\text{gen}} \times n_{\mathcal{I}}^{\text{rec}}$ bins of the observable $\mathcal{I}$ a migration matrix of the observable $\mathcal{J}$ is defined. The multidimensional bin grid in k observables is then given by $n_{\text{gen}} = \prod_k n_k^{\text{gen}}$ and $n_{\text{rec}} = \prod_k n_k^{\text{rec}}$ bins on generator level and detector

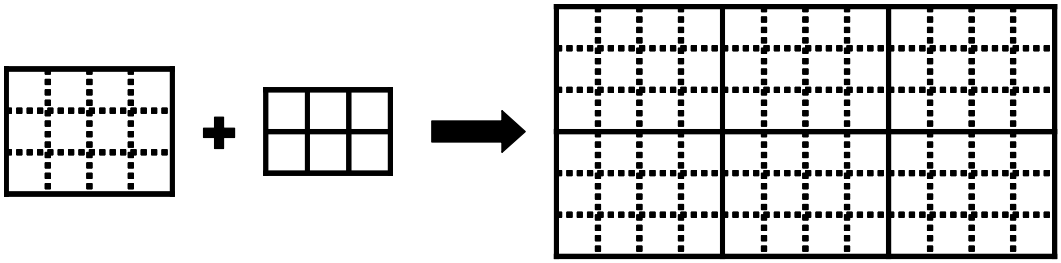


Figure 9.1: Schematic illustration of a (full) two-dimensional unfolding matrix. The migration scheme of one observable (with e.g. a 3×2 bins) is inserted in every bin of a matrix of another variable (with e.g. 4×3 bins).

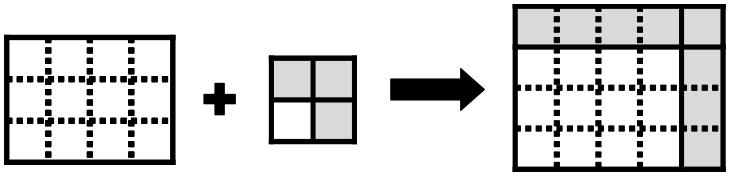


Figure 9.2: Schematic illustration of the concept of 'additional side-bins'. A migration matrix (with e.g. 4×3 bins) is extended with additional side-bins, representing for instance events outside of the measurement phase space (illustrated through a grey shaded pattern). These side-bins typically represent variables which define the phase phase.

level, respectively. Thereby, the order of the dimensions is irrelevant. A schematic illustration of this method is given in figure 9.1. The limiting factor of this method is the heavily increasing number of bins, with each additional observable. Too many bins results in poor statistics of the data, but also the MC statistics may become relevant. Furthermore, limitations on the computing power and memory may have to be considered. For instance, the $n_{\text{rec}} \times n_{\text{rec}}$ covariance matrix can become very large with consequences for the numerical inversion of the matrix.

- **Additional side-bins**
  Migrations in multiple variables can also be considered by extending the migration scheme with one or more entries. A schematic illustration is given in figure 9.2. For instance, the $n_{\mathcal{I}}^{\text{gen}} \times n_{\mathcal{I}}^{\text{rec}}$ migration scheme of an observable $\mathcal{I}$ may be extended with additional bins on generator and/or on detector level to a $(n_{\mathcal{I}}^{\text{gen}} + m_{\mathcal{J}}^{\text{gen}}) \times (n_{\mathcal{I}}^{\text{rec}} + m_{\mathcal{J}}^{\text{rec}})$ migration scheme. These bins may represent events, which are outside of the phase space of the initial event selection.

Several caveats must be mentioned for the concept of additional side-bins. The additional m bins for an observable $\mathcal{J}$ do not measure the full $\mathcal{I}$-distribution on detector level. Therefore, the migrations from an m-bin on detector level into the $\mathcal{I}$ phase space on generator level are under-determined. This may become relevant, if e.g. the $\mathcal{I}$-distribution is essentially differently in an m bin[1]. Therefore, we always include multiple bins, which represent several bins in $\mathcal{I}$, for each variable $\mathcal{J}$.

[1]Consider the following example: A dijet migration scheme, where migrations in $\langle p_{\text{T}} \rangle$ are considered in a $n_{\langle p_{\text{T}} \rangle}^{\text{rec}} \times n_{\langle p_{\text{T}} \rangle}^{\text{gen}}$ bin grid. The dijet phase space is also constrained by a cut on $p_T^{\text{jet2}} > 5\,\text{GeV}$, and the resulting migrations should also be considered. The migration scheme is then extended with one additional bin to $(n_{\langle p_{\text{T}} \rangle}^{\text{rec}} + 1) \times (n_{\langle p_{\text{T}} \rangle}^{\text{gen}} + 1)$, which takes into account dijet events where $p_T^{\text{jet2}} < 5\,\text{GeV}$. The $\langle p_{\text{T}} \rangle$ distribution within the p_T^{jet2}-bin is integrated over, but has an essentially different spectrum than the $\langle p_{\text{T}} \rangle$ distribution in the "main" phase space. Hence, these migrations could be biased, when the model predicts different distributions in these side-bins.

Furthermore, the concept of additional side-bins also often violates the rule of thumb 9.3, since the migrations in variable $\mathcal{J}$ are basically represented by a 2×2 migration scheme, while one out of these four bins is measured differentially in $\mathcal{I}$[2].
The concept of additional side-bins is often used for quantities, which are added solely on generator level (i.e. a $n_{\mathcal{I}}^{\text{rec}} \times (n_{\mathcal{I}}^{\text{gen}}+1)$ migration scheme). Such an implementation is circumvented in this analysis by fixing the detector-level-only entries, using the NC DIS events. However, such unfolding schemes are frequently used in other analyses (e.g. [209, 210]). The migrations are then predominantly determined from the Monte Carlo predictions. This may result in large positively correlated statistical uncertainties.

In this work, both concepts for a multidimensional unfolding are combined. The full multidimensional unfolding is predominantly employed for variables, where migrations are large and which are also measured differentially. The concept of additional side-bins is employed for migrations in quantities, which define the measurement phase space, like M_{12} and p_T^{jet2} in case of the dijet measurement, or $\eta_{\text{lab}}^{\text{jet}}$ in case of the inclusive jet unfolding.

9.2.2 Study of the inclusive jet unfolding scheme

The unfolding matrix of the inclusive jet measurement refers to the matrix $\boldsymbol{J}_1$ as illustrated in figure 6.1. The detector level bin grid is also used for the matrix $\boldsymbol{B}_1$.

In order to find the optimal unfolding scheme, i.e. the one which results in the smallest χ^2/n_{dof}-value, several unfolding schemes have been tested. The unfolding schemes differ in the number of dimensions and variables considered, in the number of bins in each variable and in their implementation. An overview of the relevant unfolding trials are shown in table 9.3 and the notation is explained in table 9.2. The different unfolding schemes are assigned a short label, e.g. $vI4{+}\text{F}_{gr}$ refers to the finally employed unfolding, as outlined in chapter 6.
The unfolding schemes considered are three or four-dimensional multidimensional unfolding schemes, where also additional side-bins in $\eta_{\text{lab}}^{\text{jet}}$ may be considered.
The starting point for this study is the "simplest" three-dimensional unfolding, in y, Q^2 and $p_{\text{T}}^{\text{jet}}$ labeled $vI3$. All three variables are extended with side-bins on both sides for low and high values. On generator level the bin grid is chosen such, that the 24 bins of the measurement phase space are directly determined. A value of $\chi^2/n_{\text{dof}} = 3.73$ is found.
The χ^2/n_{dof}-values for all unfolding schemes are illustrated in figure 9.3. With this figure we can draw the following conclusions.

When migrations in $\eta_{\text{lab}}^{\text{jet}}$ are included as a full additional dimension, no significant improvement is observed (compare $vI4$a and $vI4$b vs. $vI3$). The versions $vI4$a and $vI4$b differ only in the number of $\eta_{\text{lab}}^{\text{jet}}$ bins on detector level. On generator level, there are always two side-bins and one bin representing the measurement phase space in $-2.5 < \eta_{\text{lab}}^{\text{jet}} < 1.0$. We further observe that the unfolding with 5 $\eta_{\text{lab}}^{\text{jet}}$ bins is becoming unstable, which may arise from the small statistics in the $\eta_{\text{lab}}^{\text{jet}}$ side-bins, since these include further the three-dimensional bin grid in Q^2, $p_{\text{T}}$and y.

[2]If the migration matrix in $\mathcal{I}$ and $\mathcal{J}$ are part of a full multidimensional unfolding scheme, the rule of thumb 9.3 is typically fulfilled. However, the migrations in variable $\mathcal{J}$ are somewhat under-determined.

Example element

Dimension
$\mathcal{I}+\mathcal{J}$
$n^{\mathcal{I}}_{\text{rec}}+n^{\mathcal{J}}_{\text{rec}} \rightarrow n^{\mathcal{I}}_{\text{gen}}\ [n_{\text{MPS}}(n_{\text{Pub}})]+n^{\mathcal{J}}_{\text{gen}}$

Notation

Var.	Description
$\mathcal{I}$	Observable where multidimensional migrations are considered
$\mathcal{J}$	(Optional) observable(s) where additional side-bins are added to $\mathcal{I}$
n_{rec}	Number of bins on detector level
n_{gen}	Number of bins on generator level
n_{MPS}	Number of bins (out of n_{gen}) in measurement phase space (MPS)
n_{Pub}	Number of bins (out of n_{MPS}) which are published

Table 9.2: Legend for the tables and figures of various unfolding schemes. The number of bins in the measurement phase space on generator level n_{MPS} (quoted in square brackets), may be combined in order to determine the final number of bins in that dimension n_{Pub} (quoted in curved brackets). The two concepts of multidimensional unfolding are denoted differently: as individual columns *Dimension* for a full *multidimensional unfolding* in the variable $\mathcal{I}$ or as $\mathcal{I}+\mathcal{J}+\ldots$ if the concept of *additional side-bins* is used. In the latter case the number of bins for each dimension are quoted. Only migrations in and out of side-bins are implemented as additional side-bins. The numbers n_{MPS} and n_{Pub} are therefore attached to the initial observable $\mathcal{I}$.

The unfolding scheme $vI3.5+$ includes $\eta^{\text{jet}}_{\text{lab}}$ side-bins via *additional side-bins* in Q^2. In $vI3.5+$ six additional bins are defined, which represent jets with $\eta^{\text{jet}}_{\text{lab}} < -1.0$ and $\eta^{\text{jet}}_{\text{lab}} > 2.5$ for each of three Q^2 regions (denoted as $Q^2(\eta^-)$ and $Q^2(\eta^+)$, respectively, in table 9.3). Migrations within the $\eta^{\text{jet}}_{\text{lab}}$ phase space $-1.0 < \eta^{\text{jet}}_{\text{lab}} < 2.5$ are represented by a 2×1 migration scheme[3] and is denoted by $\eta^{\text{MPS}}_{\text{lab}}$.
We observe no significant change of $vI3.5+$ compared to $vI3$, although additional migrations are now respected.

We use $vI3.5+$ as the scheme for other studies.
The version $vI3.5+$(JM) uses the identical unfolding scheme as $vI3.5+$. The difference is a jet-matching procedure, where a distance measure of $\Delta p^{\text{jet}}_{\text{T}} := |p^{\text{gen}}_T - p^{\text{rec}}_T|$ is employed for the closest-pair algorithm. We observe no significant difference to $vI3.5+$ and conclude that the geometrical measure ΔR (see section 6.6.1), which is unbiased in the kinematical observables, is preferred.

The version $vI3.5+$(LJ) uses the identical unfolding scheme as $vI3.5+$, but the jet-matching is performed in ΔR_{lab} instead of ΔR_{Breit}. This reduces the sensitivity to wrongly reconstructed boost variables [79] (see section 6.6.1). Although,the improvement is not significant, we apply this procedure for all following unfolding schemes.

In version $vI3.5+\text{f}_r$ a finer bin grid with more bins on detector level in $p^{\text{jet}}_{\text{T}}$ is used compared to $vI3.5+$. A significant improvement could be observed for $\chi^2_{\text{Dj(Rg)}}$, but not for $\chi^2_{\text{Rg(Dj)}}$. Since the generator level bin grid has not changed compared to $vI3.5+$, we conclude that the relevant migrations take place, not surprisingly, in $p^{\text{jet}}_{\text{T}}$ but that they cannot be un-

[3]Since this $\eta^{\text{jet}}_{\text{lab}}$ migration scheme does not have the necessary degree of freedom on generator level, we refer to it as a '3.5' instead of 4 dimensional unfolding.

Inclusive jet unfolding schemes

Label	Dim. 1	Dimension 2	Dimension 3	Dim. 4
$vI3$	y	Q^2	$p_{\mathrm{T}}^{\mathrm{jet}}$	
[211]	$3 \to 2$ [1(1)]	$15 \to 8$ [6(6)]	$10 \to 6$ [4(4)]	
$vI4$a	y	Q^2	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15 \to 8$ [6(6)]	$10 \to 6$ [4(4)]	$4 \to 3$ [1(1)]
$vI4$b	y	Q^2	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15 \to 8$ [6(6)]	$10 \to 6$ [4(4)]	$5 \to 3$ [1(1)]
$vI3.5+$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$10 \to 6$ [4(4)]	$2 \to 1$ [1(1)]
$vI3.5+$(JM)	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$10 \to 6$ [4(4)]	$2 \to 1$ [1(1)]
$vI3.5+$(LJ)	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$10 \to 6$ [4(4)]	$2 \to 1$ [1(1)]
$vI3.5+\mathrm{f}_r$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$13 \to 6$ [4(4)]	$2 \to 1$ [1(1)]
$vI4+$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$10 \to 6$ [4(4)]	$3 \to 2$ [2(1)]
$vI4_H+$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$p_z^{\mathrm{HFS(LAr)}}$
	$3 \to 2$ [1(1)]	$15+3+3 \to 8$ [6(6)] $+3+3$	$10 \to 6$ [4(4)]	$3 \to 2$ [2(1)]
$vI4_H+\mathrm{f}_{gr}$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$p_z^{\mathrm{HFS(LAr)}}$
	$3 \to 2$ [1(1)]	$13+3+3 \to 8$ [6(6)] $+2+2$	$14 \to 8$ [5(4)]	$3 \to 2$ [2(1)]
$vI4+\mathrm{f}_{gr}$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
	$3 \to 2$ [1(1)]	$13+3+3 \to 8$ [6(6)] $+2+2$	$13 \to 8$ [5(4)]	$3 \to 2$ [2(1)]
$vI4+\mathrm{F}_{gr}$	y	$Q^2 + Q^2(\eta^-) + Q^2(\eta^+)$	$p_{\mathrm{T}}^{\mathrm{jet}}$	$\eta_{\mathrm{lab}}^{\mathrm{jet}}$
$= v13$	$3 \to 2$ [1(1)]	$14+3+3 \to 8$ [6(6)] $+2+2$	$16 \to 10$ [7(4)]	$3 \to 2$ [2(1)]

Table 9.3: Various unfolding schemes for the inclusive jet unfolding. All schemes are set up as full three or four-dimensional unfolding schemes. Additional side-bins may be included in the 'Q^2'-dimension. The left column indicates a unique name for easy reference. The syntax of the elements in this table is explained in table 9.2. The relevant differences of the schemes are discussed in the text.

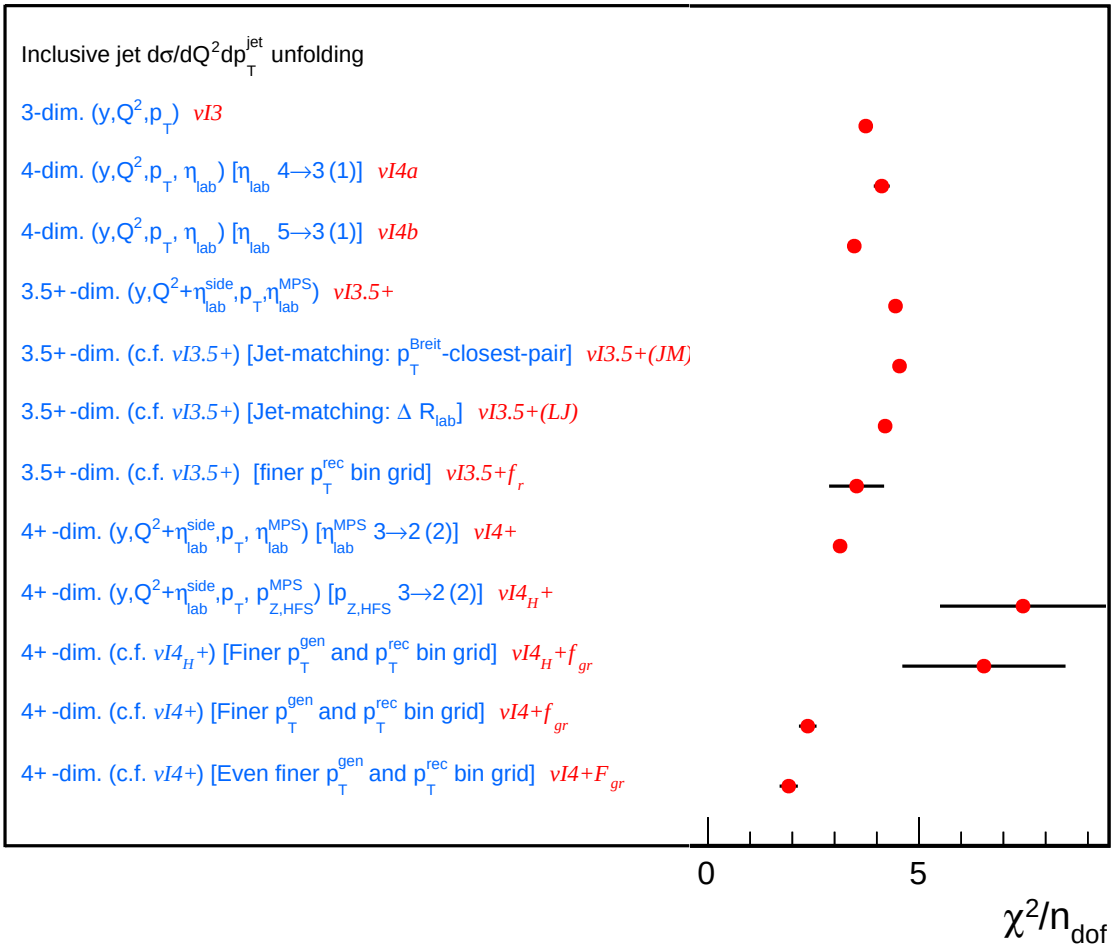


Figure 9.3: Values of χ^2/n_{dof} (c.f. equation 9.2) for various unfolding schemes of the inclusive jet measurement, determined from two different Monte Carlo generators. The χ^2-values are calculated for the 24 phase space bins, including the information of the covariance matrix. More details on the various unfolding schemes are outlined in the text and in table 9.3. The last version $vI4{+}\mathrm{F}_{gr}$ is employed for the final measurement.

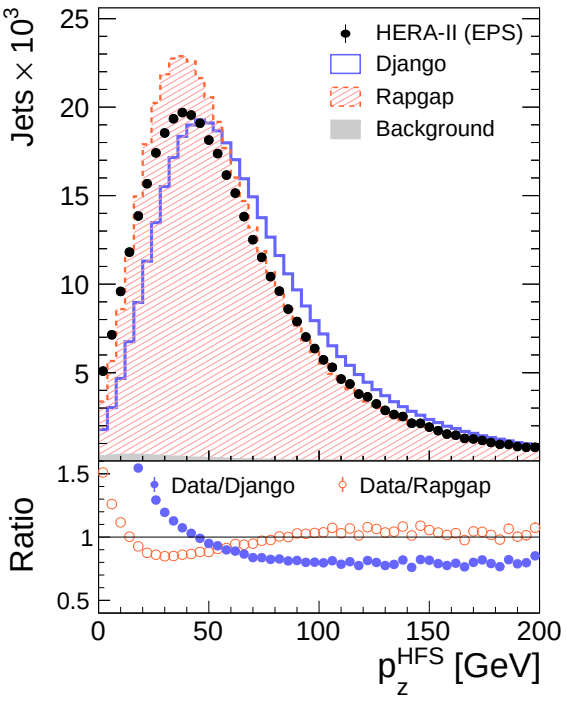


Figure 9.4: Distributions of data of the longitudinal momentum sum of the HFS within the acceptance of the LAr for each inclusive jet in an event compared to model predictions. The two models and the data differ in some phase space regions.

folded properly, since the $p_{\mathrm{T}}^{\mathrm{jet}}$ bin grid of the final measurement is also used on generator level.

The version $vI4+$ maps migrations within the $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ MPS from three detector level bins to two generator level bins. Migrations from the outside phase space regions in $\eta_{\mathrm{lab}}^{\mathrm{jet}}$, however, are still represented by additional side-bins[4] in the dimension Q^2. The performance of version $vI4+$ is similar to $vI4$b, since to some extend migrations of the same observables are taken into account.

We use the version $vI4+$ to study the inclusion of another variables instead of $\eta_{\mathrm{lab}}^{\mathrm{jet}}$. We consider in $vI4_H+$ the sum of the transverse momenta of the HFS within the LAr acceptance $p_z^{\mathrm{HFS(LAr)}}$ as an additional variable, where migrations take place (see figure 9.4). Also on generator level, this observable is restricted to the θ-acceptance of the LAr. A 3×2 migration scheme is employed for this observable, and the two generator level bins are summed for determining the cross sections. We find that the χ^2/n_{dof}-values increase significantly and differ considerably for $\chi^2_{\mathrm{Dj(Rg)}}$ and $\chi^2_{\mathrm{Rg(Dj)}}$. We conclude that this observable is not suited for the unfolding. The large χ^2 persists when increasing the $p_{\mathrm{T}}^{\mathrm{jet}}$-bin grid (compare $vI4_H$+f$_{rg}$).

Two conclusions can be drawn:

1. A finer generator level bin-grid, which could be different from the measurement bin grid, opens the necessary freedom for an unbiased unfolding.

[4]Since $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ migrations within the phase space are represented by a full dimension, but migrations in $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ from the outside of the measurement phase space are included as additional side-bins in the 'Q^2'-dimension, in only one element of the $\eta_{\mathrm{lab}}^{\mathrm{jet}}$ matrix these additional bins are present.

2. The dominant migrations take place in $p_{\mathrm{T}}^{\mathrm{jet}}$.

As a consequence, a finer bin grid is employed on detector and on generator level in the unfolding scheme $vI4{+}\mathrm{f}_{gr}$. An even finer binning in $p_{\mathrm{T}}^{\mathrm{jet}}$ on both levels, i.e. version $vI4{+}\mathrm{F}_{gr}$, yield even smaller values of χ^2/n_{dof}. It is refrained from implementing an even finer binning, since this would be then too small compared to the detector resolution in $p_{\mathrm{T}}^{\mathrm{jet}}$ and it would also lead to too small statistics in many bins.
The unfolding scheme $vI4{+}\mathrm{F}_{rg}$ (also denoted $v13$) is used in this work for the determination of the inclusive jet 'cross sections. The exact definition of the bin grids are outlined in chapter 6.

9.2.3 Study of the dijet unfolding scheme

The unfolding matrix of the dijet measurement is the matrix $\boldsymbol{J}_2$ as illustrated in figure 6.1. Its detector level bin grid is also used for the matrix $\boldsymbol{B}_2$, which fixes events which migrate into the extended detector level phase space.

The unfolding schemes studied always take migrations in the kinematic variables y, Q^2 and $\langle p_{\mathrm{T}} \rangle$ into account. In contrast to the inclusive jet measurement, also migrations in variables which define the dijet phase space have to be considered. These are migrations near the phase space boundaries $M_{12} < 16\,\mathrm{GeV}$, $p_T^{\mathrm{jet2}} < 5\,\mathrm{GeV}$, $\eta_{\mathrm{lab}}^{\mathrm{jet}} < -1.0$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}} > 2.5$. Also the requirement of at least two jets $n_{\mathrm{jet}} \geq 2$ is studied. The migrations at the phase space boundary $p_{\mathrm{T}}^{\mathrm{jet}} = 50\,\mathrm{GeV}$ are considered to be small because of the decreasing statistics and are taken into account only indirectly through the matrix $\boldsymbol{B}_2$.

The unfolding schemes studied are briefly summarized in table 9.4 and their χ^2/n_{dof}-values on non-radiative hadron level are summarized in figure 9.5.

The initial unfolding for this study is labeled $v6$. It is a three dimensional unfolding scheme in y, Q^2 and $\langle p_{\mathrm{T}} \rangle$. The migrations at the dijet phase space boundary are added in the $\langle p_{\mathrm{T}} \rangle$-dimension via 10 additional side-bins on detector level and 5 side-bins on generator level, which furthermore distinguish between different $\langle p_{\mathrm{T}} \rangle$ events. Exemplary, one bin of the (y,Q^2)-matrix is shown in figure 9.6. The χ^2-values are $\chi^2_{\mathrm{Dj(Rg)}}/n_{\mathrm{dof}} = 4.92$ and $\chi^2_{\mathrm{Rg(Dj)}}/n_{\mathrm{dof}} = 5.243$, which indicates the need for improvements.

In version $v7$ we include additional side-bins for *mono-jet* events, i.e. events with exactly one jet in the extended phase space. These bins may represent migrations among the requirement of $n_{\mathrm{jet}} \geq 2$. Five bins for different $p_{\mathrm{T}}^{\mathrm{jet}}$ values are added on detector level and three bins on generator level. Eight bins represent the migrations at the other phase space boundaries.
The inclusion of monojet events reduces the χ^2/n_{dof} value. However, this idea is not pursued further for several reasons. The number of monojet events in the lowest $p_{\mathrm{T}}^{\mathrm{jet}}$ bin is about 6 times larger than the largest number of entries in the dijet phase space. This could reduce the numerical precision of the matrix inversion. Furthermore, most of these monojet events are at the lower phase space boundary at $p_{\mathrm{T}}^{\mathrm{jet}} \gtrsim 3.0\,\mathrm{GeV}$, where almost 50 % of these monojet events on detector level migrate into the selection from outside of the phase space region and have to be fixed by the NC DIS events. On the other hand, less than 1 % of the monojet events migrate into the generator level MPS of interest. The amount of migrations from measured dijet events into monojet events on generator level is

Dijet unfolding schemes

Label	Dim. 1	Dimension 2	Dimension 3	Dim. 4
$v6$ [211]	y $3 \to 2$ [1(1)]	Q^2 $15 \to 8$ [6(6)]	$\langle p_\mathrm{T} \rangle + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}} + \eta_\mathrm{lab}^{\mathrm{jet}})$ $10 + 10 \to 5$ [4(4)] + 5	
$v7$	y $3 \to 2$ [1(1)]	Q^2 $15 \to 8$ [6(6)]	$\langle p_\mathrm{T} \rangle + \mathrm{EXT}_{n_\mathrm{jet} \geq 2} + (n_\mathrm{jet} = 1)$ $10 + 8 + 5 \to 5$ [4(4)] + 4 + 3	
$v8$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $15 + n_\mathrm{ext} \to 8$ [6(6)] + n_ext	$\langle p_\mathrm{T} \rangle$ $10 \to 6$ [4(4)]	
$v10$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $15 + n_\mathrm{ext} \to 8$ [6(6)] + n_ext	$\langle p_\mathrm{T} \rangle$ $11 \to 7$ [4(4)]	ΔR_{12} $3 \to 2$ [2(1)]
$v11$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $13 + n_\mathrm{ext} \to 8$ [6(6)] + m_ext	$\langle p_\mathrm{T} \rangle$ $11 \to 7$ [4(4)]	M_{12} $3 \to 2$ [2(1)]
$v9$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $15 + n_\mathrm{ext} \to 8$ [6(6)] + n_ext	$\langle p_\mathrm{T} \rangle$ $15 \to 9$ [6(4)]	
$v12$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $13 + n_\mathrm{ext} \to 8$ [6(6)] + m_ext	$\langle p_\mathrm{T} \rangle$ $16 \to 10$ [7(4)]	
$v13$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_\mathrm{T}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $14 + n_\mathrm{ext} \to 8$ [6(6)] + m_ext	$\langle p_\mathrm{T} \rangle$ $18 \to 11$ [8(4)]	

Table 9.4: Various dijet unfolding schemes. The unfolding schemes are set up as three or four dimensional unfolding schemes, where migrations near dijet phase space cuts are taken into account as additional side-bins. The shorthand notations $n_\mathrm{ext} := (2 \times 2) + 1 + 1 = 6$ and $m_\mathrm{ext} = 1 + 1 + 1 + 1$ are used. The relevant differences are discussed in the text, and the χ^2/n_dof-values achieved are summarized in figure 9.5.

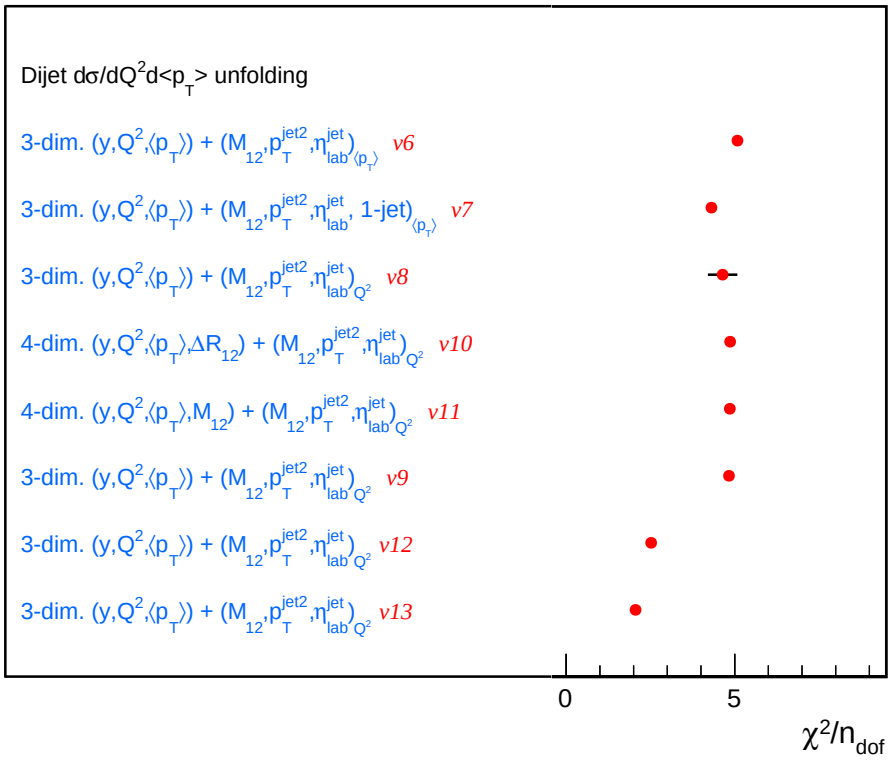


Figure 9.5: Values of χ^2/n_{dof} for various dijet unfolding schemes. All unfolding schemes are based on a three or four-dimensional implementation, where additional side-bins are included for quantities which define a 'dijet'-event. The differences of the various schemes are outlined in detail in table 9.4, and are discussed in the text. The different schemes are referred to using 'versions', which are shown as red labels.

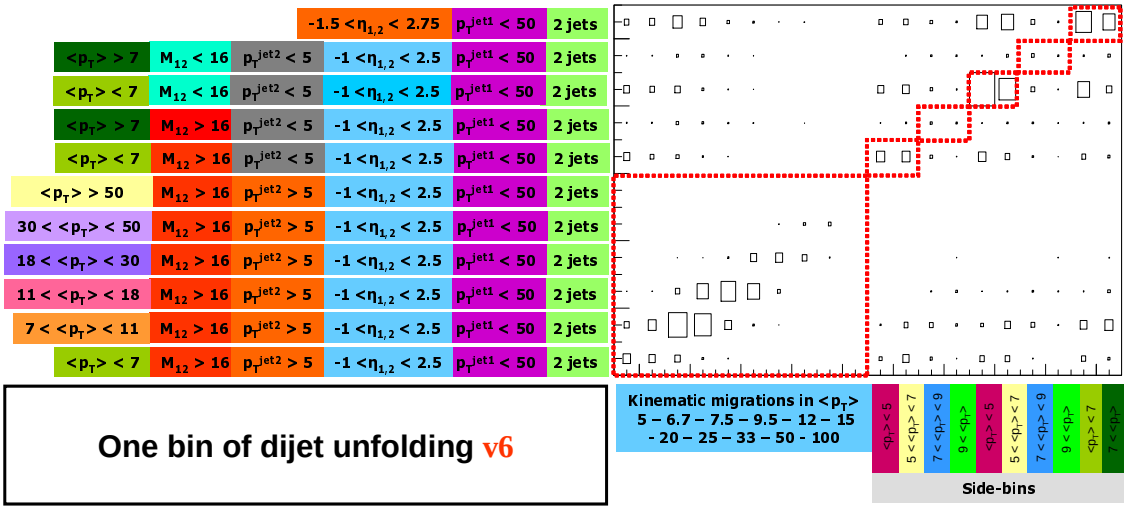


Figure 9.6: One bin (out of 720) of the dijet (y,Q^2) unfolding histogram of the unfolding scheme $v6$. The $\langle p_{\mathrm{T}}\rangle$ migration matrix is extended with additional side-bins, which define the dijet phase space. For each row on generator level all phase space cuts are listed. The red dashed rectangles indicate elements, where identical cuts are applied to the detector and generator level, but the differing $\langle p_{\mathrm{T}}\rangle$ are shown at the bottom of these columns.

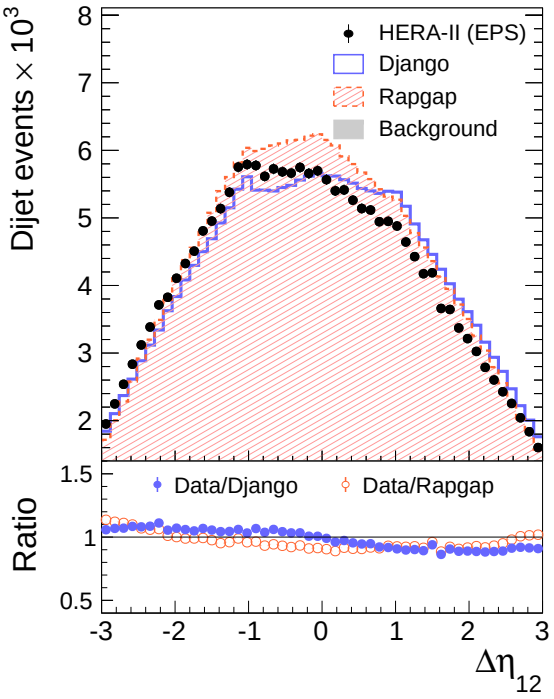


Figure 9.7: Distributions of data of the η-separation between the two leading jets in the dijet event sample compared to model predictions.

also below 1 %. Hence, the measured monojets are essentially estimated by the MC model, but contribute only insignificantly to the dijet measurement.

One disadvantage of $v6$ and $v7$ is obvious: The additional side-bins for the dijet phase space (M_{12}, $p_{\mathrm{T}}^{\mathrm{jet2}}$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$) are related to jet-quantities, and since these bins are added in the $\langle p_{\mathrm{T}} \rangle$-dimension they are differential in y and Q^2. The unfolding scheme $v8$ instead includes such migrations into the Q^2-dimension of the three-dimensional unfolding scheme. Thus, these additional bins are now fully differential in the $\langle p_{\mathrm{T}} \rangle$ bin grid. Since kinematic migrations of jet-quantities due to detector effects are expected to be almost independent of the electron kinematics, no additional bin grid for these bins are required. The χ^2/n_{dof}-value of $v8$ improve slightly compared to $v6$, however less bins are measured, which supports our assumptions.

In version $v8$ and the following versions, six side-bins are defined to take the dijet boundaries into account. One bin representing dijet events with $M_{12} < 16\,\mathrm{GeV}$, two bins for dijet events in the extended phase space of $p_{\mathrm{T}}^{\mathrm{jet2}} < 5\,\mathrm{GeV}$ for $Q^2 < 200\,\mathrm{GeV}^2$ and $Q^2 > 200\,\mathrm{GeV}^2$, one bin for dijet events with $M_{12} < 16\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\mathrm{jet2}} < 5\,\mathrm{GeV}$, one bin for the phase space extended at high rapidities of $-1.0 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75$ and one bin for phase space extended at low jet rapidities $-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.5$.

The unfolding schemes $v10$ and $v11$ include an additional dimension for the variables $\Delta R_{12} = \sqrt{\Delta\eta_{12}^2 + \Delta\phi_{12}^2}$ and M_{12}, respectively. Control distributions of these observables on detector level are shown in figure 4.7 for M_{12}, and figure 9.7 for $\Delta\eta_{12}$. Both variables are represented by a 3×2 matrix, where all bins are within the MPS. In order to ensure sufficient statistics, these two variables are only considered in the central phase space region within $7 < \langle p_{\mathrm{T}} \rangle < 30\,\mathrm{GeV}$ on generator as well as detector level.

The inclusion of additional variables into the unfolding scheme does not reduce the χ^2.

Trijet unfolding schemes

Label	Dimension 1	Dimension 2	Dimension 3
$v6$	y $3 \to 2$ [1(1)]	Q^2 $15 \to 8$ [6(6)]	$\langle p_{\mathrm{T}} \rangle + (M_{12} + p_{\mathrm{T}}^{\mathrm{jet2}} + \eta_{\mathrm{lab}}^{\mathrm{jet}})$ $7 + 10 \to 4$ [3(3)] + 5
$v7$	y $3 \to 2$ [1(1)]	Q^2 $15 \to 8$ [6(6)]	$\langle p_{\mathrm{T}} \rangle + \mathrm{EXT}_{n_{\mathrm{jet}} \geq 3} + (n_{\mathrm{jet}} = 2)$ $8 + 8 + 6 \to 5$ [3(3)] + 4 + 3
$v8$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_{\mathrm{T}}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $15 + 6 \to 8$ [6(6)] + 6	$\langle p_{\mathrm{T}} \rangle$ $8 \to 5$ [3(3)]
$v13$	y $3 \to 2$ [1(1)]	$Q^2 + (M_{12} + p_{\mathrm{T}}^{\mathrm{jet2}}) + \eta^+ + \eta^-$ $14 + 6 \to 8$ [6(6)] + 4	$\langle p_{\mathrm{T}} \rangle$ $9 \to 6$ [3(3)]

Table 9.5: Various trijet unfolding schemes. All schemes are three dimensional unfolding schemes in y, Q^2 and $\langle p_{\mathrm{T}} \rangle$, with additional side-bins for quantities which define the trijet phase space. More details are given in the text.

Therefore, we retain from such an inclusion in order to have the most simple unfolding scheme.

Similarly as for the inclusive jet measurement, the migrations in $\langle p_{\mathrm{T}} \rangle$ are dominant. We therefore increase the number of $\langle p_{\mathrm{T}} \rangle$-bins on generator and detector level successively in versions $v9$, $v12$ and $v13$[5]. Although, version $v9$ does not result in a reduced χ^2 (see figure 9.5), significant improvement is observed for versions $v12$ and $v13$.
The large χ^2 of version $v8$ is mostly caused by the two $\langle p_{\mathrm{T}} \rangle$-bins between $11 < \langle p_{\mathrm{T}} \rangle < 30\,\mathrm{GeV}$ [6]. Therefore, especially in this phase space region, a considerably finer bin grid on detector and on generator level was chosen. The small χ^2-value of $v13$ compared to $v8$ is then especially achieved in the bin of $11 < \langle p_{\mathrm{T}} \rangle < 18\,\mathrm{GeV}$, which is subdivided on generator level into three bins and thus corrects for the migrations more accurately.
Since $v13$ shows the smallest χ^2-value, this unfolding scheme is used for the dijet measurement.
The detailed description of the bin grids are outlined in chapter 6. The reason for the comparatively large χ^2 of version $v9$ compared to $v8$ is not clear.

9.2.4 Study of the trijet unfolding scheme

The unfolding scheme for the trijet measurement is closely related to the dijet unfolding scheme, because of the similarities of the measurements. A summary of the various unfolding schemes for the trijets is given in table 9.5. The corresponding χ^2/n_{dof}-values from the Monte Carlo studies are summarized in figure 9.8.
The version $v6$ of the trijet unfolding is a three-dimensional unfolding in the variables y, $Q^2$and $\langle p_{\mathrm{T}} \rangle$. Ten additional side-bins to take care of migrations near the trijet phase space boundaries are included in the $\langle p_{\mathrm{T}} \rangle$-dimension and are defined for $M_{12} < 16\,\mathrm{GeV}$,

[5]The version $v12$ and $v13$ incorporate also a different Q^2 bin grid and a reduced number of only four additional side-bins on detector level. This is due to the limited memory size of $\approx 3\,\mathrm{GB}$ on the 32-bit operating system (SL5) used.

[6]A precise measure of the χ^2/n_{dof}-value for subsets of data points is not trivial, due to the correlations. For these studies, the bins of the Q^2-dimension are integrated and an error propagation is performed to calculate the new covariance matrix $\boldsymbol{V}_x$ of the four bins for $\langle p_{\mathrm{T}} \rangle$.

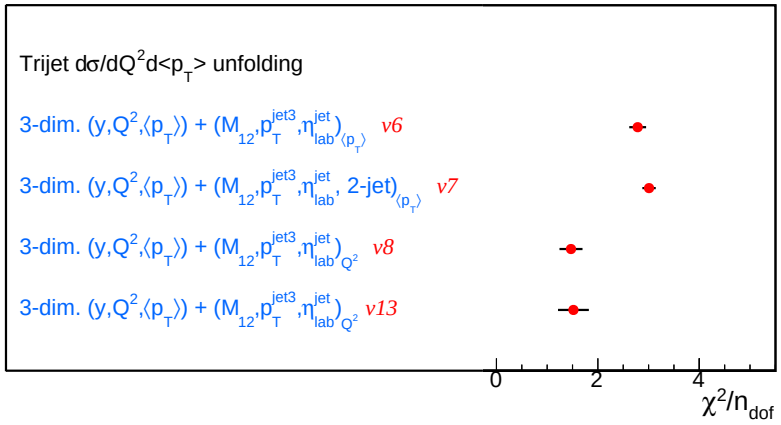


Figure 9.8: Values of χ^2/n_{dof} for the various trijet unfolding schemes. All unfolding schemes are based on a three-dimensional implementation, where additional side-bins are included for quantities which define a 'trijet'-event. The differences of the various schemes are outlined in detail in table 9.5 and are discussed in the text.

$p_{\mathrm{T}}^{\mathrm{jet2}} < 5\,\mathrm{GeV}$, $\eta_{\mathrm{lab}}^{\mathrm{jet}} < -1.0$ and $\eta_{\mathrm{lab}}^{\mathrm{jet}} > 2.5$.

The trijet unfolding scheme $v7$ includes six additional side-bins on detector level and three on generator level for events with exactly two jets. These bins do not result in a reduced χ^2/n_{dof}, compared to $v6$. These two-jet bins have around ten times the statistics than the trijet phase space bins. However, below 1 % of the two-jet events migrate into the trijet generator level phase space.

The χ^2/n_{dof}-value is reduced for version $v8$ and $v13$, where the trijet phase space migrations are included in the Q^2-dimension instead of the $\langle p_{\mathrm{T}}\rangle$-dimension, thus making these bins differential in $\langle p_{\mathrm{T}}\rangle$. Just like in the dijet case described before, this reduces significantly the χ^2/n_{dof}.
The small χ^2/n_{dof}-value for the trijet measurement is caused partly by the small statistical precision, which also enters the χ^2 formula. In order to assure sufficient statistics in the three-dimensional unfolding, the number of bins, i.e. in $\langle p_{\mathrm{T}}\rangle$, is not reduced further. The detailed description of the unfolding employed is given in chapter 6.

9.2.5 Comparison of correlation matrices from different unfolding schemes

The study of various unfolding schemes has revealed that changes in the unfolding scheme may influence the unfolded result significantly, although, the input data are always identical. This means that the detector effects, which are corrected for, are corrected differently when using different migrations. In figure 9.9 two correlation matrices, using 'real' data now, are compared. The correlation matrix for the inclusive jet unfolding $vI3$ and the dijet and trijet unfolding $v6$ is compared to the final correlation matrix of the inclusive jet unfolding scheme $vI4\mathrm{F}_{gr}$ (denoted in figure 9.9 $v13$) and $v13$ for the dijet and trijet unfolding.
In the inclusive jet measurement the correlation coefficients differ between adjacent $p_{\mathrm{T}}^{\mathrm{jet}}$-bins by up to 0.2, with the size of the negative correlations being smaller for $v13$. The negative correlations between the $\langle p_{\mathrm{T}}\rangle$-bin 1 and 2 of the dijet measurement increase by

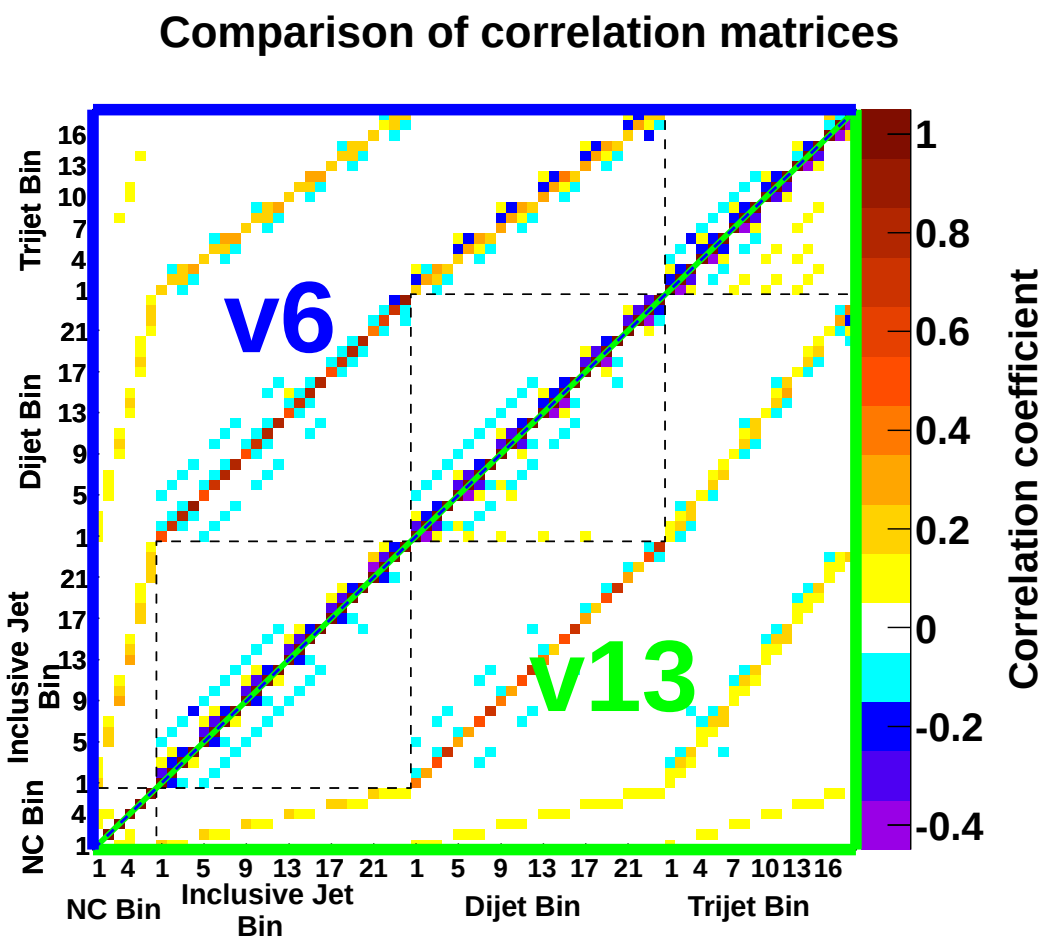


Figure 9.9: Two correlation matrices obtained from two different unfolding schemes applied to data. The upper-left half-matrix shows the correlation matrix for data from the inclusive jet unfolding $vI3$ and the dijet and trijet unfolding $v6$. The lower-right half-matrix shows the final correlation matrix for data from the inclusive jet unfolding scheme $vI4\mathrm{F}_{gr}$ and $v13$ for the dijet and trijet unfolding. The black dashed boxes indicate the four individual measurements. The binning is identical to figure 7.5.

-0.1 for $v13$ compared to $v6$, whereas the negative correlations decreases and almost vanish the between $\langle p_{\mathrm{T}} \rangle$ bins 3 and 4 in $v13$ (see table 7.1 for the bin labels). The dijet and trijet measurements show small positive correlations between adjacent Q^2 bins in $v13$.

Also differences in the correlations between the single measurements can be observed. This can be explained by the differences in the migrations to the outside of the phase space region.

The correlations are of comparable size for MC pseudo-data and show the same differences between $v6$ and $v13$ and no significant difference between Rapgap and Django pseudo-data could be observed, if unfolded with the other generator.

9.3 Regularization condition and strength

The regularization term of the unfolding equations 5.4 to 5.6 consists of two free parameters: the *regularization parameter* τ (or *strength*) and the *regularization condition* $\boldsymbol{L}^2$. Both can be chosen freely. Here, we want to study the influence of these parameter on unfolded pseudo-data and on the unfolded data cross sections, their uncertainties and their correlations in order to find the optimal choice.

9.3.1 Study of different regularization conditions

The regularization condition enters the unfolding procedure through the matrix $\boldsymbol{L}^2$ of size $n_{\mathrm{gen}} \times n_{\mathrm{gen}}$.
The simplest form of the regularization condition is referred to as *size regularization* and is defined through a unity matrix

$$\boldsymbol{L}^2_{\mathrm{size}} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}. \tag{9.4}$$

The size regularization suppresses large fluctuations of the amplitude or size of each bin content individually, i.e. the value of the adjacent bins are not considered. The size regularization is the simplest form of regularization and can be applied easily to a complex multidimensional unfolding, and it is well studied in the literature. A large regularization strength τ pushes the unfolded result towards zero.

More complex regularization conditions can be realized by mimicking discrete derivatives. The *first derivative* is approximated by differences of the entries in adjacent bins, i.e. $x_{i+1} - x_i$. A *second derivative* (or *curvature*) is approximated by $(x_{i+1} - x_i) - (x_i - x_{i-1})$ [181].

The curvature regularization condition for a one-dimensional unfolding thus reads

$$\boldsymbol{L}^2_{\text{Curvature,1-dim.}} = \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}. \tag{9.5}$$

The generalization of the discrete first or second derivative to multiple variables is straight forward and is conveniently implemented in TUnfold [181]. The definition of a curvature measure is only reasonable for bins, with at least two adjacent bins in both directions in the considered variable. The advantage of the derivative or curvature regularization are the introduction of "smoothness" conditions, since the amplitude of each bin is related to the size of the adjacent bins. However, as caveat, the definition of the discrete derivative is insensitive to the bin widths. Since the bin-widths are not included in this definition of the discrete derivatives, the curvature constraint may introduce unphysical shapes when the bin grid is not chosen to be equidistant[7]. Furthermore, in a multidimensional unfolding the identical curvature is assumed in each dimension of the problem. This is for instance obviously inappropriate when considering the shape of the y and Q^2 distributions.

We study five different regularization conditions

- $\boldsymbol{L}_{\text{size}}$: Size regularization
  This is the default regularization condition, where $\boldsymbol{L}^2 = 1$.

- $\boldsymbol{L}_{\text{1D-Deriv.}}$: 1-dim. derivative regularization in p_{T}
  A one-dimensional discrete derivative regularization is realized between neighboring p_{T} bins. The regularization pattern is set up for the p_{T} bin grids in every y, Q^2 and $\eta^{\text{jet}}_{\text{lab}}$ element. The NC DIS pseud-data are regularized using the size regularization.

- $\boldsymbol{L}_{\text{1D-Curv.}}$: 1-dim. curvature regularization in p_{T}
  A similar regularization to $\boldsymbol{L}_{\text{1D-Deriv.}}$, but where the p_{T}-bins are regularized using the curvature condition.

- $\boldsymbol{L}_{\text{2D-Deriv.}}$: 2-dim. derivative regularization in Q^2 and p_{T}
  A two-dimensional derivative regularization is realized in the variables Q^2 and p_{T}. The 2-dim. pattern is set up for each element of the y and $\eta^{\text{jet}}_{\text{lab}}$-matrices and separately for each of the three jet measurements. Hence, there are no constraints for adjacent $\eta^{\text{jet}}_{\text{lab}}$ or y bins, between different jet measurements or among additional side-bins. Since additional side-bins are realized in $v13$ in the Q^2-dimension, there is a 1-dim. derivative regularization pattern in the p_{T}-dimension present. The NC DIS pseudo-data are regularized using a one-dimensional derivative regularization for adjacent Q^2-bins.

[7] This problem is solved, as soon as the bin widths are considered. Such methods are e.g. implemented in the follow-up development in `TUnfoldDensity`.

- $\boldsymbol{L}_{\mathrm{2D-Deriv.}}$: 2-dim. curvature regularization in Q^2 and p_{T}
  Similarly to $\boldsymbol{L}_{\mathrm{2D-Deriv.}}$, but where neighboring Q^2 and p_{T}-bins are regularized using a curvature condition. The NC DIS pseudo-data are used with a one-dimensional curvature condition for each y-bin.

The χ^2-method is used to quantify the performance of the different regularization conditions. The regularization strength is determined using the L-Curve method. The χ^2/n_{dof}-values for the unfolding scheme $v13$ are illustrated in figure 9.10. We observe that the

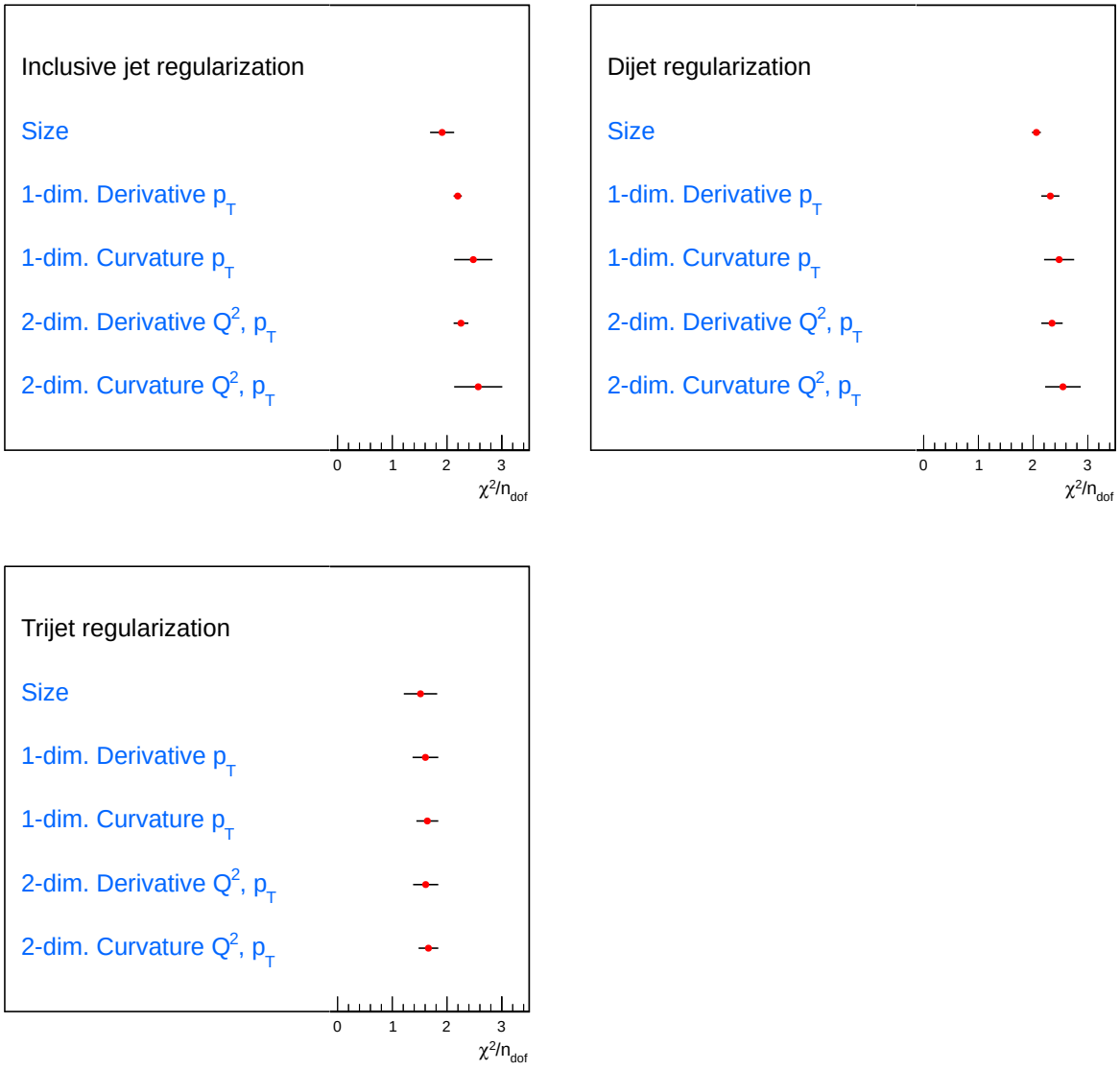


Figure 9.10: χ^2/n_{dof}-values for an unfolding of pseud-data using different regularization conditions. The regularization strength τ is determined using the L-Curve method.

χ^2/n_{dof}-values do not improve with increasing complexity of the regularization condition[8]. The observed differences are not significant. We therefore prefer the size regularization for the data unfolding because of its simplicity.

[8]Surprisingly, the value $\chi^2_{\mathrm{Rg(Dj)}}$ stays almost constant for the different regularization conditions for all three jet measurements, while changes are only seen in $\chi^2_{\mathrm{Dj(Rg)}}$. No reason for this curiosity could be found.

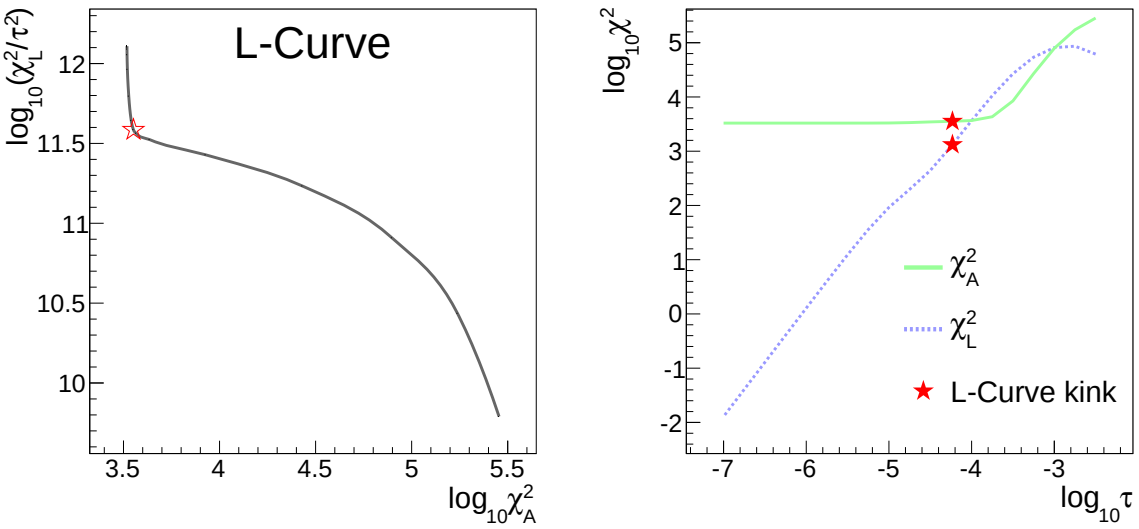


Figure 9.11: The left pad shows the L-curve for the regularization parameter τ in the range of $-7.0 < \log_{10}\tau < -2.5$. The kink, which is defined by the maximum curvature of the L-curve is chosen as optimal regularization parameter and is highlighted with the red star. The right pad shows the ingredients for the L-curve, which are the values of $\log_{10}\chi^2_A$ and $\log_{10}\chi^2_L$ as functions of $\log_{10}\tau$. The L-curve kink is not distinct in these curves.

A three-, four- or even higher multidimensional regularization condition on the 1st or 2nd derivative is not considered, since there are at most three bins in y or $\eta^{\text{jet}}_{\text{lab}}$, or exactly two bins for additional side-bins. Therefore, no reasonable discrete derivative can be defined in these variables. Furthermore, the migrations within these variables are assumed to be small w.r.t. migrations in p_{T} or Q^2.

Further regularization conditions have been studied, for instance, where the regularization condition is only applied for phase space bins. These results are not discussed here, since no significant differences could be observed.

9.3.2 Study of the regularization strength

In this section the impact of the regularization strength on the unfolded cross section from true data, their statistical and systematic uncertainties is studied. There are several proposals in the literature for the optimal determination of the regularization parameter τ, which are partly only valid for derivative or curvature regularization conditions. Therefore, we consider the size regularization and the 2-dim. curvature regularization condition, as a representation of a complex regularization pattern.
Two prominent methods for the determination of τ are proposed in the literature.

- **The L-Curve method** [212]
  The L-Curve is defined as the graph of the two variables $L_y(\tau) = \log_{10}\chi^2_L(\tau)/\tau^2$ versus $L_x(\tau) = \log_{10}\chi^2_A(\tau)$. The values are calculated from the full hadron level vector, i.e. using all bins (see equation 5.4). This curve which has similarity to an L-shaped curve typically shows a *kink*. The algorithm determines the curvature of the L-curve by a cubic spline interpolation and finds the maximum of the L-curve by another cubic spline interpolation [181]. The maximum is identified by the kink of the L-curve, which is found for the employed unfolding to be $\log_{10}\tau = -4.231$. The L-curve is given in figure 9.11, where the regularization parameter τ in the range of

$-7.0 < \log_{10}\tau < -2.5$ is used for plotting.

- **Minimizing the (global) correlation coefficient** [189]
  The idea of minimizing the global correlation coefficient provides an alternative approach for finding the appropriate regularization strength. A *correlation coefficient*[9] is defined for each bin through

$$\rho_i = \sqrt{1 - \frac{1}{(\boldsymbol{V}_x)_{ii}(\boldsymbol{V}_x^{-1})_{ii}}} \; . \tag{9.6}$$

  The global correlation coefficient is then given as the average of all bins N:

$$< \rho >= \frac{1}{N}\sum_{i}^{N} \rho_i \; . \tag{9.7}$$

  This method is better suited for derivative and curvature regularization conditions, which introduce artificial correlations between adjacent bins through $\boldsymbol{L}^2$. Therefore, it was suggested that minimizing the global correlation coefficient is the preferred choice for setting the regularization strength, since the influence of $\boldsymbol{L}^2$ is reduced. We will see that this proposal is not reasonable for this analysis.

In this analysis, the global correlation coefficient is not a reasonable estimate, since the correlations are dominated by the positive migrations between the single measurements. Especially the correlations between the inclusive jet and dijet measurement are large. Also the inclusive jet measurement has positive correlations already on detector level.
Instead of the global correlation coefficient, we therefore determine the average correlation coefficient for each measurement and only in the measurement phase space[10]. Furthermore, we calculate the correlation coefficient for single differential measurements $\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}$ and $\mathrm{d}\sigma/\mathrm{d}Q^2$. These quantities give insight into the correlations which arise from the Q^2 or the p_{T}-migrations.

As soon as the regularization condition $\boldsymbol{L}^2$ is chosen, only one single free parameter persists: the regularization strength τ. Without any technical or mathematical method, we therefore can study the dependence of relevant parameters on τ without any bias from the determination of τ.

The average correlation coefficient as function of τ is shown in figure 9.12 for the size regularization and the curvature regularization. We observe that the average correlations are constant at small values of τ. With increasing regularization strength the correlations are becoming smaller, except for the inclusive jet measurement when performed differentially in $p_{\mathrm{T}}^{\mathrm{jet}}$. Since the inclusive jet data has has positive statistical correlations from the jet multiplicities, these positive coefficients are in contradiction to the method of minimizing the global correlation coefficient. All other measurements have a minimum average correlation coefficient at quite large values of τ. No globally valid minimum of $< \rho >$ for all measurements can be found.

[9] We also refer to "correlation coefficients" as elements of the correlation matrix. These coefficients ϱ_{ij}, however, are denoted with two superscripts for the two bins, which they refer to.

[10] We use the bin grid of the measurement. This bin grid has combined several bins of the hadron level vector x.

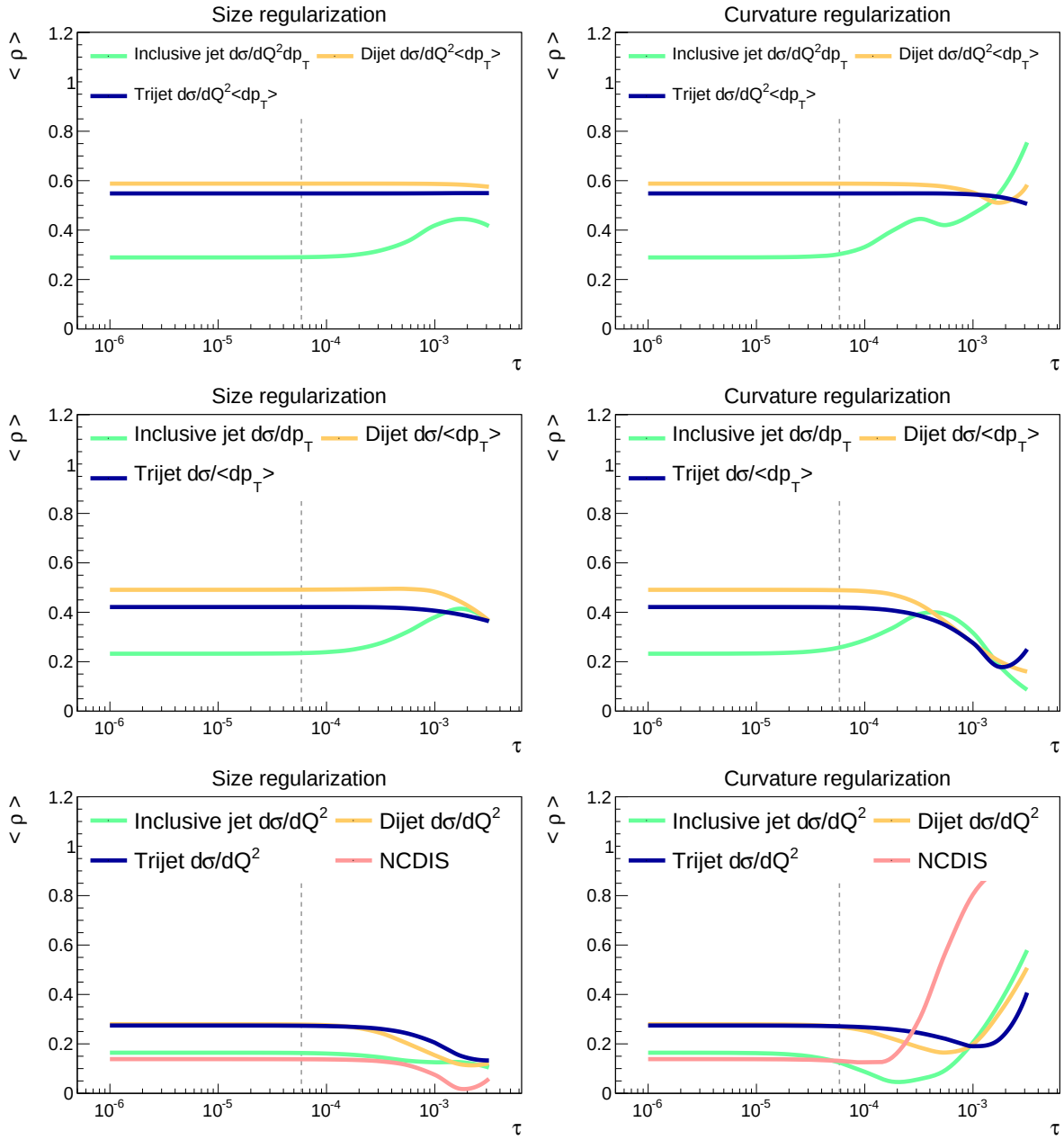


Figure 9.12: Average correlation coefficient $< \rho >$ as function of the regularization strength τ. The left pads show $< \rho >$ for the size and the right pads for the curvature regularization. The upper row shows$< \rho >$ for the three double differential jet measurements, the central pads for the single differential measurements as function of p_{T} and the lower pads show $< \rho >$ for the single differential jet measurements as function of Q^2 including the NC DIS measurement. The grey dashed line shows the τ-parameter as determined from the L-curve scan.

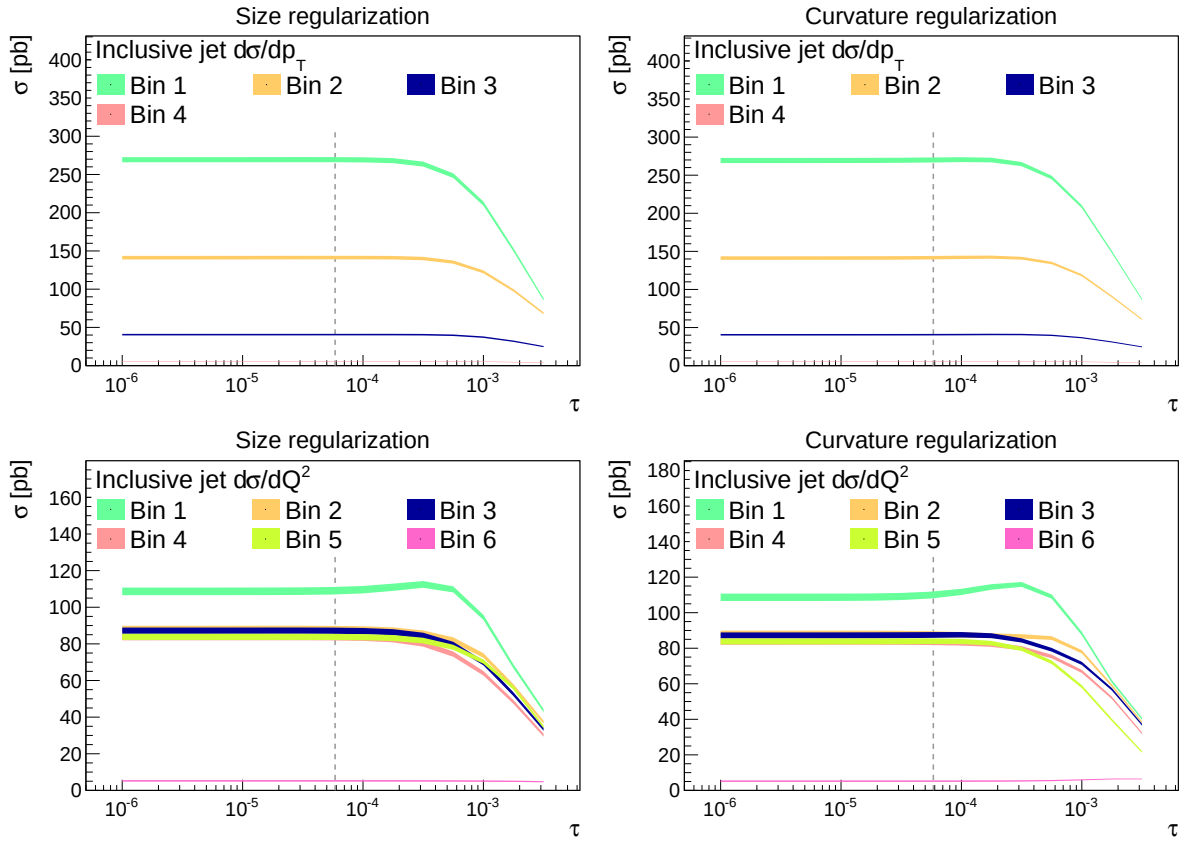


Figure 9.13: Single differential inclusive jet cross sections versus the regularization strength τ for $\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{\mathrm{jet}}$ (top) and $\mathrm{d}\sigma/\mathrm{d}Q^2$ (bottom). The left plots are calculated using size regularization, the right plots using curvature regularization. The width of the colored lines indicates the statistical uncertainty. The grey shaded band, typically too small to be seen, indicates the model uncertainty added in quadrature to the statistical uncertainty. The grey dashed line indicates the τ parameter from the L-Curve scan.

The determined τ-value from the L-curve, however, is in the region, where the average correlation coefficients are almost constant. When $\tau \to 0$, i.e. the regularization becomes negligible, both regularization conditions predict the same correlation average coefficients.

The dependence of the inclusive jet cross section on the regularization strength τ is shown in figure 9.13 and for the dijet cross section in figure 9.14. Shown are the single differential cross sections as function of p_{T} and Q^2 for the two regularization conditions studied. For large values of τ the cross sections are pushed towards zero for the size regularization, while for a negligible regularization the cross sections approach a constant value, which is identic for the two regularization conditions. The statistical uncertainty and the model uncertainty, illustrated by the width of the curves in figures 9.13 and 9.14 are found to be independent of the regularization strength. The τ-value from the L-curve scan is in the region, where the cross section approaches a negligible dependence on τ and hence these cross sections can be considered as independent from the regularization.

If a value for τ from the minimal global correlation method would be chosen, the cross section would be largely dominated by the regularization and would be unreasonably small. The cross section dependence of the trijet measurement and the NC DIS measurement is similar to the dijet measurement and therefore not shown.

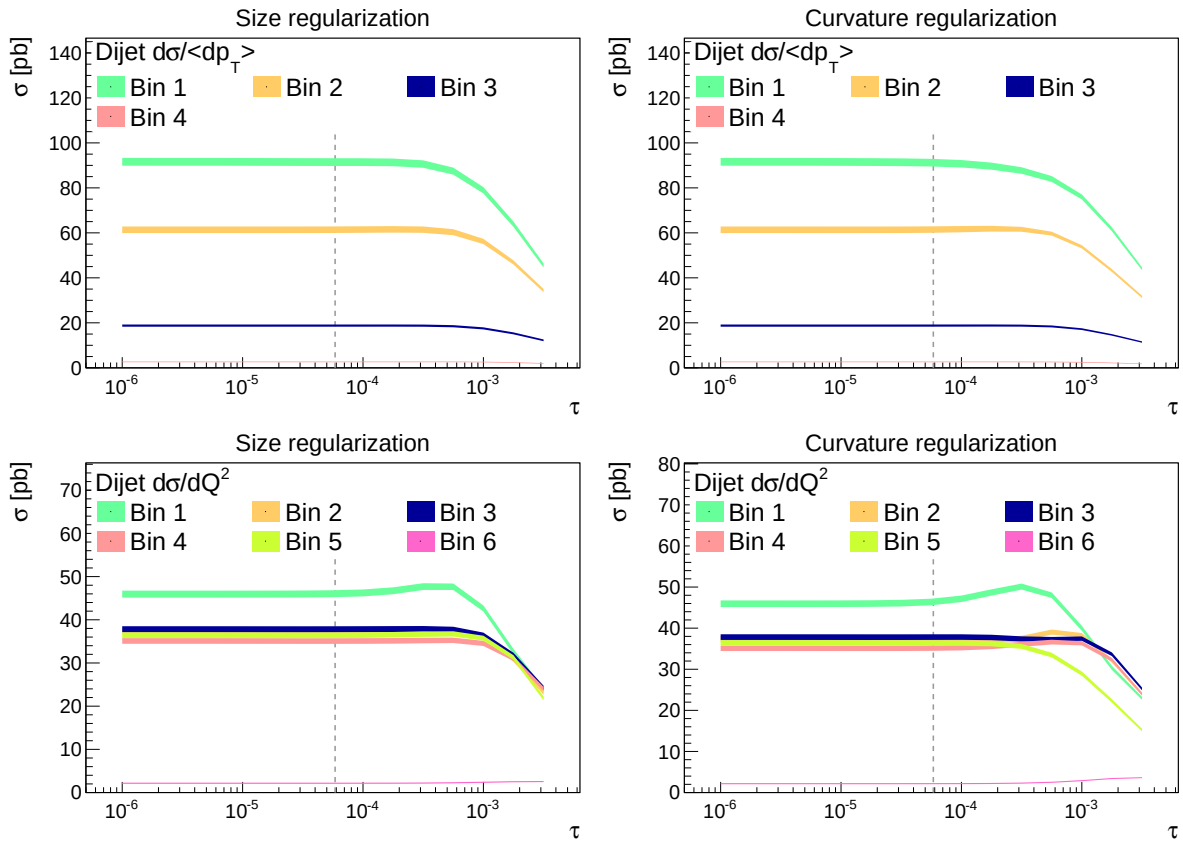


Figure 9.14: Single differential dijet cross sections versus the regularization strength τ for $\mathrm{d}\sigma/\mathrm{d}\langle p_\mathrm{T}\rangle$ (top) and $\mathrm{d}\sigma/\mathrm{d}Q^2$ (bottom). The left plots are calculated using size regularization, the right plots using curvature regularization. The width of the colored lines indicates the statistical uncertainty.

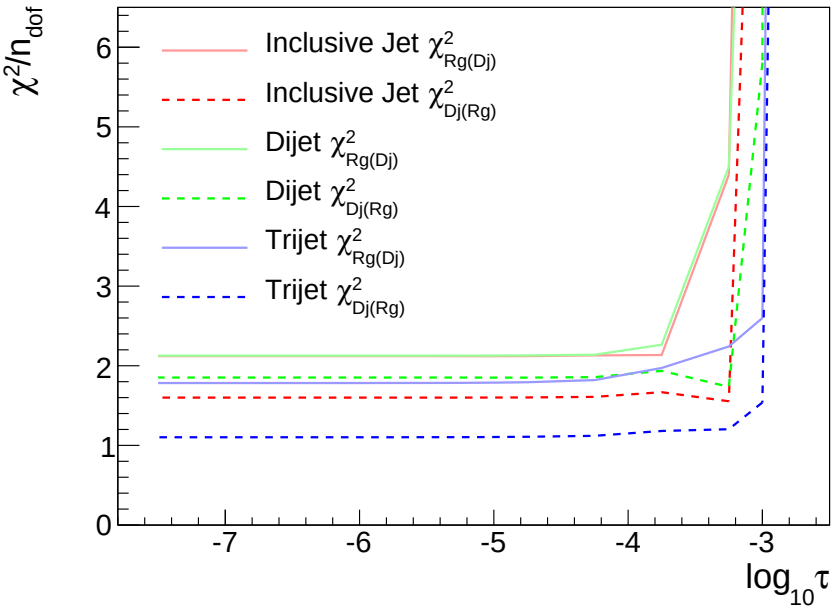


Figure 9.15: Values for χ^2/n_{dof} determined from different Monte Carlo pseudo-data as function of the regularization parameter τ. The size regularization condition is employed.

The values of χ^2/n_{dof} (c.f. equation 9.2) are also studied for Monte Carlos pseudo-data[11] as function of τ in figure 9.15. The value of χ^2/n_{dof} is constant for small values of τ for all three measurements. As soon as the regularization strength increases, the χ^2/n_{dof}-values increase dramatically. This proves that only a small regularization parameter is capable of reproducing the true distribution from Monte Carlo pseudo-data. The χ^2/n_{dof}-values of the L-curve scans (see table 9.1) are larger by 0.05 to 0.1 than those for the smallest τ value. Since the L-curve method is a well established method to determine τ, and the χ^2/n_{dof} values cannot be studied for data, we stick to the L-curve method.

9.4 Unfolding of the ξ measurements

The measurement of multijet production as function of Q^2 and ξ (see section 7.5) is also corrected for detector effects with a regularized unfolding. The employed unfolding scheme and its study are discussed in this section.

The unfolding scheme for the ξ-measurement is closely related to the unfolding of the $\langle p_{\mathrm{T}} \rangle$-measurements. It uses the overall definition of the migration matrix with a 4×4 structure for four measurements (NC DIS, inclusive jet, dijet and trijet measurement). The employed unfolding scheme of the NC DIS measurement is explained in section 6.4. The unfolding of the inclusive jet measurement is closely related to the unfolding scheme $vI4+$, with the $\eta^{\mathrm{jet}}_{\mathrm{lab}}$-dimension removed[12]. The inclusive jet data are only used in the unfolding because to determine the correlations between the inclusive jet measurement and the dijet and trjet ξ-measurements.

[11] The value of χ^2/n_{dof} should not be mistaken with χ^2_A. The former is calculated from MC pseudo data with its true hadron level, while the latter one is calculated on detector level from the vectors y and $\boldsymbol{A}x$.

[12] Discarding the $\eta^{\mathrm{jet}}_{\mathrm{lab}}$-dimension reduces the size of the migration histogram by a factor of 6 and the covariance matrix by a factor of 9. This became necessary due to limited memory allocation capability, since the dijet and trijet unfolding for the ξ-measurement employs an additional forth dimension and the matrices are huge.

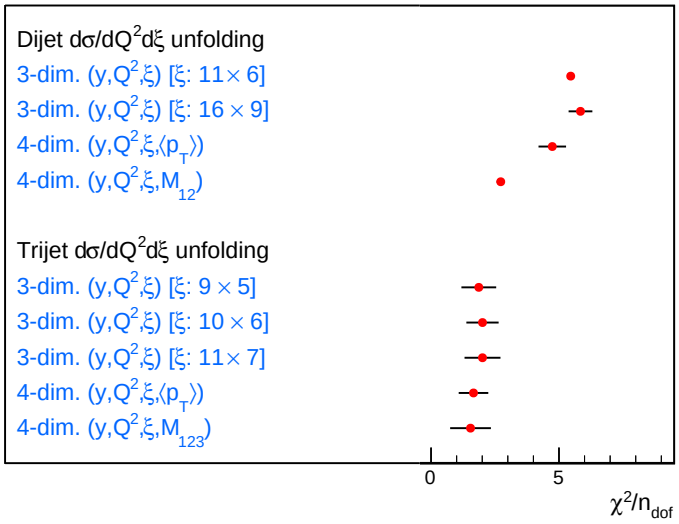


Figure 9.16: Values for χ^2/n_{dof} for various unfolding schemes for the dijet and trijet pseudo-data $\mathrm{d}\sigma/\mathrm{d}Q^2\mathrm{d}\xi$. All unfolding schemes include additional side-bins in the Q^2-dimension to take migrations into account at the phase space boundaries for dijet (trijet) events with $M_{12} < 16\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\mathrm{jet}} < 5\,\mathrm{GeV}$, $-1.0 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.75$ and $-1.5 < \eta_{\mathrm{lab}}^{\mathrm{jet}} < 2.5$.

9.4.1 Unfolding of the dijet ξ measurement

The unfolding scheme of the dijet ξ-measurement is defined similarly to the unfolding scheme of the dijet $\langle p_{\mathrm{T}}\rangle$-measurement, where the $\langle p_{\mathrm{T}}\rangle$-dimension is substituted by an 11×6 migration scheme in ξ. The ξ-phase space is extended with one lower and one upper side-bin to $0 < \xi < 1$, and the four bins of the ξ-measurement are defined the same on generator level. This three-dimensional unfolding scheme includes also additional side-bins in the Q^2-dimension for migrations into/out of the MPS (c.f. table 6.4). The resulting χ^2/n_{dof}-value of 5.5 is shown in figure 9.16.

Employing a finer bin grid of 16×9 bins in the measured variable ξ, where bins of the measurement are represented by 2 bins each on generator level[13] does not improve the χ^2/n_{dof}-value, and hence migrations in other variables than ξ are obviously relevant. In a further trial a the fourth dimension[14] is included for the variable $\langle p_{\mathrm{T}}\rangle$ with a 4×2 migration scheme[15] for the phase space $3 < \langle p_{\mathrm{T}}\rangle < 50\,\mathrm{GeV}$. The still considerably large value of χ^2/n_{dof} of 4.7 suggests that the differences of the migrations between the two MC models must arise from a different variable. This is also reasonable, since $\langle p_{\mathrm{T}}\rangle$ is integrated out in the ξ measurement and does not essentially constrain the phase space.
Since the observable ξ is calculated using $\xi = x_{\mathrm{Bj}} + (M_{12}^2/(ys))$[16] it is reasonable to include the variable M_{12} with a 4×2 bin grid into the unfolding[17]. This enables to disentangle the

[13]The highest ξ bin ($0.08 < \xi < 0.316$) is identically defined for the generator level bin grid as for the measurement.

[14]The $\langle p_{\mathrm{T}}\rangle$-bin grid is not realized for the ξ side-bins $\xi < 0.006$ and $\xi > 0.316$ due to the small statistics in these bins.

[15]In this unfolding scheme migrations in eight variables are considered. It is only known about a five-dimensional unfolding up to date [213].

[16]For the trijet measurement it is defined as $\xi = x_{\mathrm{Bj}} + \left(M_{123}^2/(ys)\right)$, where M_{123} is the invariant mass of the three leading jets.

[17]The bin grid for M_{12} is defined as $0 - 20 - 28 - 40 - 1000\,\mathrm{GeV}$ on detector level and $0 - 28 - 1000\,\mathrm{GeV}$

migrations of the different variables x_{Bj}, M_{12} and y, since also ξ itself is considered. The inclusion of the migrations in M_{12} reduce the χ^2/n_{dof} significantly w.r.t. previous unfolding schemes to $\chi^2/n_{\mathrm{dof}} = 2.8$. This unfolding scheme is employed for the unfolding of data.

9.4.2 Unfolding of the trijet ξ measurement

The unfolding of the trijet ξ-measurement is initially based on the unfolding of the $\langle p_{\mathrm{T}} \rangle$-measurement (see section 6.8), where the $\langle p_{\mathrm{T}} \rangle$-bin grid is substituted with a 9×5 migration scheme in ξ. We optimize the unfolding by employing a finer bin grid on detector and generator level with 10×6 and 11×7 bins. All these three unfolding schemes show similar values for χ^2/n_{dof} of $\chi^2_{\mathrm{Dj(Rg)}}/n_{\mathrm{dof}} \approx 2.6$ and $\chi^2_{\mathrm{Rg(Dj)}}/n_{\mathrm{dof}} \approx 1.3$. Furthermore, many bins of the measurement show very large model uncertainties of more than 100 %.
For an optimization, we include $\langle p_{\mathrm{T}} \rangle$ or M_{123} as a forth dimension in the unfolding scheme. The χ^2 does not improve, because of the already quite low χ^2 and the large statistical uncertainties. Using the M_{123}-unfolding the not significant smallest value of χ^2/n_{dof} is obtained and we employ this unfolding scheme to unfold the data.

The cross section result is expected to be reliable, since also the model uncertainty is reduced and reasonable values of 10 - 30 % for most bins of the measurement (see figure 7.16). However, the bin 4A and 6C has a model uncertainty of 40 % and 65 %, respectively, and the model uncertainty shows large fluctuations between adjacent bins. We expect that migrations in M_{123} and y are relevant between the phase space bins. Furthermore, there are substantial migrations among the phase space boundaries in $p_{\mathrm{T}}^{\mathrm{jet}} > 5\,\mathrm{GeV}$ and also in $M_{12} > 16\,\mathrm{GeV}$. These phase space constraints, give a rather complicated phase space boundary for each bin of the measurement and migrations become difficult to determine. Since the statistical precision of the trijet measurement is small, no further improvements can be achieved.

9.5 Further studies

The method of regularized unfolding has been developed further with some options, which we studied and want to mention here.

- **Area constraint**
  The method of an *area constraint* adds an additional term [181]

$$\chi^2_{\mathrm{Area}} = \lambda \left(\sum_i y_i - (\sum_i A_{ij})^{\mathrm{T}} x \right) \tag{9.8}$$

  to the χ^2 definition of the unfolding equation 5.4, where λ is a free Lagrange multiplier, which is set to its stationary point in the unfolding procedure. The area constraint enforces the unfolded result to match the total event count and thus preserves the normalization. This becomes relevant, when the data follows poissionian

on generator level is included in the region $0.006 < \xi < 0.316$. The additional side-bins for dijet events with $M_{12} < 16\,\mathrm{GeV}$ and $p_{\mathrm{T}}^{\mathrm{jet2}} < 5\,\mathrm{GeV}$ remain in the Q^2-dimension, and thus these bins also include the M_{12}-migration scheme and migrations in M_{12} are essentially represented through a 5×3 migration scheme. We further optimize the six additional side-bins for migrations in M_{12}. The Q^2-detector level bin 17 (see table 6.4) is subdivided at $M_{12} = 13\,\mathrm{GeV}$ into 2 bins and the bins 16 and 20 (both represent dijet events with $M_{12} > 16\,\mathrm{GeV}$) are combined, because of the additional M_{12} dimension in these bins.

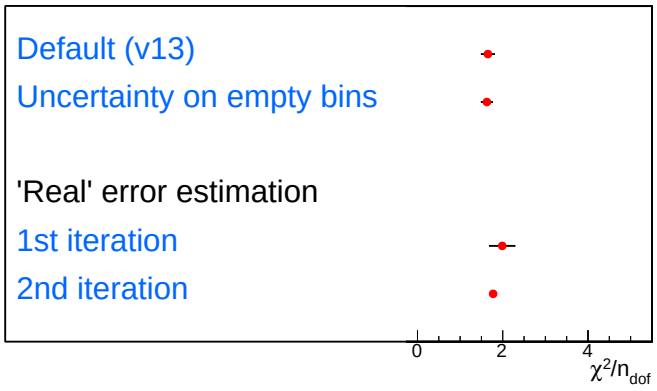


Figure 9.17: Values of χ^2/n_{dof} for all phase space bins of the three jet measurements using pseudo-data. Compared are the standard unfolding procedure to results where an uncertainty on empty bins is estimated and where the "real" statistical uncertainty is estimated from the determined true level.

statistics. However, no visible difference is observed, when the area constraint is employed.

- **Uncertainty of empty bins**
  Bins with no entries and hence zero uncertainty, are assigned an uncertainty, where there is a non-vanishing Monte Carlo prediction. This is reasonable for phase space regions, where the Monte Carlo prediction is small, but no entries are measured due to statistical fluctuations. When the uncertainty is zero, this may bias the result. For this study, we assign to all bins with zero error an uncertainty of one, i.e. $(\boldsymbol{V}_y)_{ii} = 1$. The calculated χ^2/n_{dof} value for the pseudo-data is compared to the default result without area constraint in figure 9.17. No significant change is observed. Therefore, the statistical uncertainties and the unfolded cross sections are expected to be insensitive to these bins.

- **Assumption of 'real' errors**
  The unfolding equation 5.4 leads to a biased results, which underestimates the cross section when the statistical precision is poor [191], since the least square ansatz of the regularized unfolding is stricly valid only for normal distributed measurements. A corrected result can be obtained by estimating the 'real' uncertainties in generalizing the approach of K. Nowak [191] towards a correlated measurement. The covariance matrix on detector level is corrected by the ratio of the folded-unfolded result $\tilde{y} = \boldsymbol{AB}y$ to the initial detector level distribution y, which is assumed to be the 'true' distribution. The covariance matrix for the first iteration is then given by

$$\tilde{\boldsymbol{V}}_y = (\boldsymbol{V}_y)_{ij} \sqrt{\frac{\tilde{y}_i}{y_i}} \sqrt{\frac{\tilde{y}_j}{y_j}}, \tag{9.9}$$

  which us used in the unfolding procedure together with the measured data vector y. This procedure can be applied iteratively.

  The χ^2/n_{dof}-values for the first and second iterations are shown in figure 9.17. Only a very small improvement of the χ^2 value is observed. Therefore, this procedure is not employed for data.

9.6 Individual unfolding of the NC DIS measurement

The definition of the migration matrix (see figure 6.1) consists of the three jet measurements individually, i.e. without any connection between the matrices $\boldsymbol{J}_1$, $\boldsymbol{J}_2$ and $\boldsymbol{J}_3$. The idea of fixing detector-level-only entries, using the NC DIS measurement, helps to preserve the normalization of these measurements, which is especially important for the inclusive jet measurement, due to different jet multiplicities on generator and detector level.
Since the NC DIS measurement is employed for the determination of the normalized jet measurements, we have to ensure, that the detector-level-only entries in the matrices $\boldsymbol{B}_1$, $\boldsymbol{B}_2$ and $\boldsymbol{B}_3$ do not influence the NC DIS measurement. Therefore, we perform an unfolding of the NC DIS measurement separately, i.e. where $\boldsymbol{A} = \boldsymbol{E}$ and thus no jet measurement is present, to determine the *unbiased NC DIS cross section* $\hat{\sigma}_{\mathrm{NC}}$. We compare this cross section to the NC DIS cross section σ_{NC} determined from the full unfolding procedure, by calculating χ^2/n_{dof} between these two cross sections. The value of χ^2/n_{dof} using σ_{NC} and $\hat{\sigma}_{\mathrm{NC}}$ yields a value of 0.9. When only the NC DIS, the dijet and trijet measurement are unfolded together, i.e. without the inclusive jet matrices $\boldsymbol{J}_1$ and $\boldsymbol{B}_1$, the χ^2/n_{dof}-value decreases to 0.15. When only the NC DIS and the inclusive jet measurement are unfolded together the χ^2/n_{dof}-value from σ_{NC} and $\hat{\sigma}_{\mathrm{NC}}$ yields 0.6. Hence, any potential bias is mainly caused by the detector-level-only jets of the inclusive jet measurement. We can conclude that due to the high statistical precision of the NC DIS measurement, the impact of the detector-level-only entries on the NC DIS cross section is within the statistical uncertainty. We can therefore use the unfolded NC DIS measurement for the determination of the normalized jet measurements.

Chapter 10

Determination of the strong coupling constant

In this chapter, the extraction of the strong coupling constant at the scale of the mass of the Z^0 boson from the unfolded jet cross sections is presented. First, a brief theoretical recapitulation is given, and then the fit-strategy and the least-square method employed are outlined. After studying the consistency of data to theory, the strength of the strong coupling is determined for the inclusive jet, the dijet and the trijet cross sections individually. Then a combined determination of $\alpha_s(M_Z)$ from the multijet measurement is performed. The theoretical uncertainties from scale variations and from hadronization corrections are determined. The dependence on the PDF is investigated in detail. We close the chapter with a summary of the values obtained and their uncertainties.
In this work, the strong coupling constant $\alpha_s(M_Z)$ is always quoted at the scale of the Z^0-mass. For better readability we drop the scale dependence in the notation and henceforth write α_s for $\alpha_s(M_Z)$, and '$\alpha_s(M_Z)$' is only used for explicit highlighting.

10.1 Introduction

The QCD predictions for jet production in DIS are dependent on α_s and the parton density functions of the proton. If these predictions are fitted to the data with α_s as a free parameter, the value of the strong coupling constant can be determined.

The coefficients of the perturbative expansion are known up to next-to-leading order, which is the first order, where the perturbative coefficients contain counter terms to the scale dependence of the renormalized coupling, and hence predictive cross sections can be obtained. The effective coupling strength as function of μ_r is given by the SU(3)-structure of QCD and by one free parameter of the theory. For convenience the coupling strength is usually expressed through its value at the Z^0-mass[1] as $\alpha_s(M_Z)$. For its determination, we will assume that the RGE of QCD is exactly valid, and we compute the running of the strong coupling in 4-loop precision [195] in a 3-step iterative Newton procedure[2]. The β-functions

[1] A mass of the Z^0-boson of $M_Z = 91.1876\,\text{GeV}$ [12] is used.

[2] We apply the 4-loop solution of the RGE consistently to calculations of every order in the perturbative expansion. Some theorists have the point of view that the evolution of the effective coupling should be of order $\mathcal{O}(n+1)$ separately for each order of the perturbative expansion. The impact, however, would only be of the order $\mathcal{O}(10^{-3})$ for small scales of $\mu_r \approx 7\,\text{GeV}$ and without any noteworthy impact for higher scales or for trijet cross sections.

of QCD are evaluated with five quark flavors $n_f = 5$ for the calculation of the perturbative coefficients and the effective coupling strength $\alpha_s(\mu_r)$, and neither quark masses nor threshold effects are considered. Soft contributions to the hard scattering process are factorized. In the initial state they are factorized into the PDF, whereas the final state hadronization processes are considered via a correction factor. The data have been corrected for QED effects, whereas the predictions are corrected for Z^0 and γZ-interference contributions. For the efficient repeated calculation of the QCD predictions in an iterative fit, we employ the fastNLO framework which we have improved to allow different choices for the renormalization and factorization scales (see appendix C).

The experimental input to the determination of α_s are the double-differential inclusive jet measurements as function of Q^2 and $p_{\mathrm{T}}^{\mathrm{jet}}$, the measurements of dijet production as function of Q^2 and $\langle p_{\mathrm{T}} \rangle_2$, and the measurements of trijet production as function of Q^2 and $\langle p_{\mathrm{T}} \rangle_3$. The correlations which arise from the unfolding process due to migrations because of detector effects and the finite resolution are considered in the fit by the covariance matrix. We assume that the measured observables Q^2 and p_{T} can be directly associated to the renormalization scale. Therefore, individual bins of the measurement can be directly related to the scale of the effective coupling. The goal of this α_s-determination is the simultaneous determination of α_s from all three jet measurements. We also extract the value of α_s from normalized jet cross sections, where all normalization uncertainties and some other experimental uncertainties cancel partly. The sensitivity to $\alpha_s(M_Z)$ from normalized jet cross sections is only reduced slightly by the next-to-leading order correction to the LO α_s-independent inclusive NC DIS cross section. The experimental systematic uncertainties are considered in the fit as free parameters, whereas uncertainties on α_s from the theoretical predictions are determined separately with a newly developed method for linear error propagation. The theoretical uncertainties on α_s obtained are compared to uncertainties determined with other methods for error propagation.

10.2 The central fit

In this section the 'central value(s)' of $\alpha_s(M_Z)$ are presented. The input data for the determination of α_s have been presented and visually compared to theory predictions in chapter 7. The theory calculations are outlined in section 7.2. For the central fit, the renormalization scale is chosen to be $\mu_r^2 = (Q^2 + p_{\mathrm{T}}^2)/2$, with $p_{\mathrm{T}} = p_{\mathrm{T}}^{\mathrm{jet}}$ for inclusive jets and $p_{\mathrm{T}} = \langle p_{\mathrm{T}} \rangle_2$ for dijet and $p_{\mathrm{T}} = \langle p_{\mathrm{T}} \rangle_3$ for trijet events, and the factorization scale to be $\mu_f^2 = Q^2$. The MSTW2008 PDF set is employed[3], which was determined in NLO at $\alpha_s(M_Z) = 0.118$ [61].

10.2.1 Fit strategy

The strong coupling constant is determined in an iterative least square minimization procedure, colloquially referred to as 'α_s-fit'. For this procedure several assumptions and choices on the theory prediction, the fitting technique and the data treatment must be made. These are outlined and briefly discussed in this section. All relevant ingredients to the α_s-fit are then studied separately in more detail in the successive sections and are compared to the 'central fit' from this section.

[3]LHAPDF filename: 'MSTW2008nlo_asmzrange.LHgrid', PDF member 9.

10.2.2 The least-square method

The strong coupling constant is determined in a comparison of data to theory, where $\alpha_s(M_Z)$ is a free parameter. The agreement of theory and data is estimated using a χ^2 definition of [12, 214]

$$\chi^2 = \vec{p}^{\mathrm{T}} \boldsymbol{V}^{-1} \vec{p} + \mathcal{N} \ , \tag{10.1}$$

where the elements of the vector $\vec{p}$ are the differences in each bin i between the α_s-dependent theory prediction $t_i = t_i(\alpha_s(M_Z))$ and the measurement m_i, i.e.

$$p_i = m_i - t_i \Theta_i \ , \tag{10.2}$$

where Θ_i is the estimated shift from the systematic uncertainties on the theory predictions and defined below. The inverse covariance matrix $\boldsymbol{V}$ includes (statistical) correlations between bins of the measurement.
Systematic uncertainties are accounted for in the estimate either in the covariance matrix $\boldsymbol{V}$ or by introducing nuisance parameters ε_k for each source of systematic uncertainty k. The nuisance parameters are further free parameters in the χ^2 minimization procedure and are included in a penalty term of the form

$$\mathcal{N} = \sum_k^{N_{\mathrm{sys}}} \varepsilon_k^2 \ . \tag{10.3}$$

Three different methods for the inclusion of systematic uncertainties in the definition of χ^2 are considered. These are referred to as *relative* or *absolute correlated systematic uncertainties* or as *uncorrelated systematic uncertainties*. Furthermore, it is possible to split-up each source of systematic uncertainty into these three categories. The *splitting parameters* μ are denoted as μ^{R} or μ^{A} for the fractions which are treated as relative or absolute systematic uncertainties, respectively, and as μ^{U} for the fraction which is treated as uncorrelated systematic uncertainty. The splitting parameters have to fulfill

$$0 \leq \mu^{\mathrm{U}}, \mu^{\mathrm{R}}, \mu^{\mathrm{A}} \leq 1 \quad \text{and} \quad \mu^{\mathrm{U}} + \mu^{\mathrm{R}} + \mu^{\mathrm{A}} = 1 \ , \tag{10.4}$$

such that the size of the total uncertainty is preserved.

A *relative correlated uncertainty* is included in the definition of χ^2 using the nuisance parameters ε_k as Gaussian random variables and by shifting the theory prediction t_i by a factor proportional to $E_{i,k} \simeq \delta_k m_i \cdot \varepsilon_k$, representing the relative size of the uncertainty. The value t_i is 'shifted' by factors Θ_i, defined as

$$\Theta_i^{\mathrm{Normal}} := 1 - \sum_k^{N_{\mathrm{sys}}} E_{i,k} \tag{10.5}$$

for normal distributed uncertainties, where k runs over each source of systematic uncertainty. If the theory prediction t_i is small compared to the size of the uncertainty, this leads to biases in the χ^2-determination or even to negative theory predictions. These biases are avoided for log-normal distributed uncertainties. These are, as a consequence of the central limit theorem, for instance normalization uncertainties and they ensure a constant relative uncertainty [215]. When assuming log-normal distributed uncertainties Θ_i reads [216]

$$\Theta_i^{\mathrm{LogNormal}} := \prod_k^{N_{\mathrm{sys}}} e^{-E_{i,k}} \ . \tag{10.6}$$

This ensures a constant relative uncertainty. Since Θ_i acts on the theory t_i (instead of m_i), the statistical uncertainties in the covariance matrix do not have to be rescaled. The uncertainties considered through nuisance parameters are therefore referred to as 'relative' uncertainty because of their relative dependence on the measured value m_i, and since the factors ε_k are independent of the bin i, the uncertainty is bin-to-bin 'correlated'. The general expression of a relative shift for an asymmetric uncertainty[4] reads

$$E_{i,k} = \sqrt{\mu_k^{\mathrm{R}}} \left(E_{i,k}^{\mathrm{Sym}} + E_{i,k}^{\mathrm{Asym}} \right) \ , \tag{10.7}$$

where $E_{i,k}^{\mathrm{Sym}}$ represents the symmetric part of the up and down error[5]

$$E_{i,k}^{\mathrm{Sym}}(\varepsilon_k) := \frac{\delta_k^+ m_i - \delta_k^- m_i}{2} \varepsilon_k \simeq \frac{|\delta_k^+ m_i| + |\delta_k^- m_i|}{2} \varepsilon_k \ , \tag{10.8}$$

and $E_{i,k}^{\mathrm{Asym}}$ accounts for the asymmetric component by

$$E_{i,k}^{\mathrm{Asym}}(\varepsilon_k) := \frac{\delta_k^+ m_i + \delta_k^- m_i}{2} \varepsilon_k^2 \simeq \frac{|\delta_k^+ m_i| - |\delta_k^- m_i|}{2} \varepsilon_k^2 \ . \tag{10.9}$$

The second term in 10.2.2 and 10.9 is only valid if the up and down errors have always opposite sign in each bin, which is mostly the case in this analysis. In case of symmetric uncertainties, like for instance all normalization uncertainties, the expressions simplify to

$$E_{i,k}^{\mathrm{Sym}} = \delta_k m_i \cdot \varepsilon_k \quad \text{and} \quad E_{i,k}^{\mathrm{Asym}} = 0 \ . \tag{10.10}$$

An *absolute correlated uncertainty* or an *uncorrelated uncertainty* is included in the definition of χ^2 through the covariance matrix by

$$\boldsymbol{V} = \boldsymbol{V}^{\mathrm{stat}} + \sum_k^{N_{\mathrm{sys}}} \left(\mu_k^{\mathrm{U}} \boldsymbol{V}_k^{\mathrm{U}} + \mu_k^{\mathrm{A}} \boldsymbol{V}_k^{\mathrm{C}} \right) \ , \tag{10.11}$$

where $\boldsymbol{V}^{\mathrm{stat}}$ denotes the covariance matrix of statistical uncertainties from the unfolding (it also includes the uncertainty due to the background sources). The matrices $\boldsymbol{V}_k^{\mathrm{C}}$ represent absolute correlated uncertainties and are defined as

$$V_{k,ij}^{\mathrm{C}} = \Delta_k m_i \cdot \Delta_k m_j \ . \tag{10.12}$$

In case of asymmetric uncertainties the values are symmetrized by $\Delta_k m_i = \frac{1}{2}(\Delta_k^+ m_i - \Delta_k^- m_i)$. The matrices $\boldsymbol{V}_k^{\mathrm{U}}$ represent uncorrelated uncertainties and are defined as

$$V_{k,ij}^{\mathrm{U}} = \delta_{ij} (\Delta_k m_i)^2 = \delta_{ij} V_{k,ij}^{\mathrm{C}} \ . \tag{10.13}$$

The uncorrelated definition of the covariance matrix $\boldsymbol{V}^{\mathrm{C}}$ does not involve any bin-to-bin correlations, i.e. all off-diagonal elements are zero. Both, the uncorrelated and correlated

[4] Several systematic uncertainties are determined through simulating the '1σ-size' of the effect in the 'up' and 'down' direction, which yields an 'up'- and 'down' error denoted through the superscripts '+' and '−', respectively.

[5] Remember that $\delta_k \mathcal{R}$ denotes the relative uncertainty of the quantity $\mathcal{R}$ from the error source k, while $\Delta_k \mathcal{R}$, denotes its absolute value, and both are connected through $\delta_k \mathcal{R} = \Delta_k \mathcal{R} / \mathcal{R}$. k is named explicitly, like e.g. k='Norm' or k='JES', if the normalization or jet energy scale uncertainty is meant, respectively.

covariance matrices, contain 'absolute' uncertainties and do not depend on the value of the measured data m_i[6]. Since the theory is in general in good agreement with data, no relevant difference between absolute and relative uncertainties is expected, and we abstain from an error rescaling. However, we will prefer the usage of relative correlated uncertainties over absolute uncertainties to avoid this bias. The covariance matrix of the statistical uncertainty $\boldsymbol{V}^{\text{stat}}$ is always taken from the unfolding and includes all correlations which arise from detector effects and from statistical correlations between the measurements.
For the determination of α_s the χ^2 in equation 10.1 is minimized in an iterative fitting procedure with α_s and ε_k as free parameters. The minimization algorithm MIGRAD is used as implemented in the MINUIT program [217, 218][7]. The errors on the fit parameters are determined as the square root of the diagonal elements of the inverse of the matrix of second derivatives of χ^2 (the 'Hesse'-matrix). These uncertainties take into account all parameter correlations and also the correlations between the free parameters are accessible. The degrees of freedom of the fit n_{dof} is calculated as the number of bins N_{bins} minus the number of free parameters in the fit, where, however, the nuisance parameters do not contribute, and hence

$$n_{\text{dof}} = N_{\text{bins}} - 1 \ . \tag{10.14}$$

The agreement between the theoretical predictions and the data is quantified by the ratio of χ^2 and the number of degrees of freedom χ^2/n_{dof}. One assumes that the errors are Gaussian random variables with standard deviation Δm_i and average m_i. If the theory describes the data within the experimental uncertainty, then the quantity χ can be regarded as a random variable that follows a Gaussian distribution with variance $\chi^2 = n_{\text{dof}}$. The fit can therefore be considered of good quality if $\chi^2/n_{\text{dof}} \approx 1$.

10.2.3 Treatment of systematic uncertainties

In general every source of systematic uncertainties should be studied in such detail that all relevant bin-to-bin correlations become precisely known (for instance between different $p_{\text{T}}^{\text{jet}}$-regions or between $\eta_{\text{lab}}^{\text{jet}}$-regions)[8]. Such a detailed determination, however, is typically not possible, and only the absolute size of the uncertainty in each bin of the measurement is determined. As a compromise, we split the absolute size of each systematic uncertainty into a *relative correlated fraction* ($\mu^{\text{R}} > 0$) and an *uncorrelated fraction* ($\mu^{\text{U}} > 0$), where the absolute size of the uncertainty on each bin is preserved in a quadratic summation.

The split-up of systematic errors into correlated and uncorrelated parts is however not uncritical. Only the treatment as correlated uncertainties is mathematically fully consistent. We already made the assumption that the systematic uncertainties are fully correlated during their determination, when adjacent bins on the generator level bin grid have been recombined for the measurement bin grid after the unfolding. This assumption is certainly reasonable for adjacent p_{T}-bins, where the bin-sizes are small, but may be less valid for

[6] This becomes more obvious, when regarding the alternative notation of the absolute correlated uncertainty. The definition of the absolute correlated uncertainties through $\boldsymbol{V}_k^{\text{C}}$ is exactly equivalent to a notation similar to the relative correlated uncertainties, using nuisance parameters, where, in contrast, the differences p_i are defined as $\tilde{p}_i := m_i\Theta_i - t_i$. This would change the value $m_i \to m_i\Theta_i$ and consequently, also the statistical uncertainty $\boldsymbol{V}^{\text{stat}}$ would have to be rescaled for this modified data.

[7] We use the implementation TMinuit, which is a C++ translation of the original Fortran code and is distributed with the ROOT framework [219].

[8] Such a knowledge would be considered in the fit as multiple correlated uncertainties, yielding a large number of nuisance parameters.

Error treatment

Error source	**Relative fraction μ^{R}**	**Absolute fraction μ^{A}**	**Uncorrelated fraction μ^{U}**	**Uncorrelated between Q^2 bins**
Jet energy scale Δ_{JES}	0.5	0	0.5	
Rem.-cluster energy scale Δ_{RCES}	0.5	0	0.5	
LAr Noise $\Delta_{\mathrm{LArNoise}}$	0.5	0	0.5	
Electron energy $\Delta_{\mathrm{E_e}}$	1	0	0	✓
Electron polar angle $\Delta_{\theta_{\mathrm{e}}}$	1	0	0	✓
Electron ID $\Delta_{\mathrm{ID(e)}}$	1	0	0	✓
Normalization Δ_{Norm}	1	0	0	
Model Δ_{Model}	0.25	0	0.75	✓

Table 10.1: Split-up of systematic uncertainties in the α_s-fit.

instance in the inclusive jet measurement, where bins in different $\eta^{\mathrm{jet}}_{\mathrm{lab}}$-regions have been combined. Moreover, uncorrelated uncertainties add more freedom to the fit than correlated uncertainties, since each bin may vary independently. This decreases the minimum χ^2 and also leads to a decreased experimental uncertainty on the fit parameters from that uncorrelated uncertainty[9].
When systematic uncertainties are treated as uncorrelated in a fit, it is assumed that all bins are uncorrelated to each other, which however is not always true for measurements, where there are statistical correlations present between bins. On the one hand, positive statistical correlations arise from the simultaneous measurement of multiple quantities in one event. For instance, one single event contributes to the inclusive NC DIS, inclusive jet, dijet and trijet measurement, hence it is obvious that the uncertainties on the NC DIS observables (i.e. y and Q^2) have to be correlated. On the other hand, positive statistical correlations are also present in the inclusive jet measurement, where more than one jet may be found in the same event. In such a case, the systematic uncertainty on the individual jet can be different, since e.g. these jets are naturally located in different regions of the detector or may have different $p^{\mathrm{jet}}_{\mathrm{T}}$. Therefore, the correlations of systematic uncertainties have not to be mistaken with the statistical correlations which are represented through $\boldsymbol{V}^{\mathrm{stat}}$.

The error treatment is discussed individually for each source of systematic uncertainty k in order to select the uncorrelated fraction μ^{U}_k. The splitting parameters for each source of systematic uncertainty are listed in table 10.1 and are motivated as follows:

- The uncertainties on the reconstruction of the HFS (Δ_{JES},Δ_{RCES},$\Delta_{\mathrm{LArNoise}}$) are treated as 50 % correlated and as 50 % uncorrelated. The jet energy scale uncertainty Δ_{JES} accounts for the uncertainty on the energy of calibrated clusters within a laboratory-frame jet. These calibration constants are functions of $\eta^{\mathrm{jet}}_{\mathrm{lab}}$ and $p^{\mathrm{jet}}_{\mathrm{T,lab}}$ and account

[9] In a primitive example, where the systematic uncertainty is a normalization uncertainty (i.e. the relative uncertainty is identical in every bin) and the sensitivity to the fit parameters are identical in every bin, the reduction of the uncertainty on the fit parameters compared to a correlated treatment is proportional to $\sqrt{N_{\mathrm{bins}}}$. In a correlated treatment, the fit literally shifts all bins simultaneously by the size of the uncertainty, whereas for uncorrelated uncertainties the fit shifts each single data point once by 1σ and adds these effects on $\Delta\alpha_s$ up in quadrature.

for the fact that the true value of these constants may deviate differently from the determined value in different $p^{\text{jet}}_{\text{T,lab}}$ regions 50 % of Δ_{JES} is treated as uncorrelated. The remaining-cluster energy scale uncertainty Δ_{RCES} represents clusters, which are not associated to a calibrated jet. The relevant calibration constants for these clusters may lead to different shifts in different p_{T} (and $\eta^{\text{jet}}_{\text{lab}}$) regions and therefore are treated as 50 % uncorrelated. The uncertainty on the LAr noise Δ_{LArNoise} is also assumed to be 50 % uncorrelated to account for differences in the individual LAr segments. Previous analyses treated the uncertainty on the hadronic final state as 50 % uncorrelated [74, 102, 103] or as 75 % uncorrelated [79, 220].

- The uncertainties on the reconstruction of the scattered electron ($\Delta_{\text{E}_\text{e}}$, Δ_{θ_e} and $\Delta_{\text{ID(e)}}$) are treated as fully correlated, but only for bins belonging to the same Q^2-bin. Thus, in fits to the multiple jet measurements, every event in the same Q^2 bin receives the same shift from non-zero nuisance parameters, since these events have the same reconstructed scattered lepton. For the six different Q^2-bins of this measurement the scattered lepton is typically found in well separated detector regions, and therefore these bins are considered to be uncorrelated[10]. Compared to previous H1 analyses [102, 103, 165][11] this error split-up is more consistent with studies from the inclusive NC DIS measurements.

- The uncertainties on the normalization from the trigger uncertainty Δ_{Trig}, the Track-cluster link uncertainty Δ_{TrkCl} and the luminosity measurement Δ_{Lumi} are added in quadrature to a combined *normalization uncertainty* Δ_{Norm} and are treated in the fit as fully correlated between all bins.

- The model uncertainty is treated as 75 % uncorrelated and as 25 % correlated, whereby the correlated fraction is treated as uncorrelated between Q^2-bins and correlated between adjacent p_{T}-bins. This split-up accounts for the lack of knowledge of the 'true' model, but the small correlated fraction could principally favor either the Rapgap or the Django predictions. This correlated fraction also accounts for the statistical correlations between the different observables. The large uncorrelated fraction is also motivated by the observation that the model uncertainty shows large fluctuations between adjacent bins (whose origin could not be found), which are assumed to be unphysical. In previous analyses, the model uncertainty was treated as uncorrelated [74, 102], and therefore this split-up is more conservative, i.e. it will thus increase the total experimental error on α_s.

- The correlation of the statistical uncertainties on detector level are respected in the fit through the covariance matrix and has been propagated analytically in the unfolding process.

The individual contribution of the systematic uncertainties on the experimental uncertainty on α_s are studied in section 10.2.7, where also the effect of the uncorrelated fractions $\mu^{\text{U}} > 0$ is studied.

[10] The correlations within bins of the identical Q^2-bin and the uncorrelated treatment of different Q^2-bins is technically realized by one nuisance parameter and a dedicated vector of uncertainties $\delta^q_k\sigma$ for each Q^2-bin q. The bin-wise uncertainty is given by $\delta^q_k\sigma = \delta_k\sigma$, if the bin i belongs to the Q^2-bin q, and otherwise $\delta^q_k\sigma = 0$. Thus, the uncertainty is preserved by $\sum_q^{Q^2\text{-bins}} \delta^q_k\sigma = \delta_k\sigma$.

[11] In [103] the electron angle was uncorrelated and $\Delta_{\text{E}_\text{e}}$ and $\Delta_{\text{ID(e)}}$ half-half. In [102, 165] the electron energy was assumed to be $\mu^{\text{U}}_{\text{E}_\text{e}} = 75\,\%$ uncorrelated and Δ_{θ_e} fully correlated.

α_s from double-differential jet cross sections

Measurement	Best fit ± exp. uncertainty	n_{dof}	χ^2/n_{dof}
Inclusive jet σ_{jet}	0.1172 ± 0.0022	23	1.380
Dijet σ_{dijet}	0.1129 ± 0.0024	23	1.274
Trijet σ_{trijet}	0.1170 ± 0.0018	15	0.850
Normalized inclusive jet $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	0.1180 ± 0.0010	23	1.341
Normalized dijet $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	0.1138 ± 0.0011	23	1.420
Normalized trijet $\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	0.1180 ± 0.0014	15	0.693

Table 10.2: Values of α_s obtained in a fit of NLO calculations including hadronization corrections to absolute and normalized double-differential jet cross sections. The experimental uncertainty includes the statistical uncertainty and all uncertainties from the systematic sources.

10.2.4 Fit to individual data points and sub-sets of data points

In a first step the strong coupling is extracted from each individual data point of the absolute jet cross sections. Since one free parameter is being determined from one input point, the 'fit' has no degree of freedom and therefore always $\chi^2 = 0$. As a consequence, also the nuisance parameters cannot be determined and any systematic shift, which may arise from the systematic uncertainties, is zero. The values are compared to values obtained for fits to each Q^2-bin of a measurement in figure 10.1. For comparison, also the values of α_s from the three jet measurements are shown, which will be derived in the next section. The α_s-values from the individual data points show in general good agreement within the experimental uncertainty. Only a few bins with large uncertainties deviate from the result from individual Q^2-bins, which are mostly dominated by bins at low p_{T} due to their larger statistical precision. The fits to individual Q^2-bins show an overall very good compatibility and no systematic tension in any phase space region is observed. Although, the statistical precision of the trijet measurement is small, the sensitivity to α_s is higher, which results in small experimental uncertainties on α_s compared to the dijet measurement.

10.2.5 Fit to single measurements

Before α_s can be determined in a fit to all jet measurements, the consistency of theory and data for each measurement has to be confirmed. An α_s-fit is performed individually to the absolute inclusive jet σ_{jet}, the dijet σ_{dijet} and trijet measurement σ_{trijet} (all measurements double-differentially as function of Q^2 and p_{T}) and to their normalized, i.e. w.r.t. to the inclusive NC DIS measurement σ_{NC}, jet cross sections. The results for α_s obtained are presented in table 10.2 and can be summarized as follows:

- The quality of the fits is judged by χ^2/n_{dof} and is between 0.7 and 1.4. These are considered to be reasonable values, especially, since the full covariance matrix including the correlations from the detector correction is employed. The χ^2/n_{dof} for σ_{trijet} and $\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$ is below 1, which is quite small and could point towards an overestimation or a non-optimal treatment of the systematic uncertainties. However, since also the statistical uncertainties on the trijet observables are large, this is not considered to be worrisome.
  The χ^2-parabolas as function of α_s are shown in figure 10.2, where either the nuisance parameters are fixed to their optimal value or χ^2 is minimized with ε_k as free parameters.

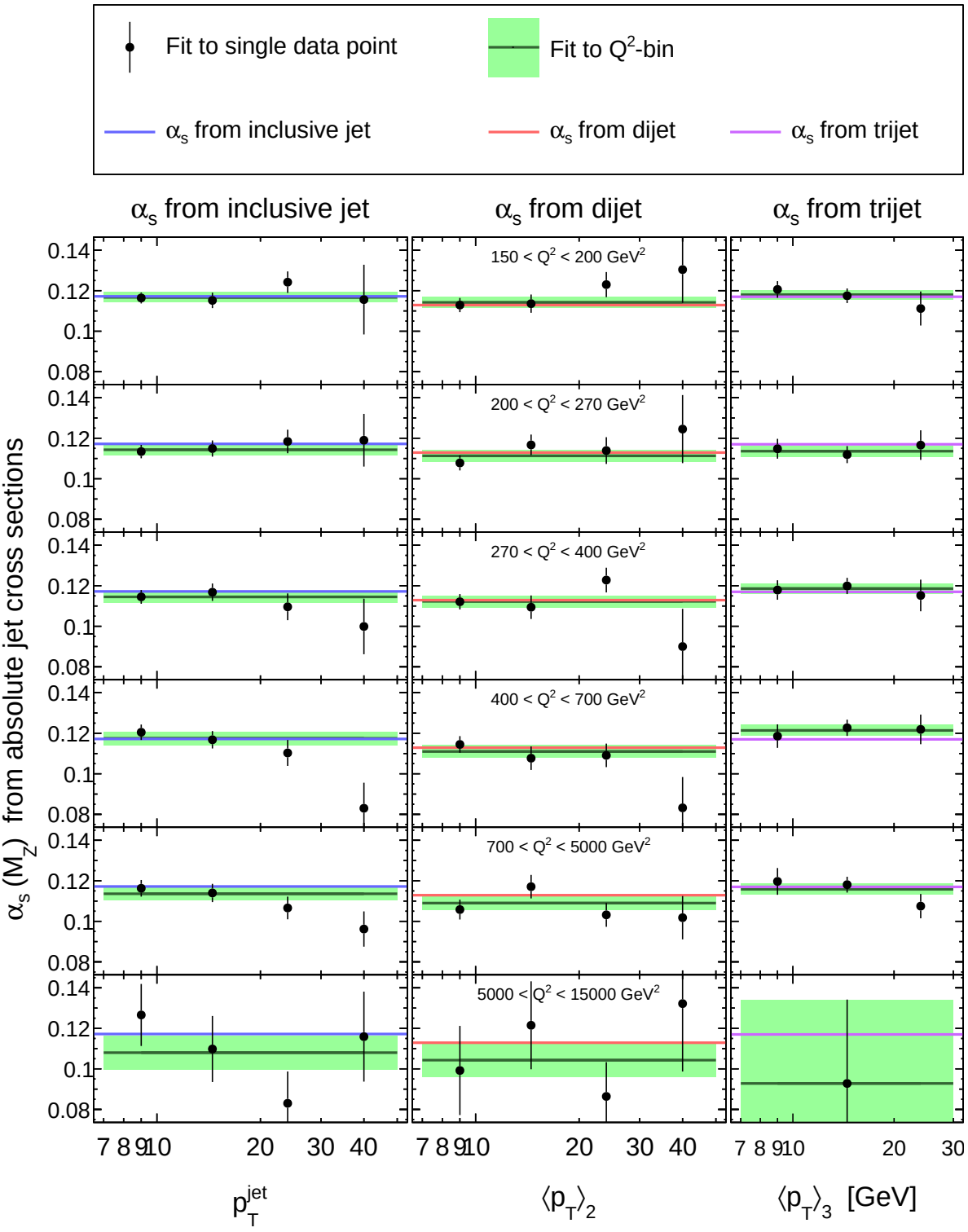


Figure 10.1: Values of α_s from every individual data point of the absolute jet cross sections. The values of α_s are compared to α_s from a simultaneous fit to the data points of the corresponding Q^2-bin (green line and band) and to the α_s value determined from a fit to all the data points of the jet measurements (colored lines). The size of the uncertainty bar indicates the total experimental uncertainty, when α_s is derived from one single data point. The green band indicates the experimental uncertainty on α_s from fits to individual Q^2-bins.

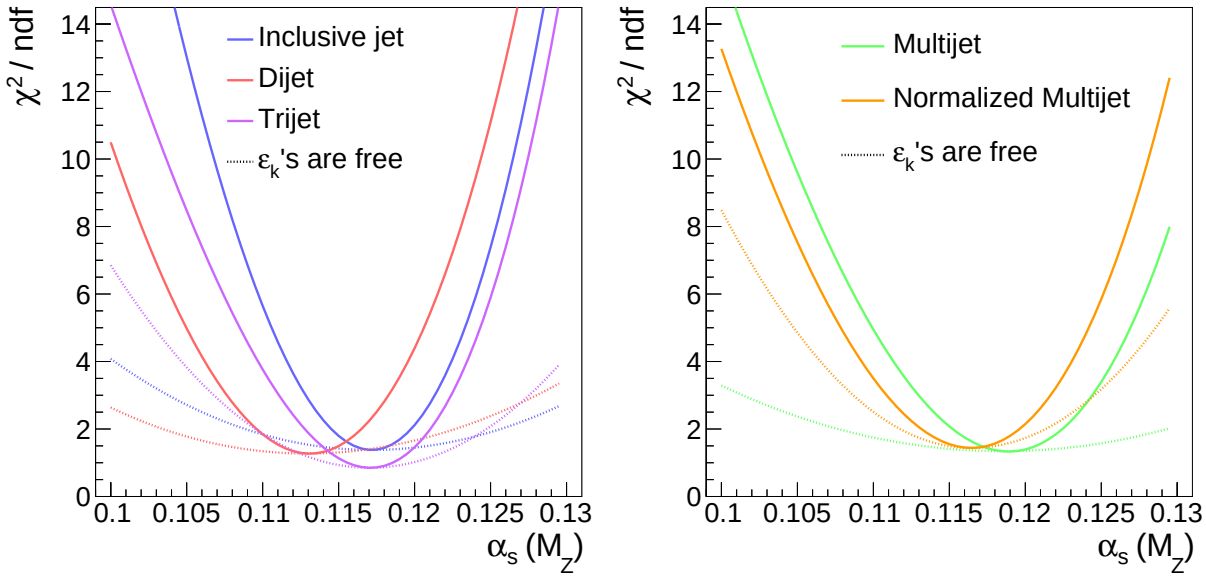


Figure 10.2: χ^2/n_{dof}-parabolas as function of α_s in the α_s-fit to single absolute jet measurements (left) and to multijet and normalized multijet measurements (right). The full lines show the χ^2-dependence when all nuisance parameters are kept fixed at their optimal values. The dotted lines indicate the χ^2-dependence of α_s, if the χ^2 is minimized with free nuisance parameters for each value of α_s.

- The three values obtained for α_s from the different absolute jet cross sections are in agreement within the experimental uncertainty. The inclusive jet and trijet measurements yield almost identical values of α_s, although they are based on different theory calculations[12]. The α_s-value obtained from σ_{dijet} is smaller by approximately 0.004 compared to the previously mentioned values.

- The α_s-values extracted from the normalized jet cross section measurements are in general well compatible with the values from the absolute cross sections within the experimental uncertainties. Also here, the result of the dijet observable $\frac{\sigma_{\mathrm{dijet}}}{\sigma_{\mathrm{NC}}}$ results in a lower value of α_s than from $\frac{\sigma_{\mathrm{jet}}}{\sigma_{\mathrm{NC}}}$ or $\frac{\sigma_{\mathrm{trijet}}}{\sigma_{\mathrm{NC}}}$. Due to the reduced experimental uncertainty from the cancellation of normalization uncertainties, the α_s-value from $\frac{\sigma_{\mathrm{dijet}}}{\sigma_{\mathrm{NC}}}$ is not compatible with the values from $\frac{\sigma_{\mathrm{jet}}}{\sigma_{\mathrm{NC}}}$ or $\frac{\sigma_{\mathrm{trijet}}}{\sigma_{\mathrm{NC}}}$ within the experimental uncertainty. The α_s-values extracted from normalized jet cross sections are typically larger by 0.001 compared to those from the absolute cross sections. This points towards an overestimated normalization, e.g. from the luminosity, as it was also observed in an other H1 analysis of HERA-II data [201].

- The experimental uncertainties express the uncertainty from limited statistical precision, from the uncertainty on the background subtraction, and from all sources of systematic uncertainties, including also the uncertainty due to different MC generators for the determination of the unfolding matrix. For the absolute cross sections a precision of $\Delta_{\mathrm{exp}}\alpha_s \simeq 0.002$ and for the normalized cross sections of $\Delta_{\mathrm{exp}}\alpha_s \simeq 0.001$ is obtained. The smaller value for normalized cross sections is due to the full cancellation of normalization uncertainties in the cross section ratio. The uncertainty on α_s from σ_{trijet} is smaller than from σ_{jet} or σ_{dijet}, since the trijet NLO calculations

[12]The inclusive jet calculations are of order $\mathcal{O}(\alpha_s)$ in leading-order, whereas the trijet calculations are of order $\mathcal{O}(\alpha_s^2)$ in LO.

are of $\mathcal{O}(\alpha_s^2)$ in leading-order and hence have a higher sensitivity to α_s, which compensates the smaller statistical precision. Due to the larger sensitivity of σ_{trijet} to α_s also the correlation between α_s and $\varepsilon_{\text{Norm}}$ is reduced, which accounts for the large uncertainty on α_s from σ_{jet} and σ_{dijet}. In the α_s-fits to normalized cross sections the normalization cancels, and therefore the observables σ_{jet} and σ_{dijet} are preferred over σ_{trijet} due to the higher statistical precision.

- The nuisance parameters ε_k of the correlated systematic uncertainties are equally distributed around zero and $|\varepsilon_k| \lesssim 1$ mostly and with uncertainties of approximately ± 0.9. The nuisance parameter of the normalization uncertainty $\varepsilon_{\text{Norm}}$ is $+0.5$ in the fit to σ_{jet} and σ_{dijet} and $+0.04$ to σ_{trijet}. The nuisance parameters ε_k and uncertainties Δ_k are defined such that a positive nuisance parameter can usually[13] be interpreted as an 'up-shift' of data. Therefore, the positive nuisance parameter for the normalization uncertainty, together with the observation that the α_s-values from normalized jets are smaller than from absolute jet cross sections, can be interpreted such that the luminosity is overestimated by at least 1 to 2 %.
  The correlation coefficient of $\varepsilon_{\text{Norm}}$ and α_s is 0.87[14]. Due to this large correlation, the α_s-value obtained in the fit is strongly dependent on the normalization, which is represented by the term $E_{\text{Norm}} = \varepsilon_{\text{Norm}} \cdot \delta_{\text{Norm}} m$. As a consequence of the large correlation, the experimental uncertainties on α_s from absolute jet cross sections are comparatively large when compared to the ones from normalized jet cross sections. The fit to normalized jet cross sections is of course independent of this correlation.

The α_s-values obtained from the normalized jet cross sections are smaller by 0.0015 for $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$ and 0.004 for $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$ than in a previous analysis of HERA-I and HERA-II data [74][15]. The α_s-values from $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$ are smaller by 0.0017 and are larger for $\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$ by 0.0008 compared to [102], where the same data as in [74] was used, but a renormalization scale identical to this analysis was chosen. The χ^2-values are comparable to the ones in [102].
The α_s-values obtained from the inclusive jet measurements are compatible within the experimental uncertainty to values determined from the HERA-I data in a similar phase space [103][16]. Also in this previous independent analysis, the value of α_s from normalized jet cross sections was found to be larger than the one from the absolute jet cross sections. Although some differences between the values of α_s from the dijet measurement compared to the inclusive jet and trijet measurements are observed, we consider these values to be consistent and will proceed with α_s-fits to multiple measurements simultaneously.

[13] If the systematic uncertainty has a complicated structure with different signs for adjacent bins, then this is no longer true.

[14] It is 0.6 in the fit to σ_{trijet}.

[15] The values of α_s in [74] are $0.1195 \pm 0.0010\,(\text{exp})$, $0.1170 \pm 0.0009\,(\text{exp})$ and $0.1183 \pm 0.0013\,(\text{exp})$ for the normalized inclusive, normalized dijet and normalized trijet measurement, respectively. The experimental differences of this analysis to [74] are an improved hadronic energy calibration, an updated luminosity measurement, an enlarged phase space in $\eta^{\text{jet}}_{\text{lab}}$, the regularized unfolding procedure, and the trijet measurement is performed now double-differentially. The previous analysis further includes data taken during the HERA-I period. The relevant differences in the theoretical calculations in [74] to this analysis are the calculation of σ_{NC}, which was performed using the DISENT program [96], the renormalization scale μ_r, which for the calculation of the dijet and trijet cross sections was set to $\mu_r^2 = Q^2$ and for the inclusive jet measurements to $\mu_r^2 = p_{\text{T}}^2 + Q^2$ [221], an older evolution code for $\alpha_{em}(\mu)$ compared to this analysis, and a different PDF set (CTEQ6.5M) [222].

[16] The values in [103] from the inclusive jet measurement is $\alpha_s = 0.1179 \pm 0.0024\,(\text{exp})$ and for the normalized inclusive jet measurement it is $\alpha_s = 0.1193 \pm 0.0014\,(\text{exp})$

α_s from multijet jet cross sections

Measurement	Best fit ± exp. uncert.	n_{dof}	χ^2/n_{dof}
$[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$	0.1162 ± 0.0022	47	1.401
$[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$	0.1182 ± 0.0017	39	1.152
$[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	0.1174 ± 0.0018	39	1.341
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$	0.1160 ± 0.0009	47	1.410
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1175 ± 0.0008	39	1.193
$[\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1149 ± 0.0009	39	1.614
Multijet $[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	0.1185 ± 0.0017	63	1.345
Normalized multijet $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1165 ± 0.0008	63	1.439

Table 10.3: α_s-values from combined fits to two or all three absolute as well as normalized double-differential jet cross sections. The experimental uncertainty includes the statistical uncertainty and all uncertainties arising from the systematic uncertainties.

10.2.6 Simultaneous α_s-determination from multijet cross sections

Simultaneous fits to respectively two out of the three jet measurements and to all three jet measurements are performed. For these *combined α_s-fits*, the statistical correlations between the single measurements are taken into account by the statistical covariance matrix, which was determined on detector level and propagated through the regularized unfolding procedure (c.f. sections 5.4.1 and 6.3). The systematic uncertainties are treated using the error splitting from table 10.1, and the application of the statistical correlations between the measurements have already been considered in the choice of the correlated fractions. These have been chosen to be larger than in previous analyses [74, 79, 102, 220].
The values for α_s obtained in the various combined fits are presented in table 10.3. The values from fits to respectively two absolute jet measurements are consistent within the experimental uncertainties. It is surprising that the fits to $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$ and $[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$ predict larger values of α_s than the fits to the individual measurements. Also the fit to $[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$ predicts a larger value of α_s than expected from the average using σ_{jet} and σ_{dijet}. This is explained by the large values of $\varepsilon_{\text{Norm}}$ of 0.9 for $[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$ and $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$ and even of $\varepsilon_{\text{Norm}} = 2.2$ for $[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$. Also other nuisance parameters can become large, for instance $\varepsilon_{\text{JES}} = 1.3$ in the fit to $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$. Since the data are statistically correlated, the fit tries to reduce the tension between the measurements with large nuisance parameters, since there is only little freedom from the covariance matrix due to the statistical correlations. As a result of the large positive nuisance parameter for the normalization uncertainty the value of α_s increases[17]. Also the χ^2/n_{dof}-values increase compared to fits to individual measurements. This reflects the tension between the measurements, however, it cannot be figured out, if the dominant tension is between $[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$, $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$ or $[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$. The total experimental uncertainty in fits to two measurements is almost the same as the experimental uncertainty of the single measurement with the smaller uncertainty. This is explained by the statistical correlations, since only little additional uncorrelated informa-

[17]The nuisance parameter of the normalization can easily become large compared to other nuisance parameters because of its large correlation to α_s (of $\rho_{\varepsilon_{\text{Norm}}\alpha_s} \approx 0.9$) which can compensate this shift for most of the data points, since only the contribution from the penalty term in the χ^2-calculation becomes relevant.

tion is added when two measurements are considered simultaneously.

The fits to respectively two normalized jet measurements yield values of α_s which are close to the average value of α_s from the single normalized measurements. The nuisance parameters are always below 0.5, except for the fit to $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$, where $\varepsilon_{\text{JES}} = 1.0$, but its correlation to α_s is only 0.2. The fit to $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$ yields a very consistent result with $\chi^2/n_{\text{dof}} = 1.2$, since already the individual measurements predict a similar value of α_s. The combined fits including $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$ result in large values of χ^2/n_{dof}, which illustrates that the fit cannot disentangle the discrepancies between the measurements which became already visible in the differing α_s from the individual measurements. The measurement $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$ seems to be even more inconsistent with $\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$ than with $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$, because of its large value of $\chi^2/n_{\text{dof}} = 1.6$.

The combined fit to all three absolute jet cross sections, denoted *multijet*, yields a value of $\alpha_s = 0.1185 \pm 0.0017$ with $\chi^2/n_{\text{dof}} = 1.3$. The fit to normalized multijet measurements yields $\alpha_s = 0.1165 \pm 0.0008$ with $\chi^2/n_{\text{dof}} = 1.4$. The χ^2-parabolas are displayed in figure 10.2 and show a distinct minimum at the values found. In the fit to multijet cross sections the nuisance parameter of the normalization uncertainty becomes $\varepsilon_{\text{Norm}} = 1.9$. Since this value shifts the data up, this explains the larger α_s-value from the multijet cross sections compared to the one from the normalized multijet cross sections and since this nuisance parameter is highly correlated with α_s in the fit ($\rho_{\varepsilon_{\text{Norm}}\alpha_s} \approx 0.8$) it cannot be reasonably constrained in the minimization procedure. In turn, this also explains the higher experimental uncertainty on α_s, compared to the value from the normalized multijet cross sections.
The experimental uncertainty of the α_s-values from the multijet measurement is not decreased compared to the smallest value from respectively two measurements alone, which is always the case without the dijet measurement, i.e. for $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$ and $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$, since the dijet measurement is statistically a full subsample of the inclusive jet measurement and hence does not increase the statistical precision. On the other hand, also the trijet data are a subsample of the inclusive jet data, but this measurements has an increased sensitivity to α_s due to the $\mathcal{O}(\alpha_s^2)$ sensitivity already in leading order. The fits to the inclusive jet and trijet data, i.e. without the dijets, therefore already achieve the maximum sensitivity from the theoretical and from the experimental perspective[18].

10.2.7 Experimental uncertainties on α_s

The values of the relevant experimental uncertainties are studied seperately in table 10.4. These values are derived by linear error propagation[19]. The normalization uncertainty accounts for the largest contribution to the experimental uncertainty for absolute cross sections. Besides its value of $\delta_{\text{Norm}} m_i = 2.95\,\%$, the uncertainty on α_s is increased due to the correlation of the normalization with α_s in the fit.

[18] If systematic uncertainties are treated as uncorrelated (which is in general inconsistent for correlated quantities), this gives an unphysical increase of the experimental precision, due to the neglect of the statistical correlations. This becomes also obvious for the fits of $[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$ and $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$ which achieve almost the identical precision as the inclusive jets individually. The difference of $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$ to $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$ is such an unphysical increase of the uncorrelated treatement of some of the systematic uncertainties, which, however, is small. Hence, these results prove that our splitting of systematic uncertainties is proper.

[19] i.e. using equation 10.15 and the splitting factors from table 10.1.

Experimental uncertainties on α_s $\times 10^4$

Measurement	$\Delta_{\text{exp}}\alpha_s$	$\Delta_{\text{Norm}}\alpha_s$	$\Delta_{\text{RCES}}\alpha_s$	$\Delta_{\text{JES}}\alpha_s$	$\Delta_{\text{Model}}\alpha_s$	$\Delta_{\text{exp}}^{(\mu_k^{\text{R}}=1)}\alpha_s$
σ_{jet}	22.3	18.7	5.7	6.5	5.3	23.1
σ_{dijet}	23.6	19.7	7.6	6.1	6.0	25.1
σ_{trijet}	18.4	11.4	8.4	5.8	4.5	21.0
$\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	10.0	–	2.9	5.5	5.6	14.2
$\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	10.7	–	3.8	2.3	5.8	16.3
$\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	13.8	–	7.1	4.8	5.8	19.6
$[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	17.0	10.7	8.5	4.8	2.1	16.4
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	8.1	–	3.5	2.4	4.6	11.2

Table 10.4: The total experimental uncertainty on α_s from fits to different observables and the contributions from the most relevant uncertainties are shown. These are the normalization uncertainty, the uncertainties on the reconstruction of the HFS (Δ_{RCES} and Δ_{JES}) and the model uncertainty. The last column illustrates, for comparison, the size of the total experimental uncertainty, if all systematic uncertainties would be treated as relative correlated uncertainty.

The uncertainty on the value of α_s due to the reconstruction of the HFS is an important source of uncertainty. Both, the JES and RCES uncertainty contribute with a comparable size to the total uncertainty, where the contribution from RCES is sometimes larger, since this uncertainty is more distinct for bins at low scales, which have more impact on the value of α_s due to the higher statistical precision and an increased sensitivity on α_s.
The contribution of the model uncertainty Δ_{Model} on α_s is approximately 0.0005 to 0.0007[20]. The model uncertainty is mostly larger for α_s-values from normalized jet cross sections than from absolute jet cross sections, since the size of $\delta_{\text{Model}}\sigma$ is larger for normalized than for absolute jet cross sections. This is due to the error propagation of the model uncertainty for the normalized jet cross sections, where all bins have been considered as uncorrelated (c.f. section 7.4.2.3). If the unfolding would have eliminated the model uncertainty, the total uncertainty on the α_s-value from normalized multijets would be reduced by approximately 0.00013, whereas the reduction for absolute jet cross sections would be negligible.

For comparisons, the total experimental uncertainty on α_s is shown in the last column of table 10.4, if all systematic uncertainties would be treated as relative correlated uncertainties in the fit (i.e. $\mu_k^{\text{R}} = 1$). This would increase the experimental uncertainty and also yield different results for α_s. Furthermore, the χ^2/n_{dof} increases drastically, especially when multiple jet measurements are fitted simultaneously[21]. Since the observables differ for the inclusive jet, dijet and trijet measurements, their theoretical predictions may suffer from different inaccuracies (e.g. different corrections beyond NLO). In these combined fits, the fits do not have the corresponding relevant degrees of freedom to disentangle these (theoretical) differences, since the data are statistically correlated. Possible discrepancies between the individual measurements (which arise most likely from non-experimental sources) are dealt with in the fit by unphysically large shifts of the nuisance parameters. For example, in the fit to multijet cross sections the minimization finds $\varepsilon_{\text{Norm}} = 2.3$ and

[20] Unfortunately, within the scope of this thesis, the differences between Rapgap and Django could not be located, which cause the differences in the migration matrix and in the radiative corrections.

[21] The fits to $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$, $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$ and $[\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$ yield values of $\chi^2/n_{\text{dof}} = 1.8$, 1.5 and 2.5, respectively. The fit to multijet (normalized multijet) cross sections yields $\chi^2/n_{\text{dof}} = 1.8$ (2.1).

$\varepsilon_{E_e} = 1.6$. Several experimental arguments for a reasonable splitting into correlated and uncorrelated fractions are given in the previous section.

10.2.8 The dependence of α_s on the MC generators

The measurement is corrected for detector and radiative effects using the averaged predictions of Rapgap and Django. Also the hadronization corrections are obtained from the average of the correction factors predicted. If the measurements would be corrected for these three effects using only one generator, the result on α_s is for most of the fits consistent with the central value within the model uncertainty[22]. However, there are some peculiarities. For the inclusive jet and dijet data the χ^2/n_{dof} values are smaller, but for the trijet data they are larger, for data corrected with Django instead of with Rapgap. The fit to inclusive jet cross sections yields a value of $\alpha_s = 0.1159$ from Django corrected data and 0.1180 from Rapgap corrected data with $\chi^2/n_{\mathrm{dof}} = 1.28$ and 1.56, respectively. On the other hand, the α_s from $\frac{\sigma_{\mathrm{jet}}}{\sigma_{\mathrm{NC}}}$ yields for both models consistent values of $\alpha_s = 0.1180$ and 0.1181.
The results on α_s from fits to multijet cross sections differ only by ≈ 0.0002 for the two different models and are therefore well within the model uncertainty.

10.3 Theoretical uncertainties of α_s

All experimental uncertainties on α_s are represented by the total experimental uncertainty $\Delta_{\mathrm{exp}}\alpha_s$. In this and the following sections, we will determine uncertainties on α_s from the limited knowledge of the theory calculations. These are uncertainties due to the truncation of the perturbative series and are meant to represent the missing higher orders, uncertainties of the hadronization corrections and uncertainties of the PDFs. The PDF dependence is studied separately in more detail.

10.3.1 Method of error propagation of theoretical uncertainties on α_s

10.3.1.1 Linear error propagation

Following the formula for linearized error propagation (c.f. equation 7.12 or [12]) the uncertainty on the fit result α_s from any quantity τ can be generally expressed by

$$\Delta_\tau \alpha_s := \left.\frac{\partial \alpha_s}{\partial \tau}\right|_{\alpha_0} \cdot \Delta\tau \ , \tag{10.15}$$

where $\Delta\tau$ denotes the absolute uncertainty of the (theoretical) quantity τ and the partial derivative represents the sensitivity of α_s on the quantity τ, which has to be determined at the value of the central fit result α_0.

[22] For this study, the migration histogram, which corrects for detector and radiative effects, was determined only from one MC generator, and also the hadronization corrections have been determined only from this generator. The model uncertainty, which also enters the fit, is calculated from the alternative migration matrix from the other MC generator, while this uncertainty was divided by 2 to have a value consistent with the average cross sections.

Let τ now be the theory prediction used in the fit, represented by the vector $\vec{t}$ with uncertainties on each bin of the size of Δt_i. The uncertainty $\Delta_t \alpha_s$ from this source can then be either expressed by an uncorrelated error propagation with $\tau = t_i$ (i.e. the uncertainties Δt_i are independent of adjacent bins), or alternatively, where the uncertainties Δt_i are fully correlated and hence $\tau = \vec{t}$ (i.e. the uncertainty Δ arises from a well-defined source and is orthogonal to all other sources of uncertainty on t). Since the uncertainties on the theoretical predictions are only an estimate, and the correlations are unknown or even not definable, we allow the uncertainty Δt to be split into a correlated and an uncorrelated fraction using two splitting parameters μ^{C} and μ^{U}, respectively. These have to fulfill $\mu_t^{\mathrm{C}} + \mu_t^{\mathrm{U}} = 1$. One obtains uncorrelated uncertainties on each bin i of t by

$$\Delta t_i^{\mathrm{U}} := \sqrt{\mu_t^{\mathrm{U}}} \Delta t_i \ , \tag{10.16}$$

and one fully correlated uncertainty, where the bin-wise components read

$$(\Delta \vec{t}^{\mathrm{C}})_i := \sqrt{\mu_t^{\mathrm{C}}} \Delta t_i \ . \tag{10.17}$$

The total uncertainty is preserved in a quadratic sum of both. The uncertainty on α_s from Δt is then determined by linear error-propagation assuming Gaussian errors by

$$\Delta_t \alpha_s = \sqrt{\sum_i^{N_{\mathrm{bins}}} \left(\left. \frac{\partial \alpha_s}{\partial t_i^{\mathrm{U}}} \right|_{\alpha_0} \cdot \Delta t_i^{\mathrm{U}} \right)^2 + \left(\left. \frac{\partial \alpha_s}{\partial \vec{t}^{\mathrm{C}}} \right|_{\alpha_0} \cdot \Delta \vec{t}^{\mathrm{C}} \right)^2} \ , \tag{10.18}$$

where the sum runs over all bins i.

10.3.1.2 The non-linear offset method

Since the value for α_s is determined in an iterative fitting procedure with typically non-trivial dependence on the quantity τ, the partial derivatives have to be determined numerically. We therefore introduce the so-called *offset method*. An absolute uncertainty on α_s may be defined through repeating the α_s-fit with 'offset' quantity by replacing $\tau \to \tau \pm \Delta\tau$ in the calculation of equation 10.1 (where τ could be for instance again $\vec{t}$). The difference to the central value α_0 may then represent an uncertainty

$$\Delta_\tau^{\pm} \alpha_s = \alpha_{(\tau \to \tau \pm \Delta\tau)} - \alpha_0 \ . \tag{10.19}$$

Since the fit has to be repeated with an essentially different theory prediction, the uncertainty obtained, especially when $\Delta\tau$ is large, includes also non-linear terms and can therefore not be considered as a 1σ uncertainty. If the non-linear effects are large, the uncertainty $\Delta_\tau^{\mathrm{offset}} \alpha_s$ may also accidentally vanish [223].

10.3.1.3 The infinitesimal offset method for numerical derivatives

If the size of the offset is becoming infinitesimal, i.e. $\Delta\tau \to \mathrm{d}\tau$, the offset method becomes linear. The offset method can then be employed for the numerical calculation of the partial derivatives by defining

$$\left. \frac{\partial \alpha_s}{\partial \tau} \right|_{\alpha_0} := \alpha_{(\tau \to \tau \pm \mathrm{d}\tau)} - \alpha_0 \ , \tag{10.20}$$

where $\mathrm{d}\tau$ is now only an infinitesimal shift. Therefore the infinitesimal offset method can be considered as a linear equation[23].

[23] The technical implementation estimates $\mathrm{d}\tau$ by $\mathrm{d}\tau := \frac{\Delta\tau}{k}$, where k is a large number, such that $\mathrm{d}\tau$ is getting small. The uncertainty on τ is then expressed by $\Delta\tau = k$.

10.3.1.4 Theoretical uncertainties as additional degrees of freedom

We will compare the values obtained for $\Delta\alpha_s$ from the error propagation to other methods, such as the offset method, which do not assume linear error propagation and hence may be biased. An uncertainty on the theory predictions Δt can alternatively be included as additional systematic uncertainty in the fit (see section 10.2.3). The uncertainty on α_s is then determined by

$$\Delta_t\alpha_s = \sqrt{(\Delta_{\text{exp}}\alpha_s)^2 - (\Delta_{\text{exp}}\alpha_0)^2}\ , \tag{10.21}$$

where the (increased) total experimental uncertainty $\Delta_{\text{exp}}\alpha_s$ is compared to the value from the central fit, i.e. without considering the source of uncertainty Δt. For the inclusion of Δ_t as additional systematic uncertainty, again three distinct opportunities are distinguished by the splitting parameters (μ_t^{R}, μ_t^{A} and μ_t^{U}). Since a relative uncertainty on the theory corresponds to an absolute uncertainty on the data in the χ^2-formula (and vice-versa), the treatment of a 'relative uncertainty on the theory' refers to $\mu_t^{\text{A}} = 1$. Otherwise, the inclusion as nuisance parameters ($\mu_t^{\text{R}} > 0$) can give additional insights in the determination of ε_t in the fit, indicating whether the data are sensitive to this new degree of freedom or not and thus which theory may be preferred.

10.3.2 Uncertainty on α_s due to hadronization corrections

The uncertainty of the hadronization correction Δ_{had} is obtained from the half-difference of the predicted hadronization correction factors of the Lund String model and the Cluster fragmentation model, using the identical parton level from the Sherpa event generator (see section 7.2.3). The uncertainties on α_s are determined from full error propagation following the method from the previous section.
We assume that the uncertainty of the hadronization correction is half correlated and half uncorrelated, and hence the splitting factors are $\mu_{\text{had}}^{\text{U}} = 0.5$ and $\mu_{\text{had}}^{\text{C}} = 0.5$. This accounts for possible differences of the correction in adjacent or distant bins, but also the consistency of the models are considered with the correlated fraction. The values obtained for $\Delta_{\text{had}}\alpha_s$ are presented in table 10.5. The values from the error propagation are compared to values obtained from the offset method $\Delta_{\text{had}}^{\text{offset}}\alpha_s$ and to values, where the uncertainty on the hadronization correction is included as additional systematic uncertainty in the fit.

We observe that the offset method leads to occasionally very asymmetric uncertainties, which may be due to the non-linearity of the method. The nuisance parameter ε_{had} proves that the estimate of hadronization correction and its uncertainty is quite reasonable. If the hadronization corrections are included in the fit as uncorrelated uncertainty $\Delta_{\text{had}}^{\text{Uncor}}\alpha_s$, the size of the uncertainty on α_s decreases as expected (c.f. section 10.2.3). The effect of different split-up parameters for the linear error propagation $\mu_{\text{had}}^{\text{C}}$ and $\mu_{\text{had}}^{\text{U}}$ is of a comparable size as the difference between $\Delta_{\text{had}}^{\text{Cov}}\alpha_s$ and $\Delta_{\text{had}}^{\text{Uncor}}\alpha_s$. The uncertainty $\Delta_{\text{had}}^{\text{offset}}\alpha_s$ is found to be smaller for the normalized cross sections than for the absolute cross sections, although the relative uncertainty $\delta_{\text{had}}\sigma$ is identical.

The values obtained for α_s from a fit to theory predictions t_i, which are derived using c_{had} from either Django or Rapgap (these α_s-values are not listed) are consistent with the central value within the hadronization uncertainty. Compared to the central value of α_s these values differ between 0.0002 and 0.0008, where the values are always smaller (larger) if the hadronization corrections from Rapgap (Django) are applied. The fits using $c_{\text{had}}^{\text{Rg}}$

Uncertainty on α_s from hadronization correction uncertainty $\times 10^4$

Measurement	$\Delta_{\text{had}}\alpha_s$	$\Delta^{\text{offset}}_{\text{had}}\alpha_s$	$\Delta^{\text{Nuisance}}_{\text{had}}\alpha_s(\varepsilon_{\text{had}})$	$\Delta^{\text{Cov}}_{\text{had}}\alpha_s$	$\Delta^{\text{Uncor}}_{\text{had}}\alpha_s$
σ_{jet}	9.6	$^{+10}_{-15}$	12 (0.2)	12	6
σ_{dijet}	7.0	$^{+\ 3}_{-15}$	9 (0.0)	9	4
σ_{trijet}	11.0	$^{+17}_{-13}$	14 (-0.1)	14	3
$\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	7.8	$^{+14}_{-\ 7}$	11 (-0.1)	11	5
$\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	5.8	$^{+\ 9}_{-\ 7}$	9 (0.0)	10	6
$\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	10.8	$^{+16}_{-13}$	15 (-0.3)	15	6
$[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$	9.1	$^{+10}_{-14}$	10 (-0.2)	11	5
$[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$	11.7	$^{+18}_{-14}$	14 (0.2)	15	3
$[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	11.0	$^{+20}_{-10}$	13 (-0.4)	15	3
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$	6.6	$^{+18}_{-\ 2}$	10 (-0.2)	10	6
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	8.3	$^{+15}_{-\ 8}$	11 (-0.3)	11	5
$[\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	6.8	$^{+\ 8}_{-13}$	9 (0.1)	11	6
$[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	12.6	$^{+23}_{-11}$	13 (0.8)	15	2
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	7.9	$^{+17}_{-\ 6}$	10 (-0.3)	10	5

Table 10.5: Uncertainty on $\alpha_s \times 10^4$ from the uncertainty of the hadronization corrections. The values from the linear error propagation $\Delta_{\text{had}}\alpha_s$ are compared to values obtained from other methods (see text).

yield always smaller values of χ^2/n_{dof} by approximately 0.1 to 0.3 compared to the ones from Django, while the central values are in between.

If the hadronization correction is derived from non-reweighted Monte Carlo generators and the α_s-fit is repeated with these data (these α_s-values are not listed), the values for χ^2/n_{dof} increase compared to the central result. Especially the value for α_s from fits to trijet measurements differs by up to +0.0014 with larger values of χ^2/n_{dof} up to 1.9. This is because the reweighting constants applied to the trijet topologies are large, since the employed MC generators are only based on 2-parton tree level matrix elements, and 3-jet topologies are only generated in the parton shower. We have ensured that the reweighted MC parton level shows an overall good agreement with the NLO predictions, which is not the case for the non-reweighted MC models.

10.3.3 Uncertainties on α_s due to missing higher orders

In this section, the uncertainty on α_s due to missing higher orders is determined. The immediate dependence of α_s on the renormalization scale factor c_{μ_r} is studied separately in section 10.5.

The corrections to the jet cross section predictions beyond NLO are unknown. The missing higher orders are traditionally estimated by a variation of the scales. Their effect on the jet cross sections is of the size of the next higher order contribution, i.e. $\frac{\mathrm{d}\sigma}{\mathrm{d}\log\mu} \propto \mathcal{O}(\alpha_s^{n+1})$ (see section 2.1.4.3). By convention the scales (i.e. the renormalization scale μ_r and the factorization scale μ_f) are varied by a factor of 2 'up' and 'down' around the nominal scale choice μ_0, corresponding to a variation of the scale factor $c_\mu := \log_2 \frac{\mu}{\mu_0}$ of $-1 \leq c_\mu \leq 1$. By

that it also accounts for a different choice of the functional form of the scale (e.g. $\mu_r = p_{\mathrm{T}}^{\mathrm{jet}}$ or $\mu_r = Q$)[24].

In many analyses, the uncertainty on α_s from missing higher orders are determined by the offset method. The biases of this non-linear method have already been illustrated in the previous sections, and therefore we want to apply the linear method of error propagation instead.

The uncertainty on α_s from a variation of the scales can be expressed by

$$\Delta_\mu \alpha_s = \frac{\partial \alpha_s}{\partial \mu} \Delta\mu \approx \frac{\partial \alpha_s}{\partial t} \frac{\partial t}{\partial \mu} \Delta\mu. \tag{10.22}$$

The dependence of the NLO calculation on the scale choice and the scale variations have already been studied in detail in sections 2.1.4 and 2.1.5. We have observed that in some phase space regions, the quantity $\left.\frac{\partial t}{\partial \mu}\right|_{\mu_0}$ may vanish, if the nominal scale μ_0 is near the value suggested by the 'principle of minimal sensitivity'[25]. As a consequence, a meaningful uncertainty on the theory prediction $\Delta_\mu t$ has to be defined beforehand.

Therefore, we define the uncertainty from the scale dependence $\Delta_\mu \alpha_s$ directly as a function of the uncertainty on the theoretical cross section from scale variations $\Delta_\mu t$ by

$$\Delta_\mu \alpha_s := \left.\frac{\partial \alpha_s}{\partial t}\right|_{\alpha_0} \cdot \Delta_\mu t \ . \tag{10.23}$$

We define the '1σ theoretical uncertainty' on t from scale variations by[26] [37]

$$\Delta_\mu t_i := \max\left(\left|t_{i,(\mu)} - t_{i,(\mu=\mu_0)}\right|\right)_{-1 \le c_\mu \le 1} . \tag{10.24}$$

The maximum absolute deviation from the nominal cross section at μ_0 in the interval $-1 \le c_\mu \le 1$ is taken to avoid unreasonably small uncertainties from possible accidental cancellations [224]. These are present for variations of $c_\mu < 0$ whenever the PMS scale is passed.

The uncertainties $\Delta_\mu t$ are determined independently for the renormalization and the factorization scale μ_r and μ_f. We assume that the contributions from missing higher orders can be expressed by a term proportional to $\Delta_\mu t$ for all bins and by an independent bin-wise contribution proportional to $\Delta_\mu t_i$. Therefore, we choose the correlated and uncorrelated fractions of $\Delta_\mu t$ as half-and-half to be $\mu_\mu^{\mathrm{U}} = 0.5$ and $\mu_\mu^{\mathrm{C}} = 0.5$, respectively. The correlated fraction of $\mu_\mu^{\mathrm{C}} > 0$ also represents the statistical correlations between the measurement in fits to multijets and is therefore a conservative estimate[27]. In case of normalized cross sections, the uncertainty $\Delta_\mu t$ is determined in a simultaneous variation of the scale in the numerator and denominator, where the scale dependence of the inclusive NC DIS calculation is small compared to the scale dependence of the jet cross sections.

The uncertainties from the renormalization and factorization scale variation on the strong coupling $\Delta_{\mu_r}\alpha_s$ and $\Delta_{\mu_f}\alpha_s$, respectively, are presented in table 10.6. The uncertainty from

[24] We have ensured in chapter 2 that a variation by a factor of 2, covers the predictions from different scale choices.

[25] The choice in this analysis of $\mu_r = \sqrt{(Q^2 + p_{\mathrm{T}}^2)/2}$ always yields a non-vanishing infinitesimal scale-dependence (compare figures 2.3, 2.4 and 2.6).

[26] We developed a scale independent method for the fast evaluation of perturbative cross sections, which enables to perform this scan of $-1 \le c_\mu \le 1$ for the first time also for the factorization scale μ_f and even for different scale choices [100].

[27] The uncertainty on α_s from the variation of the renormalization scale $\Delta_{\mu_r}\alpha_s$ is strongly dependent on the choice of the uncorrelated fraction. When choosing an uncorrelated fraction of bin-to-bin correlations of $\mu_\mu^{\mathrm{U}} = 0$, 0.5, 0.75 or 1, the uncertainty on the multijet (normalized multijet) cross sections are $\Delta_{\mu_r}\alpha_s =$ 0.0056 (0.0047), 0.0044 (0.0034), 0.0036 (0.0025) and 0.0017 (0.0010), respectively.

Uncertainty on α_s due to missing higher orders $\times 10^4$

Measurement	$\Delta_{\mu_r}\alpha_s$	$\Delta_{\mu_f}\alpha_s$	$\Delta^{\text{offset}}_{\mu_r}\alpha_s$	$\Delta^{\text{Nuis.}}_{\mu_r}\alpha_s(\varepsilon_{\mu_r})$	$\Delta^{\text{Cov}}_{\mu_r}\alpha_s$	$\Delta^{\text{Uncor}}_{\mu_r}\alpha_s$
σ_{jet}	47.0	6.0	$^{+60}_{-58}$	36 (-0.8)	34	17
σ_{dijet}	37.3	6.7	$^{+42}_{-50}$	32 (-0.9)	32	13
σ_{trijet}	32.8	3.9	$^{+49}_{-39}$	42 (-0.1)	44	12
$\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	40.8	5.7	$^{+61}_{-48}$	34 (-0.7)	33	15
$\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	32.1	6.8	$^{+47}_{-41}$	32 (-0.9)	32	13
$\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	33.3	4.0	$^{+49}_{-41}$	43 (-0.2)	46	13
$[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	44.2	3.7	$^{+60}_{-41}$	24 (-1.8)	30	11
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	34.0	5.3	$^{+55}_{-39}$	24 (-1.1)	23	9

Table 10.6: Uncertainties on α_s due to missing higher orders in fits to different observables. The values are derived separately for the renormalization $\Delta_{\mu_r}\alpha_s$ and factorization scale $\Delta_{\mu_f}\alpha_s$ and are determined by linear error propagation of the uncertainty on the theory cross section estimated by a variation of the scales. The values of $\Delta_{\mu_r}\alpha_s$ are compared to values obtained with other methods.

a variation of the renormalization scale $\Delta_{\mu_r}\alpha_s$ ranges from 0.0032 to 0.0047, and hence is by far the largest uncertainty on α_s. The uncertainty from a variation of the factorization scale $\Delta_{\mu_f}\alpha_s$ is 0.0004 to 0.0006. The uncertainties are smaller for the α_s-fits to trijet data.

When other methods than the linear error propagation are employed for determining $\Delta_{\mu_r}\alpha_s$, the uncertainty differs significantly between the methods. The values obtained from the offset method are larger than the one from the error propagation, since the uncertainty is treated as fully correlated. We further observe that for each of the fits in the offset method, the χ^2/n_{dof} increases compared to the central values and reaches up to 2.3. Additionally we note that the $\Delta^{\text{offset}}_{\mu_r}\alpha_s$ values are almost always smaller for the 'down' shift than for the 'up' shift. This is consistent with the fits, where $\Delta_{\mu_r}t$ is alternatively included as an additional nuisance parameter in the fit, and the nuisance parameter ε_{μ_r} is determined to be negative with a large value. However, the results for α_s (which are not used) of these fits are definitely unreliable, since the correlation coefficient[28] of α_s is close to one ($\rho_\alpha \approx 0.99$), and therefore all errors are not meaningful because they are so high correlated [218]. The large value is caused by the large correlation with the nuisance parameter ε_{μ_f}, which is for instance -0.85 in the fit to σ_{jet}. This result has two implications:

- Since the correlations between the nuisance parameter of $\Delta_\mu t$ and α_s is large, and the global correlation coefficient of α_s is $\rho_\alpha \approx 0.99$, it is impossible to determine the scale factor c_{μ_r} and α_s simultaneously in a fit.

- As a direct consequence, this implies that the uncertainty from the μ_r-variation is always becoming large for calculations in next-to-leading order precision.

The uncertainties $\Delta_{\mu_r}\alpha_s$ from fits, where $\Delta_{\mu_r}t$ is included in the covariance matrix either as correlated or as uncorrelated uncertainty, illustrate the effect of reduced errors, when these are treated as uncorrelated. Also in these cases the results on $\Delta_\mu\alpha_s$ are not reliable, because the values of α_s determined in these fits are essentially different, and hence the prescription to calculate α_s has changed compared to the central value, and equation 10.21

[28] See equation 9.7 for the definition of the global correlation coefficient of a parameter from a covariance matrix.

cannot be employed for the determination of $\Delta_\mu \alpha_s$.

When the fits are repeated with different functional choices for the scales (for instance $\mu^2 = p_{\mathrm{T}}^2$, $\mu^2 = Q^2$ or $\mu^2 = \sqrt{p_{\mathrm{T}}^2 + Q^2}$ and/or $\mu_f = \mu_r$), all values obtained for α_s are within the uncertainties from the scale variations. For the various combinations of scale choices for μ_r and μ_f no scale was found which yields smaller values for χ^2/n_{dof} for all of the fits to the measurements.

10.4 Uncertainty of α_s due to the uncertainty of the PDFs

The employed PDFs for the calculation of the NLO cross section are known only with limited precision. We have to discuss three distinct effects from the PDFs on α_s.
Firstly, the knowledge of the PDFs is only of limited precision, since these have been determined from experimental data. This implies some uncertainty on the theoretical predictions and translates into an uncertainty on α_s. We will therefore perform an error propagation from the uncertainties on the PDFs $\Delta f_{q,\bar{q},g}$ on α_s.
Secondly, the strong coupling and the PDFs f_i are always coupled. For all QCD induced processes (or processes which include higher-order QCD corrections), the cross sections are proportional to terms of the form $\propto \alpha_s \cdot f_i$, where especially the gluon density g always has to couple through a QCD vertex to the hard matrix element $\propto \alpha_s f_g$. This implies that in the determination of the PDFs the theoretical predictions which enter the PDF-fit have some dependence on α_s, and in turn also the PDFs determined have an α_s-dependence. The PDF fitting groups provide different PDFs, for different values of α_s.
Thirdly, the PDFs are determined by several groups and partly from different data. As we have seen in this work, a fit to one or more data sets can be a difficult task and already small differences yield different results. The PDF groups feature different philosophies on almost all parameters of a PDF fit. These are, for instance, the choice of data sets and theoretical calculations and parameters, such as the scale choices, the parametrization of the functional form of the PDFs at the starting scale, the χ^2 definition, the error treatment, additional constraints on the fit and on theoretical properties, etc. Therefore, the PDFs determined from different groups yield different results on the cross section and hence on α_s.

10.4.1 The PDF dependence of the cross sections

The dependence on the PDF of the NLO calculations for inclusive jet, dijet and trijet cross sections, as well as for LO and gluon induced contributions, are shown in figure 10.3 for the MSTW2008 PDF. In this illustration, small intervals in $\log(x)$ are integrated over, and relative contributions to the total jet cross section in that bin are displayed. Although, the observable 'p_{T}' for inclusive jets and dijet differs, the x-dependence of the cross section in comparable bins is obvious, due to the leading-order correspondence of $p_{\mathrm{T}}^{\mathrm{jet}}$ and $\langle p_{\mathrm{T}} \rangle_2$ in the Breit frame. The dependence of the cross sections on the PDFs is therefore expected to be similar for the inclusive jet and dijet observables. The cross sections have larger sensitivity to the low-x region for smaller scales (Q^2,p_{T}) and are more sensitive to the high-x region for larger scales. The measurements at small p_{T} and lower Q^2 are dominated by gluon induced processes, since at small x-values the gluon density is already significantly larger than the

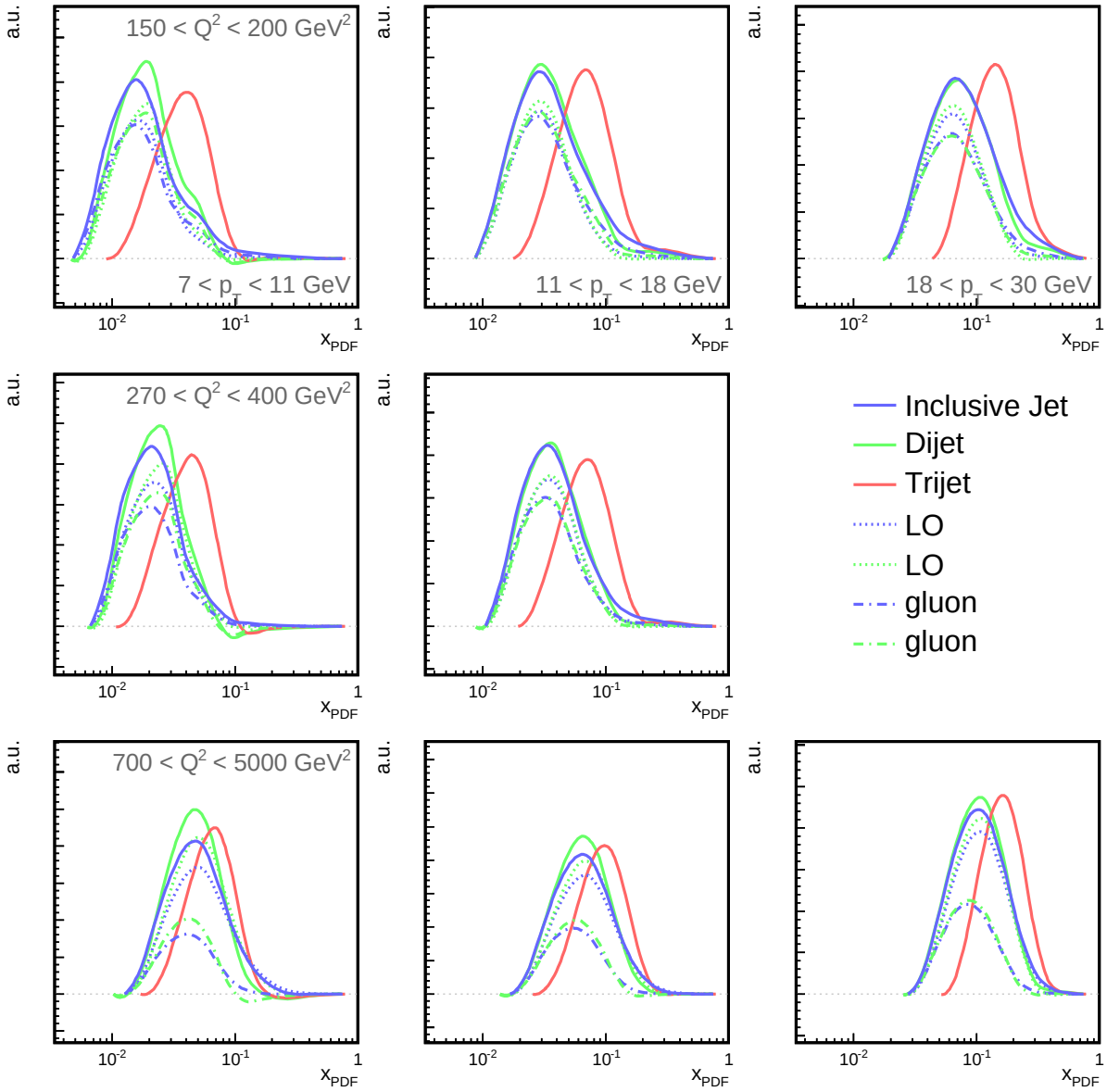


Figure 10.3: Jet cross sections and contributions to them as function of the incoming parton momentum for selected bins of the three jet measurements. Displayed are the relative contributions from small x-intervals to the total cross section (of that measurement bin), where x is the momentum fraction of the incident parton w.r.t. incoming proton momentum. The values from the set of the discrete x-values are smoothed for better readability. The gluon induced fraction is displayed separately for the inclusive jet and dijet cross sections, where also the leading-order contribution is shown. (Please note: the observable labeled 'p_{T}' is different for the inclusive jet, dijet and trijet measurement.)

quark densities. At large values of p_{T} and Q^2, where the incident partons require large momentum fractions, the gluon content in the proton is small, and the contribution from valence quarks dominates. The measurement of $\langle p_{\mathrm{T}} \rangle_3$ is only sensitive to larger x-values because of the requirement for three hard jets and does generally not reach values below $x < 10^{-2}$, and hence it is more sensitive to the valence quarks.
In every Q^2-bin, the sensitivity on the high x-region increases with increasing p_{T}. On the other hand, the inclusive NC DIS cross section integrates this p_{T}-dependence and furthermore has as the dominant leading-order contribution the QPM matrix element. Therefore, in each Q^2-bin, the x-dependence of σ_{NC} is significantly different from that than in every single p_{T}-bin of the jet measurements, and hence the PDF-dependence does not cancel for normalized jet cross sections. The dependence on the PDF can even be increased in the calculation of normalized jet cross sections compared to absolute jet cross sections[29].

10.4.2 Error propagation for the uncertainty on α_s due to the PDF

The PDFs are determined in a fit to data, where multiple parameters of the functional forms of the parton density functions at the starting scale are determined, and a covariance matrix for these parameters is obtained from the fit. This covariance matrix can be diagonalized, and the resulting eigenvectors are orthogonal and represent the uncertainty on the PDFs [30].
The uncertainty on the NLO predictions from the PDFs $\Delta^{\pm}_{\mathrm{PDF}}t$ is determined as asymmetric uncertainty from all N eigenvector pairs i following the 'Master equation' [225]

$$\Delta^{\pm}_{\mathrm{PDF}}t = \sqrt{\sum_{i=1}^{N} \left[\max\left(\pm\Delta^{+}_{i}t, \pm\Delta^{-}_{i}t, 0\right)\right]^2}, \tag{10.25}$$

where i runs over all pairs of eigenvectors and $\Delta^{\pm}_{i}t$ denote the difference from the prediction $t^{\pm}_{i}$ of the two eigenvectors i to the prediction of the central result t_0 ($\Delta^{\pm}_{i}t = t^{\pm}_{i} - t_0$).
The uncertainty $\delta_{\mathrm{PDF}}t$ with a confidence level of 68 % on the NLO predictions is around 0.5 to 1.5 % for the absolute as well as for the normalized jet cross sections (see figure 7.4, where it is also compared to predictions from other PDFs.). The uncertainty $\Delta_{\mathrm{PDF}}t$ can be directly propagated to α_s by linear error propagation. We denote this uncertainty on α_s as $\Delta_{\Delta_{\mathrm{PDF}}t}\alpha_s$, where we treat all bins as fully correlated (i.e. $\mu^{\mathrm{C}}_{\Delta_{\mathrm{PDF}}t} = 1$).
However, the knowledge of the individual orthogonal eigenvectors, and hence the correlations between the bins, is lost in this expression, and therefore it is preferred to propagate the uncertainties from the PDF eigenvectors directly to obtain the uncertainty on α_s. The asymmetric uncertainties from the PDFs are defined by

$$\Delta^{\pm}_{\mathrm{PDF}}\alpha_s = \sqrt{\sum_{i=1}^{N} \left[\max\left(\pm\Delta^{+}_{i}\alpha_s, \pm\Delta^{-}_{i}\alpha_s, 0\right)\right]^2}, \tag{10.26}$$

where the uncertainty on α_s from one single eigenvector $\Delta^{\pm}_{i}\alpha_s$ is determined from linear

[29] For instance, consider a PDF-eigenvector with increased gluon density, but where the quark densities are decreased due to the sum-rules: as a consequence such an eigenvector yields larger jet-cross sections but smaller σ_{NC} and in this case the PDF uncertainties even increase for normalized jet cross sections.

[30] The HERAPDF group further estimates an uncertainty due to the parameterization. The NNPDF group estimates the uncertainty from an average of all PDFs.

Uncertainty on α_s from the PDF $\times 10^4$

Measurement	$\langle \Delta_{\text{PDF}} \alpha_s \rangle$	$\Delta^{\pm}_{\text{PDF}} \alpha_s$	$\Delta_{\Delta_{\text{PDF}} t} \alpha_s$	$\Delta_{\text{PDFset}} \alpha_s$
σ_{jet}	6.7	$^{+8.5}_{-6.1}$	$^{+7.4}_{-5.4}$	7.4
σ_{dijet}	6.2	$^{+7.6}_{-5.2}$	$^{+8.0}_{-6.0}$	9.3
σ_{trijet}	3.0	$^{+3.1}_{-2.8}$	$^{+4.1}_{-3.6}$	7.2
$\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	5.2	$^{+6.2}_{-4.2}$	$^{+7.1}_{-5.6}$	10.9
$\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	5.0	$^{+6.0}_{-4.0}$	$^{+7.6}_{-6.1}$	12.8
$\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	3.0	$^{+3.1}_{-2.9}$	$^{+4.1}_{-3.6}$	8.8
$[\sigma_{\text{jet}}, \sigma_{\text{dijet}}, \sigma_{\text{trijet}}]$	3.6	$^{+4.2}_{-2.9}$	$^{+4.6}_{-3.6}$	6.3
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	4.0	$^{+4.7}_{-2.9}$	$^{+6.4}_{-5.2}$	10.6

Table 10.7: The symmetrized $\langle \Delta_{\text{PDF}} \alpha_s \rangle$ and asymmetric uncertainty $\Delta^{\pm}_{\text{PDF}} \alpha_s$ from the PDF eigenvectors (at 68 % CL.) on α_s compared to $\Delta_{\Delta_{\text{PDF}} t} \alpha_s$, where the uncertainty on the NLO prediction is propagated to α_s assuming them to be fully correlated. The uncertainty $\Delta_{\text{PDFset}} \alpha_s$ is defined as the half-difference between the α_s-result in fits to calculations employing the NNPDF2.3 and CT10 set.

error propagation of the uncertainty from the two eigenvectors i on t (denoted $\Delta^{\pm}_i t$)[31] following equation 10.15. The eigenvectors are orthogonal and therefore they have to be treated as fully correlated (i.e. $\mu^{\text{C}}_i = 1$).
We define the symmetric uncertainty from the uncertainty of the PDFs on α_s by

$$\langle \Delta_{\text{PDF}} \alpha_s \rangle = \sqrt{\sum_{i=1}^{N} \left[\frac{|\Delta^{+}_i \alpha_s| + |\Delta^{-}_i \alpha_s|}{2} \right]^2} , \tag{10.27}$$

where again the single uncertainties $\Delta^{\pm}_i \alpha_s$ are calculated by linear error propagation. The uncertainties on α_s from the PDF eigenvectors are listed in table 10.7. The symmetric uncertainty from the linear error propagation gives a reasonable average of the asymmetric uncertainty from a single eigenvector, and we will quote that value as the final value. The uncertainty from $\Delta_{\text{PDF}} t$ is larger than the one from the direct error propagation, since the bin-to-bin correlations from different eigenvectors was lost and we always assumed correlated uncertainties[32].
The uncertainties on α_s from $\Delta^{\text{offset}}_{\text{PDF}} t$ (these values are not listed), when the offset method is employed, are comparable to $\Delta_{\Delta_{\text{PDF}} t} \alpha_s$, since the size of the offset shift is rather small and the dependence is still linear. The uncertainties from methods, where $\Delta^{\pm}_i t$ is included as additional systematic uncertainty in the fit are mostly comparable to $\Delta^{\pm}_{\text{PDF}} \alpha_s$, but in fits which include σ_{trijet} the uncertainties are not determinable, since equation 10.21 is not calculable, since the expression under the square-root becomes negative due to a smaller overall uncertainty $\Delta_{\text{exp}} \alpha_s$ compared to the uncertainty on the central value $\Delta_{\text{exp}} \alpha_0$. If the nuisance parameter method is employed, no eigenvector with a large nuisance parameter

[31] For a consistent comparison of various PDF sets in this work, we only use PDF sets, which have been determined at the world average value of $\alpha_s = 0.118$. Since for the employed PDF set MSTW2008 no eigenvector set is available, we use the eigenvector set from the nominal MSTW2008 set, which was determined at $\alpha_s = 0.12018$ and assume that the uncertainties $\Delta^{\pm}_i t$ (which are determined w.r.t. the nominal theory prediction t_0 from MSTW2008 at $\alpha_s = 0.12018$) provide an adequate estimate.

[32] This observation in turn confirms the necessity of estimating reasonably the correlated and uncorrelated fractions of an uncertainty as we have done for the systematic experimental uncertainties and the uncertainties from scale variations.

was identified. This can be interpreted such that this measurement does not favor any definite direction in the eigenvector space of the MSTW2008 fit.

10.4.3 α_s-fits to different PDF sets

All α_s-fits are repeated for theory calculations using PDFs from different PDF groups. In order to be insensitive to the value of α_s in the PDF fit, all PDF sets considered have a set determined at the world average value of $\alpha_s = 0.118$ in NLO precision. We compare PDFs from the MSTW group [61] (which is used for the central α_s value) to predictions using the most recent PDFs from the CT group [59] (former CTEQ collaboration), the NNPDF group [65], the HERAFitter group [58] and from ABM [62][33]. The results for α_s obtained in the fits are visualized in figure 10.4. From the different values obtained the following conclusions can be drawn:

- The uncertainty on α_s derived from the MSTW2008 eigenvalues does not cover the results from different PDF sets. This uncertainty accounts only for the experimental precision of the data [61] in the MSTW PDF determination, but does not account for differences to other groups in the theoretical assumption which enter the PDF fit. These could be, for instance, differences in the scale choices or in the hadronization corrections, or also in the PDF fitting method, i.e. the error treatment, which becomes relevant in complicated PDF fits.

- The most comprehensively studied PDF sets are the ones from the MSTW, CT and NNPDF groups. These sets have also been employed in many LHC analyses so far. We observe that the α_s-value from predictions using CT10 is always larger than the one using NNPDF, whereas 'our' central value from using MSTW is always in between. The α_s-value from predictions with MSTW is almost centered between the NNPDF and CT10 results for absolute jet cross sections, whereas it is closer to the NNPDF predictions for normalized jet cross sections.
  Since these three PDFs have been determined from the largest number of measurements and do not show any differences besides these constant disagreements, we believe that the predictions from these three groups are most reliable, and we assume that the 'true' PDF is somewhere in between. In order to account for the differences on α_s from the various PDF sets, we add the half-difference between the NNPDF and CT10 results as an additional uncertainty on the α_s-value.

- For none of the observables the value of α_s from MSTW is the most extreme value. For instance, the α_s-value from CT10 is always the largest value from absolute jet cross sections, whereas the α_s-value from NNPDF is always the smallest value from normalized jet cross sections. We therefore believe that our 'central value', which was determined with the MSTW PDF set, represents a trustworthy result.

- The α_s-results obtained from the inclusive jet or dijet measurement show always the same tension for all PDF sets. This suggests that this tension does not result from differences in the PDF.

- The HERAPDF 1.5 is determined from NC and CC DIS data taken at HERA only. For the set HERAPDF 1.5 no measurements of jet-cross sections were used. This

[33] A discussion of the differences between these PDFs is a manifold field and would go beyond the scope of this thesis, and therefore we refer the reader to the individual publications and the references therein.

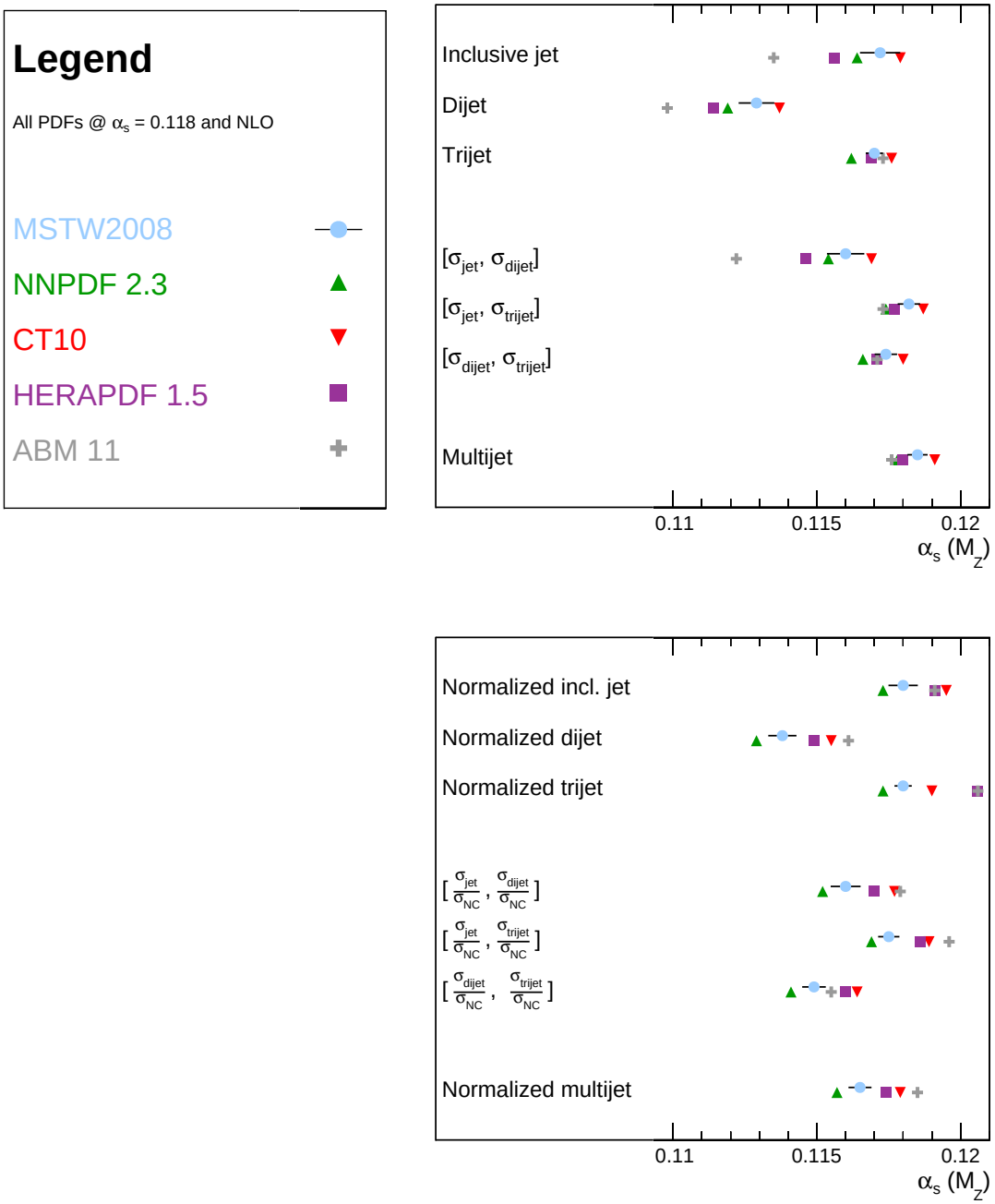


Figure 10.4: α_s from fits to NLO theory predictions obtained with PDFs from different PDF groups. All PDFs have been determined at $\alpha_s = 0.118$ and in NLO precision. The uncertainty bar shown together with the MSTW2008 value shows the uncertainty obtained from the MSTW2008 eigenvectors. The half difference between NNPDF2.3 and CT10W is introduced as additional uncertainty on α_s due to the PDF sets. The numerical values are given in table B.2.

implies that the gluon density cannot be constrained directly, but only indirectly via sum-rules and scaling violations as observed in the measurements of $F_2(x_{\mathrm{Bj}}, Q^2)$, whereas the quark densities are better constrained by the inclusive DIS data. As a consequence, the gluon and quark densities are slightly different than in PDFs which included jet data in their determination, like MSTW, CT10 and NNPDF. Therefore, the result on α_s from the HERAPDF 1.5 set behaves differently compared to the other PDF sets, when α_s-values from absolute jet cross sections are compared to values from normalized jet cross section.

- Similarly as HERAPDF 1.5, the ABM11 PDF set does not include data sets from jet-production cross sections in their determination. The differences of α_s-values between absolute jet and normalized jet cross sections observed compared to MSTW2008 changes are even larger than for the HERAPDF 1.5. The ABM11 PDF sets yield the smallest value of α_s from σ_{dijet}, whereas it yields the largest value from $\frac{\sigma_{\mathrm{dijet}}}{\sigma_{\mathrm{NC}}}$.

No PDF set was identified which yields smaller values of χ^2/n_{dof} for all of the individual α_s-fits .

In order to account for the differing results from various PDF sets, we add an additional uncertainty on the value of α_s, which is given by the half difference between the α_s-values from NNPDF 2.3 and CT10. We refer to it as *PDF set uncertainty* denoted $\Delta_{\mathrm{PDFset}}\alpha_s$. The values for $\Delta_{\mathrm{PDFset}}\alpha_s$ are listed in table 10.7 and are in the range of 0.0007 to 0.0012.

10.4.4 The α_s-dependence of the PDFs

In the determination of the PDFs the value for $\alpha_s(M_Z)$ has to be chosen or determined simultaneously with the PDF. The strong coupling constant enters a PDF-fit in two places. On the one hand in the QCD splitting functions $\mathcal{P}$ which are employed for the evolution equations, and on the other hand in the calculation of the cross sections which are considered in the PDF-fit. The two dimensional parton density functions $f_{q,\bar{q},g}(x, \mu_f)$ determined are therefore different for different assumptions of $\alpha_s(M_Z)$ in the PDF-fit. In figure 10.5 we show the results on α_s from our α_s-fit using different MSTW2008 PDFs, which have been determined at values from $\alpha_s(M_Z) = 0.110$ to 0.130. Due to the correlation of α_s and $f_{q,\bar{q},g}$ in the cross section calculation via the hard coefficients c (i.e. $\sigma \propto c(\alpha_s) \cdot f$), the PDFs decrease with increasing fixed α_s in the PDF-fit[34]. Therefore, the value of α_s determined in the fit to the jet cross sections increases with increasing input α_s to the PDF-fit. The trijet cross sections are proportional to $\alpha_s^2 \cdot f$ and hence have a larger sensitivity on α_s than on a change in f. The α_s-value obtained in the fit to trijet cross sections shows a minimum for PDFs determined at values $\alpha_s \approx 0.119$.
The PDFs determined at $\alpha_s \approx 0.115$ yield consistently the smallest value of χ^2/n_{dof} for all observables. The resulting values for α_s at the minimum χ^2/n_{dof} are consistent with the central values within the uncertainty from the factorization scale variation $\Delta_{\mu_f}\alpha_s$, which accounts for the reshuffling between the hard and the soft part in the perturbative calculation. The PDF sets from different α_s-values do not shed any light on the differences between the inclusive jet and dijet results.

[34]This seems to be in contradiction to the effect of increasing splitting functions $\mathcal{P}(\alpha_s)$. However, at all scales μ_f, where the PDFs are determined from data, this correlation is expected.

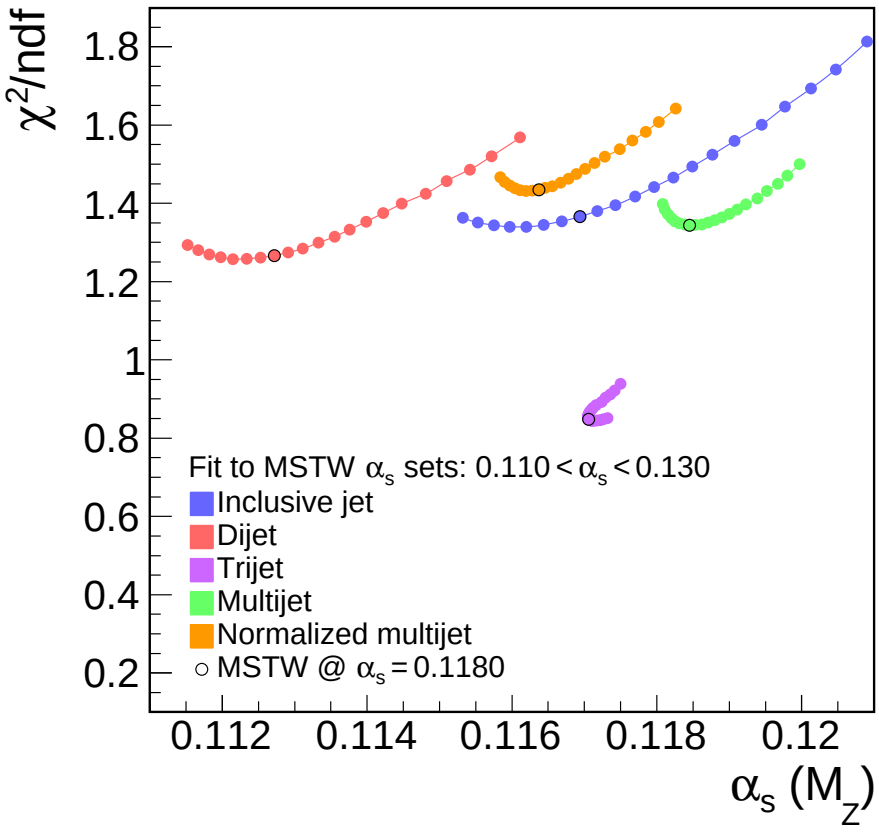


Figure 10.5: α_s-fits to PDFs which have been determined at different values of α_s. In the α_s-fits with PDFs determined at $\alpha_s = 0.110$ small values of α_s are obtained, whereas with PDFs which have been determined at large values of α_s also large values of α_s are extracted. This correlation could be due to the inclusion of inclusive jet measurements from HERA-I [103] in the PDF-fit. The α_s-fit to the trijet measurement does not show this correlation. The χ^2/n_{dof} shows a minimum, consistently for all observables for the PDF set determined at $\alpha_s \approx 0.115$.

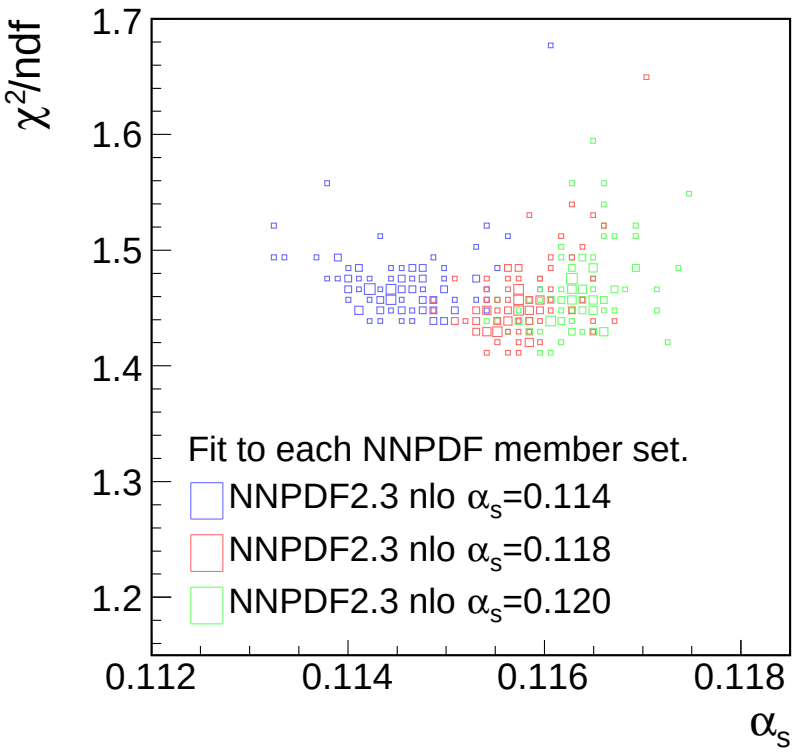


Figure 10.6: Values of χ^2/n_{dof} versus the α_s obtained from fits of normalized multijet cross sections to 300 PDF sets of different parametrizations and determined at three values of α_s.

10.4.5 The dependence of the value of α_s on the PDF parametrization

The PDFs have to be parametrized with some reasonable function at the starting scale and are then evolved with the evolution equations to higher scales in order to determine $f_i(x, \mu_f)$. This parametrization, however, is arbitrary and could bias the PDF. The NNPDF group provides unbiased PDFs which are (almost) independent on the input parametrization.

We repeat the α_s-fit to absolute multijet cross sections using 100 PDFs from the NNPDF2.3 PDF set (and for three different input values of α_s to the PDF fit). Each PDF features a different parametrization at the starting scale, but yields an optimal fit result for all considered data sets in the PDF determination. The values of χ^2/n_{dof} versus the fitted α_s-values are displayed in figure 10.6. The values for χ^2/n_{dof} obtained from the 300 fits show comparable values of $\chi^2/n_{\text{dof}} \approx 1.45$, with only a few PDF sets yielding large values. No PDF set with considerably smaller values of χ^2/n_{dof} was found. We therefore conclude that the tension between α_s-values obtained from inclusive jet and from dijets is not primarily originating from the PDF parametrization.

The dependence of the fitted value of α_s on the input α_s for the PDF determination is quite different than for the MSTW PDF sets. The median α_s-values $\tilde{\alpha}_s$ obtained from the mulijet measurement in 100 fits to PDFs determined at $\alpha_s = 0.114$, 0.118, and 0.120 are found to be $\tilde{\alpha}_s = 0.11454$, 0.11573 and 0.11631, respectively. The median $\tilde{\chi}^2/n_{\text{dof}}$ yields the smallest value for the PDF sets determined at $\alpha_s = 0.118$[35].

[35]The median values of $\tilde{\chi}^2/n_{\text{dof}}$ are 1.467, 1.447 and 1.457, for the PDF sets determined at $\alpha_s = 0.114$, 0.118 and 0.120, respectively.

10.4.6 α_s-fits to predictions including α_s-dependencies of the PDFs

We repeated the α_s-fits, where the α_s-dependence of the PDF is considered in the fit. The theory calculation $t(\alpha_s, f_i)$ in equation 10.1 is then replaced by a parametrized function $t(\alpha_s) = t(\alpha_s, f_i(\alpha_s))$, where $f_i(\alpha_s)$ are the PDFs determined at the values α_s.[36] These fits, however, yield misleading results, since on the one hand the experimental uncertainties are no longer reliable due to the inclusion of an additional theoretical dependence[37]. On the other hand, the α_s-fit does not include the χ^2-constraints from other data sets, which have been used in the PDF determination and could therefore lead to inconsistent values with other measurements. The only consistent method is a combined PDF fit, which, however, was not possible due to technical limitations in the treatment of correlated uncertainties in available fitting tools up to now.

10.5 Study of the dependence of α_s on the scale variation

The uncertainty on α_s due to missing higher orders has been determined in section 10.3.3 by a variation of the renormalization and factorization scales, whereas the cross section dependence on a scale factor has been studied in section 2.1. In this section, the direct dependence of α_s on a scale factor applied to the renormalization scale is studied, where the nominal renormalization scale is chosen to be $\mu_{r,0}^2 = \frac{Q^2+p_{\mathrm{T}}^2}{2}$.
The dependence of α_s on the renormalization scale factor c_{μ_r} is displayed in figure 10.7, where the fits are repeated for scale factors between $-2 < c_{\mu_r} < 2$. This range corresponds to a variation of the renormalization scale between $0.25\mu_{r,0}$ to $4\mu_{r,0}$. The renormalization scale enters the cross section calculation directly through the running of the coupling with terms of the form $\ln \frac{\mu_r}{\Lambda_{\mathrm{QCD}}}$ and consequently also the cross section in the perturbative expansion (c.f. equations 2.9 and 2.5).
With an increasing scale factor the value of $\alpha_s(\mu_r)$ decreases and hence the theoretical cross sections decrease. Therefore, the best value of $\alpha_s(M_Z)$ from the α_s-fit increases in order to compensate this effect. A combined fit of α_s and c_{μ_r}, where c_{μ_r} is an additional free parameter in the fit without any constraint, yields a correlation between these parameters of 0.8 (0.9), 0.6 (0.6) and 0.7 (0.2) for the fit to (normalized) inclusive jet, dijet and trijet cross sections, respectively.

[36] The α_s-dependence of the PDF is ambiguous, since it arises on the one hand from the included data sets and the α_s-dependence of the perturbative expansion of the corresponding processes as well as the scales of these processes. On the other hand, this dependence is in contradiction to the α_s-dependence of the splitting functions. As a consequence, also the α_s-dependence of σ_{NC} from α_s-dependent PDFs is obscure. The global PDF groups predict an increasing σ_{NC} cross section with increasing α_s. This is caused by the inclusion of jet-data in their PDF determination which has larger sensitivity to α_s. If trijet cross sections of $\mathcal{O}(\alpha_s^2)$ would be considered in the PDF-fit, this dependence would be even increased. The HERAPDF on the other hand, does not show any α_s-dependence of σ_{NC}, since only inclusive processes constrain the PDF at the scale of our measurement and therefore the input α_s-dependence is always absorbed in the PDF. The sensitivity of the normalized jet cross sections on α_s is therefore ambiguous and yields inconsistent results and even uncertainties for different PDF sets (since the α_s-dependence of normalized cross sections is unclear).

[37] The factorization theorem states no α_s-dependence of the soft factorized expressions $f_i(x, \mu_f)$.

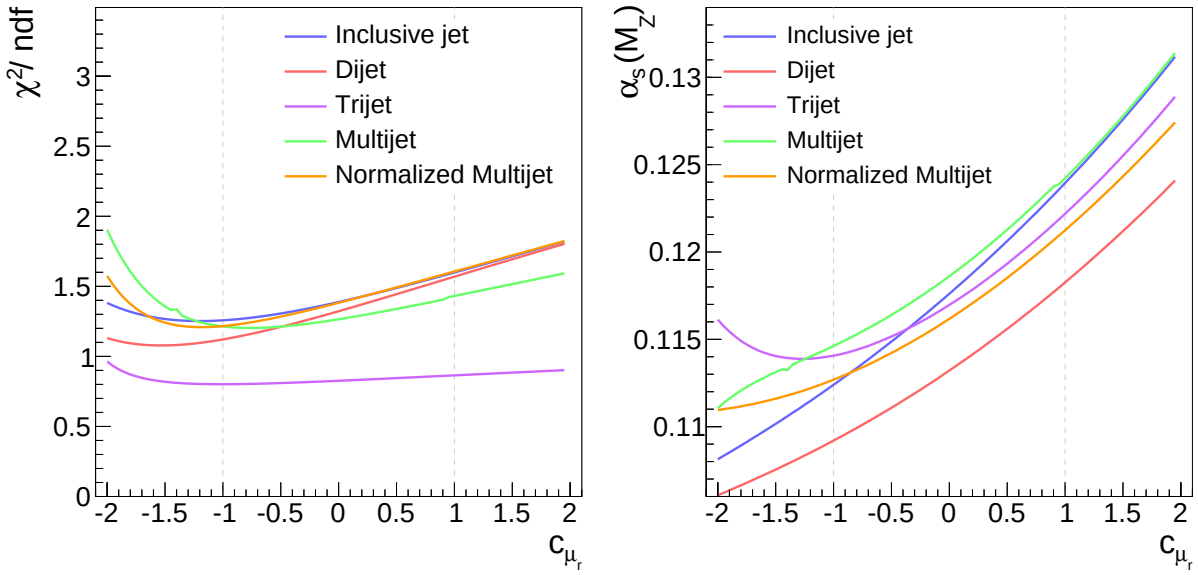


Figure 10.7: Results from α_s-fits as a function of a renormalization scale factor c_{μ_r}. The χ^2/n_{dof} dependence is non-trivial, and the minima found at negative scale factors cannot be regarded as physical. The fits show a strong correlation between the scale factor and the value of $\alpha_s(M_Z)$ found.

Compared to the scale dependence of α_s from dijets, the α_s from the inclusive jet data is more sensitive to the scale factor. This is probably due to the larger k-factors observed in predictions of inclusive jet cross sections than dijet cross sections (see figures 10.9 and 10.10).

The values of χ^2/n_{dof} show unincisive minima at small scale factors. An explanation for this effect is non-trivial. Following equations 2.12 and 2.13, the scale dependence enters the calculation in leading-order only through $\alpha_s^n(\mu_r)$, whereas in NLO it enters through $\alpha_s^{n+1}(\mu_r)$, but also through a scale dependent compensating contribution of the form $\beta_0 k_n \ln \mu_r$, where k_n denotes the (scale-independent) contribution from the leading-order calculation. In an α_s-fit any change of the scale factor can be compensated exactly by changing $\alpha_s(M_Z)$ via terms of the form $\ln \frac{\mu_r}{\Lambda_{\text{QCD}}}$ (see equation 2.9) which is demonstrated in figure 10.8. Hence, the only non-compensating contribution in the fit is the scale-dependence of order NLO which is proportional to the leading-order matrix elements. This contribution is therefore responsible for any change of χ^2/n_{dof} and hence also for the weak minima. However, this contribution neither represents any reasonable quantifier for the missing higher orders, nor is it meaningfully interpretable in the perturbative calculation and hence does not have any physical meaning. Furthermore, it cannot not be linked to specific phase space regions, since its absolute value is dependent on the reference scale of the calculation. Therefore, these χ^2-minima are without any meaning and as a consequence, it is not reasonable to determine simultaneously the renormalization scale and $\alpha_s(M_Z)$ in a fit to data. It also becomes obvious that any choice of the renormalization scale factor c_{μ_r} cannot give exact NNLO (or higher orders) predictions for the full phase space[38].

[38]Therefore, in the determination of $\Delta_{\mu_r}\alpha_s$, which should represent an estimate of the uncertainty due to missing higher orders, the uncertainties on $\Delta_{\mu_r} t_i$ have been treated as half correlated and half uncorrelated.

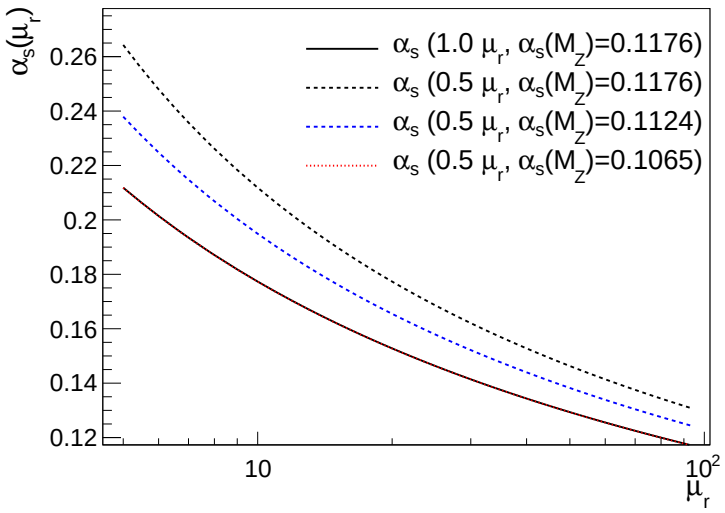


Figure 10.8: The running of the strong coupling constant $\alpha_s(\mu_r)$, for three different values of $\alpha_s(M_Z)$ and two scale factors $c_{\mu_r} = -1$ and $c_{\mu_r} = 0$. The values for $\alpha_s(M_Z)$ have been determined from the inclusive jet data. Any change from the scale factor c_{μ_r} on $\alpha_s(\mu_r)$ can exactly be compensated by Λ_{QCD}, or alternatively by $\alpha_s(M_Z)$.

10.6 Study on k-factor cut

The differences of the results on α_s in the fits to various observables, and the resulting tension in the α_s-fit to multijets between the single measurements is most likely caused by shortcomings of the theoretical predictions. We have searched for but have not found any experimental reasons for them. The calculations miss higher order contributions beyond NLO, which may be relevant to make these differences disappear.
A valuable parameter to estimate the size of contributions of higher orders is the *k-factor*, defined in each bin i as the ratio of the next-to-leading order to the leading order prediction

$$k_i = \frac{\sigma_i^{\mathrm{NLO}}}{\sigma_i^{\mathrm{LO}}} \ , \tag{10.28}$$

representing the size of the first order correction to the LO calculation. The assumption is made now that the k-factor can also be interpreted as an estimate for the unknown higher-order corrections beyond NLO. A large k-factor is thus interpreted such that missing higher orders give relevant contributions to the cross section in that bin. The k-factors are displayed for the absolute jet cross sections in figure 10.9. The k-factors of the inclusive jet calculations are between 1.09 and 1.49[39], for the dijet calculations between 1.04 and 1.42 and for the trijet calculations between 1.16 and 1.40.
The α_s-fits are repeated with a constraint on the k-factor. Bins with relevant contributions from missing higher orders are neglected in the fit and only bins where the perturbative series is close to convergence are considered. The dependence of the resulting degrees of freedom $n_{\mathrm{dof}} = n_{\mathrm{bins}} - 1$ on the allowed maximum k-factor k_{max}, as well as the values of α_s from the α_s-fit are displayed in figure 10.10.
Fits where at least half of the bins are considered are close to the nominal result with all bins included. The fit quality is almost constant for the dijet, trijet and multijet fits, but

[39]Following the perurbative expansion of the cross section from equation 2.5, the k-factor has a proportionality to α_s by $k \propto 1 + \alpha_s(\mu_r)\frac{c^{\mathrm{NLO}}}{c^{\mathrm{LO}}}$ and a small dependence on the employed PDFs. Here, the k-factors are calculated with the MSTW2008 PDF set and $\alpha_s(M_Z) = 0.1184$.

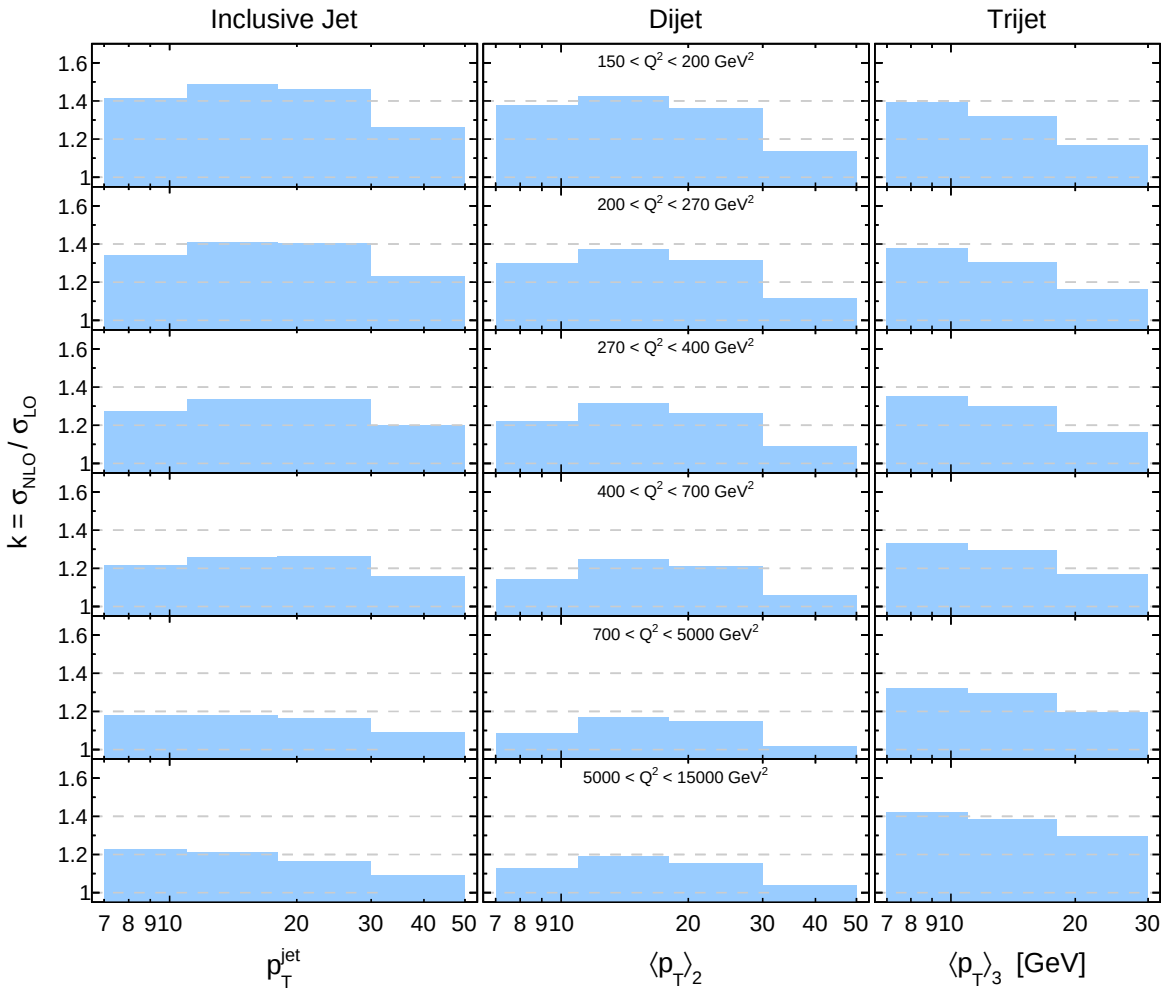


Figure 10.9: The k-factors of the inclusive jet, dijet and trijet NLO calculations using MSTW2008.

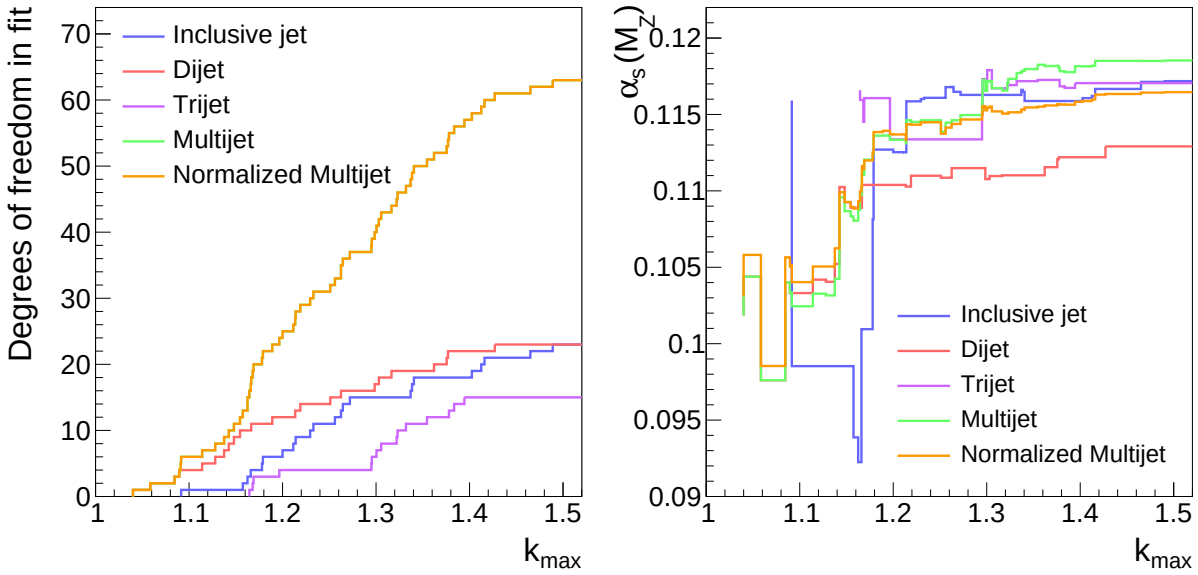


Figure 10.10: Relevant parameters of α_s-fits to datasets with a constraint on the k-factor. The α_s-fit is repeated for all values of k_{max} where one additional bin is discarded due to $k_i > k_{\text{max}}$. The left pad shows the resulting degrees of freedom of the fit to inclusive jet, dijet, trijet, multijet and normalized multijet cross sections. The right plot shows the resulting values of α_s as a function of k_{max}. Whenever only very few bins are left, the experimental uncertainty on the fit result is very large. The values of α_s show a small tendency to increase with increasing k_{max}, which points towards missing contributions beyond NLO.

the fit to inclusive jets has a large value of χ^2/n_{dof} if only bins with $k < 1.3$ are considered. Fits, where the k-factor is demanded to be small and only few bins contribute have large statistical uncertainties, since these bins are mostly related to high scale processes where the statistics becomes small.
The fit to normalized multijets[40] results in almost identical values of α_s, if k_{max} is below 1.3. At this value, bins from the trijet measurement with relevant statistical precision contribute, and the tension between the dijet and trijet measurement becomes present and is reduced in the fit by increased nuisance parameters.

With decreasing k-factors, the value of α_s is expected to be reduced, since the contribution from missing higher orders becomes small and are not necessarily compensated by an increased value of α_s. On the other side, we also have observed (see section 10.2.4 and figure 10.1) that the bins remaining at small k-factors, i.e. bins in the high Q^2 and high p_{T} region, tend to be below the theory prediction and hence result in small values of α_s. It cannot be figured out if this due to an accidental statistical fluctuation[41].

A cut on the k-factor has the advantage of a reduced theory uncertainty from scale variations, since bins with small values of k are located at large scales. At large values of the renormalization scale, the running coupling is less sensitive to a variation of the renormalization scale and due to the proportionality of $\sigma \propto \alpha_s(\mu_r)$, the theory uncertainty on α_s is reduced, since bins with large $\Delta_{\mu_r} t_i$ are removed from the fit.

[40]The k-factor for normalized jets is defined identical to the absolute jet cross sections, i.e. without considering the inclusive NC DIS calculation in the denominator.

[41]The previous analysis of HERA-II data [102] also observed the tendency for similar fluctuations. The H1 analysis of HERA-I data only does not show this behavior [103].

Since this analysis aims to determine α_s from the full phase space of the multijet measurement, no constraint on the k-factor is employed, especially, since bins within the phase space would be discarded instead of bins at the phase space boundary. Furthermore, the k-factors of different bins are often quite close, especially for bins from different measurements. Since the k-factor is dependent on $\alpha_s(M_Z)$, the scale choice as well as the PDF, it is not uniquely defined and therefore can lead to misleading results. Moreover, a large emphasis was put on a consistent analysis of inclusive jet, dijet and trijet cross sections, whereas a cut on the k-factor would impair this experimental approach with a theoretically motivated parameter.

10.7 Summary and conclusion on the α_s-fit

Fits of the strong coupling constant $\alpha_s(M_Z)$ under the assumption of the renormalization group equation of QCD have been performed to the unfolded double-differential inclusive jet σ_{jet}, dijet σ_{dijet} and trijet cross sections σ_{trijet} as functions of Q^2 and $p_{\text{T}}^{\text{jet}}$, $\langle p_{\text{T}} \rangle_2$ and $\langle p_{\text{T}} \rangle_3$, respectively. The coupling strength has further been determined from jet cross sections, which are normalized to the inclusive NC DIS cross sections.
The consistency of the individual data points from the different observables have been confirmed for α_s-determinations from individual data points and fits to sub-sets of data points. These are mostly consistent within the experimental uncertainty, and only some α_s-values at large scales deviate towards lower values. These measurements refer to a renormalization scale range of $10 < \mu_r < 94\,\text{GeV}$, and hence these α_s-values also confirm the assumption of the RGE of QCD within this range.
The values for $\alpha_s(M_Z)$ obtained from fits to absolute jet cross sections are summarized in figure 10.11 and are listed here in full detail:

$$\begin{aligned}
\alpha_s \text{ from } \sigma_{\text{jet}}: &\quad 0.1172\,(22)_{\text{exp}} \pm (7)_{\text{PDF}} \pm (7)_{\text{PDFset}} \pm (10)_{\text{had}} \pm (47)_{\mu_r} \pm (6)_{\mu_f} \\
\alpha_s \text{ from } \sigma_{\text{dijet}}: &\quad 0.1129\,(24)_{\text{exp}} \pm (6)_{\text{PDF}} \pm (9)_{\text{PDFset}} \pm\ (7)_{\text{had}}\ \pm (37)_{\mu_r} \pm (7)_{\mu_f} \\
\alpha_s \text{ from } \sigma_{\text{trijet}}: &\quad 0.1170\,(18)_{\text{exp}} \pm (3)_{\text{PDF}} \pm (7)_{\text{PDFset}} \pm (11)_{\text{had}} \pm (33)_{\mu_r} \pm (4)_{\mu_f}\ .
\end{aligned}$$

The experimental as well as the theoretical uncertainties have been determined consistently by linear error propagation, but only the PDFset uncertainty is estimated from the differences in the results obtained from the CT10 and NNPDF2.3 PDF sets. The dominant uncertainty on $\alpha_s(M_Z)$ arises from knowing the perturbative coefficients up to next-to-leading order only. The experimental uncertainty is dominated by the uncertainty on the luminosity measurement. The α_s-values from the inclusive jet and trijet measurement are almost identical, while the value from the dijet measurement is lower by the size of the total theoretical uncertainty. Since the dijet events are a subsample of the inclusive jet measurement, and both calculations are based on the same matrix elements[42], this points towards differing contributions from the missing higher orders to the respective theoretical predictions. The value of α_s from σ_{trijet} has the smallest experimental uncertainty due to the larger sensitivity to $\alpha_s(M_Z)$ and due to a smaller correlation with the nuisance parameter of the normalization, although, the statistical precision is lower than for the inclusive jet and dijet observables.
The values of $\alpha_s(M_Z)$ obtained from fits to normalized jet cross sections, summarized in figure 10.12, feature a substantially reduced experimental uncertainty compared to the

[42] Due to the requirements $-1.0 < \eta_{\text{lab}}^{\text{jet}} < 2.5$ for the jets, and also $M_{12} > 16\,\text{GeV}$ for dijets, there are configurations present, which contribute only to the inclusive jet, but not to the dijet measurement.

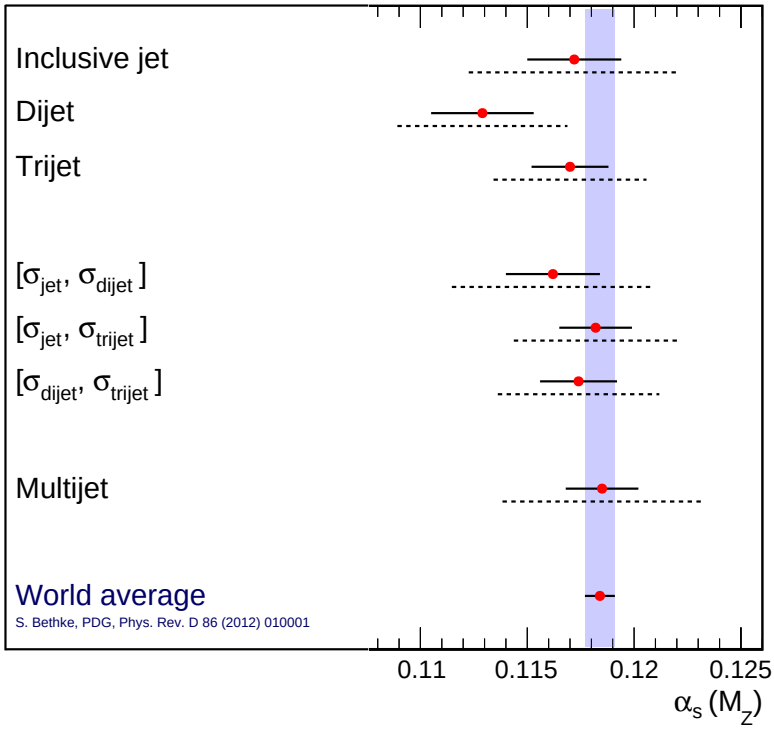


Figure 10.11: Values of $\alpha_s(M_Z)$ obtained from individual and simultaneous fits to absolute jet cross sections. The full error bars indicate the experimental uncertainty, and the dashed error bars illustrate the theoretical uncertainty, where all individual sources of theoretical uncertainties are added in quadrature. The values are compared to the world average value [12,67], which is briefly described in section 2.2.

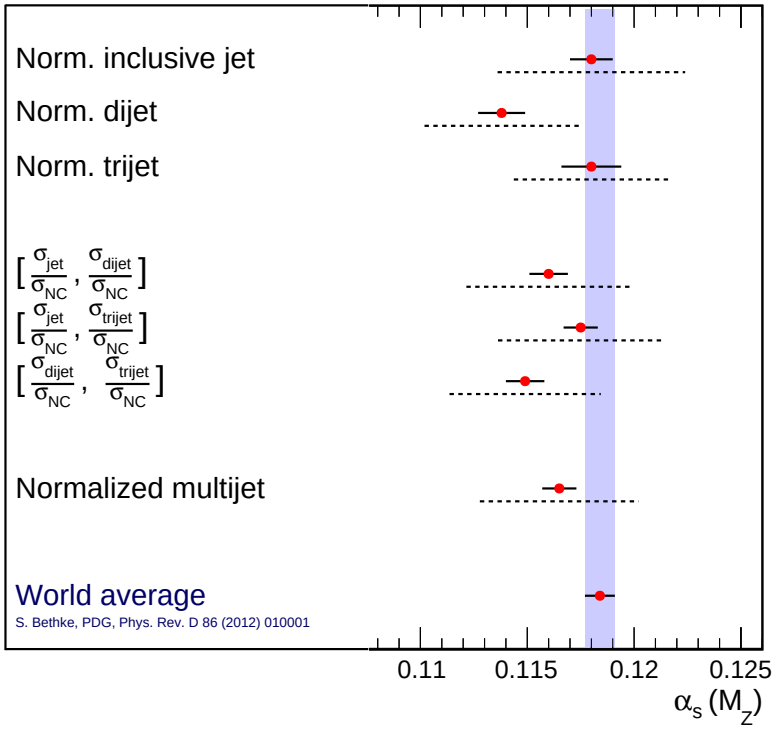


Figure 10.12: Values of $\alpha_s(M_Z)$ obtained from individual and simultaneous fits to normalized jet cross sections. See figure 10.11 and the text for more details.

absolute jet cross sections, due to the cancellation of uncertainties on the normalization. The values are:

$$\begin{aligned}
\alpha_s \text{ from } \tfrac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}: &\quad 0.1180\,(10)_{\text{exp}} \pm (5)_{\text{PDF}} \pm (11)_{\text{PDFset}} \pm\ (8)_{\text{had}}\ \pm (41)_{\mu_r} \pm (6)_{\mu_f} \\
\alpha_s \text{ from } \tfrac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}: &\quad 0.1138\,(11)_{\text{exp}} \pm (5)_{\text{PDF}} \pm (13)_{\text{PDFset}} \pm\ (6)_{\text{had}}\ \pm (32)_{\mu_r} \pm (7)_{\mu_f} \\
\alpha_s \text{ from } \tfrac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}: &\quad 0.1180\,(14)_{\text{exp}} \pm (3)_{\text{PDF}} \pm\ (9)_{\text{PDFset}}\ \pm (11)_{\text{had}} \pm (33)_{\mu_r} \pm (4)_{\mu_f}\ .
\end{aligned}$$

The highest experimental precision is obtained from the normalized inclusive jet measurement due to the highest statistical precision. The same tension between the α_s-values from σ_{jet} and σ_{dijet} is also observed for $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$ and $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$. The α_s-values are increased by around 0.001 compared to those from the absolute jet cross sections. This could be explained by a shift of the estimated recorded luminosity by around 1 to 2 % as suggested from the nuisance parameter of the normalization uncertainty in the fits to absolute jet cross sections.
The α_s-determination from multiple measurements considers more data points, which are obtained by the simultaneous measurement of different observables. Such a fit benefits from obtaining the maximum sensitivity to $\alpha_s(M_Z)$ from all of the considered observables. On the other hand, the simultaneous α_s-fit to multiple measurements gains only little in experimental accuracy due to the statistical correlation of the data. The value of $\alpha_s(M_Z)$ obtained from absolute multijet cross sections ($[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$) is[43]

$$\alpha_s(M_Z) = 0.1185\,(17)_{\text{exp}} \pm (4)_{\text{PDF}} \pm (6)_{\text{PDFset}} \pm (13)_{\text{had}} \pm (44)_{\mu_r} \pm (4)_{\mu_f}\ ,$$

and from normalized multijet cross sections ($[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$) it is

$$\alpha_s(M_Z) = 0.1165\,(8)_{\text{exp}} \pm (4)_{\text{PDF}} \pm (11)_{\text{PDFset}} \pm (8)_{\text{had}} \pm (34)_{\mu_r} \pm (5)_{\mu_f}\ .$$

The α_s-value from absolute multijet cross sections is larger by 0.002 compared to the value from normalized multijet cross sections. This is mainly caused by a large nuisance parameter on the normalization uncertainty in the fit to multijets, which shifts the data effectively to larger values. The fit quality in simultaneous fits is comparably poor. This is due to the deficits in the theoretical predictions for the different observables, which cannot be disentangled in the fit since only experimental uncertainties are considered, and the data is furthermore statistically correlated.
The experimental accuracy on $\alpha_s(M_Z)$ from the absolute jet cross sections is limited by the uncertainty on the normalization. The main components of the experimental uncertainty on $\alpha_s(M_Z)$ from normalized cross sections are the statistical precision, the uncertainty on the measurement of the hadronic final state and the model uncertainty. The latter could not be eliminated by the unfolding procedure.
However, the dominant uncertainties result from limited knowledge of the perturbative series estimated by variation of the scales. Also uncertainties on the hadronization corrections and the experimental precision of the PDF as well as results on α_s using various PDF sets, which differ by more than their respective uncertainties would allow, are non-negligible limitations of the accuracy achieved in the determination of $\alpha_s(M_Z)$ from this data.
Compared to a previous analysis of similar data [102], our result features an enlarged phase space[44], an improved calibration of the HFS [130], an updated luminosity determination [114], and an unbiased regularized unfolding method (see section 6). The values of

[43] The values obtained in fits to two measurements are listed in table B.1.

[44] In this analysis, in contrast, no data from the HERA-I period are considered, which would contribute an additional 9 % of luminosity and would benefit from an independent luminosity measurement.

$\alpha_s(M_Z)$ obtained are consistent within the experimental uncertainty and the experimental uncertainties are very similar. No improvement of the experimental uncertainty could be achieved due to the more consistent treatment of systematic uncertainties in the fit, but also due to the residual model dependence of the unfolding procedure. The theoretical uncertainties are consistently determined by linear error propagation and may therefore be interpreted as a $1\,\sigma$ uncertainty. They turn out to be of similar size as those determined previously.
All values of $\alpha_s(M_Z)$ obtained are consistent with each other within their uncertainties. The values are compared to the world average value of $\alpha_s(M_Z) = 0.1184 \pm 0.0007$ in figures 10.11 and 10.12. The values from multijet cross sections are compatible with the world average value of α_s within the uncertainties. However, the values from inclusive dijet production cross sections prefer a lower value of $\alpha_s(M_Z)$. Overall, the experimental uncertainty from normalized cross sections is very competitive with the world average, whose precision is dominated by lattice QCD predictions, but higher-order QCD predictions beyond NLO would be necessary for taking full advantage of the precision of the data.

Chapter 11

Summary and outlook

In this work double-differential cross sections for jet production in neutral current deep-inelastic $e^{\pm}p$ scattering (NC DIS) in the kinematic range $150 < Q^2 < 15\,000\,\mathrm{GeV}^2$ and $0.2 < y < 0.7$ have been measured and the strong coupling constant $\alpha_s(M_Z)$ has been determined. Detector effects, such as limited acceptance and migrations from limited resolution, have been corrected for with a regularized unfolding method, based on a detector response matrix.

The migration matrix represents the detector effects of the inclusive jet, dijet and trijet measurement, as well as of the inclusive NC DIS cross sections. Thus the four measurements have been corrected for detector effects simultaneously and also the statistical correlations on detector level of these measurements have been considered. The prescriptions of the kinematic migrations have been optimized in a Monte Carlo (MC) study using the two MC event generators Django and Rapgap. For the description of migrations into and out of the phase space, it was enlarged in all variables where migrations are taking place, which also increased the stability of the measurement. The detector response of the inclusive jet measurement has been represented by a four-dimensional unfolding scheme in the kinematic variables Q^2, y, $p_{\mathrm{T}}^{\mathrm{jet}}$, and $\eta_{\mathrm{lab}}^{\mathrm{jet}}$. The unfolding schemes of the dijet and trijet measurements have been represented by a three-dimensional unfolding scheme, and additionally migrations in the kinematic variables defining the dijet or trijet phase space have been included. For the correction of the NC DIS measurement a two-dimensional unfolding scheme was used. A method has been developed for determining detector-level-only contributions, which are present for instance from different jet-multiplicities or from migrations into the extended phase space, using information of the underlying NC DIS kinematics.

For the first time multiple measurements have been corrected for with a regularized unfolding method simultaneously, where all statistical correlations are considered and are propagated analytically through the inverse detector response. A further feature of the unfolding method are the extensive studies which were performed on the variables and the highly multidimensional unfolding schemes with up to eight observables for a single measurement. Moreover, the correction method has been compared to the frequently applied bin-by-bin method, which showed a tendency to yield higher cross sections. It could be confirmed that this is not caused by the regularization condition or strength employed. The simultaneous unfolding further enables the determination of cross section ratios due to the availability of the statistical correlations. The ratios which have been measured are jet production cross sections normalized to the inclusive NC DIS cross sections and the three-jet to two-jet production cross sections. All cross sections measured are well-described by

theoretical predictions obtained from pQCD calculations in next-to-leading order with corrections for hadronization effects, applied. The jet cross sections presented will probably be the most precise measurement of these processes in DIS in the given kinematical region for the next decade, and probably even beyond.

The unfolded jet cross sections were inputs to the determination of the strong coupling constant $\alpha_s(M_Z)$ at the scale of the Z^0-mass with calculations in next-to-leading order precision. It is for the first time that fully unfolded cross sections with their covariance matrix have been employed in such a phenomenological analysis. Values of $\alpha_s(M_Z)$ have been determined from the individual jet and normalized jet cross sections and also simultaneously from multiple jet cross sections, where the statistical correlations between the jet measurements have been taken into account. For determining the uncertainties on $\alpha_s(M_Z)$ from limitations of the theoretical predictions a novel linear error propagation method has been developed.
The values for $\alpha_s(M_Z)$ obtained from the inclusive jet, dijet and trijet measurements yield an experimental precision of 1.5 to 2.1 %. The ones from the inclusive jet and trijet cross sections are $\alpha_s(M_Z) = 0.1172 \pm 0.0022$ and $\alpha_s(M_Z) = 0.1170 \pm 0.0018$, respectively, and they are consistent with the world average value, whereas the dijet cross section favors a significantly smaller value of $\alpha_s(M_Z) = 0.1129 \pm 0.0024$. The experimental precision of $\alpha_s(M_Z)$ is significantly higher for normalized cross sections, due to the cancellation of normalization uncertainties, which are highly correlated with the value of $\alpha_s(M_Z)$. The values obtained from normalized jet cross sections are larger by around 0.001 compared to the ones from absolute jet cross sections. This is due to a shift of the normalization in the fit of 1 to 2 %. The experimental uncertainty on $\alpha_s(M_Z)$ from the normalized jet cross sections reaches an accuracy of 0.8 to 1.2 %. In the fits to absolute jet as well as to normalized jet cross sections significant tension in the results of $\alpha_s(M_Z)$ from the dijet cross sections with the inclusive jet and trijet cross sections was observed. This is most likely caused by missing contributions of higher orders in the perturbative calculations.
The most precise value for $\alpha_s(M_Z)$ is obtained from the simultaneous fit to multiple cross sections, where the accuracy benefits from the highest statistical precision of the inclusive jet data and from the increased sensitivity on $\alpha_s(M_Z)$ of the trijet cross section. The value obtained for $\alpha_s(M_Z)$ in the fit to the three absolute jet cross sections is $\alpha_s(M_Z) = 0.1185 \pm 0.0017$ and the value obtained from the three normalized jet cross sections with its uncertainties from the theoretical predictions is

$$\alpha_s(M_Z) = 0.1165\,(8)_{\mathrm{exp}} \pm (4)_{\mathrm{PDF}} \pm (11)_{\mathrm{PDFset}} \pm (8)_{\mathrm{had}} \pm (34)_{\mu_r} \pm (5)_{\mu_f}\ .$$

Despite the high experimental precision and efforts, the uncertainty on $\alpha_s(M_Z)$ is entirely dominated by uncertainties from the missing higher orders in the perturbative calculation and hence the value of $\alpha_s(M_Z)$ cannot be determined better than 3 to 4 % in fits to the currently available theoretical predictions. As soon as higher-order calculations become available, the data presented will play an important role in the precise determination of PDFs of the proton and of the strong coupling constant $\alpha_s(M_Z)$.

Appendix A

Data tables

In this chapter the tables with the numerical values of the measured cross sections are given. Also the correlations between the bins of each measurement, and the correlations between the bins of the different measurements are listed. The correlations of the normalized cross section measurements are not given explicitly, since these are comparable to the correlations of the absolute cross section within 10 % (compare figure 7.5 and figure 7.9). The bin labels are defined for the measurements as a function of p_T in tables 7.1 and for the measurements as a function of ξ in table 7.2. An overview of the measurement phase space is given in table A.1.

Phase space

Measurement	**NC DIS phase space**	**Phase space of jets**	
Inclusive jet	$150 < Q^2 < 15\,000\,\text{GeV}^2$ $0.2 < y < 0.7$	$7 < p_T^{jet} < 50\,\text{GeV}$ $-1.0 < \eta_{lab}^{jet} < 2.5$	
Dijet $d\sigma/dQ^2 d\langle p_T\rangle_2$		$5 < p_T^{jet} < 50\,\text{GeV}$ $-1.0 < \eta_{lab}^{jet} < 2.5$ $M_{12} > 16\,\text{GeV}$	$N_{jet} \geq 2$ $7 < \langle p_T\rangle_2 < 50\,\text{GeV}$
Trijet $d\sigma/dQ^2 d\langle p_T\rangle_3$			$N_{jet} \geq 3$ $7 < \langle p_T\rangle_3 < 30\,\text{GeV}$
Dijet $d\sigma/dQ^2 d\xi_2$			$N_{jet} \geq 2$ $0.006 < \xi_2 < 0.316$
Trijet $d\sigma/dQ^2 d\xi_3$			$N_{jet} \geq 3$ $0.01 < \xi_3 < 0.50$

Table A.1: Summary of the phase space boundaries of the measurements presented.

Inclusive jet cross sections in bins of Q^2 and p_T^{jet}

Bin label	CS [pb]	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	$\delta_{ID(e)}$ [%]	c_{had}	δ_{had} [%]	c_{ew}
1α	$6.98\ 10^{1}$	2.7	3.0	+0.9	$^{+1.2}_{-1.2}$	$^{+1.2}_{-1.1}$	$^{-0.4}_{+0.6}$	$^{-0.3}_{+0.5}$	$^{+0.5}_{-0.5}$	0.93	2.2	1.00
1β	$3.05\ 10^{1}$	4.0	4.3	+2.4	$^{+2.8}_{-2.4}$	$^{+0.8}_{-0.7}$	$^{-0.1}_{+1.0}$	$^{-0.1}_{+0.7}$	$^{+0.5}_{-0.5}$	0.97	1.7	1.00
1γ	$7.82\ 10^{0}$	6.4	4.1	+1.7	$^{+2.5}_{-3.3}$	$^{+0.3}_{-0.3}$	$^{-0.9}_{-0.3}$	$^{-0.7}_{-0.2}$	$^{+0.5}_{-0.5}$	0.96	1.1	1.00
1δ	$8.36\ 10^{-1}$	19.5	10.3	+8.6	$^{+5.6}_{-4.5}$	$^{-0.7}_{-0.6}$	$^{-0.9}_{-0.8}$	$^{-0.4}_{-1.2}$	$^{+0.5}_{-0.5}$	0.95	0.7	1.00
2α	$5.37\ 10^{1}$	3.0	3.6	+2.0	$^{+0.9}_{-1.4}$	$^{+1.8}_{-1.2}$	$^{-0.6}_{+0.6}$	$^{-0.5}_{+0.5}$	$^{+0.5}_{-0.5}$	0.93	2.1	1.00
2β	$2.64\ 10^{1}$	4.1	4.2	+1.9	$^{+3.1}_{-2.7}$	$^{+0.5}_{-0.6}$	$^{-0.2}_{-0.2}$	$^{-0.5}_{-0.3}$	$^{+0.5}_{-0.5}$	0.97	1.7	1.00
2γ	$6.80\ 10^{0}$	6.6	4.8	+2.5	$^{+3.4}_{-3.3}$	$^{+0.7}_{-0.1}$	$^{+0.3}_{-0.0}$	$^{-0.2}_{+0.2}$	$^{+0.5}_{-0.5}$	0.97	1.3	1.00
2δ	$8.67\ 10^{-1}$	14.5	6.9	+2.4	$^{+5.8}_{-6.2}$	$^{+0.0}_{-0.1}$	$^{+0.3}_{-0.7}$	$^{-0.8}_{+0.3}$	$^{+0.5}_{-0.5}$	0.96	1.2	1.00
3α	$5.15\ 10^{1}$	3.0	3.4	+1.7	$^{+1.0}_{-1.0}$	$^{+1.5}_{-1.2}$	$^{-0.8}_{+0.5}$	$^{-0.3}_{+0.3}$	$^{+0.5}_{-0.5}$	0.93	1.5	1.00
3β	$2.78\ 10^{1}$	3.9	4.6	+3.3	$^{+1.8}_{-2.4}$	$^{+0.7}_{-0.5}$	$^{-0.5}_{+0.8}$	$^{-0.4}_{-0.2}$	$^{+0.5}_{-0.5}$	0.97	1.1	1.00
3γ	$6.80\ 10^{0}$	6.9	6.4	+4.2	$^{+3.8}_{-4.4}$	$^{+0.2}_{-0.9}$	$^{-0.8}_{-0.1}$	$^{-0.2}_{-0.3}$	$^{+0.5}_{-0.5}$	0.97	0.9	1.00
3δ	$8.11\ 10^{-1}$	16.2	8.8	−6.0	$^{+5.9}_{-5.8}$	$^{-0.4}_{-0.4}$	$^{+1.2}_{-2.2}$	$^{-0.5}_{+0.2}$	$^{+0.5}_{-0.5}$	0.95	0.5	1.00
4α	$4.83\ 10^{1}$	3.2	3.5	+1.6	$^{+1.5}_{-1.4}$	$^{+1.4}_{-1.0}$	$^{-0.9}_{+1.2}$	$^{-0.2}_{+0.2}$	$^{+0.4}_{-0.4}$	0.93	1.2	1.00
4β	$2.65\ 10^{1}$	4.1	3.7	+2.1	$^{+2.2}_{-1.4}$	$^{+1.0}_{-0.8}$	$^{-0.1}_{-0.2}$	$^{-0.2}_{-0.2}$	$^{+0.4}_{-0.4}$	0.97	1.0	1.00
4γ	$7.66\ 10^{0}$	6.2	5.7	+3.3	$^{+3.1}_{-4.6}$	$^{-0.6}_{-0.3}$	$^{-1.7}_{+0.9}$	$^{-0.1}_{-0.5}$	$^{+0.4}_{-0.4}$	0.97	0.5	1.00
4δ	$8.11\ 10^{-1}$	16.7	8.1	−4.8	$^{+6.8}_{-4.8}$	$^{+1.6}_{-1.7}$	$^{+1.8}_{+0.3}$	$^{+0.2}_{+0.7}$	$^{+0.4}_{-0.4}$	0.96	0.4	1.00
5α	$4.27\ 10^{1}$	3.5	2.9	+0.7	$^{+0.8}_{-1.1}$	$^{+0.7}_{-0.8}$	$^{-0.4}_{+0.3}$	$^{-0.5}_{+0.5}$	$^{+1.2}_{-1.2}$	0.92	0.9	1.02
5β	$2.81\ 10^{1}$	3.9	3.4	+1.4	$^{+1.5}_{-1.6}$	$^{+0.5}_{-0.1}$	$^{-0.9}_{+0.8}$	$^{-0.9}_{+0.3}$	$^{+1.2}_{-1.2}$	0.97	0.5	1.02
5γ	$1.07\ 10^{1}$	4.9	4.5	+2.2	$^{+3.1}_{-2.8}$	$^{+0.5}_{-0.2}$	$^{-0.7}_{+0.5}$	$^{-0.6}_{+0.4}$	$^{+1.3}_{-1.3}$	0.97	0.4	1.03
5δ	$1.97\ 10^{0}$	8.9	5.9	−2.3	$^{+4.0}_{-5.4}$	$^{-0.9}_{-0.4}$	$^{-0.5}_{-0.1}$	$^{-0.3}_{+0.1}$	$^{+1.4}_{-1.4}$	0.96	0.3	1.02
6α	$2.57\ 10^{0}$	14.4	6.3	−4.8	$^{-1.1}_{-0.7}$	$^{-0.8}_{-0.3}$	$^{-3.0}_{-2.2}$	$^{+0.1}_{-0.7}$	$^{+2.0}_{-2.0}$	0.91	0.6	1.11
6β	$1.66\ 10^{0}$	16.2	8.5	−7.6	$^{+0.3}_{-0.0}$	$^{-0.4}_{-0.6}$	$^{+2.0}_{-0.8}$	$^{+1.5}_{-1.6}$	$^{+2.0}_{-2.0}$	0.96	0.6	1.11
6γ	$6.58\ 10^{-1}$	20.4	7.4	−6.1	$^{+1.6}_{-3.6}$	$^{-0.3}_{-0.5}$	$^{-0.9}_{-0.7}$	$^{+0.4}_{-1.2}$	$^{+2.0}_{-2.0}$	0.99	1.1	1.11
6δ	$2.95\ 10^{-1}$	20.1	8.1	−6.5	$^{+5.1}_{-1.8}$	$^{+0.3}_{+1.5}$	$^{-0.7}_{+2.7}$	$^{-0.4}_{-0.4}$	$^{+2.0}_{-2.0}$	0.98	0.8	1.11

Table A.2: Double differential inclusive jet cross sections (CS) measured as function of Q^2 and p_T^{jet}. The bin labels are defined in table 7.1. The data points are statistically correlated, and the bin-to-bin correlations are shown in the correlation matrix in table A.12. The correlation with the dijet measurements as function of $\langle p_T \rangle$ and ξ are shown in table A.15 and A.20, respectively. The correlations with the trijet measurements as function of $\langle p_T \rangle$ and ξ are shown in table A.16 and A.21, respectively. The total systematic uncertainty δ_{sys} sums all systematic uncertainties in quadrature (the linear average of absolute values is taken in case of asymmetric uncertainties), where also the (not listed) uncertainty due to the LAr noise of 0.5 % and the total normalization uncertainty of $\delta_{Norm} = 2.9\,\%$ are considered. The latter consists of the uncertainty on the luminosity of $\delta_{Lumi} = 2.5\,\%$, the track-cluster-link uncertainty of $\delta_{TrkCl} = 1.0\,\%$ and the trigger uncertainty of $\delta_{Trig} = 1.2\,\%$. The size of the model uncertainty δ_{Model} is identical for the 'up' and 'down' shift, however, the sign can be opposite in adjacent bins and is therefore given explicitly. The correction factors on the theoretical cross sections c_{ew} and c_{had} are listed separately together with the uncertainty on the hadronization correction δ_{had} in the rightmost columns.

Normalized inclusive jet cross sections in bins of Q^2 and p_T^{jet}

Bin label	CS	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	c_{had}	δ_{had} [%]
1α	$1.64\ 10^{-1}$	2.7	1.9	+1.4	$^{+0.9}_{-0.9}$	$^{+0.6}_{-0.4}$	$^{+0.5}_{-0.7}$	$^{-0.2}_{+0.3}$	0.93	2.2
1β	$7.17\ 10^{-2}$	3.9	3.6	+2.7	$^{+2.5}_{-2.0}$	$^{+0.3}_{-0.0}$	$^{+0.2}_{-1.1}$	$^{-0.4}_{+0.3}$	0.97	1.7
1γ	$1.84\ 10^{-2}$	6.4	3.4	+2.1	$^{+2.3}_{-3.0}$	$^{-0.3}_{+0.4}$	$^{+1.0}_{+0.1}$	$^{+0.2}_{+0.1}$	0.96	1.1
1δ	$2.00\ 10^{-3}$	19.5	10.0	+8.7	$^{+5.3}_{-4.1}$	$^{-1.3}_{+0.0}$	$^{+1.0}_{+0.6}$	$^{-0.1}_{+0.3}$	0.95	0.7
2α	$1.73\ 10^{-1}$	2.9	2.9	+2.4	$^{+0.5}_{-1.0}$	$^{+1.3}_{-0.7}$	$^{+0.7}_{-0.8}$	$^{-0.1}_{-0.1}$	0.93	2.1
2β	$8.48\ 10^{-2}$	4.1	3.5	+2.3	$^{+2.7}_{-2.3}$	$^{-0.0}_{-0.1}$	$^{+0.4}_{+0.0}$	$^{-0.0}_{-0.5}$	0.97	1.7
2γ	$2.19\ 10^{-2}$	6.6	4.3	+3.0	$^{+3.0}_{-2.9}$	$^{+0.2}_{+0.5}$	$^{-0.1}_{-0.1}$	$^{-0.3}_{+0.0}$	0.97	1.3
2δ	$2.80\ 10^{-3}$	14.5	6.3	+2.8	$^{+5.3}_{-5.8}$	$^{-0.5}_{+0.4}$	$^{-0.1}_{+0.5}$	$^{+0.3}_{+0.1}$	0.96	1.2
3α	$1.90\ 10^{-1}$	3.0	2.6	+2.1	$^{+0.6}_{-0.5}$	$^{+1.1}_{-0.9}$	$^{+0.9}_{-0.6}$	$^{-0.1}_{+0.1}$	0.93	1.5
3β	$1.03\ 10^{-1}$	3.8	4.0	+3.5	$^{+1.3}_{-1.9}$	$^{+0.4}_{-0.2}$	$^{+0.6}_{-0.9}$	$^{+0.0}_{-0.4}$	0.97	1.1
3γ	$2.51\ 10^{-2}$	6.9	5.8	+4.4	$^{+3.4}_{-3.9}$	$^{-0.1}_{-0.5}$	$^{+0.9}_{-0.0}$	$^{-0.1}_{-0.0}$	0.97	0.9
3δ	$3.00\ 10^{-3}$	16.2	8.3	+6.1	$^{+5.5}_{-5.3}$	$^{-0.7}_{-0.1}$	$^{-1.1}_{+2.1}$	$^{+0.1}_{-0.2}$	0.95	0.5
4α	$2.23\ 10^{-1}$	3.1	2.7	+2.0	$^{+1.0}_{-0.8}$	$^{+1.2}_{-0.7}$	$^{+0.9}_{-1.3}$	$^{-0.1}_{-0.1}$	0.93	1.2
4β	$1.23\ 10^{-1}$	4.1	2.9	+2.5	$^{+1.6}_{-0.8}$	$^{+0.7}_{-0.6}$	$^{+0.1}_{+0.1}$	$^{+0.0}_{-0.3}$	0.97	1.0
4γ	$3.54\ 10^{-2}$	6.1	5.2	+3.7	$^{+2.6}_{-4.1}$	$^{-0.9}_{-0.1}$	$^{+1.7}_{-0.9}$	$^{-0.2}_{+0.4}$	0.97	0.5
4δ	$3.80\ 10^{-3}$	16.7	7.5	+5.0	$^{+6.2}_{-4.2}$	$^{+1.4}_{-1.4}$	$^{-1.8}_{-0.4}$	$^{-0.4}_{-0.2}$	0.96	0.4
5α	$2.40\ 10^{-1}$	3.4	2.0	+1.6	$^{+0.3}_{-0.6}$	$^{+0.6}_{-0.7}$	$^{+0.6}_{-0.5}$	$^{+0.0}_{+0.0}$	0.92	0.9
5β	$1.58\ 10^{-1}$	3.9	2.6	+2.0	$^{+1.1}_{-1.1}$	$^{+0.4}_{+0.0}$	$^{+1.0}_{-1.0}$	$^{+0.4}_{-0.4}$	0.97	0.5
5γ	$6.02\ 10^{-2}$	4.9	3.8	+2.7	$^{+2.6}_{-2.3}$	$^{+0.4}_{-0.1}$	$^{+0.8}_{-0.7}$	$^{+0.1}_{+0.1}$	0.97	0.4
5δ	$1.11\ 10^{-2}$	8.9	5.2	+2.9	$^{+3.6}_{-4.9}$	$^{-1.0}_{-0.3}$	$^{+0.7}_{-0.1}$	$^{-0.2}_{+0.2}$	0.96	0.3
6α	$3.05\ 10^{-1}$	14.2	6.0	+5.3	$^{-1.5}_{-0.2}$	$^{-0.8}_{-0.2}$	$^{+3.4}_{+1.3}$	$^{-0.3}_{+1.8}$	0.91	0.6
6β	$1.97\ 10^{-1}$	16.1	8.2	+8.0	$^{-0.0}_{+0.5}$	$^{-0.4}_{-0.5}$	$^{-1.6}_{-0.1}$	$^{-1.8}_{-0.9}$	0.96	0.6
6γ	$7.80\ 10^{-2}$	20.3	7.3	+6.8	$^{+1.3}_{-3.1}$	$^{-0.3}_{-0.5}$	$^{+1.3}_{-0.1}$	$^{-0.7}_{-0.4}$	0.99	1.1
6δ	$3.50\ 10^{-2}$	19.9	8.3	+7.3	$^{+4.7}_{-1.3}$	$^{+0.3}_{+1.6}$	$^{+1.1}_{-3.6}$	$^{+0.1}_{+0.1}$	0.98	0.8

Table A.3: Double differential normalized inclusive jet cross sections measured as function of Q^2 and p_T^{jet}. The bin labels are defined in table 7.1. The statistical uncertainty is shown in the third column. The bin-to-bin correlations are of comparable size to the ones from the absolute jet cross section from table A.12. The total systematic uncertainty sums all systematic uncertainties in quadrature, where also the (not listed) uncertainty due to the LAr noise of 0.5 % is considered. The normalization uncertainties cancel in the cross section ratio. The hadronization correction factors c_{had} applied to the theory cross sections are listed separately together with their uncertainty in the rightmost columns.

Dijet cross sections in bins of Q^2 and $\langle p_T \rangle_2$

Bin label	CS [pb]	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	$\delta_{ID(e)}$ [%]	c_{had}	δ_{had} [%]	c_{ew}
1α	$2.32\ 10^{1}$	3.6	3.9	+1.9	$^{+0.5}_{-0.2}$	$^{+2.5}_{-2.2}$	$^{+0.2}_{+0.4}$	$^{-0.1}_{+0.7}$	$^{+0.5}_{-0.5}$	0.94	2.0	1.00
1β	$1.33\ 10^{1}$	5.8	3.8	+2.4	$^{+1.2}_{-2.4}$	$^{+0.1}_{-0.8}$	$^{-0.2}_{+0.1}$	$^{-0.3}_{+0.8}$	$^{+0.5}_{-0.5}$	0.97	1.4	1.00
1γ	$3.46\ 10^{0}$	6.7	5.4	+2.5	$^{+4.2}_{-4.1}$	$^{+0.3}_{-0.2}$	$^{-0.2}_{-0.6}$	$^{-0.5}_{-0.1}$	$^{+0.5}_{-0.5}$	0.96	1.0	1.00
1δ	$4.36\ 10^{-1}$	16.0	8.7	+6.2	$^{+5.2}_{-5.8}$	$^{-0.4}_{+0.0}$	$^{-0.6}_{-0.5}$	$^{+0.1}_{-0.9}$	$^{+0.5}_{-0.5}$	0.96	1.2	1.00
2α	$1.77\ 10^{1}$	4.1	3.6	+1.0	$^{+0.8}_{-0.4}$	$^{+2.3}_{-2.2}$	$^{+0.2}_{+1.3}$	$^{+0.1}_{+1.5}$	$^{+0.5}_{-0.5}$	0.94	1.7	1.00
2β	$1.23\ 10^{1}$	5.4	5.4	+4.3	$^{+1.3}_{-2.3}$	$^{+0.4}_{-0.5}$	$^{-1.2}_{-0.8}$	$^{-1.1}_{-1.3}$	$^{+0.5}_{-0.5}$	0.98	1.6	1.00
2γ	$2.88\ 10^{0}$	7.2	5.7	+3.7	$^{+3.7}_{-3.6}$	$^{+0.0}_{-0.2}$	$^{-0.1}_{-0.2}$	$^{-0.2}_{+0.3}$	$^{+0.5}_{-0.5}$	0.97	1.0	1.00
2δ	$3.97\ 10^{-1}$	17.1	7.8	+4.4	$^{+7.8}_{-4.2}$	$^{+0.1}_{+0.3}$	$^{+0.9}_{-0.7}$	$^{-0.6}_{+0.2}$	$^{+0.5}_{-0.5}$	0.95	1.9	1.00
3α	$1.82\ 10^{1}$	3.9	3.4	−0.6	$^{+0.5}_{-0.0}$	$^{+2.3}_{-2.1}$	$^{+0.1}_{+1.3}$	$^{+0.2}_{+1.3}$	$^{+0.5}_{-0.5}$	0.93	1.2	1.00
3β	$1.12\ 10^{1}$	6.0	6.1	+5.2	$^{+1.3}_{-2.0}$	$^{+0.7}_{-0.5}$	$^{-1.0}_{-0.5}$	$^{-0.9}_{-1.4}$	$^{+0.5}_{-0.5}$	0.98	0.9	1.00
3γ	$3.73\ 10^{0}$	6.0	5.1	+2.6	$^{+3.6}_{-4.0}$	$^{-0.0}_{-0.2}$	$^{-0.4}_{-0.3}$	$^{-0.3}_{-0.2}$	$^{+0.5}_{-0.5}$	0.97	0.8	1.00
3δ	$3.19\ 10^{-1}$	22.8	11.2	−9.5	$^{+4.6}_{-6.0}$	$^{-0.7}_{-0.4}$	$^{+1.2}_{-1.2}$	$^{-0.2}_{+0.2}$	$^{+0.5}_{-0.5}$	0.96	0.4	1.00
4α	$1.66\ 10^{1}$	4.1	3.2	−0.5	$^{+0.5}_{+0.2}$	$^{+2.1}_{-1.3}$	$^{+0.7}_{+1.6}$	$^{+0.4}_{+1.5}$	$^{+0.4}_{-0.4}$	0.92	1.1	1.00
4β	$1.05\ 10^{1}$	6.4	4.9	+3.4	$^{+2.0}_{-2.4}$	$^{+0.0}_{-0.5}$	$^{-1.3}_{-1.0}$	$^{-0.6}_{-1.5}$	$^{+0.5}_{-0.5}$	0.97	0.9	1.00
4γ	$3.52\ 10^{0}$	6.4	3.9	+1.1	$^{+2.9}_{-3.0}$	$^{+1.0}_{-0.3}$	$^{-0.1}_{+0.6}$	$^{-0.1}_{+0.3}$	$^{+0.4}_{-0.4}$	0.98	0.5	1.00
4δ	$3.68\ 10^{-1}$	20.3	7.2	+3.9	$^{+2.9}_{-7.8}$	$^{-0.1}_{-1.3}$	$^{+2.3}_{-0.1}$	$^{+0.2}_{+0.1}$	$^{+0.4}_{-0.4}$	0.96	0.3	1.00
5α	$1.47\ 10^{1}$	4.5	3.5	−1.5	$^{-0.2}_{+0.6}$	$^{+1.5}_{-1.2}$	$^{+1.7}_{+0.3}$	$^{+0.2}_{+1.5}$	$^{+1.2}_{-1.2}$	0.92	0.6	1.02
5β	$1.31\ 10^{1}$	5.2	4.2	+2.8	$^{+1.3}_{-1.7}$	$^{+0.3}_{-0.5}$	$^{-1.1}_{-0.6}$	$^{-0.8}_{-0.7}$	$^{+1.2}_{-1.2}$	0.96	0.3	1.02
5γ	$4.80\ 10^{0}$	5.4	4.7	+3.2	$^{+2.2}_{-2.4}$	$^{+0.3}_{-0.1}$	$^{-0.2}_{+0.1}$	$^{-0.7}_{+0.2}$	$^{+1.3}_{-1.3}$	0.98	0.4	1.03
5δ	$9.46\ 10^{-1}$	10.4	5.9	−1.6	$^{+3.4}_{-6.5}$	$^{-1.5}_{-0.2}$	$^{-1.0}_{-0.0}$	$^{-0.2}_{+0.5}$	$^{+1.4}_{-1.4}$	0.96	0.7	1.01
6α	$7.33\ 10^{-1}$	22.3	12.8	+10.6	$^{-1.2}_{+3.5}$	$^{-1.9}_{+0.5}$	$^{+6.2}_{-2.6}$	$^{+6.6}_{+1.1}$	$^{+2.2}_{-2.2}$	0.89	0.2	1.11
6β	$8.46\ 10^{-1}$	19.8	10.2	−8.7	$^{+4.0}_{+0.1}$	$^{+2.7}_{+2.1}$	$^{-2.3}_{+1.5}$	$^{-3.6}_{-0.6}$	$^{+2.0}_{-2.0}$	0.95	0.5	1.11
6γ	$3.16\ 10^{-1}$	21.7	5.5	−3.6	$^{-0.2}_{-3.8}$	$^{-0.4}_{-1.3}$	$^{-0.1}_{-3.5}$	$^{+0.9}_{-0.9}$	$^{+2.0}_{-2.0}$	0.97	0.8	1.11
6δ	$1.55\ 10^{-1}$	24.4	14.0	−12.8	$^{+5.6}_{+2.0}$	$^{-1.2}_{+3.6}$	$^{+0.7}_{+0.6}$	$^{+0.9}_{-1.4}$	$^{+2.0}_{-2.0}$	0.98	1.0	1.11

Table A.4: Double differential dijet cross sections measured as function of Q^2 and $\langle p_T \rangle_2$. Not shown are the total normalization uncertainty of $\delta_{Norm} = 2.9\,\%$ and the LAr noise uncertainty of $\delta_{LArNoise} = 0.6\,\%$. The correlations between the data points are listed in table A.13. The statistical correlations with the trijet measurement as function of $\langle p_T \rangle$ are listed in table A.17. Further details are given in the caption of table A.2.

Dijet cross sections in bins of Q^2 and ξ_2

Bin label	CS [pb]	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	$\delta_{\text{ID(e)}}$ [%]	c_{had}	δ_{had} [%]	c_{ew}
1a	2.09 10^1	3.9	5.9	+4.8	$^{+0.1}_{+0.0}$	$^{+2.5}_{-2.8}$	$^{-0.2}_{-0.1}$	$^{-0.5}_{+0.3}$	$^{+0.5}_{-0.5}$	0.94	2.1	1.00
1b	1.79 10^1	3.5	5.0	+3.6	$^{+1.6}_{-1.9}$	$^{+2.3}_{-2.0}$	$^{-0.0}_{-0.2}$	$^{+0.1}_{+0.1}$	$^{+0.5}_{-0.5}$	0.94	1.7	1.00
1c	5.77 10^0	7.5	8.7	−7.8	$^{+3.1}_{-2.9}$	$^{-0.1}_{+0.3}$	$^{-0.1}_{-0.1}$	$^{-1.0}_{-0.4}$	$^{+0.5}_{-0.5}$	0.94	1.3	1.00
1d	1.87 10^0	9.4	13.7	+12.8	$^{+3.8}_{-4.4}$	$^{-0.8}_{+0.5}$	$^{-2.0}_{+1.0}$	$^{-0.4}_{+0.2}$	$^{+0.5}_{-0.5}$	0.92	0.7	1.00
2a	1.48 10^1	3.9	6.3	−5.4	$^{+1.0}_{-0.2}$	$^{+2.1}_{-2.3}$	$^{-0.4}_{+0.3}$	$^{-0.3}_{+0.4}$	$^{+0.5}_{-0.5}$	0.94	1.8	1.00
2b	1.54 10^1	3.6	4.3	+2.8	$^{+1.2}_{-1.6}$	$^{+1.9}_{-1.8}$	$^{+0.5}_{-0.8}$	$^{-0.4}_{+0.4}$	$^{+0.5}_{-0.5}$	0.94	1.7	1.00
2c	6.01 10^0	6.3	9.4	+8.7	$^{+2.6}_{-2.7}$	$^{+0.4}_{-0.2}$	$^{+0.4}_{+0.3}$	$^{-0.0}_{+0.0}$	$^{+0.5}_{-0.5}$	0.94	1.1	1.00
2d	1.64 10^0	9.5	27.5	+27.2	$^{+3.5}_{-2.2}$	$^{+0.1}_{+1.2}$	$^{-0.7}_{+1.4}$	$^{-0.2}_{+1.3}$	$^{+0.5}_{-0.5}$	0.93	0.6	1.00
3a	1.08 10^1	5.0	5.5	+4.5	$^{+1.1}_{-0.4}$	$^{+2.2}_{-1.7}$	$^{-0.2}_{+0.8}$	$^{-0.2}_{+0.4}$	$^{+0.5}_{-0.5}$	0.93	1.4	1.00
3b	1.74 10^1	3.1	5.0	+3.8	$^{+1.2}_{-1.5}$	$^{+1.7}_{-2.2}$	$^{+0.2}_{-0.7}$	$^{-0.3}_{+0.2}$	$^{+0.5}_{-0.5}$	0.94	1.2	1.00
3c	8.20 10^0	4.6	3.8	−2.2	$^{+2.1}_{-1.9}$	$^{+1.0}_{-0.5}$	$^{+0.7}_{-0.3}$	$^{-0.2}_{+0.3}$	$^{+0.5}_{-0.5}$	0.94	0.9	1.00
3d	1.98 10^0	8.3	5.8	+4.2	$^{+3.2}_{-3.3}$	$^{-0.8}_{+0.1}$	$^{-0.8}_{+0.0}$	$^{-0.7}_{+0.2}$	$^{+0.5}_{-0.5}$	0.94	0.4	1.00
4a	4.91 10^0	7.7	4.9	+3.7	$^{+1.4}_{-1.0}$	$^{+2.1}_{-1.3}$	$^{-0.8}_{+0.6}$	$^{-0.1}_{+0.2}$	$^{+0.5}_{-0.5}$	0.92	1.4	1.00
4b	1.70 10^1	3.1	4.5	+3.5	$^{+0.6}_{-1.0}$	$^{+1.3}_{-1.7}$	$^{+0.1}_{-0.3}$	$^{-0.1}_{+0.1}$	$^{+0.4}_{-0.4}$	0.93	1.2	1.00
4c	1.12 10^1	3.7	7.0	+6.2	$^{+1.7}_{-1.7}$	$^{+1.4}_{-1.1}$	$^{+0.7}_{-0.7}$	$^{-0.0}_{+0.2}$	$^{+0.4}_{-0.4}$	0.94	0.8	1.00
4d	2.38 10^0	7.8	9.8	+8.9	$^{+3.6}_{-3.3}$	$^{+0.7}_{-0.3}$	$^{-0.2}_{+0.6}$	$^{-0.2}_{+0.3}$	$^{+0.4}_{-0.4}$	0.95	0.5	1.00
5b	8.66 10^0	3.7	4.4	+3.3	$^{+1.1}_{-1.1}$	$^{+1.1}_{-1.0}$	$^{+0.1}_{+0.1}$	$^{-0.4}_{+0.5}$	$^{+1.2}_{-1.2}$	0.92	0.5	1.01
5c	1.68 10^1	3.0	4.0	+2.6	$^{+1.0}_{-1.0}$	$^{+1.0}_{-1.5}$	$^{+0.6}_{-0.8}$	$^{-0.5}_{+0.2}$	$^{+1.1}_{-1.1}$	0.93	0.5	1.02
5d	1.10 10^1	3.0	3.7	+1.9	$^{+1.8}_{-1.6}$	$^{+1.1}_{-0.8}$	$^{-0.2}_{-0.2}$	$^{-0.4}_{+0.2}$	$^{+1.2}_{-1.2}$	0.94	0.4	1.03
6d	1.91 10^0	6.7	6.1	+5.2	$^{+0.3}_{-0.2}$	$^{+1.0}_{-1.0}$	$^{+0.6}_{-1.0}$	$^{+0.6}_{+0.2}$	$^{+2.0}_{-2.0}$	0.93	0.8	1.11

Table A.5: Double differential dijet cross sections measured as function of Q^2 and ξ_2. The bin labels of the first column are defined in table 7.2. Not shown are the total normalization uncertainty of $\delta_{\text{Norm}} = 2.9\,\%$ and the LAr noise uncertainty of $\delta_{\text{LArNoise}} = 0.6\,\%$. The correlations between the data points are listed in table A.18. The statistical correlations with the trijet measurement as function of ξ are listed in table A.22, the correlations with the inclusive jet measurement are given in table A.20.

Normalized dijet cross sections in bins of Q^2 and $\langle p_T \rangle_2$

Bin label	CS	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	c_{had}	δ_{had} [%]
1α	5.46 10^{-2}	3.5	3.0	+2.2	$^{+0.4}_{-0.2}$	$^{+2.0}_{-1.6}$	$^{+0.2}_{-0.6}$	$^{-0.4}_{+0.4}$	0.94	2.0
1β	3.13 10^{-2}	5.8	2.9	+2.7	$^{+0.4}_{+0.7}$	$^{-0.5}_{-0.1}$	$^{-0.1}_{-0.3}$	$^{-0.2}_{+0.6}$	0.97	1.4
1γ	8.10 10^{-3}	6.7	3.5	+2.8	$^{+1.6}_{+2.3}$	$^{-0.2}_{+0.4}$	$^{-0.0}_{+0.4}$	$^{-0.0}_{+0.1}$	0.96	1.0
1δ	1.00 10^{-3}	16.0	6.8	+6.3	$^{+1.3}_{-3.4}$	$^{-0.9}_{+0.7}$	$^{-0.3}_{+0.3}$	$^{-0.6}_{+0.9}$	0.96	1.2
2α	5.70 10^{-2}	4.0	2.9	+1.7	$^{+1.1}_{-0.2}$	$^{+1.8}_{-1.7}$	$^{+0.5}_{-1.5}$	$^{-0.6}_{+1.1}$	0.94	1.7
2β	3.96 10^{-2}	5.4	4.8	+4.5	$^{-0.3}_{+0.4}$	$^{-0.2}_{+0.0}$	$^{-1.1}_{+0.6}$	$^{+0.6}_{-1.9}$	0.98	1.6
2γ	9.30 10^{-3}	7.2	4.2	+3.9	$^{+1.3}_{+1.3}$	$^{-0.5}_{+0.4}$	$^{-0.0}_{-0.0}$	$^{-0.3}_{+0.1}$	0.97	1.0
2δ	1.30 10^{-3}	17.1	5.2	+4.6	$^{+2.8}_{+1.4}$	$^{-0.4}_{+0.8}$	$^{-0.1}_{+0.5}$	$^{+0.1}_{+0.4}$	0.95	1.9
3α	6.71 10^{-2}	3.8	2.8	+1.4	$^{+0.6}_{-0.1}$	$^{+2.0}_{-1.8}$	$^{+0.5}_{-1.4}$	$^{-0.6}_{+1.1}$	0.93	1.2
3β	4.15 10^{-2}	6.0	5.5	+5.3	$^{-0.2}_{+0.1}$	$^{+0.3}_{-0.2}$	$^{-1.1}_{+0.4}$	$^{+0.6}_{-1.5}$	0.98	0.9
3γ	1.38 10^{-2}	6.0	3.2	+2.9	$^{+1.1}_{+1.4}$	$^{-0.4}_{+0.1}$	$^{+0.2}_{+0.2}$	$^{-0.0}_{-0.1}$	0.97	0.8
3δ	1.20 10^{-3}	22.8	10.1	+9.6	$^{+1.4}_{+4.4}$	$^{-1.1}_{-0.1}$	$^{-0.4}_{+1.1}$	$^{-0.2}_{-0.1}$	0.96	0.4
4α	7.69 10^{-2}	4.0	2.6	+1.4	$^{+0.4}_{-0.6}$	$^{+1.9}_{-1.0}$	$^{+0.5}_{-1.6}$	$^{-0.6}_{+1.0}$	0.92	1.1
4β	4.84 10^{-2}	6.3	4.1	+3.7	$^{+1.0}_{+0.7}$	$^{-0.2}_{-0.3}$	$^{-0.9}_{+0.9}$	$^{+0.4}_{-1.2}$	0.97	0.9
4γ	1.63 10^{-2}	6.3	2.1	+1.8	$^{+0.4}_{+1.0}$	$^{+0.8}_{-0.1}$	$^{+0.2}_{-0.7}$	$^{-0.2}_{+0.0}$	0.98	0.5
4δ	1.70 10^{-3}	20.3	5.2	+4.2	$^{-0.7}_{+4.7}$	$^{-0.4}_{-1.0}$	$^{+1.5}_{+0.1}$	$^{-0.4}_{+0.2}$	0.96	0.3
5α	8.26 10^{-2}	4.4	3.0	+2.2	$^{+0.5}_{-0.6}$	$^{+1.4}_{-1.1}$	$^{+1.1}_{-0.4}$	$^{-0.7}_{+1.3}$	0.92	0.6
5β	7.36 10^{-2}	5.2	3.5	+3.1	$^{-0.5}_{+0.8}$	$^{+0.2}_{-0.4}$	$^{-1.2}_{+0.4}$	$^{+0.3}_{-1.4}$	0.96	0.3
5γ	2.70 10^{-2}	5.3	3.6	+3.5	$^{+0.4}_{+0.6}$	$^{+0.2}_{+0.0}$	$^{+0.5}_{-0.3}$	$^{+0.2}_{+0.2}$	0.98	0.4
5δ	5.30 10^{-3}	10.4	3.2	+2.5	$^{-0.1}_{+3.4}$	$^{-1.6}_{-0.1}$	$^{+0.2}_{-0.2}$	$^{-0.3}_{+0.2}$	0.96	0.7
6α	8.70 10^{-2}	22.1	11.9	+10.8	$^{-1.3}_{-1.8}$	$^{-1.9}_{+0.6}$	$^{+1.7}_{+1.8}$	$^{-6.9}_{+1.1}$	0.89	0.2
6β	1.00 10^{-1}	19.7	10.5	+9.1	$^{+4.3}_{-0.1}$	$^{+2.7}_{+2.2}$	$^{-3.1}_{-2.4}$	$^{+3.4}_{+2.4}$	0.95	0.5
6γ	3.75 10^{-2}	21.5	5.5	+4.5	$^{-2.6}_{+1.8}$	$^{-0.4}_{-1.2}$	$^{-1.1}_{+2.7}$	$^{-1.1}_{-0.3}$	0.97	0.8
6δ	1.84 10^{-2}	24.3	13.8	+13.1	$^{-1.0}_{-4.3}$	$^{-1.3}_{+3.7}$	$^{+1.7}_{-1.5}$	$^{-1.2}_{-1.3}$	0.98	1.0

Table A.6: Double differential normalized dijet cross sections measured as function of Q^2 and $\langle p_T \rangle_2$. Not shown is the LAr noise uncertainty of $\delta_{LArNoise} = 0.6\,\%$. Further details are given in the caption of table A.3.

Normalized dijet cross sections in bins of Q^2 and ξ_2

Bin label	CS	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	$\delta_{\text{E}_\text{e}}$ [%]	δ_{θ_e} [%]	c_{had}	δ_{had} [%]
1a	$4.92\ 10^{-2}$	3.9	5.3	+4.9	$^{+0.3}_{-0.6}$	$^{+2.0}_{-2.2}$	$^{+0.1}_{-0.4}$	$^{-0.1}_{+0.2}$	0.94	2.1
1b	$4.23\ 10^{-2}$	3.4	4.3	+3.9	$^{-0.5}_{+0.8}$	$^{+1.7}_{-1.4}$	$^{-0.2}_{+0.3}$	$^{+0.2}_{+0.3}$	0.94	1.7
1c	$1.36\ 10^{-2}$	7.5	8.2	+8.1	$^{-0.8}_{+0.5}$	$^{-0.6}_{+1.0}$	$^{+0.2}_{-0.3}$	$^{-0.2}_{+0.8}$	0.94	1.3
1d	$4.40\ 10^{-3}$	9.4	13.1	+13.0	$^{-0.8}_{+0.9}$	$^{-1.4}_{+1.2}$	$^{+1.1}_{+0.2}$	$^{+0.3}_{+0.3}$	0.92	0.7
2a	$4.78\ 10^{-2}$	3.9	5.9	+5.6	$^{-0.4}_{-0.2}$	$^{+1.5}_{-1.8}$	$^{+0.4}_{-0.2}$	$^{+0.1}_{+0.1}$	0.94	1.8
2b	$4.97\ 10^{-2}$	3.6	3.4	+3.0	$^{-0.2}_{+0.4}$	$^{+1.4}_{-1.2}$	$^{-0.3}_{+0.2}$	$^{+0.1}_{+0.0}$	0.94	1.7
2c	$1.93\ 10^{-2}$	6.3	8.9	+8.8	$^{-0.2}_{+0.2}$	$^{-0.2}_{+0.4}$	$^{-0.5}_{+0.1}$	$^{+0.1}_{+0.4}$	0.94	1.1
2d	$5.30\ 10^{-3}$	9.5	27.3	+27.3	$^{-0.2}_{+1.2}$	$^{-0.5}_{+1.8}$	$^{-0.0}_{+0.2}$	$^{+0.1}_{-0.8}$	0.93	0.6
3a	$3.98\ 10^{-2}$	5.0	5.0	+4.7	$^{-0.1}_{-0.2}$	$^{+1.8}_{-1.3}$	$^{-0.4}_{-0.9}$	$^{+0.0}_{-0.1}$	0.93	1.4
3b	$6.42\ 10^{-2}$	3.0	4.3	+4.0	$^{-0.3}_{+0.3}$	$^{+1.3}_{-1.8}$	$^{-0.0}_{+0.4}$	$^{-0.1}_{+0.2}$	0.94	1.2
3c	$3.03\ 10^{-2}$	4.6	2.5	+2.4	$^{-0.3}_{-0.1}$	$^{+0.6}_{-0.1}$	$^{-0.2}_{+0.2}$	$^{+0.1}_{+0.0}$	0.94	0.9
3d	$7.30\ 10^{-3}$	8.3	4.6	+4.3	$^{-1.1}_{+1.2}$	$^{-1.2}_{+0.5}$	$^{+0.7}_{+0.2}$	$^{-0.2}_{+0.2}$	0.94	0.4
4a	$2.27\ 10^{-2}$	7.6	4.5	+4.2	$^{-0.8}_{+0.1}$	$^{+1.8}_{-1.0}$	$^{+0.4}_{-0.6}$	$^{-0.1}_{-0.0}$	0.92	1.4
4b	$7.87\ 10^{-2}$	3.1	4.2	+4.0	$^{+0.0}_{+0.1}$	$^{+1.0}_{-1.4}$	$^{-0.2}_{+0.2}$	$^{-0.0}_{+0.1}$	0.93	1.2
4c	$5.16\ 10^{-2}$	3.7	6.5	+6.4	$^{-0.1}_{+0.2}$	$^{+1.1}_{-0.8}$	$^{-0.0}_{+0.2}$	$^{-0.0}_{+0.0}$	0.94	0.8
4d	$1.10\ 10^{-2}$	7.8	9.1	+9.1	$^{-0.5}_{+0.8}$	$^{+0.4}_{+0.0}$	$^{+0.4}_{+0.2}$	$^{+0.1}_{-0.1}$	0.95	0.5
5b	$4.86\ 10^{-2}$	3.6	3.9	+3.8	$^{-0.2}_{+0.4}$	$^{+0.9}_{-0.9}$	$^{+0.0}_{+0.0}$	$^{-0.0}_{-0.1}$	0.92	0.5
5c	$9.45\ 10^{-2}$	3.0	3.5	+3.3	$^{+0.1}_{-0.1}$	$^{+0.8}_{-1.3}$	$^{-0.1}_{+0.4}$	$^{-0.0}_{+0.2}$	0.93	0.5
5d	$6.17\ 10^{-2}$	2.9	2.8	+2.6	$^{-0.5}_{+0.5}$	$^{+0.9}_{-0.7}$	$^{-0.1}_{-0.1}$	$^{-0.0}_{+0.3}$	0.94	0.4
6d	$2.26\ 10^{-1}$	6.3	6.4	+6.2	$^{-0.9}_{-0.9}$	$^{+1.0}_{-0.9}$	$^{-0.1}_{-1.0}$	$^{+0.0}_{+0.0}$	0.93	0.8

Table A.7: Double differential normalized inclusive dijet cross sections measured as function of Q^2 and ξ_2. The bin labels of the first column are defined in table 7.2. Not shown is the LAr noise uncertainty of $\delta_{\text{LArNoise}} = 0.6\,\%$. Further details are given in the caption of tables A.20.

Trijet cross sections in bins of Q^2 and $\langle p_T \rangle_3$

Bin label	CS [pb]	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	$\delta_{ID(e)}$ [%]	c_{had}	δ_{had} [%]	c_{ew}
1α	4.84 10^0	8.7	9.1	+1.2	$^{-2.2}_{+1.1}$	$^{+7.4}_{-6.5}$	$^{-2.4}_{-4.8}$	$^{-3.0}_{-4.0}$	$^{+0.4}_{-0.4}$	0.79	5.3	1.00
1β	2.65 10^0	8.4	5.1	+2.3	$^{+3.2}_{-3.4}$	$^{+1.8}_{-1.5}$	$^{+0.6}_{+0.6}$	$^{+0.4}_{+1.2}$	$^{+0.5}_{-0.5}$	0.85	4.3	1.00
1γ	4.02 10^{-1}	19.9	7.5	+4.2	$^{+4.0}_{-6.3}$	$^{+0.6}_{-2.8}$	$^{-1.1}_{-1.8}$	$^{-1.0}_{-1.0}$	$^{+0.5}_{-0.5}$	0.89	3.6	1.00
2α	3.34 10^0	10.9	8.4	−3.6	$^{-1.7}_{+0.6}$	$^{+7.0}_{-7.2}$	$^{-1.3}_{-0.1}$	$^{-0.8}_{-1.4}$	$^{+0.5}_{-0.5}$	0.78	5.0	1.00
2β	2.00 10^0	9.4	6.5	+4.8	$^{+2.8}_{-3.0}$	$^{+2.6}_{-1.9}$	$^{+0.4}_{+0.7}$	$^{-0.2}_{+0.7}$	$^{+0.5}_{-0.5}$	0.84	4.4	1.00
2γ	4.38 10^{-1}	16.7	6.0	−2.2	$^{+5.6}_{-4.3}$	$^{+0.6}_{-0.8}$	$^{+1.1}_{-0.3}$	$^{+0.1}_{+0.6}$	$^{+0.5}_{-0.5}$	0.89	2.7	1.00
3α	3.50 10^0	10.2	8.1	−2.9	$^{-1.8}_{+0.3}$	$^{+6.8}_{-6.9}$	$^{-2.0}_{-0.6}$	$^{-1.4}_{-1.4}$	$^{+0.5}_{-0.5}$	0.78	4.6	1.00
3β	2.64 10^0	8.0	6.1	+4.3	$^{+3.3}_{-2.3}$	$^{+2.1}_{-1.5}$	$^{-0.6}_{+2.0}$	$^{+0.8}_{+0.4}$	$^{+0.5}_{-0.5}$	0.85	3.7	1.00
3γ	4.80 10^{-1}	17.3	7.5	−4.3	$^{+4.3}_{-6.2}$	$^{+0.5}_{-2.6}$	$^{-1.1}_{-1.2}$	$^{-0.1}_{-1.0}$	$^{+0.5}_{-0.5}$	0.87	2.3	1.00
4α	3.13 10^0	10.7	10.5	−7.7	$^{-1.7}_{+0.5}$	$^{+6.3}_{-6.6}$	$^{-1.4}_{-1.5}$	$^{-0.8}_{-1.6}$	$^{+0.4}_{-0.4}$	0.77	4.1	1.00
4β	2.82 10^0	7.3	6.3	+5.2	$^{+2.4}_{-1.5}$	$^{+1.3}_{-2.2}$	$^{-0.2}_{+0.9}$	$^{+0.2}_{+0.8}$	$^{+0.4}_{-0.4}$	0.85	3.6	1.00
4γ	6.49 10^{-1}	14.4	8.5	+3.5	$^{+6.3}_{-7.6}$	$^{+3.1}_{-1.8}$	$^{+0.2}_{+1.0}$	$^{-0.1}_{-0.5}$	$^{+0.4}_{-0.4}$	0.87	2.3	1.00
5α	3.19 10^0	9.9	11.9	−10.5	$^{-2.2}_{+1.2}$	$^{+4.0}_{-4.3}$	$^{-0.6}_{-2.5}$	$^{-1.6}_{-0.8}$	$^{+1.6}_{-1.6}$	0.77	3.5	1.03
5β	2.92 10^0	7.3	5.0	+3.2	$^{+1.9}_{-1.9}$	$^{+1.9}_{-1.7}$	$^{+0.2}_{+1.3}$	$^{-0.2}_{+1.2}$	$^{+1.4}_{-1.4}$	0.83	2.9	1.03
5γ	7.52 10^{-1}	12.9	5.6	+2.2	$^{+4.3}_{-4.1}$	$^{+0.6}_{-0.5}$	$^{-1.6}_{+0.4}$	$^{-0.1}_{-0.6}$	$^{+1.4}_{-1.4}$	0.86	2.2	1.03
6β	1.40 10^{-1}	34.9	11.1	−8.5	$^{+0.6}_{+3.5}$	$^{+3.0}_{+2.2}$	$^{+4.7}_{+5.5}$	$^{-0.1}_{+3.8}$	$^{+2.3}_{-2.3}$	0.82	0.8	1.12

Table A.8: Double differential trijet cross sections measured as function of Q^2 and $\langle p_T \rangle_3$. Not shown are the total normalization uncertainty of $\delta_{Norm} = 2.9\,\%$ and the LAr noise uncertainty of $\delta_{LArNoise} = 0.9\,\%$. The correlations between the data points are listed in table A.14. Further details are given in the caption of table A.2.

Trijet cross sections in bins of Q^2 and ξ_3

Bin label	CS [pb]	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	$\delta_{ID(e)}$ [%]	c_{had}	δ_{had} [%]	c_{ew}
1A	3.09 10^0	11.1	25.9	+24.1	$^{-0.2}_{-0.1}$	$^{+7.9}_{-10.1}$	$^{+0.9}_{-1.9}$	$^{-1.0}_{+0.7}$	$^{+0.5}_{-0.5}$	0.81	6.5	1.00
1B	2.95 10^0	11.2	12.3	−11.4	$^{-0.2}_{-1.7}$	$^{+3.9}_{-2.0}$	$^{-3.1}_{-1.1}$	$^{-0.1}_{-2.0}$	$^{+0.4}_{-0.4}$	0.81	5.3	1.00
1C	1.30 10^0	13.9	13.9	+12.8	$^{+5.7}_{-3.5}$	$^{-0.3}_{-0.4}$	$^{+2.0}_{-0.4}$	$^{-0.7}_{+0.9}$	$^{+0.4}_{-0.4}$	0.81	3.7	1.00
2A	1.67 10^0	19.6	43.6	+42.8	$^{+1.0}_{+0.2}$	$^{+6.5}_{-9.1}$	$^{+0.4}_{+0.1}$	$^{-0.5}_{+1.4}$	$^{+0.5}_{-0.5}$	0.80	5.7	1.00
2B	2.52 10^0	10.9	20.6	−19.9	$^{+0.4}_{-1.1}$	$^{+5.2}_{-3.0}$	$^{+2.7}_{+0.2}$	$^{-0.4}_{+2.1}$	$^{+0.5}_{-0.5}$	0.81	4.9	1.00
2C	1.05 10^0	13.5	13.2	+11.7	$^{+7.1}_{-4.1}$	$^{-0.6}_{-0.5}$	$^{+0.6}_{-0.8}$	$^{-0.4}_{-0.1}$	$^{+0.4}_{-0.4}$	0.80	3.5	1.00
3A	1.77 10^0	15.6	10.9	+7.3	$^{-1.1}_{+0.2}$	$^{+8.1}_{-7.2}$	$^{-0.9}_{+0.2}$	$^{-0.7}_{+0.8}$	$^{+0.5}_{-0.5}$	0.80	5.1	1.00
3B	3.38 10^0	8.6	8.6	+6.7	$^{+1.6}_{-1.1}$	$^{+5.3}_{-4.2}$	$^{-0.4}_{+0.3}$	$^{+0.1}_{-0.3}$	$^{+0.4}_{-0.4}$	0.81	4.5	1.00
3C	1.45 10^0	11.0	38.1	+37.9	$^{+4.7}_{-1.6}$	$^{+0.3}_{+0.6}$	$^{+0.9}_{-1.0}$	$^{+0.2}_{-0.5}$	$^{+0.4}_{-0.4}$	0.80	3.0	1.00
4A	1.47 10^0	16.6	21.6	+20.9	$^{+2.2}_{-0.5}$	$^{+4.5}_{-5.2}$	$^{+0.4}_{-0.5}$	$^{-0.6}_{-0.4}$	$^{+0.5}_{-0.5}$	0.80	5.1	1.00
4B	3.13 10^0	9.7	27.8	−27.2	$^{+0.6}_{-0.3}$	$^{+5.6}_{-4.2}$	$^{+0.0}_{+0.3}$	$^{+0.3}_{+0.5}$	$^{+0.4}_{-0.4}$	0.81	4.5	1.00
4C	2.07 10^0	9.0	63.0	−62.9	$^{+2.2}_{-2.2}$	$^{+1.1}_{-0.1}$	$^{+1.4}_{-0.6}$	$^{-0.1}_{+0.1}$	$^{+0.3}_{-0.3}$	0.81	3.1	1.00
5B	2.95 10^0	8.9	34.4	−34.0	$^{+0.4}_{-0.1}$	$^{+5.9}_{-2.2}$	$^{-1.2}_{+0.8}$	$^{-0.5}_{+1.0}$	$^{+1.5}_{-1.5}$	0.80	2.9	1.03
5C	3.45 10^0	7.1	55.9	+55.8	$^{+1.2}_{-1.2}$	$^{+0.1}_{-1.4}$	$^{+0.9}_{-1.6}$	$^{-0.0}_{-0.8}$	$^{+1.4}_{-1.4}$	0.80	2.8	1.04
6C	3.80 10^{-1}	17.0	32.7	+32.3	$^{-0.2}_{+0.4}$	$^{+3.4}_{+0.8}$	$^{+0.1}_{-5.0}$	$^{-0.7}_{-2.3}$	$^{+2.3}_{-2.3}$	0.79	1.1	1.11

Table A.9: Double differential trijet cross sections measured as function of Q^2 and ξ_3. The bin labels of the first column are defined in table 7.2. Not shown is the LAr noise uncertainty of $\delta_{LArNoise} = 0.9\,\%$. The correlations between the data points are listed in table A.19. Further details are given in the caption of table A.5.

Normalized trijet cross sections in bins of Q^2 and $\langle p_T \rangle_3$

Bin label	CS	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	c_{had}	δ_{had} [%]
1α	$1.14\ 10^{-2}$	8.7	8.7	+1.8	$^{+2.5}_{-1.4}$	$^{+6.9}_{-5.9}$	$^{+2.8}_{-4.6}$	$^{-2.5}_{+4.9}$	0.79	5.3
1β	$6.20\ 10^{-3}$	8.4	4.3	+2.6	$^{-2.9}_{+3.0}$	$^{+1.2}_{-0.8}$	$^{-0.3}_{+0.7}$	$^{+0.9}_{-0.6}$	0.85	4.3
1γ	$9.00\ 10^{-4}$	19.9	6.8	+4.3	$^{-3.7}_{+6.0}$	$^{+0.0}_{-2.1}$	$^{+0.9}_{-1.7}$	$^{-0.5}_{+0.8}$	0.89	3.6
2α	$1.07\ 10^{-2}$	10.9	7.9	+3.9	$^{+2.1}_{-1.0}$	$^{+6.5}_{-6.6}$	$^{+0.9}_{+0.1}$	$^{-0.3}_{+1.8}$	0.78	5.0
2β	$6.40\ 10^{-3}$	9.4	6.0	+5.0	$^{-2.4}_{+2.6}$	$^{+2.0}_{-1.4}$	$^{-0.2}_{+0.9}$	$^{+0.3}_{-0.3}$	0.84	4.4
2γ	$1.40\ 10^{-3}$	16.7	5.3	+2.5	$^{-5.2}_{+3.9}$	$^{+0.0}_{-0.3}$	$^{+0.9}_{-0.1}$	$^{+0.6}_{+0.5}$	0.89	2.7
3α	$1.29\ 10^{-2}$	10.2	7.8	+3.5	$^{+2.2}_{-0.8}$	$^{+6.5}_{-6.5}$	$^{+1.2}_{-0.5}$	$^{-1.0}_{+1.9}$	0.78	4.6
3β	$9.80\ 10^{-3}$	8.0	5.5	+4.5	$^{-2.9}_{+1.9}$	$^{+1.7}_{-1.1}$	$^{-0.2}_{+2.2}$	$^{+1.1}_{-0.4}$	0.85	3.7
3γ	$1.80\ 10^{-3}$	17.3	6.8	+4.5	$^{-3.9}_{+5.7}$	$^{+0.1}_{-2.2}$	$^{+0.2}_{-1.1}$	$^{+0.2}_{+0.3}$	0.87	2.3
4α	$1.45\ 10^{-2}$	10.6	10.4	+7.9	$^{+2.3}_{-1.1}$	$^{+6.1}_{-6.4}$	$^{+1.2}_{-1.4}$	$^{-0.6}_{+1.9}$	0.77	4.1
4β	$1.30\ 10^{-2}$	7.3	5.9	+5.3	$^{-1.9}_{+0.9}$	$^{+1.1}_{-1.9}$	$^{-0.3}_{+1.0}$	$^{+0.4}_{-0.5}$	0.85	3.6
4γ	$3.00\ 10^{-3}$	14.4	8.1	+4.1	$^{-5.8}_{+7.1}$	$^{+2.9}_{-1.5}$	$^{-0.7}_{+1.1}$	$^{+0.1}_{-0.1}$	0.87	2.3
5α	$1.80\ 10^{-2}$	9.9	11.8	+10.7	$^{+2.7}_{-1.7}$	$^{+3.9}_{-4.2}$	$^{+0.3}_{-2.4}$	$^{-1.1}_{+1.5}$	0.77	3.5
5β	$1.64\ 10^{-2}$	7.3	4.6	+3.8	$^{-1.5}_{+1.4}$	$^{+1.8}_{-1.6}$	$^{+0.0}_{+1.5}$	$^{+0.3}_{-0.5}$	0.83	2.9
5γ	$4.20\ 10^{-3}$	12.8	4.8	+2.7	$^{-3.9}_{+3.6}$	$^{+0.5}_{-0.4}$	$^{+0.0}_{+0.6}$	$^{+0.4}_{+0.6}$	0.86	2.2
6β	$1.66\ 10^{-2}$	34.8	10.1	+8.8	$^{-0.2}_{-4.0}$	$^{+3.0}_{+2.3}$	$^{-0.6}_{+6.3}$	$^{+0.2}_{-0.7}$	0.82	0.8

Table A.10: Double differential normalized trijet cross sections measured as function of Q^2 and $\langle p_T \rangle_3$. Not shown are the normalization uncertainty of $\delta_{Norm} = 2.9\,\%$ and the LAr noise uncertainty of $\delta_{LArNoise} = 0.9\,\%$. Further details are given in the caption of table A.3.

Normalized trijet cross sections in bins of Q^2 and ξ_3

Bin label	CS	δ_{stat} [%]	δ_{sys} [%]	δ_{Model} [%]	δ_{JES} [%]	δ_{RCES} [%]	δ_{E_e} [%]	δ_{θ_e} [%]	c_{had}	δ_{had} [%]
1A	$7.30\ 10^{-3}$	11.1	25.6	+24.2	$^{+0.7}_{-0.4}$	$^{+7.4}_{-9.4}$	$^{+0.2}_{-1.8}$	$^{-0.5}_{+0.7}$	0.81	6.5
1B	$7.00\ 10^{-3}$	11.2	12.0	+11.5	$^{+1.4}_{+1.1}$	$^{+3.4}_{-1.3}$	$^{+0.2}_{-1.0}$	$^{+0.4}_{-2.8}$	0.81	5.3
1C	$3.10\ 10^{-3}$	13.9	13.0	+12.8	$^{-3.7}_{-0.2}$	$^{-0.9}_{+0.2}$	$^{-0.3}_{-0.2}$	$^{-0.3}_{+0.9}$	0.81	3.7
2A	$5.40\ 10^{-3}$	19.6	43.5	+42.9	$^{+0.0}_{-0.5}$	$^{+5.9}_{-8.5}$	$^{+0.3}_{+0.1}$	$^{-0.0}_{+0.8}$	0.80	5.7
2B	$8.10\ 10^{-3}$	10.9	20.4	+20.0	$^{+0.1}_{+0.4}$	$^{+4.6}_{-2.4}$	$^{+0.4}_{+0.3}$	$^{+0.1}_{+2.0}$	0.81	4.9
2C	$3.40\ 10^{-3}$	13.5	12.0	+11.8	$^{-2.7}_{+1.3}$	$^{-1.2}_{+0.1}$	$^{+0.7}_{-0.7}$	$^{+0.1}_{-0.3}$	0.80	3.5
3A	$6.50\ 10^{-3}$	15.6	10.4	+7.3	$^{+1.7}_{+0.2}$	$^{+7.7}_{-6.8}$	$^{+0.0}_{+0.2}$	$^{-0.3}_{+0.3}$	0.80	5.1
3B	$1.25\ 10^{-2}$	8.6	8.2	+6.8	$^{-0.5}_{-0.3}$	$^{+4.9}_{-3.7}$	$^{+0.1}_{+0.3}$	$^{+0.4}_{-0.5}$	0.81	4.5
3C	$5.40\ 10^{-3}$	11.0	38.0	+37.9	$^{-2.8}_{+1.4}$	$^{-0.1}_{+1.0}$	$^{-0.0}_{-1.0}$	$^{+0.5}_{-0.2}$	0.80	3.0
4A	$6.80\ 10^{-3}$	16.5	21.6	+21.1	$^{+0.7}_{+0.2}$	$^{+4.2}_{-4.9}$	$^{-0.6}_{-0.4}$	$^{-0.4}_{+0.2}$	0.80	5.1
4B	$1.45\ 10^{-2}$	9.7	27.7	+27.3	$^{-0.1}_{-0.1}$	$^{+5.3}_{-3.9}$	$^{+0.6}_{+0.4}$	$^{+0.6}_{+0.1}$	0.81	4.5
4C	$9.60\ 10^{-3}$	9.0	63.0	+63.0	$^{-0.6}_{+0.1}$	$^{+0.8}_{+0.2}$	$^{-0.7}_{-0.5}$	$^{+0.2}_{-0.0}$	0.81	3.1
5B	$1.65\ 10^{-2}$	8.9	34.4	+34.2	$^{-0.3}_{-0.1}$	$^{+5.8}_{-2.0}$	$^{+0.0}_{+0.9}$	$^{-0.1}_{-0.2}$	0.80	2.9
5C	$1.94\ 10^{-2}$	7.1	56.0	+55.9	$^{+0.4}_{+0.0}$	$^{-0.1}_{-1.3}$	$^{-0.2}_{-1.5}$	$^{+0.5}_{+0.7}$	0.80	2.8
6C	$4.49\ 10^{-2}$	16.7	33.2	+32.9	$^{+1.5}_{-0.0}$	$^{+3.4}_{+0.8}$	$^{+0.1}_{-4.4}$	$^{-0.4}_{+2.9}$	0.79	1.1

Table A.11: Double differential normalized trijet cross sections measured as function of Q^2 and ξ_3. The bin labels of the first column are defined in table 7.2. Not shown is the LAr noise uncertainty of $\delta_{LArNoise} = 0.9\,\%$. Further details are given in the caption of table A.7.

Inclusive jet as function of Q^2 and p_T^{jet}																							
	1α	1β	1γ	1δ	2α	2β	2γ	2δ	3α	3β	3γ	3δ	4α	4β	4γ	4δ	5α	5β	5γ	5δ	6α	6β	6γ	6δ
1α	100	-20	-10	-2	-13	2	1		1															
1β		100	1	-2	4	-13				2														
1γ			100	5	1		-13	-1			2								1					
1δ				100				-15				2								1				1
2α					100	-22	-10	-2	-11	2	1		-1				-1							
2β						100	2	-1	3	-10								-1						
2γ							100	7	1	1	-12	-1							-1					
2δ								100			-1	-11								-1				
3α									100	-22	-12	-2	-8	1	1		-1							
3β										100	-1	-2	2	-8				-1						
3γ											100	6	1	1	-8				-1					
3δ												100				-9								
4α													100	-23	-11	-2	-4							
4β														100	-1	-2	1	-4						
4γ															100	4		1	-4					
4δ																100				-5				
5α																	100	-23	-12	-2	-1			
5β																		100		-2		-1		
5γ																			100	3			-1	
5δ																				100				-2
6α																					100	-25	-13	-3
6β																						100	-1	
6γ																							100	-5
6δ																								100

Table A.12: Correlation coefficients between data points of the inclusive jet measurement as function of Q^2 and p_T^{jet}. All values are multiplied by a factor 100.

Dijet as function of Q^2 and $\langle p_T \rangle_2$																								
	1α	1β	1γ	1δ	2α	2β	2γ	2δ	3α	3β	3γ	3δ	4α	4β	4γ	4δ	5α	5β	5γ	5δ	6α	6β	6γ	6δ
1α	100	-43	10	3	-3	6	-1		11	-1			9				8				2			
1β		100	-35	-9	6	-12	4	1	-1	2	-1			1				1						
1γ			100	3	-1	3	-12				2								1				1	
1δ				100		1	1	-14				2				1				1		-1	1	
2α					100	-44	9	2	-5	6	-1		4	1			4	1			1			
2β						100	-31	-8	6	-11	3	1	1	-1			2	-1			1			
2γ							100	2	-1	3	-11	-1				1								
2δ								100		1		-11				-1				-2		1	-1	
3α									100	-47	11	3	-3	4	-1		4	1			1			
3β										100	-34	-11	5	-8	3	1	1							1
3γ											100	2	-1	2	-8				-1	1				
3δ												100		1		-9				-1			-1	
4α													100	-45	11	3		3			1			
4β														100	-37	-11	3	-4	2					
4γ															100	3		2	-5	1				-1
4δ																100			1	-5		1	-1	
5α																	100	-48	11	2	1			
5β																		100	-36	-8	1	-1		
5γ																			100	-2				-1
5δ																				100		1		-1
6α																					100	-47	10	2
6β																						100	-38	-7
6γ																							100	-20
6δ																								100

Table A.13: Correlation coefficients between data points of the dijet measurement as function of Q^2 and $\langle p_T \rangle_2$. All values are multiplied by a factor 100.

Trijet as function of Q^2 and $\langle p_T \rangle_3$	1α	1β	1γ	2α	2β	2γ	3α	3β	3γ	4α	4β	4γ	5α	5β	5γ	6β
1α	100	-37	9	1	2	-1	14	-4	1	12	-3	1	12	-3	1	-1
1β		100	-27	1	-8	3	-3	7	-2	-2	5	-1	-3	6	-2	2
1γ			100	-1	3	-13	1	-2	5	1	-1	3	1	-1	5	1
2α				100	-36	8	2	1	-1	11	-3	1	10	-2	1	
2β					100	-25	1	-6	2	-3	5	-1	-2	5	-1	2
2γ						100		2	-9	1	-1	3	1	-1	4	
3α							100	-36	10	2	1		10	-3	1	-1
3β								100	-28		-2	1	-2	6	-2	2
3γ									100		1	-4		-1	5	1
4α										100	-35	9	5			
4β											100	-27	-1	2		2
4γ												100			1	
5α													100	-36	9	
5β														100	-29	1
5γ															100	1
6β																100

Table A.14: Correlation coefficients between data points of the trijet measurement as function of Q^2 and $\langle p_T \rangle_3$. All values are multiplied by a factor 100.

Inclusive jet as function of Q^2 and p_T^{jet}	1α	1β	1γ	1δ	2α	2β	2γ	2δ	3α	3β	3γ	3δ	4α	4β	4γ	4δ	5α	5β	5γ	5δ	6α	6β	6γ	6δ
1α	35	1	-1		-5				1															
1β	-6	25	-1	-1	1	-3				1														
1γ	-1	-2	49				-6				1													
1δ		2	-9	67		1	-1	-9			-1	2			1		1	-1						1
2α	-5				35		-1		-4	1			-1			-1								
2β	1	-4			-7	28			1	-3								-1						
2γ			-7	1	-1	-2	51	-2		1	-7								-1					
2δ				-11	1	-1	-1	69			1	-7			-1					-1			-1	
3α	1				-5	1			35	1	-1		-3											
3β		1			1	-4	1		-6	25		-1		-3				-1						
3γ			1				-7		-1	-5	50	-1			-5				-1					
3δ				2				-7		-1	-3	66				-7				-1			-1	
4α					-1				-3	1			35		-1		-2							
4β										-2			-6	25	-1	-1		-2						
4γ											-4			-3	48	-3			-3					
4δ				1				-1			1	-6		-1	-2	69			1	-5		1	-1	1
5α	1				-1								-2				32	1	-1					
5β						-1					1			-2	1		-7	24		-1				
5γ							-1				-1				-3	1		-4	49	-2			-1	
5δ				1				-2			1				1	-4			-7	73		1	-1	-1
6α																					29	3		-1
6β																				1	-6	19	-2	-1
6γ																			-1		-3	-4	45	-9
6δ																				-1	-1	-1	-6	67

Table A.15: Correlation coefficients between data points of the inclusive jet measurement as function of Q^2 and p_T^{jet} and of the dijet measurement as function of Q^2 and $\langle p_T \rangle_2$. All values are multiplied by a factor 100.

Inclusive jet as function of Q^2 and p_T^{jet}	Trijet as function of Q^2 and $\langle p_T \rangle_3$														
	1α	1β	1γ	2α	2β	2γ	3α	3β	3γ	4α	4β	4γ	5α	5β	5γ	6β
1α	11	2	1	-2						-1			-1			
1β	10	12	1	-3	-2		-1			-1			-2			
1γ	-6	18	12	1	-4	-2		-1			-1			-2		-1
1δ	2	-5	21		1	-5			-1		1	-2		1	-2	-1
2α	-2			10	1	1	-2						-1			
2β	-3	-2		10	13		-2	-1					-1			
2γ	1	-3	-2	-6	17	17	1	-3	-2		-1			-1		-1
2δ		1	-5	2	-7	21		1	-2						-2	1
3α	-1			-2			12	1	2	-1						
3β	-2			-2	-2		8	13		-1	-1		-1			
3γ		-1		1	-3	-2	-6	15	12	1	-2	-1		-1		-1
3δ			-1		1	-2	2	-6	23		1	-3			-1	1
4α	-1						-1			8	3	2	-1			
4β	-1			-1			-1	-1		8	12	1	-1	-1		
4γ		-1			-1		1	-2	-1	-6	16	17		-2	-1	
4δ			-1			-1		1	-3	2	-7	22			-2	1
5α	-1			-1						-1			9	3		
5β	-1			-1						-1	-1		8	10	2	
5γ	1	-1			-1			-1		1	-2	-1	-5	13	14	-1
5δ			-2			-1			-1		1	-2	2	-6	16	2
6α																2
6β																10
6γ			-1												-1	11
6δ																-3

Table A.16: Correlation coefficients between data points of the inclusive jet measurement as function of Q^2 and p_T^{jet} and of the trijet measurement as function of Q^2 and $\langle p_T \rangle_3$. All values are multiplied by a factor 100.

Dijet as function of Q^2 and $\langle p_T \rangle_2$	Trijet as function of Q^2 and $\langle p_T \rangle_3$															
	1α	1β	1γ	2α	2β	2γ	3α	3β	3γ	4α	4β	4γ	5α	5β	5γ	6β
1α	13	-7	2	-2	1											
1β	6	16	-5	-2	-3	1	-1			-1			-1			
1γ	-4	12	21	1	-3	-3	1	-1		1	-1		1	-2		
1δ	2	-6	16		1	-3			-1			-2			-2	-1
2α	-2	1		13	-7	2	-1	1								
2β	-2	-2	1	5	17	-5	-1	-2	1							
2γ	1	-3	-4	-3	8	24	1	-2	-3		-1			-1		
2δ		1	-3	1	-4	13		1	-3			-1			-2	2
3α				-1			14	-7	3	-1						
3β	-1			-1	-2	1	3	17	-7	-1	-1					
3γ	1	-2		1	-2	-4	-4	11	24	1	-2	-2		-1		-1
3δ					1	-2	2	-7	20			-2			-1	1
4α							-1			13	-6	3				
4β	-1							-1	1	4	17	-7		-1		
4γ	1	-1			-1		1	-2	-2	-4	8	28	1	-1	-1	-1
4δ			-1			-1		1	-2	2	-6	17			-2	2
5α													13	-7	2	
5β	-1										-1		3	16	-5	
5γ	1	-2	-1	1	-1		1	-1		1	-1	-1	-4	10	19	-1
5δ		1	-2			-1			-1		1	-2	2	-5	14	2
6α																-13
6β			1												1	27
6γ			-1			-1			-1			-1			-2	-6
6δ			1													

Table A.17: Correlation coefficients between data points of the dijet measurement as function of Q^2 and $\langle p_T \rangle_2$ and of the trijet measurement as function of Q^2 and $\langle p_T \rangle_3$. All values are multiplied by a factor 100.

Dijet as function of Q^2 and ξ_2	1a	1b	1c	1d	2a	2b	2c	2d	3a	3b	3c	3d	4a	4b	4c	4d	5b	5c	5d	6d
1a	100	-37	6	-29	-10	3		2	5	-1			3				-1			
1b		100	-42	16	7	-11	3	-1		4	-3			3	-1	1	3	-2	1	
1c			100	-23	-1	5	-9	2		-1	5	-1		1	1	-2	1	3	-2	
1d				100	3	-2	4	-12			-1	2				1			1	
2a					100	-30	9	-18		-1		2					-1			
2b						100	-36	9	3	-7	3	-1	1	4	-1		4	-2	1	
2c							100	-24	1	4	-6	2			1	-1		3	-2	
2d								100	-3		2	-10	1			1				
3a									100	-32	15	-13	-8	2	-1		-2			
3b										100	-35	9	3	-3	1		6	-2	1	
3c											100	-26	-2	3	-3	1	1	5	-2	
3d												100		-1	2	-6		1		
4a													100	-26	14	-11	1	-1		
4b														100	-28	-1	2	1	-1	
4c															100	-21	3	10	-3	-1
4d																100			-1	
5b																	100	-15	20	
5c																		100	-25	-3
5d																			100	12
6d																				100

Table A.18: Correlation coefficients between data points of the dijet measurement as function of Q^2 and ξ_2. All values are multiplied by a factor 100.

Trijet as function of Q^2 and ξ_3	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C	5B	5C	6C
1A	100	-37	13	-4	3	-1	10	-3	1	6	-3	1	-3	1	1
1B		100	-47	9	-10	1	-1	6	-3	-1	5	-1	7	-3	-2
1C			100	-5	5	-5	1	-1	3		-1	3	-1	6	4
2A				100	-36	9		1	-4	4	-1		-2	2	1
2B					100	-35	1	-5	3	-1	6	-1	7	-3	-2
2C						100	-3	3	-3		-1	3	-2	7	3
3A							100	-38	7	-4	1	-1	-2	1	1
3B								100	-34	1		1	8	-3	-2
3C									100		1	-3	-2	7	5
4A										100	-41	6			
4B											100	-36	5	-2	-2
4C												100	-1	5	5
5B													100	-31	-3
5C														100	12
6C															100

Table A.19: Correlation coefficients between data points of the trijet measurement as function of Q^2 and ξ_2. All values are multiplied by a factor 100.

Inclusive jet as function of Q^2 and p_T^{jet}	Dijet as function of Q^2 and ξ_2																			
	1a	1b	1c	1d	2a	2b	2c	2d	3a	3b	3c	3d	4a	4b	4c	4d	5b	5c	5d	6d
1α	28	8	6		-4	-1	-1		1				1							
1β	7	20	8	2	-1	-2	-1		1	1										
1γ	-2	5	11	23		-1	-2	-3				1								
1δ		1	-3	29				-4				1								
2α	-3	-1	-1		32	10	7		-2	-2	-1									
2β	-1	-3	-1		5	20	12	1	-2	-2	-2									
2γ		-1	-2	-3	-1	4	12	27	1	-1	-2	-4								
2δ				-5	4	1	-3	31	1	-1		-3								
3α	1				-2	-1	-1		27	14	9		-1	-1	-1					
3β					-1	-2	-2		4	22	12	1		-2	-1					
3γ			1	1			-1	-3	1	2	15	29			-1	-3				
3δ				1	1		-1	-4	-10	2	-4	35				-2				
4α									-2	-1	-1		22	22	12	-1	-1	-1	-1	
4β										-2	-1		-1	19	14	5	-1	-1	-1	
4γ											-1	-2		1	16	28			-1	
4δ			-1						1			-2	2	2	-2	26				
5α										1			-1	-1	-1		28	19	15	
5β			1												-1	-1	11	20	21	
5γ								-1								-2		9	24	-1
5δ																-1	2	-1	18	
6α																		-1	3	35
6β																				29
6γ																				16
6δ																			-1	7

Table A.20: Correlation coefficients between data points of the inclusive jet measurement as function of Q^2 and p_T^{jet} and of the dijet measurement as function of Q^2 and ξ_2. All values are multiplied by a factor 100.

Inclusive jet as function of Q^2 and p_T^{jet}	Trijet as function of Q^2 and ξ_3													
	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C	5B	5C	6C
1α	18	5	2	-3	-1										
1β	8	14	7	-2	-3	-1	-1	-1			-1		-1		
1γ		7	19		-1	-3						-1	-1	-2	-1
1δ	1	-1	10			-2								-1	-2
2α	-3	-1		14	6	3	-2	-1							
2β	-2	-3	-1	6	15	5	-1	-2	-1		-1		-1		
2γ	1	-1	-3	-2	7	22		-1	-3					-1	-1
2δ			-2			13			-2						
3α				-2	-1		17	8	2	-1	-1				
3β	-1				-2	-1	5	16	6		-2	-1	-1		
3γ				1	-1	-4		5	20			-2		-1	-1
3δ						-2			12			-1			
4α							-1	-1		14	9	4	-1		
4β		-1						-2	-1	3	15	8	-1	-1	
4γ			-1						-2		5	25		-2	
4δ									-1		1	8		-1	
5α											-1		12	9	
5β		-1	-1			-1			-1		-1	-1	12	12	-2
5γ									-1			-1	3	20	-1
5δ												-1		7	
6α															24
6β			-1			-1			-1			-1		-2	17
6γ															23
6δ															13

Table A.21: Correlation coefficients between data points of the inclusive jet measurement as function of Q^2 and p_T^{jet} and the data points of the trijet measurement as function of Q^2 and ξ_3. All values are multiplied by a factor 100.

Dijet as function of Q^2 and ξ_2	Trijet as function of Q^2 and ξ_3														
	1A	1B	1C	2A	2B	2C	3A	3B	3C	4A	4B	4C	5B	5C	6C
1a	18	-3		-4	1		-1			-1					
1b		17	-2		-3	1									
1c	-1	4	8		-2	-2	1	-1	1	1	-1		-2		
1d	3	-4	15		1	-3		1			1	-1	1	-2	-2
2a	-6	1	-1	18	-6	2	-3								
2b		-3	1		17	-4		-2					-1	1	
2c	1	-1	-2		3	11		-1	-1		-1		-1		
2d			-2	2	-1	15		1	-2				1	-1	-1
3a	-2			-1		1	20	-6	2	-2					
3b		-1		-1	-2	2	1	19	-5		-2		-1		
3c	1	-1		1	-1	-2	-1	5	13	1	-1	-1	-1		
3d					1	-2	3	-3	15		1	-1	1	-1	-1
4a	-2						-2			18	-7	1		-1	
4b		-1		-1	-1		-1	-2	1	6	15	-3	-2		
4c	1	-1			-1		1	-2	-1	-3	8	12	-2		
4d		1	-1					1	-2	3	-3	17	1	-2	-1
5b	-1	-1						-1			-1		10	1	-1
5c	1	-1			-1			-1		1	-2	-1	14	6	
5d	-1	1	-1			-1		1	-1		1	-2	3	17	-1
6d			-1			-1			-1			-1		-2	27

Table A.22: Correlation coefficients between data points of the dijet measurement as function of Q^2 and ξ_2 and of the trijet measurement as function of Q^2 and ξ_3. All values are multiplied by a factor 100.

Appendix B

Tables of α_s-values

In this chapter the values of the strong coupling constant $\alpha_s(M_Z)$ determined in this work are summarized. In table B.1 the α_s-values from fits to single or multiple measurements are given with all uncertainties. In table B.2 the values of α_s from fits to theoretical predictions involving PDF sets from various PDF groups are summarized.

$\alpha_s(M_Z)$ from fits to multiple measurements

α_s from σ_{jet}:	$0.1172\,(22)_{\text{exp}} \pm (7)_{\text{PDF}} \pm\ (7)_{\text{PDFset}}\ \pm (10)_{\text{had}} \pm (47)_{\mu_r} \pm (6)_{\mu_f}$
α_s from σ_{dijet}:	$0.1129\,(24)_{\text{exp}} \pm (6)_{\text{PDF}} \pm\ (9)_{\text{PDFset}}\ \pm\ (7)_{\text{had}}\ \pm (37)_{\mu_r} \pm (7)_{\mu_f}$
α_s from σ_{trijet}:	$0.1170\,(18)_{\text{exp}} \pm (3)_{\text{PDF}} \pm\ (7)_{\text{PDFset}}\ \pm (11)_{\text{had}} \pm (33)_{\mu_r} \pm (4)_{\mu_f}$
α_s from $\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$:	$0.1180\,(10)_{\text{exp}} \pm (5)_{\text{PDF}} \pm (11)_{\text{PDFset}} \pm\ (8)_{\text{had}}\ \pm (41)_{\mu_r} \pm (6)_{\mu_f}$
α_s from $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$:	$0.1138\,(11)_{\text{exp}} \pm (5)_{\text{PDF}} \pm (13)_{\text{PDFset}} \pm\ (6)_{\text{had}}\ \pm (32)_{\mu_r} \pm (7)_{\mu_f}$
α_s from $\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$:	$0.1180\,(14)_{\text{exp}} \pm (3)_{\text{PDF}} \pm\ (9)_{\text{PDFset}}\ \pm (11)_{\text{had}} \pm (33)_{\mu_r} \pm (4)_{\mu_f}$
α_s from $[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$:	$0.1162\,(22)_{\text{exp}} \pm (7)_{\text{PDF}} \pm\ (7)_{\text{PDFset}}\ \pm\ (9)_{\text{had}}\ \pm (45)_{\mu_r} \pm (6)_{\mu_f}$
α_s from $[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$:	$0.1182\,(17)_{\text{exp}} \pm (3)_{\text{PDF}} \pm\ (6)_{\text{PDFset}}\ \pm\ (6)_{\text{had}}\ \pm (37)_{\mu_r} \pm (4)_{\mu_f}$
α_s from $[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$:	$0.1174\,(18)_{\text{exp}} \pm (3)_{\text{PDF}} \pm\ (7)_{\text{PDFset}}\ \pm (12)_{\text{had}} \pm (35)_{\mu_r} \pm (4)_{\mu_f}$
α_s from $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$:	$0.1160\,(9)_{\text{exp}}\ \pm (5)_{\text{PDF}} \pm (12)_{\text{PDFset}} \pm\ (7)_{\text{had}}\ \pm (35)_{\mu_r} \pm (6)_{\mu_f}$
α_s from $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$:	$0.1175\,(8)_{\text{exp}}\ \pm (4)_{\text{PDF}} \pm (10)_{\text{PDFset}} \pm\ (8)_{\text{had}}\ \pm (36)_{\mu_r} \pm (5)_{\mu_f}$
α_s from $[\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$:	$0.1149\,(9)_{\text{exp}}\ \pm (4)_{\text{PDF}} \pm (11)_{\text{PDFset}} \pm (11)_{\text{had}} \pm (31)_{\mu_r} \pm (6)_{\mu_f}$
α_s from $[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$:	$0.1185\,(17)_{\text{exp}} \pm (4)_{\text{PDF}} \pm\ (6)_{\text{PDFset}}\ \pm (13)_{\text{had}} \pm (44)_{\mu_r} \pm (4)_{\mu_f}$
α_s from $[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}},\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$:	$0.1165\,(8)_{\text{exp}}\ \pm (4)_{\text{PDF}} \pm (11)_{\text{PDFset}} \pm\ (8)_{\text{had}}\ \pm (34)_{\mu_r} \pm (5)_{\mu_f}$

Table B.1: Values of $\alpha_s(M_Z)$ from fits to single measurements and to multiple measurements simultaneously.

$\alpha_s(M_Z)$ using different PDF sets

Measurement	$\alpha_s^{\text{MSTW2008}}$	$\alpha_s^{\text{NNPDF2.3}}$	α_s^{CT10}	$\alpha_s^{\text{HERAPDF1.5}}$	α_s^{ABM11}
	All PDF sets used were determined with $\alpha_s(M_Z) = 0.1180$				
σ_{jet}	0.1172	0.1164	0.1179	0.1156	0.1135
σ_{dijet}	0.1129	0.1119	0.1137	0.1114	0.1098
σ_{trijet}	0.1170	0.1162	0.1176	0.1169	0.1173
$\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}$	0.1180	0.1173	0.1195	0.1188	0.1191
$\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}$	0.1138	0.1129	0.1155	0.1149	0.1161
$\frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}$	0.1180	0.1173	0.1190	0.1206	0.1206
$[\sigma_{\text{jet}},\sigma_{\text{dijet}}]$	0.1162	0.1154	0.1169	0.1146	0.1122
$[\sigma_{\text{jet}},\sigma_{\text{trijet}}]$	0.1182	0.1174	0.1187	0.1177	0.1173
$[\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	0.1174	0.1166	0.1180	0.1171	0.1171
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}]$	0.1160	0.1152	0.1177	0.1170	0.1179
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1175	0.1169	0.1189	0.1186	0.1196
$[\frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1149	0.1141	0.1164	0.1160	0.1155
$[\sigma_{\text{jet}},\sigma_{\text{dijet}},\sigma_{\text{trijet}}]$	0.1185	0.1178	0.1191	0.1180	0.1176
$[\frac{\sigma_{\text{jet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{dijet}}}{\sigma_{\text{NC}}}, \frac{\sigma_{\text{trijet}}}{\sigma_{\text{NC}}}]$	0.1165	0.1157	0.1179	0.1174	0.1185

Table B.2: Values for $\alpha_s(M_Z)$ from fits to theoretical predictions obtained with PDFs from various PDF groups. For consistency, all PDF sets used have been determined at a value of $\alpha_s(M_Z) = 0.118$. The values are displayed in figure 10.4.

Appendix C

Flexible scale concept for fastNLO

The calculation of higher-order calculations of QCD cross sections in hadron-induced collisions are time-consuming and can need up to several years of CPU time for achieving a high precision of the Monte Carlo integration. On the other hand, in the iterative α_s-fitting procedure, the jet cross sections have to be repeatedly calculated for varying α_s-values. Also when jet cross sections are included in PDF-fits, the calculations have to be repeated for varying PDFs, but the perturbative coefficients do not change. Therefore, an efficient technique for a repeated calculation of jet cross sections with varying values of $\alpha_s(M_Z)$ and PDFs becomes necessary. In this work, we use the fastNLO method [79, 98–100], which stores the perturbative coefficients, which are calculated using nlojet++ [93, 94], in a table, such that $\alpha_s(M_Z)$ and the PDFs can be chosen without a recalculation of the matrix elements. The jet cross sections can then be obtained in a timescale of the order of $\mathcal{O}(\text{ms})$, which makes them suitable for the usage in an iterative fit.
Furthermore, it is of interest, to change or study the renormalization and factorization scale dependence of the cross sections. Following equation 2.12, the renormalization can be changed easily by a constant factor, if the LO coefficients are separately accessible. However, so far it was not possible, to vary the factorization scale within the framework of fastNLO[1], to choose a different functional definition of the scales, or to have different choices for the renormalization scale and the factorization scale. Therefore, a method for a scale-independent storage of the perturbative coefficients, based on the fastNLO method was developed, which is valid in any order perturbation theory.

C.1 The fastNLO concept

The pQCD cross section for jet production in DIS is given by $\sigma = \sum_{n,a} \sigma_{n,a} \otimes f_a$ (c.f. equation 2.53), where the partonic cross sections $\sigma_{n,a}$ (equation 2.50) have to be determined in a time consuming Monte Carlo integration.
In order to factorize the explicit PDF dependence from the convolution, the PDFs are expressed on a grid in the convolution variable x. Around each grid node i (ι) a cubic eigenfunction $E_i(x)$ (or also called *interpolation kernel*) is defined which fulfills

$$E_i(x_\iota) = \begin{cases} 1 & (i = \iota) \\ 0 & (i \neq \iota) \end{cases} , \tag{C.1}$$

[1] In [226], it was shown that an 'a-posteriori' variation of the factorization scale is in principle possible without a recalculation of the perturbative coefficients, which however is only possible with little computing costs up to NLO, since the QCD splitting functions have to be integrated numerically.

and the sum of all eigenfunctions are normalized to one

$$\sum_i E_i(x) = 1 \quad \text{for all } x. \tag{C.2}$$

The PDFs[2] $f_{a,i}$ are then approximately expressed through its values at every node x_i by a linear combination of eigenfunctions

$$f_a(x) \simeq \sum_i f_a(x_i) E_i(x) \ . \tag{C.3}$$

Similarly, the scale dependence of the partonic cross section can be approximately expressed on a grid of discrete values μ_j of the renormalization and/or factorization scale[3] μ by[4]

$$\sigma_{n,a} \simeq \sum_j E_j(\mu) \sigma_{n,a} \tag{C.4}$$

Using the perturbative expansion $\sigma_{n,a} = \alpha_s^n(\mu) c_{n,a}$, the cross section for any perturbative order n can then be expressed by a sum

$$\sigma_n = \sum_i \sum_j \sum_{a=q,\bar{q},g} \alpha_s^n(\mu_j) f_a(x_i, \mu_j) \left(c_{n,a} E_j(\mu) \otimes E_i(x) \right) \ , \tag{C.5}$$

where the integration of the partonic cross sections becomes independent from the explicit content of the PDFs, and these were pulled out from the convolution integral. The time-consuming MC integrations of x and the phase space can now be performed only once with high statistical precision, and the discrete values $\mathcal{C}_{n,a,i,j}$ are stored in a table

$$\mathcal{C}_{n,a,i,j} := c_{n,a} E_j(\mu) \otimes E_i(x) \ . \tag{C.6}$$

The cross section can then be obtained from the simple sum

$$\sigma = \sum_{n,a,i,j} \alpha_s^n(\mu_j) f_a(x_i, \mu_j) \mathcal{C}_{n,a,i,j} \ , \tag{C.7}$$

where the values for the strong coupling and the PDFs can be chosen without a recalculation of $\mathcal{C}$. For a more efficient storage of $\mathcal{C}$ and a faster evaluation of σ the coefficients are stored for the PDF linear combinations of the specific process (for jet production in DIS these are $a \in \{g, \Sigma, \Delta\}$), instead of for every single PDF flavor.
In hadron-hadron collisions, where two PDFs are involved, also the second PDFs f_b of all flavors b, are approximated on an additional x-grid x_k by $f_b(x_2) \simeq \sum_k f_b(x_k) E_k(x_2)$ and in μ_j. The coefficients which need to be calculated once and stored are then given by

$$\mathcal{C}^{\text{hh}}_{n,a,b,i,j,k} := c_{n,a,b} E_j(\mu) \otimes E_i(x_1) \otimes E_k(x_2) \ . \tag{C.8}$$

The cross section in hadron-hadron collisions σ_{hh} is obtained from

$$\sigma_{hh} = \sum_{n,a,b,i,j,k} \alpha_s^n(\mu_j) f_a(x_i, \mu_j) f_b(x_k, \mu_j) \mathcal{C}^{\text{hh}}_{n,a,b,i,j,k} \ . \tag{C.9}$$

The $13 \cdot 13$ linear combinations of the two PDFs f_a and f_b are more efficiently expressed by calculating and storing the coefficients for the independent subprocesses only.

[2] For a more accurate expression of the interpolation, the PDFs are reweighted with a simple function $w(x) = x^{-3/2}(1 - 0.99x)^3$ where the inverse $w(x)^{-1}$ is absorbed in the interpolation kernels E_i. This reduces the curvature for all scales and improves the high-x gluon interpolation.

[3] The generalization for a different choice of the renormalization and factorization scale is given by replacing equation C.4 by two interpolations for μ_r and μ_f, like $\sigma_{n,a} \simeq \sum_{j_r} E_{j_r}(\mu_r) \sigma_{n,a}$ and $\sigma_{n,a} \simeq \sum_{j_r} E_{j_r}(\mu) \sigma_{n,a}$, respectively.

[4] The interpolation kernels E_j can be of course differently from the ones used for the x-interpolation (E_i), and only have to fulfill equations C.1 and C.2.

C.2 Scale flexibility for the fastNLO concept

For the calculation of the cross sections, the renormalization and factorization scale have to be specified for the calculation of $\mathcal{C}$, and cannot be changed without a time-consuming recalculation. The above outlined fastNLO concept can be extended by two modifications towards a concept, where also the scales can be changed without a recalculation of the matrix elements.
Firstly, the perturbative coefficients[5] c can be expressed at next-to-leading order perturbation theory in the form[6]

$$c = c_0 + \log(\mu_r^2)c_r + \log(\mu_f^2)c_f \ , \tag{C.10}$$

where only in LO the coefficients c_r and c_f are zero. At higher orders, further logarithms of μ_r and μ_f appear, where the generalization is straight forward and only the NLO case is outlined here. The coefficients c_0, c_r and c_f factorize when calculating $\mathcal{C}$ using equations C.6 or C.8 and hence they can be determined and stored separately as $\mathcal{C}^0$, $\mathcal{C}^r$ and $\mathcal{C}^f$. Using a separate interpolation grid for the renormalization and factorization scale according to equation C.4 (denoted by the indices j_r and j_f, respectively), the cross section in DIS can be expressed as

$$\sigma = \sum_{n,a,i,j_r,j_f} \alpha_s^n(\mu_{j_r}) f_a(x_i,\mu_{j_f}) \left(\mathcal{C}^0_{n,a,i,j_r,j_f} + \log(\mu^2_{j_r})\mathcal{C}^r_{n,a,i,j_r,j_f} + \log(\mu^2_{j_f})\mathcal{C}^f_{n,a,i,j_r,j_f}\right) \ . \tag{C.11}$$

The equation for the cross section in hadron-hadron collisions σ^{hh} is obvious.
Using this idea it becomes possible to change the renormalization scale and the factorization scale by any factor c_{μ_r} and c_{μ_f}, respectively, through the replacements

$$\mu_{j_r} \rightarrow c_{\mu_r}\mu_{j_r} \tag{C.12}$$

$$\mu_{j_f} \rightarrow c_{\mu_f}\mu_{j_f} \tag{C.13}$$

in equation C.11. The two scale factors c_{μ_r} and c_{μ_f} can be chosen freely and independent from each other. However, the nominal renormalization and factorization scale has to be chosen before the calculation of the coefficients $\mathcal{C}$.

For a full flexibility in the choice of the scales, we consider that the scales μ_r and μ_f are expressed as functions of observables O_κ

$$\mu = \mu(O_1,\ldots,O_m). \tag{C.14}$$

These observables are for instance Q^2 or p_{T}, which can constitute the functional form $\mu_r^2 = (Q^2 + p_{\mathrm{T}}^2)/2$.
Instead of an interpolation of the variable of the scale (μ_r and μ_f) itself in equation C.4, the interpolation is performed in up to m observables O_κ

$$\sigma_{n,a} \simeq \sum_{o_1,\ldots,o_m} E_{o_1}(O_1)\cdot\ldots\cdot E_{o_m}(O_m)\cdot\sigma_{n,a} \ , \tag{C.15}$$

[5] We drop the subscripts n and a (and b) for better readability.
[6] Independent from this development, also the BlackHat collaboration made usage of this scale independent expression [227].

where the discrete values of the grid of the Observable O_κ are denoted o_κ. For a precise approximation of equation C.15, it is necessary, that either the grid spacing is chosen according to the shape of the observable, or the observable is reweighted. For instance, a linear distance between the grid points o_κ is reasonable for an interpolation in the pseudo-rapidity, while a logarithmic (or double logarithmic) spacing is preferred for instance for Q^2 or p_{T}.
Using equation C.15, the scales μ_r and μ_f can be expressed approximately at the grid point $o_1, \ldots, o_m$ through

$$\mu^{r/f}_{o_1,\ldots,o_m} = \mu(O_{1,o_1}, \ldots, O_{m,o_m}) \ . \tag{C.16}$$

The NLO cross section in DIS can then be calculated by

$$\sigma = \sum_{n,a,i,o_1,\ldots,o_m} \alpha_s^n(\mu_r) f_a(x_i,\mu_f) \left(\mathcal{C}^0_{n,a,i,o_1,\ldots,o_m} + \log(\mu_r^2)\mathcal{C}^r_{n,a,i,o_1,\ldots,o_m} + \log(\mu_f^2)\mathcal{C}^f_{n,a,i,o_1,\ldots,o_m} \right) \ . \tag{C.17}$$

where the scales $\mu_r = \mu^r_{o_1,\ldots,o_m}$ and $\mu_f = \mu^f_{o_1,\ldots,o_m}$ are determined at the points $o_1, \ldots, o_m$ by equation C.16. Since the choice of the functional form is free in equation C.16, the scales μ_r and μ_f can be defined and calculated from any function which takes the previously stored observables $O_1, \ldots, O_m$ as parameters. Only the discrete perturbative coefficients $\mathcal{C}_{n,a,i,o_1,\ldots,o_m}$ need to be calculated and stored once with high statistical precision but can then be read from a table for repeated calculations.

Using the notation from above the cross section in hadron-hadron collisions can be calculated in NLO by

$$\sigma_{hh} = \sum_{n,a,b,i,k,o_1,\ldots,o_m} \alpha_s^n f_a f_b \left(\mathcal{C}^0_{n,a,b,i,k,o_1,\ldots,o_m} + \log(\mu_r^2)\mathcal{C}^r_{n,a,b,i,k,o_1,\ldots,o_m} + \log(\mu_f^2)\mathcal{C}^f_{n,a,b,i,k,o_1,\ldots,o_m} \right) . \tag{C.18}$$

For an efficient technical implementation, the coefficients $\mathcal{C}_{a,b}$ are stored for the l individual subprocesses of the reaction[7] which reduces the two sums $\sum_a \sum_b$ to only one sum $\sum_l$, as well as the number of indices are reduced for all coefficients. The $13 \cdot 13$ PDF combinations $f_a f_b$ constitute the l linear combinations according to the single subprocesses. In the current implementation of fastNLO, the number of stored observables m is chosen to be $m = 2$. In the framework of fastNLO, the interpolation grid is optimized for every bin of the measurement, and spans exactly the covered phase phase in x and $o_1, \ldots, o_m$. The necessary minimum and maximum values are obtained in a so-called 'warm-up run'. The LO coefficients are independent from μ_f and μ_r and therefore the coefficients $\mathcal{C}^r$ and $\mathcal{C}^f$ are zero and are not stored.

The code for the evaluation of standardized fastNLO tables is publicly available through `http://fastnlo.hepforge.org`.

C.3 Study and application of the flexible scale concept

The concept for flexible scales in fastNLO is studied with the calculations for the inclusive jet measurement, where the perturbative coefficients are obtained from nlojet++. The

[7] In jet production in hadron-hadron collisions, these are $l = 6$ in LO, and $l = 7$ in NLO. In DIS there are $n = 2$ indpendent subprocesses in LO and $n = 3$ in NLO.

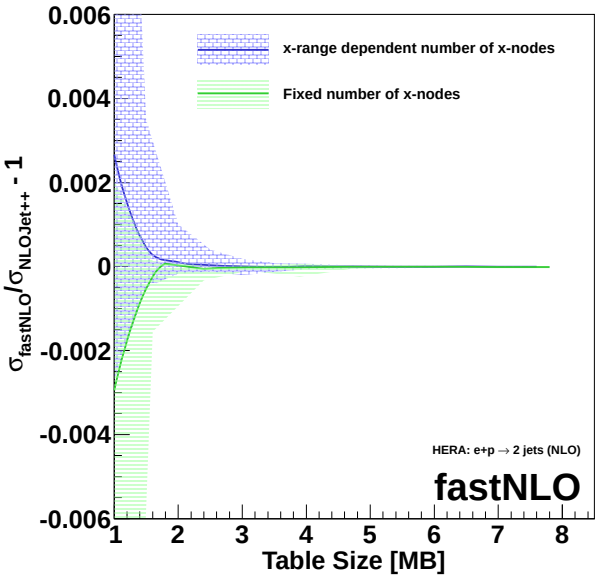


Figure C.1: Comparison of cross sections calculated with the approximated formula from fastNLO to direct calculations from nlojet++ for inclusive jet cross sections in DIS at high Q^2. Shown are the average deviation of all bins from the direct calculations as a function of the table size. The size of the table is increased due to a larger number of x-nodes and an increased number of nodes for Q^2 and p_{T}, which increases the precision of the interpolations. The shaded areas show the deviation of the bin with the maximum deviation. Already at very small table sizes, the precision of the interpolation is better than $\mathcal{O}(10^{-4})$ and can reach a precision of up to $\mathcal{O}(10^{-6})$ for a larger number of nodes.

cross sections calculated with the approximated fastNLO formulae are compared to the calculations from nlojet++ in figure C.1. One observes, that already with very few nodes, which corresponds to small table sizes, the cross sections approximated with fastNLO are identical to the ones from nlojet++ within an agreement better than $\mathcal{O}(10^{-4})$ and can reach an agreement up to $\mathcal{O}(10^{-6})$. This proves, that the interpolated observables $o_1, \ldots, o_m$ can indeed being used like continuous variables.

The concept is further studied for calculations of ATLAS dijet cross sections measured as function of the invariant mass M_{12} in proton-proton collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ [228]. In figure C.2, the cross sections obtained with fastNLO are compared to the calculations from nlojet++. The scale μ_r and μ_f are chosen to be $\mu_r = \mu_f = \langle p_{\mathrm{T}} \rangle e^{(0.3y^*)}$, where the observable $\langle p_{\mathrm{T}} \rangle$ and y^* are stored in the table ($\langle p_{\mathrm{T}} \rangle$ is the average transverse momentum of the two leading jets and y^* is their rapidity separation). The cross section predictions agree with each other within 0.2 %, despite the complicated functional form of the scale. Only at very high y^* the deviation is up to 0.4 %, which, however, can be improved with an increased number of nodes for y^* in these bins. Also, the precision of the fastNLO interpolation can be improved with an increased number of nodes for the x-interpolation.

C.4 Exemplary application for LHC jet production

The flexible-scale concept can be used for studying the choice of the ATALS collaboration for their measurement of dijet cross sections as function of M_{12} [228]. The choice of the renormalization and factorization scale in the article was motivated by a scale of $\mu_f = \mu_r =$

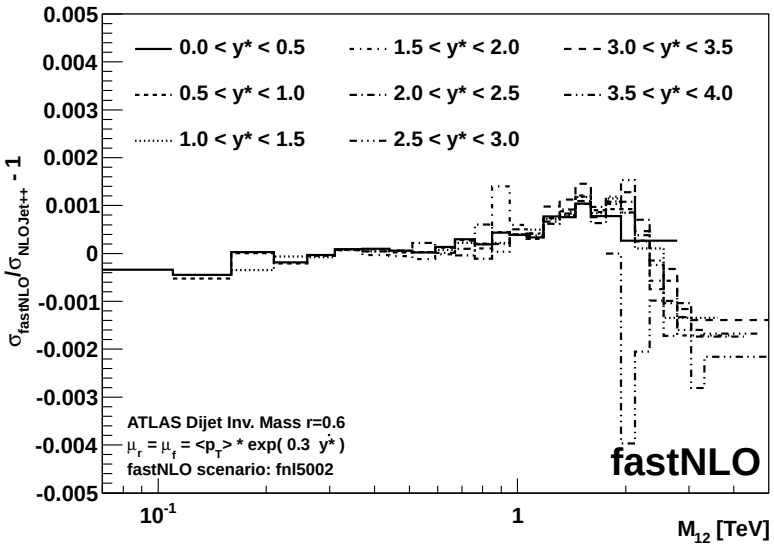


Figure C.2: Ratio of cross sections calculated with fastNLO to direct calculations with nlojet++ [229] for ATLAS data. The scales μ_r and μ_f are set to $\mu_r = \mu_f = \langle p_\mathrm{T}\rangle e^{(0.3y^*)}$. This is possible, since the fastNLO perturbative coefficients are interpolated in $\langle p_\mathrm{T}\rangle$ and $y*$.

$M_{12}/(2\cosh(0.7y^*))$ [46], which then was approximated by $\mu_f = \mu_r = p_\mathrm{T}e^{0.3y^*}$ [228][8]. The factors of 0.3 and 0.7 seem to be quite arbitrary and the choice was motivated, that the cross sections become 'more stable' if also the rapidity separation is used.
The flexible-scale concept enables to study the dependence of the cross section prediction as function of the parameter a in a scale formula of the form $\mu_f = \mu_r = p_\mathrm{T}e^{ay^*}$. The cross sections ratio to a scale choice of $\mu = p_\mathrm{T}e^{0.3y^*}$ for two bins of M_{12} are shown as function of the rapidity separation y^* in figure C.3 together with data for several choices of a. Also the theoretical uncertainty from scale variations of a factor of 2 for the scale choice $\mu_f = \mu_r = p_\mathrm{T}e^{0.3y^*}$ is shown, and furthermore, the cross section predictions for a scale choice of $\mu_f = \mu_r = p_{\mathrm{T,max}}e^{0.3y^*}$ is displayed. Relevant changes in the predictions can of course only be observed for large values of y^*, where infact the cross section changes by about 20 % compared to the nominal choice of $\mu = p_\mathrm{T}e^{0.3y^*}$. This is larger than the theory uncertainty as estimated from scale variations, since at these values of $y^* \approx 3.5$ this translates into a scale factor of approximately $e^{0.3\cdot 3.5} \approx 2.9$. On the other hand we observe, that the scale variations always yield smaller cross sections for the up as well as the down variation, since the PMS scale is in between these ranges.
However, the observation that the cross sections become unstable or even negative for large values of y^* could not be confirmed for any of the the scale choices studied. Furthermore, the promised 'plateau'-region [228] in the ATLAS choice of $a = 0.3$ is observed at smaller values of approximately $a \approx 0.15$. The flexible-scale concept now enables to optimize the parameter a such that the desired PMS scale is indeed realized.

[8] In the article [228], it does not become clear if the momentum of the hardest jet, or the average jet momentum of the two leading jets $\langle p_\mathrm{T}\rangle$ is used for the calculation. However, these two quantities are equivalent in leading-order approximation. Here, we identify p_T with $\langle p_\mathrm{T}\rangle$.

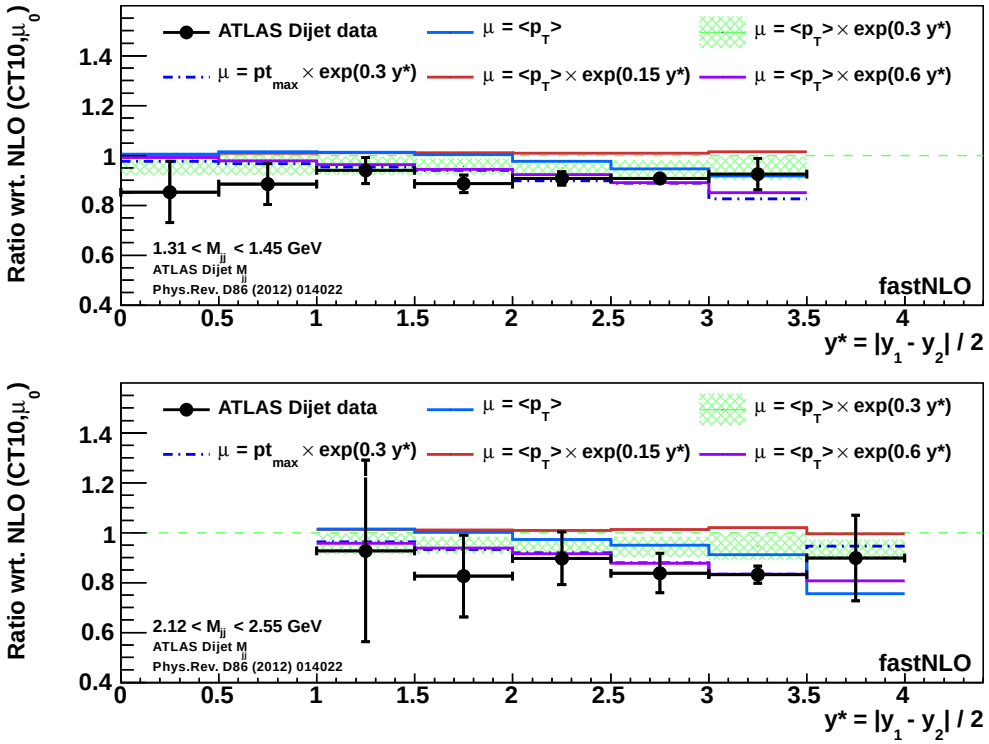


Figure C.3: Cross section ratio of various NLO predictions and ATLAS dijet data to NLO calculations with a scale choice of $\mu = p_{\mathrm{T}} e^{0.3y^*}$. The dijet cross section is shown for the ranges $1.31 < M_{12} < 1.45\,\mathrm{GeV}$ (upper plot) and $2.12 < M_{12} < 2.55\,\mathrm{GeV}$ (lower plot) as function of the rapidity separation of the two leading jets y^*. The theory uncertainty from scale variation by a factor of 2 on the reference scale choice is shown with the green hatched area. The full lines give predictions obtained from factors of a in the scale formula $\mu = p_{\mathrm{T}} e^{ay^*}$ when calculating the cross sections, and are all calculated from the same fastNLO table. The data points are for illustration and show the statistical uncertainties.

Appendix D

A method for the fast evaluation of jet cross sections in diffractive DIS

The fastNLO concept for the fast repeated calculation of jet cross sections can also be adapted for jet production cross sections in diffractive DIS. There, it is possible to improve the previously applied 'slicing method' such that the $x_{I\!P}$-integration can be performed on the same matrix elements. This enables to calculate these coefficients with very high statistical precision.

D.1 Jet production in diffractive DIS

One of the LO Feynman diagrams for jet production in diffractive DIS is shown in figure D.1. The cross section for jet production in diffractive DIS for one bin of a differential measurement is expressed in perturbative QCD through

$$\sigma = \sum_{a,n} \int \mathrm{d}x_{I\!P} \int \mathrm{d}z_{I\!P} \alpha_s^n c_{a,n} f_{I\!P/a}(x_{I\!P}, z_{I\!P}, \mu_f) \ , \tag{D.1}$$

where it was integrated over t and the notation from the previous chapter is used. The integrals have to be performed over $x_{I\!P}$ and $z_{I\!P}$, which express the momentum fraction of the pomeron w.r.t. the incoming proton momentum and the parton momentum fraction w.r.t. the pomeron momentum, respectively.

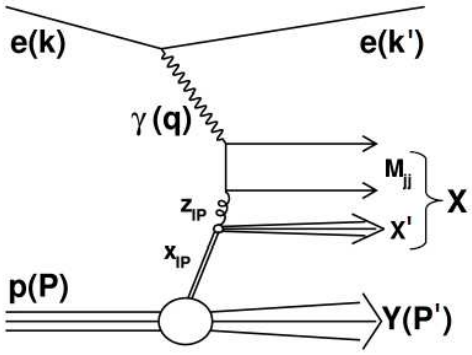


Figure D.1: The leading order boson-gluon-fusion diagram for dijet production in diffractive DIS [230].

In standard methods for calculating the cross section D.1, the $x_{I\!P}$-integration is approximated numerically by a Riemannian sum. Hence, the cross section can be calculated by the so-called *slicing method* using the approximation [231, 232]

$$\int \mathrm{d}x_{I\!P} f_{I\!P/a}(x_{I\!P})\sigma_{I\!P}(x_{I\!P}) \simeq \sum_k \Delta x_{I\!P,k} f_{I\!P/a}(x_{I\!P,k})\sigma_{I\!P}(x_{I\!P,k}) \ , \tag{D.2}$$

where the pomeron-lepton cross section $\sigma_{I\!P}$ is calculated for reduced center of mass energies for every $x_{I\!P}$-slice of width $\Delta x_{I\!P,k}$ at $x_{I\!P,k}$.

D.2 A concept for the fast and efficient evaluation of cross sections in diffractive DIS

According to the factorization theorem, the calculation of the hard contribution $\sigma_{I\!P}(x_{I\!P,k})$ is independent from the soft contribution(s) $f_{I\!P/a}$ or f_a, and hence all values of $\sigma_{I\!P}(x_{I\!P,k})$ calculate essentially repeatedly the identical matrix elements. For an efficient calculation of the cross section in diffractive DIS, we make full usage of the factorization theorem and the partonic cross section D.1 ($\sigma_{n,a}$) in order n can be more suitably being approximated by[1]

$$\sigma_{n,a} = \sum_k \Delta x_{I\!P,k} \int_0^{x_{I\!P,k}} \frac{\mathrm{d}x}{x_{I\!P,k}} \alpha_s^n c_0(x) f_a(x_{I\!P,k}, z_{I\!P} = \frac{x}{x_{I\!P,k}}, \mu_f) \ , \tag{D.3}$$

where c_0 is the perturbative coefficient calculated at the proton-lepton center-of-mass energy and the PDF has to be evaluated at $x_{I\!P,k}$ and $z_{I\!P}$ which is calculated using the integration variable by $z_{I\!P} = \frac{x}{x_{I\!P,k}}$. The correctness of equation D.3 is studied in figure D.2.
If $x_{I\!P,k}$ is pulled out from the integral, the integration of equation D.2 can be expressed according to the fastNLO approximation by equation C.17, where only the upper integration range of $x_{I\!P,k}$ has to be considered.
The restriction of the upper integration limit, however, is not without any problems. The perturbative coefficients have been interpolated in x, and thus are discrete in the integration variable x and are only accessible at the particular x_i-values of the grid. For an accurate approximation of the cross section, the contribution of the last x-node, where the integration is constrained, has therefore to be weighted according to the spanned x-range. This is done by a linear weight, which is calculated from the last x-node within the integration range[2].
This approach enables to calculate jet production cross sections in diffractive DIS very efficiently for different pomeron PDFs and different values of $\alpha_s(M_Z)$. Only the table of perturbative coefficients have to be evaluated once with high statistical precision. Also the full flexibility of the choice of the scales is given. Cross sections obtained from the here presented method are compared to measured cross sections from [230] and are shown in figure D.3.

[1] A method for the reweighting to a different center-of-mass energy in hadron-hadron collisions of such calculations was proposed in [226], where however the biases from the discrete values of x at the upper (and lower) integration bound is not pointed out.

[2] For technical limitations, the first x-node outside of the integration range cannot be accessed, for an even more precise weighting.

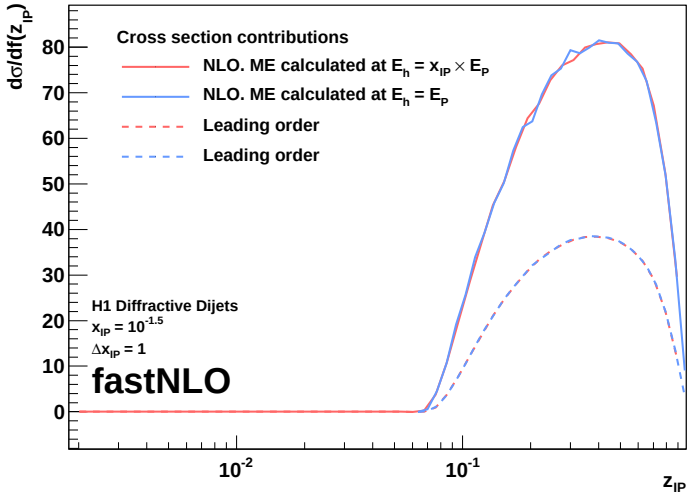


Figure D.2: Proof of concept for the calculation of jet cross sections in diffractive DIS. The matrix elements are calculated using nlojet++ once with an incident hadron energy of $E_h = x_{I\!P} E_p$, where the hadron is the pomeron, as it is done in the slicing-method, and once with an hadron energy equal to the proton energy $E_h = E_p$, where the incident hadron is the proton. In the latter case, the contribution of the $x_{I\!P}$-slice at $x_{I\!P} = 10^{-1.5}$ is 'down-scaled' towards the reduced center-of-mass energy. The contribution to the cross section is shown as function of $z_{I\!P}$, while $z_{I\!P}$ being the momentum fraction of the incoming parton to the hard interaction w.r.t. pomeron momentum for the slicing method, and $z_{I\!P} = x_{\mathrm{hadron}}/x_{I\!P} < 1$, where x_{hadron} is the parton's momentum fraction w.r.t. the incoming proton and $z_{I\!P}$ is constraint to values smaller than 1. The LO contributions agree perfectly. The NLO contributions show some fluctuations because of the low statistical precision of this example-calculation, and since the phase space sampling of the MC integration is less optimal for the calculation with $E_h = E_p$ for these low scale processes.

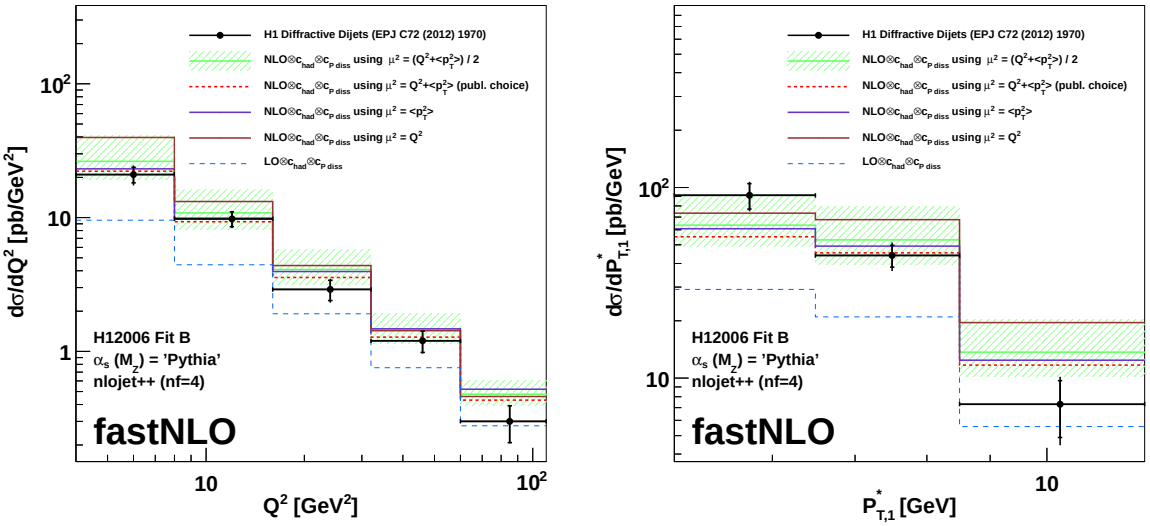


Figure D.3: Comparison of inclusive dijet cross section in diffractive DIS measured by H1 [230] as function of Q^2 (left) and the leading-jet transverse momentum $P^*_{T,1}$ (right) compared to NLO calculations obtained with the fastNLO method for diffractive DIS. The calculations are displayed for various choices for the renormalization and factorization scale. Also the LO predictions are shown. The hadronization correction and proton dissociation correction factor are taken from the publication. All calculations for each observables are obtained from only one calculation of matrix elements. The green shaded area shows the uncertainty on the theory predictions, which are obtained from the variation of the renormalization and factorization scale, which is now possible without the recalculation of the matrix elements.

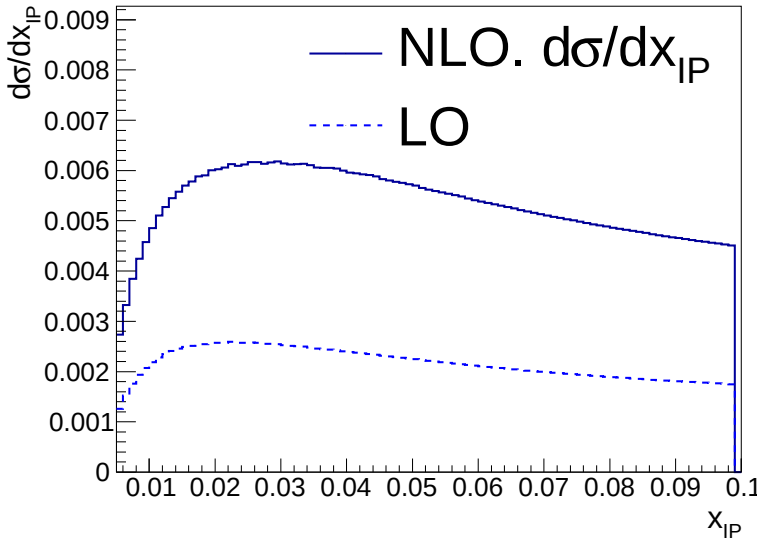


Figure D.4: The cross section for inclusive dijet production as function of $x_{I\!P}$. We observe, that the cross section is mostly flat in $x_{I\!P}$ and drops only for very small values of $x_{I\!P}$.

D.3 Study of the $x_{I\!P}$-integration

The new approach for the $x_{I\!P}$-integration now enables to study the accuracy of the previously used slicing-method in detail. The slicing can be performed with any slice widths $\Delta x_{I\!P,k}$, where one typically chooses equidistant widths in either a logarithmic or a linear measure. The cross section prediction as function of $x_{I\!P}$ is shown in figure D.4. We observe, that the cross section is mostly flat in $x_{I\!P}$. This can be interpreted such that a linear slicing in $x_{I\!P}$ is preferred over a logarithmic slicing.
The accuracy of the calculation as function of the number of slices in $x_{I\!P}$ is studied for the cross section as function of $P^*_{T,1}$ (as shown in the right figure of figure D.3) in figure D.5. We observe, that already with very few slices of $n > 10$ the cross section becomes independent from the applied slicing method. The remaining fluctuations are due to the biases introduced at the upper integration bound at $x_{I\!P,k}$. These biases, however, can be reduced by more nodes of the fastNLO grid in x. It is surprising, that a logarithmic slicing in $x_{I\!P}$ converges with less number of slices than the linear slicing[3]. However, for the logarithmic slicing method, a lower bound of $x_{I\!P}^{\min}$ must be known, since otherwise the calculations are biased, or the number of relevant slices may be small.

D.4 Cross section calculation as function of $x_{I\!P}$

The cross section for every bin i of a differential measurement is calculated using equation D.3. Moreover, the cross section as function of $x_{I\!P}$ can also be obtained from equation D.3, if the $x_{I\!P}$-integration is restricted according to the bin boundaries and by summing all bins i. This, however, is only possible for coefficient tables, which cover the full phase space of the $x_{I\!P}$ measurement. A calculation of cross sections as function of $x_{I\!P}$ is shown in figure D.6 and compared to data [230]. This calculation is obtained from the fastNLO coefficients calculated for the cross sections as function of Q^2: $\mathrm{d}\sigma/dQ^2$ (c.f. left figure in D.3). For a precise determination of these cross sections, also the lower boundary of the $x_{I\!P}$

[3]The apparent deviation of bin 1 with the logarithmic slicing is due to an accidental downward fluctuation of the reference cross section with $n = 48$ slices.

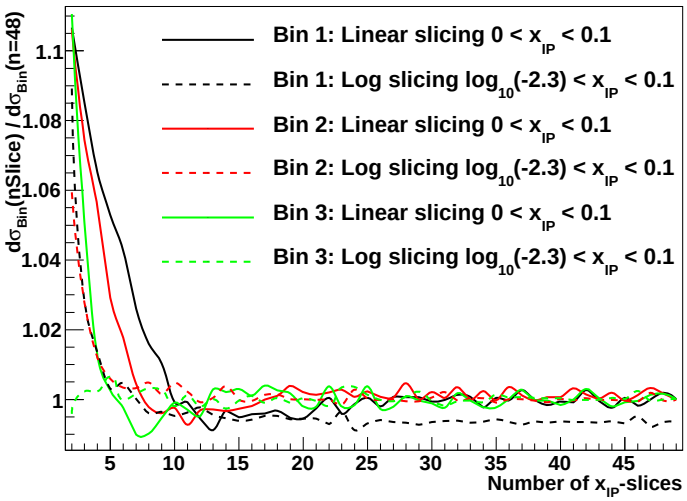


Figure D.5: Study of the accuracy of the Riemann-integration of $x_{I\!P}$ for the three bins of the measurement as function of $P^*_{T,1}$. Compared are a linear and a logarithmic slicing and are illustrated as function of the number of slices. The ratio of the cross section prediction is calculated w.r.t. the prediction from a slicing with a large number of slices (n=48).

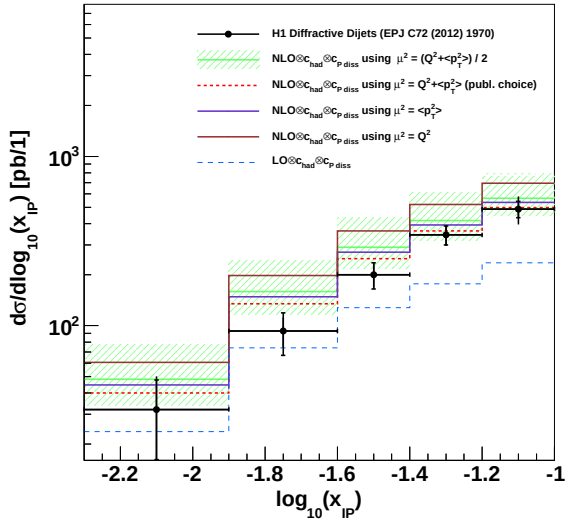


Figure D.6: Calculations of cross sections for inclusive dijet production in diffractive DIS as function of $x_{I\!P}$ using fastNLO. The cross sections are obtained from a calculation of cross sections as function of Q^2, where all bins are summed, but the $x_{I\!P}$ integration is restricted according to the bin intervals. More details are given in the caption of figure D.3.

integration has to be weighted according to the x-width of the underlying fastNLO grid. The cross sections as function of $z_{I\!P}$, however, cannot be obtained with such a method, since the integration variable $z_{I\!P}$ is differently from the measured observable '$z_{I\!P}$', and both agree only in LO approximation.

The generalization for the calculation in diffractive processes in hadron-hadron collisions is obvious. The code for the evaluation of the fastNLO tables is publicly available from `http://fastnlo.hepforge.org`.

References

[1] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys. Rev. Lett. 10* (1963) 531–533.

[2] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog. Theor. Phys. 49* (1973) 652–657.

[3] B. Pontecorvo, "Mesonium and anti-mesonium," *Sov. Phys. JETP 6* (1957) 429.

[4] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge," *Sov. Phys. JETP 7* (1958) 172–173.

[5] B. Pontecorvo, "Neutrino Experiments and the Problem of Conservation of Leptonic Charge," *Sov. Phys. JETP 26* (1968) 984–988.

[6] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles," *Prog. Theor. Phys. 28* (1962) 870–880.

[7] G. Aad *et al.* [ATLAS Collaboration], "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B 716* (2012) 1–29, arXiv:1207.7214.

[8] S. Chatrchyan *et al.* [CMS Collaboration], "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B 716* (2012) 30–61, arXiv:1207.7235.

[9] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking," Report number: CERN-TH-401 (1964).

[10] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys. Lett. B 47* (1973) 365–368.

[11] C.-N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," *Phys. Rev. 96* (1954) 191–195.

[12] J. Beringer *et al.* [Particle Data Group], "Review of Particle Physics (RPP)," *Phys. Rev. D 86* (2012) 010001.

[13] L. Faddeev and V. Popov, "Feynman Diagrams for the Yang-Mills Field," *Phys. Lett. B 25* (1967) 29–30.

[14] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and Collider Physics," Cambridge University Press, Cambridge (1996).

[15] D. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories," *Phys. Rev. Lett. 30* (1973) 1343–1346.

[16] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?," *Phys. Rev. Lett. 30* (1973) 1346–1349.

[17] R. Brock *et al.* [CTEQ Collaboration], "Handbook of perturbative QCD: Version 1.0," *Rev. Mod. Phys. 67* (1995) 157–248.

[18] B. Wiik, "First Results from PETRA," *Conf. Proc. C7906181* (1979) 113–154.

[19] P. Söding, "Jet Analysis," *Conf. Proc. C790627* (1979) 271–281.

[20] P. Söding, "On the discovery of the gluon," *Eur. Phys. J. H 35* (2010) 3–28.

[21] G. 't Hooft and M. Veltman, "Regularization and Renormalization of Gauge Fields," *Nucl. Phys. B 44* (1972) 189–213.

[22] J. C. Ward, "An Identity in Quantum Electrodynamics," *Phys. Rev. 78* (1950) 182.

[23] W. A. Bardeen, A. Buras, D. Duke, and T. Muta, "Deep Inelastic Scattering Beyond the Leading Order in Asymptotically Free Gauge Theories," *Phys. Rev. D 18* (1978) 3998.

[24] G. Sterman and S. Weinberg, "Jets from Quantum Chromodynamics," *Phys. Rev. Lett. 39* (1977) 1436–1439.

[25] F. Bloch and A. Nordsieck, "Note on the Radiation Field of the electron," *Phys. Rev. 52* (1937) 54–59.

[26] T. Kinoshita, "Mass singularities of Feynman amplitudes," *J. Math. Phys. 3* (1962) 650–677.

[27] T. D. Lee and M. Nauenberg, "Degenerate Systems and Mass Singularities," *Phys. Rev. 133* (1964) 1549.

[28] G. 't Hooft and M. Veltman, "Regularization and Renormalization of Gauge Fields," *Nucl. Phys. B 44* (1972) 189–213.

[29] S. Kluth, "Tests of Quantum Chromo Dynamics at e^+e^- Colliders," *Rept. Prog. Phys. 69* (2006) 1771–1846, arXiv:hep-ex/0603011.

[30] W. E. Caswell, "Asymptotic Behavior of Nonabelian Gauge Theories to Two Loop Order," *Phys. Rev. Lett. 33* (1974) 244–246.

[31] O. Tarasov, A. Vladimirov, and A. Zharkov, "The Gell-Mann-Low function of QCD in the three-loop approximation," *Phys. Lett. B 93* no. 4, (1980) 429 – 432.

[32] S. Larin and J. Vermaseren, "The Three loop QCD Beta function and anomalous dimensions," *Phys. Lett. B 303* (1993) 334–336, arXiv:hep-ph/9302208.

[33] T. van Ritbergen, J. Vermaseren, and S. Larin, "The Four loop beta function in quantum chromodynamics," *Phys. Lett. B 400* (1997) 379–384, arXiv:hep-ph/9701390.

[34] [ALEPH Collaboration, DELPHI Collaboration , L3 Collaboration , OPAL Collaboration , SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group], "Precision electroweak measurements on the Z resonance," *Phys. Rept. 427* (2006) 257–454, arXiv:hep-ex/0509008.

[35] K. Chetyrkin, J. H. Kuhn, and M. Steinhauser, "RunDec: A Mathematica package for running and decoupling of the strong coupling and quark masses," *Comput. Phys. Commun. 133* (2000) 43–65, arXiv:hep-ph/0004189.

[36] C. Glasman, "Precision measurements of alpha(S) at HERA," *AIP Conf. Proc. 792* (2005) 689–692, arXiv:hep-ex/0506035.

[37] D. E. Soper, "Basics of QCD perturbation theory," Lecture notes (Aug. 1996). Lecture at SLAC Summer Institute. arXiv:hep-ph/9702203.

[38] P. M. Stevenson, "Resolution of the Renormalization Scheme Ambiguity in Perturbative QCD," *Phys. Lett. B 100* (1981) 61.

[39] G. Grunberg, "Renormalization Group Improved Perturbative QCD," *Phys. Lett. B 95* (1980) 70.

[40] S. J. Brodsky, G. P. Lepage, and P. B. Mackenzie, "On the Elimination of Scale Ambiguities in Perturbative Quantum Chromodynamics," *Phys. Rev. D 28* (1983) 228.

[41] S. J. Brodsky and X.-G. Wu, "Application of the Principle of Maximum Conformality to Top-Pair Production," *Phys. Rev. D 86* (2012) 014021, arXiv:1204.1405.

[42] B. L. Ioffe, V. S. Fadin, and L. N. Lipatov, "Quantum chromodynamics: Perturbative and nonperturbative aspects," Cambridge University Press, Cambridge (2010).

[43] S. J. Brodsky and X.-G. Wu, "Self-Consistency Requirements of the Renormalization Group for Setting the Renormalization Scale," *Phys. Rev. D 86* (2012) 054018, arXiv:1208.0700.

[44] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and J. Pires, "Second order QCD corrections to jet production at hadron colliders: the all-gluon contribution," *Phys. Rev. Lett. 110* (2013) 162003, arXiv:1301.7310.

[45] E. Mirkes and D. Zeppenfeld, "QCD corrections to jet cross-sections in DIS," *Nucl. Phys. Proc. Suppl. 51C* (1996) 273–280, arXiv:hep-ph/9606332.

[46] S. D. Ellis, Z. Kunszt, and D. E. Soper, "Two jet production in hadron collisions at order α_s^3 in QCD," *Phys. Rev. Lett. 69* (1992) 1496–1499.

[47] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD," *Adv. Ser. Direct. High Energy Phys. 5* (1988) 1–91, arXiv:hep-ph/0409313.

[48] W. Tung, "Bjorken scaling," *Scholarpedia 4* no. 3, (2009) 7412.

[49] V. Gribov and L. Lipatov, "Deep inelastic ep scattering in perturbation theory," *Sov. J. Nucl. Phys. 15* (1972) 438–450.

[50] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys. B 126* (1977) 298.

[51] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.," *Sov. Phys. JETP 46* (1977) 641–653.

[52] R. Herrod and S. Wada, "Altarelli-Parisi equation in the next-to-leading order," *Phys. Lett. B 96* (1980) 195.

[53] W. Furmanski and R. Petronzio, "Singlet Parton Densities Beyond Leading Order," *Phys. Lett. B 97* (1980) 437.

[54] A. Vogt, S. Moch, and J. Vermaseren, "The Three-loop splitting functions in QCD: The Singlet case," *Nucl. Phys. B 691* (2004) 129–181, arXiv:hep-ph/0404111.

[55] S. Moch, J. Vermaseren, and A. Vogt, "The Three loop splitting functions in QCD: The Nonsinglet case," *Nucl. Phys. B 688* (2004) 101–134, arXiv:hep-ph/0403192.

[56] K. Kovarik, T. Stavreva, A. Kusina, T. Jezo, F. Olness, *et al.*, "A Survey of Heavy Quark Theory for PDF Analyses," *Nucl. Phys. Proc. Suppl. 222-224* (2012) 52–60, arXiv:1201.1946.

[57] J. Blümlein, "The Theory of Deeply Inelastic Scattering," *Prog. Part. Nucl. Phys. 69* (2013) 28–84, arXiv:1208.6087.

[58] F. Aaron *et al.* [H1 and ZEUS Collaboration], "Combined Measurement and QCD Analysis of the Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA," *JHEP 1001* (2010) 109, arXiv:0911.0884.

[59] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, *et al.*, "New parton distributions for collider physics," *Phys. Rev. D 82* (2010) 074024, arXiv:1007.2241.

[60] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J. C 63* (2009) 189–285, arXiv:0901.0002.

[61] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections," *Eur. Phys. J. C 64* (2009) 653–680, arXiv:0905.3531.

[62] S. Alekhin, J. Blümlein, and S. Moch, "Parton Distribution Functions and Benchmark Cross Sections at NNLO," *Phys. Rev. D 86* (2012) 054009, arXiv:1202.2281.

[63] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, *et al.*, "A first unbiased global NLO determination of parton distributions and their uncertainties," *Nucl. Phys. B 838* (2010) 136–206, arXiv:1002.4407.

[64] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, *et al.*, "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology," *Nucl. Phys. B 849* (2011) 296–363, arXiv:1101.1300.

[65] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, *et al.*, "Parton distributions with LHC data," *Nucl. Phys. B 867* (2013) 244–289, arXiv:1207.1303.

[66] M. Botje, "QCDNUM: Fast QCD Evolution and Convolution," *Comput. Phys. Commun. 182* (2011) 490–532, arXiv:1005.1481.

[67] S. Bethke, "World Summary of α_s (2012)," *Nucl. Phys. Proc. Suppl. 234* (2013) 229–234, arXiv:1210.0325.

[68] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP 1006* (2010) 043, arXiv:1002.2581.

[69] S. Alioli, K. Hamilton, P. Nason, C. Oleari, and E. Re, "Jet pair production in POWHEG," *JHEP 1104* (2011) 081, arXiv:1012.3380.

[70] B. Webber, "Parton shower Monte Carlo event generators," *Scholarpedia 6* no. 12, (2011) 10662.

[71] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton Fragmentation and String Dynamics," *Phys. Rept. 97* (1983) 31–145.

[72] B. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference," *Nucl. Phys. B 238* (1984) 492.

[73] D. Amati and G. Veneziano, "Preconfinement as a Property of Perturbative QCD," *Phys. Lett. B 83* (1979) 87–92.

[74] M. Gouzevitch, "Mesure de la constante de couplage forte α_s avec les jets hadroniques en Diffusion Inélastique Profonde," Dissertation, Ecole Polytechnique Palaiseau Paris (2008). Available as DESY-THESIS-2008-047.

[75] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. Webber, "New clustering algorithm for multi - jet cross-sections in e^+e^- annihilation," *Phys. Lett. B 269* (1991) 432–438.

[76] S. Catani, Y. L. Dokshitzer, and B. Webber, "The $k_\perp$ clustering algorithm for jets in deep inelastic scattering and hadron collisions," *Phys. Lett. B 285* (1992) 291–299.

[77] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms," *JHEP 9708* (1997) 001, arXiv:hep-ph/9707323.

[78] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering," *Conf. Proc. C980427* (1998) 270–279, arXiv:hep-ph/9907280.

[79] M. Wobisch, "Measurement and QCD Analysis of Jet Cross Sections in Deep-Inelastic Positron-Proton Collisions at $\sqrt{s} = 300$ GeV," Dissertation, RWTH Aachen (2000). Available as DESY-THESIS-2000-049.

[80] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP 0804* (2008) 063, arXiv:0802.1189.

[81] J. Blümlein and A. Tkabladze, "Target mass corrections for polarized structure functions and new sum rules," *Nucl. Phys. B 553* (1999) 427–464, arXiv:hep-ph/9812478.

[82] W. Furmanski and R. Petronzio, "Lepton - Hadron Processes Beyond Leading Order in Quantum Chromodynamics," *Z. Phys. C 11* (1982) 293.

[83] W. van Neerven and E. Zijlstra, "Order α_s^2 contributions to the deep inelastic Wilson coefficient," *Phys. Lett. B 272* (1991) 127–133.

[84] E. Zijlstra and W. van Neerven, "Contribution of the second order gluonic Wilson coefficient to the deep inelastic structure function," *Phys. Lett. B 273* (1991) 476–482.

[85] E. Zijlstra and W. van Neerven, "Order α_s^2 correction to the structure function F_3 (x, Q^2) in deep inelastic neutrino - hadron scattering," *Phys. Lett. B 297* (1992) 377–384.

[86] J. Sanchez Guillen, J. Miramontes, M. Miramontes, G. Parente, and O. Sampayo, "Next-to-leading order analysis of the deep inelastic $R = \sigma_L/\sigma_T$," *Nucl. Phys. B 353* (1991) 337–345.

[87] J. Callan, Curtis G. and D. J. Gross, "High-energy electroproduction and the constitution of the electric current," *Phys. Rev. Lett. 22* (1969) 156–159.

[88] R. Feynman, "Very high-energy collisions of hadrons," *Phys. Rev. Lett. 23* (1969) 1415–1417.

[89] R. Feynman, "Photon-hadron interactions," Benjamin, Reading, MA (1973).

[90] J. Bjorken, "Asymptotic Sum Rules at Infinite Momentum," *Phys. Rev. 179* (1969) 1547–1553.

[91] L. Hand, *Conf. Proc. C670923* (1967) 128.

[92] K. Streng, T. Walsh, and P. Zerwas, "Quark and gluon jets in the breit frame of lepton-nucleon scattering," *Z. Phys. C 2* (1979) 237.

[93] Z. Nagy and Z. Trocsanyi, "Next-to-leading order calculation of four jet observables in electron positron annihilation," *Phys. Rev. D 59* (1999) 014020, arXiv:hep-ph/9806317.

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

[95] Z. Nagy, "Next-to-leading order calculation of three jet observables in hadron hadron collision," *Phys. Rev. D 68* (2003) 094002, arXiv:hep-ph/0307268.

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

[97] Z. Nagy and Z. Trocsanyi, "Calculation of QCD jet cross-sections at next-to-leading order," *Nucl. Phys. B 486* (1997) 189–226, arXiv:hep-ph/9610498.

[98] T. Kluge, K. Rabbertz, and M. Wobisch, "FastNLO: Fast pQCD calculations for PDF fits," *Conf. Proc. C0604201* (2006) 483–486, arXiv:hep-ph/0609285.

[99] M. Wobisch, D. Britzger, T. Kluge, K. Rabbertz, and F. Stober [fastNLO Group], "Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes," Report numbers: DESY-11-150 and FERMILAB-PUB-11-418-PPD (2011). arXiv:1109.1310.

[100] D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch [fastNLO Group], "New features in version 2 of the fastNLO project," *Conf. Proc. C1203261* (2012) 217–221, arXiv:1208.3641.

[101] F. Aaron *et al.* [H1 Collaboration], "Jet Production in ep Collisions at Low Q^2 and Determination of alpha(s)," *Eur. Phys. J. C 67* (2010) 1–24, arXiv:0911.5678.

[102] F. Aaron *et al.* [H1 Collaboration], "Jet Production in ep Collisions at High Q^2 and Determination of alpha(s)," *Eur. Phys. J. C 65* (2010) 363–383, arXiv:0904.3870.

[103] A. Aktas *et al.* [H1 Collaboration], "Measurement of inclusive jet production in deep-inelastic scattering at high Q^2 and determination of the strong coupling," *Phys. Lett. B 653* (2007) 134–144, arXiv:0706.3722.

[104] S. Chekanov *et al.* [ZEUS Collaboration], "Jet-radius dependence of inclusive-jet cross-sections in deep inelastic scattering at HERA," *Phys. Lett. B 649* (2007) 12–24, arXiv:hep-ex/0701039.

[105] S. Chekanov *et al.* [ZEUS Collaboration], "Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA," *Nucl. Phys. B 765* (2007) 1–30, arXiv:hep-ex/0608048.

[106] [ZEUS Collaboration], "Inclusive-jet production in NC DIS with HERA II," ZEUS preliminary report: ZEUS-prel-10-002 (2010).

[107] J. Breitweg *et al.* [ZEUS Collaboration], "Measurement of dijet production in neutral current deep inelastic scattering at high Q^2 and determination of alpha(s)," *Phys. Lett. B 507* (2001) 70–88, arXiv:hep-ex/0102042.

[108] J. Huang, "Determination of the Strong Coupling Constant with Jet Data," DESY Summerstudent report (2011).

[109] H. Abramowicz *et al.* [ZEUS Collaboration], "Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA," *Eur. Phys. J. C 70* (2010) 965–982, arXiv:1010.6167.

[110] H. Abramowicz *et al.* [ZEUS Collaboration], "Inclusive-jet photoproduction at HERA and determination of alphas," *Nucl. Phys. B 864* (2012) 1–37, arXiv:1205.6153.

[111] S. Chekanov *et al.* [ZEUS Collaboration], "Multijet production in neutral current deep inelastic scattering at HERA and determination of alpha(s)," *Eur. Phys. J. C 44* (2005) 183–193, arXiv:hep-ex/0502007.

[112] D. Hoffmann, "HERA Drawings," H1 internal website (2013).

[113] [H1 Collaboration], "List of luminosity summary tables," H1 internal website (2013).

[114] F. Aaron *et al.* [H1 Collaboration], "Determination of the Integrated Luminosity at HERA using Elastic QED Compton Events," *Eur. Phys. J. C 72* (2012) 2163, arXiv:1205.2448.

[115] I. Abt *et al.* [H1 Collaboration], "The Tracking, calorimeter and muon detectors of the H1 experiment at HERA," *Nucl. Instrum. Meth. A 386* (1997) 348–396.

[116] I. Abt *et al.* [H1 Collaboration], "The H1 detector at HERA," *Nucl. Instrum. Meth. A 386* (1997) 310–347.

[117] D. Hoffmann, "Central Tracker drawing menu," H1 internal website (2013).

[118] D. Pitzl, O. Behnke, M. Biddulph, K. Bosiger, R. Eichler, *et al.*, "The H1 silicon vertex detector," *Nucl. Instrum. Meth. A 454* (2000) 334–349, arXiv:hep-ex/0002044.

[119] B. List, "The H1 central silicon tracker," *Nucl. Instrum. Meth. A 501* (2001) 49–53.

[120] J. Kretzschmar, "A Precision Measurement of the Proton Structure Function F_2 with the H1 Experiment," Dissertation, Humboldt Universität Berlin (2008).

[121] I. Glushkov, "D* Meson Production in Deep Inelastic Electron-Proton Scattering with the Forward and Backward Silicon Trackers of the H1 Experiment at HERA," Dissertation, Humboldt Universität Berlin (2007).

[122] J. Becker, K. Bosiger, L. Lindfeld, K. Muller, P. Robmann, *et al.*, "A Vertex trigger based on cylindrical multiwire proportional chambers," *Nucl. Instrum. Meth. A 586* (2008) 190–203, arXiv:physics/0701002.

[123] M. Urban, J. Becker, S. Schmitt, and U. Straumann, "The CIP2k first-level trigger system at the H1 experiment at HERA," *IEEE Trans. Nucl. Sci. 50* (2003) 903–908.

[124] C. Alexa *et al.* [H1 Collaboration], "Measurement of Charged Particle Spectra in Deep-Inelastic ep Scattering at HERA," arXiv:1302.1321. to be published in Eur. Phys. J. C.

[125] P. Laycock, R. Henderson, S. Maxfield, J. Morris, G. Patel, *et al.*, "The H1 Forward Track Detector at HERA II," *JINST 7* (2012) T08003, arXiv:1206.4068.

[126] B. Andrieu *et al.* [H1 Calorimeter Group], "Results from pion calibration runs for the H1 liquid argon calorimeter and comparisons with simulations," *Nucl. Instrum. Meth. A 336* (1993) 499–509.

[127] L. Goerlich and H. Wellisch, "Documentation of the LAr Clustering," H1 internal report: H1-12/91-204 (1991).

[128] H. Wellisch, J. Kubenka, H. Oberlack, and P. Schacht, "Hadronic calibration of the H1 LAr calorimeter using software weighting techniques," H1 internal note: H1-IN-346 (1994).

[129] V. Chekelian [H1 Collaboration], "Simulation and reconstruction in H1 liquid argon calorimetry," *Conf. Proc. C930222* (1993) 302–316.

[130] R. Kogler, "Measurement of jet production in deep-inelastic *ep* scattering at HERA," Dissertation, Universität Hamburg (2011). Available as MPP-2010-175.

[131] R. Appuhn *et al.* [H1 SPACAL Group], "The H1 lead / scintillating fiber calorimeter," *Nucl. Instrum. Meth. A 386* (1997) 397–408.

[132] T. Nicholls *et al.* [H1 SPACAL Group], "Performance of an electromagnetic lead / scintillating fiber calorimeter for the H1 detector," *Nucl. Instrum. Meth. A 374* (1996) 149–156.

[133] H. Bethe and W. Heitler, "On the Stopping of fast particles and on the creation of positive electrons," *Proc. Roy. Soc. Lond. A 146* (1934) 83–112.

[134] C. M. Kiesling, J. Fent, W. Frochtenicht, G. Grindhammer, W. Haberer, *et al.*, "The H1 neural network trigger project," *AIP Conf. Proc. 583* (2001) 36–44.

[135] A. Baird, E. Elsen, Y. Fleming, M. Kolander, S. Kolya, *et al.*, "A Fast high resolution track trigger for the H1 experiment," *IEEE Trans. Nucl. Sci. 48* (2001) 1276–1285, arXiv:hep-ex/0104010.

[136] A. Schoning [H1 Collaboration], "The Fast Track Trigger at the H1 experiment design concepts and algorithms," *Nucl. Instrum. Meth. A 566* (2006) 130–132.

[137] H. Jung, "Hard diffractive scattering in high-energy e p collisions and the Monte Carlo generator RAPGAP," *Comput. Phys. Commun. 86* (1995) 147–161.

[138] T. Sjöstrand and M. Bengtsson, "The Lund Monte Carlo for Jet Fragmentation and e^+e^- Physics. Jetset Version 6.3: An Update," *Comput. Phys. Commun. 43* (1987) 367.

[139] A. Kwiatkowski, H. Spiesberger, and H. Möhring, "HERACLES: an event generator for ep interactions at HERA energies including radiative processes," *Comput. Phys. Commun. 69* (1992) 155–172.

[140] K. Charchula, G. Schuler, and H. Spiesberger, "Combined QED and QCD radiative effects in deep inelastic lepton - proton scattering: The Monte Carlo generator DJANGO6," *Comput. Phys. Commun. 81* (1994) 381–402.

[141] G. Ingelman, A. Edin, and J. Rathsman, "LEPTO 6.5: A Monte Carlo generator for deep inelastic lepton - nucleon scattering," *Comput. Phys. Commun. 101* (1997) 108–134, arXiv:hep-ph/9605286.

[142] G. Gustafson, "Dual Description of a Confined Color Field," *Phys. Lett. B 175* (1986) 453.

[143] L. Lönnblad, "ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model," *Comput. Phys. Commun. 71* (1992) 15–31.

[144] L. Lönnblad, "Rapidity gaps and other final state properties in the color dipole model for deep inelastic scattering," *Z. Phys. C 65* (1995) 285–292.

[145] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, *et al.*, "Event generation with SHERPA 1.1," *JHEP 0902* (2009) 007, arXiv:0811.4622.

[146] T. Sjöstrand, P. Eden, C. Friberg, L. Lönnblad, G. Miu, *et al.*, "High-energy physics event generation with PYTHIA 6.1," *Comput. Phys. Commun. 135* (2001) 238–259, arXiv:hep-ph/0010017.

[147] T. Sjöstrand, L. Lönnblad, and S. Mrenna, "PYTHIA 6.2: Physics and manual," Report number LU-TP-01-21 (2001). arXiv:hep-ph/0108264.

[148] A. Courau, S. Kermiche, T. Carli, and P. Kessler, "Quasireal QED Compton Monte Carlo for HERA," *Conf. Proc. C911029* (1991) 902–915.

[149] A. Courau and P. Kessler, "QED Compton scattering in high-energy electron - proton collisions," *Phys. Rev. D 46* (1992) 117–124.

[150] T. Abe, "GRAPE dilepton (Version1.1): A Generator for dilepton production in ep collisions," *Comput. Phys. Commun. 136* (2001) 126–147, arXiv:hep-ph/0012029.

[151] R. Brun, F. Bruyant, M. Maire, A. McPherson, and P. Zanarini, "GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version," Report number: CERN-DD-EE-84-1 (1987).

[152] J. Meyer, "Guide for the H1 simulation program H1SIM," H1 internal Software-Note: 03-11/89 (1989).

[153] G. Grindhammer, M. Rudowicz, and S. Peters, "The Fast Simulation Of Electromagnetic And Hadronic Showers," *Nucl. Instrum. Meth. A 290* (1990) 469–499.

[154] G. Grindhammer and S. Peters, "The Parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters," *Conf. Proc. C930222* (1993) 294–337, arXiv:hep-ex/0001020.

[155] A. Nikiforov, "Measurements of the neutral current $e^{\pm}$p cross sections using longitudinally polarised lepton beams at HERA II," Dissertation, Ludwig-Maximilians-Universität München (2007). Available as MPP-2007-71.

[156] S. Z. Habib, "Unpolarized neutral current $e^{\pm}p$ cross section measurements at the H1 experiment, HERA," Dissertation, Universität Hamburg (2009). Available as DESY-THESIS-2009-039.

[157] S. Shushkevich, "Measurement of neutral current cross sections with longitudinally polarised leptons at HERA," Dissertation, Ludwig-Maximilians-Universität, München (2011). Available as MPP-2012-583.

[158] C. Kleinwort [H1 Collaboration], "H1 alignment experience," *Conf. Proc. C0609045* (2006) 41–50.

[159] V. Blobel, "Software alignment for tracking detectors," *Nucl. Instrum. Meth. A 566* (2006) 5–13.

[160] T. H. Tran, "Precision measurement of cross sections of charged and neutral current processes at high Q^2 at HERA," Dissertation, University Paris-Sud 11, Paris (2010). Available as LAL-10-28.

[161] V. Blobel, "Central Track Reconstruction," Workgroup presentation (Mar. 2004). Talk at H1 tracking group meeting.

[162] V. Blobel, "A new fast track-fit algorithm based on broken lines," *Nucl. Instrum. Meth. A 566* (2006) 14–17.

[163] P. Bruel, Y. Sirois, and E. Perez, "The standard model challenged at HERA.," *Recherche 298* (1997) 37–39.

[164] B. Andrieu, V. Boudry, T. Carli, and Y. Sirois, "Electron identification in H1 LAr calorimeters," *Nucl. Phys. Proc. Suppl. 32* (1998) 97–105.

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

[166] B. Portheault, M. Peez, and E. Sauvan, "An energy flow algorithm for hadronic reconstruction in OO: Hadroo2," H1 internal note: H1-IN-616 (2005).

[167] U. Bassler and G. Bernardi, "Structure function measurements and kinematic reconstruction at HERA," *Nucl. Instrum. Meth. A 426* (1999) 583–598, arXiv:hep-ex/9801017.

[168] R. Hofstadter and R. McAllister, "Electron Scattering from the Proton," *Phys. Rev. 98* (1955) 217–218.

[169] U. Bassler and G. Bernardi, "On the kinematic reconstruction of deep inelastic scattering at HERA: The Sigma method," *Nucl. Instrum. Meth. A 361* (1995) 197–208, arXiv:hep-ex/9412004.

[170] C. Velkeen, "H1NonepBgFinder - Rejection of Cosmic Muon and Beam-halo Events in the H1OO Framework," H1 internal note: H1-IN-603 (2002).

[171] R. Kogler, "Jets at high Q^2 - Finalising the data treatment," Workgroup presentation (Jun. 2011). Talk given at HaQ workgroup meeting.

[172] R. Kogler, "Reweighting in an extended phase space," Workgroup presentation (Aug. 2011). Talk given at HaQ workgroup meeting.

[173] R. Kogler, D. Britzger, G. Grindhammer, *et al.* [H1 Collaboration], "Measurement of Multijet Production in Deep-Inelastic ep Scattering at high Q^2 at HERA," H1 preliminary document: H1prelim-11-032 (2011).

[174] C. Adloff *et al.* [H1 Collaboration], "Measurement of internal jet structure in dijet production in deep inelastic scattering at HERA," *Nucl. Phys. B 545* (1999) 3–20, arXiv:hep-ex/9901010.

[175] M. Seymour, "Jet shapes in hadron collisions: Higher orders, resummation and hadronization," *Nucl. Phys. B 513* (1998) 269–300, arXiv:hep-ph/9707338.

[176] M. Martisikova, "Jet Shapes in Charm Photoproduction at HERA," Dissertation, Universität Hamburg (2005). Available as DESY-THESIS-2005-047.

[177] J. Breitweg *et al.* [ZEUS Collaboration], "Measurement of jet shapes in photoproduction at HERA," *Eur. Phys. J. C 2* (1998) 61–75, arXiv:hep-ex/9710002.

[178] S. Chekanov *et al.* [ZEUS Collaboration], "Substructure dependence of jet cross sections at HERA and determination of alpha(s)," *Nucl. Phys. B 700* (2004) 3–50, arXiv:hep-ex/0405065.

[179] G. Alexander *et al.* [OPAL Collaboration], "A Comparison of b and (uds) quark jets to gluon jets," *Z. Phys. C 69* (1996) 543–560.

[180] G. P. Salam, "Towards Jetography," *Eur. Phys. J. C 67* (2010) 637–686, arXiv:0906.1833.

[181] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics," *JINST 7* (2012) T10003, arXiv:1205.6201.

[182] V. Blobel, "Unfolding methods in high-energy physics experiments," Lecture notes (1984). Lecture at CERN Summer School for Computing, Aiguablava.

[183] G. D'Agostini, "A Multidimensional unfolding method based on Bayes' theorem," *Nucl. Instrum. Meth. A 362* (1995) 487–498.

[184] A. Höcker and V. Kartvelishvili, "SVD approach to data unfolding," *Nucl. Instrum. Meth. A 372* (1996) 469–481, arXiv:hep-ph/9509307.

[185] V. Blobel and E. Lohrmann, "Statistische und numerische Methoden der Datenanalyse," Teubner, Stuttgart (1998).

[186] G. Cowan, "A survey of unfolding methods for particle physics," *Conf. Proc. C0203181* (2002) 248–257.

[187] V. Blobel, "An Unfolding method for high-energy physics experiments," *Conf. Proc. C0203181* (2002) 258–267, arXiv:hep-ex/0208022.

[188] P. C. Hansen, "Rank-deficient and discrete ill-posed problems," Society for Industrial and Applied Mathematics, Philadelphia, PA (1998).

[189] V. Blobel, "Data unfolding," Lecture notes (Mar. 2010). Lecture given at Terascale statistics school.

[190] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math. Dokl. 4* (1963) 1035–1038.

[191] K. Nowak, "Prompt photon production in photoproduction at HERA," Dissertation, Universität Zürich (2010). Available as DESY-THESIS-2010-011.

[192] D. Britzger, "Unfolding: Jets at High Q^2," Workgroup presentation (Jan. 2011). Talk given at the Miniworkshop on unfolding in H1.

[193] D.-J. Fischer, "Inclusive Neutral Current *ep* Cross Sections with HERA II and Two-dimensional Unfolding," Dissertation, Universität Hamburg (2011). Available as DESY-THESIS-2011-025.

[194] M. Whalley, D. Bourilkov, and R. Group, "The Les Houches accord PDFs (LHAPDF) and LHAGLUE," *Conf. Proc. C0503211* (2005) 575–581, arXiv:hep-ph/0508110.

[195] M. Gluck, E. Reya, and A. Vogt, "Dynamical parton distributions revisited," *Eur. Phys. J. C 5* (1998) 461–470, arXiv:hep-ph/9806404.

[196] B. Schmidt and M. Steinhauser, "CRunDec: a C++ package for running and decoupling of the strong coupling and quark masses," *Comput. Phys. Commun. 183* (2012) 1845–1848, arXiv:1201.6149.

[197] H. Spiesberger, "Program EPRC91," Unpublished program (Jan. 1991). See also: H. Spiesberger, Conf. Proc. C980427 (1998) 530-538.

[198] H. Spiesberger, "Program EPRC91," Unpublished program (May 2012). Modified version for HeraFitter. Using code also from ref. [200] and references therein.

[199] H. Abramowicz *et al.* [ZEUS Collaboration], "Inclusive-jet cross sections in NC DIS at HERA and a comparison of the kT, anti-kT and SIScone jet algorithms," *Phys. Lett. B 691* (2010) 127–137, arXiv:1003.2923.

[200] F. Jegerlehner, "The Running fine structure constant alpha(E) via the Adler function," *Nucl. Phys. Proc. Suppl. 181-182* (2008) 135–140, arXiv:0807.4206.

[201] F. Aaron *et al.* [H1 Collaboration], "Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA," *JHEP 1209* (2012) 061, arXiv:1206.7007.

[202] V. M. Abazov *et al.* [D0 Collaboration], "Measurement of angular correlations of jets at $\sqrt{s} = 1.96$ TeV and determination of the strong coupling at high momentum transfers," *Phys. Lett. B 718* (2012) 56–63, arXiv:1207.4957.

[203] M. Wobisch, K. Chakravarthula, R. Dhullipudi, L. Sawyer, and M. Tamsett, "A new quantity for studies of dijet azimuthal decorrelations," *JHEP 1301* (2013) 172, arXiv:1211.6773.

[204] R. Brown, "Smoothing, Forecasting and Prediction of Discrete Time Series," Dover Publications, Mineola, NY (2004).

[205] S. Schmitt, "Correction of detector effects: bin-by-bin and unfolding," H1 internal note: H1-03/11-633 (2011).

[206] R. Barlow, "SLUO Lectures on Statistics and Numerical Methods in HEP - Lecture 9: Unfolding," Lecture notes (2000). Lecture given at SLUO School, Menlo Park, CA.

[207] V. Anykeyev, A. Spiridonov, and V. Zhigunov, "Correcting factors method as an unfolding technique," *Nucl. Instrum. Meth. A 322* no. 2, (1992) 280 – 285.

[208] V. Anykeyev, A. Spiridonov, and V. Zhigunov, "Comparative investigation of unfolding methods," *Nucl. Instrum. Meth. A 303* no. 2, (1991) 350 – 369.

[209] M. D. Sauter, "Measurement of Beauty Photoproduction at Threshold using Di-Electron Events with the H1 Detector at HERA," Dissertation, ETH Zürich (2009). Available as Diss. ETH No. 18652.

[210] R. Polifka, "Analysis of dijet events in diffractive ep interactions with tagged leading proton at the H1 experiment," Dissertation, Charles University of Prague (2011). Available as DESY-THESIS-2011-025.

[211] D. Britzger, G. Grindhammer, R. Kogler, *et al.* [H1 Collaboration], "Normalised Multi-jet Cross Sections using Regularised Unfolding and Extractions of $\alpha_s(M_Z)$ in Deep-Inelastic Scattering at high Q^2 at HERA,"
H1 preliminary document: H1prelim-12-031 (2012).

[212] P. C. Hansen, "The L-Curve and its Use in the Numerical Treatment of Inverse Problems," in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering, WIT PRESS, 119–142 (2000).

[213] A. Airapetian *et al.* [HERMES Collaboration], "Azimuthal distributions of charged hadrons, pions, and kaons produced in deep-inelastic scattering off unpolarized protons and deuterons," *Phys. Rev. D 87* (2013) 012010, arXiv:1204.4161.

[214] V. Barone, C. Pascaud, and F. Zomer, "A New global analysis of deep inelastic scattering data," *Eur. Phys. J. C 12* (2000) 243–262, arXiv:hep-ph/9907512.

[215] V. Blobel, "Constrained Least Squares Methods with Correlated Data and Systematic Uncertainties," Seminar lecture (Mar. 2010). Lecture given at Terascale statistics school.

[216] S. Schmitt. Private communication (Mar. 2013).

[217] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations,"
Comput. Phys. Commun. 10 (1975) 343–367.

[218] F. James, "Minuit v94.1," CERN Program Library Long Writeup D506 (1994).

[219] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth. A 389* (1997) 81–86.

[220] C. Adloff *et al.* [H1 Collaboration], "Measurement and QCD analysis of jet cross-sections in deep inelastic positron-proton collisions at $\sqrt{s}$ of 300 GeV,"
Eur. Phys. J. C 19 (2001) 289–311, arXiv:hep-ex/0010054.

[221] M. Gouzevitch. Private communication (Mar. 2012).

[222] W. Tung, H. Lai, A. Belyaev, J. Pumplin, D. Stump, *et al.*, "Heavy Quark Mass Effects in Deep Inelastic Scattering and Global QCD Analysis,"
JHEP 0702 (2007) 053, arXiv:hep-ph/0611254.

[223] G. Lobo, C. Pascaud, and F. Zomer, "On parton density error band calculations in QCD analysis of Proton Structure," H1 internal note: H1-01/98-536 (1998).

[224] D. E. Soper, "Basics of QCD perturbation theory," Lecture notes (Jun. 2000). Lecture at TASI Summer School. arXiv:hep-ph/0011256.

[225] J. Campbell, J. Huston, and W. Stirling, "Hard interactions of quarks and gluons: a primer for LHC physics," *Rep. Prog. Phys. 70* no. 1, (2007) 89.

[226] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, *et al.*, "A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project," *Eur. Phys. J. C 66* (2010) 503–524, arXiv:0911.2985.

[227] Z. Bern, L. Dixon, F. Febres Cordero, S. Höche, H. Ita, D. Kosower, D. Maitre, and K. Ozeren, "Multi-jet physics with BlackHat," Conference presentation (Aug. 2012). Talk given by K. Ozeren at QCD@LHC , East Lansing, MI.

[228] G. Aad *et al.* [ATLAS Collaboration], "Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector," *Phys. Rev. D 86* (2012) 014022, arXiv:1112.6297.

[229] Z. Nagy, "Three jet cross-sections in hadron hadron collisions at next-to-leading order," *Phys. Rev. Lett. 88* (2002) 122003, arXiv:hep-ph/0110315.

[230] F. Aaron *et al.* [H1 Collaboration], "Measurement of Dijet Production in Diffractive Deep-Inelastic Scattering with a Leading Proton at HERA," *Eur. Phys. J. C 72* (2012) 1970, arXiv:1111.0584.

[231] S. Schaetzel, "NLO calculations for diffractive dijets," H1 internal note: H1-07/07-628 (2007).

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

Acknowledgement

At the end of my thesis, I would like to take the opportunity to thank really all the people that have helped, motivated, supported and taught me during the last years, but also during my studies. These are all my supervisors, mentors, professors, colleagues, collaborators, fellows, friends, family members, and people who belong to more than one or probably to none of these groups. In particular, I want to express my gratefulness to Günter Grindhammer for his pleasant supervision and teaching, his enduring support, for giving me the opportunity to work within H1 and of course also for the proposal of this topic. Furthermore, I want to thank Klaus Rabbertz and Markus Wobisch for taking me into the fastNLO team and the numerous profound discussions about QCD that followed. A big thank goes also to Stefan Schmitt for giving enormously valuable suggestions within shortest timescales on the unfolding topic, but also on probably all other issues addressed in this work. A big thank goes also to my friends and office mates Aziz, Philipp, Roman, and Stanislav, who motivated and encouraged me and made life (work) much more joyful, and for being friends, colleagues and also supervisors altogether. Thank you all for your support, and last but not least, I want to thank you, the reader, for reading this thesis and sharing the interest on QCD and this topic with me.