Measurement of the differential Drell-Yan production cross-section and application of deep convolutional neural networks on event images in the context of pileup mitigation

Dissertation
zur Erlangung des Grades

"Doktor
der Naturwissenschaften"

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

Bernard Brickwedde

geb. in Mainz
Mainz, den 18.05.2020

1. Berichterstatter: [redacted]
2. Berichterstatter: [redacted]
Datum der mündlichen Prüfung: 20.08.2020

Ich versichere, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Mainz, den 18.05.2020

Zur Publikation wurde die Titelseite den Anforderungen entsprechend angepasst. Zusätzlich wurden kleine Korrekturen eingearbeitet. Der Inhalt bleibt dabei unverändert. Eine Liste der Änderungen findet sich im Anhang A.7.

Zusammenfassung

Der Large Hadron Collider (LHC) am Forschungszentrum CERN in der Nähe von Genf ist der Teilchenbeschleuniger mit der zum Zeitpunkt dieser Arbeit höchsten Schwerpunktsenergie. 2012 verkündeten die beiden LHC-Experimente ATLAS und CMS die Entdeckung des vom Standardmodell der Teilchenphysik vorhergesagten Higgs-Bosons. Diese Entdeckung vervollständigt das Standardmodell, das jedoch noch einige wesentliche Defizite aufweist, darunter die fehlende Beschreibung der Gravitation oder eine Erklärung für die dunkle Materie im Universum. Daher suchen Physiker in Hochenergie-Experimenten nach neuer Physik jenseits des Standardmodells: Einerseits wird direkt nach neuen Effekten gesucht, andererseits werden Präzisionsmessungen durchgeführt, die die Vorhersagen des Standardmodells testen und dadurch Inkonsistenzen aufdecken können. Zu letzterem zählt die Messung der W Boson Masse, deren Präzision wiederum stark von der genauen Kenntnis der Transversalimpulsverteilung des W Bosons abhängt. Diese Verteilung kann sowohl direkt als auch indirekt gemessen werden.

Die direkte Messung nutzt Kalorimeterdaten und wird daher stark von zusätzlichen Teilchenkollisionen, so genanntem Pileup, beeinflusst. Der erste Teil dieser Arbeit beschreibt hier einen neuartigen Ansatz die Auswirkungen von Pileup durch maschinelles Lernen und Bilderkennungstechniken zu unterdrücken. Dabei werden künstliche neuronale Netze ausgehend von Bilddaten der Detektorsignale darauf trainiert zwischen Pileup und Primärkollision zu unterscheiden.

Die indirekte Messung der W Boson Transversalimpulsverteilung hängt wiederum direkt vom differentiellen Wirkungsquerschnitt des Drell-Yan Prozesses ab. Die Präzisionsmessungen sowohl des integrierten als auch des normierten differentiellen Wirkungsquerschnitts dieses Prozesses werden im zweiten Teil der Arbeit beschrieben. Diese beiden Messungen, letztere eine der bisher genauesten Messungen der ATLAS Kollaboration, verwenden $\mathcal{L} = 36.1\,\mathrm{fb}^{-1}$ an Daten, die in den Jahren 2015 und 2016 bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\mathrm{TeV}$ mit dem ATLAS Detektor aufgenommen wurden.

Beide Teile dieser Arbeit führten zu eigenen Publikationen, die als Referenz [34] und Referenz [54] im Literaturverzeichnis aufgeführt sind.

Abstract

The Large Hadron Collider (LHC), currently the most powerful particle accelerator, is hosted at the CERN research complex near Geneva. In 2012, two LHC experiments, namely ATLAS and CMS, presented the discovery of the Higgs boson, the last till then undiscovered particle predicted by the Standard Model of Particle Physics. Although this discovery completes the Standard Model, it has some critical deficits like, among others, the missing inclusion of gravity or an explanation of the dark matter in the universe. Hence, physicists search for new physics beyond the Standard Model in high energy physics experiments. Such searches can be performed following two approaches, on the one hand, direct searches for new effects and, on the other hand, precision measurements. The latter probe predictions of the Standard Model and can discover inconsistencies between predictions and observations. The measurement of the mass of the W boson is such a probe, which relies on the precise knowledge of the W boson transverse momentum distribution to achieve high precision. This distribution can be measured either directly or indirectly.

The direct measurement uses calorimeter data and is profoundly affected by pileup activity, a collective term for additional particle collisions happening in parallel to the collision under investigation. The first part of this thesis describes a novel approach aiming to mitigate the effects of pileup activity using machine learning and image recognition techniques. Deep convolutional neural networks are trained on images created from several detector signals to differ between the primary collision and pileup activity.

The indirect measurement, on the other hand, directly depends on the measurement of the differential cross-section of the Drell-Yan Z boson production process. The second part presents and discusses a precision measurement of both the integrated and the normalized differential cross-section of this process. These measurements, the latter being one of the most precise measurements within ATLAS thus far, was performed using $\mathcal{L} = 36.1\,\mathrm{fb}^{-1}$ of data collected in 2015 and 2016 at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector.

Both parts of this thesis resulted in dedicated publications, listed as Ref. [34] and Ref. [54].

Table of Contents

Chapter 1

Overview

The Standard Model of Particle Physics (SM) describes the fundamental laws of physics. While qualitative descriptions of physical observations and the internal structure of matter can be dated back to ancient Greece, well defined mathematical concepts and symmetry groups providing precise predictions of particle physics processes form the basis of the SM. These predictions offer the possibility to test this model in dedicated experiments like ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**), the Higgs boson being the most recent and publicly received discovery. Although measurements continuously test the predictions of the SM, no result significantly differs from its predictions so far.

This thesis describes two approaches to improve the W boson mass measurement by reducing its model dependency, both using data created from Z bosons decaying into a pair of leptons, the Drell-Yan process. Data from proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ is analyzed in this thesis. The data is collected by the ATLAS detector, a multi-purpose detector at the **L**arge **H**adron **C**ollider (LHC), a storage ring facility at CERN in Switzerland.

The first approach focuses on improving the direct measurement of the boson transverse momentum through the hadronic recoil. Particles created by initial-state radiation are analyzed using calorimeter clusters to reconstruct the kinematics of the boson. This measurement method is profoundly affected by so-called pileup activity, additional collision happening in the same bunch-crossing. The performance of a novel approach aiming to mitigate the effect of pileup by applying deep convolutional neural networks on event images is studied in simulated data. It has been partially published as Ref. [34].

The second approach focusses on the precise measurement of the differential cross-section of the production of Drell-Yan lepton pairs in the Z boson peak region, a process well described by the SM. The results of this measurement are published in Ref. [54]. Since the production processes of W and Z bosons in proton-proton collisions are similar, the results can be used to tune event generators to constrain model uncertainties for the measurement of the W boson mass.

This thesis is structured as follows. Following this chapter, three dedicated chapters discuss the theoretical foundations of this thesis. Chapter 2 provides an introduction to machine learning and core concepts of deep neural networks, emphasizing the description of convolutional layers, the gradient descent algorithm, padding methods, and advanced optimizers, as the machine learning project covered in Chapter 8 uses these. Afterward, the SM, together with its particles and fundamental interactions, are presented in Chapter 3. Last but not least, Chapter 4 discusses the core concepts of particle collisions at the LHC.

Before discussing the results of the projects covered in this thesis, the experimental setup, namely LHC and ATLAS detector, is introduced. Chapter 5 provides a summary of CERN and its mission and the LHC accelerator complex. Chapter 6 focusses on the ATLAS detector and its subsystems. The description of the ATLAS detector includes its subdetectors, the trigger system, detector simulations, kinematic variables and the coordinate system used to describe the position and movement of particles within the detector.

Chapter 7 describes the W boson mass measurement strategy at hadron colliders and provides additional context for the two projects described afterward. The application of image recognition techniques combined with deep convolutional neural networks on event images for pileup mitigation and, therefore, the direct measurement of the W boson transverse momentum is discussed in Chapter 8. The chapter also includes a discussion of established pileup mitigation techniques as well as the dataset. Building upon the discussion of perks and limitations of the pileup mitigation approach, the precise measurement of the integrated and normalized differential fiducial cross-section in bins of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ and ϕ_{η}^{*} of Drell-Yan lepton pairs in the Z peak region $66\,\mathrm{GeV} < m^{\ell\ell} < 116\,\mathrm{GeV}$ and $\left|\eta^{\ell}\right| < 2.5$ is described in Chapter 9. This chapter includes a description of the analysis strategy and the fiducial volume, the datasets used in both data and simulation, the correction of detector effects, and a discussion of uncertainties limiting the precision of the measurement, using both the Drell-Yan electron and muon pair decay channels. A combined measurement of both the differential and integrated cross-section, using the results of these two channels, is performed and compared to theoretical predictions.

Finally, Chapter 10 recapitulates the content of the work described in this thesis and reflects on possible follow-up studies building up on the obtained results.

Chapter 2

Introduction to machine learning

Contents

Machine learning, a subset of artificial intelligence, is a computational model inspired by the structure of the human brain and its ability to learn from processed information. So-called **A**rtificial **N**eural **N**etworks (ANN)s, neural networks, in short, consist of single computational units called neurons that can form complex architectures. They are trained on (usually large) datasets utilizing dedicated frameworks like Keras [93] and Tensorflow [123]. Thereby the networks learn to mimic the underlying behavior of the dataset by providing only input and output data. ANNs are a standard tool in both industry and many branches of science, including particle physics. Several studies discuss and evaluate possible applications in particle physics, like, e.g., PUMML [112] or ANN based unfolding [100], but there are no commonly used tools so far.

Defining a suitable architecture is, besides data preparation and data processing, the main challenge of machine learning. During the training process, the neurons try to learn the systematic behavior behind a provided dataset and form connections within the network to fit its structure to the general concept behind the dataset. There are different approaches to training a network. When applying supervised learning, the network is provided the desired output during training and tries to generalize this by minimizing a so-called loss function (putting the networks' current output to the desired output) using the gradient descent algorithm. Both are described later. Contrary to this, unsupervised learning is a method of adapting to data without providing a direct measure but a scoring mechanism. Only supervised learning techniques are used in this thesis. Generally speaking, there are two types of applications.

The first one, called classification, provides a probability of a particular input to be a member of a specific output class or, if provided with more than one class, its probability distribution over all classes. For the training, one only has to provide the class of the given input. The network then tries to distinguish between the classes. Classification problems can solve a variety of different issues reaching from particle identification within a detector to image classification or fraud detection.

The other application field is called prediction. The network tries to compute an output value to a given input as precisely as possible, imitating the behavior of the dataset used during training. Typically this is used in image processing, e.g., applying filters or object identification. For every machine learning-based application, the training dataset has to cover as many possible input cases as possible as inputs vastly different from the training dataset might not yield the correct result.

In this Chapter, only a subset of neural networks, called feedforward neural networks, are discussed in detail, like these used in Chapter 8. Ref. [124, 153] provide a concise introduction to other types of networks and training methods commonly used.

2.1 Feedforward neural networks

Feedforward **N**eural **N**etworks (FNN) [148] are a subclass of neural networks, where information only flows in one direction from the input to the output nodes passing an arbitrary network architecture between those. Its inner connections do not form any cycles or loops. As a result, a neuron only has inputs coming from the output of neurons closer to the input nodes. The purest form and, at the same time, the historical starting point of machine learning, are so-called perceptrons. A perceptron itself is a FNN and, at the same time, the basic structure of more complex networks.

2.1.1 The perceptron

The most simple form of a FNN is a single layer perceptron [144], or just perceptron in short. A perceptron, shown in Figure 2.1, can be described by the following variables:

- An arbitrary number $i \geq 1$ of inputs x_i with corresponding input weights w_i
- The combination function $\sum$ with the bias b
- The activation function σ

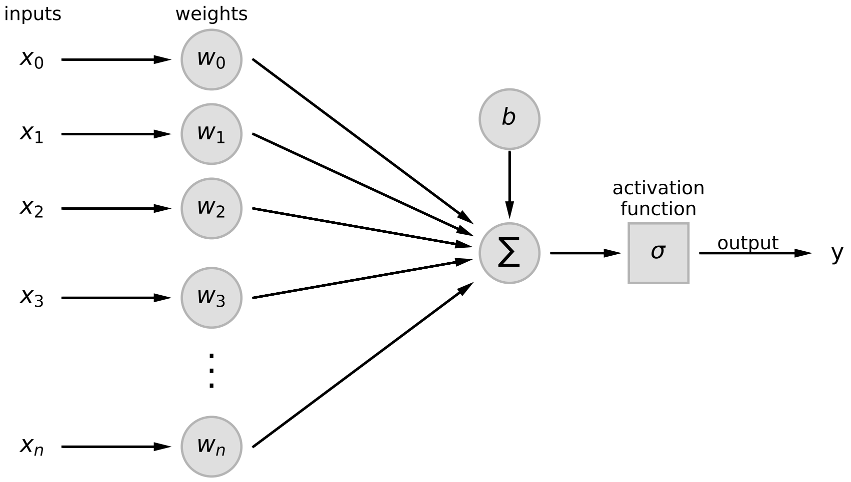


Figure 2.1: Visualization of a perceptron with all its components. The inputs x_i are combined with the perceptrons internal weights w_i and bias b via the combination function $\sum$. This is then fed to the activation function σ.

Each input x_i is weighted with its corresponding input weight w_i and combined by the combination function $\sum$. Note that this usually means to sum all these weighted inputs and the bias, but this may differ in some corner cases.

The result of the combination function is then fed to the activation function σ, resulting in the definition of the output y of the perceptron as:

$$y = \sigma\left(\sum_i (w_i \cdot x_i) + b\right). \tag{2.1}$$

Depending on the problem at hand, the initialization of weights and biases changes. While commonly a uniform or Gaussian distribution over all variables are chosen, also other approaches like all zero or all one are possible. When training huge or complex networks, there are more advanced ways to initialize weights and bias like Xavier [102] or Kaiming [105] initialization. As the ANNs used in this thesis are rather simple, these methods are not covered as these are not applied in the context of this thesis. Assuming a perceptron with only two inputs and step function as activation function, one can already implement basic logical AND, OR and NOT gates, from which more complex calculations can be built, which shows the computing capabilities of neural networks. The truth table of these can be seen in Table 2.1, while the corresponding perceptrons can be seen in Figure 2.2.

Table 2.1: Truth table for different logic gates

A	B	AND gate	OR gate	NOT A
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	0

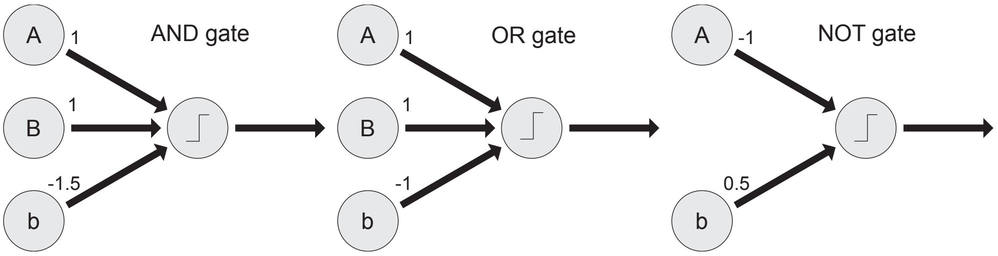


Figure 2.2: Perceptrons implementing logic AND(left), OR(mid) and NOT(right) gates. A and B refer to inputs, b to the bias.

Several perceptrons are required to implement more complex gates like XOR, as XOR is defined as a combination of the gates defines above $(XOR(A, B) = (A \vee B) \wedge (\overline{A \wedge B})$. Alternatively, a more sophisticated architecture like a multilayer perceptron, which is better able to generalize based on the training data due to the higher number of degrees of freedom, can solve this problem.

2.1.2 The multilayer perceptron

A multilayer perceptron [146] is at its heart just a combination of perceptrons. As information in a FNN only flows in one direction, these perceptrons naturally form layers containing several perceptrons with an equal distance to the input layer. All multilayer perceptrons must have at least one layer as a starting point, called the input layer, one layer as its endpoint, called the output layer, and at least one of so-called hidden layers, layers not being input or output, between those. Network architectures containing more than one hidden layer are called deep neural networks. An example is shown in Figure 2.3.

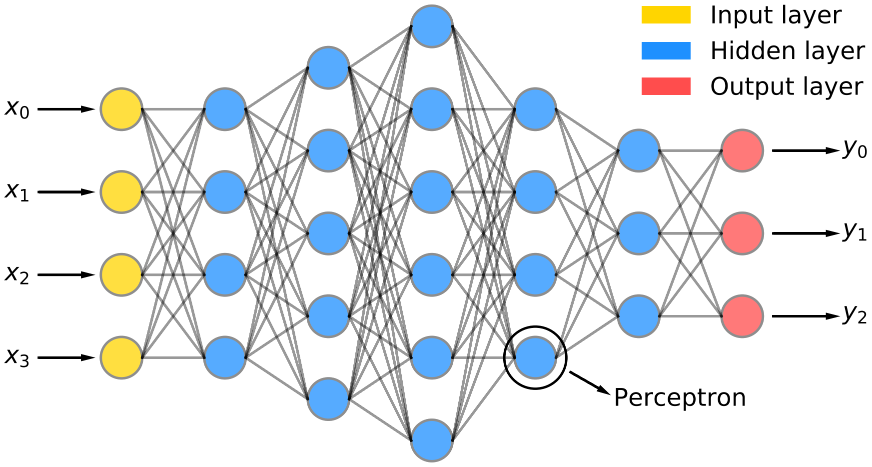


Figure 2.3: Possible architecture of a multilayer perceptron with four inputs, three outputs, and five hidden layers of perceptrons. As these are feedforward neural networks, the information can only flow from the input through the hidden layers to the output. Note that it is still possible for the information to ignore layers and flow directly to the output or another hidden layer that is not directly adjacent.

The network passes information from the input layers through all hidden layers to the output layers using the output of all neurons of the previous layer as an input of all neurons of a certain layer.

As stated above, these networks can implement a variety of complex algorithms. Starting from these basics concepts, different types of layers are possible, such as Dropout layers, convolutional layers, pooling layers, or dense layers. These layer types, which are all used in the projects outlined in this thesis, are introduced in Section 2.2 below. After this, the process of training ANNs, the so-called gradient descent, is described in Section 2.5.

2.2 Neural network layer types

Building on the basics introduced so far, different layer types for ANNs can be defined. This section introduces the most important layer types typically used, but obviously, one can define customized layer types as well, as can be seen later. Depending on the problem, certain layers are more suited than others to be used as part of a suitable architecture.

2.2.1 Dense layers

Dense layers, depending on the framework used, also known as fully connected layers, are defined as a layer with weighted connections between every input and every output. Every output unit provides a bias, which is, alike the connection weights, a trainable parameter. This is visualized in Figure 2.4. One can compare such a layer to a multilayer perceptron with just an input and output layer.

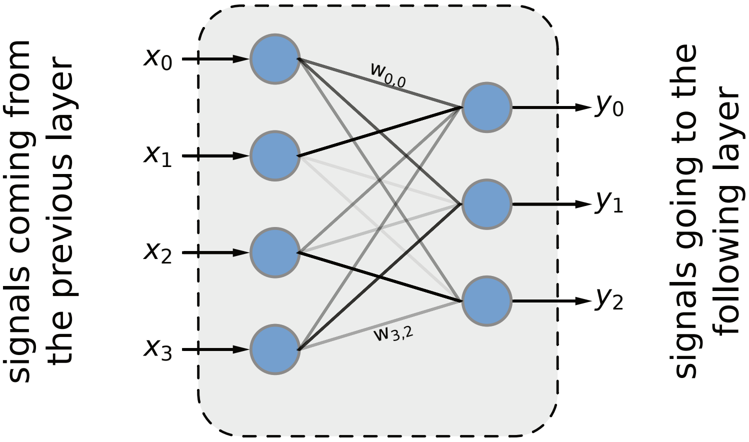


Figure 2.4: Visualization of a dense layer. In this examples four inputs x_i are mapped to three outputs y_j via the matrix w_{ij}. Some of the weights are marked for better understanding while the weight of a connection between input and output is given by its line width.

These layers can be seen as the brute force approach to finding a working architecture, as all inputs are connected to all outputs. The layer differs between relevant and not relevant connections between an input x_i and an output y_j via its weight w_{ij}. If a specific input does not affect a particular output, the connecting weight decreases to zero over the training period, basically removing the connection between input and output. The biggest weakness of dense layers is the exponentially growing number of trainable parameters with the number of hidden layers.

Given the inputs x_i, the weights w_{ij} and the activation function σ the output of unit i can be calculated by:

$$y_i = \sigma\left(\left[\sum_j w_{ij} \cdot x_j\right] + b_i\right), \tag{2.2}$$

with w_{ij} being the connection weight of neurons i and j and b_i the bias of neuron i. A dense layer with n_{in} inputs and n_{out} output neurons has $n_{in} \cdot n_{out}$ trainable variables for the weight matrix as well as n_{out} variables for the biases of the output neurons.

2.2.2 Dropout layers

Dropout layers [156] try to tackle the problem of overfitting, as discussed in Section 2.5.3. A dropout layer connects outputs of the layer in front to inputs of the layer behind the dropout layer. It has no trainable parameters, but randomly deactivates a manually defined fraction of its connections for each training sample during training. Hence, dropout layers are ignored when using a trained network to perform predictions or classifications.

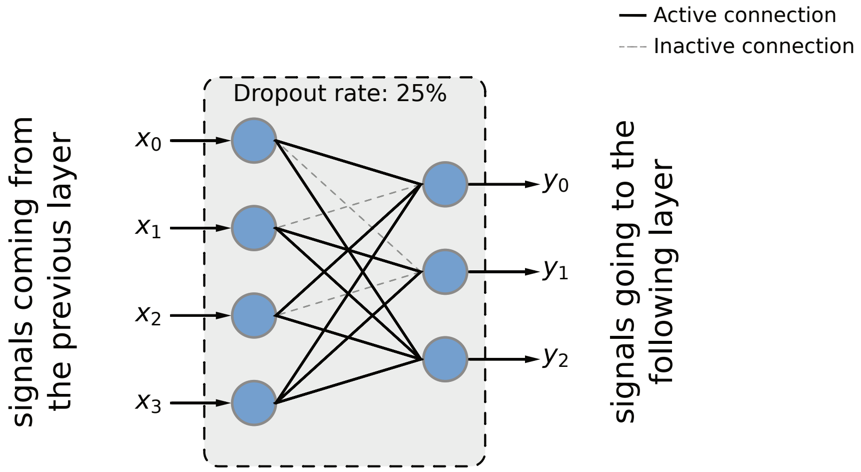


Figure 2.5: Visualization of the working principle of dropout layers. All inputs are connected to all the outputs like a dense layer, but for each training step a predefined fraction of these connections are disabled. This is shown by the dashed lines in this example setup.

Using dropout fundamentally changes the concept of training ANNs, as instead of training one architecture, each dropout layer creates a new architecture per training step. These new architectures are always a substructure of the original architecture and, given enough statistics, each part of the network is trained as often as the other parts. Using dropout layers usually results in a trade-off between a higher training effort, as some parts are excluded from training and a higher generalization of the problem.

2.2.3 Convolutional layers

Convolutional layers are mainly used in image processing and pattern recognition. Unlike fully connected layers, convolutional layers scan the input in slices with so-called kernels of predefined size. Each kernel works like a filter that can learn different features within the data, like, for example, edges or lines. Depending on the input, there are one-dimensional and two-dimensional convolutional layers. These layers require two-dimensional and three-dimensional input, respectively. The reason for this kind of implementation is versatility, as the same convolutional layer is then able to process multiple input images stacked in the additional dimension.

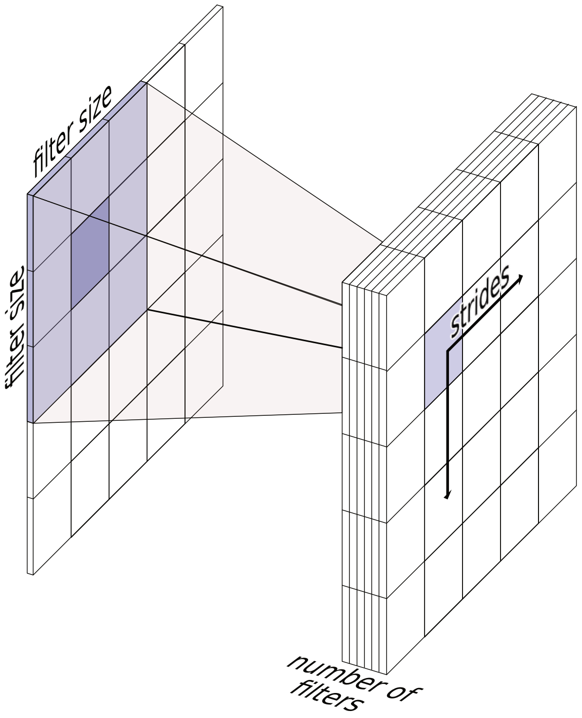


Figure 2.6: Visualization of convolutional layers. The kernel, shown in purple, is moved over the input in steps defined by strides, resulting in an output image with depth equal to the number of filters applied in the kernel.

In the context of this work, these can, for example, be inputs from different detector systems, all covering the same event, which gives the network additional information to process. The kernel is moved over the entire input by a certain step size, called stride. This is shown in Figure 2.6. Given a two dimensional convolutional layer, the kernel size and the stride can be defined for each of its dimensions. The remaining, uncovered dimension of the input is affected by the number of filters n_{filters}. The output size of this dimension is completely independent of the input size of this dimension, always n_{filters}.

Convolutional layers can be fully described by the following sets of parameters:

- Number of filters
- Filter size
- strides
- padding

The trainable variables of such a convolutional layer are independent of the actual input but are defined by the *filter size* and the number of filters $n_{filters}$. A convolutional layer using an one-dimensional input has only $n_{filters} \cdot size$ trainable parameters. In the two-dimensional case, there would be $n_{filters} \cdot size^2$ parameters, given a symmetric filter size. The movement of the kernel over the input is defined by the **stride** and **padding** parameters. While stride simply defines the step size the filter is moved over the input, padding describes the behavior along the edges of the input image. Additionally, these two parameters define the output size of the convolutional layer, for example, a stride of two would cut in half the input dimensions because every second position is skipped.

There are two standard ways of padding called **valid** and **zero padding**, but one could think of other approaches, depending on the task at hand. Both valid and zero-padding are visualized in Figure 2.7

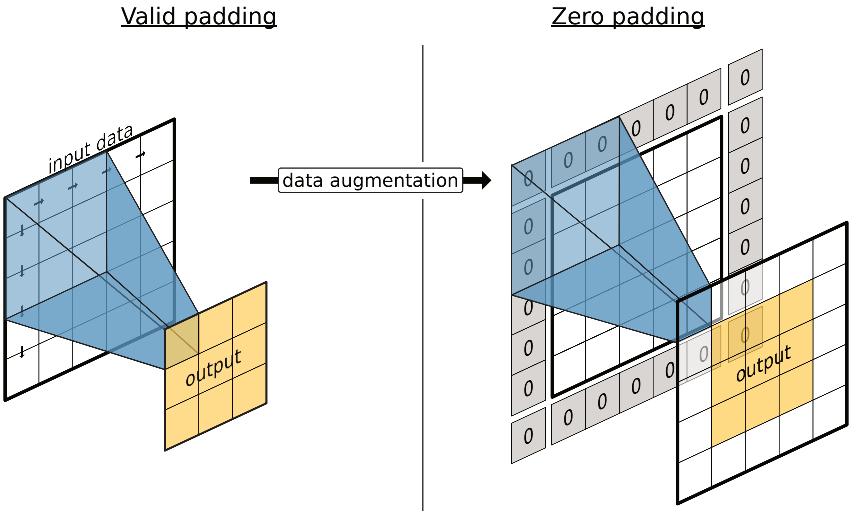


Figure 2.7: Visualization of both valid padding and zero padding. Valid padding uses the input presented, while zero padding augments this input. As a result using zero padding results in a better control of the ANNs output. Note that the yellow areas are identical.

Valid padding uses only the input, having a size W, as presented to the ANN. The kernel of the corresponding layer moves over the input as long as it still entirely fits inside.

Given a filter size K larger than one and a stride of S, larger than 1 in at least one dimension, results in a smaller output image compared to the dimensions of the input. Using the Padding parameter $P = \frac{K-1}{2}$, one can calculate the output size O of each filter by:

$$O = 1 + \frac{W - K + 2P}{S} = 1 + \frac{W-1}{S}. \tag{2.3}$$

Zero padding augments the input as such as adding rows and columns filled with zeros around the input if needed. If the kernel laps over the actual input, the missing entries are filled with zeros. Using zeros removes the effect of the weights connected to these entries and introduces a mechanism to adjust the size of the output of a convolutional layer. Stride 1, for example, preserves all input dimensions, except the output depth, which is defined by the number of filters.

To summarize, a convolutional layer can be seen as a stack of n_{filters} small dense layers that are slid over the input image, thereby computing the new image.

2.2.4 Pooling layers

Pooling layers are used to collect either local or global features like the maximum or the average of a layer (global) or parts of it (local). This is visualized in Figure 2.8.

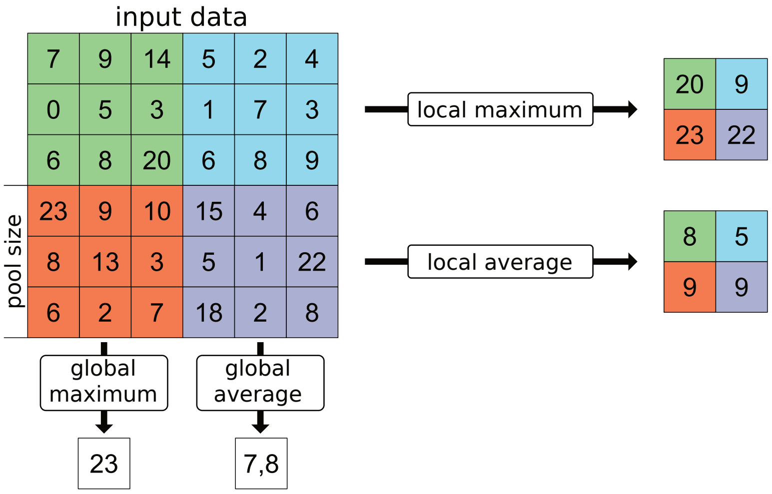


Figure 2.8: Example of different pooling layers and their results. Both the results of global and local maximum and average pooling are provided.

Collecting local information results in a resized multidimensional output with its dimensions depending on the pooling size. Within the area defined by the pooling size, the chosen metric, e.g., maximum or average, is collected and written to the output. Contrary to this, collecting global information results in just one output, as the metric applied to the complete input.

2.3 Activation functions

So far, different layer types from the very basic perceptron to more complex types of layers have been described. Except for Dropout layers, the output of these layer types is determined by its weights, biases, and activation function. The latter has been treated rather generically so far. There are several well suited standard activation functions commonly used among a variety of applications. A selection of these are listed in Table 2.2 and are plotted in Figure 2.9.

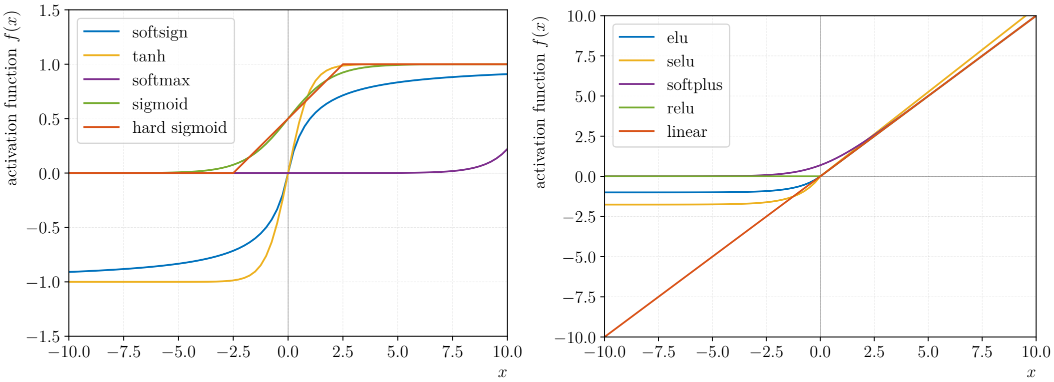


Figure 2.9: A selection of commonly used activation functions for ANN layers. Left shows functions limited to a certain value. Some of these limited to the range of 0 to 1 are commonly used for last layers of classification problems. Right shows activation functions with no upper limit which can be used for a variety of use cases.

The choice of activation functions highly depends on the problem at hand and if a layer is hidden or an output layer. The function for a classification problem usually aims for a probability distribution as output and should hence be limited to 1 as a maximum value. Good examples for these are sigmoid or step functions. Usually, only the last layer of a network aiming to solve a classification problem has to have an activation function limited to the described above, as the last layer produces the final output of the network.

Generally speaking, there are no of these limitations mentioned above for a prediction problem. The topology of the input dataset defines which activation functions are suitable. Having negative values present, using activation functions that are 0 for negative input values, removes information from the network and is therefore not an appropriate choice. This approach is inverted when working on a dataset spanning over several magnitudes of values. Here an activation function introducing an upper limit can be the right choice. However, there are no general rules on how to approach but more simple rules to get started. If computation time is an issue, as it is, for example, a trigger system, simpler functions like a linear function are used as these are computational lightweight.

Table 2.2: Commonly used activation functions together with their corresponding definition and derivative.

Activation function	Definition	Derivative
Rectified linear unit (ReLu) [128]	$f(x) = \begin{cases} x \text{ for } x > 0 \\ 0 \text{ else} \end{cases}$	$f'(x) = \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ else} \end{cases}$
elu	$f(x) = \begin{cases} x \text{ for } x > 0 \\ \alpha \cdot (e^x - 1) \text{ else} \end{cases}$	$f'(x) = \begin{cases} 1 \text{ for } x > 0 \\ \alpha \cdot e^x \text{ else} \end{cases}$
leaky ReLu	$f(x) = \begin{cases} x \text{ for } x > 0 \\ \alpha \cdot x \text{ else} \end{cases}$	$f'(x) = \begin{cases} 1 \text{ for } x > 0 \\ \alpha \text{ else} \end{cases}$
Softplus	$f(x) = \ln(1 + e^x)$	$f'(x) = \frac{1}{(1+e^{-x})}$
Linear	$f(x) = x$	$f'(x) = 1$
Softsign	$f(x) = \frac{x}{1+\lvert x\rvert}$	$f'(x) = \frac{1}{(1+\lvert x\rvert)^2}$
Sigmoid	$f(x) = \frac{1}{(1+e^{-x})}$	$f'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$
Tangens hyperbolicus	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$f'(x) = 1 - f(x)^2$
Softmax	$f_j(x) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$ for $j = 1....K$ given K output neurons	$\Delta_i f(x)_j = f_i(x)(\delta_{ih} - f_j(x))$
Hard sigmoid	$f(x) = \min(1, \max(0, 0.2 \cdot x + 0.5))$	$f'(x) = \begin{cases} 0 \text{ if } \lvert x\rvert > 2.5 \\ 0.2 \text{ else} \end{cases}$

2.4 Data preparation

One of the most important tasks when training an ANN is the preparation of the dataset. It should fulfill the following criteria:

- Have enough statistics for proper training. The amount of data highly differs depending on the size of the network and the problem under investigation. If applicable, e.g. if invariants exist, data can be rotated or mirrored to enlarge statistics.
- Does not contain duplicates of the same data points as the influence of this data point is then overestimated.
- Cover as many edge cases as possible.
- Should have no bias or in reality close to no bias

Generally speaking, three independent datasets are used when training an ANN. These datasets can be either obtained by splitting the available dataset into three parts or come from different sources all describing the same underlying process and probability distributions and are called:

- Training dataset,
- Validation dataset,
- Test dataset.

The training dataset is used for everything related to the training process, as described in Section 2.5. This process includes updating the weights and biasses of the ANN. During training, the training dataset is processed by the ANN several times, called epochs.

The performance of the network is measured on the validation dataset after each iteration over the training dataset. Using an independent dataset enables an unbiased performance review of the network, as this dataset has never been seen by the network before. This way, issues like under- or overfitting, as discussed in Ref. 2.5.3, can be spotted.

The last dataset, called the test dataset, is used as a final, unbiased measure of performance for fully trained models. This is especially important when applying machine learning training techniques that use the performance on the validation dataset for evaluation. Such techniques are e.g. *Checkpointing* or *Early-Stopping*.

Checkpointing only stores the model with the best performance on the validation dataset instead of the model after the last training epoch, while Early-Stopping terminates the complete training process once the performance on the validation dataset worsens.

2.5 The training process

The process of updating the neurons weights and bias of an ANN is called training. It is a time-intensive process running over several epochs, ideally leading to a good generalization of the training dataset. The training dataset usually gets shuffled after each epoch to achieve further generalization. The ANNs performance is evaluated on the validation and test dataset, as described above. A set of so-called hyperparameters describes ANNs. These include the architecture itself with all layers and their respective activation functions but also further tunable parameters. Examples of these tunable parameters are, on the one hand, fixed parameters not updated through the training process like the dropout rate but also parameters affecting the training process like the learning rate α or the optimizer used for training.

2.5.1 Gradient descent

The algorithm commonly used to train ANNs is called gradient descent. It aims to minimize the networks' so-called loss function, sometimes called the error function, $\mathcal{L}(y, \hat{y})$ during the training process. This loss function is a performance metric describing the difference between the output of the network y and the desired output $\hat{y}$. It directly depends on the weights and biasses stored in the network's layers, as the network's output is computed using these parameters. The weights and biasses are updated after each training step as the loss is propagated back through the network.

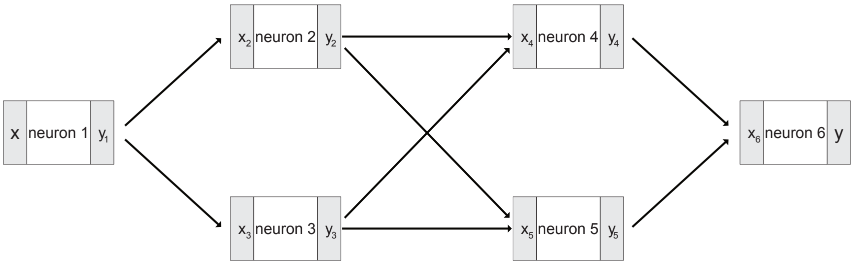


Figure 2.10: Exemplary 6 neuron ANN with inputs x and outputs y.

Given a differentiable multi-variable function F, Gradient descent is based on the idea that the fastest way to minimize this function is to move from a given initial point a in the functions hyperparameter space along the negative gradient of $-\nabla F(a)$. On this assumption, one can write the process of gradient descent as a monotonically decreasing series of a_n as:

$$a_{n+1} = a_n - \nabla F(a_n). \tag{2.4}$$

This process is repeated until the loss function reaches sufficiently small values, ideally 0, given such fix points exist at all.

Gradient descent is split into two phases, **forward pass** and **back propagation**. Assuming a network like shown in Figure 2.10 with inputs x, outputs y and activation function $\sigma(x)$, these can be described as follows.

During the forward pass, the output of a neuron j, and as a result finally the output of the network is calculated by:

$$y_j = \sigma(x_j), \tag{2.5}$$

with the input in neuron j

$$x_j = \sum_{i \in \text{inputs}(j)} w_{ij} \cdot y_i + b_j. \tag{2.6}$$

Hereby the inputs of x_j can be either the input of the network or the output of the neurons connected to neuron j.

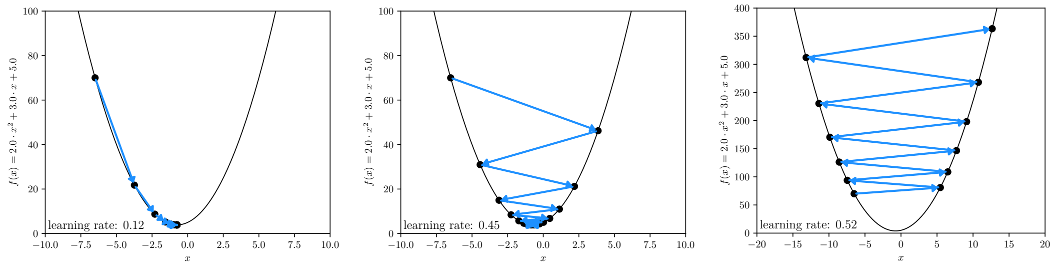


Figure 2.11: Gradient descent for different learning rates exemplary shown on a parabola. Using a sufficient learning rate as shown in the left picture, gradient descent efficiently converges to the minimum. A too large learning rate can result in jumpy behavior (middle) or no convergence at all (right). As the function is given, its derivate cane be directly used for back propagation.

During back-propagation, the weights are updated depending on the comparison of the networks' actual output y and the designated output $\hat{y}$. From there, the loss of the network $\mathcal{L}(y, \hat{y})$ is calculated. Starting from Equation 2.4, the weights are updated according to:

$$w_{ij}^{\text{new}} = w_{ij} - \alpha \frac{d\mathcal{L}(y, \hat{y})}{dw_{ij}}, \tag{2.7}$$

with α being the so-called learning rate, a tunable parameter describing which fraction of the actual gradient should be used for the next step in the series. Figure 2.11 visualizes the influence of an appropriate or inappropriate choice of learning rate.

Given that machine learning is usually used to iteratively learn to generalize data too complicated to be fully described, the learning rate is one of the tunable parameters of a network together with, for example, the architecture.

The change of the loss function depending on the output of the network is given by the derivative $\frac{d\mathcal{L}(y,\hat{y})}{dy}$. With $f'(x)$ being the derivative of the activation function, the next step is describes the dependence of the loss function of a certain input x_j:

$$\frac{d\mathcal{L}(y,\hat{y})}{dx_j} = \frac{dy}{dx_j}\frac{d\mathcal{L}(y,\hat{y})}{dy} = \frac{d\sigma(x)}{dx_j} \cdot \frac{d\mathcal{L}(y,\hat{y})}{dy} = \frac{d\mathcal{L}(y,\hat{y})}{dy} \cdot \sigma'(x_j) \tag{2.8}$$

and from a certain weight, which is what we are looking for:

$$\frac{d\mathcal{L}(y,\hat{y})}{dw_{ij}} = \frac{dx_j}{dw_{ij}}\frac{d\mathcal{L}(y,\hat{y})}{dx_j} = y_i \frac{d\mathcal{L}(y,\hat{y})}{dx_j}. \tag{2.9}$$

Furthermore, the loss can be propagated through all layers of the network as:

$$\frac{d\mathcal{L}(y,\hat{y})}{dy_i} = \frac{dx_j}{dy_i}\frac{d\mathcal{L}(y,\hat{y})}{dx_j} = \sum_j w_{ij} \frac{d\mathcal{L}(y,\hat{y})}{dx_j}. \tag{2.10}$$

The change in bias can be computed the same way. Any differentiable function can be used as a loss function. While each framework comes with a set of predefined loss functions commonly used in the machine learning community, also problem-specific, custom loss function can be applied as well.

There are several implementations of gradient descent, which all base on principles described above. **Stochastic gradient descent** performs the back-propagation after each data point in the training dataset, resulting in a computationally intense, fine granular update of the network parameters also finding small local minima. On the other hand, this might lead to the network being kept in a local minimum and not able to pass the barrier around it. **Batch gradient descent** performs the back-propagation step once per epoch after processing the complete training dataset, resulting in a very coarse, global update of the weights and biases of the network. Combining the best of both worlds, **Minibatch gradient descent** performs the back-propagation step after processing a certain amount of data points, called minibatch. Minibatches save computing time and give the network a chance to find both local and global minima.

2.5.2 Accuracy as a performance metric

As stated above, the performance of a fully trained network is usually evaluated on the test dataset using problem-specific metrics. In addition to this, a general metric to measure the performance of an ANN called *accuracy* exists. This metric can be applied to both the training and validation dataset already during the training process. These are then called training accuracy and validation accuracy, respectively. Generally speaking, the accuracy is defined as the fraction of correct predictions over all predictions. Depending on the problem at hand, variations of this approach might be applied. Given e.g. a multiclass classification problem, only the class with the highest probability is compared to the designated output class, called categorical accuracy [93]. The accuracy can be used to spot a variety of problems, which are outlined below.

2.5.3 Under– and overfitting

Overfitting refers to a model starting to store the provided training dataset in its architecture, losing its ability to generalize the training dataset. It negatively impacts the performance of the model when applied to, for example, the validation or test dataset or when used in production. As can easily be seen, the chance of this happening rises with the networks' size relative to the size of the training dataset. Overfitting results in a (usually much) higher training accuracy compared to the validation accuracy. It can be mitigated by lowering the number of trainable parameters, the number of training epochs or by adding Dropout layers. Contrary to an overfitted network, an underfitted network did neither learn the training dataset nor was able to generalize it, resulting in a weak performance on all the training, validation, and test dataset. Standard solutions to this problem can be a more flexible or at least more massive network architecture or alternating the loss function. Before approaching the latter solution, a more detailed study should be made to understand if the inputs used for the ANN have a sufficient correlation to the outputs after all.

2.5.4 Vanishing gradient

The vanishing gradient problem can occur in larger ANN architectures applying a particular type of activation functions. The output value of activation functions limiting their output to a specific upper and or lower boundary, e.g., sigmoid$(x) \in [0, 1]$, applied over and over again through the network can lead to a so-called vanishing gradient. During back-propagation, the gradient of the loss function, depending on the derivative of the activation function, tends to get smaller with every layer.

This is shown in Figure 2.12 for the Sigmoid activation function, which is limited to 0.25. As a result, the first few layers of a network, learn slower than the later layers closer to the output. The solution to this problem is rather pragmatic and straightforward. Some activation functions are better suited than others to deal with the vanishing gradient problem, as their derivative is not limited to a specific value, e.g., the *relu* activation function. The best solution is to adapt if the problem occurs and switch activation functions or apply batch normalization [107].

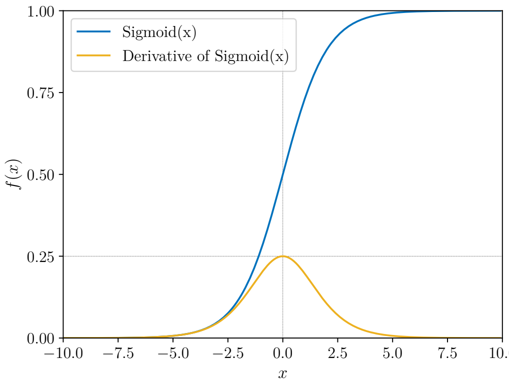


Figure 2.12: Sigmoid and its derivative.

There can also be a so-called exploding gradient problem, which is nearly the same but with a growing instead of vanishing gradient during back-propagation.

2.6 Optimizers

In the context of machine learning, more advanced derivations of the gradient descent algorithm are applied to increase the training performance further. These derivations, so-called optimizers, try to compensate for intrinsic weaknesses of gradient descent, which are discussed below in the context of different optimization strategies. This section is based on Ref. [145], which provides an overview of different optimizing techniques beyond the scope of this thesis.

2.6.1 Gradient descent with momentum

The concept of momentum [140] is an expansion of the classic gradient descent algorithm, introducing an additional weighted term when updating weights and biases. It is defined in analogy to a particle traveling through the network parameter space, accelerating from the gradient of the loss function.

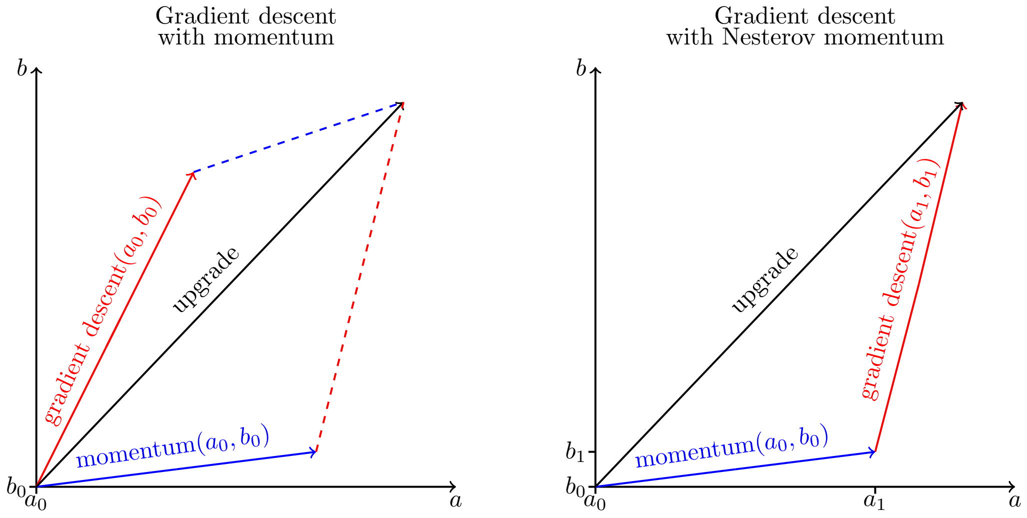


Figure 2.13: Different derivations of Gradient descent with momentum. Classic momentum (left) uses the vector sum of the gradient descent step and the momentum step evaluated at the current phase space position, while Nesterov momentum uses the momentum step to determine the point used for gradient descent calculations. Shown in an arbitrary two-variable phase space.

Due to this momentum, the direction of the minimization is more stable, preventing oscillations within the parameter space. Starting from Equation 2.4, the change of a certain parameter a (e.g., bias or weight) can be written as:

$$a_{n+1} = a_n - \alpha \cdot \nabla F(a_n). \tag{2.11}$$

Introducing the concept of momentum m_t at a certain gradient descent step t, an additional term is added to this equation. This term depends on the momentum m_{t-1} gained in the previous update step and the momentum parameter $\gamma \approx 0.9$. The value of the parameter to be updated is then calculated as a linear combination of classic gradient descent and this momentum:

$$m_n = \gamma m_{n-1} + \alpha \nabla F(a) \tag{2.12}$$

$$a_{n+1} = a_n - m_n. \tag{2.13}$$

On the other hand, so called Nesterov momentum or Nesterov Accelerated Gradient [131, 145] uses the momentum as an additional movement in the parameter space. Hereby the momentum parameter is added to the weight (and bias respectively) before calculating the updated values using gradient descent:

$$m_n = \gamma m_{n-1} + \alpha \nabla F(a - \gamma m_{n-1}) \tag{2.14}$$

$$a_{n+1} = a_n - m_n. \tag{2.15}$$

Both are valid ways of defining momentum in the context of machine learning and are visualized in Figure 2.13.

2.6.2 The Adaptive Moment Estimation optimizer

Adaptive **M**oment Estimation (Adam) [108] calculates an adaptive learning rate for each parameter of a network. An exponentially decaying average of past squared gradients v_t as well as an exponentially decaying average of past gradients m_t, which is similar to the momentum defined in the previous section, are utilized:

$$m_n = \beta_1 m_{n-1} + (1 - \beta_1) \cdot \nabla F(a_n) \tag{2.16}$$

$$v_n = \beta_2 v_{n-1} + (1 - \beta_2) \left(\nabla F(a_n)\right)^2 . \tag{2.17}$$

v_n and m_n are estimates of the mean (or first moment) and the uncentered variance (second moment) of the gradient, being the namesake of this method. These values are initialized as 0 and were found to have a bias towards this initial value for the first time steps. To counter this bias, the so called bias-corrected first and second moment estimates are computed as:

$$\hat{m}_n = \frac{m_n}{1 - \beta_1^n} \tag{2.18}$$

$$\hat{v}_n = \frac{b_n}{1 - \beta_2^n} \tag{2.19}$$

and then used to update the parameter value:

$$a_{n+1} = a_n - \frac{\eta \cdot \hat{m}_n}{\sqrt{\hat{v}_n} + \epsilon}. \tag{2.20}$$

According to the authors, default values of $\epsilon = 10^{-8}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ work well in practice, but have to be seen as additional tunable parameters in the context of machine learning.

Chapter 3

The Standard Model of Particle Physics

Contents

This chapter focusses on the fundamentals of particle physics, namely the SM and the particles and interactions included in it. The SM, shown in Figure 3.1, describes all known fundamental particles and their interactions. It consists of six quarks, six leptons as well as five gauge bosons and the Higgs boson. Additionally, all fundamental interactions, except gravity, are covered together with the gauge bosons being the force carriers of the different forces: the strong, weak, and electromagnetic force. This chapter gives a short introduction to the SM based on Ref. [138] and Ref. [160], which provide further reading.

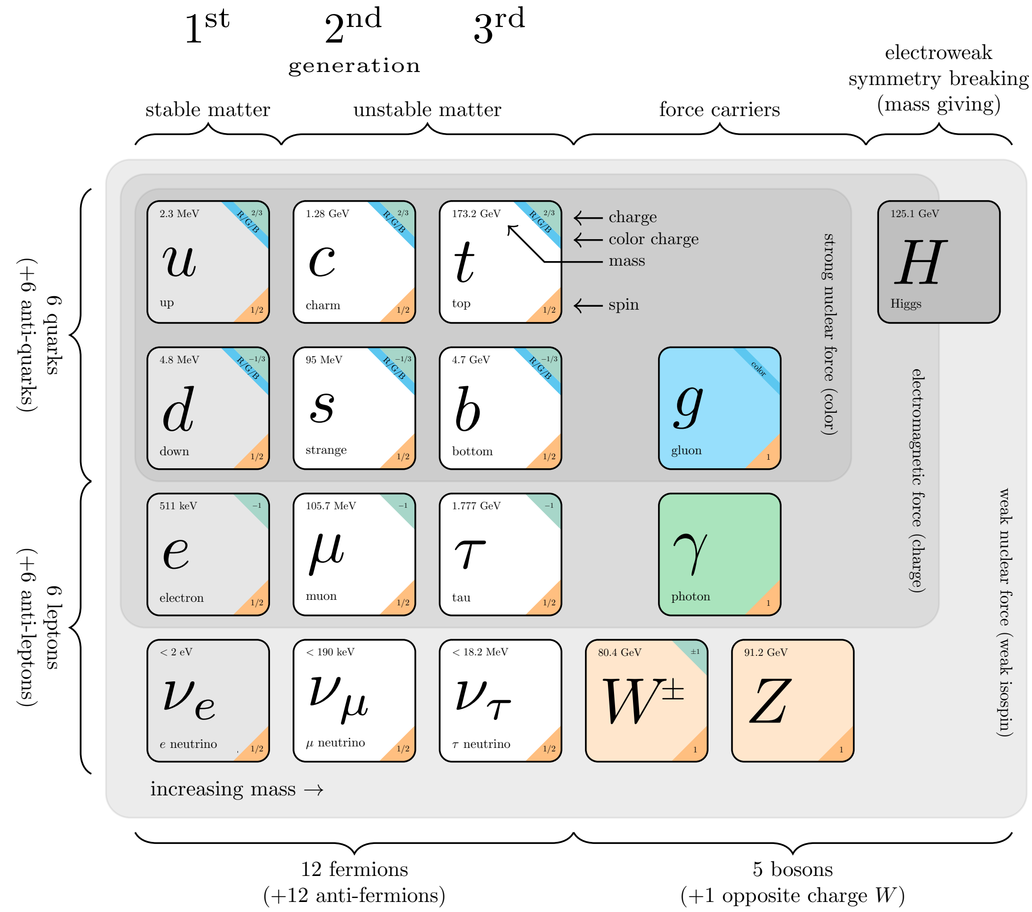


Figure 3.1: Standard Model of Particle Physics. All fundamental particles described by the SM, separated by particle type, are included. It includes the force carriers, the Higgs boson as well as 12 fermions. The fermions are split into quarks and leptons as well as generations. Based on Ref. [35].

Mathematically, the SM is a relativistic quantum field theory based on the non-Abelian symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. C refers to the quark color, L to left-handed fields, and Y to the hypercharge of the particles.

The theories described by the SM are:

- **Q**uantum **C**hromo**d**ynamics (QCD), describing the strong interaction.
- Electroweak interactions, describing both the weak interaction and **Q**uantum **E**lectro**d**ynamics (QED).
- Higgs mechanism

All subatomic particles and processes observed thus far, including decay and scattering processes, can be described by the SM. The fundamental particles can be split into four groups, two groups of spin $1/2$ fermions, called quarks and leptons, one group of spin 1 bosons, the gauge bosons, and the spin 0 Higgs boson.

3.1 Quarks and leptons of the Standard Model

The SM predicts twelve matter-particles and their corresponding antiparticles. These spin $1/2$ matter-particles are fermions and respect the Pauli exclusion principle, following the spin-statistics theorem. These matter-particles are grouped into quarks and leptons, each containing six particles. Inside each group, pairs of particles form a generation, grouping particles with similar characteristics as shown in Figure 3.1. Each member of a generation has greater mass than the corresponding particles of lower generations. Additionally, the first-generation charged particles do not decay.

3.1.1 Quarks

Quarks carry a color charge and, as a result, interact via the strong interaction. This includes interactions with other quarks. Quarks also carry an electric charge and, in the case of left-handed particles, weak isospin. Three up-type quarks, namely *up*, *charm* and *top* quark, carry an electric charge of $2/3$, while the down-type *down*, *strange* and *bottom* quarks carry an electric charge of $-1/3$.

Due to color confinement, quarks are strongly bound to other quarks via gluons. As a result, quarks can not be observed as free particles but form color-neutral particles called hadrons. These hadrons can either be baryons, containing three quarks, or mesons, containing a quark-antiquark pair. Quarks, generally all color-charged particles, can never be observed in nature, as color confinement states that energy put into the separation of quarks ultimately leads to the creation of quark-antiquark pairs.

3.1.2 Leptons

Fermions carrying no color charge are called leptons. Electron, muon, and tau carry an electric charge of -1 and, in the case of left-handed particles, a weak isospin of $-1/2$. Each of these leptons forms a generation with a corresponding lepton neutrino, e.g. muon and muon-neutrino.

The neutrinos carry no electrical charge and, again only if left-handed, a weak isospin of $1/2$. Neutrinos, as a result, only interact via the weak interaction, while charged leptons also interact electromagnetically. Electrons, as well as all neutrinos, do not decay, while muons and taus can decay.

3.2 Forces and gauge bosons in the Standard Model

Interactions between particles, namely strong, weak, and electromagnetic interactions, are described by the exchange of force-mediating particles in the SM. These spin 1 force carriers follow the Bose-Einstein statistics and are called gauge bosons.

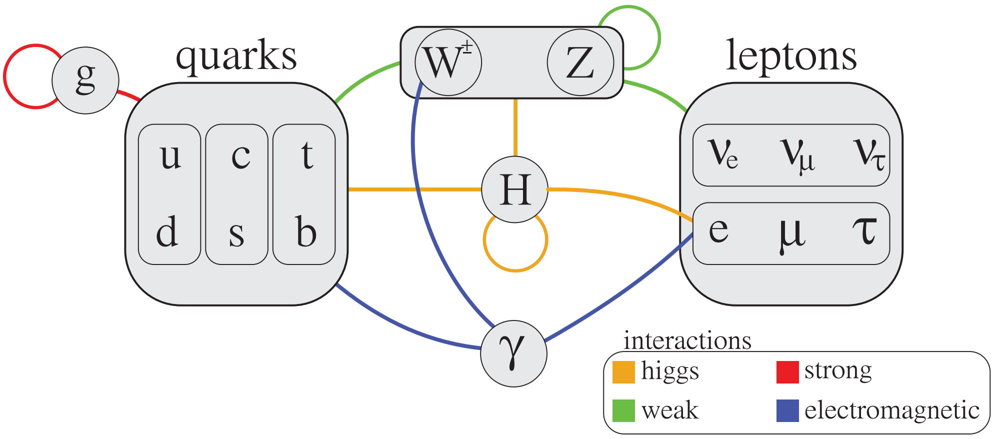


Figure 3.2: Summary of the possible interactions of the different fundamental particles of the SM. Differently colored lines mark the different kinds of interactions. These lines connect the gauge bosons corresponding to its interaction with the particles it is acting on. Based on Ref. [75].

Table 3.1 summarizes all interactions, the property of the particle they act on, the mediating particles, and its relative strength. Additionally, Figure 3.2 illustrates which particles of the SM are affected by which interaction. The following part provides a brief introduction and overview of the SM forces. Ref. [133] provides further reading.

3.2.1 Quantum electrodynamics

Electromagnetism describes the interaction of electrically charged particles by the exchange of photons. The photon γ, itself carrying no charge, is the massless gauge boson of electromagnetism and mediates the force between affected particles. It is not limited to a certain range.

Table 3.1: The four fundamental forces, taken from Ref. [94]. The strong force is split into fundamental and residual contributions, as on a bigger scale, residual contributions of the quarks in protons and neutrons bind nuclei together.

Property	Gravitation	Electroweak		Strong	
		Weak	Electromagnetic	Fundamental	Residual
Acts on:	Mass -Energy	Flavor	Electric charge	Color charge	Atomic nuclei
Particles affected:	All	Quarks, Leptons	Electrically charged	Quarks, Gluons	Hadrons
Particles mediating:	-	W^+,W^-,Z^0	γ(photon)	Gluons	Mesons
Strength at the scale of quarks:	10^{-41}(predicted)	10^{-4}	1	60	-
Strength at the scale of protons/neutrons:	10^{-36}(predicted)	10^{-7}	1	-	20

3.2.2 Weak interaction

The weak interaction describes the interactions between leptons and quarks, including quark conversions within the same generation and lepton decays switching generations. Figure 3.3 visualizes examples of these processes. This force has a range of 10^{-16} m and is mediated by the massive $W^{\pm}$ and Z bosons, with the Z bosons mass $M_Z = 91.2\,\mathrm{GeV}$ being higher than the W boson mass $M_W = 80.4\,\mathrm{GeV}$.

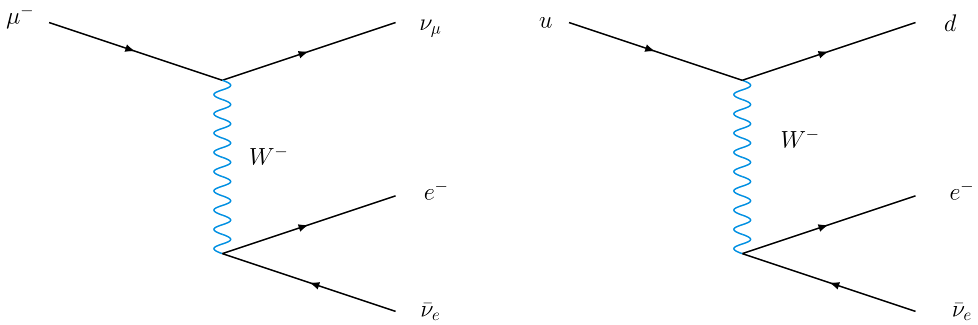


Figure 3.3: Muon decay into an electron (left) and conversion of up into a down quark (right).

In the context of this theory, quarks and leptons are divided into left-handed particles, only participating in the charged-current interactions involving the $W^{\pm}$ bosons, and right-handed particles, only participating in the neutral-current interactions via Z bosons. Thereby, left-handed particles form doublets, while right-handed particles form singlets. These can be described using projection operators as:

$$\Psi_{L,R} = \frac{1}{2}(1 \pm \gamma^5)\Psi, \tag{3.1}$$

with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ the fifth Dirac matrix.

Furthermore, weak interactions involving the $W^{\pm}$ bosons only act on left-handed particles and right-handed antiparticles. These bosons, as they carry an electric charge of ± 1, can additionally interact electromagnetically. The weak interaction is the only interaction able to trigger a particle to switch its type.

3.2.3 Electroweak interaction

Electromagnetism and weak interaction unify at the electroweak scale (or Fermi scale) at $\nu \approx 245\,\mathrm{GeV}$ and can both be described by a Yang-Mills gauge field theory with the symmetry group $U(1) \times SU(2)$. The generators of these groups are called weak hypercharge Y_W and weak isospin T, respectively. The invariant Lagrangian describing the electroweak sector is defined as:

$$\mathcal{L}_{\mathrm{EW}} = \sum_{\psi} \bar{\psi}\gamma^{\mu} \left(i\partial_{\mu} - g'\frac{1}{2}Y_{\psi}B_{\mu} - g\vec{\tau}_a\vec{W}_{\mu} \right) \psi - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \qquad (3.2)$$

with B_{μ} being the $U(1)$ gauge field, $\vec{W}_{\mu}$ the $SU(2)$ gauge field, Y_{ψ} the hypercharge of fermion ψ and τ_a the generator of the $SU(2)$ representation of ψ. For doublets, τ_a refers to the Paul matrices. Furthermore, g' and g are coupling constants and $B^{\mu\nu}$ and $W^{a\mu\nu}$ two field strength tensors. Photons and Z bosons are linear combinations of B and W_3 bosons, which can be written using the weak mixing angle Θ_W as:

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\Theta_W & \sin\Theta_W \\ -\sin\Theta_W & \cos\Theta_W \end{pmatrix} \cdot \begin{pmatrix} B \\ W_3 \end{pmatrix}, \qquad (3.3)$$

The $W^{\pm}$ bosons is defined by:

$$W^{\pm} = \frac{1}{\sqrt{2}} \left(W_1 \mp iW_2 \right). \qquad (3.4)$$

The bosons have to be massless to preserve the electroweak gauge symmetry but are observed to have non-zero masses. The bosons acquire their mass by a process called *spontaneous symmetry breaking* via the Higgs mechanism. In the context of this theory, mass terms for the fermions can be generated by the Higgs mechanism via so-called Yukawa couplings between fermions and the Higgs field.

3.2.4 Strong interaction

The strong interaction, as described by QCD, refers to the force between color-charged particles, namely quarks and gluons, and has an effective range of $10^{-15}\,\mathrm{m}$, similar to the size of hadrons. It is a non-Abelian gauge theory based on a $SU(3)$ gauge symmetry. 8 gluons, massless gauge bosons, mediate the strong force.

Each gluon carries one of eight possible color-anticolor pair combinations and therefore can, due to net color charge of the pair, interact with each other.

Two effects are important in the context of strong interactions: *Confinement* and *Hadronization*. Generally speaking, processes described by the SM can result in quarks and gluons in the final-state. Both quarks and gluons can not exist as free particles but form colorless bound states directly after their production under conditions below the Hagedorn temperature $\mathrm{T_H} \approx 1 \cdot 10^{12}\,\mathrm{K}$ [26].

This phenomenon is called color confinement and is true for all color charged particles. Hadronization is a consequence of color confinement and refers to the formation of hadrons. This is achieved by singular quark or gluons immediately forming combinations with quarks and antiquarks or more gluons, for example created by vacuum fluctuations to form baryons or mesons.

Chapter 4

Theoretical description of particle collisions at the LHC

Contents

4.1 Particle collisions at the LHC

Proton–proton collisions as produced at the LHC, which is discussed in Chapter 5, involve complex, composite particles consisting of three valence quarks, gluons, and quark-antiquark pairs. Data created in such collisions, especially through the Drell-Yan process, are used in this thesis. Hence the theoretical description of the underlying processes is given here. This section is based in chapters 10.9.1 and 10.9.2 of Ref. [160], while sections 4.2 and 4.3 are based on Ref. [42].

Protons do not collide as one object, but each parton of the proton, carrying a fraction of the proton momentum, can but does not have to participate in the scattering processes. The kinematics of such a proton–proton collision can be described in its center-of-mass frame. Assuming two colliding protons A and B with four-momenta P_A and P_B while ignoring the proton mass, which plays only a neglectable role at high energies, the squared total energy s in the center-of-mass frame is given by:

$$s = (P_A + P_B)^2 \approx 2P_A P_B. \tag{4.1}$$

Looking at the interacting partons a and b of proton A and B, respecively, with momenta p_a and p_b, the square of the total energy is defined as:

$$\begin{aligned} \hat{s} &= (p_a + p_b)^2 \\ &= (x_a P_A + x_b P_B)^2 \\ &\approx 2x_a x_b P_A P_B \\ &= x_a x_b s. \end{aligned} \tag{4.2}$$

Hereby x denotes the *Bjorken scale*, the fraction of the quark momentum with respect to the proton momentum. In order to be able to produce a certain particle in this collision, e.g. a Z boson, the centre-of-mass energy $\sqrt{s}$ has to be equal to the invariant mass of this particle M. Using this, one can rewrite the parton momenta to

$$\begin{aligned} p_{a,b} &= x_{a,b} \frac{\sqrt{s}}{2} (1, 0, 0, \pm 1)^{\mathrm{T}} \\ &= x_{a,b} \frac{M}{2} (1, 0, 0, \pm 1)^{\mathrm{T}}. \end{aligned} \tag{4.3}$$

The rapidity y of the produced particle, a measure of the boost of a particle along the beam axis, can be defined using its energy E and momentum along the z-axis p_z as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{4.4}$$

Assuming the parton mass to be 0, this can be rewritten using Equation 4.3 to:

$$y = \frac{1}{2} \ln \left(\frac{x_a}{x_b} \right). \tag{4.5}$$

Using the rapidity, Equation 4.3 can be rewritten as:

$$x_{a,b} = \frac{M}{\sqrt{s}} e^{\pm y}. \tag{4.6}$$

Thus, different invariant masses M and rapidties y of the final-state particle probe different x. This relationship is visualized in Figure 4.1 for a center-of-mass energy of 13 TeV. In addition, sensitivity regions of other experiments are marked.

4.2 Parton distribution functions

The kinematics of a proton–proton collisions depend on the interacting partons, as described above. So-called **P**arton **D**istribution **F**unctions (PDF) describe the structure of a proton and hence quantify the probability density of finding a parton (quark or gluon) inside the proton. These PDFs depend on the fraction of the proton momentum, described by the Bjorken scale x, the parton carries as well as the energy scale Q^2.

As PDFs can't be calculated perturbatively, they are extracted from data from different experimental measurements using global fits. The experimental data used for these fits include, among many other sources, data from HERA (**H**adron-**E**lectron **R**ing **A**ccelerator), Tevatron, and the LHC. PDFs are generically parameterized with a starting scale $Q_0^2 = 1\,\text{GeV}^2$ as [42]:

$$xf(x, Q_0^2) = Ax^B(1-x)^C P(x, D). \tag{4.7}$$

Hereby the normalization parameter A is constrained by the quark number sum-rules [98] and momentum sum-rule, while B and C are free fit parameters and $P(x, D)$ are polynomials in x. Using the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations [8, 83, 104], PDFs can be evolved to any other energy scale $Q^2 > Q_0^2$:

$$\frac{\partial f(x, \mu_F^2)}{\partial \log \mu_F^2} = \frac{\alpha_S(\mu_R^2)}{2\pi} \sum_{b=q,\hat{q},g} \int_x^1 \frac{dz}{z} P_{ab}(\frac{x}{z}, \mu_F^2) f_b(z, \mu_F^2), \tag{4.8}$$

where μ_F is the factorization scale, which separates the long- and short-distance regimes. $\alpha_S(\mu_R^2)$ is the running strong coupling constant depending on the renormalization scale μ_R for the QCD running coupling and $P_{ab}(\frac{x}{z}, \mu_F^2)$ are so-called splitting functions. The latter describe the probability of the parton b to emit a parton a carrying a fraction $\frac{x}{z}$ of the momentum of the parton b. These functions can occur in four different forms:

- P_{qq} – a quark radiating a quark,
- P_{qg} – a gluon radiating a quark,
- P_{gq} – a quark radiating a gluon,
- P_{gg} – a gluon radiating a gluon.

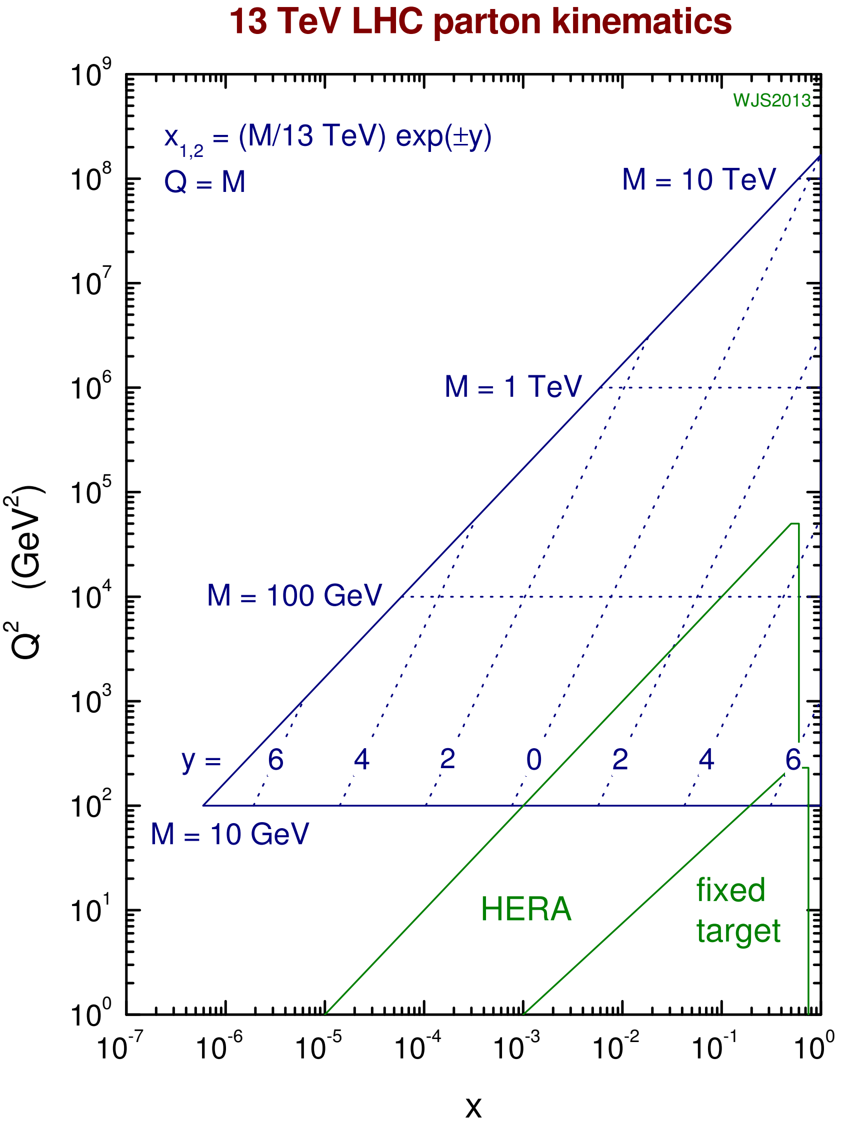


Figure 4.1: Graphical representation of the relationship between a partons x and Q^2 and a collisions final-state described by its mass M and rapidity y at the LHC at a center-of-mass energy of 13 TeV. This Figure is taken from Ref. [158].

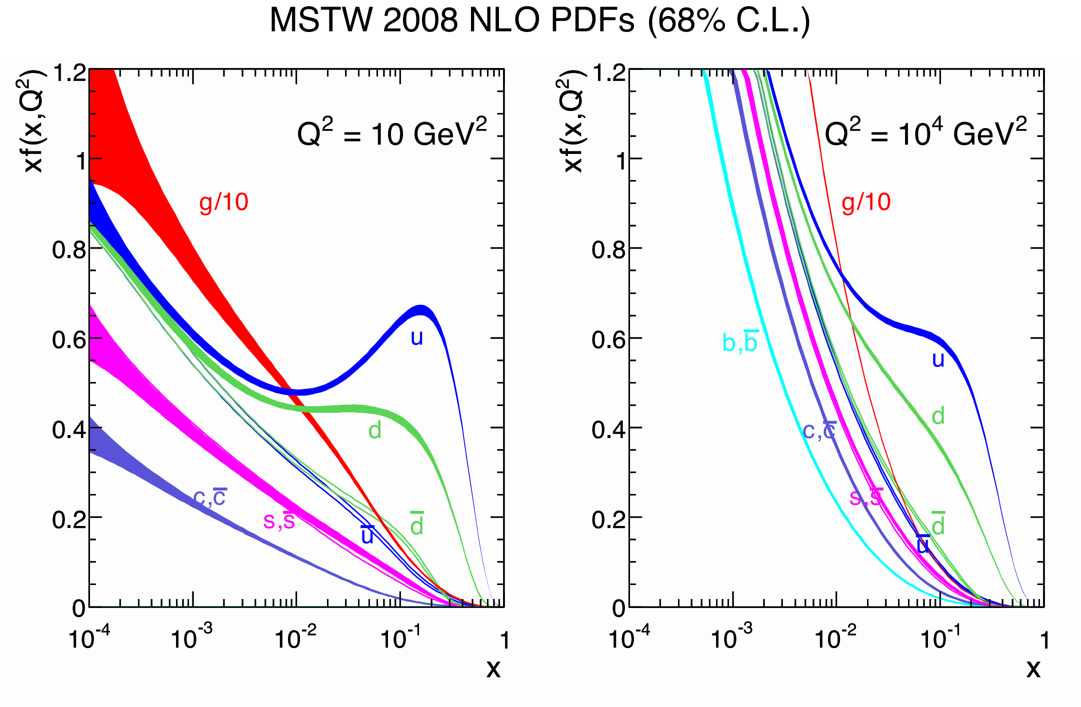


Figure 4.2: *MSTW2008NNLO* PDFs as a function of Bjorken-x for quarks and gluons. Uncertainties are indicated as an uncertainty band. The gluon PDF is divided by a 10. The PDFs are shown for a scale of $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right). This Figure is taken from Ref. [121].

Figure 4.2 shows a PDF calculated up to **N**ext-to-**N**ext-to-**L**eading **O**rder (NNLO) for two energy scales $Q^2 = 10\,\mathrm{GeV}^2$ and $Q^2 = 10^4\,\mathrm{GeV}^2$ together with its corresponding uncertainties [121]. Overall, the distributions decrease with increasing x. The up quark and down quark distributions show a local maximum at $\approx 1/3$, which refers to the valence quarks of the proton. These peaks are less prominent for higher Q^2, while sea quark contributions are higher for low x.

PDFs are used to calculate cross-section predictions of various processes. These predictions are widely used within the experimental communities at the LHC to simulate collision events. Such simulations are then combined with suitable detector simulations and used to differentiate between background and signal contributions within measured data as well as to correct for technical effects limiting the precision of a measurement, such as handling pileup. Hence, a good knowledge of the PDFs is essential, but the precision of such predictions usually suffer from significant PDF-related uncertainties, which can only be reduced by adding new precise data, covering different x ranges, to the QCD fits.

4.3 Partonic cross-section

Both hard- and soft-scattering processes appear in hadronic collisions. Hereby, the term hard-scattering process refers to processes with high momentum transfer Q^2, which can be calculated using pertubative QCD, while the soft-scattering processes correspond to lower Q^2 and are dominated by non-pertubative QCD effects. Using the QCD factorization theorem [74], the total cross-section of a specific hard-scattering process in a proton–proton collision, producing a final-state X, can be calculated by:

$$\sigma(AB \to X) = \sum_{a,b} \int dx_a dx_b \; f_{a/A}(x_a, \mu_F^2) \; f_{b/B}(x_b, \mu_F^2) \; \hat{\sigma}_{ab\to X}, \tag{4.9}$$

summing over all partons a of hadron A and partons b of hadron B contributing to the process. Hereby $f_{a/A}$ and $f_{b/B}$ are PDFs for the partons a and b, and $\hat{\sigma}_{ab\to X}$ the partonic cross section describing interactions between two partons.

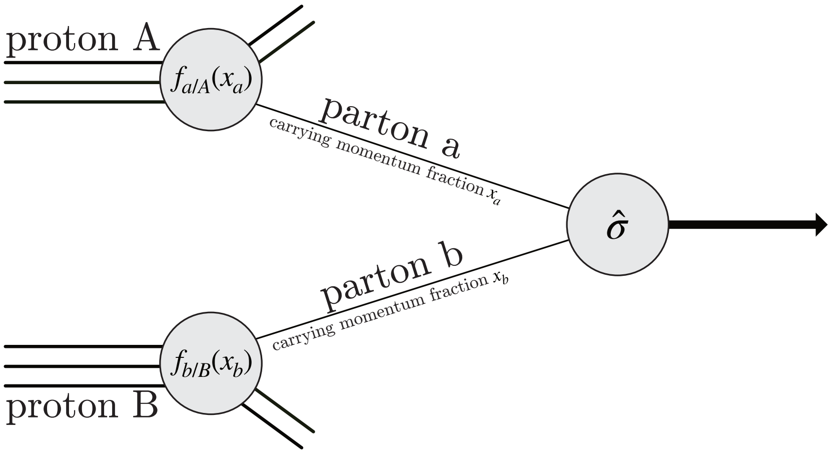


Figure 4.3: Schematic view of a hard scattering process of cross-section $\hat{\sigma}$. The partons a and b involved in this process carry a momentum fraction $x_{a,b}$ of its corresponding proton. The probability to find a certain parton with a certain x is given by the corresponding PDF. These are denoted $f_{a/A}(x_a)$ for parton a coming from proton A and $f_{b/B}(x_b)$ for parton b coming from proton B. Based on Figure 1 of Ref. [42].

The latter can be calculated perturbatively using QCD and expressed as a power series expansion of the α_S coupling constant, with **L**eading **O**rder (LO), NLO and NNLO and even higher-order contributions as:

$$\hat{\sigma}_{ab\to X} = \underbrace{\hat{\sigma}_0}_{\text{LO}} + \underbrace{\alpha_S(\mu_R^2)\hat{\sigma}_1}_{\text{NLO}} + \underbrace{\alpha_S^2(\mu_R^2)\hat{\sigma}_2}_{\text{NNLO}} + \ldots \tag{4.10}$$

Hereby μ_R is the renormalization scale of the QCD running coupling. Figure 4.3 schematically visualizes a hard scatter process with cross-section $\hat{\sigma}$. The probability of finding a parton with the two colliding protons are given by the corresponding PDF.

4.4 The Drell-Yan process

Z bosons, first observed at the UA1 and UA2 experiments in 1983 [97, 135], can be created in proton–proton collisions at the LHC. Table 4.1 lists the branching ratios, which can be understood as a decaying probability, of a Z boson into a pair of leptons as well as a pair of hadrons.

Table 4.1: Branching ratios of Z boson decays at the LHC. Taken from Ref. [134].

Process	branching ratio [%]
$Z \to q\bar{q}$	69.91 ± 0.06
$Z \to \nu\bar{\nu}$	20.00 ± 0.06
$Z \to \mu^-\mu^+$	3.366 ± 0.007
$Z \to \tau^-\tau^+$	3.370 ± 0.008
$Z \to e^-e^+$	3.363 ± 0.004

The branching ratios are similar for all three non-neutrino lepton decays, meaning each type of charged lepton is created equally often in decay processes. This phenomenon is called *lepton universality.*

Z bosons at the LHC are dominantly produced through the Drell–Yan process [84]. In this process one quark and one antiquark from the incoming protons annihilate and form a photon γ or Z boson, which then instantaneously ($\tau_Z \approx 10^{-25}\,\text{sec}$) decays into a pair of leptons. This can be described as:

$$q\bar{q} \to Z/\gamma^* \to \ell^+\ell^-. \tag{4.11}$$

Feynman diagrams of the Drell-Yan process in LO, **N**ext-to-**L**eading **O**rder (NLO) and NNLO are shown in Figure 4.4, each order adding a radiated gluon or quark creating **I**nitial-**S**tate **R**adiation (ISR). The leading order cross-section of the Drell-Yan process [160] can be calculated by:

$$\hat{\sigma}(q\bar{q} \to Z/\gamma^* \to l^+l^-) = \frac{4\pi\alpha^2}{3\hat{s}} \frac{1}{N_C} Q_q^2. \tag{4.12}$$

Hereby Q_q denotes the charge of the quarks and $N_C = 3$, the number of same color combinations out of the nine possible color-combinations. Table 4.2 lists experimentally determined fiducial cross-sections of the Drell-Yan process for different center-of-mass energies $\sqrt{s}$. The results are taken from Ref. [56].

Table 4.2: Fiducial cross-section of the leptonic Drell-Yan process for both muon and electron pairs for different center-of-mass energies. The fiducial volume of the 13 TeV and 7 TeV measurement is defined by the lepton $|\eta_\ell| < 2.5$ and the reconstructed Z invariant mass $m_{\ell\ell}$ 66 and 116 GeV. For the 8 TeV result, the lepon cut changes to $|\eta_\ell| < 2.4$ and an additional cut on the rapidity of the Z boson $|y_{\ell\ell}| < 2.4$ is applied. The content is taken from Ref. [56]. The kinematic cuts are explained in Section 6.2.

$\sqrt{s}$ [TeV]	13	8	7
$\sigma^{fid}_{Z\rightarrow ee}$ value±stat±sys [pb]	$778.3 \pm 0.7 \pm 17.7$	$507.0 \pm 0.2 \pm 11.0$	$451.2 \pm 0.5 \pm 8.7$
$\sigma^{fid}_{Z\rightarrow \mu\mu}$ value±stat±sys [pb]	$774.4 \pm 0.6 \pm 18.2$	$504.7 \pm 0.2 \pm 10.8$	$450.0 \pm 0.3 \pm 8.8$

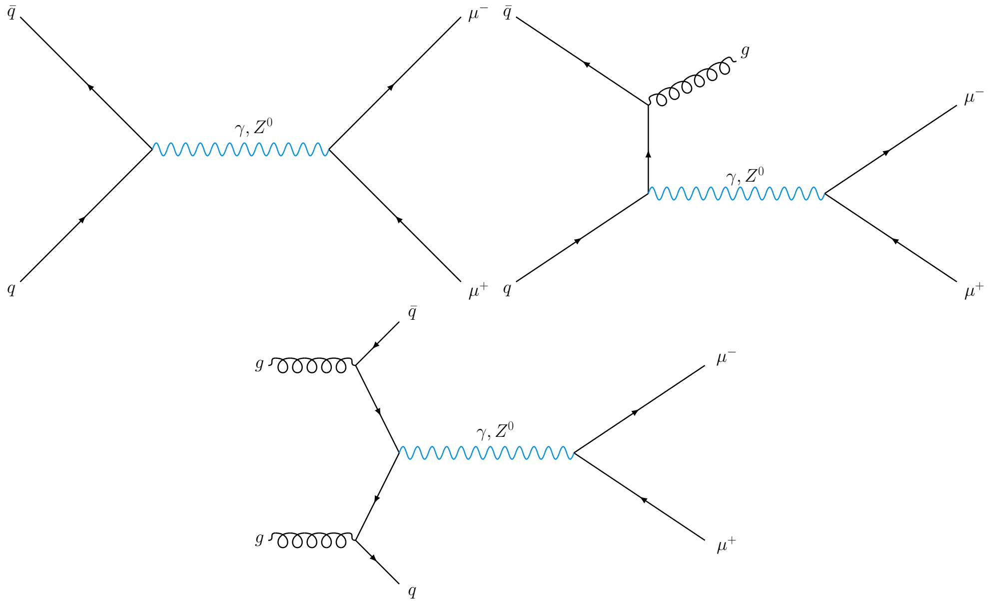


Figure 4.4: Leading order (upper left) as well as examples of NLO (upper right) and NNLO (lower) Feynman diagrams of the Drell-Yan process, describing two quarks annihilating to form a charge-neutral Z boson. This boson further decays into a pair of muons. The gluon radiated from the quark in the NLO diagram as well as the two remaining quarks in the NNLO diagram hadronize and form jets. This is referred to as ISR. Radiation can also happen in the final-state from the muons, which then radiate photons. This is then called **F**inal-**S**tate **R**adiation (FSR).

4.5 Initial- and final-state radiation

In leading-order, Z bosons created in proton–proton collisions have no momentum in the transverse plane as the protons collide head-on. In higher-order, the interacting quarks can gain transverse momentum through the radiation of gluons or photons. These radiations are collectively referred to as ISR as it concerns the quarks involved in the initial interaction. Figure 4.4 shows examples of such processes. FSR, on the other hand, refers to the radiation of photons or gluons from final-state particles created from the decaying Z boson. In the case of Drell-Yan lepton pairs, the FSR does not include the radiation of gluons, as leptons carry no color charge.

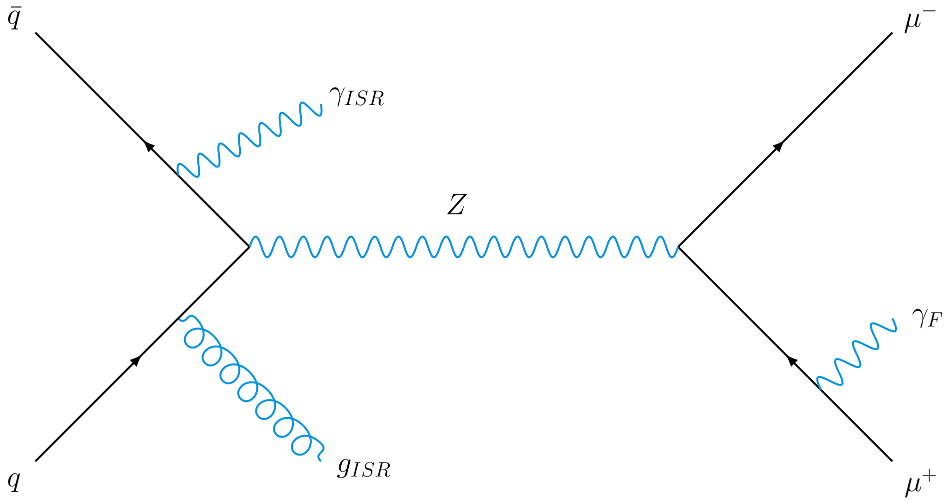


Figure 4.5: Visualization of both ISR and FSR using a Drell-Yan Z boson decay into a pair of muons. Gluon and photon are radiated from the initial quarks, creating ISR, while one of the final-state muons radiated a photon as final-state radation.

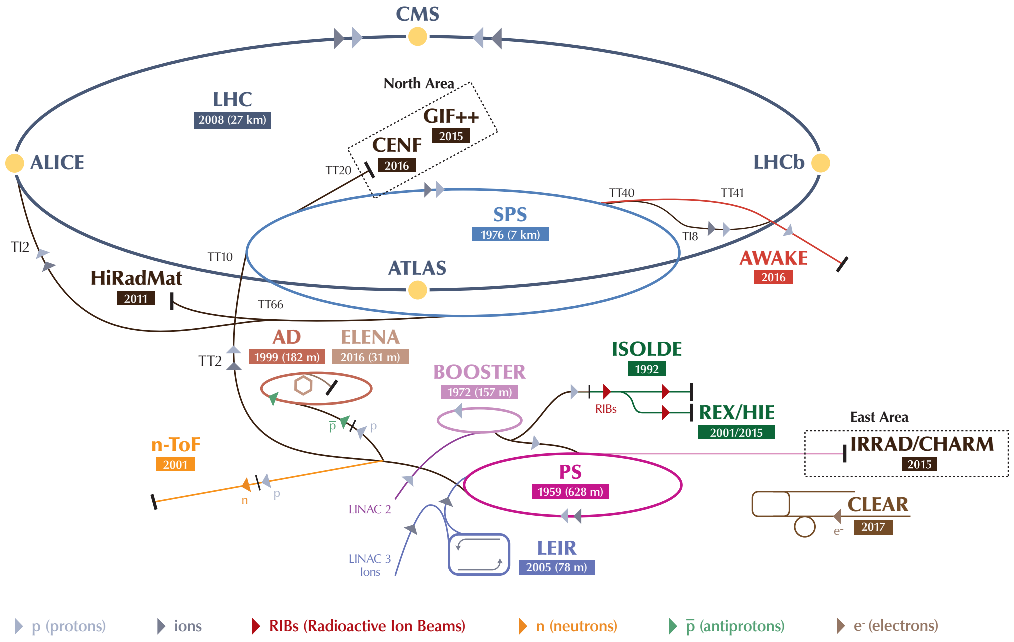
Figure 4.5 visualizes both initial-state and FSR using a Z boson decaying into a pair of muons as an example process. Both a gluon and a photon are radiated as ISR, while one of the final-state muons emits a photon as FSR.

Chapter 5

Particle physics at CERN

Contents

All topics covered in this thesis are related to the ATLAS [62] detector, an all-purpose particle detector at the LHC [92] built at European Organization for Nuclear Research (CERN) near Geneva, Switzerland. This chapter introduces CERN, the LHC and in particular the ATLAS detector and its subsystems.



LHC - Large Hadron Collider // **SPS** - Super Proton Synchrotron // **PS** - Proton Synchrotron // **AD** - Antiproton Decelerator // **CLEAR** - CERN Linear Electron Accelerator for Research // **AWAKE** - Advanced WAKefield Experiment // **ISOLDE** - Isotope Separator OnLine // **REX/HIE** - Radioactive EXperiment/High Intensity and Energy ISOLDE // **LEIR** - Low Energy Ion Ring // **LINAC** - LINear ACcelerator // **n-ToF** - Neutrons Time Of Flight // **HiRadMat** - High-Radiation to Materials // **CHARM** - Cern High energy AcceleRator Mixed field facility // **IRRAD** - proton IRRADiation facility // **GIF++** - Gamma Irradiation Facility // **CENF** - CErn Neutrino platForm

Figure 5.1: The CERN accelerator complex together with the experiments. Taken from Ref. [127].

5.1 CERN

CERN is a research organization with 22 member states at the time of this thesis submission. It operates the largest particle physics laboratory in the world. Based in Geneva, Switzerland, CERN provides particle accelerators for high-energy physics research and the necessary infrastructure to maintain as well as improve them. It operates several different accelerator complexes, each focusing on different areas of study. One of these machines is the LHC, a terascale hadron collider facility together with its preaccelerators. The next section introduces this machine complex, while Figure 5.1 shows a schematic of it.

5.2 The Large Hadron Collider

The LHC, developed at CERN, is the largest energy particle accelerator built thus far and able to collide both protons as well as heavy ions. Its current primary objective is to investigate electroweak symmetry breaking, measure the properties of the Higgs boson, and search for physics beyond the Standard Model. It consists of 1232 dipole magnets to keep the particles beams on track as well as 392 quadrupole magnets to keep the beams focused. In addition to this higher pole order magnets are used to preserve the beam geometry.

Located between 50 m to 175 m underground, it is built in a 26.7 km long tunnel formerly used for the Large Electron Proton Collider. Over the years of operation, the LHC increased its center of mass energy from 7 TeV (3.5 TeV per beam) in 2011 till up to 13 TeV from 2015 to 2018, when the machine was shut down for further upgrades.

Table 5.1: LHC Run–II parameters during different years of operation [157].

Parameter	2015	2016	2017	2018
Beam Energy [TeV]	6.5	6.5	6.5	6.5
Bunches per beam	2244	2220	1916	2556
Protons per bunch [10^{11}]	1.1	1.1	1.2	1.2
Peak luminosity [$10^{34}\text{cm}^{-2}\,\text{s}^{-1}$]	0.51	1.44	1.9	2.1
Average interactions per bunch crossing $\langle\mu\rangle$	13.4	25.1	37.8	36.1
LHC integrated luminosity [fb^{-1}]	4.2	39.7	50.2	66

At the LHC, two particle-beams consisting of several bunches with several thousand particles are running in the opposite direction through the tunnel. These beams intersect at four distinct interaction points. There, particle bunches collide every 25 ns, creating observable events for the four main experiments at the LHC. These four experiments are the two multipurpose experiments ATLAS and CMS (**C**ompact **M**uon **S**olenoid) [71], the **LHC** **B**eauty experiment (LHCb) [72], focusing on b physics, and ALICE (**A** **L**HC **I**on **C**ollider **E**xperiment) [70], a heavy-ion experiment.

As the LHC is a storage ring, particles entering the LHC have to have a minimum kinetic energy of 450 GeV to be able to be stored within the LHC. The LHC is fed by a series of accelerator complexes, preparing the particles to enter the LHC. For proton runs, the particles are first accelerated to 50 MeV in the LINAC2 (**Lin**ear **AC**celerator **2**) linear accelerator. From there, the protons are passed to the PSB (**P**roton **S**ynchrotron **B**ooster) and, after being accelerated to 1.4 GeV, enter the PS (**P**roton **S**ynchrotron). Reaching 26 GeV, the protons are fed into the last stage of preacceleration, the SPS (**S**uper **P**roton **S**ynchrotron). In there, particles are accelerated up to the LHC's injection kinetic energy and are fed into the LHC.

Once all particle bunches entered the LHC, these are further accelerated and, after reaching the designated final kinetic energy, collide at the interaction points.

The so-called instantaneous luminosity is a measure of performance or intensity of particle beams and describes the number of particle interactions per unit time per unit area. It only depends on the beam characteristics and is defined as

$$\mathscr{L} = \frac{N_b^2\, n_b\, f_r\, \gamma_r}{4\pi\, \epsilon_n\, \beta^\star} \cdot \left(1 + \left(\frac{\Theta_c\, \sigma_Z}{2\, \sigma^\star}\right)\right)^{-\frac{1}{2}}, \tag{5.1}$$

with

- the particles per bunch N_b,
- the number of bunches n_b,
- the revolution frequency f_r,
- the relativistic gamma factor γ_r,
- the transverse normalized beam emittance ϵ_n,
- the beam squeezability at the interaction point $\beta^\star$,
- the crossing angle between the beams Θ_c,
- the **R**oot **M**ean **S**quare (RMS) of the bunch length σ_Z,
- the RMS of the beam size at the interaction point $\sigma^\star$.

An overview of the LHC Run–II operational parameters, which is the scope of this thesis, is additionally given in Table 5.1. These parameters all have to be experimentally optimized for a given task by technicians to run the LHC at its peak performance.

A measure of the overall particle interactions per area produced by the accelerator is the so-called integrated luminosity, defined as:

$$L = \int \mathscr{L} dt. \tag{5.2}$$

Figure 5.2 exemplary shows the luminosity delivered in Run–II (2015-2018). From the luminosity, the expected number of events for a process with a given cross-section $\sigma_{process}$ can be calculated by:

$$N_{\text{process}} = \sigma_{\text{process}} \cdot L \cdot A \cdot \epsilon, \tag{5.3}$$

with A being the geometrical acceptance and ϵ the detector efficiency. Data taken with the ATLAS detector from proton-proton collisions at the LHC is used in the context of this thesis, as outlined in Chapter 9. This data, taken during LHC Run–II in 2015 and 2016, features a centre-of-mass energy of $\sqrt{S} = 13\,\text{TeV}$ and a bunch crossing every 25 nss. Further run conditions are summarized in Table 5.1 for all years of Run–II, 2015 to 2018.

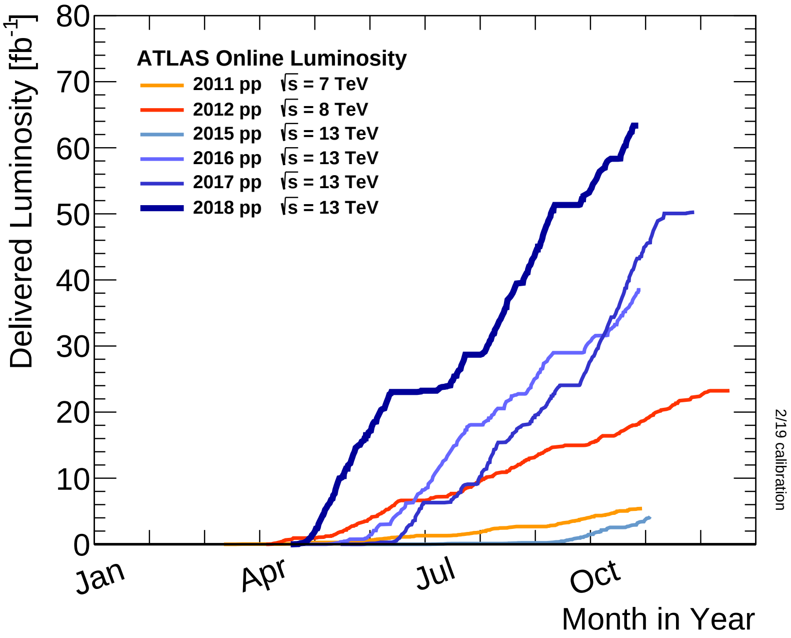


Figure 5.2: The integrated (delivered) luminosity for the ATLAS detector in the years 2015 to 2018 are shown. This plot is an ATLAS public result taken from Ref. [20].

5.3 Pileup activity at the LHC

When colliding particles at the LHC, usually more than one interaction per bunch crossing appears. The collision of interest, the so-called hard-scatter, is defined as the collision with the highest $\sum p_T^2$ over all particles created. Additional activity in the detector besides the hard-scatter is collectively referred to as pileup.

Figure 5.3 visualizes the rising average number of interactions over the years of Run–II. Pileup activity is a significant challenge for the LHC experiments as it complicates particle detection and reconstruction. Pileup is not limited to further collisions during a particular bunch crossing but can include contributions from earlier or later collisions as well as particle beam and background effects. Five categories of pileup exist.

Additional proton-proton collisions occurring during the same proton-proton bunch-crossing are referred to as in-time pileup and present a significant background for all physics objects. This kind of collision can happen throughout a detector volume and can typically be distinguished from the hard scatter. The rising number of particles to be reconstructed and particles being too close to each other to be identified are the main problems for this class of pileup.

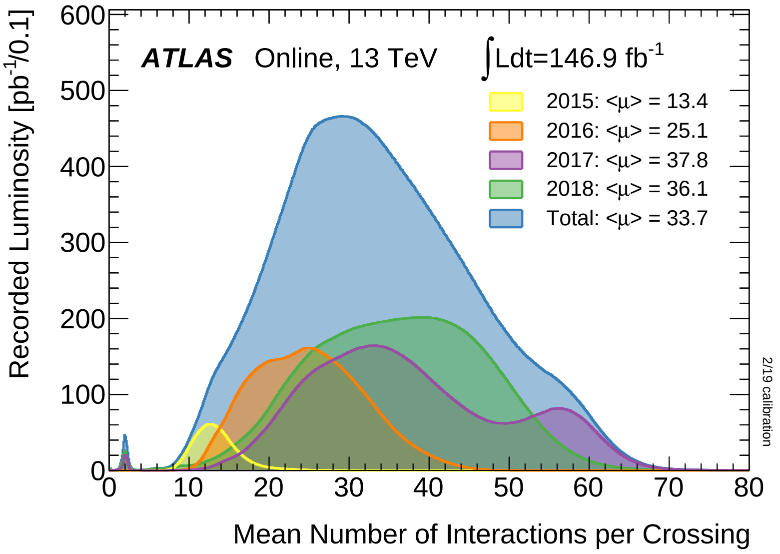


Figure 5.3: The mean interactions per bunch crossing for 2015 to 2018 are shown together with the provided and collected luminosity. This plot is an ATLAS public result taken from Ref. [20]

On the other hand, out-of-time pileup refers to contributions from proton-proton collisions before or after the in-time collision. The effect of this type of pileup on detector systems highly depends on the detectors working principle. Depending on factors like pulse shapes or integration time, out-of-time pileup can induce signals in detector systems. Additionally, out-of-time pileup can add signals to tracking detectors, which may shadow particles from the in-time collision.

During LHC operation, detector caverns fill with a gas of thermal neutrons and photons. This cavern background, its particles to be more precise, has a long lifetime (particles can propagate for seconds without interacting) and consists mostly of low-energy particles with kinetic energies at or below 1 MeV, creating signals in detectors.

Proton bunches scraping against collimators far upstream of an interaction point create so-called beam halo events. Particles created by this process, reaching the detectors at the interaction point, are mainly muons, which can pass through the entire detector nearly parallel to the particle beams. These muons can induce signals in both the muon system and the calorimeters.

Collisions between protons with residual gas in the beam-pipe are called beam-gas events, which create detector signals if the collisions happen near an interaction point. These proton-carbon, proton-oxygen, proton-nitrogen, or proton-proton (coming from hydrogen) collisions create additional collision vertices, usually outside of the standard interaction region. Hence these collisions can easily be removed.

Chapter 6

The ATLAS detector

Contents

The ATLAS detector [62] is a multipurpose detector system capable of tracking particles originating from the LHC interaction point located at its center. Figure 6.1 gives a general overview of the detector, while Figure 6.2 visualizes how the different particles interact with the detector subsystems. Due to its versatility and performance of all subsystems, not only discoveries but also precision measurements like the measurement of the W boson mass are possible.

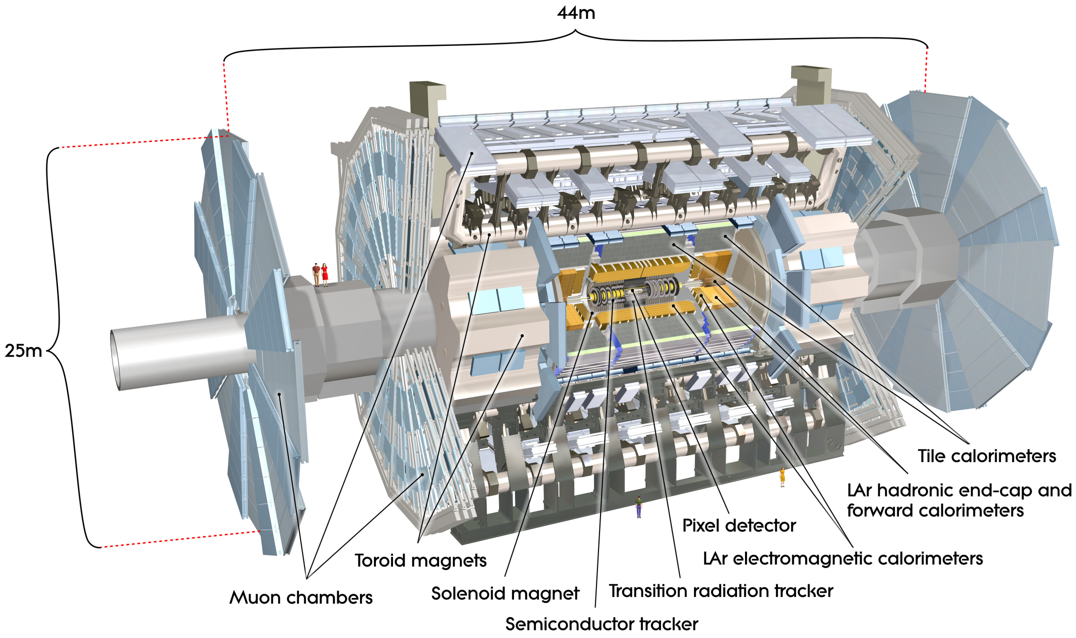


Figure 6.1: Overview of the ATLAS detector [136]. Built following the onion shell principle, it consists of several detector subsystems shown here. The innermost part tracks particles coming from the interaction point, while the calorimeters detect electrons, photons and hadrons. Only muons penetrate the outermost layer, the muon system and are detected there.

It is built in the so-called onion shell principle commonly used in particle physics, having several layers of detector systems around the interaction point. It is constructed symmetrically around the interaction point in its center and divided into three regions. The barrel region is a cylindrical structure built around the beam pipe, measuring in the transverse plane. It is enclosed at both ends by the end-cap (or forward) regions.

Two superconducting magnet systems, which are described in more detail in subsection 6.4.1, are built into the ATLAS detector. The magnet systems are used to bend the tracks of charged particles and thus measure their transverse momenta in both the **I**nner **D**etector (ID) and the **M**uon **S**ystem (MS). Except for the ID, all detector systems are, to some extent, present in the forward regions.

The ID, being the innermost detector system, consists of silicon-based detectors and the **T**ransition **R**adiation **T**racker (TRT). It is used to track particles and assign them to a certain point of collision.

All particles except muons and neutrinos are absorbed in the next layer, the calorimeters, which consists of two parts. The inner layer is formed by the electromagnetic calorimeter, which absorbs electrons and photons. Hadron showers (together with muons and neutrinos) travel through this inner part to the hadronic calorimeter, which consists of a higher density material. All remaining particles except muons and neutrinos are then absorbed in the hadronic calorimeter.

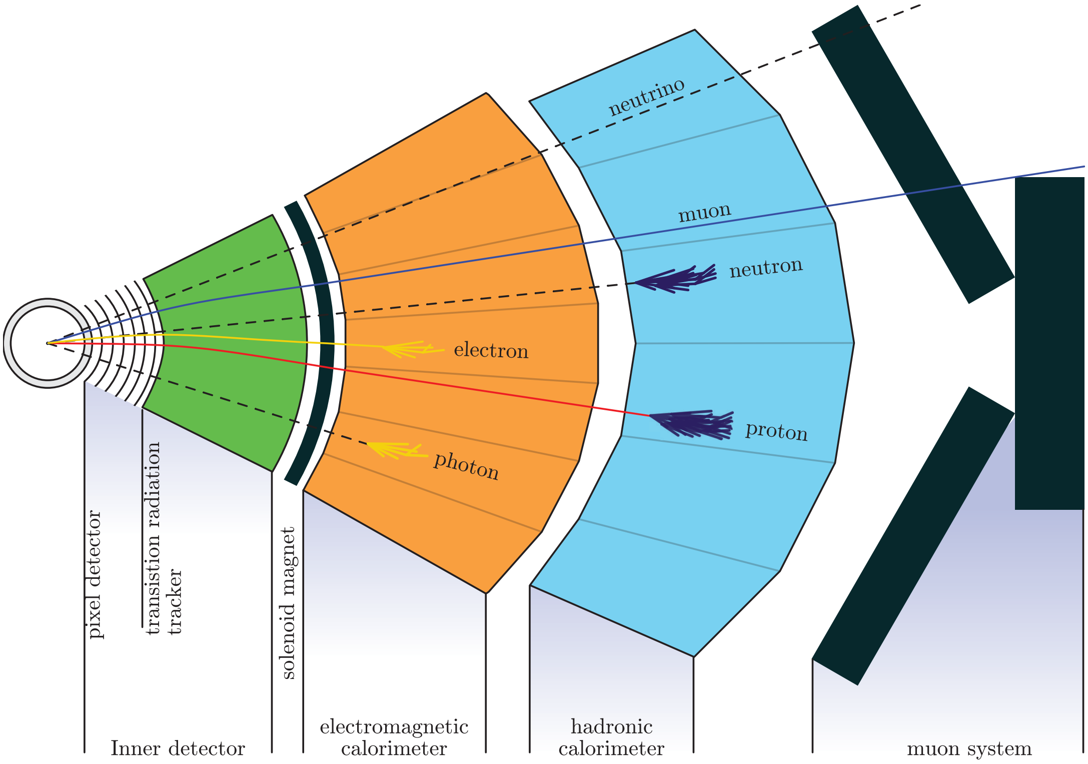


Figure 6.2: Visualization of the detection of different types of particles within the ATLAS detector subsystems. Based on Figure 8 of Ref. [113].

Muons are reconstructed using the MS, the outermost layer of ATLAS, consisting of gaseous detectors. In an ideal case, only muons and neutrinos, the latter not being detectable with the ATLAS detector, penetrate through the calorimeters. In reality, some other particles penetrate the MS and create fake muon signatures. The MS is a dedicated tracking system used to reconstruct the tracks of muons. The muons momentum, and their charge, can be calculated from the curvature of the corresponding track, which is caused by the magnetic field present in the MS.

Quarks and gluons created in collisions are not detectable in the detector but instantly hadronize and create jets of particles, consisting of both leptons and hadrons. The constituents of jets are not detected as single particles in the detector as these particles are created too close to each other but leave signals in the corresponding detector systems. In the context of this thesis, every jet is reconstructed using the anti-k_t jet reconstruction algorithm [41] with a cone-radius of $R = 0.4$. This radius, as well as other kinematic variables, are described below.

6.1 The ATLAS Coordinate System

ATLAS uses several parameters to describe locations within the detector. These are visualized in the following Figure 6.3 and are defined below.

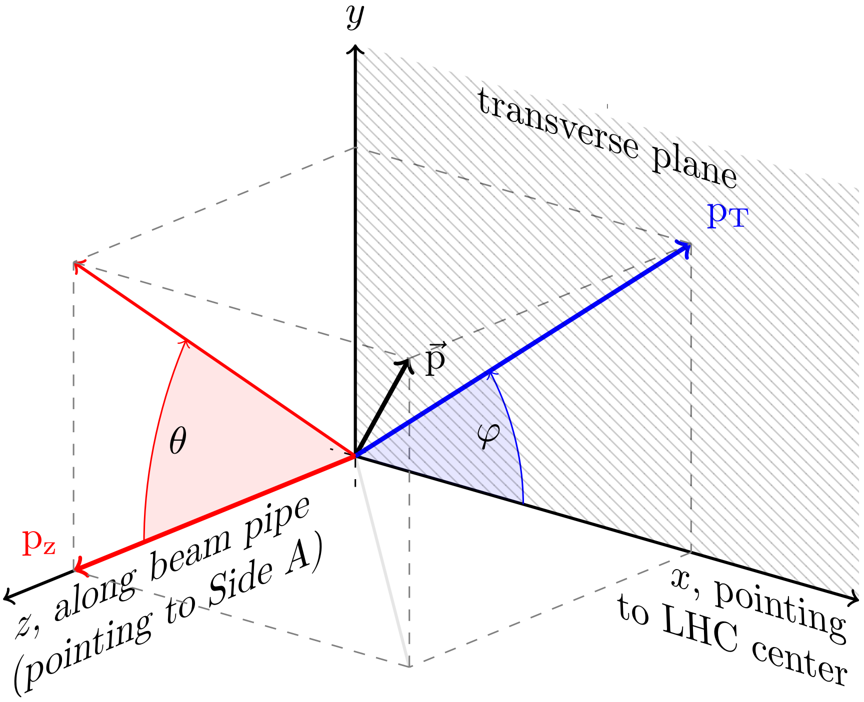


Figure 6.3: Definition of the ATLAS lab frame with the z-axis along the beam pipe and the x-axis pointing towards the LHC center. Side A refers to the side of the detector pointing to the alps. Also shown are kinematic parameters of a particle with momentum $\vec{\mathrm{p}}$.

The ATLAS lab frame is defined as a right-handed coordinate system with its z-axis pointing along the beam pipe, while x points towards the center of the LHC ring. The y-axis is perpendicular to these. Within this frame, two angles are defined. These are the polar angle Θ in the XZ plane and the azimuth angle φ in the XY plane.

6.2 Kinematic variables

Ther rapidity, a measure of the boost of a particles along the beam axis, has already been defined in Equation 4.5. For highly relativistic particles, this approximates through the pseudorapidity η, a measure of a particles track independent of its Lorentz boost along the z-axis, being defined as:

$$\eta = -\ln\left(\tan\frac{\Theta}{2}\right). \tag{6.1}$$

η is a spatial coordinate to measure a particles angle in relation to the beam axis. It is also used to differ between the different detector regions, namely the barrel region ($|\eta| \leq 1.05$) and the forward regions ($|\eta| > 1.05$). Using η and φ, the angular distance ΔR of two particles, described by their angular vectors (η_i, φ_i), can be defined as:

$$\Delta R = \sqrt{(\eta_2 - \eta_1)^2 - (\varphi_2 - \varphi_1)^2}. \tag{6.2}$$

Furthermore, ΔR can be used to define so-called isolation criteria, which are applied to particles. If no other particles are present within a ΔR cone with radius R defined by a given isolation criterium, a particle is considered isolated.

Using proton–proton collisions to perform measurements comes with some limitations as the initial collision energy of the partons involved in the hard-scatter are unknown in beam direction. On the other hand, the momentum can be assumed to be balanced in the transverse plane as the protons collide heads-on. Knowing a particles momentum $\vec{p}$, one can define the transverse momentum

$$\begin{aligned} \vec{\mathrm{p}}_{\mathrm{T}} &= (\mathrm{p_x}, \mathrm{p_y})^{\mathrm{T}} \\ &= (\mathrm{p_T} \cdot \sin\varphi, \mathrm{p_T} \cdot \cos\varphi)^{\mathrm{T}} \end{aligned} \tag{6.3}$$

with its absolute value

$$\mathrm{p_T} = \sqrt{\mathrm{p}_x^2 + \mathrm{p}_y^2}. \tag{6.4}$$

The transverse energy $\mathrm{E_T}$ can be defined using the mass m and $\mathrm{p_T}$ of a particle as:

$$\mathrm{E_T} = \sqrt{\mathrm{p_T^2} + m^2}. \tag{6.5}$$

$\mathrm{p_T}$ and $\mathrm{E_T}$ are identical for massless particles. Each particle within the detector can be fully described by the tuple $(\eta, \varphi, \mathrm{p_T}, \mathrm{E_T}, q)$ using the variables defined above and the particle charge q. Two more variables, the transverse impact parameter d_0 and the longitudinal impact parameter z_0, are needed to define the trajectory of a particle. d_0 is defined as the shortest distance between a track and the beamline in the transverse plane. From this point, the longitudinal impact parameter z_0 can be calculated. It is defined as the distance between the collision position and the point used to evaluate d_0 in z-direction. The corresponding uncertainties are denoted $\sigma(d_0)$ and $\sigma(z_0)$.

6.3 Missing transverse momentum reconstruction

Having ideal head-to-head particle collisions within the ATLAS detector, the vectorial sum of the transverse momentum over all particles created by this collision should be zero due to momentum conservation. This statement is true assuming the detector can detect all particles and match them to the corresponding production vertex, which is not necessarily true in reality. Neutrinos in general, for example, are not detectable by the ATLAS detector at all, as these barely interact with matter. Besides particle characteristics, there are other sources like a malfunction in the detector, insensitivities, detector resolution effects, gaps between detector parts or noise from pileup vertices, as discussed before, contributing to the so-called missing transverse momentum. The following short introduction is based on Ref. [18].

6.3.1 Missing transverse momentum definition

Depending on $\vec{\mathrm{p}}_{\mathrm{T}}$ of all reconstructed particles, the missing transverse momentum vector $\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as:

$$\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} = (\mathrm{E}_{\mathrm{T,x}}^{\mathrm{miss}}, \mathrm{E}_{\mathrm{T,y}}^{\mathrm{miss}})^{\mathrm{T}}, \tag{6.6}$$

with its components

$$\mathrm{E}_{\mathrm{T,x}}^{\mathrm{miss}} = -\sum_{i\in[\text{hard objects}]} \mathrm{p}_{\mathrm{x},i} - \sum_{i\in[\text{soft objects}]} \mathrm{p}_{\mathrm{x},i} \tag{6.7}$$

$$\mathrm{E}_{\mathrm{T,y}}^{\mathrm{miss}} = -\sum_{i\in[\text{hard objects}]} \mathrm{p}_{\mathrm{y},i} - \sum_{i\in[\text{soft objects}]} \mathrm{p}_{\mathrm{y},i} \tag{6.8}$$

and the absolute missing transverse momentum being

$$\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}\right| = \sqrt{(\mathrm{E}_{\mathrm{T,x}}^{\mathrm{miss}})^2 + (\mathrm{E}_{\mathrm{T,y}}^{\mathrm{miss}})^2}. \tag{6.9}$$

Furthermore, one can define its angle in the transverse plain:

$$\phi_{\mathrm{T}}^{\mathrm{miss}} = \tan^{-1}\left(\frac{\mathrm{E}_{\mathrm{T,x}}^{\mathrm{miss}}}{\mathrm{E}_{\mathrm{T,y}}^{\mathrm{miss}}}\right). \tag{6.10}$$

The scalar sum of all contributions, called $\sum \mathrm{E}_{\mathrm{T}}$, is defined as:

$$\sum \mathrm{E}_{\mathrm{T}} = -\sum_{i\in[\text{hard objects}]} \mathrm{p}_{\mathrm{T},i} - \sum_{i\in[\text{soft objects}]} \mathrm{p}_{\mathrm{T},i}. \tag{6.11}$$

This variable gives a measure of the overall hardness of the hard-scatter collision event.

6.3.2 Missing transverse momentum contributions

In reality, the reconstruction of the missing transverse momentum is more complicated than presented so far. The hard-scatter contribution is split into contributions from different particles, as the reconstruction is different for all of them. Depending on the specific requirements of a given measurement, the hard-scatter contributions can change, depending on the accepted particles within an event. Generally speaking, the missing transverse momentum can be defined as:

$$\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{\underbrace{-\sum_{\substack{\text{selected}\\ \text{electrons}}} \vec{\mathrm{p}}_{\mathrm{T}}^{e}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss},e}} \underbrace{-\sum_{\substack{\text{selected}\\ \text{muons}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\mu}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} \underbrace{-\sum_{\substack{\text{selected}\\ \text{photons}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\gamma}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} \underbrace{-\sum_{\substack{\text{accepted}\\ \text{jets}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,\ jet}}} \underbrace{-\sum_{\substack{\text{accepted}\\ \tau\text{-leptons}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\tau_{\mathrm{had}}}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss},\ \tau_{\mathrm{had}}}}}_{\text{hard-scatter}} \underbrace{\underbrace{-\sum_{\substack{\text{unaccepted}\\ \text{tracks}}} \vec{\mathrm{p}}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,soft}}}}_{\text{soft term}} . \tag{6.12}$$

The same holds true for both hard and soft term of $\sum \mathrm{E}_{\mathrm{T}}$:

$$\sum \mathrm{E}_{\mathrm{T}} = \underbrace{-\sum_{\substack{\text{selected}\\ \text{electrons}}} \mathrm{p}_{\mathrm{T}}^{e} - \sum_{\substack{\text{selected}\\ \text{muons}}} \mathrm{p}_{\mathrm{T}}^{\mu} - \sum_{\substack{\text{selected}\\ \text{photons}}} \mathrm{p}_{\mathrm{T}}^{\gamma} - \sum_{\substack{\text{accepted}\\ \text{jets}}} \mathrm{p}_{\mathrm{T}}^{\mathrm{jet}} - \sum_{\substack{\text{accepted}\\ \tau\text{-leptons}}} \mathrm{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}}}_{\text{hard-scatter}} \underbrace{-\sum_{\substack{\text{unaccepted}\\ \text{tracks}}} \mathrm{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\text{soft term}} . \tag{6.13}$$

The hard-scatter terms show only a small dependence on in-time pileup, as it is formed from only fully calibrated particles. These calibrations include a correction for pileup as well as the removal of objects not coming from the hard-scatter vertex.

The missing transverse momentum soft term is solely reconstructed from ID tracks coming from the hard-scatter vertex. This **T**rack **S**oft **T**erm (TST), the contribution of tracks to the $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,soft}}$, is measured by analyzing tracks from the hard-scatter vertex which are not assigned to any detected particle in the dectector and are hence not selected to be part of the actual hard-scatter process. These tracks are required to have a high reconstruction quality and can not be associated with an electron, τ-lepton, or jet already contributing to the corresponding $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ term. The signal-overlap removal process removes

- ID tracks associated with muons
- ID tracks associated with contributing jets
- ID tracks with $\Delta R(\text{track}, \text{electron}) < 0.05$
- ID tracks with $\Delta R(\text{track}, \gamma) < 0.05$
- ID tracks with $\Delta R(\text{track}, \tau - \text{leptons}) < 0.2$

The TST includes contributions from several kinds of ID tracks, e.g., tracks being associated with jets rejected by overlap removal or jets tagged to be probably coming from pileup vertices, as described above. This term is widely used within ATLAS for analysis, as it is mostly independent of pileup processes due to only considering hard-scatter tracks. Additionally, it is not affected by any soft neutral particles coming from any type of vertex.

6.4 The ATLAS Detector Subsystems

ATLAS consists of several subsystems, including detector systems, as already mentioned at the beginning of this chapter, as well as the magnet system and the trigger system, which accepts or rejects collision events. These systems are introduced in this section.

6.4.1 The Magnet System

The ATLAS magnet system [21] consists of four separate, superconducting magnet subsystems. These are the central solenoid magnet and the barrel toroid in the central region, as well as two end-cap toroids, consisting of coils made of aluminum stabilized Cu-NbTi. It is continuously kept at a temperature of 4.5 K. As the momentum of a particle is measured by its deflection by the magnetic field, this field is continuously mapped and monitored to ensure stable operation.

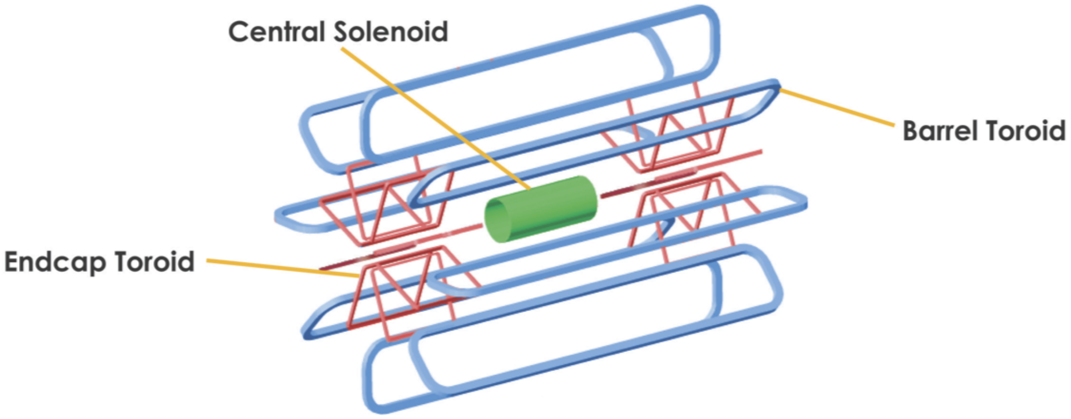


Figure 6.4: Coil structure of the ATLAS magnet system. Taken from [141].

The central solenoid, installed within a cryostat together with the liquid argon calorimeters, is wrapped around the ID and provides 0.9 to 2 T strong magnetic field. The end-cap toroids, built from 8 coils, produce a magnetic field of 0.2 to 3.5 Tesla and are enclosed in individual cryostats. Finally, the barrel toroid provides a magnetic field of 0.2 to 2.5 Tesla. These systems are shown in Figure 6.4.

6.4.2 The Inner Detector

The main purpose of the ID [65] consists of the identification of interaction points (primary vertices), where protons collided in an event, as well as the measurements of tracks coming from these primary vertices.

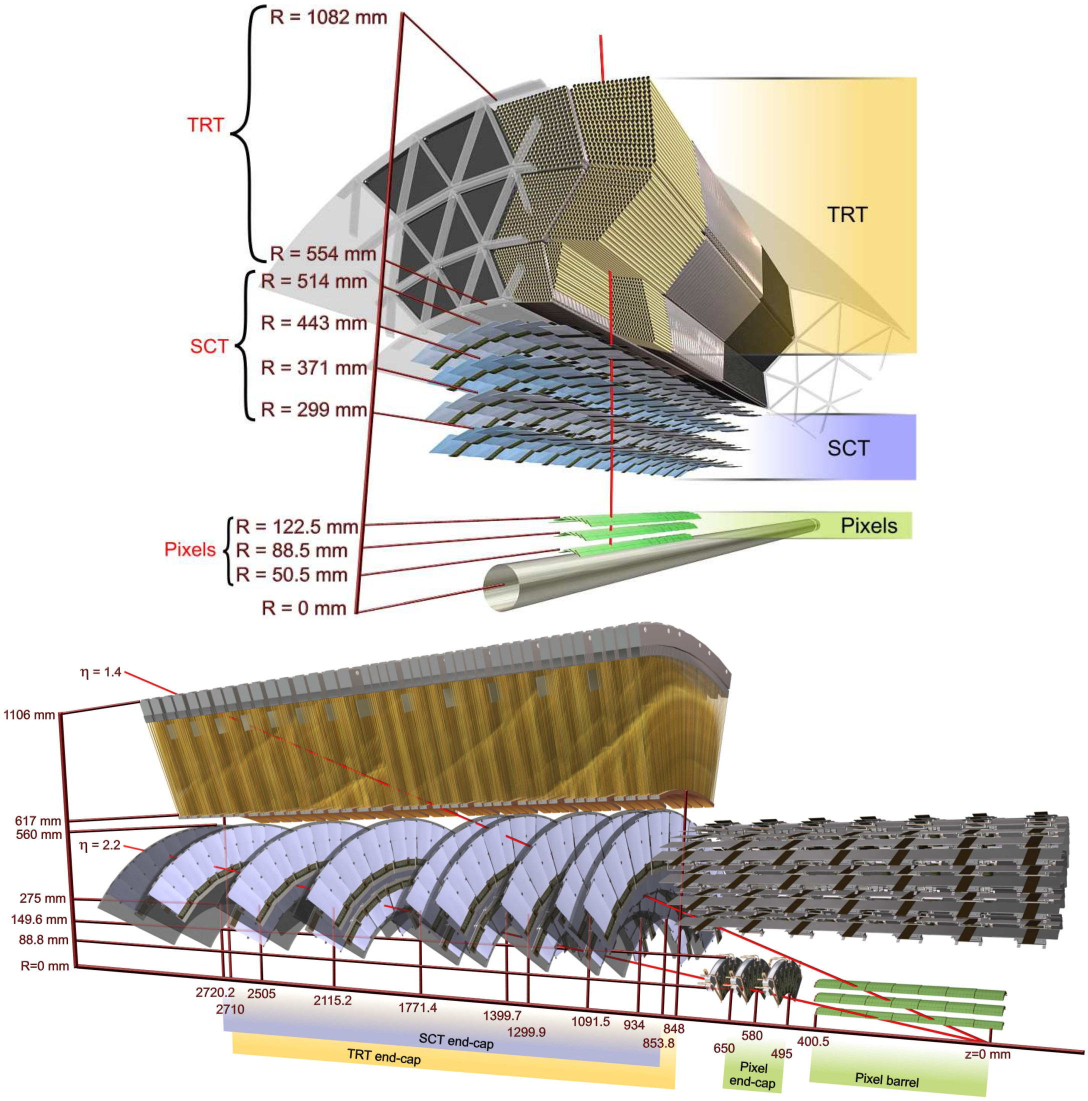


Figure 6.5: The ATLAS inner detector schematic [62], showing all subsystems together with its radial distance from the interaction point R as well as the specific η coverage.

It is 6.2 m long and 2.1 m in diameter and wrapped around the beamline. To measure track properties, high resolution and granularity are required as close to the beams as possible.

The inner detector is split into three subsystems. Directly next to the beam pipe, three layers of pixel detectors are mounted. Behind those, four layers of **S**ilicon **M**icrostrip **L**ayers (SCT) are installed. The TRT is the final layer of the inner detector. Both pixel detector and SCT provide coverage till up to $|\eta| \leq 2.5$, while TRT covers a η range of $|\eta| \leq 2.0$.

6.4.2.1 The Pixel Detector

Due to its position close to the beamline, the pixel detector [47, 106] has to withstand a high radiation environment and, at the same time, provide excellent resolution to track particles. It consists of 1744 sensor modules with a surface area of $19 \times 63\,\mathrm{mm}^2$, mounted around the beam pipe.

Each sensor houses 46080 $50 \times 400\,\mu\mathrm{m}^2$ large readout pixels. It provides a resolution of $12\,\mu\mathrm{m}$ in the (R, φ) plane, with R being the radial distance from the interaction point, and $110\,\mu\mathrm{m}$ parallel to the LHC beam line [106].

6.4.2.2 The Semiconductor Tracker

The SCT [126] is a silicon microstrip detector consisting of 15912 sensors. Being further away from the beam pipe and hence suffering fewer radiation effects allows for this more cost-efficient approach compared to the pixel detector. Each of these sensors has a thickness of 285 µm, and a strip pitch of 80 µm, housing 768 12 cm long readout strips. The strip pitch of 80 µm is slightly varied for the sensors covering the forward region to keep a constant azimuth pitch. For the barrel region, this design leads to an intrinsic resolution of 17 µm in the (R, φ) plane, with R being the distance from the drift tube center and 580 µm in the z-direction. It provides a hit efficiency of $99.74 \pm 0.04\,\%$ [126].

6.4.2.3 The Transition Radiation Tracker

The TRT [125] is built from 72 layers of drift tubes. These drift tubes, being 4 mm in diameter, are oriented in parallel to the beam pipe in the barrel region and the radial direction for the two forward regions. Such a drift tube consists of a 31 µm diameter tungsten wire, stretched in the middle of a multilayer polyimide tube. Being operated with an electron collection time of 48 ns, and a gas mixture of Xe:CO_2:O_2 (70:27:3) leads to a position resolution of roughly 130 µm. During operation, each track creates roughly 30 hits in the TRT, which then have to be matched to single objects.

In addition to tracking, the TRT is used to differentiate between electrons and pions. Due to a polyethylene content in the tubes, the production of transition radiation photons is enhanced. These can efficiently be detected, and from the signal characteristics, these particles can be classified.

6.4.3 The Calorimeter System

Calorimeters measure the energy of particles passing through it by absorption. In the case of ATLAS, so-called sampling calorimeters are used for both calorimeter systems, namely the electromagnetic and hadronic calorimeters.

Calorimeters use a destructive process, as these detectors absorb the particles, which shields detector systems behind the calorimeter from these particles, but makes precision measurements even more important.

These calorimeters consist of several layers, alternating between high-density absorption layers and detection layers. The penetration depth of a particle is a measure for its energy, while the structure and width of the particle shower created from the particle absorption are used to classify the particle type.

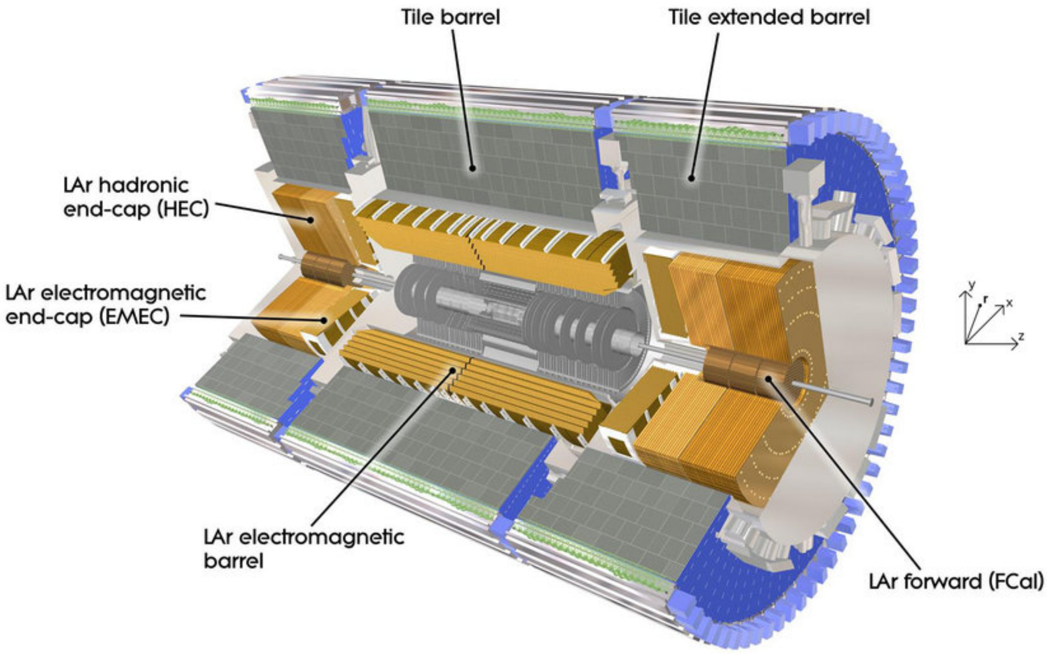


Figure 6.6: The ATLAS calorimeter system [3]. Shown here are both the electromagnetic and hadronic calorimeter subsystems for both the barrel and the forward region. The calorimeters are built around the inner detector, which is shown in the center of the image in grey.

For the operation of ATLAS, and especially high precision measurements, it is crucial to reconstruct as many particles as possible as precisely as possible. To achieve this the ATLAS calorimeter [45] is built with 99.5 % geometric coverage.

6.4.3.1 The Electromagnetic Calorimeter

The innermost calorimeter, the electromagnetic calorimeter, is used to detect photons and electrons, while other particles pass through it nearly no energy deposition. When passing through the calorimeter, electrons lose energy by Bremsstrahlung, which creates photons. Photons undergo pair-production within the material. Hence the energy of electrons and photons is subsequently absorbed by the calorimeter material until the particles are completely absorbed.

The electromagnetic calorimeter is split into the electromagnetic barrel calorimeter, covering $|\eta| < 1.475$, and two **E**lectro**m**agnetic **e**nd-cap **C**alorimeters (EMEC), covering $1.35 < |\eta| < 3.2$. These detectors use liquid argon as a detection medium, which provides high signal linearity and radiation hardness. The argon is ionized by the shower create from absorbed particles in the absorber material. These showers create electron-ion pairs in the argon, which are then collected on an electrode. This electrode is shaped in an accordion geometry, leading to a full φ coverage, avoiding dead regions in the detector. To minimize the chance of not absorbing an electron or photon, the barrel and the forward region of the electromagnetic calorimeter have a depth of 22 and 24 radiation lengths, respectively. The radiation length X_0 is given by the mean length, where an electron loses a fraction of $1/e$ of its energy or for photons as 7/9 of the mean free path for pair production.

The electromagnetic calorimeter is divided into three longitudinal regions for $|\eta| < 2.5$. The first layer, being 4 to 5 radiation lengths thick, is finely segmented in η to provide photon direction measurements and π^0 separation. It uses lead as absorber material. For the barrel region, the cell size in η is 0.003125. The second layer is used to collect most of the shower energy. It has a cell size of $\Delta\eta \times \Delta\varphi = 0.025 \times 0.025$ and covers $|\eta| < 2.5$. Copper is used to absorbing particles in this region. The last region is used to correct for leakage and uses tungsten to absorb particles.

6.4.3.2 The Hadronic Calorimeter

Other particles, except muons and neutrinos, are absorbed by the hadronic calorimeter. While passing through, these particles create further hadronic showers, which leads to wider signals in comparison to the electromagnetic calorimeter. Hence it is built with broader segments and a depth of 11 times the interaction length, leading to a negligible punch-through of hadronic particles into the MS.

The barrel hadronic calorimeter is built from alternating plastic scintillator plates and steel absorber plates, covering $0 < |\eta| < 1.7$. In the forward region the hadronic end-cap calorimeter, covering $1.5 < |\eta| < 3.2$, uses copper as absorber material. To realize a good geometric coverage, this is extended by the liquid argon forward calorimeter is installed directly around the beam pipe, covering $3.1 < |\eta| < 4.9$. It uses copper and tungsten as absorber materials.

6.4.4 The Muon System

The ATLAS MS [46] forms the outermost layer of the ATLAS detector, measuring the properties of muons passing through. It consists of different gaseous detector technologies, optimized for their specific use case. The MS covers a range of $|\eta| < 2.7$ and provides muon triggers to up to $|\eta| < 2.4$. The MS is able to measure the momentum of muons with a precision of $\sigma(\mathrm{p_T})/\mathrm{p_T} < 10\,\%$ for 1 TeV particles. This precision is achieved by, on the one hand, accurately monitored detectors providing precise measurements, but on the other hand, an optical alignment system, which measures the position of the muon detectors within the ATLAS detector with a precision of 30 µm. Without knowing the position of a detector within the ATLAS frame of reference, a track can not be reconstructed using several detectors, which would lead to a decrease in precision.

Monitored **D**rift **T**ubes (MDT)s are used for precision tracking all over the MS. In the barrel region, **R**esistive **P**late **C**hamber (RPC)s are used to trigger events. In this region, the muons are not influenced by magnetic field deflection effects. For the forward region, **T**hin **G**ap **C**hamber (TGC)s are used for this purpose. Both of these chamber types provide a worse spatial but excellent time resolution. The spatial resolution is good enough to match the particles to trigger objects, which makes them ideal for trigger purposes. The innermost forward detector system, called small wheel, is equipped with **C**athode **S**trip **C**hamber (CSC)s, covering $2.0 < |\eta| < 2.7$.

All detector types utilized in the MS are gaseous detectors, in which a counting gas is ionized by muons passing through the detector volume. An electric field splits the electrons and ions, leading the electrons to a dedicated readout structure. Near the readout structure, the electrons are further accelerated by the high electric field, creating avalanches through further collisions with gas atoms, leading to signal amplification. During the drift in the electric field, both ions and electrons induce signals in the readout structure and the cathode. Both types of signals can be measured, which is utilized by some detector types.

6.4.4.1 The Monitored Drift Tubes

For precision measurements of muon tracks, or to be more precise, the bending of the muon track to later calculate its energy, ATLAS houses 1150 monitored drift tube chambers in 3 layers in each direction [57]. These chambers house several layers of MDTs. Each MDT consists out of a 3 cm diameter aluminum tube with a 50 µm tungsten-rhenium wire spanned in its center. The tubes are filled with a gas mixture of Ar:$\mathrm{CO_2}$ of 93:7 at a 3 bar overpressure, which is carefully monitored, leading to the name of this detector type. Utilizing these chambers, 20 independent measurements of the bending coordinate of the track within the magnetic field, are performed, leading to a resolution of 35 µm.

6.4.4.2 The Cathode Strip Chambers

16 CSCs per end-cap cover a range of $2.0 < |\eta| < 2.7$. CSCs are multi-wire proportional chambers, each housing four detection layers [10]. Such a layer consists of a gas volume, enclosed by cathode surfaces segmented in readout strips with a pitch of 5.5 mm to achieve spatial resolution.

In the middle of the gas volume, several 30 µm thick wires, creating a high electric field for avalanche formation, are spanned. Each layer has a resolution in the precision coordinate of 60 µm, which can be combined to further enhance the resolution.

6.4.4.3 The Resistive Plate Chambers

RPCs [44] consist of two flat, conducting surfaces coated with a highly resistive material, creating an electric field between those plates. The electric field is chosen as such, that traversing muons create avalanches in the 2 mm high gas volume between those plates.

The gas volume is filled with a mixture of $C_2H_2F_4(94.7\,\%)$:$C_4H_{10}(5\,\%)$:$SF_6(0.3\,\%)$. The avalanches travel to the plates, where it induces a signal. This signal is capacitively coupled to readout strips beneath the insulating coating.

The readout strips beneath the insulating coat are oriented in the X direction and the Y direction for the anode and the cathode plate, respectively. This setup ensures short drift times and leads to a time resolution of 1.5 ns, and 10 mm of spatial resolution, which is enough for triggering purposes.

6.4.4.4 The Thin Gap Chambers

TGCs [9], like CSCs, are multi-wire proportional chambers. Within the 3600 chambers installed in ATLAS, the distance between wires is larger than the distance to the cathode, which leads to a good time resolution of 4 ns, but a worse spatial resolution.

In this case, the cathode is built from a graphite coated insulator with a stripped copper coating on its backside, pointing towards the gas volume. These chambers, positioned in the forward region of ATLAS, are used to trigger low-p_T muons.

6.4.5 The Trigger System

Due to the high collision rate at the LHC of 40 MHz, corresponding to a bunch crossing every 25 ns, not all data collected by ATLAS can be stored. A dedicated trigger system [64] is implemented to decide which events to store. This system reduces the data rate from 40 MHz collision rate to 1 kHz of accepted events. In Run–II, from 2015 to 2018, the trigger chain consists of a hardware-based trigger system called first-level trigger (L1 in short), followed by a **H**igh-**L**evel **T**rigger (HLT).

The ATLAS Level 1 Trigger is implemented purely in hardware and aims to reduce the data rate to 100 kHz with a fixed latency of 2.5 µs. Specialized hardware finds so-called regions of interest (RoI), regions around a potential particle track, utilizing coarse calorimeter and muon data.

It consists of two subsystems, called L1MUON [11], L1CALO [4], L1TOPO [143] and the **C**entral **T**rigger **P**rocessor (CTP). Hereby L1CALO and L1MUON feed preprocessed information to L1TOPO. The results of L1TOPO are then, together with information from L1CALO, used by the CTP to accept or reject an event. This is also shown in the Figure 6.7.

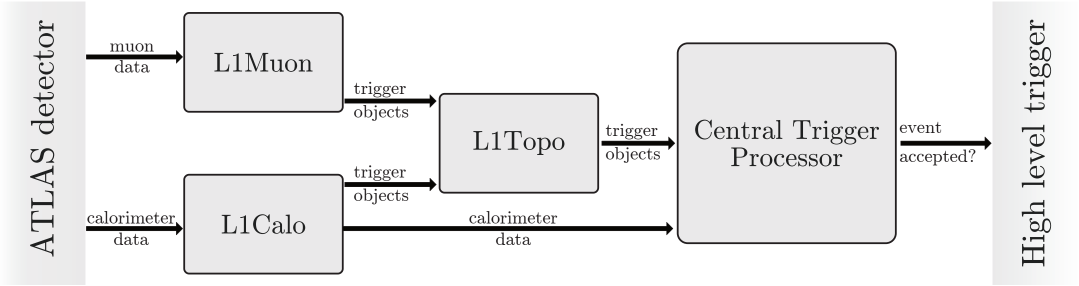


Figure 6.7: Visualization of the ATLAS Level 1 Trigger. It consists of several subsystems, namely L1CALO, L1MUON and L1TOPO. Hereby L1CALO and L1MUON feed preprocessed information to L1TOPO. The CTP uses information from these systems to accept or reject an event. This system reduces the data rate to 100 kHz.

L1Calo, as the name implies, uses information from both the electromagnetic and the hadronic calorimeter. A sliding-window algorithm is applied to find local transverse energy maxima in a range of $|\eta| < 4.9$. The data is taken from two grids of trigger towers, one grid from the electromagnetic, the other coming from the hadronic calorimeter, with each tower covering 0.1×0.1 in $\eta \times \varphi$.

In parallel to L1CALO, L1Muon utilizes data from fast RPCs in the barrel region and TGCs in the forward region to locate coincidences within the MS as well as coincidences between the muon spectrometer and the inner detector.

Both L1CALO and L1MUON pass information to L1TOPO in the form of trigger objects, which include information about particle type, location, isolation status, and energy of identified objects in the detector. L1TOPO performs selections based on kinematic and geometric information.

In a last step within the level 1 trigger, CTP uses the output of L1TOPO and L1CALO information to reject or preliminary accept an event, computing several different triggers in parallel.

In the case of the level 1 trigger accepting an event, the data collected about the event by the detector subsystems is buffered in the so-called Read-Out System (ROS). The trigger information collected by the level 1 trigger is passed to the high-level trigger together with data in the ROS for further processing.

The HLT executes its reconstruction and trigger algorithms in parallel on approximately 40000 CPU cores, aiming to further reduce the rate of recorded events to 1 kHz. Additionally, as the data is already being processed and present in the high-level trigger, the data is also used for trigger-level analysis, detector monitoring, detector calibration purposes, and monitoring.

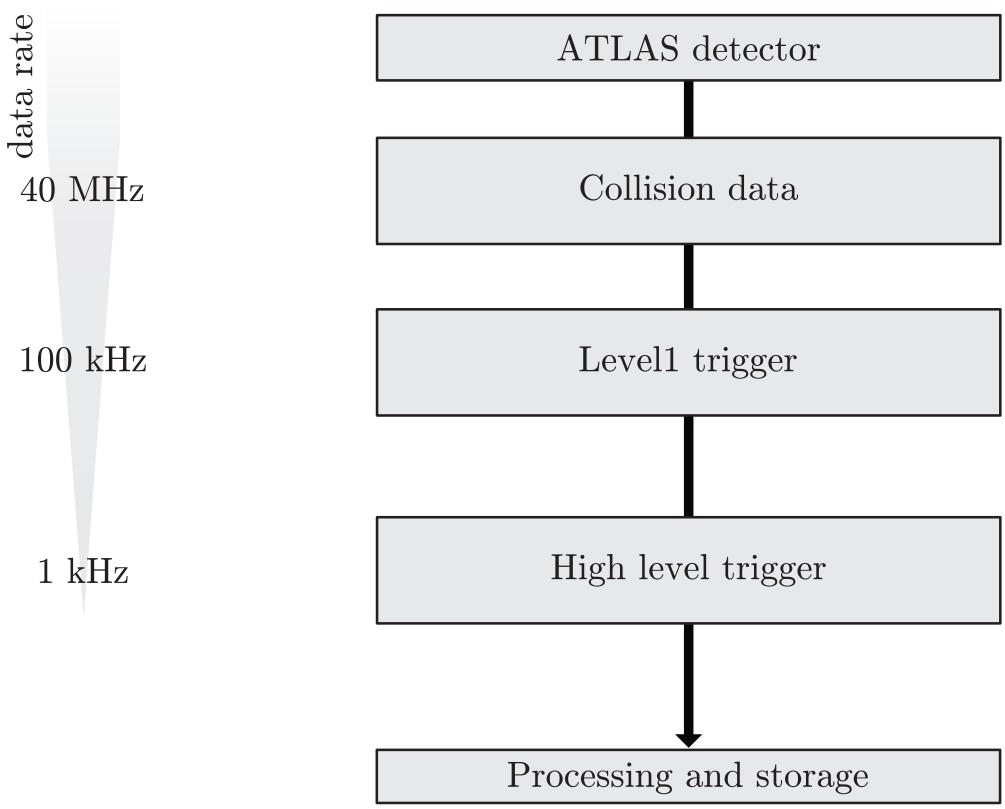


Figure 6.8: Schematic of the ATLAS high-level trigger chain. Data coming from the level 1 trigger is processed further, aiming to reduce the data rate in each step. The remaining data rate after each step is shown on the left.

After the HLT accepts an event, the data collected about this event by all subsystems is passed to local storage at the detector site via data streams. From there, the central facility of the LHC Grid stores the data for offline reconstruction. There the full event is reconstructed.

6.5 Vertexing within the ATLAS detector

As already mentioned, more than one interaction per bunch crossing, so-called primary vertices, appear in an event during proton-proton runs at the LHC. The number of primary vertices per event is called N_{PV}. From the collision data, the positions and hence the vertices themselves are reconstructed from ID signals. Vertices coming from collision products decaying further into other particles are called secondary vertices but are not of interest in this context.

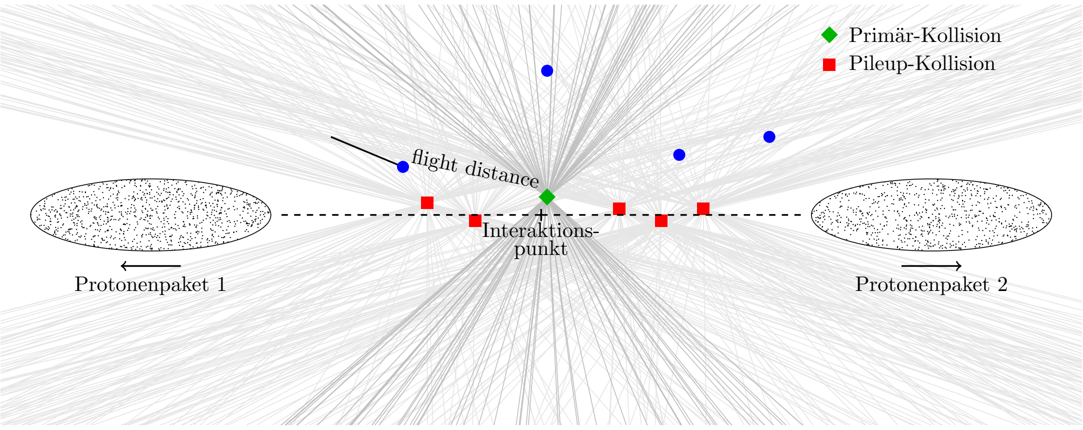


Figure 6.9: Visualization of in-time pileup within the ATLAS detector. Several primary vertices are created, while only the hard-scatter is of interest. This image is loosely based on Figure 2 of Ref. [161].

For a primary vertex to be reconstructed, a certain set of cuts have to be fulfilled:

- Tracks are required to have $\text{p}_\text{T} > 400\,\text{MeV}$.
- Tracks are required to fullfill the tight isolation criteria [17].
- Tracks are needed to fulfill $|d_0| < 4\,\text{mm}$, $\sigma(d_0) < 5\,\text{mm}$ and $\sigma(z_0) < 10\,\text{mm}$
- Further reconstruction quality requirements are applied, which are detailed in [61].

The term hard-scatter vertex refers to the primary vertex of the so-called hard-scatter, the particle interaction with the largest sum of p_T^2 of tracks associated with it. All other primary vertices are considered in-time pileup vertices. This is schematically shown in Figure 6.9.

As can be additionally seen in this Figure, there are constellations where vertices can not be reconstructed, for example, if these appear within a jet. Additionally, having tracks that overlap with other vertices does make the reconstruction more challenging in practice. Algorithms to limit the effect of pileup are present and under development within the ATLAS collaboration. Some of them are discussed in Section 8.1.

6.6 Luminosity Measurement

A precise knowledge of the integrated luminosity is crucial to calculate the number of events of a given process, as stated in Equation 5.3. The luminosity often is, especially in the case of cross-section measurements, as presented in Chapter 9 of this thesis, one of the leading sources of uncertainties (2.0 % to 2.4 %for Run–II [51]). This section, which is based on Ref. [51], provides a short overview of different luminosity measurements within ATLAS.

The luminosity measurement of the ATLAS experiment is based on two key components. Luminosity-sensitive detectors are calibrated in low-luminosity runs, while **v**an **d**er **M**eer (vdM) scans [36, 137] are performed in specially-tailored LHC runs.

Multiple redundant luminosity detectors with complementary capabilities are utilized for this measurement, each using specialized algorithms and hence having different systematic uncertainties. During Run–II, the LUCID2 Cherenkov detector [66] provided the primary bunch-by-bunch luminosity measurement. The measurements of this detector were complemented by the ATLAS **B**eam **C**onditions **M**onitor (BCM) diamond detectors, providing measurements on a bunch-by-bunch basis, as well as offline track counting, a method using the multiplicity of reconstructed charged particles in randomly-selected bunch-crossings. Additional measurements are provided by cluster counting in a network of TimePix readout sensors, which are distributed throughout the ATLAS cavern [155]. Ref. [51, 155] provide further information.

The LUCID2 detector consists of two measurement stations, one per forward arm of the ATLAS detector placed $\approx 17\,\mathrm{m}$ from interaction point 2, each containing 16 photomultiplier tubes. Passing particles produce Cherenkov radiation in the quartz windows of the photomultipliers. These windows are coated with ^{207}Bi radioactive sources, which provide an additional signal used for calibration. The detector parts data is collected using dedicated electronics, which provide so-called luminosity counts for each of the 3564 LHC bunch slots. These counts are then integrated over distinct time periods, so-called luminosity blocks, having a typical length of 60 seconds.

The absolute luminosity calibration of both LUCID2 and BCM, both based on the determination of the visible cross-section σ_{vis} for each of the detection algorithms, is performed using dedicated van der Meer scan sessions in special LHC fills in each data-taking year. The calibration methodology and sources of uncertainty are the same as for the $\sqrt{s} = 8\,\mathrm{TeV}$ measurement and are described in Ref. [155] in more detail.

Besides these dedicated devices, parts of the ATLAS calorimeters can be used to provide bunch-integrated measurements. These bunch-integrated measurements are based on quantities proportional to instantaneous luminosity, for example, liquid-argon gap currents within the end-cap electromagnetic and forward calorimeters, or photomultiplier currents from the hadronic calorimeter.

6.7 Detector Simulation

In order to study several effects coming from the detector systems involved in particle reconstruction, not only **M**onte **C**arlo (MC) simulations of physics processes are utilized, but also how these events are detected by the ATLAS detector as it would for measured data. The term MC generators describes the software used to simulate particle collisions. A simulation is split into four steps:

1. Event Generation
2. Detector Simulation
3. Digitization
4. Reconstruction

MC event generators simulate proton–proton collisions. This simulation includes the hard-scatter process as well as the so-called underlying event, additional parton-parton scatter processes creating additional particles in the same event. This step also includes hadronization and decay processes of unstable particles and is called particle-level.

All particles present after this first simulation step are considered stable. Some examples of such generator are POWHEG [6], PYTHIA8 [154] and SHERPA [32]. Due to the complexity of such an event generator, some generators apply certain simplifications and approximations and later have to be tuned towards measured data.

The final-state particles coming from the event generator are used as input for the detector simulation, which describes the interaction of particles with the ATLAS detector systems. Generally speaking, this step computes particle trajectories, the interaction of particles with matter, and particle decays within the detector system volumes. The most commonly used and detailed detector simulation is the GEANT4 [5] toolkit. In addition to GEANT4, other faster methods, e.g., Fatras [86] and FastCaloSim [67] are available, which can also be combined with Geant4.

Detector simulations are ended once all particles have left the detector volume or remaining particles have energy below a predefined threshold, chosen as such, that no contributions relevant to particle reconstruction are missed. This step results in a number of so-called detector hits, each referring to an interaction of a particle with an ATLAS detector subsystem leading to its detection by the ATLAS detector.

Following the detector simulation, the simulated detector hits are converted into signals similar to those retrieved from a detector system by the detector digitization step. Hereby interactions of particles with sensitive detector material are converted into measurable quantities, such as collected charges, drift times, or photomultiplier signals. Besides computing output signals, event pileup can be added during this step. This state is called detector-level and is technically equal to data collected by the detector.

6.8 Particle-level definitions

Three different particle-level definitions exist, called *born*, *bare* and *dressed* particle-level. On born-level, objects directly coming from the hard-scatter matrix element calculation, i.e. additional FSR is not yet considered. Using this defintion simplifies the comparisson of several decay channels, e.g. electrons and muons created by the Drell-Yan process, as e.g. muons do not have as strong FSR as electrons.

Bare-level particles are stable particles after all FSR has been evaluated. This definition is closer to detectable particles compared to the born-level, but this definition complicates the comparison, following the same argument as for born-level particles.

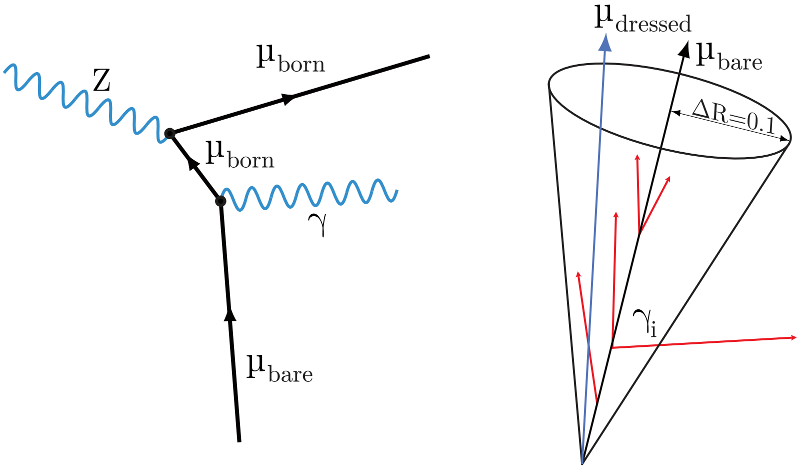


Figure 6.10: Visualization of different particle-levels (left). To obtain dressed-level (right), photons, shown in red, are added to the muon four-vector in a cone of $\Delta R < 0.1$ around the bare-level muon. The image is based on Figure 12 in Ref. [151].

Finally, dressed-level particles are defined in similarity to jets by taking the bare-level particle and add all bare-level photons radiated from it to the lepton four-vector. This is done by defining a cone of $\Delta R < 0.1$ around the lepton vector, as:

$$\mathrm{p}^{\mu}_{\mathrm{T,\ dressed}} = \mathrm{p}^{\mu}_{\mathrm{T,\ bare}} + \sum_{\Delta R(\mathrm{bare}\ \mu,\gamma)<0.1} \mathrm{p}^{\gamma}_{\mathrm{T,\ bare}}. \tag{6.14}$$

Chapter 7

W Boson mass measurement in proton collisions at the LHC

The SM predicts two relations between the W boson mass and SM parameters. Firstly, the weak mixing angle Θ_W relates the masses of the Z boson M_Z and the W boson M_W are related by :

$$M_Z = \frac{M_W}{\cos\Theta_W}. \tag{7.1}$$

Secondly, the mass of the W boson can be calculated at the lowest order from other Standard Model parameters, namely M_Z, the fine-structure constant α, and the Fermi constant G_μ. Higher-order loop-corrections introduce further dependencies on the gauge couplings and the masses of the heavy particles of the SM, especially the Higgs boson and top quark masses. This can be expressed as:

$$M_W^2\left(1-\frac{M_W^2}{M_Z^2}\right) = \frac{\pi\alpha}{\sqrt{2}\,G_\mu}\left(1+\Delta r\left(M_H, M_{top}\right)\right), \tag{7.2}$$

where $\Delta r\left(M_H, M_{top}\right)$ depends on the Higgs Boson mass M_H as well as the top quark mass M_{top} and incorporates all higher-order corrections [22]. All of these parameters were measured by several collaborations before and are summarized in Table 7.1 below. M_W can be predicted using the formulas presented above, as all parameters are known. This calculation results in a leading-order prediction of $M_W = 80.358 \pm 0.008\,\text{GeV}$ [55]. Measuring M_W in data and comparing it to the Standard Model predictions provides the opportunity to test the consistency of the SM. Additionally, discrepancies between prediction and measurement could hint to physics beyond the Standard Model.

Table 7.1: SM parameters and its values, taken from Ref. [159]. The particle mass values are obtained by global fits of different measurements, while the uncertainties of the physics constants are provided in parentheses after the values are the 1-standard-deviation uncertainties in the last digits.

SM Parameter	Measured value
Z boson mass M_Z	$91.1876 \pm 0.0021\,\mathrm{GeV}$
W boson mass M_W	$80.379 \pm 0.012\,\mathrm{GeV}$
Higgs boson mass M_H	$125.10 \pm 0.14\,\mathrm{GeV}$
Top quark mass M_{top}	$172.9 \pm 0.4\,\mathrm{GeV}$
fine-structure constant $\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c}$	$7.2973525693(11) \cdot 10^{-3}\,\mathrm{GeV}$
Fermi constant ${G_\mu}/{(\hbar c)^3}$	$1.1663787(6) \cdot 10^{-5}\,\mathrm{GeV}^{-2}$
Conversion factor $\hbar c$	197.3269804... MeV fm

A first precision measurement of the W boson mass M_W was published by the four experiments ALEPH, DELPHI, L3, and OPAL [87] at the LEP2 accelerator. This electron-positron collider reached the energy threshold for W^+W^- production. Later, the CDF [1] and DØ [2] Collaborations at the Tevatron, a proton-antiproton collider, published the first precision measurements of the W boson mass at a hadron collider. In 2018, the ATLAS collaboration published the result of its W boson mass measurement of $M_W^{\mathrm{ATLAS}} = 80.370 \pm 0.019\,\mathrm{GeV}$ [55].

The W boson decays into a lepton-neutrino pair with a probability of ≈ 11 % for each of the three lepton types and a hadronic decay probability of ≈ 67 %. Due to the extensive multijet background in the hadronic decay channel, only the leptonic decay channels are investigated as shown in Figure 7.1. To be more precise, only the muon and electron decay channels are analysed due to the short lifetime of the tauons. The processes under investigation are:

- $W^+ \rightarrow \mu^+ + \nu_\mu$
- $W^+ \rightarrow e^+ + \nu_e$
- $W^- \rightarrow \mu^- + \bar{\nu}_\mu$
- $W^- \rightarrow e^- + \bar{\nu}_e$

The decay products of these processes are detected and analyzed as the W boson can not be directly observed. The W boson mass M_W can then be derived from the kinematic distributions of the decay products. This is realized by a template fit approach.

A selection of final-state distributions sensitive to M_W are predicted using MC simulations and used as templates. This is done for a range of potential W boson masses. The results of the variations are then compared to the kinematic distributions observed in data to find the W boson mass prediction that fits best to data.

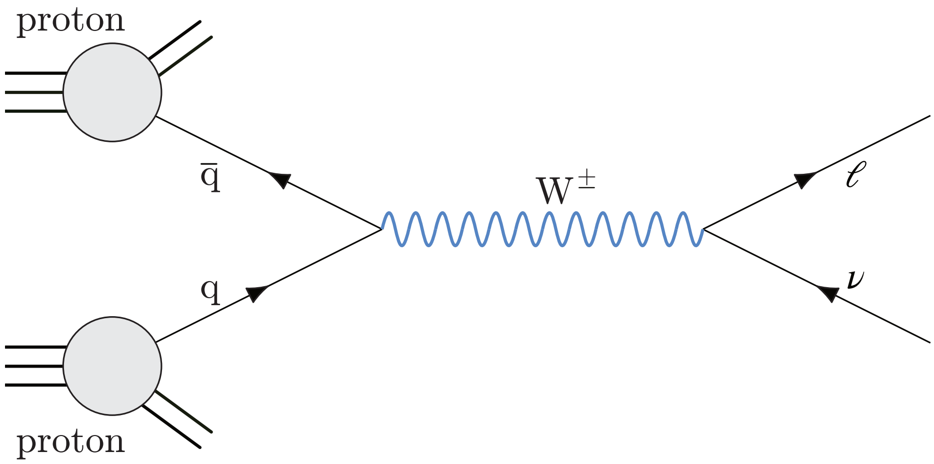


Figure 7.1: Leading order Feynman diagram of a W boson, produced by two quarks, decaying into a lepton-neutrino pair.

In the context of ATLAS, the distributions sensitive to M_W are the transverse momentum spectrum of the charged decay lepton $\vec{\mathrm{p}}_{\mathrm{T}}^{\ell}$, the missing transverse momentum

$$\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}} = -\vec{\mathrm{p}}_{\mathrm{T}}^{\ell} + \vec{\mathrm{p}}_{\mathrm{T}}^{W} \tag{7.3}$$

with the transverse momentum of the W boson $\vec{\mathrm{p}}_{\mathrm{T}}^{W}$ as well as the transverse mass distribution m_{T}. The latter is defined as

$$m_{\mathrm{T}} = \sqrt{2\vec{\mathrm{p}}_{\mathrm{T}}^{\ell}\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}\left(1 - \cos\Delta\varphi(\vec{\mathrm{p}}_{\mathrm{T}}^{\ell}, \vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}})\right)}. \tag{7.4}$$

and hence can be interpreted as the invariant mass of the dilepton system, if the W boson decays in the transverse plane. $\Delta\varphi$ denotes the azimuthal angle between the lepton and the missing transverse momentum. Generally speaking, both $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathrm{p}_{\mathrm{T}}^{\ell}$ peak at $\approx \frac{M_W}{2}$, while m_{T} peaks at roughly the W boson mass.

As these distributions depend on different aspects of both the detector response and the underlying physics modeling, they provide partially uncorrelated measurements of M_W, which can be used as a cross-check for each other. Nevertheless, all of these methods to measure M_W carry its own challenges and limitations, which have to be solved independently, especially in the case of the transverse mass, as it directly depends on $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and $\mathrm{p}_{\mathrm{T}}^{\ell}$.

The $\mathrm{p}_{\mathrm{T}}^{\ell}$ distribution is limited by the knowledge of the momentum scale of the tracking system as well as the energy scale of the electromagnetic calorimeter or muon system. The same holds true for the $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution, which is additionally affected by the calorimeter response parameters and pileup, leading to a reduced sensitivity to M_W.

This problem is even more prominent for the transverse mass as it depends on both of these quantities and additionally requires precise knowledge of the azimuthal angle between lepton and missing transverse momentum.

The calibration of the detector response is typically performed using well-described resonances, such as the Z boson decay, which is kinematically very close to the decay of W bosons in its leptonic decay channels. Utilizing the leptonic Z boson decay, corrections of the momentum and energy scales of the leptonic decay products, including lepton identification- and reconstruction efficiencies, can be derived. Several analysis procedures were specially developed to transfer the corrections from the Z to the W boson events [91].

Furthermore, the transverse momentum distribution of the decay leptons $\mathrm{p}_\mathrm{T}^{\ell}$ is broadened by the intrinsic transverse momentum distribution of the W boson $\mathrm{p}_\mathrm{T}^{W}$, which is created by ISR.

Uncertainties of the $\mathrm{p}_\mathrm{T}^{W}$ modeling, coming from approximations, free parameters in the theoretical calculations, and the PDF choice, influence the lepton transverse momentum spectrum and hence have a direct effect on M_W.

The ATLAS Collaboration follows two approaches to model $\mathrm{p}_\mathrm{T}^{W}$. Measurements of the transverse momentum distribution of leptonic Z boson decays $Z \to \ell\ell$ can be used to tune MC generator predictions, following the same argumentation as for using this process to calibrate the detector response. This tuning process can mitigate the model uncertainties of the $\mathrm{p}_\mathrm{T}^{W}$ measurement but highly relies on precise measurements of the Z boson transverse momentum $\mathrm{p}_\mathrm{T}^{\mu\mu}$. Such a measurement is described in Chapter 9.

Alternatively, $\mathrm{p}_\mathrm{T}^{W}$ can be experimentally measured by analyzing calorimeter signals created by the hadronic recoil of the W boson. This approach is profoundly affected by pileup and can hence only be used as a rough estimate for $\mathrm{p}_\mathrm{T}^{W}$.

The hadronic recoil vector is defined in similarity to $\mathrm{E}_\mathrm{T}^\mathrm{miss}$ in Equation 6.9 as the sum over all 4-vectors of the particles coming from the W boson. Experimentally, it is obtained by the vectorial sum over reconstructed energy clusters in the calorimeters of the detector:

$$\vec{\mathrm{u}}_\mathrm{T} = \sum_{\substack{\text{calorimeter}\\ \text{signals}}} \vec{\mathrm{E}}_\mathrm{T}^\mathrm{calo}. \tag{7.5}$$

In W boson events, $\mathrm{E}_\mathrm{T}^\mathrm{miss}$ and $\vec{\mathrm{u}}_\mathrm{T}$ are essentially the same experimental observable and are directly related by the reconstructed charged lepton transverse momentum $vec\mathrm{p}_\mathrm{T}^{l}$ as:

$$\vec{\mathrm{u}}_\mathrm{T} = \vec{\mathrm{p}}_\mathrm{T}^{l} + \vec{\mathrm{E}}_\mathrm{T}^\mathrm{miss}. \tag{7.6}$$

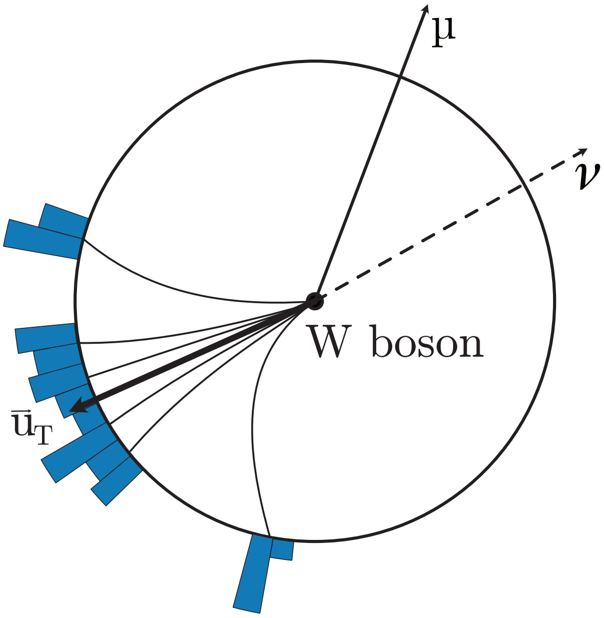


Figure 7.2: Hadronic recoil in a W boson leptonic decay resulting in a muon and muon-neutrino, shown in the transverse plane. The transverse momentum of the W boson p_T^W is created by ISR from the colliding quarks. The hadronic recoil $\vec{u}_T$ can be used to measure p_T^W. The same holds true for decays of the Z boson, but with a muon pair instead of a muon-neutrino pair.

This approach is visualized in Figure 7.2 for a W boson decaying leptonically into a muon and muon-neutrino. Generally speaking, the same method can be applied to events containing a Z boson decaying via the Drell-Yan process to measure p_T^Z. In practice, the Z boson is reconstructed from the leptons created in the decay, as these can be detected with high precision. At the same time, the hadronic recoil is profoundly affected by pileup activity due to additional scattering processes creating more particles in the detector and hence more signals in the calorimeters. The detector itself can not completely mitigate these pileup effects due to the limited resolution of the calorimeters. Therefore Z boson decays are a perfect process to study new pileup mitigation methodologies, as it allows the comparison of the reconstructed $p_T^{\mu\mu}$ via the decay leptons with the hadronic recoil, in principle even in measured data.

Both methods have been studied in a MC $Z \to \mu\mu$ + jets sample, described in more detail in Chapter 8, to visualize the limitations of the direct measurement via the hadronic recoil. Events are required to have two opposite charged muons with a combined four-vector mass between 66 GeV and 116 GeV. Additionally, one muon has to have a transverse momentum above 27 GeV to trigger the event.

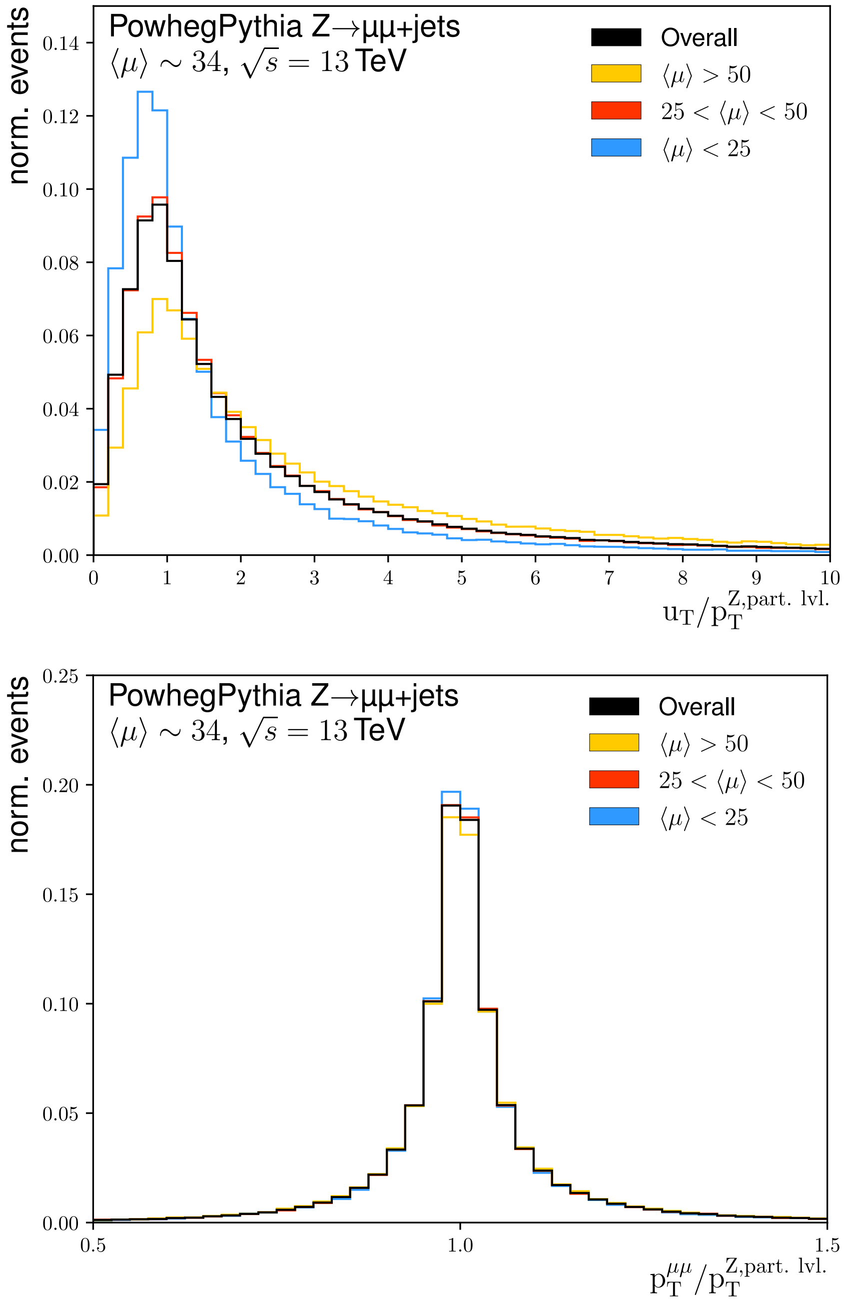


Figure 7.3: Z boson reconstruction performance when using the hadronic recoil (left) or the two muons coming from the Z boson decay. The plots show the ratio of the reconstructed Z boson transverse momentum over its particle-level value $p_T^{Z,\text{part. lvl.}}$ for different pileup activities. The reconstruction using muons is largely unaffected by pileup and more precise, while the performance using the hadronic recoil worsens with higher pileup.

As only ISR creates the boson transverse momentum, the muons are not used to calculate it. Hence, only calorimeter clusters having a $\Delta R \geq 0.2$ to the muons are taken into account. To estimate the underlying ISR activity covered by this ΔR cone, a different region in φ outside the muon cones with the same pseudorapidity η as the muon under investigation is randomly selected. All calorimeter clusters within the randomly selected regions are analyzed and the resulting hadronic recoil is rotated back to the muon cone. This is visualized in Figure 7.4.

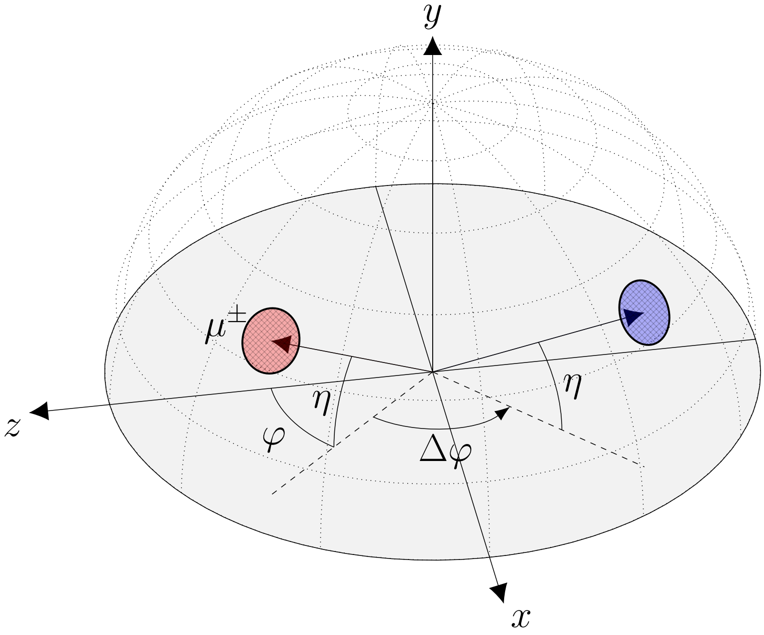


Figure 7.4: Estimation of the underlying event activity. The cone around the muon (red) is rotated randomly around φ. The hadronic recoil in the rotated region (blue) is measured from calorimeter cluster and rotate back.

Figure 7.3 visualizes the reconstruction performance using the hadronic recoil. The ratio of the transverse momentum reconstructed from calorimeter cells over the particle-level transverse momentum is shown. The distribution broadens with higher pileup, having events with ten times higher reconstructed transverse momentum. The same Figure additionally shows the reconstruction performance when using the muons to provide context. These distributions show a neglectable dependence on the pileup activity and are not as broad as the distributions of the hadronic recoil.

To conclude, measuring the hadronic recoil from calorimeter signal can thus far only be seen as a rough estimate of the W boson transverse momentum. The precision and hence the effectiveness of such a measurement highly depends on the mitigation of pileup effects. A novel approach to pileup mitigation using Deep Convolutional Neural Networks, which can potentially increase the precision of such a measurement, is presented in Chapter 8.

Chapter 8

ANN based pileup mitigation

Contents

With higher instantaneous luminosity and more interactions per bunch crossing with every upgrade of the accelerator complex, limiting the effects of pileup is a versatile task. Solving this issue includes better reconstruction algorithms, better detector subsystems, and better triggers, but also applying techniques to classify detector activity and suppress all unwanted activity. The latter is called pileup mitigation.

Currently, ATLAS deploys a variety of pileup mitigation techniques to correct for pileup related effects. On hardware level, the calorimeters calibration and minimal signal thresholds suppress pileup activity. In addition to this, the detectors' ability, or main purpose to be more precise, to track charged particles and associate them to a certain point of origin, the production vertex or collision point, can sort out unwanted activity not coming from the hard scatter. This approach works best for any type of lepton, as these can be tracked with a good resolution.

Jets originating from pileup vertices can also easily be removed based on the measured track information and event topology [50, 59], as long as they do not overlap with other jets. As overlapping jets coming from several different production vertices can no longer be distinguished, this approach can not be used. Hence the pileup contribution to the energy of jets is corrected on average using an area-based approach [39, 41] with a residual correction [59].

Offline reconstruction after data taking offers a way to suppress pileup activity further. A variety of techniques applied on the jet constituent level are under study or already used in ATLAS [48].

Pileup mitigation on constituent-level results in a reduction of local fluctuations and can be seen as corrections for jet substructure observables as well as any other event shapes. A variety of methods are available, all following different approaches. This includes SoftKiller (SK) [40], Constituent Subtraction (CS) [27, 28], cluster vertex fraction [48], and cluster area correction [48]. Additionally, pileup mitigation can be performed on jet level instead of jet constituents. Methods here are e.g. subjet removal via grooming [37, 40, 59, 88, 89, 115, 118] or Jet Cleansing [114].

This chapter describes a new exploratory approach to constituent subtraction pileup mitigation never used within ATLAS before, applying deep convolutional neural networks, which have been introduced in chapter 2, on detector event images. Collision events are treated and processed as images with a defined resolution within the (η, φ) plane. These images are then processed utilizing tools and approaches commonly used within the machine learning community as well as other, already established pileup mitigation techniques. The goal of this project is to predict the hard scatter particle-level image, hence correcting for pileup activity as well as detector effects. The missing transverse momentum $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$, a parameter especially sensitive to pileup, is utilized as a measure of network performance.

Although being a black-box approach, which requires extensive validation and large datasets (which is usually not a problem within particle physics), standardized and widely supported tools are used for the ANN definition, training, and optimization process. The training process is performed using Keras [93] as a high-level description language and Tensorflow [123] as a backend. Results of this study, and hence the first proof of concept, have been partially published in 2019 as Ref. [34]. Therefore, the context of this chapter is based on Ref. [34].

8.1 Commonly used pileup mitigation techniques

As already stated above, there are several pileup mitigation techniques deployed within ATLAS. Utilizing ANNs to solve a problem naturally leads to a black box kind solution, as neural networks describe a (if training successfully) generalized dataset with their trainable parameters rather than physical processes. Given a more complex and not fully describable field of application such as pileup mitigation, validation of results produced with such a trained ANN can be achieved by comparison to validated and commonly used techniques in a variety of well-understood test cases. These are introduced here. The introduction to pileup mitigation techniques shown in this section is based on Ref. [48], together with the corresponding publications mentioned in each subsection.

8.1.1 Jet-Area correction

The jet-area method [39] is the standard pileup correction method used within the ATLAS collaboration. This method uses two parameters, the area in (η, φ) occupied by a jet and the event pileup density, to correct the transverse momentum of each jet present in an event.

The jet area is measured by applying the k_t jet reconstruction algorithm [90] on a customized detector output. A large number of infinitely soft particles, so-called ghosts, are added to the detector output. The area of a certain jet is defined as the area in the (η, φ) space occupied by the original jet as well as the area covered by all ghosts clustered into this jet by the k_t algorithm. A particle clustered into a jet is not visible as an individual particle to this algorithm but is covered by a nearby jet. Knowing the areas of all jets, the transverse momentum density ρ, an event-by-event estimate of the pileup density, can be defined as:

$$\rho = \text{median}\left(\frac{\mathrm{p}_\mathrm{T}^{jet}}{A_{k_t}^{jet}}\right), \tag{8.1}$$

the median transverse momentum per jet area over all jets in the event. Using this, the p_T of each jet is then corrected by

$$\mathrm{p}_\mathrm{T}^\mathrm{corr} = \mathrm{p}_\mathrm{T}^{jet} - \rho \cdot A_{k_t}^{jet}. \tag{8.2}$$

8.1.2 Voronoi subtraction

Building upon this, Voronoi subtraction [48] transfers the concept introduced by jet area correction to the constituent level. Voronoi cells in (η, φ) are formed by constituents closer to each other than to any other constituents. Before the clustering of a jet, the $\mathrm{p_T}$ of a constituent is corrected in analogy to the jet area correction. It uses the Voronoi area A_{Voronoi} and the pileup density ρ:

$$\mathrm{p}_\mathrm{T}^{\mathrm{corr}} = \mathrm{p}_\mathrm{T} - \rho \cdot A_{\mathrm{Voronoi}}. \tag{8.3}$$

This approach can result in negative transverse momenta of some constituent, which can then not be used as input for jet reconstruction. There are several approaches to deal with negative transverse momenta. So-called negative suppression removes these from the event, while so-called spreading combines these negative momenta to nearby clusters, which is described below.

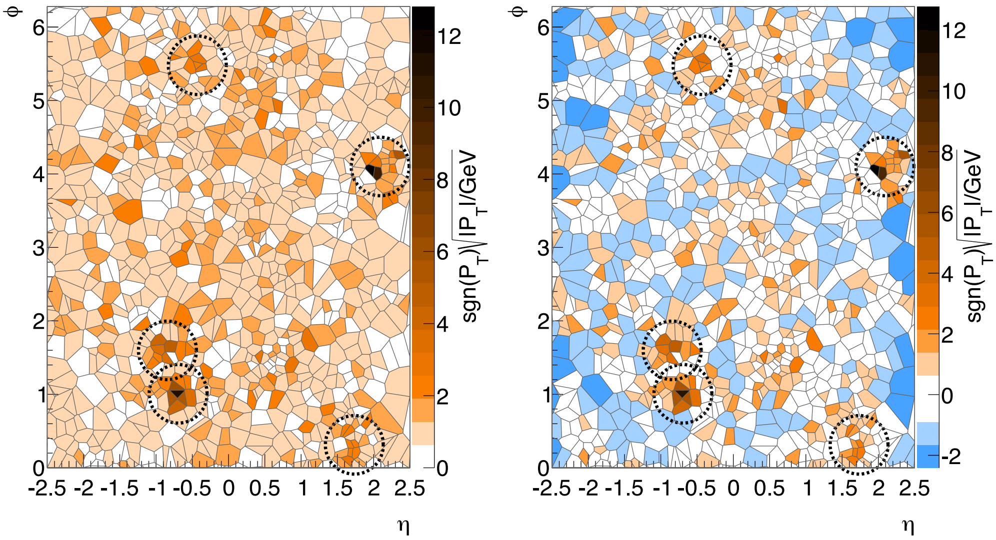


Figure 8.1: Event display before (left) and after (right) Voronoi subtraction. Please not that the right image shows the result of the pure Voronoi subtraction without any further handling of negative cell entries. Circles mark areas before and after corrections. The images are taken from Ref. [48].

For each cluster i with $\mathrm{p}_\mathrm{T}^{\mathrm{corr}} < 0\,\mathrm{GeV}$, clusters j with a distance ΔR_{ij} below 0.4 and positive transverse momentum are selected. For each cluster found this way, the weight w_{ij} i calculated as:

$$w_{ij} = \frac{1}{N} \cdot \frac{1}{\Delta R_{ij}^2} \text{ with } N = \sum_j \frac{1}{\Delta R_{ij}^2}. \tag{8.4}$$

This weight is higher for clusters with smaller ΔR_{ij} and results in $\sum_j w_{ij} = 1$.

The negative transverse momentum of cluster i is spread as:

$$\mathrm{p}_{\mathrm{T}j}^{\mathrm{corr}} = \begin{cases} \mathrm{p}_{\mathrm{T}j}^{\mathrm{corr}} + w_{ij} \cdot \mathrm{p}_{\mathrm{T}i}^{\mathrm{corr}}, & \text{if } |w_{ij} \cdot \mathrm{p}_{\mathrm{T}i}^{\mathrm{corr}}| < \mathrm{p}_{\mathrm{T}j}^{\mathrm{corr}} \\ 0\,\mathrm{GeV}, & \text{otherwise} \end{cases} \tag{8.5}$$

Finally, the transverse momentum of cluster i is set to zero. After this process is finished, there are no clusters with negative transverse momentum left in the event.

8.1.3 Constituent subtraction

In contrast to the jet area correction approaches, constituent subtraction [27] applies corrections to constituents of a jet, correcting both the 4-vector of a jet and its constituents. The contamination of a jet due to pileup is described in terms of the transverse momentum density ρ and mass density ρ_m. These are very similar to the jet area method and also use the k_t algorithm to measure the area of a jet (here denoted patch) and its p_T in the (η, φ) plane. These are defined as:

$$\rho = \mathrm{median}\left(\frac{\mathrm{p}_\mathrm{T}^{\mathrm{patch}}}{A^{\mathrm{patch}}}\right)_{\mathrm{patches}} \tag{8.6}$$

$$\rho_m = \mathrm{median}\left(\frac{\mathrm{m}_\delta^{\mathrm{patch}}}{A_{k_t}^{\mathrm{patch}}}\right)_{\mathrm{patches}} . \tag{8.7}$$

The expected pileup deposition in a small area $\Delta y \Delta\varphi$ is described using the following 4-vector, assuming ρ and ρ_m are only weakly dependent on η and φ:

$$\mathrm{p}_\mathrm{T}^{\mathrm{pileup}} = \begin{pmatrix} \rho\cos\varphi \\ \rho\sin\varphi \\ (\rho+\rho_m)\sinh y \\ (\rho+\rho_m)\cosh y \end{pmatrix} \Delta y \Delta\varphi. \tag{8.8}$$

The transverse momentum of such a patch $\mathrm{p}_\mathrm{T}^{\mathrm{patch}}$ as well as its mass $m_\delta^{\mathrm{patch}}$ are determined by summing over all particles within that patch

$$\mathrm{p}_\mathrm{T}^{\mathrm{patch}} = \sum_{i\in\mathrm{patch}} \mathrm{p}_{\mathrm{T},i} \tag{8.9}$$

$$m_\delta^{\mathrm{patch}} = \sum_{i\in\mathrm{patch}} \sqrt{\mathrm{p}_{\mathrm{T},i}^2 + m_i^2} - \mathrm{p}_{\mathrm{T},i}, \tag{8.10}$$

with $\mathrm{p}_{\mathrm{T},i}$ and m_i being the transverse momentum and the mass of particle i.

Each event is populated with massless particles with very low momentum, called ghosts, such that they uniformly cover the (y, φ) plane with high density.

Each ghost covers a fixed area A_g in (y, φ). Using $m_\delta = \sqrt{\mathrm{p}_\mathrm{T}^2 + m^2} - \mathrm{p}_\mathrm{T}$, the 4-momentum of both the particles and the ghost is expressed as:

$$\mathrm{p}^\mu = \begin{pmatrix} \mathrm{p}_\mathrm{T} \cos\varphi \\ \mathrm{p}_\mathrm{T} \sin\varphi \\ (\mathrm{p}_\mathrm{T} + m_\delta) \sinh y \\ (\mathrm{p}_\mathrm{T} + m_\delta) \cosh y \end{pmatrix} \tag{8.11}$$

Even after adding the ghosts to the events, jet clustering algorithms identify the same jets as before, but these now contain ghost particles which can be used to correct the jet structure. Starting from Equation 8.8, mass and transverse momentum of a ghost can be identified with the pileup contamination:

$$\mathrm{p}_\mathrm{T}^g = A_g \cdot \rho \tag{8.12}$$

$$m_\delta^g = A_g \cdot \rho_m \tag{8.13}$$

The mass and transverse momentum of constituents of a jet are corrected iteratively. For each pair of particle and ghost, ghosts are matched to particles utilizing their distance ΔR_{ik}:

$$\Delta R_{ik} = \mathrm{p}_\mathrm{T}^\alpha \sqrt{(\eta_i - \eta_k^g)^2 + (\phi_i - \phi_k^g)^2} \tag{8.14}$$

Hereby α can be any real number but is usually set to zero. The distances of all pairs are sorted and the pileup is corrected starting with the pair with the lowest distance. Transverse momentum and mass of the pair are corrected as:

$$\begin{array}{lll|lll} \textbf{if } \mathrm{p}_\mathrm{T,i} \geq \mathrm{p}_\mathrm{T,k}^g\textbf{:} & \begin{array}{l} \mathrm{p}_\mathrm{T,i} \to \mathrm{p}_\mathrm{T,i} - \mathrm{p}_\mathrm{T,k}^g \\ \mathrm{p}_\mathrm{T,k}^g \to 0\,\mathrm{GeV} \end{array} & & \textbf{if } m_{\delta i} \geq m_{\delta k}^g\textbf{:} & \begin{array}{l} m_{\delta i} \to m_{\delta i} - m_{\delta k}^g \\ m_{\delta k}^g \to 0\,\mathrm{GeV} \end{array} \\ & & & & \\ \textbf{else:} & \begin{array}{l} \mathrm{p}_\mathrm{T,i} \to 0\,\mathrm{GeV} \\ \mathrm{p}_\mathrm{T,k}^g \to \mathrm{p}_\mathrm{T,k}^g - \mathrm{p}_\mathrm{T,i} \end{array} & & \textbf{else:} & \begin{array}{l} m_{\delta i} \to 0\,\mathrm{GeV} \\ m_{\delta k}^g \to m_{\delta k}^g - m_{\delta i} \end{array} \end{array}$$

This is repeated until $\Delta R_{i,k} > \Delta R_{max}$, for a predefined ΔR_{max} threshold parameter.

8.1.4 The SoftKiller algorithm

The **S**oft**K**iller (SK) [40] algorithm is a straightforward method using only the p_T of an uncalibrated particle to differ between particles originating from the hard scatter and the pileup vertices.

A hard p_T cut is applied to accept or reject particles. Therefore, the detector into a (η, φ) grid. For each event, the p_T threshold of this cut is chosen to leave half the grid entries empty. This threshold sets the median particle flow per bin of this grid to zero. A more detailed explanation of the algorithm and the grid can be found in Ref. [40].

SoftKiller can be combined with both Constituent subtraction and Voronoi subtraction to enhance the performance further.

8.2 Training sample

The network is trained using a supervised learning approach, which is based on providing the desired output to the ANN during training. Hence no real data, as there is no particle-level information available, but simulated events are used for training. This study uses a Z→μμ + jets MC dataset, which is generated taking higher-order corrections into account, which results in the creation of jets from ISR and FSR. Figure 8.3 shows some example Feynman diagrams of this process. Ref. [34] describes full technical details like detector tune or used PDF.

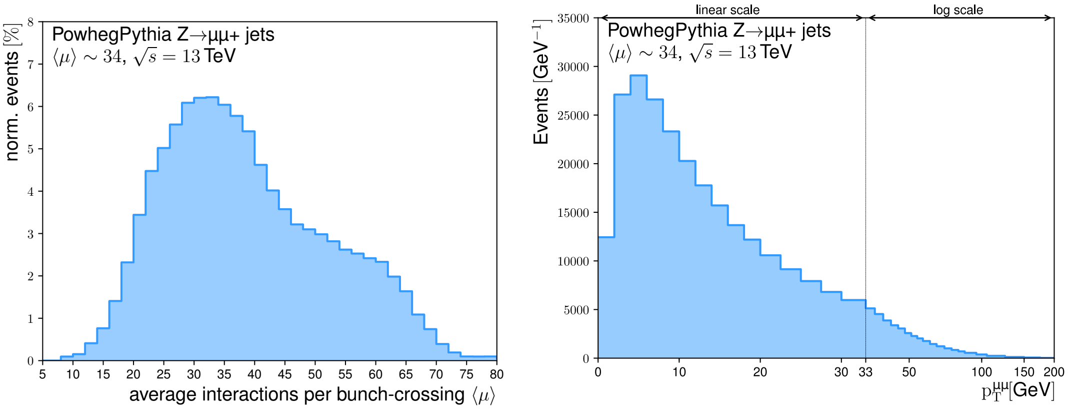


Figure 8.2: Binned particle-level transverse momentum distribution of the Z boson.

This dataset is chosen to create a controlled environment for the training process of the ANN utilizing the high muon reconstruction efficiency and precision of the ATLAS detector, as stated in Chapter 6. As a first step, this study only uses events with a Z boson reconstructed from muons on detector-level. This selection requires two opposite-charged muons with at least medium quality and a combined four-vector mass between 66 GeV and 116 GeV. To simulate a trigger, only events with a muon with transverse momentum above 27 GeV are used for this analysis.

To then add artificial missing transverse momentum, emulating events with non-zero E_T^{miss}, the two muons with combined invariant mass closest to the Z mass peak at 91 GeV are removed from an event on both particle and detector-level. This process leads to events with missing transverse momentum as it would occur from neutrinos or other weakly interacting particles, which in this case can be very well quantified and used as a crosscheck for validation. This validation could theoretically even be performed in data as the E_T^{miss} can be reconstructed from the muons with high precision. By this approach, only particles created by ISR are left in the event as the Z boson decay products are removed from the event. Hence, hadronic recoil and missing transverse momentum are equal, and results obtained are directly transferable to the W boson mass measurement.

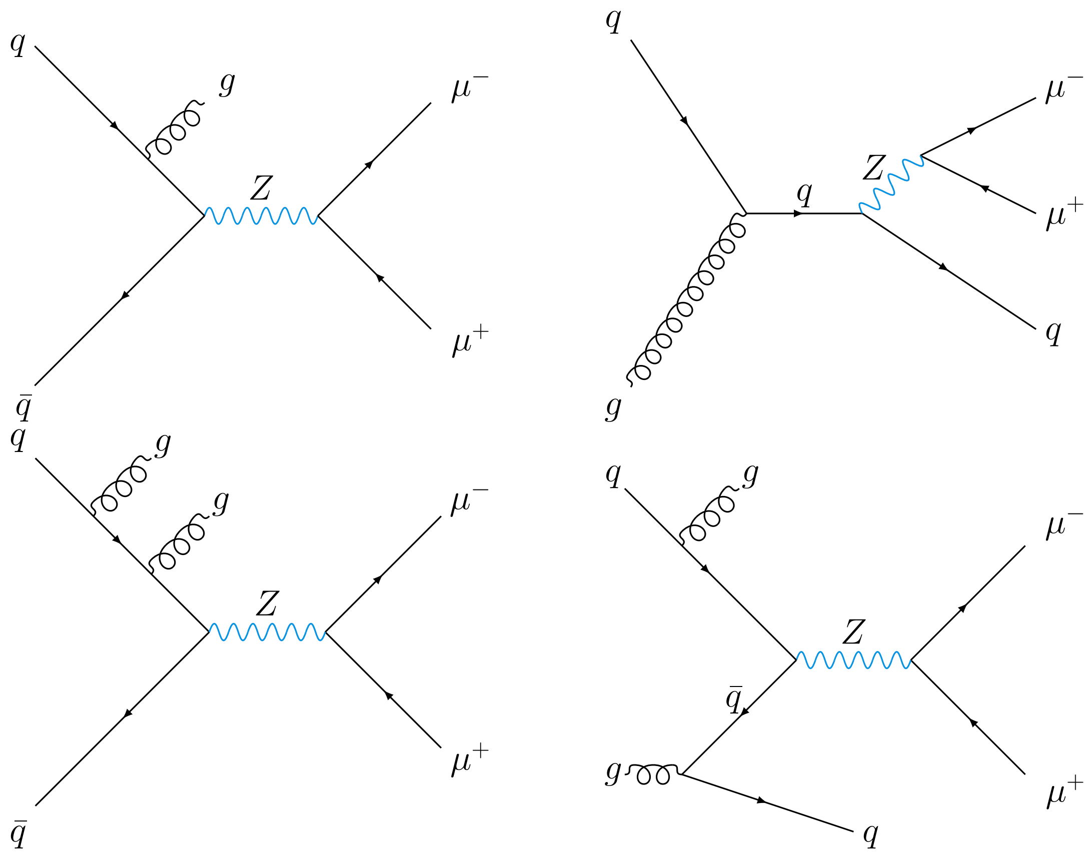


Figure 8.3: Higher order Feynman diagrams of the Z→μμ process. Hadronizing gluons and quarks create additional jets.

8.3 ANN input: Event images

Each event of the dataset is transformed into a set of histograms, the images, each having 50 pixels in η, covering $-2.5 < \eta < 2.5$, and 64 bins in φ covering the full azimuth. Each pixel covers approximately $0.1 \times 0.1\ \mathrm{rad}^2$, which corresponds to the finest single-layer granularity of the hadronic calorimeter [63]. The transverse momentum p_T deposited in the corresponding detector area during this event is stored in each pixel. In order to be as close as possible to the tracking detector acceptance, only constituents with $p_\mathrm{T} > 0.5\,\mathrm{GeV}$ are filled into the histograms. Using finer binning is, of course, possible, but the larger data size makes training a network more time and hardware consuming.

Six detector-level images are constructed and used as input:

- *Clusters* [68]: Calorimeter-cell clusters at the EM-scale.
- *Primary tracks*: charged-particle tracks matched to the hard scatter vertex. Such tracks are required to have at least loose quality [85].
- *Pileup tracks*: charged-particle tracks matched to any primary vertex but the hard scatter vertex.
- *SoftKiller* [40]: Calorimeter-cell clusters after the application of the SK algorithm, which removes all clusters below some threshold that is determined event-by-event.
- *Constituent subtraction + SoftKiller (CSSK)* [28]: Calorimeter-cell clusters after the application of CS and SK. The CS procedure locally and dynamically subtracts momentum from each cluster based on the pileup density.
- *Voronoi subtraction + SoftKiller (VorSK)* [48]: Calorimeter-cell clusters after the application of Voronoi subtraction and SK. Voronoi subtraction is a local area-based pileup subtraction technique where the Voronoi areas are used to estimate the amount of momentum that should be locally subtracted.

The operational parameters used for the SK, CSSK, and VorSK algorithms can be found in [34], following the ATLAS recommendations. Additionally, the particle-level event image is created and used as the training output for the network. In these images, all primary particles considered stable ($c\tau > 10\,\mathrm{mm}$) except neutrinos and muons (as these are artificially removed from the dataset) are included. As on detector-level a $p_\mathrm{T} > 0.5\,\mathrm{GeV}$ cut is applied.

The particle-level is chosen as output because calorimeter cell signals can not be fully distinguished between pileup and hard scatter contributions due to its non-linear response. Due to this fact, the network learns, in contrast to the other pileup mitigation techniques, to correct for the detector response and about the dependence on the particle-level energy spectrum [49, 76].

Table 8.1: Machine learning pileup mitigation network architecture. The dropout rate of each dropout layer is set to 0.1, disabling 10 % of connections.

Layer	Kernel size	Filters	Activation function	Parameters
2D convolution	9	15	ReLU	7305
Dropout and Wrap padding				0
2D convolution	7	20	ReLU	14720
Dropout and Wrap padding				0
2D convolution	5	25	ReLU	12525
Dropout and Wrap padding				0
2D convolution	1	1	ReLU	26
Total				34576

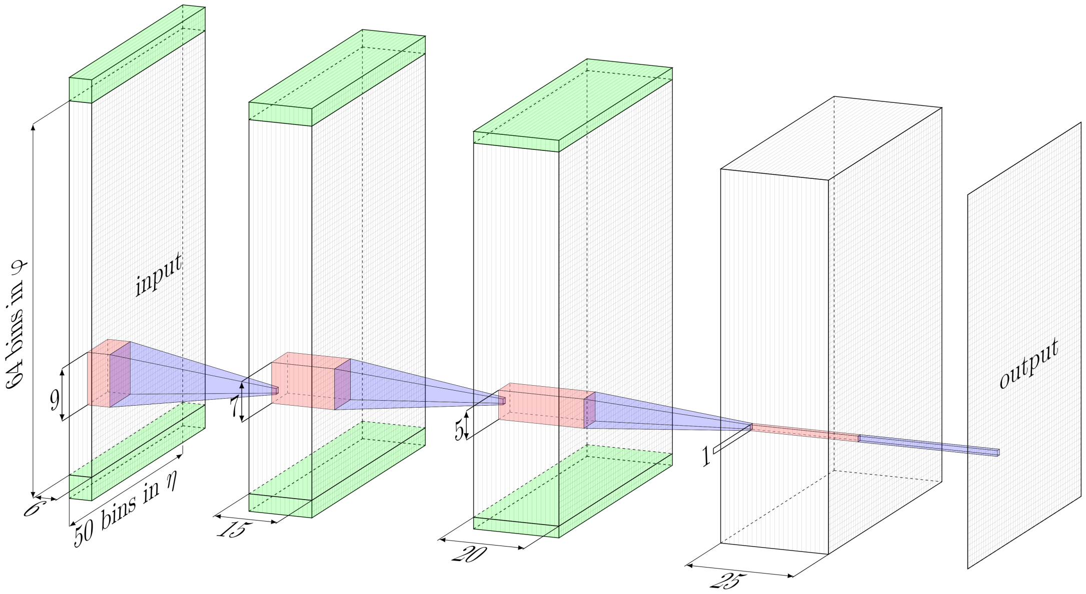


Figure 8.4: Visualization of the network architecture, together with the convolution input (red) and its projection to the next layer (blue) as well as the wrap padding layer (green). The numbers below each layer indicates the number of convolutional filters (e.g. 6 for the input), while the numbers next to the red boxes indicate the size of the filters (e.g. 9 × 9). Shown for one exemplary position.

8.4 ANN architecture

The deep convolutional neural network used in this context consists of four convolutional layers, combined with dropout and padding layers, which are described in Sec. 8.5, between each pair of convolutional layers. This is shown in Figure 8.4. Table 8.1 records the specifications of the network, including the number of trainable parameters per layer. The data stored in the input images are fed through each of the overall ten layers (4 convolutional layers, three dropout, and three padding layers). Each convolutional layer applies more filters on a smaller kernel size compared to its predecessor. The last convolutional layer applies one filter to each pixel with kernel size 1 to adjust the dimension of the output to be comparable to the particle-level event image.

8.5 Padding methods

The padding methods are chosen to preserve the dimension and spatial information of the event images all through the network to keep the output size of a layer in (η, φ) constant. In η, zero-padding is applied. Along the φ axis, so-called wrap padding is used, simulating the continuity of the detector geometry, as the event images cover the full 2π range along this axis. As this is no feature commonly provided within array implementations, a custom network layer is developed, artificially enhancing the event images in between the convolutional layer. Similar to zero padding, additional rows of pixels are added along the φ axis on top and bottom of the image. The number of rows is equal to half the kernel size, rounded down. These additional pixels mirror the pixel entries on the other side of the φ scale to imitate continuous behavior.

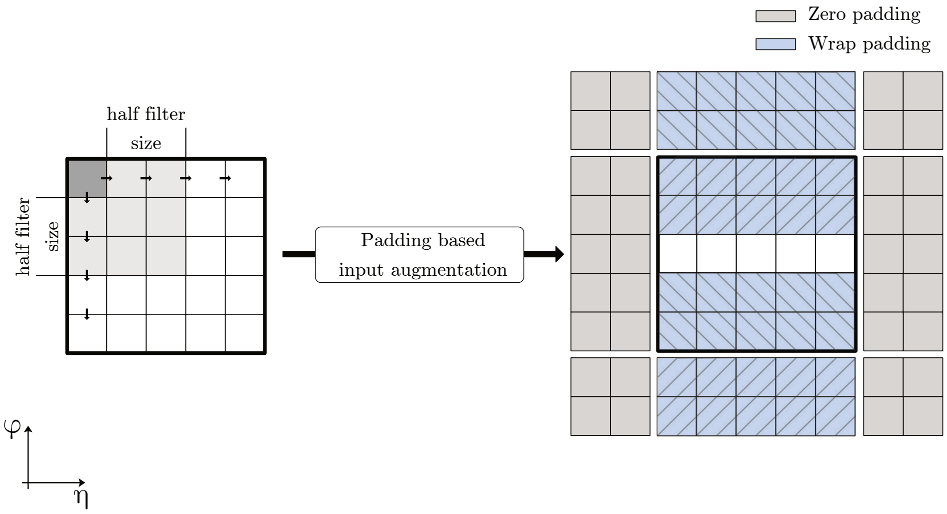


Figure 8.5: Visualization of the event image padding. Along η, zero padding is applied. Along φ, the image is cloned to emulate continuous behavior of the ATLAS geometry. Here, half the kernel size would be 2.

8.6 Training the network

The network is trained using 1.4 million events, split into three parts as:

- 0.8 million training samples
- 0.2 million validation samples
- 0.4 million test samples

One key parameter when training a network is the choice of the loss function as it determines how the weights are updated and has a significant impact on the performance of the ANN. Two ANNs are trained using different loss functions. Each loss function emphasizes different aspects of the problem at hand are trained.

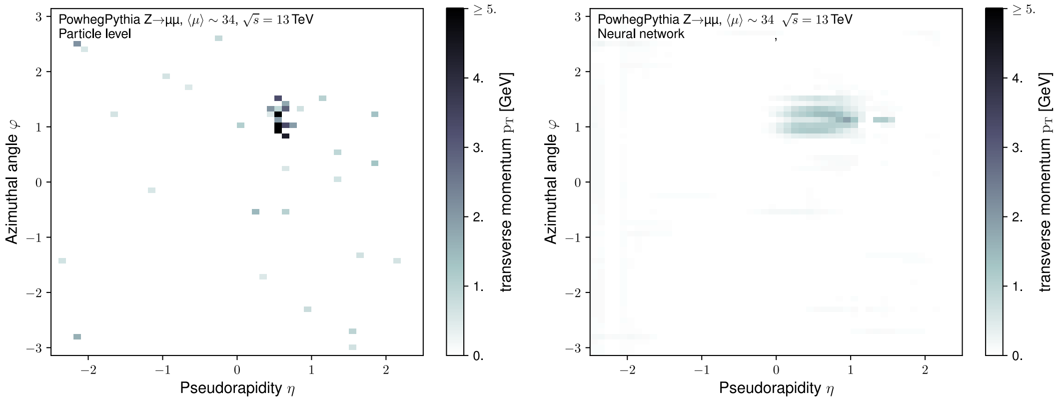


Figure 8.6: Exemplary particle-level images (left) and the corresponding predicted event image (right). The network predicts images with a similar missing transverse momentum, but not the particle-level image itself.

The first loss function directly utilizes a global variable, the missing transverse momentum, which was chosen as core quantity to measure ANN performance. Due to the binned nature of input and output data, a binned version of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ is constructed as:

$$\mathrm{E}_{\mathrm{T,binned}}^{\mathrm{miss,x}} = -\sum_{\eta}\sum_{\varphi} p_{\mathrm{T}}(\eta,\varphi)\cdot\cos\varphi \tag{8.15}$$

$$\mathrm{E}_{\mathrm{T,binned}}^{\mathrm{miss,y}} = -\sum_{\eta}\sum_{\varphi} p_{\mathrm{T}}(\eta,\varphi)\cdot\sin\varphi \tag{8.16}$$

and from this, the loss function is chosen to be:

$$\mathcal{L} = \left\langle \left(\mathrm{E}_{\mathrm{T,binned}}^{\mathrm{miss,\,x}} - \mathrm{E}_{\mathrm{T,binned,\,part.}}^{\mathrm{miss,\,x}}\right)^2 + \left(\mathrm{E}_{\mathrm{T,binned}}^{\mathrm{miss,y}} - \mathrm{E}_{\mathrm{T,binned,part.}}^{\mathrm{miss,\,y}}\right)^2 \right\rangle . \tag{8.17}$$

It uses the pixel centers η, φ coordinates to calculate the missing transverse momentum. Variables denoted *part.* are related to the particle-level event image, the designated output. The mean value of all loss values for each batch entry is chosen not to introduce a direct dependency on the batch size. The results discussed in the following section using this loss function are published as Ref. [34].

The loss function described above provides no incentive to learn the composition of the particle-level event image, but only to find a way to create images with comparable missing transverse momentum. To emphasize local, binwise differences between prediction and particle-level image, the loss function from a similar machine learning project called *PileUp Mitigation using Machine Learning* (PUMML) [112] is used to train the second network. This function is defined as:

$$\mathcal{L} = \left\langle \left\langle \log \left(\frac{\mathrm{p}_\mathrm{T}^{\text{predicted}} + \hat{\mathrm{p}}_\mathrm{T}}{\mathrm{p}_\mathrm{T}^{\text{particle-level}} + \hat{\mathrm{p}}_\mathrm{T}} \right)^2 \right\rangle_{\text{pixel}} \right\rangle_{\text{batch entries}} \tag{8.18}$$

The logarithm of the bin entries of the predicted and desired event image, after subtracting a scaling-factor $\hat{\mathrm{p}}_\mathrm{T} = 16$, is calculated. These are then subtracted. The scaling factor $\hat{\mathrm{p}}_\mathrm{T}$ thereby is an artificially introduced, tunable hyperparameter. The loss is computed per pixel and then later averaged over these pixels. Afterward, the average over all batch entries is calculated to not introduce a dependency on the batch size.

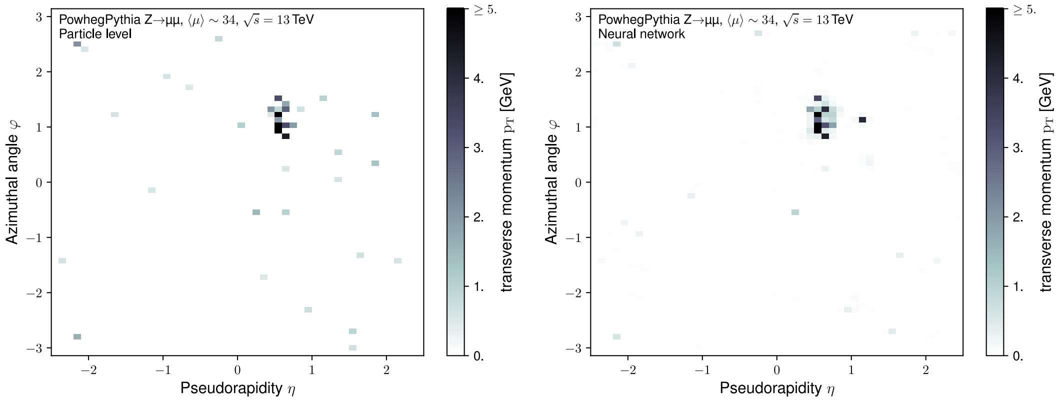


Figure 8.7: Predicted (left) and particle-level image(right). Due to the applied PUMML [112] loss function, the network is able to reproduce the maxima in the hard scatter image, while not emphasizing smaller contributions.

Figures 8.6 and 8.7 show exemplary outputs of both the global and the local loss function trained network, respectively, together with the hard scatter particle-level image the ANNs try to predict. Applying a global loss function leads to an abstract output, while a loss based on local differences results in a comparable output. Nevertheless, both ANNs provide a good measure of the missing transverse momentum within an event, as discussed in the following Section 8.7.

8.7 Performance studies and comparison

The performance of the ANNs is evaluated in direct comparison with established pileup mitigation techniques as well as the measurement of the hadronic recoil in three different test cases.

At first, the output spectrum of the methods are compared, ensuring the network has a non-zero output, which would result in a low E_T^{miss}, but counts as unwanted behavior. Furthermore both the stability and the bias of each method is evaluated by dedicated measurements. Hereby, the missing transverse momentum is measured in relation to the actual missing transverse momentum on particle-level $E_T^{miss,\ part.}$ to evaluate the recontruction performance of a given method. This is defined as:

$$\begin{aligned}\Delta E_{T,x}^{miss} &= E_{T,x}^{miss} - E_{T,x}^{miss,part.}\\ \Delta E_{T,y}^{miss} &= E_{T,y}^{miss} - E_{T,y}^{miss,part.}.\end{aligned} \tag{8.19}$$

The minimal baseline is the *tight* TST E_T^{miss}, as this is currently the most widely used definition for analyses performed within ATLAS [18]. In addition to this, *tight* TST is measured without any jets outside $-2.5 < \eta < 2.5$ as the ANN only uses information up to $|\eta| = 2.5$. This approach leads to a better comparability, as removing jets in the forward region can improve the E_T^{miss} for some topologies. This approach is, due to the missing tracking information, usually used to reject forward pileup jets. Result of SK, VorSK and CSSK methods applied on calorimeter cluster data are used for further comparison.

Table 8.2: Widths of the predicted E_T^{miss} distributions of each method for both the absolute distributions as well as with respect to the particle-level E_T^{miss}.

	x - component		y - component	
Method	abs. width	diff. width	abs. width	diff. width
Hadronic recoil	32.07 ± 0.05	38.06 ± 0.06	32.19 ± 0.05	38.24 ± 0.06
ΔE_T^{miss} based ANN	9.71 ± 0.02	10.37 ± 0.02	9.75 ± 0.02	10.41 ± 0.02
PUMML based ANN	8.30 ± 0.02	10.07 ± 0.02	8.32 ± 0.02	10.09 ± 0.02
Tight TST	14.38 ± 0.06	12.28 ± 0.02	14.26 ± 0.06	12.31 ± 0.02
η-cut tight TST	11.64 ± 0.03	11.64 ± 0.02	11.55 ± 0.03	11.67 ± 0.02
SK	17.30 ± 0.03	18.95 ± 0.03	17.41 ± 0.03	19.02 ± 0.028
Vor+SK	14.10 ± 0.02	16.58 ± 0.02	14.24 ± 0.02	16.68 ± 0.025
CS+SK	14.82 ± 0.02	17.00 ± 0.03	14.90 ± 0.02	17.07 ± 0.025
Particle-level	12.20 ± 0.03		12.21 ± 0.03	

Figure 8.8 shows the distributions of both the absolute E_T^{miss} and the difference to the particle-level value for all algorithms under investigation. The widths of these distributions, measured with the RMS of the distributions, can be found in Table 8.2. One can easily see, that the ANN provides non-trivial, non-zero predictions and results in a sharper resolution compared with the TST E_T^{miss} as well as other studied definitions.

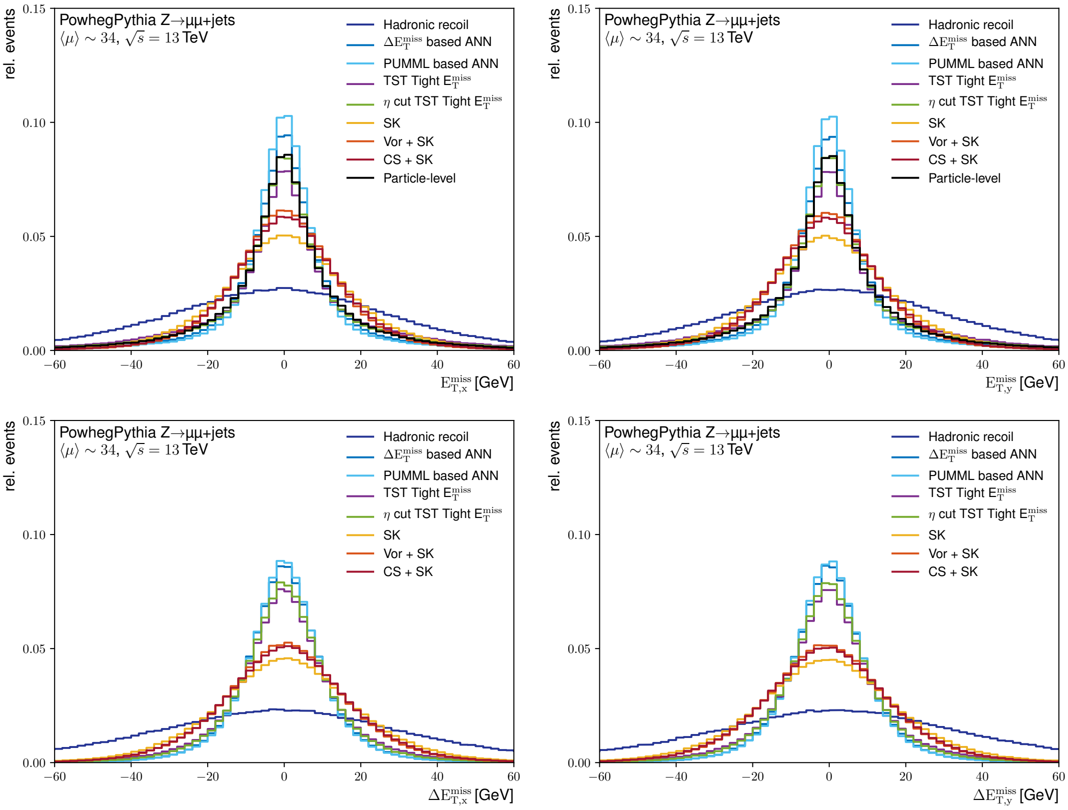


Figure 8.8: Top: The distribution of the E_T^{miss} distribution for various detector-level and particle-level definitions for the η direction (left) and ϕ direction (right). Bottom: The distribution of the difference between the predicted and particle-level E_T^{miss} distribution for various detector-level definitions.

The hadronic recoil performs the worst having the widest distribution. As the SoftKiller based constituent-level techniques do not apply any preselection but work on raw detector input, these have a worse resolution compared to the ANN and TST methods. The results measured in the x- and y-component of both E_T^{miss} and ΔE_T^{miss} are comparable, which is expected as there is no preferred direction within the transverse plane.

The stability of the various $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ methods are quantified as a function of pileup activity. $\Delta\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ distributions are measured in bins of $\langle\mu\rangle$ as well as the number of primary vertices N_{PV}. The width of the distribution in each bin is measured by its RMS.

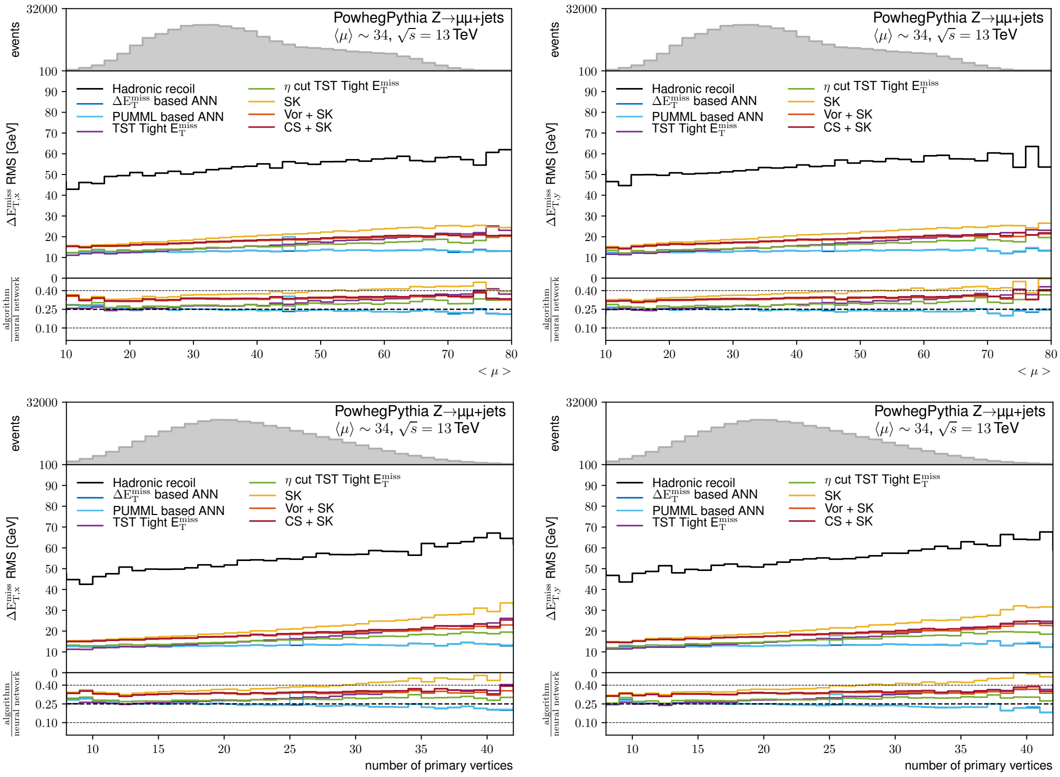


Figure 8.9: Stability of different pileup suppression methods, measured using the RMS of the difference between the predicted and particle-lebel value. The top row quantifies the performance as a function of $\langle\mu\rangle$ while the bottom row shows the dependence on the number of reconstructed primary vertices. The plots on the left are for the x direction while the plots on the right are for the y direction. The underlying distributions of the x-axis quantities is shown in the upper panel of each plot, while the lower part shows the ratio of a given method to the hadronic recoil.

Again, all methods perform more stable than the hadronic recoil, and both the x- and y-component are comparable. The ANNs performs better than the other methods, showing stable behavior over a wide range of pileup activity. Hence both ANNs can perform predictions largely unaffected by pileup activity, a feature unique to this approach. The PUMML based ANN shows a small bump, which is caused by singular outlier events that are wrongly predicted. These events are well predicted by the other ANN, as it is trained using a global function.

The performance of the hadronic recoil worsens with higher pileup activity. It shows a rise of the RMS with higher pileup activity similar to the TST and SK based methods but has a large offset in RMS from the start.

The performance of any method within the field of pileup mitigation is a trade-off between stability and bias. Figure 8.10 shows the bias (or linearity) of each method under investigation. In order to measure the bias, the $\Delta\vec{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ vector is projected along the direction of $\hat{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss,part.}}$. This parameter is measured in bins of the particle-level missing transverse momentum $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,part.}}$. The average value in each bin refers to the bias in the corresponding bin.

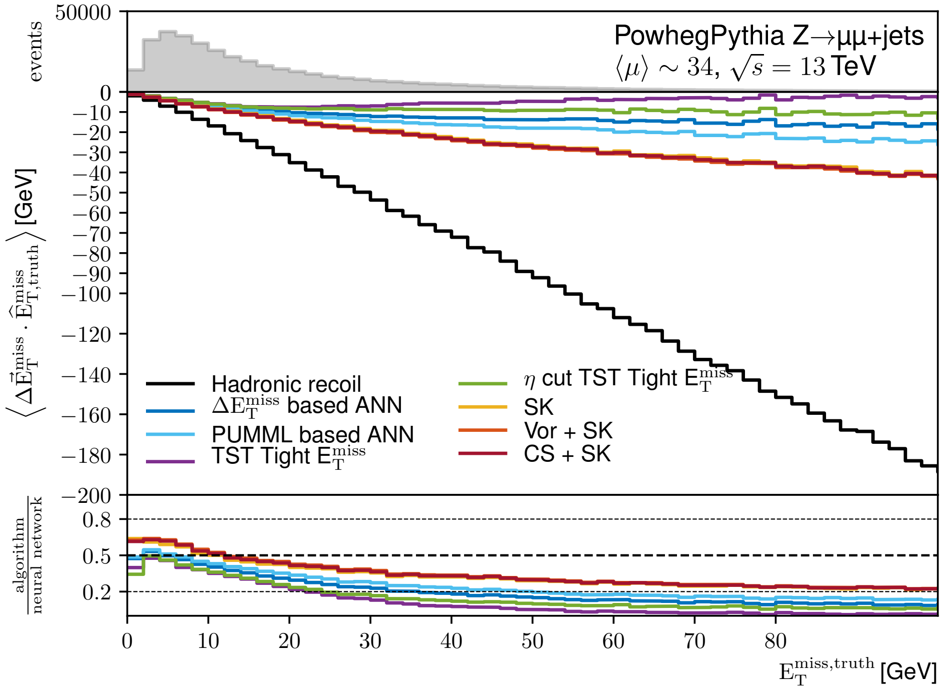


Figure 8.10: Bias of each method, measured in bins of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,part.}}$. $\Delta\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ is projected along the normalized particle-level $\hat{\mathrm{E}}_{\mathrm{T}}^{\mathrm{miss}}$ direction. The upper pannel shows the $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss,part.}}$, while the lower panel shows the bias ratio of each method with respect to the hadronic recoil bias.

The ideal value of the bias is 0, as there would be no difference between prediction and particle-level value. Depending on the angle between prediction and particle-level vector, the sign of the bias can change. For all methods, the mean bias in each bin is negative. For the ANN, the bias is more substantial compared to the TST based methods. As the SoftKiller based methods are not calibrated, the bias is expected to be larger than the TST or ANN approach.

8.8 Advantages and limitations of this approach

To summarize, the previous sections present the results of a first study of the application of deep convolutional neural networks to perform pileup mitigation. The ANNs were trained to perform eventwise corrections, making use of low-level information to improve upon existing techniques. Event information is used as input in the form of event images, histograms with 50 bins in η, and 64 bins in φ. Image recognition techniques, namely convolutional layers, are used to predict the hard scatter particle-level image from the input. An augmented $Z \to \mu\mu + \text{jets}$ dataset, where the muons are reconstructed and later removed to provide a precisely known amount of missing transverse momentum.

The resulting neural network performances are more independent of pileup than the default $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and established mitigation techniques. For $\langle\mu\rangle = 30$, the neural network $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution improves upon the default resolution by about 15 %. Given the importance of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ to the ATLAS physics program, the results presented here are promising. Other groups are looking into the applicability of this approach in a trigger environment, as the ATLAS trigger uses binned data at the first stage of the trigger chain.

Both networks are trained using different loss functions, emphasizing various aspects of the project. Although both ANNs show somewhat similar performances, some differences exist. The $\Delta\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ based loss function utilizes a global variable rather than a per-pixel comparison between the images. Additionally, the ANN has no incentive to work on the precision along the η axis as $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ only depends on the φ coordinate of a pixel. Applying a loss function utilizing local instead of global differences, like the PUMML based loss function, remedies this. One can easily see that the output of the PUMML based network is way more similar to the particle-level image. While this feature does not provide any further advantages, it confirms the concept behind the loss function.

Further improvements may be possible by using more information for training and performing a detailed hyper-parameter scan. Additionally, using finer binning or not limiting the input image to the detector region beyond $|\eta| = 2.5$ may further improve the performance. The latter argument is especially important when studying the performance of the ANNs in the context of boson kinematic reconstruction.

Its performance is measured using the so-called bin *stability* S and *purity* P, key metrics in correcting for detector effects. These parameters are derived from a so-called response matrix, a two-dimensional histogram filled with events that happened in a certain bin on particle-level (x-axis) and detector-level.

Ideally, such a matrix has non-zero entries only on the diagonal, meaning that all events created in a particular bin on particle-level are present on the same bin on detector-level. In this case, no correction has to be applied and the measurement would be independent of any MC generator.

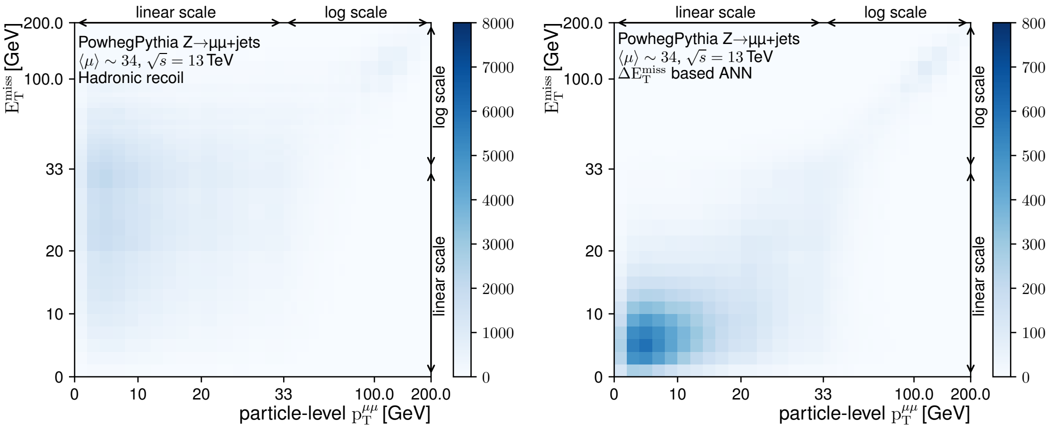


Figure 8.11: Exemplary response matrices of the hadronic recoil (left) and $\Delta\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ based ANN pileup mitigation approach. Shown is a two-dimensional histogram filled with entries of both the particle-level and detector-level Z boson transverse momentum. Ideally non-zero entries can only be found in the diagnonal entries.

In reality, off-diagonal entries, also called bin migrations, exist. The more bin-migrations exist, the stronger a measurement depends on the underlying physics model and hence suffers from a stronger model uncertainty, one key limitation the W boson mass measurement. Given a response matrix, as shown in Figure 8.11, stability, and purity of bin i can be calculated by:

$$S_i = \frac{N_i^{\text{det. \& part. lvl.}}}{N_i^{\text{part. lvl.}}}, \tag{8.20}$$

$$P_i = \frac{N_i^{\text{det. \& part. lvl.}}}{N_i^{\text{det. lvl.}}}. \tag{8.21}$$

Hereby $N_i^{\text{det. lvl.}}$ refers to the overall number of events visible in detector-level bin i, $N_i^{\text{part. lvl.}}$ the number of events created in particle-level bin i and $N^{\text{det. \& part. lvl.}}$ the number of events created in the same bin i on detector-level and particle-level (the diagonal entries). Both stability and purity are a measure of bin migrations, meaning how many events created in a certain bin on particle-level end up visible in the same bin on particle-level or detector-level, respectively. The higher the stability and purity, the more reliable is the unfolding process.

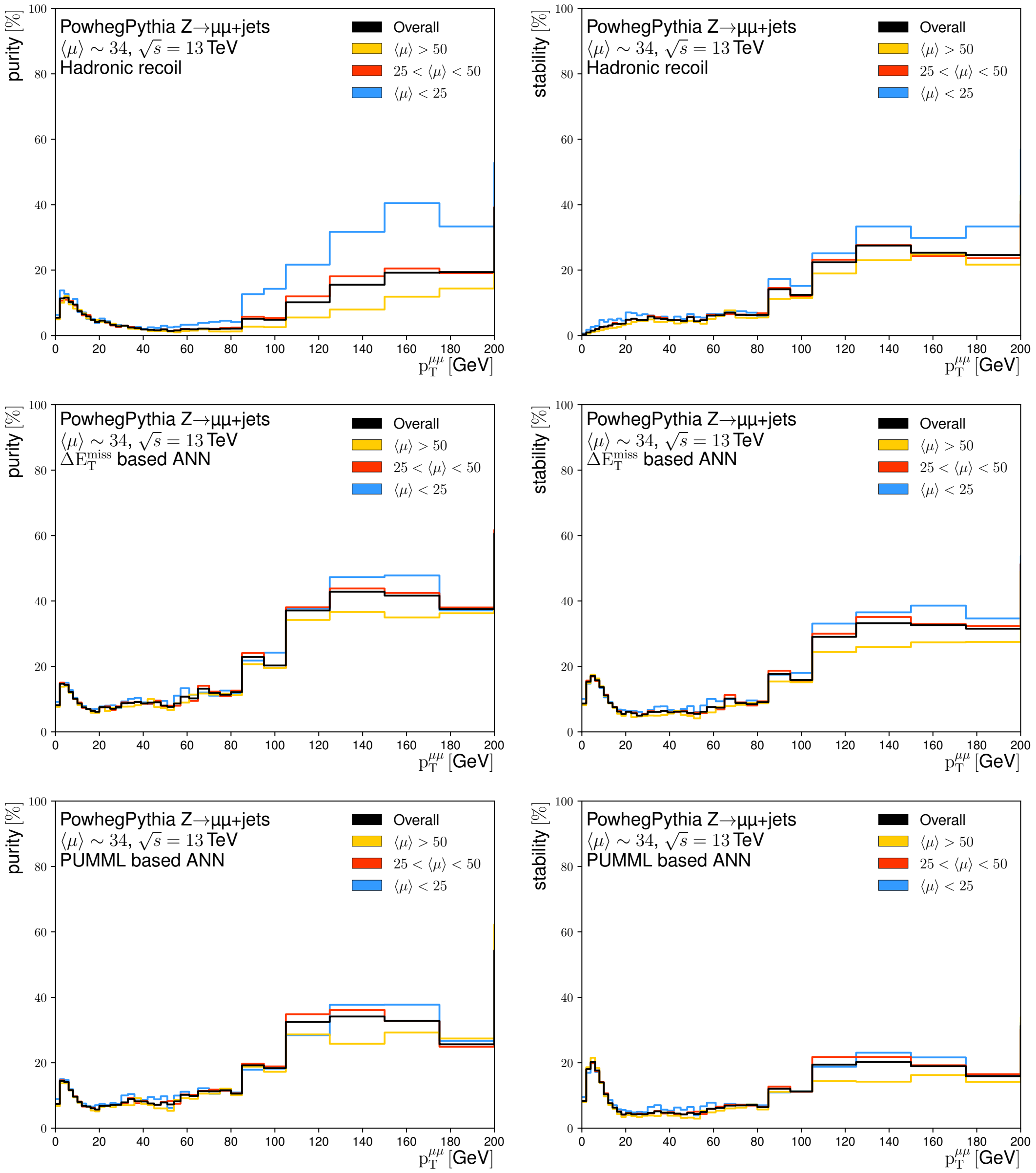


Figure 8.12: Purity (left column) and Stability (right column) for the reconstruction of $p_T^{\mu\mu}$ using the hadronic recoil (upper row) as well as the ΔE_T^{miss} (middle row) and PUMML based (lower row) ANN approaches. The variables are measured for three regions of pileup activity, defined by the corresponding $\langle\mu\rangle$ region.

Figure 8.12 shows the stability of the $p_T^{\mu\mu}$ measurement using the hadronic recoil as well as the ANN approaches for different regions of pileup activity. Some key points can be learned from these. Both the ANNs have higher stability and purity compared to the hadronic recoil measurement with the global loss function approach performing better than the local loss function approach. Furthermore, the ANN results show less dependency on pileup activity. While there is some improvement visible, stability and purity around 20 % are not good enough to be used in an analysis project as it still introduces a strong model dependence. The impact of the purity on measurement results is further discussed in Section 9.12.

The performances of both the ANNs are ultimately limited by the particle-level images' ability to describe the hard-scatter process, as the ANNs aim to recreate these images. This ability can be studied by measuring the purity and stability of the input images in bins of particle-level $p_T^{\mu\mu}$. Results of such a study are shown in Figure 8.13, emphasizing different aspects of the hard-scatter particle-level event images like the number of bins and selection criteria.

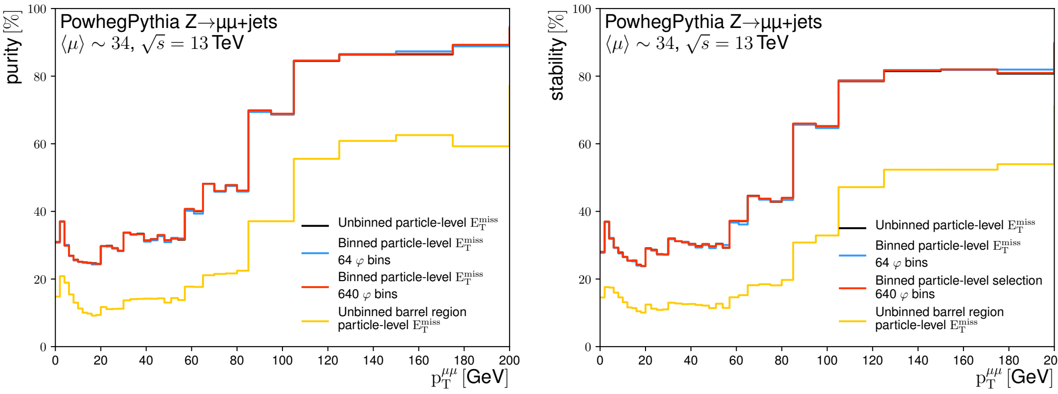


Figure 8.13: Purity and stability of different particle-level E_T^{miss} approaches in bins of particle-level $p_T^{\mu\mu}$. Stability and purity are shown for the unbinned particle-level E_T^{miss} without any η cut, two binned versions of this using 64 and 640 bins in φ as well as the unbinned E_T^{miss} limited to the barrel region $|\eta| < 2.5$. While the choice of binning has some minor impact, limiting the input to the barrel region has the biggest impact on these metrices. The overall particle-level selection to use any stable particle already results in a purity and stability of only $\approx 30\,\%$ in the lower bins. These values rise with higher bins, which is expected as the bins get bigger with higher $p_T^{\mu\mu}$.

The drop in purity and stability are mostly caused by three critical technical reasons rather than intrinsic limitations of deep artificial neural networks. These reasons are the particle-level selection criteria, the restriction of the ANN to the barrel region and the particle-level image binning, the latter being the weakest limitation.

As one can see, even the unbinned particle-level $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ does not result in a 100 %, but ≈ 30 % purity and stability, showing a lot of room for improvement to similarize these values. Building up on this, using only 64 bins in φ shows some minor differences to the unbinned $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$. This effect vanishes using 640 bins in φ. The particle-level selection presented in this chapter uses all stable born particle-level particles without any further selection criteria but does not describe $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ well. Hence, using a more specialized selection procedure can result in higher purity and stability.

The most substantial limitation is the selection of only particles present in the barrel region, causing the most significant drop in stability and purity to below 20 %. Hence extending the event images to the forward region and a redefinition of the image content can significantly improve the results of this approach to ANN based pileup mitigation. Until these problems are solved, better-suited approaches to measure the W boson mass are available.

As discussed in Chapter 7, another approach to the W boson mass measurement is to measure the cross-section of the Drell-Yan lepton-pair production cross-section to tune MC generators. As this decay is kinematically similar to the leptonic W boson decays, this can also lead to a decrease in model uncertainty. Such a measurement is presented in the following chapter.

Chapter 9

Measurement of the Drell-Yan lepton pair production cross-section

Contents

Looking at the kinematics of the Drell-Yan process, no transverse momenta of the Z boson is predicted at leading order calculations. Higher-order calculations take ISR into account and hence allow for a transverse momentum of the Z boson. Figure 4.4 shows exemplary Feynman diagrams of leading order and next to leading order Feynman diagrams, leading to one jet in the final-state, against which the Z boson recoils to balance its momentum. Next-to-next-to-leading order calculations in α_s are expected to lead to an improved description of large transverse momenta of the Z boson. However, to correctly describe the momentum range between 0 and 30 GeV, more sophisticated theoretical methods have to be deployed.

There are several approaches to model the Z boson transverse momentum correctly in the low $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ regime ($\mathrm{p}_{\mathrm{T}}^{\mu\mu} < 30\,\mathrm{GeV}$). On the one hand, parton shower models [116] can be matched to NLO order calculations, providing a good description of vector boson transverse momentum spectra, as a large variety of underlying parameters can be tuned to describe the measured data. On the other hand, resummation techniques [73] are expected to provide the most accurate description of the data at low momenta of the Z boson. Hereby resummation refers to calculating the impact of an infinite series of soft gluon emissions on the transverse momentum spectrum.

The measurement is performed using data collected in the years 2015 and 2016 by the ATLAS experiment at $\sqrt{s} = 13\,\mathrm{TeV}$ proton-proton collision at the LHC with an integrated luminosity $\mathcal{L} = 36.1\,\mathrm{fb}^{-1}$. The transverse momentum was measured using the measured spectrum of the dilepton system. In order to improve the precision of the measurement, the analysis was also performed in the electron decay channel and combined with the muon decay channel in a second step. The measurement in the electron decay channel was not performed within this thesis. The results of this project have been published as Ref. [54]. The full documentation of the analysis project can be found as Ref. [163]. Hence, the content of this chapter is based on these two references.

This chapter presents the measurement of the integrated and normalized differential fiducial cross-section of Drell-Yan lepton pairs. The Drell-Yan process describes the decay of a Z boson, or more generally Z/γ^*, into a pair of oppositely charged leptons. In the scope of this project, mainly the decay of Z bosons into a pair of muons is investigated. The measurement is performed in bins of the transverse momentum and the ϕ^*_η variable, which is defined as:

$$\phi^*_\eta = \tan\left(\frac{\pi - \Delta\varphi}{2}\right)\sin(\theta^*_\eta), \tag{9.1}$$

where $\Delta\varphi$ is the azimuth angle between the two leptons. θ^*_η is a measure of the scattering angle of the leptons to the proton beam direction in the rest frame of the dilepton system. It is defined by $\cos(\theta^*_\eta) = \tanh[(\eta^- \text{-} \eta^+)/2]$, with η^- and η^+ being the pseudorapidity of the negatively and positively charged lepton, respectively. Figure 9.1 shows the relation between ϕ^*_η and $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ measured in bins used for this analysis.

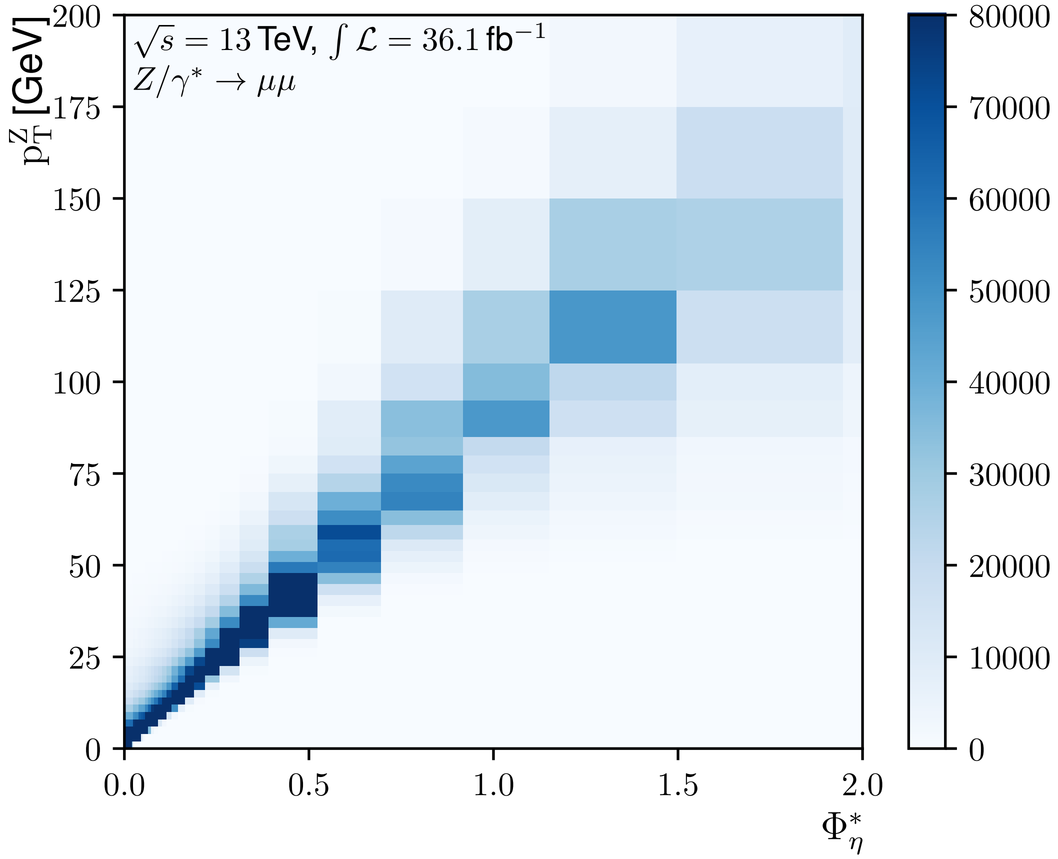


Figure 9.1: Relation between between ϕ^*_η and $\mathrm{p}_\mathrm{T}^{\ell\ell}$, measured in bins used for this analysis.

9.1 Analysis strategy

Generally speaking, an ATLAS data analysis can be split into the following steps:

1. Simulate the process under investigation and determine characteristic signatures in the recorded data. Find processes resulting in the same signature to estimate background activity.
2. Find suited triggers and kinematic cuts to select signal candidate events.
3. Process the data corresponding to the years under investigation.
4. Correct for differences between MC predictions and observed data.
5. Perform the analysis using both signal candidate events and relevant MC datasets.

These steps define the structure of the event selection process and are discussed below in further detail. The fiducial volume of this measurement is defined by two kinematic cuts applied to both the leptons and the boson system. The invariant mass of the reconstructed Z boson $m^{\mu\mu}$ has to be in the Z boson peak region between 66 GeV and 116 GeV, while the leptons are required to have a transverse momentum above 27 GeV and be in the ATLAS barrel region defined by $|\eta_\ell| < 2.5$.

The analysis is performed using data recorded during the second run of the LHC using proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ with a bunch spacing time of 25 ns during the runs of 2015 and 2016. Only data taken during stable beam conditions and with a fully operating magnet system, tracking and calorimeter are considered. The data must pass different data-quality requirements and corresponds to an integrated luminosity of $\mathcal{L} \approx 36.1\ \text{fb}^{-1}$ for the full 2015+2016 datasets. Dedicated MC simulation samples are used to describe background processes and the $Z \to \mu\mu$ process, the latter being used to correct for detector effects.

Measurements involving the $\text{p}_\text{T}^{\ell\ell}$ process, especially high precision measurements alike this analysis, require a precise understanding of both the beam conditions and the detector systems used to detect the lepton pair. Systematic uncertainties related to the particle reconstruction and measurement affect the detector resolution in $\text{p}_\text{T}^{\ell\ell}$ and hence limit the precision. The ϕ_η^* observable was introduced as an alternative probe of $\text{p}_\text{T}^{\ell\ell}$ [53], depending only on the lepton directions, to minimize the impact of the lepton resolution and scale uncertainties. The introduction of ϕ_η^* leads to a gain in precision as directions can be better measured within ATLAS compared to particle momenta.

9.2 Event reconstruction and selection

The selection in this analysis is straight forward, as the detector can fully reconstruct all relevant physics objects. The study focusses on the Z boson mass peak region between 66 and 116 GeV of the combined 4-vector boson mass of the invariant dilepton mass. The selections discussed below refer to the detector-level selection. Only a subset of these criteria are applied on particle-level and are indicated as such.

9.2.1 General event selection

All events entering the analysis are required to be recorded while the detector subsystems were operational and under nominal conditions. This selection is performed using a so-called **G**ood **R**un **L**ist (GRL). The GRL contains a list naming which data blocks to skip or to analyze depending on the detector conditions. Furthermore, all events are required to pass at least one of a set of triggers on detector-level. These triggers are defined depending on the data taking period, e.g., 2015 or 2016. In general, these triggers, presented in Table 9.1, are combined by a logical OR.

Table 9.1: Trigger chains used in the analysis

Year	2015	2016
triggers used	`HLT_mu20_iloose_L1MU15` `HLT_mu50`	`HLT_mu26_ivarmedium` `HLT_mu50`

For 2015, it is required to have at least one muon with 50 GeV of transverse momentum in the HLT or the combination of a 20 GeV loose quality muon in the HLT which is also present as a 15 GeV muon in the level 1 trigger. For 2016, due to the rise in pileup activity, the threshold changes to detecting any muon with at least 50 GeV or a 26 GeV transverse momentum muon with at least medium quality in the HLT.

Table 9.2: Muon trigger efficiencies

Year	Reconstructed bosons	Triggered events	Trigger efficiency
2015	1865162±1366	1796932±1340	96.34±0.10 %
2016	23240549±4821	22237984±4716	95.69±0.03 %

The efficiencies of these triggers were studied in the data used for this thesis and are summarized in Table 9.2. Reconstruction algorithms are applied to event data before accepting or rejecting it through the triggers listed above. If a Z boson can be reconstructed from data, the trigger corresponding to the run year is checked. The ratio of triggered events over reconstructed bosons is used as estimate for the trigger efficiency. It is found to be above 95 % for both 2015 and 2016.

Table 9.3: Cutflow of all data and MC samples. The steps are all following the definition of the fiducial volume and the $Z \rightarrow \mu\mu$ signature.

	Data	Powheg signal	electroweak background	photon induced background	$W^{\pm} \rightarrow l\nu$	$t\bar{t}$, single top
all events	246984511	77497800	9695207	566945	140993	12675994
rel all events [%]	100.00±0.01	100.00±0.02	100.00±0.05	100.00±0.19	100.00±0.38	100.00±0.04
Triggered events [%]	21.92±0.00	59.56±0.01	48.12±0.03	55.36±0.12	59.05±0.26	45.67±0.02
found Z candidate [%]	9.19±0.00	31.77±0.01	21.96±0.02	26.98±0.08	0.40±0.02	6.37±0.01
mass in peak region [%]	8.73±0.00	31.02±0.01	19.30±0.02	7.92±0.04	0.16±0.01	2.27±0.00
opp. charged muons [%]	7.35±0.00	26.49±0.01	16.36±0.01	7.11±0.04	0.01±0.00	1.36±0.00

9.2.2 Muon selection

The selection criteria for muons are applied in accordance to the recommendations of the ATLAS experiment and have to meet the following requirements:

- Medium Muon [17], i.e. a combined muon (a muon object using the combined information from the Inner detector and Muon system) [1] with at least 1 hit within the precision layers and a $\frac{q}{p}$-significance lower than 7.
- $\mathrm{p}_{\mathrm{T}}^{\mu} > 27\,\mathrm{GeV}$ and $|\eta_{\mu}| < 2.5$
- $d_0^{\mathrm{significance}} := \left|\frac{d_0}{\sigma(d_0)}\right| < 3$ with $\sigma(d_0)$ being the uncertainty of d_0.
- $|z_0 \cdot \sin(\theta)| < 0.5\,\mathrm{mm}$
- Gradient working point isolation [17], a working point that has an efficiency of 95 % for muons with $\mathrm{p}_{\mathrm{T}} = 25\,\mathrm{GeV}$ and 99 % for muons with $\mathrm{p}_{\mathrm{T}} = 60\,\mathrm{GeV}$.

The cuts on d_0 and z_0 assure that the leptons are coming from the hard scatter vertex. The isolation criterion serves the same purpose. In addition, the momentum resolution of combined muons is better than for muons recorded only within the muon spectrometer. On particle-level, only the $\mathrm{p}_{\mathrm{T}}^{\mu}$ and η cuts are applied to select muons.

9.2.3 Z boson selection

In order to select $Z \rightarrow \mu\mu$ events on detector- and particle-level, the following additional requirements are made, which are checked after identifying muon candidates:

- The event contains exactly two oppositely charged muons
- The boson invariant mass lies within $66\,\mathrm{GeV} < m_{\mu\mu} < 116\,\mathrm{GeV}$.

Table 9.3 shows the remaining events after each selection step.

---

[1] Medium includes also standalone muons outside the ID acceptance. These are rejected due to the cut on η. Section 9.6.1 presents the measurement principle of combined muons.

9.2.4 Particle-level selections

The selection of the different particle-levels is straight forward, as the levels of a particle-level object are provided by the generator, with one exception. On born-level, if the event does not contain two born-level muons, the bare-level muons are used. This procedure originates from the technical setup of MC generators, as muons if no FSR appears, are directly saved as stable, bare particles.

The goal of this study is to provide a combined measurement of the Drell-Yan lepton pair production differential and integrated cross-section using the measurements from both the electron and muon decay channels. Due to the differences of muons and electrons on bare particle-level, combined results are only presented for the born and dressed-level.

9.3 Background estimation

Dedicated kinematic selections and triggers are used to filter data collected with the ATLAS detector. Nevertheless, data of all processes resulting in a pair of oppositely charged muons within the fiducial volume are collected this way. The background contributions have to be estimated to extract Drell-Yan lepton pair production events from the collected data. These background processes include:

- Electroweak processes ($Z \rightarrow \tau\tau$, photon-induced and diboson processes)
- Top processes ($Z \rightarrow \bar{t}t$)
- Multijet processes

Nearly all processes, except for multijet contributions, contributing to this signature are modeled by MC simulations. The contribution of multijet processes, which refers to processes creating muons within jets, is estimated using a combination of MC datasets and a data-driven method. The following sections discuss these contributions in more detail, while the MC samples are presented in Section 9.4.

9.3.1 Electroweak background processes

The main background contributions are coming from electroweak processes, an umbrella term covering processes using the electroweak interactions. This covers Z bosons decaying into τs, which then decay into muons due to their short lifetime, as well as photon-induced processes, both shown in Figure 9.2. Furthermore, diboson processes involving two bosons decaying both hadronically and leptonically are covered. Hereby two bosons decay into a pair of leptons, a lepton-neutrino pair or two quarks, creating a signal including two opposite charged leptons, which are not necessarily created as a pair. The Feynman diagrams of same-type diboson processes are shown in Figure 9.3, while the Feynman diagrams of the remaining processes are shown in Figure 9.4.

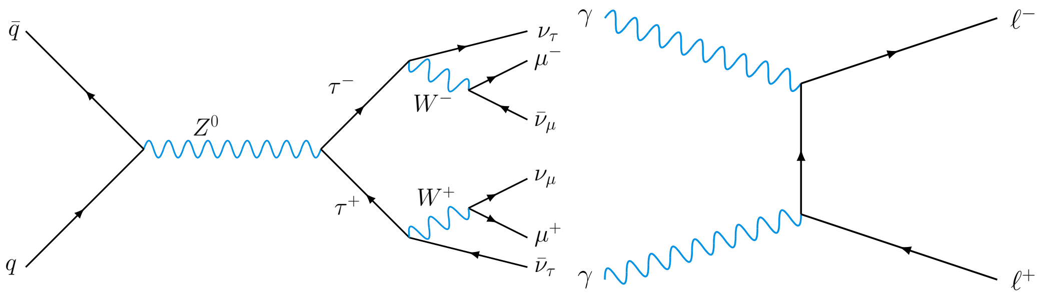


Figure 9.2: Exemplary leading order Feynman diagram of the $Z \to \tau\tau$ decay(left) and photon induced muon pair creation (right). Hereby two quarks annihilate and form a photon, that decays into a pair of tauons or muons. The tauons further decay into two opposite charged muons and four neutrinos.

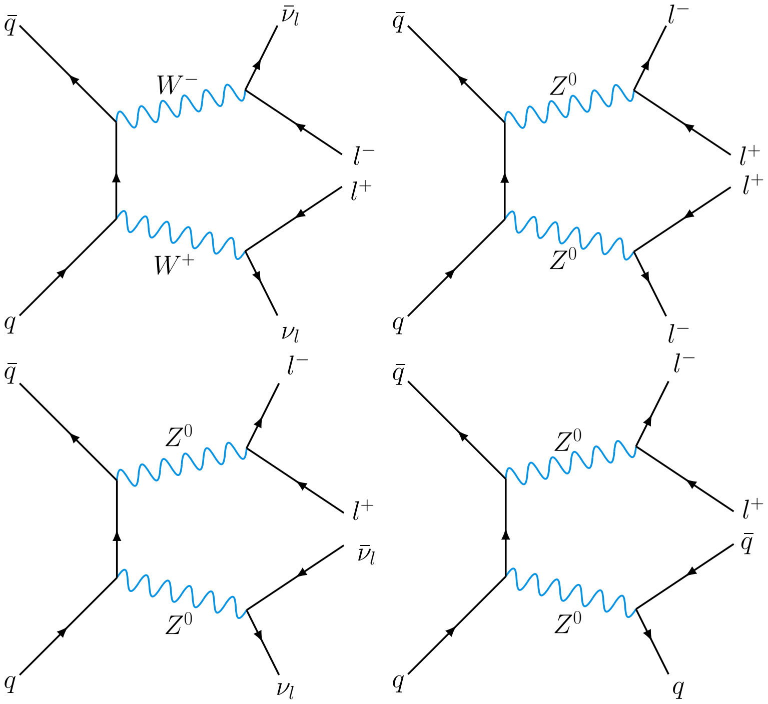


Figure 9.3: Feynman diagrams of same-type diboson processes. Diagrams of the processes $W + W \to l\nu + l\nu$ (upper left), $Z + Z \to ll + ll$ (upper right), $Z + Z \to \nu\nu + ll$ (lower left) and $Z + Z \to qq + ll$ (lower right) are shown.

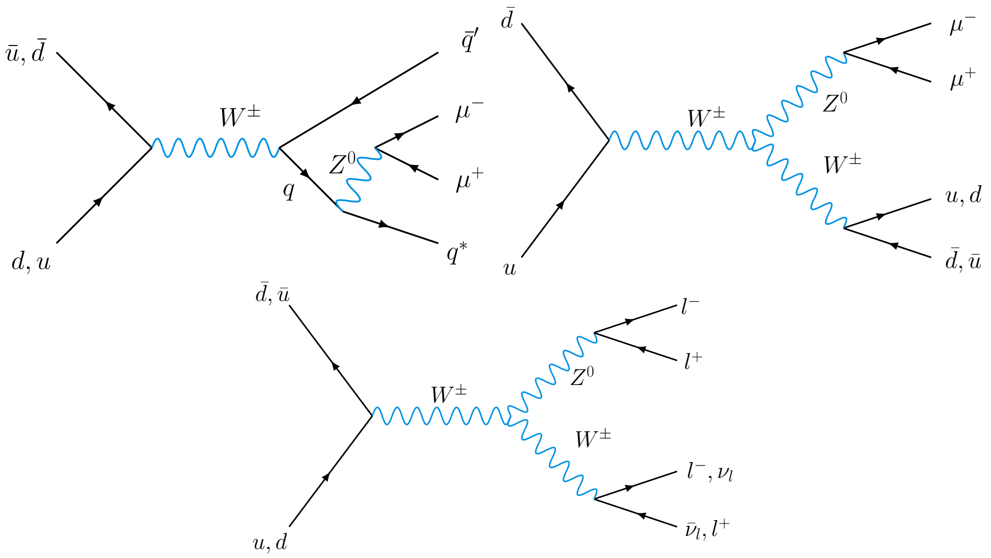


Figure 9.4: Feynman diagrams of diboson processes, namely $W \to qq, q \to Z \to ll$ (left), $W + Z \to qq + ll$ (middle) and $W + Z \to l\nu + ll$(right).

These diboson processes are:

- $W + W \to l\nu + l\nu$
- $Z + Z \to ll + ll$
- $Z + Z \to \nu\nu + ll$
- $Z + Z \to qq + ll$
- $W + Z \to l\nu + ll$
- $W + Z \to qq + ll$
- $W \to qq, q \to Z \to ll$

Hereby the processes $Z+Z \to ll+ll$ and $W+Z \to l\nu+ll$ has to be seen as special cases, as selecting events coming from this process require the rejection or misidentification of muons. Furthermore, the process $W + Z \to qq + ll$ is a mix between signal and background process. The leptons are created using the Drell-Yan process. Hence it should be seen as a signal sample. On the other hand, MC generators do not yet take this production channel into account, so it can also be defined as a background process. The impact of treating this process as a signal or as a background contribution is summarized in Table A.7. Within uncertainties, no difference can be observed. In the following, this processes is defined as background contribution.

9.3.2 Top physics background processes

The signature of Drell-Yan muon pairs can also be created by a pair of top quarks, denoted $g \rightarrow \bar{t}t$. Two annihilating quarks form a gluon that decays into a pair of top quarks. The top quarks then decay into bottom quarks by creating a W boson, which can then create lepton-neutrino pair, as shown in Figure 9.5.

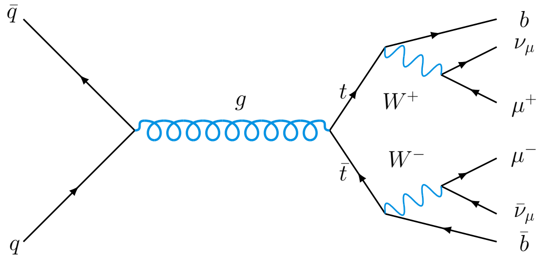


Figure 9.5: Feynman diagram of the $Z \rightarrow \bar{t}t$ process. After the creation of a top quark, this quark can irradiate a W boson, which then forms a bottom quark together with a lepton-lepton-neutrino pair.

9.3.3 Multijet background

As stated above, multijet backgrounds are processes that create at least one muon from jets due to hadronization. Multijet processes containing one prompt muon, a muon not created from hadronization, are estimated using MC datasets, while all other multijet contributions are estimated using a data-driven approach.

W boson + jets and diboson + jets contributions

In contrast to the processes discussed thus far, not only same-type quark pairs can create a muon pair signature in the detector. A charged W boson is formed in quark-antiquark annihilation processes with different quark types (e.g. $u\bar{d}$, $\bar{d}u$), which then decays into a lepton-neutrino pair of the same lepton family. As these processes, shown in Figure 9.6, create only one muon, another muon created from hadronization is needed to fake the Drell-Yan muon pair signature in the detector. W bosons decays resulting in τ leptons are also considered, as these can decay into muons:

- $W^+ \rightarrow \mu^+ + \nu_\mu$
- $W^- \rightarrow \mu^- + \bar{\nu}_\mu$
- $W^+ \rightarrow \tau^+ + \nu_\tau$
- $W^- \rightarrow \tau^- + \bar{\nu}_\tau$

The same argument holds true for the following diboson processes, shown in Figure 9.7:

- $W + W \rightarrow l\nu + qq$
- $W + Z \rightarrow l\nu + \nu\nu$
- $W + Z \rightarrow l\nu + qq$

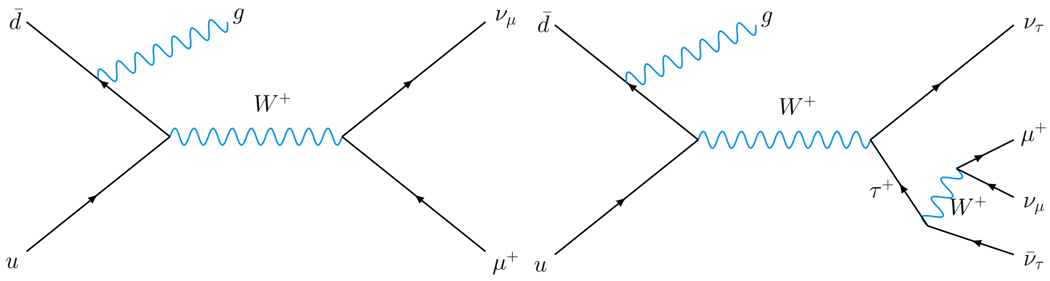


Figure 9.6: Leading order Feynman diagrams of W bosons decaying into a lepton-neutrino pair, shown for muons (left) and τ leptons (right). The τ can further decay into a muon and muon-neutrino.

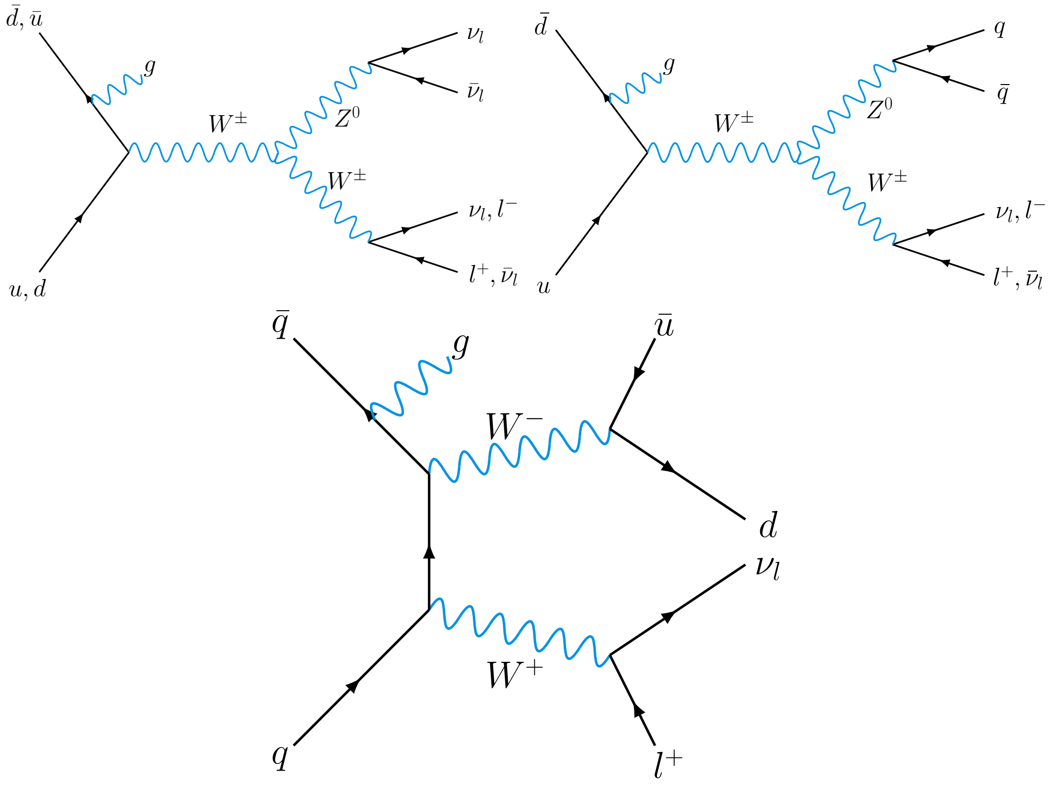


Figure 9.7: Feynman diagrams of the diboson + jets processes included in this analysis. Shown are $W + Z \to l\nu + \nu\nu$ (upper left), $W + Z \to l\nu + qq$ (upper right) and $W + W \to l\nu + qq$(bottom).

Single top + jets backgrounds

Single top processes can also create a muon pair signature in the detector if ISR creates a jet that includes a muon. Both the s-channel and the t-channel of these processes are of interest. The Feynman diagrams are shown in Figure 9.8.

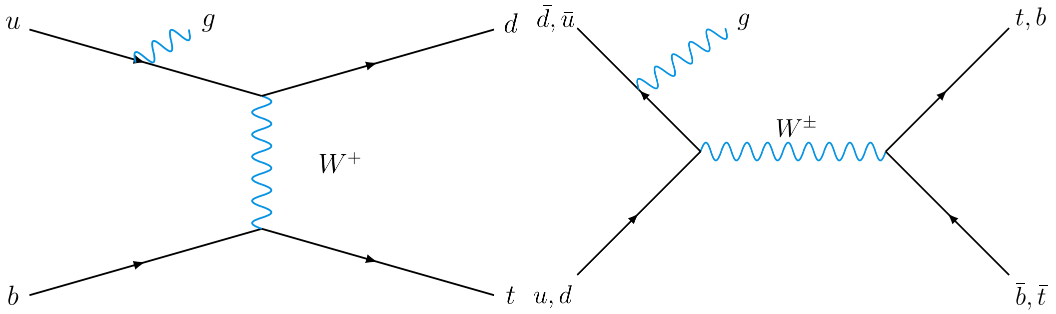


Figure 9.8: Feynman diagrams of single top + jet processes. Shown for both the s-channel, the exchange of a virtual W boson (left) and the t-channel, the annihilation of two quarks to form a W boson (right). After the creation of a top quark, this quark can irradiate a W boson, which then forms a bottom quark together with a lepton-lepton-neutrino pair. All leftover quarks will decay hadronically and form jets in the detector.

Quarks, e.g., up and bottom quarks, exchange a virtual W boson, resulting in a down and top quark in the s-channel. These will then further decay hadronically and form a jet. In the t-channel, up and down quarks annihilate to form a W boson decaying into a top-bottom quark pair. These can then create muons the same way as discussed for the $t\bar{t}$ process. Feynman diagrams of all these top processes are shown in Figure 9.8.

Data-driven estimation of remaining multijet background

Multijet background contributions created from processes without prompt muon, as exemplarily shown in Figure 9.9, are estimated using a data-driven approach. This is done using a so-called ABCD method. Two uncorrelated variables are chosen, the charge of the muons and the isolation of the muons. Four categories are defined using these variables, as listed in Table 9.4, from here denoted regions A, B, C, and D. Events fulfilling the region's criteria are counted to then ultimately estimate the multijet background.

Table 9.4: Regions used for the multijet background

region	opposite sign	same sign
isolated	A	B
non isolated	C	D

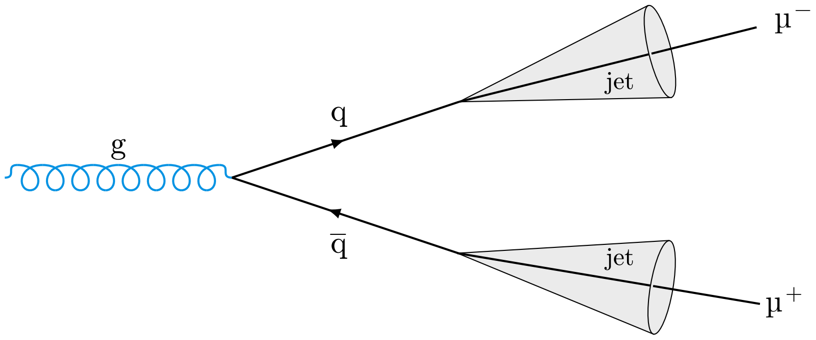


Figure 9.9: Two muons created within jets coming from two hadronizing quarks. These quarks are inititially created from a gluon.

One can assume that the ratio of multijet background in regions A and B will be the same as the ratio in regions C and D as the spatial isolation of muons within an event can be assumed independent of the muon charge. So one can estimate the multijet background in the signal region by:

$$\frac{n_A^{qcd}}{n_B^{qcd}} = \frac{n_C^{qcd}}{n_D^{qcd}} \Rightarrow n_A^{qcd} = \frac{n_C^{qcd}}{n_D^{qcd}} \cdot n_B^{qcd}. \tag{9.2}$$

The Poisson error $\sqrt{n}$ is used as the uncertainty of the number of events in each region. These uncertainties are propagated using Gaussian error propagation. The number of events in each region is measured for both data and MC samples. As this measurement aims to estimate the remaining multijet background not covered by any MC sample, the events seen in MC samples are subtracted from the events seen in data in each region. The number of data and MC background events in each of the regions is given in Table 9.5. This is done for two different $Z \to \mu\mu$ signal samples, the additional one using the Sherpa event generator. Within uncertainties, the numbers are consistent for both Sherpa and Powheg signal samples.

Table 9.5: Z→ $\mu\mu$ multijet background events. The Poisson error $\sqrt{n}$ is used as uncertainty for each region, while Gaussian error propagation is applied for the final event number uncertainties.

	Powheg			Sherpa		
sample	n_{ss}^{iso}	$n_{ss}^{\overline{iso}}$	$n_{os}^{\overline{iso}}$	n_{ss}^{iso}	$n_{ss}^{\overline{iso}}$	$n_{os}^{\overline{iso}}$
data	708.0 ± 26.6	9076.0 ± 95.3	179670.0 ± 423.9	708.0 ± 26.6	9076.0 ± 95.3	179670.0 ± 423.9
$Z \to \mu\mu$	13.5 ± 3.9	19.1 ± 6.9	141980.8 ± 487.9	18.4 ± 6.5	15.8 ± 4.6	127265.8 ± 1683.6
electroweak	312.3 ± 18.4	6.5 ± 1.0	246.8 ± 4.7	312.3 ± 18.4	6.5 ± 1.0	246.8 ± 4.7
photon induced	0.0 ± 0.0	0.0 ± 0.0	24.9 ± 8.0	0.0 ± 0.0	0.0 ± 0.0	24.9 ± 8.0
$W \to l\nu$	30.0 ± 22.4	0.0 ± 0.0	390.3 ± 245.1	30.0 ± 22.4	0.0 ± 0.0	390.3 ± 245.1
top processes	108.5 ± 6.7	873.2 ± 19.9	1472.7 ± 26.5	108.5 ± 6.7	873.2 ± 19.9	1472.7 ± 26.5
data corrected	243.6 ± 40.2	8177.2 ± 97.6	35554.5 ± 691.8	238.8 ± 40.5	8180.5 ± 97.4	50269.5 ± 1753.6
multijet events	1059.2 ± 176.3			1467.3 ± 254.6		

9.3.4 Overall background contributions

Looking at the number of events contributing to this analysis, as shown in Table 9.6, background processes yield less than 1 % contribution to the Z boson events.

Table 9.6: Inclusive background and data events for all samples under investigation.

Process	Selected events
data	18162641.0 ± 4261.8
electroweak	50063.8 ± 223.7
photon induced	4062.7 ± 63.7
$W \rightarrow l\nu$	409.5 ± 20.2
top processes	45411.2 ± 213.1
multijet	1059.2 ± 176.3

The main background processes are electroweak and top processes, every other process having a negligible contribution. These contributions are bin-wise shown in Figure 9.10 for the $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ and ϕ^*_η spectrum, showing the fractions of the overall signal. One can easily see that background becomes more dominant in the last few bins reaching up to 7 %, while in the first bins background contributions stay below 1 %. All other sources of background are negligible in comparison.

For better visibility, the backgrounds are summarized in the electroweak + top process group (also including the multijet MC samples grouped in this category) and the data-driven estimate of the remaining multijet background. Additionally, the multijet background is smoothed due to the lack of statistics available by averaging over the three neighboring bins in each direction for plotting purposes.

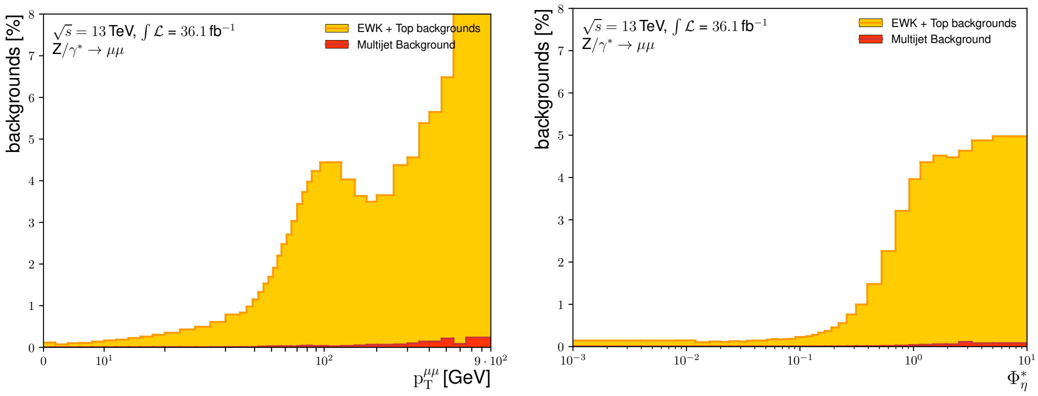


Figure 9.10: Background contributions shown as fractions of the overall signal for both the $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ spectrum(left) and the ϕ^*_η spectrum(right).

9.4 Monte Carlo datasets

This section lists the different MC samples used to estimate the signal Drell-Yan muon pair production process as well as the background contributions described above. The signal sample is used for the correction of detector effects such as inefficiencies, as described later in Section 9.12. The measured spectrum in the background samples is subtracted from the data sample to remove background contributions. Additionally, simulation samples can be systematically varied to estimate systematical uncertainties as well as probe the quality of the modeling of the physics processes and the detector, which are discussed in detail in Sectionsec:systematics.

The $Z \to \mu\mu$ signal process was generated with the Powheg-Box V1 MC event generator [6, 7, 95, 129] at next-to-leading order in α_s interfaced to Pythia version 8.186 [154] for the modeling of the parton shower, hadronization, and underlying event, with parameters set according to the AZNLO tune [15][2]. The CT10 (NLO) set of PDF [117] was used for the hard-scattering processes, whereas the CTEQ6L1 PDF set [139] was used for the parton shower.

Table 9.7: Summary of the MC simulation tunes used for this analysis.

Dataset	Generator	Type	PDF	Tune
$Z \to \mu\mu$, $Z \to \tau\tau$	Powheg +Pythia8	NLO Parton Shower	CT10 CTEQ6L1	AZNLO
top processes	Powheg +Pythia8	NLO Parton Shower	CT10 CTEQ6L1	A14
photon induced	Pythia8	Parton Shower	NNPDF2.3	
Diboson processes	Sherpa + Parton Shower	LO multi-leg matrix-element	NNPDF3.0 [23]	NNLO
	Powheg +Pythia8	NLO Parton Shower	CT10 CTEQ6L1	AZNLO

The effect of final-state photon radiation was simulated with Photos++ v3.52 [80, 103]. The EvtGen v1.2.0 program [77] was used to decay bottom and charm hadrons.

Powheg+Pythia8 was also used to simulate the majority of the background processes considered. The $Z \to \tau\tau$ and the diboson processes WW, WZ and ZZ [130] used the same tune and PDF as the signal process. The $t\bar{t}$ and single-top-quark [96, 142] backgrounds to the dimuon channel were simulated with POWHEG+PYTHIA8 with the A14 tune [12] and the NNPDF3.0 PDF [23]. The photon-induced background $\gamma\gamma \to \ell\ell$ was generated with Pythia8 using the NNPDF2.3 QED PDF [25].

Two diboson samples, namely $WqqZll$ and $ZqqZll$ with all having a Z boson decaying into two muons, were created using the Sherpa v2.2.1 [32] MC generator, a multi-purpose MC generator.

---

[2] The AZNLO tune is optimized to the low $\mathrm{p_T}$ bins of the $\mathrm{p_T^{\mu\mu}}$ measurement at $\sqrt{s} = 7\,\mathrm{TeV}$

Special emphasis is put on matrix- element and parton-shower merging, the latter being based on the Catani-Seymour dipole factorization [152]. Looking at the signal creating leptons, both real and virtual QED corrections are taken into account [149]. Additionally, this simulated process is combined with parton-level calculations of different jet multiplicity on the next-to-leading order. For this, the NNPDF3.0 PDF set is used. The shower hadronization is calculated using a phenomenological cluster-hadronization model.

The effect of multiple interactions in the same and neighboring bunch crossings (event pileup activity) was modeled by overlaying the hard-scattering event with simulated minimum-bias events generated with Pythia version 8.186 using the MSTW2008LO set of PDFs [122] and the A2 tune [19]. The description of the datasets is taken from Ref. [54]. and a summary of the generator tunes and PDF sets can be found in Table 9.7.

MC samples are created to fit a wide range of applications rather than to fit a specific dataset as computation time is limited. To describe the amount of events observed in data, MC samples are scaled by a global weight w, defined as the ratio of the luminosities of data $\mathcal{L}^{\text{data}}$ and simulation sample $\mathcal{L}^{\text{MC}}$:

$$w = k \cdot \frac{\mathcal{L}^{\text{data}}}{\mathcal{L}^{\text{MC}}}, \tag{9.3}$$

multiplied by a so-called k-factor to correct for higher order effects not included in simulation. The MC sample luminosity is defined as

$$\mathcal{L}^{\text{MC}} = \frac{\sum \text{weights}}{\sigma_{\text{sample}}}. \tag{9.4}$$

Hereby σ_{sample} is the cross-section of the simulated process and $\sum$ weights the sum over all weights coming from the MC generator. As MC do not perfectly describe both the real-life detector response as well as run conditions, These weights are used to correct for a variety of differences observed in detector performances and run conditions. These are discussed in the following Sections 9.5 and 9.6.

9.5 Efficiency corrections in Monte Carlo simulations

Several aspects of the detector and the underlying processes can not be modeled with sufficient precision. The muon reconstruction, isolation, trigger, and track-to-vertex association efficiencies are examples of such aspects directly related to this analysis. Measuring these efficiencies in data and simulations leads to the observation of non-negligible differences, which hence have to be corrected by scale factors. These scale factors are defined as the ratio of a certain efficiency in data over the corresponding efficiency in MC simulations:

$$SF = \frac{\epsilon^{\text{Data}}}{\epsilon^{\text{MC}}}. \tag{9.5}$$

The scale factors are calculated by analyzing $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events in both data and simulation using a tag-and-probe method. Each scale factor has a dedicated uncertainty, which is described in Section 9.9.

9.5.1 The Tag-and-Probe method

Processes resulting in a muon pair are used to perform efficiency measurements because of the clean signature in the detector. These are performed using a tag-and-probe method [17, 52, 58], visualized in Figure 9.11. At first, events with two muon candidates are selected. The events and the corresponding particles have to fulfill a set of criteria, ensuring the two muons originate from the process under investigation, e.g., a decayed Z boson. The exact composition of the selection criteria depends on the scale factor under investigation and is discussed in the corresponding section. In general, one muon (the tag muon) has to fulfill stricter criteria, as it is used to define suited events, compared to the second muon, called the probe muon. The efficiency can be calculated by:

$$\epsilon = \frac{N_R^{Signal} - N_R^{Bkg}}{N_P^{Signal} - N_P^{Bkg}}. \tag{9.6}$$

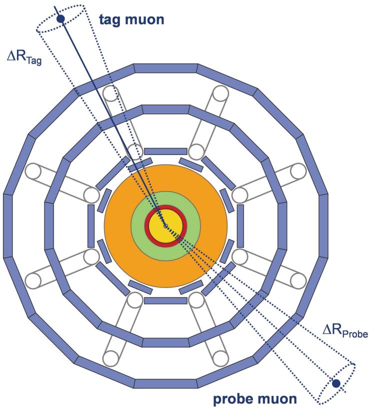


Figure 9.11: Visualization of the tag and probe method, taken from Ref. [111].

N_R denotes the number of successfully reconstructed events, passing all selection criteria, while N_P denotes the overall number of probes performed. The number of background events has to be subtracted from the number of signal events to obtain the correct number of events to compute the efficiency. The determination of the number of background events is defined differently for each dataset under investigation. Ref. [58] provides the full definition of the background processes.

A visualization of the effect of different scale factors is shown in Figure 9.12, while Table 9.8 lists typical values of these corrections.

9.5.2 Muon reconstruction efficiency

The measurement of the reconstruction efficiency [58, 109, 147] is performed for both the ID and the MS in bins of η. The tag muon requires to be a medium muon triggering the event. The probe selection depends on the detector system under investigation. If investigating the muon system, the probe is selected to be an ID track or an ID track with calorimeter tagging. When probing the ID, the probe has to be a MS track.

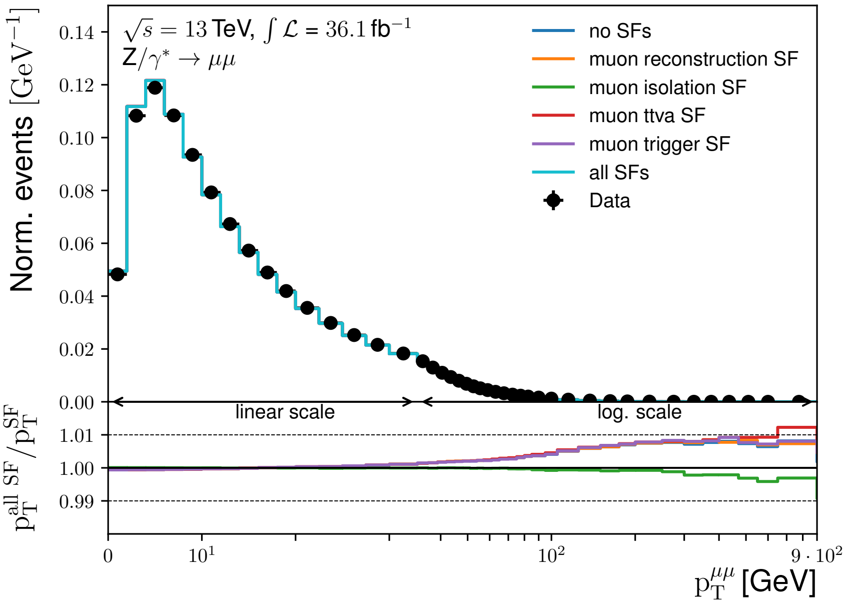


Figure 9.12: The effect of applying different scale factors (SFs) is shown together with the $p_T^{\mu\mu}$ distribution observed in data (upper part) and in relation to the application of all scale factors (lower part).

For both detector systems, the track is considered successfully reconstructed, if a reconstructed muon is found in a cone of $\Delta R < 0.05$ around the probe track.

9.5.3 Muon isolation efficiency

The muon efficiency isolation is measured in data and simulation in $Z \to \mu\mu$ decays, following the same procedure as for the muon reconstruction efficiency measurement with some additional cuts [58]. The probe muons need to have a transverse momentum higher than 20 GeV and a ΔR above 0.4 to the closest jet to avoid probe muons being covered by a jet. Furthermore, the two muons need to be separated by $\Delta R^{\mu\mu} > 0.3$.

Table 9.8: Mean value and RMS of all muon efficiency scale factors.

scale factor	Mean value	Standard deviation
Muon reconstruction efficiency	0.9939	0.0086
Muon isolation efficiency	0.9945	0.0015
Muon ttva efficiency	0.9881	0.0061
Muon trigger efficiency	0.9870	0.1026

9.5.4 Muon trigger efficiency

The muon trigger efficiency is measured using the tag-and-probe method described above in both data and simulation of the $Z \to \mu\mu$ process. It corrects for the false association of reconstructed muons as trigger objects [60].

In this case, the tag muon is required to be a combined, isolated muon firing the trigger under investigation. The probe muon, on the other hand, has to be a combined and isolated muon. Furthermore, the muons need to be of opposite charge and have an invariant mass of the two muons close to the Z peak. As the muon system only provides triggers to $|\eta| < 2.4$, this limitation is also applied to the muons. The probe muon is considered successfully reconstructed if a trigger object can be matched with it.

9.5.5 Muon track-to-vertex-association efficiency

Muons used in precision measurements, like the one described here, are required to pass the so-called track-to-vertex association requirements defined by $|z_0| \cdot \sin\Theta < 0.5\,\mathrm{mm}$ and $|d_0|/\sigma(d_0) < 3$, ensuring the muons are coming from the primary vertex. As these cuts, which are affected by the reconstruction of primary vertices as well as the measurement of d_0 and z_0, are independent of the muon identification, the track-to-vertex association efficiency [110] has to be measured separately, using the tag-and-probe method on $Z \to \mu\mu$ events in both simulation and data.

The selection of the tag muon is identical to the one described in the muon reconstruction efficiency measurement section above. Probe muons have to satisfy the loose identification criterion [58] and have to pass the $|\eta| < 2.5$ and $\mathrm{p_T} > 10\,\mathrm{GeV}$ cuts.

9.6 Muon transverse momentum calibration in simulation

Not only efficiencies but also the measurement of the transverse momentum has to be calibrated to describe the observed distributions in data better. Although the ATLAS detector is accurately modeled, the precision of the simulation is not sufficient to describe both the muon momentum scale and the momentum resolution to a level sufficient for precision measurements. Such measurements require a description accurate on per mille level for the muon momentum scale and percent level for the resolution [58]. The muon momentum is recalibrated as described below and outlined in more detail in Ref. [58] to achieve this level of precision. The corrections applied in this calibration procedure to the muon momentum scale and resolution is studied by measuring the $\mathrm{p_T}$ and η distributions in both the $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ datasets. The corrections are obtained by reweighting the distributions observed in MC to the ones observed in data.

9.6.1 Transverse momentum measurements

As the ID does not differentiate between types of particles when reconstructing the corresponding track by the procedure outlined in Ref. [65, 69], this section focusses on the description of the measurement in the muon system and the combination of the muon system and ID information.

In the muon system, the transverse momentum is measured by the curvature of the muon track due to the present magnetic field B. MDT chambers measure the sagitta of the curvature defined by hits in three MDT chambers (and shown in Figure 9.13) to reconstruct the muon track and determine the muon transverse momentum. As shown in Ref. [150], the sagitta s is defined by the reconstructed track segment of length L and angle α with radius R as:

$$s = r\left(1 - \cos\frac{\alpha}{2}\right) \approx r\frac{\alpha^2}{8} \approx \frac{1}{8}\frac{L^2 B}{\mathrm{p_T}}. \tag{9.7}$$

From there, the transverse momentum can be obtained by:

$$\mathrm{p_T} \approx \frac{1}{8}\frac{L^2 B}{s}. \tag{9.8}$$

The measurement of the transverse momentum is hereby mostly limited by the uncertainty of the magnetic field ΔB and the uncertainty of the sagitta Δs.

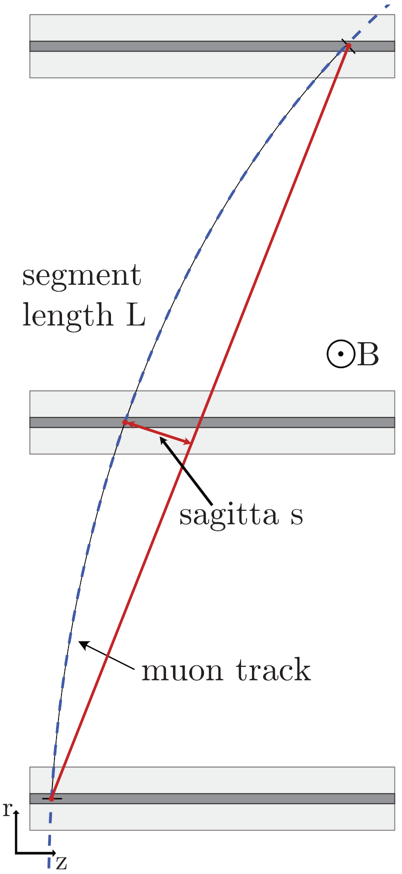


Figure 9.13: Muon track reconstruction.

In reality, not only muons but also tails of hadronic showers, cavern background, or electronic noise create signals in the muon detector, resulting in a multitude of induced signals in the MDT chambers. Due to the sheer amount of possible combinations of detector hits, not all signals in different MDT chambers can be analyzed in a reasonable time, but suitable candidates are selected by hit pattern recognition algorithms.

Furthermore, the muons are distracted due to multi-scattering in the detector material, and the magnetic field is not constant over the full detector. Hence, the track parameters are obtained by a global fit, which takes all of these factors into account, as described in Ref. [58].

This fit is used to extrapolate the muon track back to the ID, and finally, the interaction point from where it originated. The muon tracks are reconstructed independently in both the ID and the MS to obtain a combined muon transverse momentum measurement. Afterward, a combined track is formed by a global refit using the data from both detector systems. Not all muon system tracks have to be used during this procedure but can be added or removed to improve the fit quality.

9.6.2 Muon transverse momentum calibration

The correction is applied to both ID and MS measurements, as these systems are used to measure combined muon properties. Hence only combined muons are used for the calibration procedure, which utilizes the track of such a combined muon. The track consists of both an ID component $\mathrm{p_T^{ID}}$ and a muon system component $\mathrm{p_T^{MS}}$, which are reconstructed by an extrapolation procedure from detector system hits and the production vertex position.

The uncorrected transverse momentum observed in MC $\mathrm{p_T^{MC,\,Det}}$ is corrected for each detector system, namely ID and MS, by:

$$\mathrm{p_T^{Cor,\,Det}} = \frac{\mathrm{p_T^{MC,\,Det}} + \sum\limits_{n=0}^{1} s_n^{Det}(\eta,\varphi)\left(\mathrm{p_T^{MC,\,Det}}\right)^n}{\sum\limits_{m=0}^{2} \Delta r_m^{\mathrm{Det}}(\eta,\varphi)\left(\mathrm{p_T^{MC,\,Det}}\right)^{m-1} g_m}, \tag{9.9}$$

with g_m being normally distributed random variables with zero mean and unit width. Furthermore, $\Delta r_m^{\mathrm{Det}}(\eta,\varphi)$ describes the momentum resolution smearing, while $s_n^{Det}(\eta,\varphi)$ describes the scale corrections. Both are applied depending on the region within a specific $\eta - \varphi$ grid of the ATLAS detector, defined as such that bins are homogeneous in terms of detector technology and performance.

Both $\mathrm{p_T^{Cor,\,ID}}$ and $\mathrm{p_T^{Cor,\,MS}}$ are used to calculate the corrected combined muon transverse momentum $\mathrm{p_T^{Cor,\,CB}}$ as:

$$\mathrm{p_T^{Cor,\,CB}} = f \cdot \mathrm{p_T^{Cor,\,ID}} + (1-f) \cdot \mathrm{p_T^{Cor,\,MS}}. \tag{9.10}$$

Assuming that the relative contributions of the two detector systems remain unchanged under recalibration, f can be calculated utilizing the uncalibrated transverse momentum $\mathrm{p_T^{MC,\,CB}}$ of the combined muon and of the two detector systems $\mathrm{p_T^{MC,\,ID}}$, $\mathrm{p_T^{MC,\,MS}}$ by:

$$\mathrm{p_T^{MC,\,CB}} = f \cdot \mathrm{p_T^{MC,\,ID}} + (1-f) \cdot \mathrm{p_T^{MC,\,MS}}. \tag{9.11}$$

9.7 Z vertex reweighting

As the MC samples are generated to fit a wide range of applications, the samples have to be reweighted to describe the data samples under investigation. Most of this is done by tools provided by the ATLAS community and affect different scale factors, as described above.

One reweighting procedure that has been performed manually in the context of this analysis is the so-called z vertex position reweighing. The primary vertex z-position distributions for data and MC samples are measured and compared.

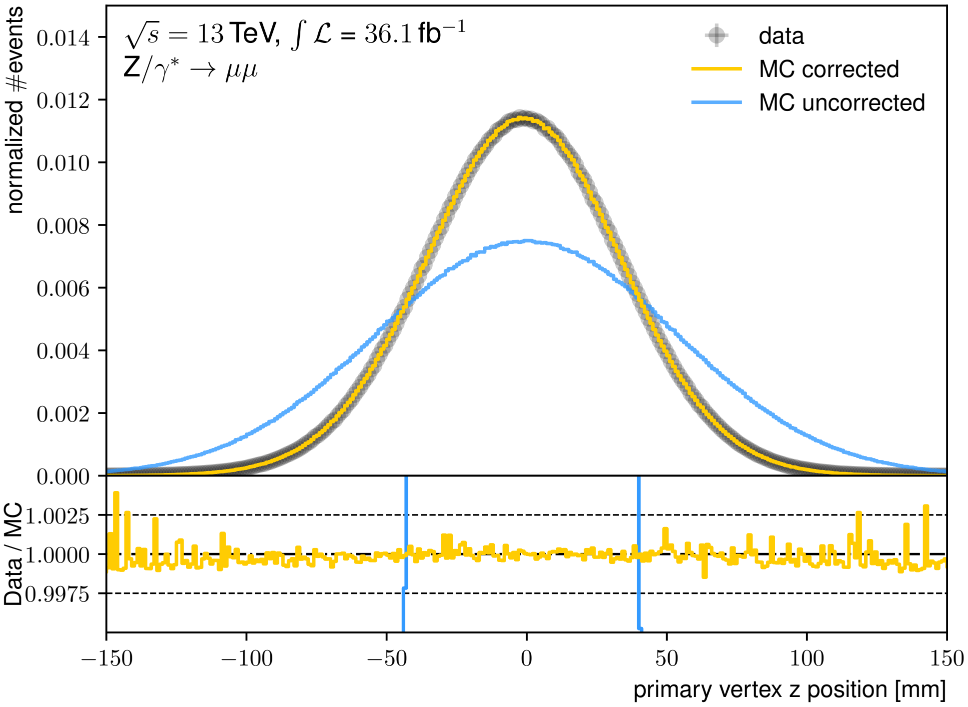


Figure 9.14: Comparison of the data and MC z vertex position distributions for both reweighed and original MC samples.

In a second run, which only affects the MC samples, an additional event weight depending on the z vertex position is introduced. It is defined as the ratio of the normalized z-vertex position distributions and given by:

$$w_{zpos}(\mathrm{pos}_{\mathrm{vertex}}) = \frac{\mathrm{zpos}_{data}(\mathrm{pos}_{\mathrm{vertex}})}{\mathrm{zpos}_{MC}(\mathrm{pos}_{\mathrm{vertex}})}, \tag{9.12}$$

where $\mathrm{pos}_{\mathrm{vertex}}$ describes the z vertex position. Applying this weight leads to nearly identical distributions of the primary vertex z-positions with only small statistical fluctuations. Figure 9.14 shows the result of this procedure.

9.8 Pileup reweighting

A dedicated pileup reweighting tool [38] provided by the ATLAS community is used to model the pileup activity of the collected data in the corresponding run periods under investigation. This tool calculates an event-based pileup weight that is applied to all distributions to reweight the MC sample, as these are simulated using a flat distribution of $\langle\mu\rangle$. A dedicated pileup uncertainty is also described in Section 9.9.4.

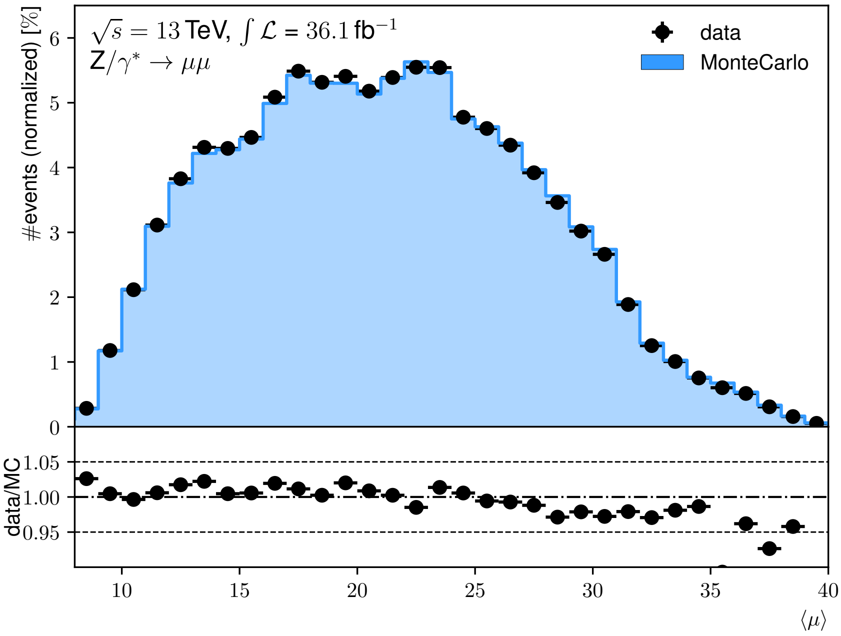


Figure 9.15: Average interactions per bunch crossing $\langle\mu\rangle$ for both data and the Powheg MC signal sample. In order to model to model the pileup activity of the data, a pileup event weight is applied.

The resulting distributions of the average interactions per bunch crossing $\langle\mu\rangle$ are shown in Figure 9.15. They agree reasonably well, differing less than 5 % in the central region.

9.9 Systematic uncertainties

Every measurement is affected and hence limited by systematical and statistical uncertainties. Generally speaking, systematic uncertainties are all uncertainties that are not directly related to the limited number of data events but represent intrinsic methodological limitations. These can originate from different sources. Examples are the missmodeling of the detector, beam conditions, or physics processes in the MC simulation, the correction of detector effects, or the missmodeling of pileup effects.

While dedicated methods are applied for the statistical uncertainties, which are described in the corresponding sections, the systematic uncertainties incorporated in this analysis are mostly taken from recommendations provided by ATLAS analysis tool developers. This is also true for physics objects (e.g., leptons) selection and most event-level corrections. Self-defined estimations of uncertainties are applied for both the pileup modeling as well as the Z vertex reweighting. The systematic uncertainties are split into experimental and theoretical systematics.

9.9.1 Treatment of systematic uncertainties

Systematic uncertainties are provided by ATLAS analysis tools as a set of two variations, called σ_{1up} and σ_{1down}. For both variations, the parameters affected by this systematic uncertainty are varied one standard deviation up and down, respectively. These variations are available for all MC samples. The full analysis process is performed again with the adapted values and the varied backgrounds subtracted from the measured data sample. The resulting distribution using the varied MC signal and background samples is then compared to the so-called nominal result $\sigma_{nominal}$ with no variation applied. The absolute difference of the varied and the nominal distribution is depicted as uncertainty. As each of these pairs is considered directly correlated, each systematic is combined bin-wise as following:

$$\sigma_{syst} = \text{mean}(|\sigma_{1up} - \sigma_{nominal}|, |\sigma_{1down} - \sigma_{nominal}|).$$

This holds for all systematic uncertainties except the pileup reweighting and background contamination. The details of the corresponding procedures are described below in the corresponding subsection. For the final result, systematic uncertainties are further combined to on the one hand packs of related uncertainties like statistical, modeling, lepton efficiency, and resolution uncertainties, but on the other hand, to the total uncertainty. Correlations between different uncertainties are taken into account.

9.9.2 Experimental uncertainties

Experimental uncertainties cover all uncertainties related to the detector. As ATLAS is a complex, modular detector system, measurements of particle properties are a combination of several steps, which all are affected by uncertainties. This is also true for the scale factors used to better match MC predictions with observed data as these are obtained from dedicated measurements.

Muon efficiencies

The analysis includes scale factors for the muon isolation (MUON_ISO) and muon reconstruction efficiencies (MUON_EFF), while Track to vertex association describes the efficiency of correctly assigning a track to an interaction vertex. For each one, the scale factors are systematically varied up and down by one standard deviation of the corresponding scale factor distribution.

As these efficiencies highly depend on the detector subsystems and their position within ATLAS, the detector is divided into a grid in the (η, φ) plane to decorrelate these efficiency uncertainties. This grid features ten regions in η, reaching from -2.5 to 2.5, and 6 regions in ϕ, reaching from $-\pi$ to π. The $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ and ϕ^*_η distributions are measured in each region separately, only applying the systematically varied efficiency scale factor in the region covering the muons (η, φ) coordinate. All other regions use the nominal scale factor value. The distributions obtained in all of these regions are then later unfolded, and the relative difference to the nominal unfolded result added in square.

Muon momentum calibration uncertainty

Muon ID and Muon MS variations smear the reconstructed transverse momentum of the ID track and the MS track by one standard deviation of its reconstruction process up and down, respectively. Hence the properties of the reconstructed combined muon change as well, leading to the systematically variated distributions.

Muon sagitta uncertainty

The muon sagitta systematic variation corrects for two effects that occur within the muon system. On the one hand, effects resulting from the misalignment of the muon detector chambers within ATLAS have to be corrected, as this limits the spacial resolution. On the other hand, charge-dependent contributions due to the magnetic field, in particular effects coming from defects in the magnet curl, have to be treated accordingly. These contributions are later quadratically summed.

9.9.3 Luminosity uncertainty

Both ATLAS and the LHC are using dedicated systems to measure the luminosity, which naturally is affected by uncertainties. The luminosity uncertainty is provided by the ATLAS luminosity group to be 2 % [66] for the 2015 and 2016 datasets. It is only used for the fiducial cross-section measurement.

9.9.4 Pileup uncertainty

A correction is applied to the MC samples such that the distribution of the average number of interactions per bunch crossing $\langle\mu\rangle$ in these samples better reflects the distribution seen in data. The 2015 data analysis team observed a double counting of isolation effects coming from both the isolation and the pileup reweighting tool, resulting in an overestimation of this uncertainty. The absolute difference between the result with and without pileup reweighting is used as uncertainty to avoid double-counting but to estimate a pileup uncertainty. Therefore the analysis is run a second time without pileup reweighting, resulting in a flat distribution in $\langle\mu\rangle$ as provided by the MC generators.

9.9.5 Z vertex position uncertainty

As Z vertex reweighting is applied, a dedicated uncertainty is assigned. The resolution of the z vertex position is better than 1 mm, so the reweighting spectrum is shifted by 1 mm up and down. The relative difference to the nominal is used as uncertainty.

9.10 Detector-level distributions

Applying all these corrections, the signal and background MC predictions should together describe the recorded data reasonably well. Furthermore, systematic variations of these MC datasets are used to estimate a binwise uncertainty of these MC predictions. From there, several detector-level distributions are measured and used as control mechanisms to investigate the agreement between MC samples and data.

Results coming from both MC signal and background samples are stacked. Ideally, the sum of all of these samples corresponds to the events observed in data in each bin. Figures 9.16 and 9.17 show muon control plots, while Figure 9.18 focusses on a variety of Z boson properties. All plots show the ratio of the normalized combined MC distributions and data distribution is shown in the lower panel of each plot. Detector-level systematic uncertainties are indicated as a grey error band in the ratio plot pannel. For plot purposes and due to a lack of statistics, the data-driven multijet background contributions are smoothed over the four neighboring bins.

Figure 9.17 shows the muon η distribution for all muons as well as separated by muon charge. These plots show good agreement between data and MC, confirming the application of muon efficiencies. This holds also true for the $m^{\mu\mu}$ and $y^{\mu\mu}$ distributions shown in Figure 9.18, all showing a difference in shape of below 5 %.

The higher bins of $\mathrm{p}_{\mathrm{T}}^{\mu}$, shown in Figure 9.16 as well as of ϕ^{*}_{η} and $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$, shown in Figure 9.18, are not well described by MC generators, showing differences of up to 20 %. These differences further emphasize the importance of such high-precision measurement, which can be used to tune MC generators to describe better the underlying physics processes leading to more realistic predictions.

As a further cross-check, the same distributions were measured using the *Sherpa* event generator. These plots can be found in Section A.5 in the appendix, as no further information is gained from these plots.

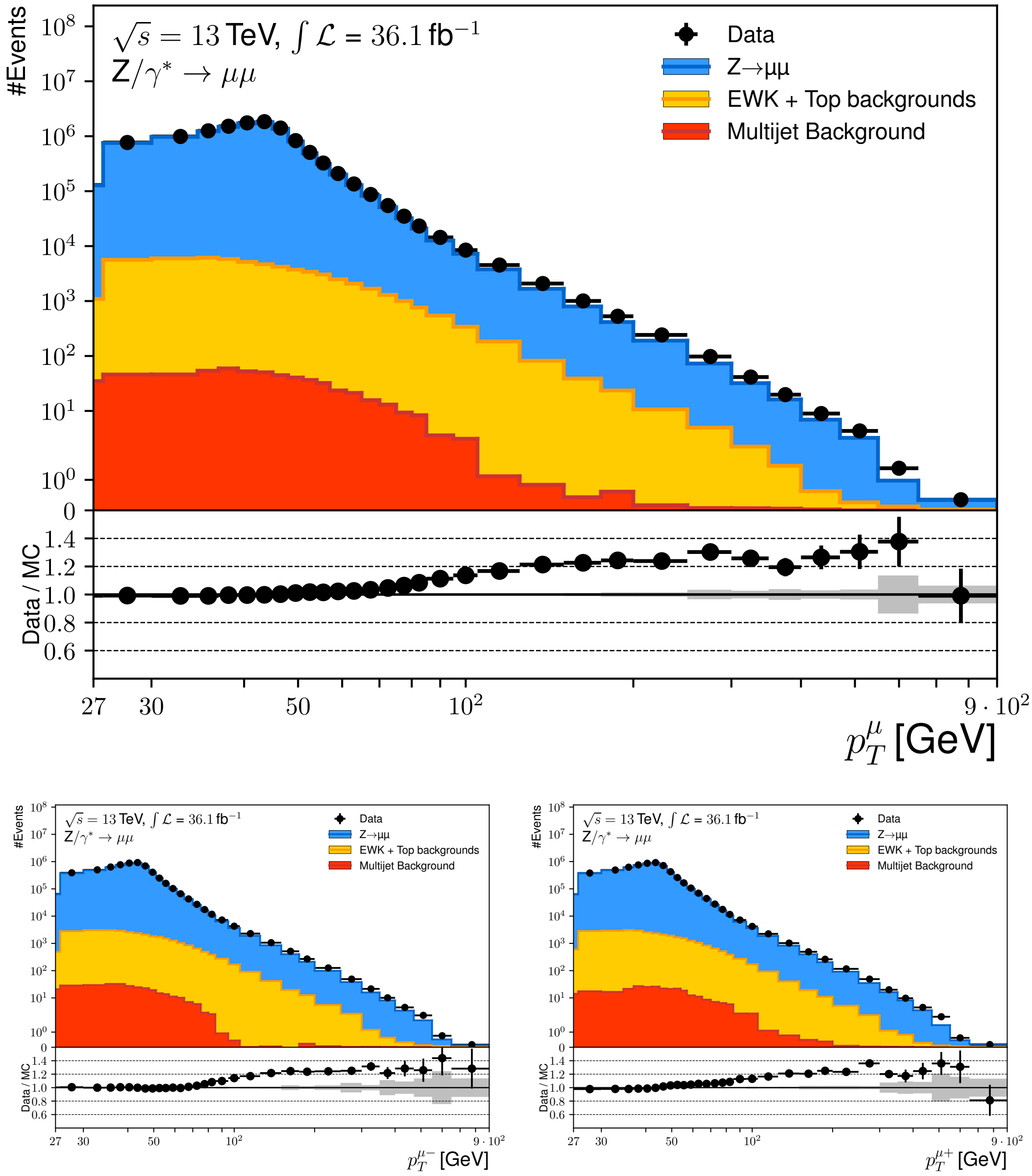


Figure 9.16: Powheg control plots of the muon transverse momentum p_T^μ (up) for all muons as well as $p_T^{\mu+}$ for μ^+ (lower left) and $p_T^{\mu+}$ for μ^- (lower right).

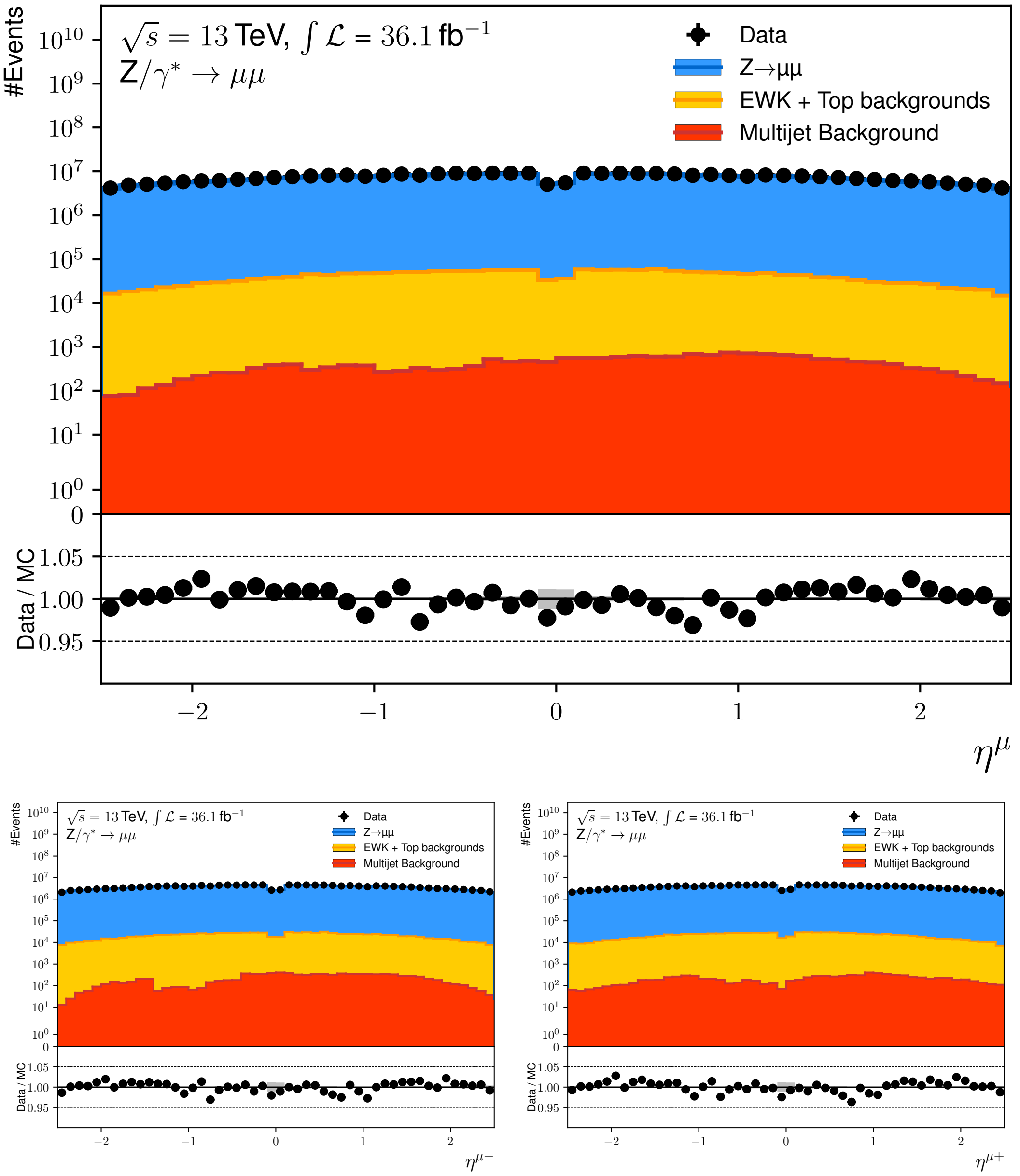


Figure 9.17: Powheg control plots of the muon pseudorapidity η_μ (up) for all muons as well as $\eta_{\mu+}$ for μ^+ (lower left) and $\eta_{\mu+}$ for μ^- (lower right).

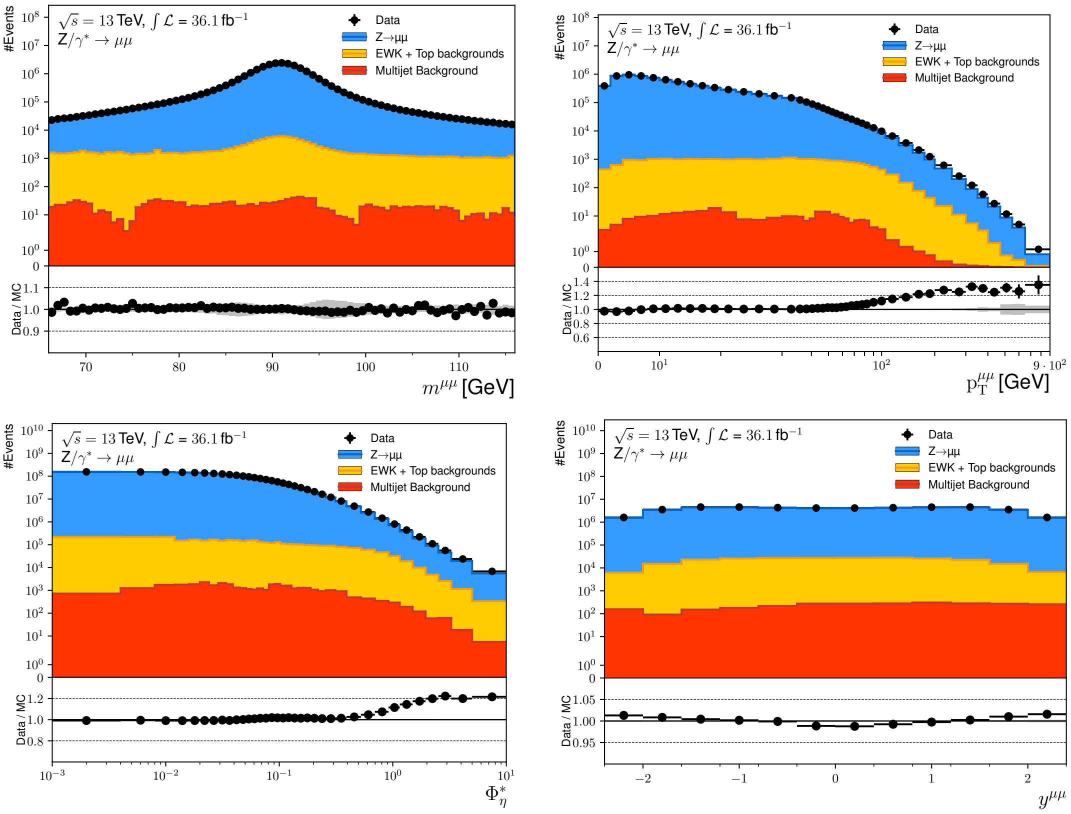


Figure 9.18: Powheg control plots for the Z mass $m_{\mu\mu}$ (upper left), Z transverse momentum $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ (upper right) as well as muon $\phi^*_{\mu\mu}$ (lower left) rapidty $y^{\mu\mu}$ (lower right). All plots show detector-level data. The MC samples are normalized as such that the MC luminosity fits the data luminosity.

Table 9.9: C-factors for different systematic variations and particle levels. The c-factor is obtained by the ratio of the number of events on detector-level over the number of events on particle-level. Shown are only the difference to nominal. *EFF* hereby refers to efficiency corrections, while *ISO* refers to muon isolation scale-factors. *STAT* and *SYS* refer to variations of statistical and systematic components of a certain scale-factor, respectively. Furthermore, *1up* and *1down* refer to up and down variations of the corresponding scale-factors.

systematic variation	C_Z^{born}	C_Z^{bare}	C_Z^{dressed}
nominal	0.6851	0.7147	0.7024
MUON EFF STAT 1down	-0.0026	-0.0027	-0.0027
MUON EFF STAT 1up	0.0026	0.0027	0.0027
MUON EFF STAT LOWPT 1down	0.0000	0.0000	0.0000
MUON EFF STAT LOWPT 1up	0.0000	0.0000	0.0000
MUON EFF SYS 1down	-0.0094	-0.0098	-0.0097
MUON EFF SYS 1up	0.0095	0.0099	0.0097
MUON EFF SYS LOWPT 1down	0.0000	0.0000	0.0000
MUON EFF SYS LOWPT 1up	0.0000	0.0000	0.0000
MUON EFF TrigStatUncertainty 1down	-0.0003	-0.0003	-0.0003
MUON EFF TrigStatUncertainty 1up	0.0003	0.0003	0.0003
MUON ID 1down	-0.0000	-0.0000	-0.0000
MUON ID 1up	0.0000	0.0000	0.0000
MUON ISO STAT 1down	-0.0002	-0.0002	-0.0002
MUON ISO STAT 1up	0.0002	0.0002	0.0002
MUON ISO SYS 1down	-0.0028	-0.0029	-0.0029
MUON ISO SYS 1up	0.0028	0.0029	0.0029
MUON MS 1down	-0.0000	-0.0000	-0.0000
MUON MS 1up	-0.0001	-0.0001	-0.0001
MUON SAGITTA RESBIAS 1down	-0.0000	-0.0000	-0.0000
MUON SAGITTA RESBIAS 1up	0.0000	0.0000	0.0000
MUON SAGITTA RHO 1down	-0.0001	-0.0001	-0.0001
MUON SAGITTA RHO 1up	-0.0001	-0.0001	-0.0001
MUON SCALE 1down	0.0004	0.0005	0.0005
MUON SCALE 1up	-0.0005	-0.0005	-0.0005
MUON TTVA STAT 1down	-0.0011	-0.0011	-0.0011
MUON TTVA STAT 1up	0.0011	0.0011	0.0011
MUON TTVA SYS 1down	-0.0008	-0.0009	-0.0009
MUON TTVA SYS 1up	0.0008	0.0009	0.0009
z vertex 1up	0.0002	0.0002	0.0002
z vertex 1down	-0.0004	-0.0004	-0.0004
Pileup	0.0020	0.0020	0.0020
error	0.0105	0.0107	0.0108

9.11 Measurement of the fiducial integrated cross-section

The fiducial integrated cross-section is obtained using Equation 9.13 with an additional factor C_Z, correcting for detector efficiencies, leading to:

$$\sigma_{fid} = \frac{N_{signal}}{\mathcal{L} \cdot C_Z}. \tag{9.13}$$

with $N_{signal} = N_{events} - N_{Background}$ referring to the number of signal events. It is calculated by subtracting the number of background events $N_{Background}$ from the number of events observed in data N_{events}. The C_Z-factor is defined as the ratio of the $Z \to \mu\mu$ events on detector-level, passing all cuts, and the $Z \to \mu\mu$ events on particle-level, passing the fiducial criteria.

This factor is calculated for all systematic variations, which are summarized in Table 9.9. Muon efficiency, muon isolation, and pileup scale-factor variations are the major contributions to the C_Z-factor uncertainties, as these directly affect the selection process. This is also true for muon track-to-vertex association variation. For the muon efficiency scale-factors, the low p_T scale-factor variations have no impact on the C_Z-factor uncertainties. Furthermore, scale-factors uncertainties related to the muon resolution like *Muon ID*, *Muon MS* and *Muon Sagitta* as well as *Muon Scale* show only a minor impact. This is expected, as the integrated fiducial cross-section measurement depends on the number of observed events rather than the highly precise reconstruction of the muon momentum.

Using the nominal C_Z-factor obtained from the comparison of particle-level and detector-level values, one can calculate the fiducial cross-section σ_{fid}, leading to the results presented in Table 9.10.

Table 9.10: C_Z-factors and fiducial cross sections for born, bare and dressed-level.

Particle-level	C_Z-factor	fiducial cross-section [pb]
Born-level	0.685 ± 0.011	$731.7 \pm 15.3^{lumi} \pm 0.2^{stat} \pm 11.3^{sys}$
Dressed-level	0.702 ± 0.011	$713.2 \pm 14.3^{lumi} \pm 0.2^{stat} \pm 11.0^{sys}$
Bare-level	0.715 ± 0.011	$701.0 \pm 14.0^{lumi} \pm 0.2^{stat} \pm 10.7^{sys}$

Hereby, The different uncertainties affecting the cross-section measurement are calculated using Gaussian error propagation. Hereby, the uncertainty of the number of events is estimated by its Poisson error $\Delta N_{signal} = \sqrt{N_{signal}}$, while the uncertainty of the luminosity measurement is provided by the ATLAS Collaboration to be 2 % [51]. The systematic uncertainty is obtained from the differences in the C_Z-factors. As outlined in Section 9.9.1, the mean absolute difference of up- and down-variation to the nominal value of these factors is calculated for all systematics. The squared sum over all systematics obtained this way is then used as the total systematic uncertainty ΔC_Z.

Table 9.11: Comparison to former measurement [56]. Compared are fiducial cross-sections with a $p_T^{\mu} > 25.$ GeV at $\sqrt{s} = 13$ TeV.

	Z→ $\mu\mu$ cross-section [pb]
this analysis	$774.9 \pm 0.2^{stat} \pm 14.6^{lumi} \pm 15.0^{syst}$
previous measurement	$774 \pm 1^{stat} \pm 6^{syst} \pm 5^{beam} \pm 16^{lumi}$

This analysis is performed one more time using a particle-level lepton transverse momentum cut $p_T^{\mu} > 25$ GeV to be able to compare the obtained results to former measurements, presented in Ref. [56]. Table 9.11 summarizes the obtained results, showing overall good agreement. The former measurement has a smaller systematic uncertainty compared to this measurement, because the muon reconstruction efficiencies were assumed to be better understood.

As stated before, Drell-Yan lepton pairs do not solely consist of muon pairs, but also electron and τ pairs, the latter decaying too fast to be detected within the ATLAS detector, while electron pairs can be detected. In parallel to this analysis, another group measured the cross-section of Drell-Yan electron pairs. These measurements are combined to achieve higher precision.

Table 9.12: Measured inclusive cross-sections in the electron and muon decay channels at born-level together with the combined fiducial cross-section. Additionally a theory prediction at NNLO in α_S using the CT14 PDF set is presented. Taken from Ref. [54].

Channel	Measured cross-section × $\mathcal{B}(Z \to \ell\ell)$ (value ± stat. ± syst. ± lumi.)	Predicted cross-section × $\mathcal{B}(Z \to \ell\ell)$ (value ± PDF ± α_s ± scale ± intrinsic)
$Z \to ee$	$738.3 \pm 0.2 \pm 7.7 \pm 15.5$ pb	
$Z \to \mu\mu$	$731.7 \pm 0.2 \pm 11.3 \pm 15.3$ pb	
$Z \to \ell\ell$	$736.2 \pm 0.2 \pm 6.4 \pm 15.5$ pb	$703^{+19}_{-24}\,^{+6}_{-8}\,^{+4}_{-6}\,^{+5}_{-5}$ pb [16]

Table 9.12 shows the measured fiducial cross-sections of both the $Z \to \mu\mu$ and $Z \to ee$ channel together with the corresponding statistical, systematic, and luminosity uncertainties together with the combined cross-section, labeled $Z \to \ell\ell$. These show good agreement within the uncertainties and are comparable to former ATLAS measurements [16] as seen in Table 9.11. Additionally, Table 9.12 shows a fixed-order theoretical prediction, which also shows good agreement with the measurement within uncertainties. Details of the prediction can be found in Ref. [54].

The two measurements are combined using Bayesian Evidence Combination. Applying this leads to a combined cross-section value calculated as the weighted average, using the squared sum of all uncertainties as full uncertainy of the cross-section.

The combined integrated fiducial cross-section is then given by:

$$\sigma_{fid}^{comb} = \frac{\frac{\sigma_{fid}^{\mu}}{\left(\Delta\sigma_{fid}^{\mu}\right)^2} + \frac{\sigma_{fid}^{e}}{\left(\Delta\sigma_{fid}^{e}\right)^2}}{\left(\Delta\sigma_{fid}^{\mu}\right)^{-2} + \left(\Delta\sigma_{fid}^{e}\right)^{-2}} \tag{9.14}$$

$\Delta\sigma_{fid,\mu}$ and $\Delta\sigma_{fid,e}$ refer to the total cross-section uncertainty for the muon and electron channel, respectively. The uncorrelated uncertainties, namely systematical and statistical uncertainty, are combined, while the luminosity uncertainty is assumed to be 100 % correlated between the channels and kept as is. Uncorrelated uncertainties are combined by:

$$\Delta\sigma_{fid}^{comb} = \sqrt{\frac{1}{\left(\Delta\sigma_{fid}^{\mu}\right)^{-2} + \left(\Delta\sigma_{fid}^{e}\right)^{-2}}}. \tag{9.15}$$

This combination is only valid assuming no correlations between the channels. Other approaches like Bayesian iterative unfolding, which is used for the differential cross-section measurement, can take correlations into account. Combining both channels finally leads to the combined fiducial integrated cross-section:

$$\sigma_{\mathrm{fid}}(pp \to Z \to \ell\ell) = 736.2 \pm 0.2^{\mathrm{stat}} \pm 6.4^{\mathrm{sys}} \pm 15.5^{\mathrm{lumi}}\,\mathrm{pb}.$$

The combination of the two channels leads to a reduction of the systematic uncertainty compared to the individual electron- and muon-channel measurements, due to the uncorrelated systematics.

The same measurement can be performed differential in bins of both $\mathrm{p_T}$ and ϕ^*_η. This measurement is discussed in the following sections. Hereby only the shape of the cross-section distribution is of interest, as the normalization is given by the integrated fiducial cross-section. Hence, only normalized distributions are analyzed.

9.12 Normalized differential cross-section measurement

As for the integrated fiducial cross-section measurement, the measurement of the normalized differential cross-section aims to provide a particle-level measurement. In this case, detector effects can not be corrected by a global factor like C_Z, as local, bin-dependent influences exist as well. This approach also depends on the binning chosen for the differentiating variables, $\mathrm{p_T^{\mu\mu}}$ and ϕ^*_η in this case. The binning has to be optimized and chosen to have high purity. For this project, the binning is chosen to be equal to former measurements [16], which studied the binning.These detector effects and limitations can be corrected using MC samples, as well-described detector models produce a smearing effect similar to the real detector. This simulation can be inverted to compute the particle-level properties from the detector-level measurements, which is called unfolding. One approach to calculate this inversion is the so-called Bayesian iterative unfolding [78, 79], which is used in the context of this project and explained below.

9.12.1 Bayesian iterative unfolding

Bayesian iterative unfolding is a statistical unfolding algorithm based on Bayes theorem to correct for detector effects. The following nomenclature is used for the discussion of the mathematical concept:

- $\mathbf{d}$ as the measured data distribution on detector-level with n bins, where the content of bin i is written as d_i.
- $\mathbf{b}$ as the expected background contribution on detector-level with n bins, where the content of bin i is written as b_i.
- $\mathbf{x}$ as the signal distribution on detector-level with n bins, where the content of bin j is written as x_j.
- μ as the unfolded distribution with n bins. μ_i refers to the content of bin i.

Two corrections are applied, denoted as:

- the fiducial factors $f_i = \frac{N_i^{\text{in fid.}}}{N_i}|$ for each bin i of the detector-level distribution. This fiducial factor corrects for those reconstructed events on detector-level, which fall outside the particle-level fiducial region. These events are also referred to as fakes. f_i is defined as the ratio of events falling in the fiducial region on particle-level $N_i^{\text{in fid.}}$ over all reconstructed and selected events N_i.
- the correction factors $m_i = \frac{N_i^{\text{detector}}}{N_i^{\text{particle}}}$ for each bin i of the distribution, which accounts for acceptance and efficiency losses on detector-level. Events missed on detector-level but present at particle-level are also referred to as misses. It is defined as the ratio of reconstructed events N_i^{detector} in bin i on detector-level over the number of particle-level events N_i^{particle} in bin i in the fiducial volume. Bin i is defined by the truth value of the unfolding variable.

Using these correction factors, the number of signal events x_j in bin j are calculated as well as the corrected number of particle-level events y_i:

$$x_j = (d_j - b_j) \cdot f_j \tag{9.16}$$

This value is then unfolded to obtain the particle-level value μ_i, which is then corrected to account for missed events to achieve the final value:

$$y_i = \mu_i \cdot m_i. \tag{9.17}$$

The full process is also depicted in Figure 9.19.

Mathematically, unfolding attempts to determine the probability distribution of particle-level events $P(\mu_i|x_j, I)$, describing the fraction of detector-level event x_j visible in particle-level bin i. From this, the number of events μ_i on particle-level can be estimated.

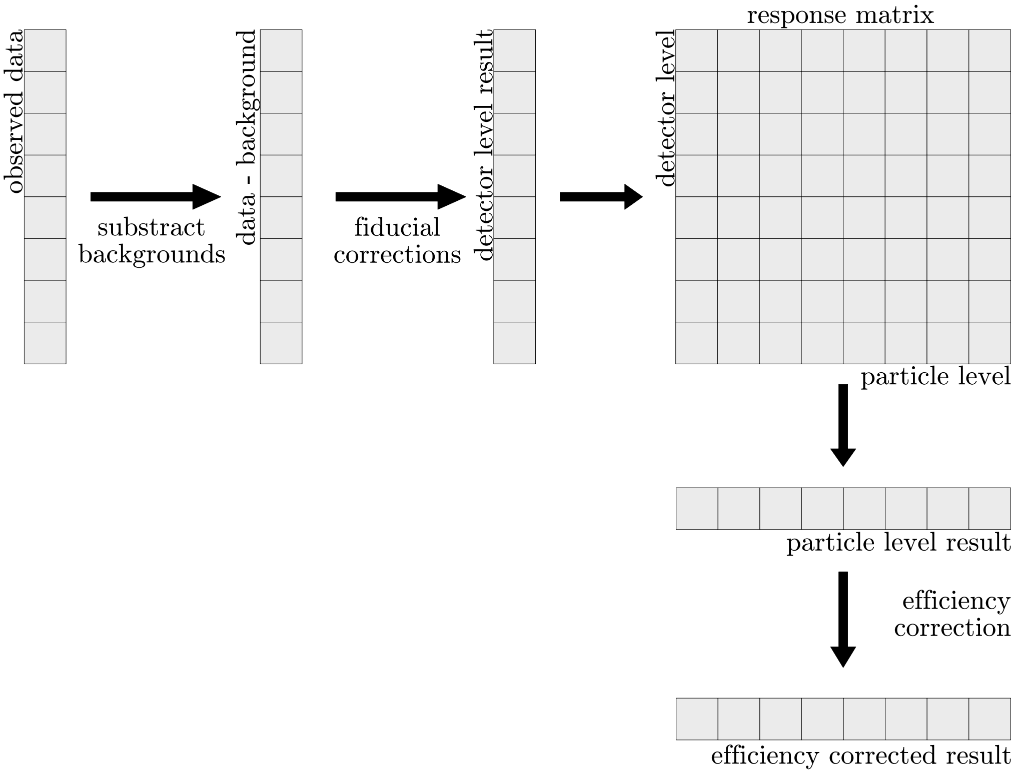


Figure 9.19: Pictorial description of the unfolding method for a distribution in a fiducial volume. The background subtracted measured distribution is unfolded and corrected to compute the particle-level distribution. Each grey square corresponds to a bin of the measured distributions. Fiducial corrections correspond to events observed only on detector-level, also called fakes. Efficiency corrections correct for events only seen on particle-level but missed on detector-level, also called misses.

The variable I hereby describes underlying assumptions postulated for the analysis at hand, which are usually left implicit.

Starting from the event generator, a given detector simulation can in this context be abstractly described as follows. The simulation describes how many of the μ_i events created in particle-level bin i are visible in detector-level bin j, denoted $x_j|\mu_i$. This fraction can be described in similarity to above by a probability function, from here on called response matrix:

$$R_{ji} := P(x_j|\mu_i, I). \tag{9.18}$$

$x_j|\mu_i$ can be calculated by

$$x_j|\mu_i = P(x_j|\mu_i, I) \cdot \mu_i = R_{ji} \cdot \mu_i \tag{9.19}$$

Using Bayes theorem, Equation 9.18 leads to

$$\Theta_{ij} := P(\mu_i|x_j, I) = \frac{P(x_j|\mu_i, I) \cdot P(\mu_i|I)}{\sum\limits_k P(x_j|\mu_k, I) \cdot P(\mu_k|I)} \tag{9.20}$$

which can be rewritten using R_{ji}

$$P(\mu_i|x_j, I) = \frac{R_{ji} \cdot P(\mu_i|I)}{\sum\limits_k R_{jk} \cdot P(\mu_k|I)} \tag{9.21}$$

The number of particle-level events in bin i can be estimated given n_j events were measured in detector-level bin j:

$$\mu_i|_{n_j} = \frac{\Theta_{ij} \cdot n_j}{\epsilon_i} \tag{9.22}$$

with ϵ being the efficiency, by design defined by:

$$\epsilon_i = \sum_{j=1}^{N} R_{ji}. \tag{9.23}$$

Summing over all detector-level bins, the total number of events estimated on particle-level in bin i can be calculated as:

$$\mu_i = \frac{1}{\epsilon_i} \sum_{j=1}^{N} \Theta_{ij} \cdot n_j \tag{9.24}$$

Hereby, the response matrix is estimated using MC simulations by:

$$R_{ji} = \frac{n_j^{MC}}{\mu_i^{MC}}, \tag{9.25}$$

as particle-level information is only available for MC samples.

The process is applied iteratively, following these steps:

- Update prior $P(\mu_i, I)$. For the first iteration, the prior is usually chosen to be the normalized particle-level distribution.
- Calculate the unfolded values.
- The resulting μ_i is used as prior for the next generation.

9.12.2 Number of iterations

As one can easily see, the results of this unfolding procedure depend on the number of iterations. There are two approaches to evaluating the suited number of iterations. On the one hand, one can find the number of iterations where the combination of statistical and model uncertainty, both described later in Section 9.12.4, is minimal. A more practical approach used in this project is to find the number of iterations after which the result does not change within the uncertainties. Four iterations were found to fulfill this criterion, which is visualized in the following Figure 9.20.

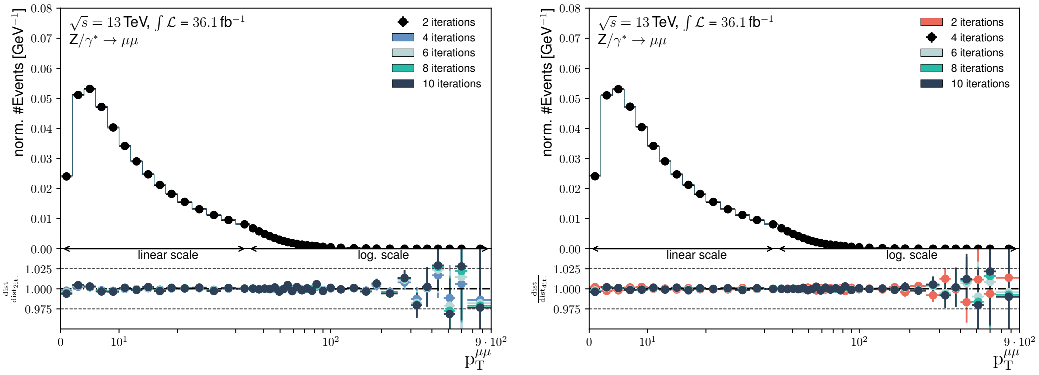


Figure 9.20: Comparison of the differential cross sections depending on the number of unfolding steps. Left is shown for 2 iterations as standard value, right for 4 iterations.

9.12.3 Closure and purity test

As done in Section 8.8, the performance of detector-level corrections, the unfolding process, is evaluated by analyzing response matrices. The purity of the response matrix is calculated to check bin migrations and the model dependence of this approach. Both the response matrix as well as the purity are shown in Figure 9.21 for both $p_T^{\mu\mu}$ and ϕ^*_η. The bin purities are found to be above 60 % for $p_T^{\mu\mu}$ and above 90 % for ϕ^*_η, both reasonably good values for Bayesian iterative unfolding, especially compared to the purities in Section 8.8.

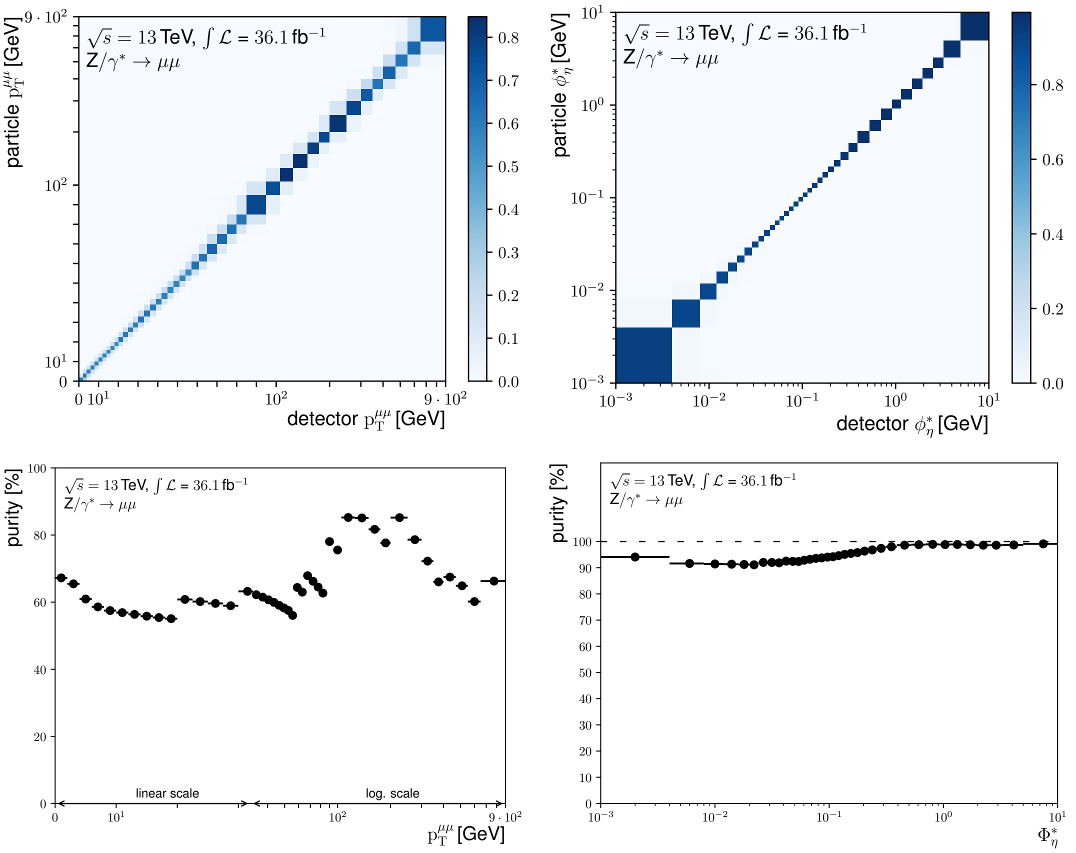


Figure 9.21: Response matrix (upper row) and bin purity(lower row) for the transverse momentum (left column) and ϕ_η^*(right column). The response matrix visualizes bin-migrations, which are measured by the purity. A high purity refers to low bin migrations and a low model dependence.

Additionally, the functionality of the unfolding is validated using a so-called technical closure test. Hereby, the objects required for unfolding, the detector-level and particle-level distributions, as well as the response matrix, are created from MC signal sample. Then, this MC detector-level distribution is unfolded using Bayesian iterative unfolding. The result, shown in Figure 9.22, is by design expected to have 100 % agreement, which is achieved.

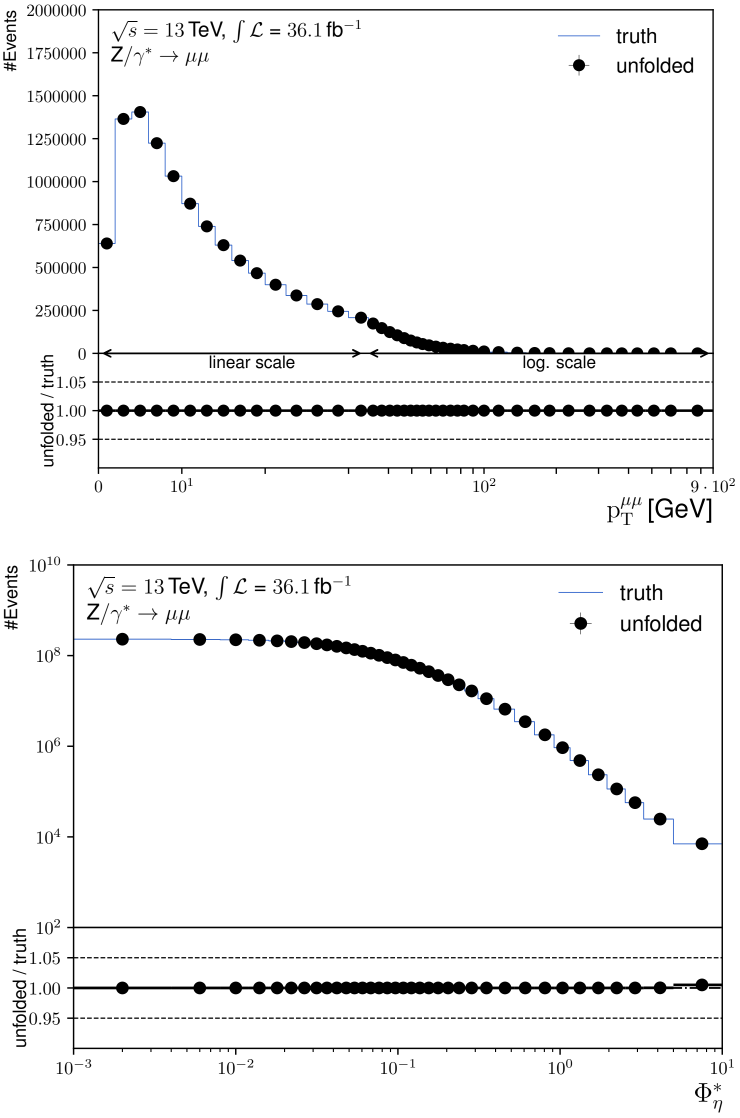


Figure 9.22: Closure test for the transverse momentum (left) and ϕ^*_η(right).

9.12.4 Additional systematic uncertainties

Further systematic uncertainties in addition to the systematics mentioned in Section 9.9 are taken into account for the normalized differential measurement. The luminosity measurement is not taken into account as only the shape is measured.

Monte Carlo Prior uncertainty

Bayesian iterative unfolding relies on an initial assumption of a MC prior $P(x_i, I)$, usually the normalized particle-level distribution measured in the MC signal sample. The MC Prior uncertainty probes the impact of this assumption by changing this prior using a well-defined data-driven closure test. This test transfers the differences between MC and recorded data on detector-level to the particle-level.

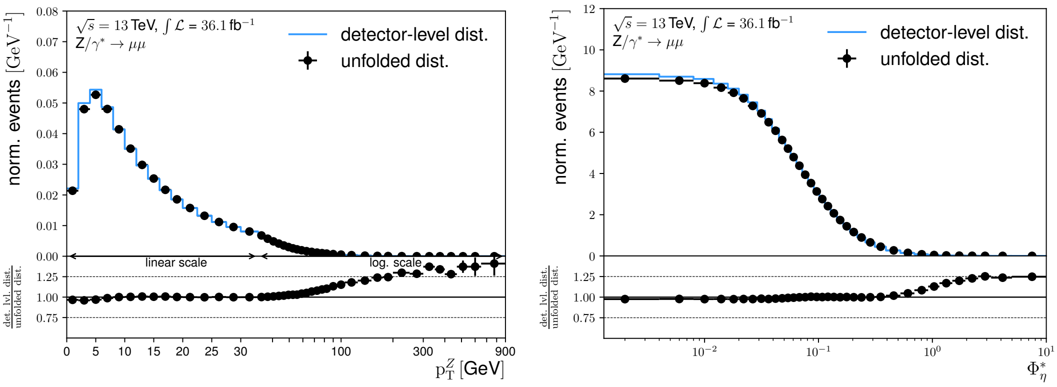


Figure 9.23: Reweighting spectrum for $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ (left) and ϕ_η^* (right), used to estimate the MC prior uncertainty. The detector-level spectrum is divided by the unfolded spectrum to calculate the bin weights, which are shown in the lower panel. These weights are then applied on an event-by-event basis depending on the event $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ or ϕ_η^*. Similar to the detector-level control plots the weights rise with high p_{T}, following the same argumentation.

For each distribution measured, the MC signal events are reweighed based on measured particle-level distributions. This results in a better agreement with the measured data distributions. The weights are applied on an event-by-event basis, depending on the particle-level value of the observable. It is applied to all distributions of the MC signal sample incorporated in the unfolding procedure. The distributions are shown in Figure 9.23.

The measured data distribution is unfolded using this reweighed MC signal sample. The systematic uncertainty of the unfolding method is then given by the difference between this distribution (reweighed and unfolded), and the data distribution unfolded with the nominal baseline MC sample without reweighting.

Background contamination

Not only the simulated processes but also the normalization to measured data are affected by uncertainties. The measured distributions of the different background samples are scaled up and down to study the effect on the unfolded result. Hereby the normalization of the background from $Z \rightarrow \tau\tau$ is varied by ±2%, all other backgrounds (diboson and $t\bar{t}$ processes) are varied by ±5% both up and down. These numbers refer to the relative uncertainty of the cross-section of these processes.

In order to evaluate a systematic uncertainty for the remaining multijet background estimated by the data-driven ABCD method, the analysis of the dataset is performed again using the *FixedCutTight*[3] [17] instead of the *Gradient* isolation working point. This is done to study the influence of the isolation criteria on the final result, changing one of the two defining parameters of the four regions used for the ABCD method. As the charge of the muons can not be varied, only the influence of the isolation is studied.

The difference of the unfolded spectrum using the corresponding multijet background distribution is used as a systematic uncertainty. The varied backgrounds are subtracted from the data distribution, which is then unfolded. The absolute difference to the nominal result is used as uncertainty for both the up- and the downscaled distributions. The upscaled and downscaled distributions are considered fully correlated, and the mean of the absolute differences is used for combination, as described above. The background contamination

9.12.5 Treatment of systematic uncertainties

Variated distributions in bins of $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ and ϕ^*_η are unfolded using Bayesian iterative unfolding. The resulting particle-level variated distributions are then combined following the procedure outlined in Section 9.9, using the difference of the particle-level nominal distribution. Both the Monte Carlo Prior, as well as the Pileup uncertainty, does not provide up- and down-variations. Hence the absolute difference to the nominal unfolded result to the nominal distribution is used as uncertainty.

9.12.6 Statistical uncertainties

Apart from systematic uncertainties also the limited statistics lead to statistical uncertainties, affecting both data and MC samples. The statistical uncertainties on the unfolded result are estimated by creating 100 Poison fluctuated variations of the measured detector-level distribution (before subtracting the background).

---

[3] This cut measures the transverse momentum of objects within a cone of radius $R = min\left(\frac{10\,\mathrm{GeV}}{\mathrm{p_T}}, 0.2\right)$ around the muon. A muon is called isolated if less than 6 % of its own transverse momentum is present in this cone.

Then the analysis steps (background subtraction, bayesian iterative unfolding) are repeated for all these varied distributions. The RMS of the relative difference between systematically varied unfolded distributions and the nominal unfolded distribution are used as uncertainty.

The MC statistical uncertainties are estimated by creating 100 Poisson varied unfolding matrices, detector-level, and particle-level distributions. The distribution measured in data is unfolded using the varied distributions, and the RMS of the differences to the nominal unfolded histogram is used as uncertainty.

9.12.7 Overall bin-wise uncertainties of the unfolded distributions

Table 9.13 and Table 9.14 show a full bin-wise breakdown of all uncertainties for the unfolded distributions of $p_T^{\mu\mu}$ and ϕ^*_η on born-level, respectively. The results are only shown on born particle-level here, as the breakdown is essentially unchanged for the bare- and dressed-level definitions. The same Tables covering the dressed and bare level can be found in the appendix as Table A.8 and Table A.9 for the bare particle-level as well as Table A.10 and Table A.11 for the dressed particle-level for $p_T^{\mu\mu}$ and ϕ^*_η, respectively.

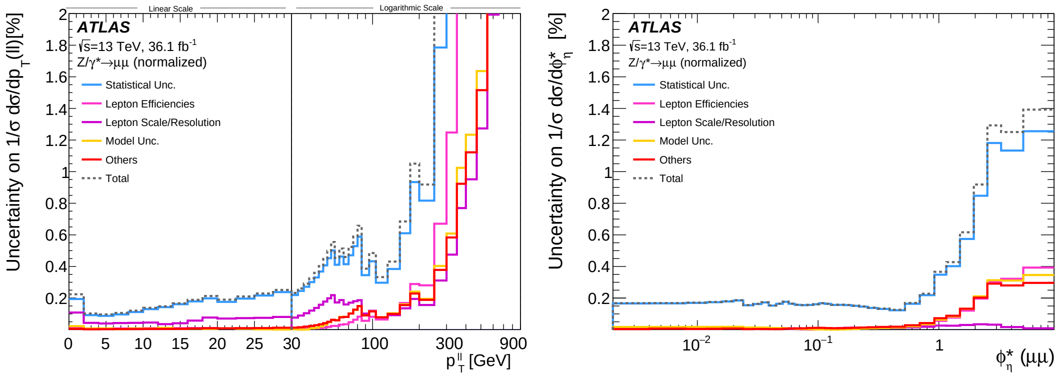


Figure 9.24: Binwise breakdown of relative uncertainties for the $p_T^{\mu\mu}$ spectrum (left) and the ϕ^*_η spectrum on born particle-level. The breakdown is essentially unchanged for bare and dressed particle-level. Overall this measurement is limited by statistical uncertainties. Taken from Ref. [54].

Figure 9.24 visualizes these bin-wise breakdowns of the unfolded distributions of $p_T^{\mu\mu}$ and ϕ^*_η. As showing a large number of uncertainties results in an overly crowded plot, the uncertainties are grouped into five categories. Statistical uncertainties cover both the MC and data statistical uncertainties, while lepton efficiencies and lepton scale/resolution cover uncertainties related to varied scale factors and momentum resolution uncertainties, respectively. Model uncertainty covers only the MC prior, while all remaining uncertainties are called Others.

Table 9.13: Measured normalized $p_T^{\mu\mu}$ cross-section on born level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 2	0.0483485	0.264	0.214	0.021	0.166	0.088	0.001	0.008	0.000	0.000	0.000	0.001	0.060	0.046	0.009
2 - 4	0.102336	0.114	0.099	0.008	0.061	0.068	0.007	0.003	0.000	0.000	0.000	0.000	0.011	0.044	0.004
4 - 6	0.106454	0.100	0.090	0.008	0.060	0.068	0.000	0.003	0.000	0.000	0.000	0.000	0.014	0.000	0.003
6 - 8	0.0949447	0.119	0.105	0.008	0.059	0.038	0.001	0.003	0.000	0.000	0.000	0.000	0.002	0.010	0.004
8 - 10	0.0811024	0.128	0.118	0.009	0.046	0.024	0.001	0.003	0.000	0.000	0.000	0.000	0.002	0.014	0.004
10 - 12	0.068568	0.153	0.138	0.010	0.028	0.085	0.007	0.002	0.000	0.000	0.000	0.000	0.000	0.016	0.004
12 - 14	0.0583926	0.176	0.160	0.011	0.010	0.104	0.005	0.002	0.000	0.000	0.000	0.000	0.013	0.025	0.005
14 - 16	0.0495932	0.195	0.184	0.010	0.005	0.100	0.001	0.002	0.000	0.000	0.000	0.000	0.013	0.007	0.006
16 - 18	0.0426003	0.218	0.207	0.009	0.014	0.077	0.006	0.002	0.000	0.000	0.000	0.000	0.001	0.019	0.006
18 - 20	0.0365929	0.218	0.233	0.009	0.030	0.071	0.005	0.003	0.000	0.000	0.000	0.000	0.008	0.015	0.008
20 - 22.5	0.039024	0.186	0.203	0.006	0.031	0.052	0.002	0.002	0.000	0.000	0.000	0.000	0.005	0.005	0.008
22.5 - 25	0.0328605	0.212	0.225	0.007	0.046	0.042	0.004	0.003	0.000	0.000	0.000	0.000	0.002	0.018	0.011
25 - 27.5	0.0280475	0.232	0.253	0.006	0.057	0.042	0.006	0.003	0.000	0.000	0.000	0.000	0.006	0.005	0.013
27.5 - 30	0.0238817	0.264	0.281	0.007	0.065	0.033	0.006	0.004	0.000	0.000	0.000	0.000	0.011	0.010	0.016
30 - 33	0.0242942	0.226	0.257	0.006	0.070	0.017	0.013	0.004	0.000	0.000	0.000	0.000	0.025	0.001	0.016
33 - 36	0.0203863	0.261	0.323	0.007	0.090	0.007	0.014	0.005	0.000	0.000	0.000	0.000	0.009	0.016	0.020
36 - 39	0.0172779	0.283	0.359	0.008	0.107	0.004	0.010	0.005	0.000	0.000	0.000	0.000	0.022	0.015	0.024
39 - 42	0.0146252	0.311	0.371	0.009	0.135	0.017	0.004	0.006	0.001	0.000	0.000	0.001	0.029	0.000	0.027
42 - 45	0.0124869	0.333	0.335	0.010	0.143	0.036	0.004	0.006	0.001	0.000	0.000	0.002	0.047	0.008	0.032
45 - 48	0.0105421	0.370	0.375	0.013	0.168	0.022	0.001	0.006	0.001	0.000	0.000	0.003	0.037	0.027	0.039
48 - 51	0.00896744	0.402	0.435	0.016	0.194	0.021	0.008	0.006	0.001	0.001	0.000	0.000	0.018	0.031	0.053
51 - 54	0.00760829	0.446	0.501	0.019	0.212	0.022	0.007	0.007	0.001	0.001	0.000	0.000	0.024	0.053	0.063
54 - 57	0.00648499	0.491	0.578	0.022	0.229	0.013	0.023	0.007	0.000	0.001	0.000	0.000	0.023	0.072	0.079
57 - 61	0.00752818	0.401	0.507	0.019	0.168	0.013	0.010	0.005	0.000	0.001	0.000	0.000	0.033	0.058	0.071
61 - 65	0.00635331	0.425	0.564	0.024	0.171	0.015	0.008	0.006	0.000	0.001	0.000	0.001	0.067	0.081	0.083
65 - 70	0.00675757	0.398	0.525	0.026	0.146	0.000	0.001	0.005	0.000	0.000	0.000	0.001	0.094	0.097	0.078
70 - 75	0.00565293	0.446	0.517	0.040	0.165	0.002	0.002	0.007	0.000	0.001	0.000	0.000	0.083	0.090	0.110
75 - 80	0.00477561	0.480	0.640	0.063	0.170	0.027	0.013	0.009	0.000	0.001	0.000	0.000	0.073	0.147	0.135
80 - 85	0.00405529	0.556	0.726	0.087	0.164	0.029	0.004	0.013	0.000	0.001	0.000	0.004	0.069	0.162	0.198
85 - 95	0.00652509	0.323	0.428	0.060	0.099	0.006	0.007	0.010	0.000	0.000	0.000	0.002	0.029	0.113	0.128
95 - 105	0.00486299	0.422	0.551	0.095	0.116	0.010	0.002	0.015	0.000	0.000	0.000	0.004	0.046	0.128	0.165
105 - 125	0.00656466	0.289	0.344	0.075	0.080	0.000	0.002	0.012	0.000	0.000	0.000	0.002	0.019	0.105	0.119
125 - 150	0.00463292	0.383	0.424	0.109	0.097	0.008	0.004	0.020	0.000	0.000	0.000	0.002	0.031	0.112	0.163
150 - 175	0.0026228	0.604	0.674	0.195	0.127	0.056	0.004	0.037	0.000	0.000	0.000	0.005	0.013	0.191	0.237
175 - 200	0.00151905	0.947	0.931	0.333	0.174	0.024	0.013	0.063	0.002	0.000	0.000	0.012	0.010	0.232	0.321
200 - 250	0.00152326	0.821	0.821	0.320	0.122	0.008	0.000	0.055	0.003	0.000	0.000	0.008	0.105	0.158	0.181
250 - 300	0.000600594	1.726	1.806	0.769	0.168	0.059	0.091	0.113	0.009	0.000	0.000	0.013	0.075	0.230	0.259
300 - 350	0.000295887	2.900	3.057	1.344	0.246	0.023	0.025	0.186	0.019	0.001	0.000	0.006	0.008	0.375	0.375
350 - 400	0.00014254	4.782	4.044	2.444	0.297	0.138	0.015	0.308	0.046	0.001	0.001	0.009	0.148	0.309	0.477
400 - 470	8.61709e-05	6.587	5.397	3.570	0.317	0.157	0.053	0.315	0.077	0.001	0.000	0.039	0.380	0.489	0.482
470 - 550	4.60894e-05	10.87	9.231	5.849	0.277	0.083	0.131	0.408	0.141	0.003	0.000	0.027	1.081	0.027	0.494
550 - 650	2.04842e-05	18.67	17.34	8.916	0.459	0.184	0.393	0.666	0.338	0.006	0.001	0.027	1.203	0.791	0.781
650 - 900	1.25171e-05	26.68	25.22	9.716	0.504	0.723	1.669	0.833	0.586	0.008	0.002	0.060	0.895	0.894	0.918

Table 9.14: Measured normalized ϕ_η^* cross-section on born level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 0.004	0.0352108	0.222	0.124	0.009	0.016	0.002	0.001	0.004	0.000	0.000	0.000	0.000	0.019	0.017	0.004
0.004 - 0.008	0.0348122	0.208	0.137	0.009	0.013	0.004	0.001	0.004	0.000	0.000	0.000	0.000	0.030	0.017	0.003
0.008 - 0.012	0.0341709	0.183	0.132	0.010	0.014	0.001	0.000	0.004	0.000	0.000	0.000	0.000	0.002	0.017	0.003
0.012 - 0.016	0.0333264	0.205	0.140	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.005	0.018	0.003
0.016 - 0.02	0.0323455	0.207	0.135	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.019	0.018	0.003
0.02 - 0.024	0.0311863	0.207	0.143	0.010	0.014	0.001	0.001	0.004	0.000	0.000	0.000	0.001	0.025	0.019	0.003
0.024 - 0.029	0.0371284	0.177	0.112	0.008	0.011	0.003	0.000	0.003	0.000	0.000	0.000	0.000	0.008	0.015	0.003
0.029 - 0.034	0.0351867	0.178	0.115	0.009	0.009	0.003	0.001	0.003	0.000	0.000	0.000	0.000	0.002	0.015	0.003
0.034 - 0.039	0.032948	0.193	0.127	0.009	0.010	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.003	0.015	0.003
0.039 - 0.045	0.0370367	0.162	0.110	0.008	0.009	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.013	0.012	0.003
0.045 - 0.051	0.0342322	0.173	0.116	0.009	0.009	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.011	0.003
0.051 - 0.057	0.0316565	0.188	0.127	0.009	0.008	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.015	0.009	0.004
0.057 - 0.064	0.033933	0.176	0.116	0.008	0.005	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.019	0.006	0.004
0.064 - 0.072	0.0353678	0.169	0.111	0.007	0.003	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.010	0.003	0.004
0.072 - 0.081	0.0357806	0.174	0.108	0.007	0.003	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.004	0.002	0.004
0.081 - 0.091	0.0355277	0.176	0.108	0.007	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.004
0.091 - 0.102	0.0345791	0.174	0.111	0.007	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.021	0.000	0.005
0.102 - 0.114	0.0331588	0.196	0.114	0.007	0.003	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.016	0.000	0.005
0.114 - 0.128	0.0338174	0.202	0.111	0.006	0.005	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.008	0.000	0.005
0.128 - 0.145	0.035083	0.191	0.107	0.006	0.008	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.006	0.000	0.006
0.145 - 0.165	0.0345822	0.210	0.106	0.005	0.011	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.007	0.001	0.006
0.165 - 0.189	0.0342832	0.199	0.107	0.006	0.014	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.003	0.001	0.009
0.189 - 0.219	0.0344528	0.192	0.104	0.006	0.015	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.005	0.001	0.012
0.219 - 0.258	0.034456	0.187	0.101	0.008	0.018	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.001	0.001	0.018
0.258 - 0.312	0.0349282	0.177	0.095	0.009	0.018	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.009	0.005	0.023
0.312 - 0.391	0.0346567	0.164	0.088	0.010	0.017	0.001	0.001	0.003	0.000	0.000	0.000	0.000	0.009	0.010	0.027
0.391 - 0.524	0.0347245	0.153	0.083	0.012	0.014	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.008	0.015	0.031
0.524 - 0.695	0.0241386	0.208	0.110	0.020	0.017	0.001	0.004	0.005	0.000	0.000	0.000	0.000	0.003	0.028	0.049
0.695 - 0.918	0.016633	0.249	0.149	0.031	0.017	0.003	0.004	0.007	0.000	0.000	0.000	0.002	0.012	0.048	0.076
0.918 - 1.153	0.0094987	0.367	0.236	0.060	0.020	0.006	0.008	0.013	0.000	0.001	0.000	0.002	0.046	0.090	0.125
1.153 - 1.496	0.0074166	0.428	0.253	0.081	0.006	0.002	0.006	0.016	0.000	0.000	0.000	0.003	0.048	0.111	0.134
1.496 - 1.947	0.00487765	0.502	0.355	0.132	0.000	0.000	0.004	0.023	0.000	0.001	0.000	0.004	0.059	0.148	0.174
1.947 - 2.522	0.00309366	0.693	0.495	0.221	0.007	0.002	0.001	0.035	0.001	0.001	0.001	0.007	0.030	0.193	0.232
2.522 - 3.277	0.00206941	0.946	0.675	0.340	0.006	0.006	0.000	0.052	0.003	0.001	0.002	0.006	0.037	0.256	0.305
3.277 - 5	0.0020097	0.879	0.637	0.352	0.010	0.005	0.005	0.051	0.005	0.001	0.002	0.006	0.066	0.220	0.279
5 - 10	0.0016914	0.945	0.654	0.418	0.006	0.014	0.007	0.056	0.007	0.001	0.002	0.004	0.032	0.230	0.235

Overall, relative bin-wise uncertainties are below 0.4 % for the majority of bins, rising with higher bin number. This is true for both $p_T^{\mu\mu}$ and ϕ_η^*. Overall, the measurement is limited by the statistical uncertainties. Lepton resolution uncertainties play a role for low-p_T bins, while lepton efficiencies are contributing to the high p_T regime.

The rise in uncertainty in higher bins can be explained by the lack of events in these bins. Statistical uncertainties rely on Poisson variated distributions, which show large fluctuations in bins with only a few events. The same holds true for the uncertainties related to muon efficiency and muon reconstruction scale-factors. Scale-factors are determined using the tag-and-probe method, hence rely on events covering the full kinematic range these factors are applied on. Similar to the statistical uncertainties, fluctuations in scale-factors get larger with fewer events used to determine these factors. Furthermore, the measurement of the muon transverse momentum relies on the curvature of the track, which decreases with higher transverse momentum. Hence the uncertainty rises with smaller curvature.

9.12.8 Results of the differential cross-section measurements

Utilizing Equation 9.13, this differential cross-section can be calculated, done by dividing the unfolded distribution by the luminosity, as the number of events can be determined by $N = \mathcal{L} \cdot \sigma$. As this Section covers the normalized differential cross-section measurements, scaling the unfolded distributions by a global factor, the luminosity, in this case, has no effect due to the normalization. Hence, the unfolded distribution is already equal to the normalized differential cross-section and the uncertainty.

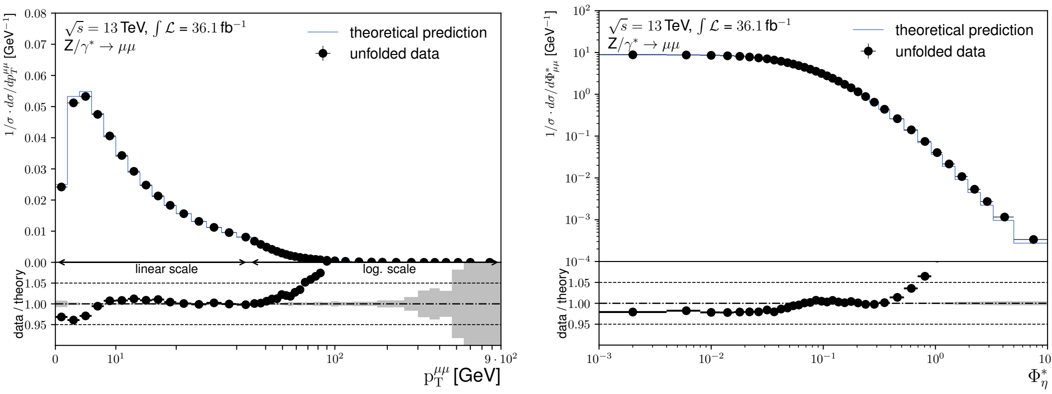


Figure 9.25: Born-level normalized differential cross-section $p_T^{\mu\mu}$ (left) and $\Phi^*_{\mu\mu}$ (right) with the relative uncertainties indicated by the grey band in the ratio plot. Bare and dressed-level results can be found in Figure A.4.

The full information can be recovered by multiplying the fiducial cross-section to the normalized cross-section. The results of the measurements of the differential $p_T^{\mu\mu}$ and ϕ_η^* cross-sections are both shown in Figure 9.25 and summarized in the Tables 9.15 and 9.16 for $p_T^{\mu\mu}$ and ϕ_η^*, respectively.

Table 9.15: Measured normalized differential cross-section vs. $p_T^{\mu\mu}$ where the results are shown for born-level, the bare-level and the dressed-level definition of leptons. The uncertainties are split-up between bin-to-bin uncorrelated uncertainties (dominated by data and MC statistics) and bin-to-bin correlated uncertainties

Binning	Born-level val ± uncor-unc ± cor-unc $[10^{-5}]$	Bare-level val ± uncor-unc ± cor-unc $[10^{-5}]$	Dressed-level val ± uncor-unc ± cor-unc $[10^{-5}]$
0 - 2	4834.85 ± 16.47 ± 9.86	4638.28 ± 15.97 ± 9.72	4727.29 ± 16.20 ± 9.76
2 - 4	10233.60 ± 15.56 ± 10.57	9949.08 ± 15.38 ± 10.27	10078.00 ± 15.42 ± 10.44
4 - 6	10645.40 ± 14.46 ± 9.88	10538.20 ± 14.35 ± 9.75	10582.60 ± 14.45 ± 9.82
6 - 8	9494.47 ± 15.16 ± 6.85	9550.20 ± 14.85 ± 6.75	9522.21 ± 15.00 ± 6.80
8 - 10	8110.24 ± 14.21 ± 4.45	8240.11 ± 13.96 ± 4.37	8173.78 ± 14.08 ± 4.40
10 - 12	6856.80 ± 14.19 ± 6.30	7009.81 ± 13.98 ± 6.01	6935.86 ± 14.07 ± 6.15
12 - 14	5839.26 ± 13.95 ± 6.38	5988.52 ± 13.55 ± 6.10	5916.19 ± 13.61 ± 6.20
14 - 16	4959.31 ± 13.32 ± 5.04	5084.26 ± 12.90 ± 4.87	5023.03 ± 12.97 ± 4.94
16 - 18	4260.03 ± 12.83 ± 3.49	4358.97 ± 12.50 ± 3.41	4313.10 ± 12.49 ± 3.45
18 - 20	3659.29 ± 11.70 ± 2.92	3718.05 ± 11.27 ± 2.88	3692.46 ± 11.34 ± 2.89
20 - 22.5	3902.40 ± 10.78 ± 2.42	3947.54 ± 10.42 ± 2.47	3927.12 ± 10.43 ± 2.45
22.5 - 25	3286.05 ± 10.18 ± 2.19	3300.98 ± 9.84 ± 2.17	3295.71 ± 9.88 ± 2.17
25 - 27.5	2804.75 ± 9.65 ± 2.06	2800.36 ± 9.25 ± 2.07	2803.97 ± 9.33 ± 2.08
27.5 - 30	2388.17 ± 9.22 ± 1.85	2383.65 ± 8.79 ± 1.88	2385.77 ± 8.86 ± 1.90
30 - 33	2429.41 ± 8.32 ± 1.93	2414.33 ± 7.85 ± 1.97	2424.91 ± 7.94 ± 2.01
33 - 36	2038.63 ± 8.48 ± 1.96	2022.66 ± 7.95 ± 1.97	2031.31 ± 8.06 ± 1.95
36 - 39	1727.79 ± 7.91 ± 1.96	1702.47 ± 7.38 ± 2.09	1716.68 ± 7.52 ± 2.03
39 - 42	1462.52 ± 7.09 ± 2.08	1441.44 ± 6.83 ± 2.13	1452.77 ± 6.95 ± 2.14
42 - 45	1248.69 ± 5.91 ± 1.98	1227.51 ± 6.37 ± 1.99	1238.56 ± 5.82 ± 2.01
45 - 48	1054.21 ± 5.56 ± 1.91	1035.66 ± 5.96 ± 1.92	1047.23 ± 5.45 ± 1.95
48 - 51	896.74 ± 5.31 ± 1.84	878.69 ± 5.75 ± 1.85	888.54 ± 5.19 ± 1.85
51 - 54	760.82 ± 5.11 ± 1.75	748.87 ± 5.64 ± 1.76	755.69 ± 5.13 ± 1.78
54 - 57	648.49 ± 4.92 ± 1.65	639.07 ± 5.49 ± 1.64	644.68 ± 4.95 ± 1.65
57 - 61	752.81 ± 4.87 ± 1.47	743.34 ± 5.52 ± 1.46	748.72 ± 4.91 ± 1.49
61 - 65	635.33 ± 4.49 ± 1.39	627.20 ± 5.14 ± 1.42	631.26 ± 4.59 ± 1.41
65 - 70	675.75 ± 4.45 ± 1.44	668.41 ± 5.11 ± 1.42	672.56 ± 4.56 ± 1.48
70 - 75	565.29 ± 3.87 ± 1.32	559.83 ± 4.39 ± 1.36	562.55 ± 3.95 ± 1.39
75 - 80	477.56 ± 3.83 ± 1.31	472.52 ± 4.38 ± 1.34	475.22 ± 3.90 ± 1.35
80 - 85	405.52 ± 3.72 ± 1.27	401.21 ± 4.14 ± 1.27	403.37 ± 3.74 ± 1.26
85 - 95	652.50 ± 3.52 ± 1.31	646.74 ± 3.94 ± 1.29	649.79 ± 3.56 ± 1.28
95 - 105	486.29 ± 3.40 ± 1.19	480.51 ± 3.81 ± 1.17	483.46 ± 3.44 ± 1.14
105 - 125	656.46 ± 2.99 ± 1.18	650.20 ± 3.31 ± 1.11	653.31 ± 3.01 ± 1.10
125 - 150	463.29 ± 2.69 ± 1.03	456.41 ± 2.92 ± 0.97	459.76 ± 2.64 ± 0.92
150 - 175	262.27 ± 2.42 ± 0.88	258.07 ± 2.54 ± 0.81	260.14 ± 2.34 ± 0.79
175 - 200	151.90 ± 2.08 ± 0.66	149.04 ± 2.09 ± 0.61	150.67 ± 1.99 ± 0.58
200 - 250	152.32 ± 1.83 ± 0.45	149.29 ± 1.84 ± 0.50	151.30 ± 1.74 ± 0.45
250 - 300	60.05 ± 1.57 ± 0.25	59.32 ± 1.57 ± 0.25	59.87 ± 1.49 ± 0.24
300 - 350	29.58 ± 1.30 ± 0.18	28.92 ± 1.33 ± 0.18	29.44 ± 1.26 ± 0.17
350 - 400	14.25 ± 0.95 ± 0.10	13.84 ± 0.95 ± 0.11	14.32 ± 0.93 ± 0.11
400 - 470	8.61 ± 0.79 ± 0.07	8.32 ± 0.81 ± 0.07	8.52 ± 0.78 ± 0.07
470 - 550	4.60 ± 0.71 ± 0.06	4.45 ± 0.71 ± 0.06	4.55 ± 0.69 ± 0.04
550 - 650	2.04 ± 0.55 ± 0.03	2.06 ± 0.55 ± 0.03	2.06 ± 0.54 ± 0.04
650 - 900	1.25 ± 0.47 ± 0.03	1.20 ± 0.47 ± 0.03	1.22 ± 0.46 ± 0.03

Table 9.16: Measured normalized differential cross-section vs. ϕ^*_η, where the results are shown for born-level, the bare-level and the dressed-level definition of leptons. The uncertainties are split-up between bin-to-bin uncorrelated uncertainties (dominated by data and MC statistics) and bin-to-bin correlated uncertainties

Binning	Born-level val ± uncor-unc ± cor-unc [10^{-5}]	Bare-level val ± uncor-unc ± cor-unc [10^{-5}]	Dressed-level val ± uncor-unc ± cor-unc [10^{-5}]
0.0 - 0.004	3521.08 ± 8.97 ± 1.11	3493.06 ± 8.03 ± 0.86	3492.33 ± 8.13 ± 0.86
0.004 - 0.008	3481.22 ± 8.70 ± 1.32	3455.20 ± 7.57 ± 1.51	3454.91 ± 7.60 ± 1.42
0.008 - 0.012	3417.09 ± 7.72 ± 0.80	3395.99 ± 6.63 ± 0.82	3396.13 ± 6.78 ± 0.80
0.012 - 0.016	3332.64 ± 8.28 ± 0.80	3312.71 ± 7.21 ± 0.86	3312.71 ± 7.38 ± 0.81
0.016 - 0.02	3234.55 ± 8.01 ± 0.98	3217.08 ± 7.02 ± 0.88	3215.93 ± 7.14 ± 0.88
0.02 - 0.024	3118.63 ± 7.88 ± 1.09	3102.76 ± 6.87 ± 0.96	3102.57 ± 7.03 ± 0.94
0.024 - 0.029	3712.84 ± 7.82 ± 0.81	3696.22 ± 7.07 ± 0.79	3695.87 ± 7.19 ± 0.80
0.029 - 0.034	3518.67 ± 7.48 ± 0.68	3502.23 ± 6.75 ± 0.70	3503.00 ± 6.90 ± 0.71
0.034 - 0.039	3294.80 ± 7.66 ± 0.64	3289.32 ± 6.92 ± 0.69	3289.38 ± 7.01 ± 0.70
0.039 - 0.045	3703.67 ± 7.29 ± 0.78	3694.78 ± 6.64 ± 0.64	3693.68 ± 6.74 ± 0.67
0.045 - 0.051	3423.22 ± 7.15 ± 0.52	3420.77 ± 6.49 ± 0.53	3419.66 ± 6.64 ± 0.53
0.051 - 0.057	3165.65 ± 7.20 ± 0.66	3163.85 ± 6.65 ± 0.73	3165.65 ± 6.84 ± 0.75
0.057 - 0.064	3393.30 ± 7.18 ± 0.72	3395.37 ± 6.66 ± 0.61	3395.55 ± 6.85 ± 0.65
0.064 - 0.072	3536.78 ± 7.17 ± 0.44	3539.48 ± 6.64 ± 0.42	3541.26 ± 6.80 ± 0.43
0.072 - 0.081	3578.06 ± 7.36 ± 0.29	3582.16 ± 6.81 ± 0.24	3584.07 ± 6.93 ± 0.23
0.081 - 0.091	3552.77 ± 7.36 ± 0.28	3559.80 ± 6.83 ± 0.32	3560.26 ± 6.96 ± 0.27
0.091 - 0.102	3457.91 ± 7.16 ± 0.76	3467.85 ± 6.60 ± 0.43	3468.78 ± 6.78 ± 0.52
0.102 - 0.114	3315.88 ± 7.53 ± 0.60	3324.24 ± 6.99 ± 0.40	3325.77 ± 7.17 ± 0.42
0.114 - 0.128	3381.74 ± 7.84 ± 0.41	3395.22 ± 7.35 ± 0.31	3395.84 ± 7.52 ± 0.34
0.128 - 0.145	3508.30 ± 7.70 ± 0.43	3518.18 ± 7.25 ± 0.38	3521.18 ± 7.45 ± 0.39
0.145 - 0.165	3458.22 ± 8.15 ± 0.53	3473.55 ± 7.73 ± 0.47	3474.68 ± 7.93 ± 0.48
0.165 - 0.189	3428.32 ± 7.77 ± 0.60	3441.94 ± 7.37 ± 0.57	3443.95 ± 7.55 ± 0.59
0.189 - 0.219	3445.28 ± 7.53 ± 0.71	3456.92 ± 7.09 ± 0.68	3460.40 ± 7.23 ± 0.67
0.219 - 0.258	3445.60 ± 7.35 ± 0.92	3459.31 ± 6.96 ± 0.84	3460.67 ± 7.08 ± 0.84
0.258 - 0.312	3492.82 ± 7.03 ± 1.12	3502.60 ± 6.73 ± 0.93	3503.48 ± 6.84 ± 0.98
0.312 - 0.391	3465.67 ± 6.46 ± 1.26	3476.51 ± 6.26 ± 1.07	3477.62 ± 6.37 ± 1.19
0.391 - 0.524	3472.45 ± 6.07 ± 1.37	3484.33 ± 5.91 ± 1.15	3483.96 ± 6.01 ± 1.33
0.524 - 0.695	2413.86 ± 5.71 ± 1.45	2423.15 ± 5.56 ± 1.23	2420.44 ± 5.63 ± 1.40
0.695 - 0.918	1663.29 ± 4.86 ± 1.56	1670.50 ± 4.70 ± 1.34	1667.23 ± 4.76 ± 1.46
0.918 - 1.153	949.87 ± 4.18 ± 1.55	955.91 ± 4.06 ± 1.39	952.13 ± 4.13 ± 1.53
1.153 - 1.496	741.66 ± 3.74 ± 1.35	746.65 ± 3.65 ± 1.27	743.88 ± 3.68 ± 1.39
1.496 - 1.947	487.76 ± 3.06 ± 1.15	491.83 ± 2.98 ± 1.12	489.56 ± 3.00 ± 1.21
1.947 - 2.522	309.36 ± 2.72 ± 0.94	311.49 ± 2.70 ± 0.91	310.40 ± 2.69 ± 0.95
2.522 - 3.277	206.94 ± 2.50 ± 0.83	208.09 ± 2.51 ± 0.80	207.30 ± 2.50 ± 0.78
3.277 - 5.0	200.97 ± 2.29 ± 0.73	201.91 ± 2.28 ± 0.69	201.06 ± 2.28 ± 0.66
5.0 - 10.0	169.14 ± 2.06 ± 0.56	168.99 ± 2.02 ± 0.54	168.61 ± 2.03 ± 0.57

9.13 Combined differential measurement

The combination of the two Drell-Yan lepton pair channels is ther main result of Ref. [54], hence this section is based on it. The combination is performed using χ^2 minimization, following the best linear unbiased estimator prescription (Blue) [120, 132, 162]. The following explanation is based on Ref. [120]. This approach searches for a combined stimate $\hat{y}$, which:

- is a linear combination of the channels y_i to be combined,
- provides an unbiassed estimate and
- has a minimal variance.

In the context of this project, channel refers to the variations of both the muon and the electron channel. Following the first criteria, $\hat{y}$ can be written as:

$$\hat{y} = \sum_i \alpha_i \cdot y_i, \tag{9.26}$$

with α_i being weighting factors of the different channels fullfilling $\sum \alpha_i = 1$, as the estimate has to be unbiassed. The variance of this approach, which is minimized, is given by:

$$\sigma^2 = \vec{\alpha}^{\mathrm{T}} \cdot \mathbf{E} \cdot \vec{\alpha} = \sum_i \sum_j Eij \cdot \alpha_i \cdot \alpha_j \tag{9.27}$$

with $\vec{\alpha}$ being the vector of weights α_i and $\mathbf{E}$ the error matrix. This matrix contains the variances of the individual estimates as diagonal elements, while the off-diagonal entries describe the correlations between pairs of estimates. The resulting $\hat{y}$ is optimized by performing χ^2 minimization. The weights α_i are applied for each bin. A detailed breakdown of the results are shown in Tables A.12, A.13 and A.14 for $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ as well as Tables A.15, A.16 and A.17 for ϕ^*_η in the appendix.

For the combination, the uncertainties are separated into two groups, one set of bin-to-bin uncorrelated sources and another set of bin-to-bin correlated sources, the latter being primarily reduced due to the normalization by the fiducial cross-section. Only the pile-up, physics modeling and luminosity uncertainties are treated as correlated between the two decay channels. The combination of the two channels leads to a reduction of the systematic uncertainty compared to the individual electron- and muon-channel measurements. The dominant detector-related systematic uncertainties are treated as largely uncorrelated.

The χ^2 minimization process results in a $\chi^2/N_{\mathrm{dof}} = 47/44$ for the combination of the $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ distribution and a $\chi^2/N_{\mathrm{dof}} = 32/36$ for the ϕ^*_η distribution. These values, being both close to 1, indicate a good agreement between the two channels. The combination is performed on born-level. All results presented here are published in Ref. [54].

The normalized differential cross-sections measured in both the $Z \to \mu\mu$ and the $Z \to ee$ channel are compared to the resulting combination in Figure 9.26. Hereby the two channels differ between 0.1% and 0.5% for $\mathrm{p}_{\mathrm{T}}^{\ell\ell} < 100\,\mathrm{GeV}$, rising to 10% for larger transverse momenta. As observed before, the overall precision in this regime is limited by the data and MC statistics. This effect can be seen in the ratio plot of this Figure. The pull distribution between the electron and muon channel is shown below the ratio plot to visualize the contributions of each channel further. The pull is defined as the difference between the two channels divided by the combined uncorrelated uncertainty.

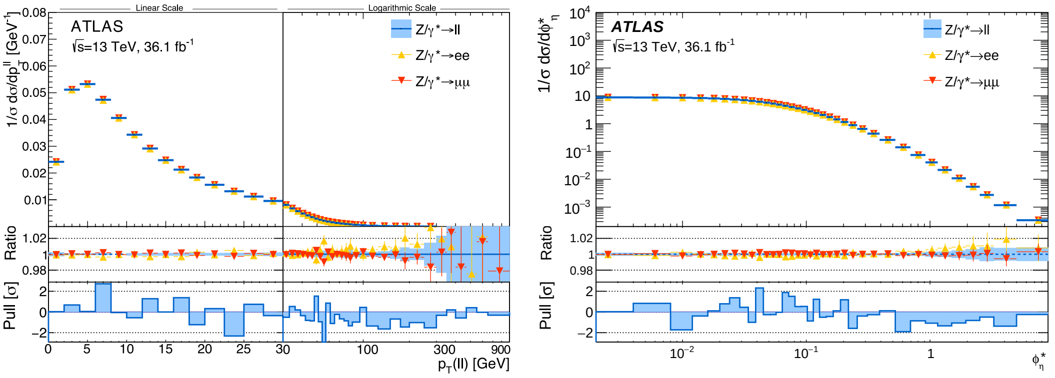


Figure 9.26: Measured normalized cross-section as a function of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$(left) and ϕ^*_η(right), taken from Ref. [54]. The results are presented for both the electron and muon channel as well as the combined result. Additionally the ratio of the channels and the combined result is shown together with the total uncertainties, shown as a blue band. The contribution of each channel is shown by the Pull distribution.

In addition to this, the values of the combined differential cross-section are bin-wise presented in Table 9.17. This table additionally provides the corresponding statistical as well as the bin-to-bin uncorrelated and correlated systematic uncertainties. In order to be able to transfer the result on born-level to the dressed-level, the k_{dr} factor is given. Further tables containing more detailed informations can be found in Section A.4.4 of the appendix.

9.14 Comparison to theoretical predictions

As stated in the motivation, differential cross-sections of the Drell-Yan lepton pairs are on the one hand needed for W boson related measurements. Additionally, they are useful to tune MC generators due to the clean signature in the detector. This section focuses on the comparison of a variety of different MC generator predictions to the obtained combined measurement. These generators are Powheg+Pythia, Sherpa, Pythia8 and RadISH. The generators are based on different theoretical approaches.

Table 9.17: The measured combined normalized differential cross-sections, divided by the bin-width, in the fiducial volume at born-level as well as a factor k_{dr} to translate from the Born particle-level to the dressed particle-level. Shown in bins of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ (left column) and ϕ^*_η (right column). The table is taken from Ref. [54].

Bin [GeV]	$1/\sigma_{\mathrm{fid}} \times$ $\mathrm{d}\sigma/\mathrm{dp}_{\mathrm{T}}^{\mu\mu}$ [1/GeV]	Corr. uncert.	Uncorr. uncert.	k_{dr}	Bin	$1/\sigma_{\mathrm{fid}} \times$ $\mathrm{d}\sigma/\mathrm{d}\phi^*_{\mu\mu}$	Corr. uncert.	Uncorr. uncert.	k_{dr}
0-2	0.024189	± 0.15%	± 0.18%	0.978	0-0.004	8.8053	± 0.03%	± 0.13%	0.992
2-4	0.051144	± 0.06%	± 0.08%	0.985	0.004-0.008	8.6969	± 0.03%	± 0.13%	0.993
4-6	0.053232	± 0.05%	± 0.08%	0.994	0.008-0.012	8.5624	± 0.02%	± 0.13%	0.993
6-8	0.047383	± 0.05%	± 0.08%	1.000	0.012-0.016	8.3378	± 0.02%	± 0.13%	0.994
8-10	0.040568	± 0.04%	± 0.09%	1.010	0.016-0.02	8.0881	± 0.03%	± 0.14%	0.994
10-12	0.034317	± 0.06%	± 0.11%	1.010	0.02-0.024	7.7920	± 0.03%	± 0.14%	0.995
12-14	0.029157	± 0.07%	± 0.12%	1.010	0.024-0.029	7.4174	± 0.02%	± 0.12%	0.995
14-16	0.024804	± 0.06%	± 0.14%	1.010	0.029-0.034	7.0360	± 0.02%	± 0.13%	0.996
16-18	0.021268	± 0.05%	± 0.15%	1.010	0.034-0.039	6.5989	± 0.02%	± 0.13%	0.998
18-20	0.018325	± 0.04%	± 0.16%	1.010	0.039-0.045	6.1608	± 0.02%	± 0.12%	0.998
20-22.5	0.015605	± 0.03%	± 0.14%	1.010	0.045-0.051	5.7085	± 0.01%	± 0.13%	0.999
22.5-25	0.01318 0	± 0.03%	± 0.15%	1.000	0.051-0.057	5.2791	± 0.02%	± 0.14%	1.000
25-27.5	0.011207	± 0.04%	± 0.17%	1.000	0.057-0.064	4.8488	± 0.02%	± 0.13%	1.000
27.5-30	0.0095568	± 0.05%	± 0.19%	0.999	0.064-0.072	4.4139	± 0.01%	± 0.12%	1.000
30-33	0.0081029	± 0.06%	± 0.17%	0.998	0.072-0.081	3.9705	± 0.01%	± 0.12%	1.000
33-36	0.0067881	± 0.08%	± 0.19%	0.996	0.081-0.091	3.5515	± 0.01%	± 0.12%	1.000
36-39	0.0057563	± 0.09%	± 0.21%	0.994	0.091-0.102	3.1421	± 0.02%	± 0.13%	1.000
39-42	0.0048769	± 0.12%	± 0.23%	0.993	0.102-0.114	2.7659	± 0.01%	± 0.13%	1.000
42-45	0.0041688	± 0.12%	± 0.25%	0.992	0.114-0.128	2.4125	± 0.01%	± 0.13%	1.000
45-48	0.0035213	± 0.14%	± 0.28%	0.993	0.128-0.145	2.0648	± 0.01%	± 0.12%	1.000
48-51	0.0029751	± 0.17%	± 0.31%	0.991	0.145-0.165	1.7299	± 0.02%	± 0.13%	1.000
51-54	0.0025433	± 0.18%	± 0.35%	0.992	0.165-0.189	1.4282	± 0.02%	± 0.13%	1.000
54-57	0.0021832	± 0.20%	± 0.38%	0.994	0.189-0.219	1.1469	± 0.02%	± 0.12%	1.000
57-61	0.0018779	± 0.15%	± 0.31%	0.994	0.219-0.258	0.8848	± 0.02%	± 0.12%	1.000
61-65	0.0015932	± 0.17%	± 0.35%	0.994	0.258-0.312	0.6470	± 0.03%	± 0.11%	1.000
65-70	0.0013519	± 0.16%	± 0.32%	0.995	0.312-0.391	0.4387	± 0.03%	± 0.11%	1.000
70-75	0.0011323	± 0.17%	± 0.37%	0.995	0.391-0.524	0.2610	± 0.03%	± 0.10%	1.000
75-80	0.0009574	± 0.20%	± 0.43%	0.995	0.524-0.695	0.1414	± 0.04%	± 0.13%	1.000
80-85	0.0008150	± 0.22%	± 0.49%	0.995	0.695-0.918	0.07462	± 0.07%	± 0.17%	1.000
85-95	0.0006537	± 0.14%	± 0.29%	0.996	0.918-1.153	0.04047	± 0.12%	± 0.27%	1.000
95-105	0.0004849	± 0.18%	± 0.37%	0.995	1.153-1.496	0.02167	± 0.14%	± 0.30%	1.000
105-125	0.0003291	± 0.12%	± 0.25%	0.996	1.496-1.947	0.01084	± 0.18%	± 0.42%	1.000
125-150	0.0001861	± 0.16%	± 0.32%	0.994	1.947-2.522	0.005386	± 0.23%	± 0.59%	1.000
150-175	0.0001050	± 0.24%	± 0.51%	0.993	2.522-3.277	0.002738	± 0.31%	± 0.79%	1.000
175-200	$6.1279{\cdot}10^{-5}$	± 0.30%	± 0.78%	0.992	3.277-5.000	0.0011730	± 0.29%	± 0.72%	1.000
200-250	$3.0584{\cdot}10^{-5}$	± 0.22%	± 0.66%	0.995	5.000-10.00	0.0003372	± 0.30%	± 0.78%	0.997
250-300	$1.2211{\cdot}10^{-5}$	± 0.34%	± 1.4%	0.997					
300-350	$5.9026{\cdot}10^{-6}$	± 0.56%	± 2.3%	0.994					
350-400	$2.7742{\cdot}10^{-6}$	± 0.90%	± 3.8%	0.991					
400-470	$1.2513{\cdot}10^{-6}$	± 0.82%	± 4.9%	0.991					
470-550	$5.5219{\cdot}10^{-7}$	± 1.2%	± 7.9%	0.994					
550-650	$2.0165{\cdot}10^{-7}$	± 1.5%	± 13%	0.995					
650-900	$5.1153{\cdot}10^{-8}$	± 1.8%	± 16%	0.990					
900-2500	$1.5735{\cdot}10^{-9}$	± 6.3%	± 60%	0.964					

This way, both the soft and hard emissions from ISR are taken into account. Unless stated otherwise, the predictions do not consider NNLO electroweak effects. The generator descriptions and prediction discussion are taken from Ref. [54], which provides in-detail information about the generators and software packages used, as the predictions were not created in the context of this thesis, but are provided by the ATLAS community. Figure 9.27 shows both the predictions and the combined measurement in bins of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ and ϕ^*_η combined cross-section measurement including the ratios of predictions and data.

The first two predictions were created using the same generator setup as used for the W-boson mass measurement at 7 TeV LHC data [14]. These were chosen as this measurement required a high-precision description of the W-boson transverse momentum spectrum at low $\mathrm{p}_{\mathrm{T}}^{\ell\nu}$. Pythia8 with matrix elements at leading-order in α_s, supplemented with a parton shower with the AZ set of tuned parameters [15], as well as Powheg+Pythia8 using NLO matrix elements with the Pythia8 parton shower parameters set and AZNLO tune [15] are used. Both tunes are optimized to describe the ATLAS 7 TeV $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ and ϕ^*_η data [13, 15].

The third prediction is simulated using the Sherpa v2.2.1 [32] generator. Used PDF and parton shower are described in Section 9.4. Furthermore, the Comix [101] and OpenLoops [43, 81] libraries are used to calculate NLO-accurate matrix elements for up to two partons as well as LO-accurate matrix elements for up to four partons. Uncertainties from missing higher-orders are evaluated [33] using seven variations of the QCD factorization and renormalization scales in the matrix elements by factors of 0.5 and 2, avoiding variations in opposite directions. The effect of each variation is fully correlated across the full spectrum. The envelope of all variations is taken as final uncertainty, resulting in uncertainties of 3–4% in the low-bins of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ and ϕ^*_η, rising to 25% in the high bins. The effects of PDF uncertainties are found to be very small, typically $< 1\%$ for low bins, and a few percent above.

The fourth prediction is created using the RadISH program [29, 30], which combines a fixed-order NNLO prediction of Z+jet production ($O(\alpha_s^3)$) from NNLOjet [99] with resummation of $\log(m^{\ell\ell}/\mathrm{p}_{\mathrm{T}}^{\mu\mu})$ terms at next-to-next-to-next- to-leading-logarithm (N^3LL) accuracy [31] using the The NNPDF3.1nnlo PDFs set [24]. Uncertainties for this prediction are derived from variations of μ_R and μ_F in the same way as for the Sherpa prediction described above and, also, two variations of Q by a factor of two up and down, assuming that the effects of scale variations are fully correlated across the full spectrum.

Pythia-based predictions describe the 13 TeV data to within 2–4% for $\mathrm{p}_{\mathrm{T}}^{\ell\ell} < 40\,\mathrm{GeV}$ and $\phi^*_\eta < 0.5$. The prediction using the AZ tune is better-suited modeling the measured data, showing that predictions based on 7 TeV collision data can be used for higher center-of-mass energies data in the low $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ regime. In higher bins, the prediction underestimates the data due to missing higher-order matrix elements.

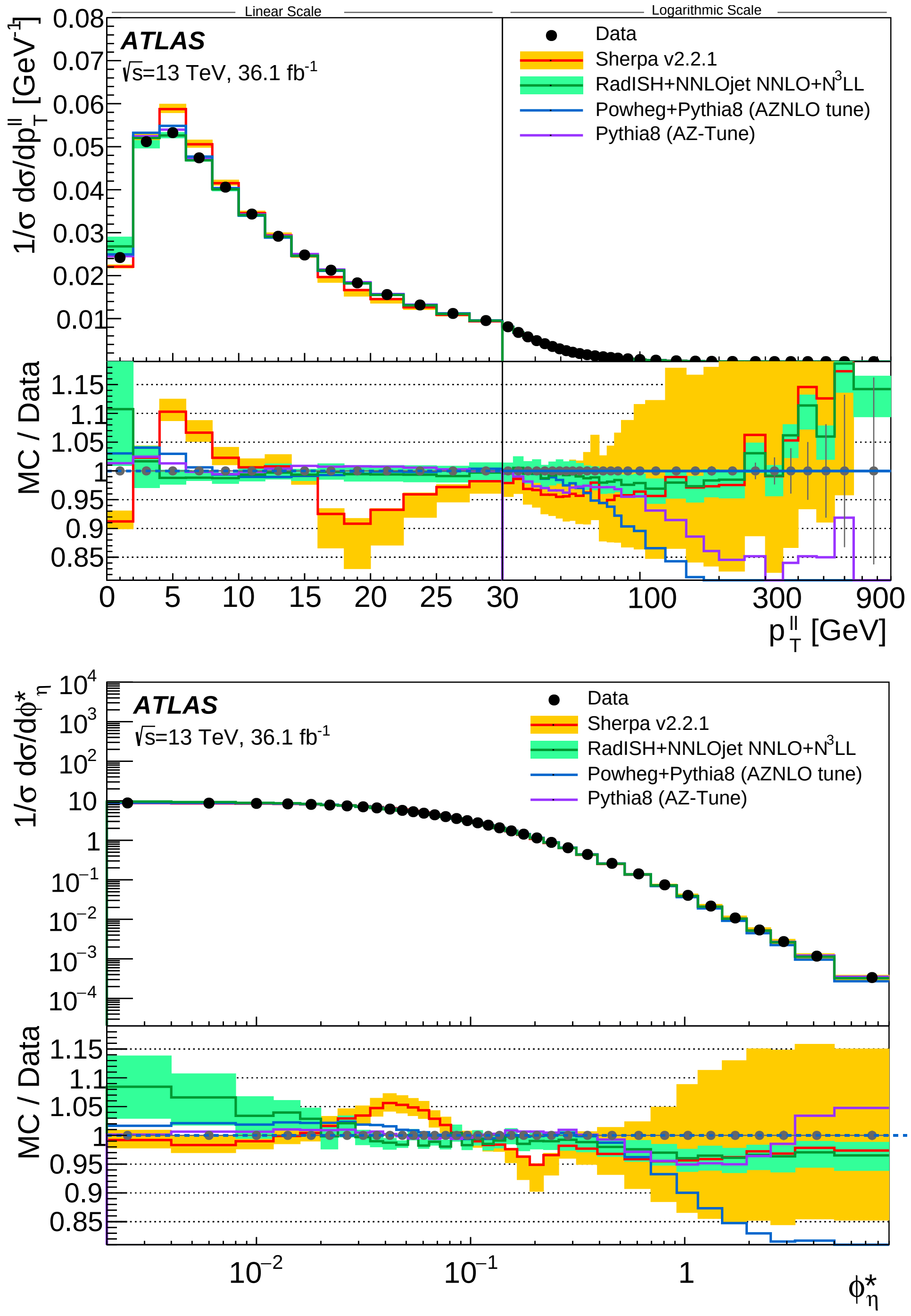


Figure 9.27: Comparison of the normalized $p_T^{\ell\ell}$(upper) and ϕ^*_η(middle) distributions predicted by different computations to the born-level combined measurement. The uncertainties of the measurement are shown as vertical bars. Taken from Ref. [54]

The SHERPA prediction does not sufficiently well model the normalized spectrum in the low-bin regime of $p_T^{\mu\mu} < 25\,\text{GeV}$ and $\phi^*_\eta < 0.1$. Above these bins, the data is modeled to within $\approx 4\,\%$ up to the point where statistical uncertainties in the data become significant, which is better than the uncertainty estimate obtained from scale variations. Lastly, the RADISH prediction describes the data well over the full spectrum with uncertainties of typically 1–3%, highlighting the benefits of this prediction.

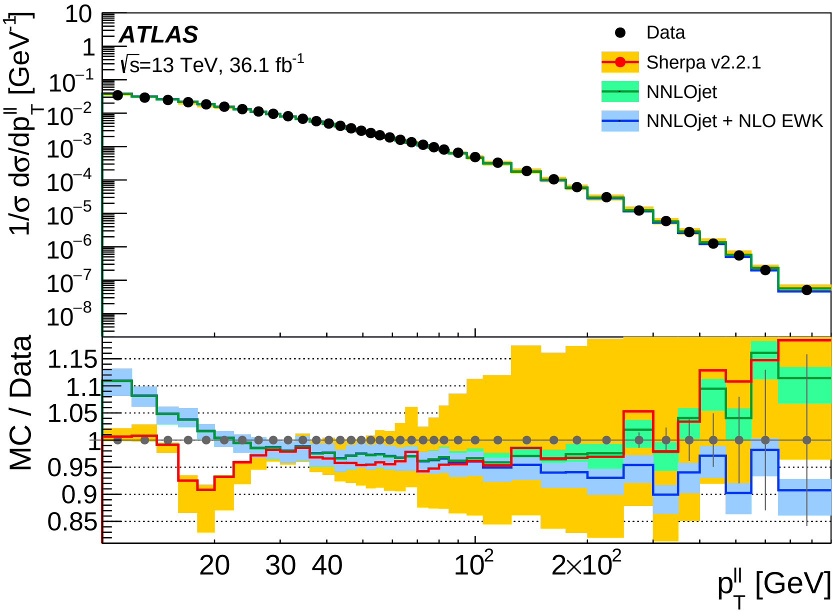


Figure 9.28: Comparison of the normalized $p_T^{\ell\ell}$ distributions predicted by different computations to the born-level combined measurement in the range $p_T^{\mu\mu} > 10\,\text{GeV}$. The uncertainties of the measurement are shown as vertical bars. The theoretical uncertainties on the textscNNLOjet predictions are not shown, but have been estimated to be up to 5% for $p_T^{\ell\ell} \approx 1\,\text{TeV}$ [119]. Taken from Ref. [54].

In a second study, the $p_T^{\ell\ell}$ measurement is compared to predictions in bins of $p_T^{\mu\mu} > 10\,\text{GeV}$, as the NNLOJET prediction used in RADISH is only expected to describe the data at sufficiently large $p_T^{\ell\ell} \gtrsim 15\,\text{GeV}$ [99]. In addition to the SHERPA prediction, the data is compared to fixed-order NNLOJET predictions both with and without NLO electroweak corrections [82]. As shown in Figure 9.28, the application of NLO electroweak corrections lead to a reduction of up to 20% at the higher end of the $p_T^{\ell\ell}$ spectrum. In this region, NNLOJET predictions without NLO EW corrections generally overestimate while including the corrections lead to an underestimation. The difference between those is not more significant than the uncertainties in the measurements.

9.15 Summary of the Cross-Section Measurement

This chapter presented the result of the high precision measurement of the cross-section of the Drell-Yan lepton pair production process. The cross-sections were measured both integrated and differential in bins of $\mathrm{p}_\mathrm{T}^{\mu\mu}$ and ϕ^*_η using $36.1\,\mathrm{fb}^{-1}$ of data from proton-proton collisions recorded in 2015 and 2016 at a center-of-mass energy of 13 TeV with the ATLAS experiment. A series of kinematic cuts define the fiducial volume. Those affect the lepton transverse momentum $\mathrm{p}_\mathrm{t}^{\ell} > 27\,\mathrm{GeV}$ as well as its $|\eta_\ell| < 2.5$ while the analysis is limited to the Z peak region $66 < m^{\ell\ell} < 116\,\mathrm{GeV}$.

Utilizing the 2015 and 2016 dataset allows this analysis to cover a kinematic range up to the TeV-regime and dedicated MC samples are used to correct for background processes and detector inefficiencies. The cross-section results from the muon and electron channels, which were both measured separately, were combined using Blue and show good agreement between the two channels. As the uncertainties limiting this measurement are mostly uncorrelated, the combination increased the precision of the analysis. Its normalized uncertainty is lower than 0.2% for $\mathrm{p}_\mathrm{T}^{\ell\ell} < 30\,\mathrm{GeV}$, providing crucial information to validate and tune MC event generators. Furthermore, it can be used to constrain models of vector-boson production, which are essential for measurements of the W-boson mass as well as searches for new physics at high $\mathrm{p}_\mathrm{T}^{\ell\ell}$.

The results of both the fiducial integrated and normalized differential cross-section measurements are compared to theoretical predictions. The measured integrated cross-section is compared to a fixed-order perturbative QCD prediction, showing good agreement within uncertainties. The measured differential cross-sections in bins of both $\mathrm{p}_\mathrm{T}^{\ell\ell}$ and ϕ^*_η are compared to a set of four different theoretical predictions.

Pythia8 parton shower based predictions using parameters tuned to 7 TeV data are found to describe the 13 TeV data well in the low-bin regime of both $\mathrm{p}_\mathrm{T}^{\ell\ell}$ and ϕ^*_η. On the other hand, both Sherpa based predictions and fixed-order NNLOjet predictions provide a good description of the data at high $\mathrm{p}_\mathrm{T}^{\ell\ell}$. Predictions calculated using RadISH agree with the measured result over the full spectrum.

Chapter 10

Conclusion and Outlook

This thesis covers two approaches to improving the measurement of the W boson mass utilizing Z boson Drell-Yan process data. Both projects aim to reduce the model uncertainty by improving measurements of the W boson transverse momentum and resulted in dedicated main-author publications, namely Ref. [34] and Ref [54].

The first approach investigates the application of deep convolutional neural networks to mitigate the effects of pileup and enables a direct measurement of the W boson transverse momentum. Therefore, a first study of the application of image recognition techniques and deep convolutional neural networks on event images has been performed. A 10 layer artificial neural network is trained on binned event images with 50 bins in η and 64 bins in φ to predict the hard scatter particle-level image without pileup contributions. Two ANNs are trained, one utilizing a global, eventwise physics parameter, the missing transverse momentum, as loss function, while the second network describes the loss by local bin fluctuations. The ANNs show a more stable behavior compared to other pileup mitigation techniques like Voronoi subtraction, Constituent Subtraction or SoftKiller, but were not able to fully reconstruct events without bias. Additional application fields can be found in a low-level trigger environment, where speed and implementation size of an algorithm is more critical than maximum precision due to the binned nature of this approach and the small size of the network.

In the context of the W boson transverse momentum measurement, this approach reduces bin migration effects compared to the measurement using the hadronic recoil. It shows a neglectable dependence on pileup activity. On the downside, this approach still shows a significant model dependence due to the choice of the particle-level definition and the limitation to the ATLAS barrel region.

Until these rather technical than intrinsic limitations are solved, a better approach of limiting the model uncertainty consists of tuning MonteCarlo generators. This is done using results obtained by Drell-Yan lepton pair production processes, which are kinematically similar to leptonic W boson decays.

Such a measurement is described in the second part of this thesis, measuring the Drell-Yan lepton pair production cross-section both differential in bins of $\mathrm{p}_{\mathrm{T}}^{\ell\ell}$ and ϕ_{η}^{*} and integrated over the fiducial volume. Hereby, the fiducial volume of the measurement is defined by the Z peak region $66\,\mathrm{GeV} < m^{\ell\ell} < 116\,\mathrm{GeV}$ and the lepton being reconstructed in the ATLAS barrel region $\left|\eta^{\ell}\right| < 2.5$. The measurement was performed for the Drell-Yan muon decay channel and later combined with the measured results of the electron decay channel.

The integrated cross-section measurement resulted in $736.2 \pm 0.2 \pm 6.4 \pm 15.5\,\mathrm{pb}$. Uncertainties covering both statistical and systematical influences limiting the precision of this measurement are studied and discussed, resulting in an uncertainty of below $0.3\,\%$ for most of the bins of the measured normalized differential cross-section, one of the most precise measurements within ATLAS thus far. The application of this result in the context of a new W boson mass measurement using the ATLAS detector is currently ongoing.

Appendix A

Drell-Yan analysis material

Contents

A.1 GoodRunLists

The GoodRunLists for 2015 and 2016 are summarized below:

Table A.1: 2015 and 2016 GoodRunLists

Year	GoddRunList
2015	`data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml`
2016	`data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml`

A.2 Data samples

Events recorded during data taking are available in containers divided per runs or per periods. The DAOD_STDM3 group derivation is used in both the electron and muon channels, re-processed with p-tag 2950. The data containers have the following format: `data1*_13TeV.*.physics_Main.merge.DAOD_STDM3.*_p2950/`.

Table A.2: Data containers included in analysis in $Z \to \mu\mu$ channel

Dataset container
data15_13TeV.periodD.physics_Main.PhysCont.DAOD_STDM3.grp15_v01_p2950
data15_13TeV.periodE.physics_Main.PhysCont.DAOD_STDM3.grp15_v02_p2950
data15_13TeV.periodF.physics_Main.PhysCont.DAOD_STDM3.grp15_v01_p2950
data15_13TeV.periodG.physics_Main.PhysCont.DAOD_STDM3.grp15_v01_p2950
data15_13TeV.periodH.physics_Main.PhysCont.DAOD_STDM3.grp15_v01_p2950
data15_13TeV.periodJ.physics_Main.PhysCont.DAOD_STDM3.grp15_v01_p2950
data16_13TeV.periodA.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodB.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodC.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodD.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodE.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodF.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodG.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodI.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodK.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
data16_13TeV.periodL.physics_Main.PhysCont.DAOD_STDM3.grp16_v01_p2950
Total: $\int \mathcal{L} = 36\,074.56\,\mathrm{pb}^{-1}$

A.3 Monte Carlo Samples

The MC samples: And the Sherpa signal samples:

Table A.3: Cross section, filtering efficiency and k-factors for Powheg signal and background samples

Process	Generator	DSID	Cross-section [pb]	GenFiltEff	k-factor	tag
$Z \to \mu\mu$	Powheg	361107	1901.2	1.0	1.026	e3601_s2576_s2132_r7725_r7676_p2952
$t\bar{t}$	Powheg	410501	730.19	0.54383	1.139	e5458_s2726_r7772_r7676_p3314
t. chan.	Powheg	410011	43.739	1.0	1.009	e3824_s2608_s2183_r7725_r7676_p3314
	Powheg	410012	25.778	1.0	1.019	e3824_s2608_s2183_r7725_r7676_p3314
Wt. chan.	Powheg	410013	34.009	1.0	1.054	e3753_s2608_s2183_r7725_r7676_p3314
	Powheg	410014	33.989	1.0	1.054	e3753_s2608_s2183_r7725_r7676_p3314
$W^+ \to e\nu$	Powheg	361100	11306	1.0	1.017	e3601_s2576_s2132_r7725_r7676_p3314
$W^+ \to \mu\nu$	Powheg	361101	11306	1.0	1.017	e3601_s2576_s2132_r7725_r7676_p3314
$W^+ \to \tau\nu$	Powheg	361102	11306	1.0	1.0172	e3601_s2576_s2132_r7725_r7676_p3314
$W^- \to e\nu$	Powheg	361103	8282.9	1.0	1.036	e3601_s2576_s2132_r7725_r7676_p3314
$W^- \to \mu\nu$	Powheg	361104	8282.9	1.0	1.036	e3601_s2576_s2132_r7725_r7676_p3314
$W^- \to \tau\nu$	Powheg	361105	8282.9	1.0	1.036	e3601_s2576_s2132_r7772_r7676_p3314
$Z \to \tau\tau$	Powheg	361108	1901.2	1.0	1.026	e3601_s2726_r7725_r7676_p3314
WWlvlv	Powheg	361600	10.637	1.0	1.0	e4616_s2726_r7772_r7676_p3314
WWlvqq	Powheg	361606	44.821	1.0	1.0	e4711_s2726_r7772_r7676_p3314
WZlvll	Powheg	361601	4.4625	1.0	1.0	e4054_s2608_s2183_r7725_r7676_p2949
WZlvvv	Powheg	361602	2.778	1.0	1.0	e4054_s2608_s2183_r7725_r7676_p3314
WZqqll	Sherpa	361607	3.2777	1.0	1.0	e4711_s2726_r7772_r7676_p3314
WZlvqq	Powheg	361609	10.086	1.0	1.0	e4711_s2726_r7772_r7676_p3314
ZZllll	Powheg	361603	1.2689	1.0	1.0	e4475_s2726_r7772_r7676_p3314
ZZvvll	Powheg	361604	0.92318	1.0	1.0	e4475_s2726_r7772_r7676_p3314
WqqZll	Sherpa	363358	3.437	1.0	1.0	e5525_s2726_r7772_r7676_p2949
ZqqZll	Sherpa	363356	15.563	0.14089	1.0	e5525_s2726_r7772_r7676_p2949
Photon induced (e channel)	Pythia	363672	17.494	0.11564	1.0	e4913_s2726_r7773_r7676_p3314
(e channel)	Pythia	363673	1.7919	0 .31443	1.0	e4913_s2726_r7773_r7676_p3314
(e channel)	Pythia	363674	0.13862	0 .37842	1.0	e4913_s2726_r7773_r7676_p3314
(mu channel)	Pythia	363675	109.170	0.072923	1.0	e4913_s2726_r7773_r7676_p3314
(mu channel)	Pythia	363676	17.477	0.26481	1.0	e4913_s2726_r7773_r7676_p3314
(mu channel)	Pythia	363677	1.791	0.34228	1.0	e4913_s2726_r7773_r7676_p3314
(mu channel)	Pythia	363678	0.13884	0.38183	1.0	e4913_s2726_r7773_r7676_p3314

Table A.4: Details of the Sherpa samples used in the analysis.

Process	DSID	$p_T^{\mu\mu}$	Veto/Filter	Cross-section [pb]	GenFiltEff	k-factor
$Z \to \mu\mu$	364100	0-70 GeV	CVetoBVeto	1982.3	0.82179	0.975
	364101	0-70 GeV	CFilterBVeto	1982.1	0.11356	0.975
	364102	0-70 GeV	BFilter	1981.6	0.06574	0.975
	364103	70-140 GeV	CVetoBVeto	109.07	0.68978	0.975
	364104	70-140 GeV	CFilterBVeto	108.94	0.19588	0.975
	364105	70-140 GeV	BFilter	108.98	0.12052	0.975
	364106	140-280 GeV	CVetoBVeto	39.884	0.60179	0.975
	364107	140-280 GeV	CFilterBVeto	39.857	0.23545	0.975
	364108	140-280 GeV	BFilter	39.892	0.15628	0.975
	364109	280-500 GeV	CVetoBVeto	8.526	0.56012	0.975
	364110	280-500 GeV	CFilterBVeto	8.5261	0.26632	0.975
	364111	280-500 GeV	BFilter	8.5276	0.17657	0.975
	364112	500-1000 GeV		1.787	1.0	0.975
	364113	1000 GeV		0.148	1.0	0.975

A.4 Drell-Yan analysis tables

A.4.1 Background contamination tables

Table A.5: Z→ $\mu\mu$ transverse momentum signal and background events (Powheg)

$p_T^{\mu\mu}$-bin	data	EWK Background (Z→ $\tau\tau$, Diboson)	Photon induced	$W \to l\nu$	Top-Quark Pairs, single top	Multijet Background
0.0-2.0	774992.0±880.3	505.4±22.5	316.8±17.8	0.0±0.0	45.0±6.7	16.0±9.8
2.0-4.0	1739841.0±1319.0	625.9±25.0	456.5±21.4	0.0±0.0	126.7±11.3	-5.6±4.9
4.0-6.0	1910029.0±1382.0	1145.9±33.9	483.3±22.0	0.0±0.0	212.7±14.6	9.1±10.8
6.0-8.0	1741572.0±1319.7	1101.3±33.2	347.7±18.6	0.0±0.0	304.2±17.4	6.5±12.0
8.0-10.0	1501796.0±1225.5	1171.8±34.2	353.1±18.8	13.7±3.7	376.1±19.4	23.4±16.3
10.0-12.0	1273443.0±1128.5	1284.2±35.8	246.3±15.7	0.0±0.0	412.9±20.3	46.0±18.7
12.0-14.0	1081168.0±1039.8	1134.3±33.7	184.0±13.6	10.1±3.2	489.7±22.1	36.5±18.7
14.0-16.0	919721.0±959.0	1088.0±33.0	175.9±13.3	16.6±4.1	544.9±23.3	26.0±16.5
16.0-18.0	786335.0±886.8	1023.5±32.0	168.4±13.0	4.8±2.2	591.9±24.3	18.9±16.0
18.0-20.0	673716.0±820.8	966.2±31.1	142.3±11.9	0.0±0.0	673.3±25.9	25.0±16.5
20.0-22.5	714042.0±845.0	1202.2±34.7	160.4±12.7	0.0±0.0	818.0±28.6	23.3±18.4
22.5-25.0	599581.0±774.3	1182.0±34.4	157.8±12.6	0.0±0.0	918.5±30.3	36.7±20.7
25.0-27.5	507623.0±712.5	1149.4±33.9	112.7±10.6	0.0±0.0	919.9±30.3	101.7±26.1
27.5-30.0	433271.0±658.2	1194.3±34.6	98.8±9.9	0.0±0.0	1008.9±31.8	7.2±17.5
30.0-33.0	440628.0±663.8	1524.7±39.0	103.0±10.2	202.5±14.2	1225.1±35.0	-77.9±95.0
33.0-36.0	370191.0±608.4	1302.1±36.1	88.1±9.4	21.0±4.6	1268.3±35.6	13.0±21.0
36.0-39.0	312420.0±558.9	1289.1±35.9	38.6±6.2	8.0±2.8	1335.0±36.5	49.3±24.0
39.0-42.0	264897.0±514.7	1271.3±35.7	55.1±7.4	0.0±0.0	1341.4±36.6	77.9±27.3
42.0-45.0	226004.0±475.4	1232.9±35.1	32.5±5.7	0.0±0.0	1347.1±36.7	28.6±23.7
45.0-48.0	192434.0±438.7	1096.1±33.1	57.5±7.6	4.5±2.1	1422.4±37.7	45.0±22.8
48.0-51.0	164631.0±405.7	990.9±31.5	24.4±4.9	0.0±0.0	1432.9±37.9	38.5±22.7
51.0-54.0	141460.0±376.1	923.6±30.4	19.8±4.5	0.0±0.0	1442.8±38.0	50.7±23.3
54.0-57.0	123215.0±351.0	910.5±30.2	17.6±4.2	0.0±0.0	1488.5±38.6	11.1±16.8
57.0-61.0	142747.0±377.8	1073.9±32.8	22.1±4.7	104.7±10.2	1969.4±44.4	45.1±24.8
61.0-65.0	121548.0±348.6	998.6±31.6	31.8±5.6	3.9±2.0	1927.7±43.9	15.0±19.7
65.0-70.0	130108.0±360.7	1107.3±33.3	24.4±4.9	6.6±2.6	2434.6±49.3	36.8±23.5
70.0-75.0	109096.0±330.3	1028.6±32.1	36.0±6.0	0.0±0.0	2319.0±48.2	64.4±26.5
75.0-80.0	93247.0±305.4	905.8±30.1	11.9±3.5	0.0±0.0	2228.8±47.2	9.0±18.8
80.0-85.0	79746.0±282.4	796.9±28.2	7.2±2.7	0.0±0.0	2061.7±45.4	22.3±21.1
85.0-95.0	128911.0±359.0	1373.7±37.1	20.4±4.5	0.0±0.0	3549.3±59.6	56.7±24.3
95.0-105.0	97469.0±312.2	1110.1±33.3	15.8±4.0	0.0±0.0	2770.7±52.6	54.6±23.4
105.0-125.0	132065.0±363.4	1708.7±41.3	23.2±4.8	0.0±0.0	3485.1±59.0	-1.0±91.5
125.0-150.0	94220.0±307.0	1386.0±37.2	12.6±3.5	0.0±0.0	1833.4±42.8	75.6±23.8
150.0-175.0	53126.0±230.5	897.5±30.0	6.2±2.5	0.0±0.0	645.4±25.4	22.1±13.9
175.0-200.0	31046.0±176.2	593.2±24.4	1.7±1.3	0.0±0.0	234.0±15.3	12.6±10.9
200.0-250.0	30908.0±175.8	662.8±25.7	5.7±2.4	13.0±3.6	156.3±12.5	9.1±9.9
250.0-300.0	12787.0±113.1	363.1±19.1	4.6±2.1	0.0±0.0	42.8±6.5	4.9±6.2
300.0-350.0	6059.0±77.8	175.2±13.2	0.7±0.8	0.0±0.0	19.1±4.4	11.5±7.5
350.0-400.0	2903.0±53.9	98.5±9.9	0.3±0.5	0.0±0.0	5.0±2.2	-0.9±0.4
400.0-470.0	1916.0±43.8	70.0±8.4	0.2±0.5	0.0±0.0	2.8±1.7	3.9±4.3
470.0-550.0	939.0±30.6	39.1±6.3	0.3±0.5	0.0±0.0	1.1±1.1	-0.2±0.2
550.0-650.0	504.0±22.4	34.8±5.9	0.0±0.1	0.0±0.0	0.4±0.7	-0.2±0.2
650.0-900.0	286.0±16.9	15.7±4.0	0.0±0.0	0.0±0.0	0.0±0.0	-0.2±0.2
900.0-2500.0	62.0±7.9	3.2±1.8	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0

Table A.6: Z$\rightarrow \mu\mu$ transverse momentum background fractions (Powheg)

$p_T^{\mu\mu}$-bin	EWK Background (Z$\rightarrow \tau\tau$, Diboson) [%]	Photon induced [%]	$W \rightarrow l\nu$ [%]	Top-Quark Pairs, single top [%]	Multijet Background [%]
0.0-2.0	0.065±0.003	0.041±0.002	0.000±0.000	0.006±0.001	0.002±0.001
2.0-4.0	0.036±0.001	0.026±0.001	0.000±0.000	0.007±0.001	-0.000±0.000
4.0-6.0	0.060±0.002	0.025±0.001	0.000±0.000	0.011±0.001	0.000±0.001
6.0-8.0	0.063±0.002	0.020±0.001	0.000±0.000	0.017±0.001	0.000±0.001
8.0-10.0	0.078±0.002	0.024±0.001	0.001±0.000	0.025±0.001	0.002±0.001
10.0-12.0	0.101±0.003	0.019±0.001	0.000±0.000	0.032±0.002	0.004±0.001
12.0-14.0	0.105±0.003	0.017±0.001	0.001±0.000	0.045±0.002	0.003±0.002
14.0-16.0	0.118±0.004	0.019±0.001	0.002±0.000	0.059±0.003	0.003±0.002
16.0-18.0	0.130±0.004	0.021±0.002	0.001±0.000	0.075±0.003	0.002±0.002
18.0-20.0	0.143±0.005	0.021±0.002	0.000±0.000	0.100±0.004	0.004±0.002
20.0-22.5	0.168±0.005	0.022±0.002	0.000±0.000	0.115±0.004	0.003±0.003
22.5-25.0	0.197±0.006	0.026±0.002	0.000±0.000	0.153±0.005	0.006±0.003
25.0-27.5	0.226±0.007	0.022±0.002	0.000±0.000	0.181±0.006	0.020±0.005
27.5-30.0	0.276±0.008	0.023±0.002	0.000±0.000	0.233±0.007	0.002±0.004
30.0-33.0	0.346±0.009	0.023±0.002	0.046±0.003	0.278±0.008	-0.018±0.022
33.0-36.0	0.352±0.010	0.024±0.003	0.006±0.001	0.343±0.010	0.004±0.006
36.0-39.0	0.413±0.012	0.012±0.002	0.003±0.001	0.427±0.012	0.016±0.008
39.0-42.0	0.480±0.013	0.021±0.003	0.000±0.000	0.506±0.014	0.029±0.010
42.0-45.0	0.546±0.016	0.014±0.003	0.000±0.000	0.596±0.016	0.013±0.010
45.0-48.0	0.570±0.017	0.030±0.004	0.002±0.001	0.739±0.020	0.023±0.012
48.0-51.0	0.602±0.019	0.015±0.003	0.000±0.000	0.870±0.023	0.023±0.014
51.0-54.0	0.653±0.022	0.014±0.003	0.000±0.000	1.020±0.027	0.036±0.017
54.0-57.0	0.739±0.025	0.014±0.003	0.000±0.000	1.208±0.032	0.009±0.014
57.0-61.0	0.752±0.023	0.015±0.003	0.073±0.007	1.380±0.031	0.032±0.017
61.0-65.0	0.822±0.026	0.026±0.005	0.003±0.002	1.586±0.036	0.012±0.016
65.0-70.0	0.851±0.026	0.019±0.004	0.005±0.002	1.871±0.038	0.028±0.018
70.0-75.0	0.943±0.030	0.033±0.006	0.000±0.000	2.126±0.045	0.059±0.024
75.0-80.0	0.971±0.032	0.013±0.004	0.000±0.000	2.390±0.051	0.010±0.020
80.0-85.0	0.999±0.036	0.009±0.003	0.000±0.000	2.585±0.058	0.028±0.027
85.0-95.0	1.066±0.029	0.016±0.004	0.000±0.000	2.753±0.047	0.044±0.019
95.0-105.0	1.139±0.034	0.016±0.004	0.000±0.000	2.843±0.055	0.056±0.024
105.0-125.0	1.294±0.032	0.018±0.004	0.000±0.000	2.639±0.045	-0.001±0.069
125.0-150.0	1.471±0.040	0.013±0.004	0.000±0.000	1.946±0.046	0.080±0.025
150.0-175.0	1.689±0.057	0.012±0.005	0.000±0.000	1.215±0.048	0.042±0.026
175.0-200.0	1.911±0.079	0.005±0.004	0.000±0.000	0.754±0.049	0.041±0.035
200.0-250.0	2.144±0.084	0.018±0.008	0.042±0.012	0.506±0.041	0.029±0.032
250.0-300.0	2.840±0.151	0.036±0.017	0.000±0.000	0.335±0.051	0.038±0.048
300.0-350.0	2.891±0.222	0.011±0.014	0.000±0.000	0.315±0.072	0.190±0.124
350.0-400.0	3.392±0.348	0.010±0.019	0.000±0.000	0.171±0.077	-0.031±0.012
400.0-470.0	3.653±0.445	0.011±0.024	0.000±0.000	0.144±0.087	0.203±0.226
470.0-550.0	4.168±0.680	0.028±0.055	0.000±0.000	0.121±0.114	-0.016±0.016
550.0-650.0	6.895±1.209	0.004±0.030	0.000±0.000	0.085±0.130	-0.038±0.038
650.0-900.0	5.483±1.422	0.000±0.000	0.000±0.000	0.000±0.000	-0.061±0.061
900.0-2500.0	5.152±2.956	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000

A.4.2 Influence of the $W + Z \rightarrow qq + ll$ sample

Table A.7: Difference in the normalized differential cross-section in bins of the Z boson transverse momentum if including the $W+Z \rightarrow qq+ll$ background process or not. Shown are bin entries of the measurement together with its uncertainty. No bin shows a difference within the uncertainties, hence no effect is observed.

Bin	$d\sigma/\sigma$ with $W + Z \rightarrow qq + ll$ [GeV]	$d\sigma/\sigma$ without $W + Z \rightarrow qq + ll$ [GeV]
0.0-2.0	4837.52 ± 30.18	4834.84 ± 30.17
2.0-4.0	10239.19 ± 29.92	10233.58 ± 29.91
4.0-6.0	10651.13 ± 9.09	10645.40 ± 9.08
6.0-8.0	9499.37 ± 11.83	9494.47 ± 11.83
8.0-10.0	8114.16 ± 13.78	8110.23 ± 13.77
10.0-12.0	6859.96 ± 6.86	6856.79 ± 6.86
12.0-14.0	5841.68 ± 7.63	5839.25 ± 7.62
14.0-16.0	4961.00 ± 5.87	4959.31 ± 5.87
16.0-18.0	4261.27 ± 8.08	4260.03 ± 8.08
18.0-20.0	3660.17 ± 4.94	3659.29 ± 4.94
20.0-22.5	3903.01 ± 3.82	3902.39 ± 3.82
22.5-25.0	3286.24 ± 3.93	3286.04 ± 3.93
25.0-27.5	2804.63 ± 3.33	2804.75 ± 3.33
27.5-30.0	2387.81 ± 1.93	2388.17 ± 1.93
30.0-33.0	2428.67 ± 2.96	2429.42 ± 2.96
33.0-36.0	2037.65 ± 2.38	2038.62 ± 2.38
36.0-39.0	1726.71 ± 2.59	1727.78 ± 2.59
39.0-42.0	1461.34 ± 2.65	1462.51 ± 2.66
42.0-45.0	1247.47 ± 1.87	1248.69 ± 1.88
45.0-48.0	1053.01 ± 2.36	1054.21 ± 2.36
48.0-51.0	895.59 ± 1.92	896.74 ± 1.92
51.0-54.0	759.70 ± 1.93	760.82 ± 1.93
54.0-57.0	647.45 ± 2.10	648.49 ± 2.10
57.0-61.0	751.45 ± 1.45	752.81 ± 1.45
61.0-65.0	634.09 ± 1.80	635.33 ± 1.80
65.0-70.0	674.30 ± 1.16	675.75 ± 1.17
70.0-75.0	563.78 ± 1.20	565.29 ± 1.20
75.0-80.0	476.27 ± 1.56	477.56 ± 1.56
80.0-85.0	404.33 ± 1.16	405.52 ± 1.17
85.0-95.0	650.39 ± 1.67	652.50 ± 1.68
95.0-105.0	484.51 ± 1.89	486.29 ± 1.90
105.0-125.0	653.64 ± 1.67	656.46 ± 1.67
125.0-150.0	460.83 ± 1.67	463.29 ± 1.68
150.0-175.0	260.60 ± 1.11	262.28 ± 1.12
175.0-200.0	150.79 ± 1.03	151.90 ± 1.04
200.0-250.0	151.03 ± 1.15	152.32 ± 1.16
250.0-300.0	59.40 ± 0.82	60.05 ± 0.83
300.0-350.0	29.22 ± 0.55	29.58 ± 0.55
350.0-400.0	14.00 ± 0.45	14.25 ± 0.45
400.0-470.0	8.46 ± 0.21	8.61 ± 0.21
470.0-550.0	4.51 ± 0.33	4.60 ± 0.34
550.0-650.0	1.99 ± 0.15	2.04 ± 0.16
650.0-900.0	1.21 ± 0.06	1.25 ± 0.06

A.4.3 Fiducial cross section tables

Table A.8: Measured normalized $p_T^{\mu\mu}$ cross-section on bare-level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 2	0.0463828	0.282	0.196	0.022	0.170	0.093	0.001	0.008	0.000	0.000	0.000	0.001	0.059	0.048	0.005
2 - 4	0.0994908	0.120	0.096	0.008	0.061	0.068	0.007	0.003	0.000	0.000	0.000	0.000	0.006	0.045	0.002
4 - 6	0.105382	0.102	0.088	0.008	0.060	0.067	0.000	0.003	0.000	0.000	0.000	0.000	0.018	0.000	0.002
6 - 8	0.095502	0.121	0.096	0.008	0.058	0.037	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.010	0.002
8 - 10	0.0824012	0.129	0.109	0.009	0.045	0.022	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.014	0.003
10 - 12	0.0700981	0.154	0.126	0.010	0.029	0.078	0.007	0.002	0.000	0.000	0.000	0.000	0.001	0.015	0.004
12 - 14	0.0598852	0.177	0.139	0.010	0.011	0.097	0.005	0.002	0.000	0.000	0.000	0.000	0.009	0.024	0.004
14 - 16	0.0508426	0.196	0.159	0.010	0.004	0.094	0.002	0.002	0.000	0.000	0.000	0.000	0.010	0.006	0.006
16 - 18	0.0435897	0.220	0.182	0.009	0.012	0.073	0.006	0.002	0.000	0.000	0.000	0.000	0.001	0.019	0.007
18 - 20	0.0371805	0.223	0.205	0.009	0.029	0.069	0.001	0.003	0.000	0.000	0.000	0.000	0.005	0.014	0.009
20 - 22.5	0.0394755	0.189	0.183	0.007	0.030	0.051	0.000	0.002	0.000	0.000	0.000	0.000	0.015	0.005	0.008
22.5 - 25	0.0330098	0.217	0.203	0.007	0.046	0.041	0.000	0.003	0.000	0.000	0.000	0.000	0.006	0.018	0.011
25 - 27.5	0.0280036	0.239	0.227	0.007	0.057	0.042	0.001	0.003	0.000	0.000	0.000	0.000	0.012	0.005	0.013
27.5 - 30	0.0238365	0.271	0.250	0.007	0.065	0.033	0.000	0.004	0.000	0.000	0.000	0.000	0.018	0.010	0.019
30 - 33	0.0241433	0.232	0.227	0.006	0.071	0.016	0.007	0.004	0.000	0.000	0.000	0.000	0.030	0.001	0.019
33 - 36	0.0202266	0.265	0.289	0.007	0.091	0.006	0.007	0.005	0.000	0.000	0.000	0.000	0.010	0.016	0.024
36 - 39	0.0170247	0.288	0.324	0.008	0.108	0.004	0.010	0.005	0.000	0.000	0.000	0.000	0.046	0.015	0.028
39 - 42	0.0144144	0.316	0.352	0.009	0.136	0.017	0.004	0.006	0.001	0.000	0.000	0.000	0.042	0.000	0.032
42 - 45	0.0122751	0.339	0.392	0.011	0.144	0.037	0.004	0.006	0.001	0.000	0.000	0.001	0.050	0.008	0.038
45 - 48	0.0103566	0.377	0.434	0.013	0.170	0.023	0.001	0.006	0.001	0.000	0.000	0.002	0.044	0.028	0.047
48 - 51	0.00878699	0.409	0.511	0.016	0.195	0.021	0.009	0.006	0.001	0.001	0.000	0.000	0.038	0.032	0.054
51 - 54	0.00748878	0.451	0.602	0.019	0.212	0.023	0.007	0.007	0.001	0.001	0.000	0.000	0.046	0.053	0.067
54 - 57	0.00639071	0.495	0.702	0.023	0.228	0.013	0.024	0.007	0.000	0.001	0.000	0.000	0.023	0.073	0.084
57 - 61	0.00743347	0.404	0.623	0.020	0.166	0.013	0.011	0.005	0.000	0.001	0.000	0.000	0.036	0.058	0.075
61 - 65	0.00627207	0.428	0.699	0.025	0.170	0.015	0.009	0.006	0.000	0.001	0.000	0.001	0.067	0.082	0.103
65 - 70	0.00668419	0.399	0.652	0.026	0.145	0.000	0.001	0.005	0.000	0.000	0.000	0.001	0.071	0.098	0.098
70 - 75	0.00559836	0.446	0.644	0.041	0.164	0.001	0.001	0.007	0.000	0.001	0.000	0.000	0.090	0.091	0.124
75 - 80	0.00472527	0.481	0.790	0.064	0.169	0.029	0.013	0.009	0.000	0.001	0.000	0.001	0.075	0.149	0.149
80 - 85	0.00401212	0.556	0.866	0.089	0.163	0.031	0.005	0.013	0.000	0.001	0.000	0.005	0.085	0.163	0.198
85 - 95	0.00646743	0.322	0.514	0.061	0.099	0.008	0.008	0.010	0.000	0.000	0.000	0.003	0.037	0.114	0.123
95 - 105	0.00480518	0.421	0.664	0.096	0.116	0.012	0.002	0.015	0.000	0.000	0.000	0.005	0.061	0.129	0.157
105 - 125	0.00650209	0.288	0.413	0.076	0.079	0.001	0.002	0.012	0.000	0.000	0.000	0.003	0.039	0.106	0.099
125 - 150	0.00456411	0.384	0.500	0.111	0.097	0.009	0.003	0.020	0.000	0.000	0.000	0.003	0.057	0.114	0.137
150 - 175	0.00258074	0.603	0.755	0.198	0.128	0.057	0.005	0.037	0.000	0.000	0.000	0.007	0.036	0.194	0.199
175 - 200	0.00149049	0.947	0.976	0.340	0.176	0.025	0.014	0.064	0.002	0.000	0.000	0.012	0.009	0.237	0.276
200 - 250	0.00149294	0.821	0.864	0.327	0.122	0.008	0.000	0.056	0.003	0.000	0.000	0.008	0.198	0.161	0.169
250 - 300	0.000593267	1.714	1.879	0.778	0.167	0.056	0.088	0.114	0.009	0.000	0.000	0.012	0.129	0.233	0.236
300 - 350	0.00028923	2.909	3.302	1.370	0.247	0.019	0.029	0.188	0.019	0.001	0.000	0.006	0.136	0.383	0.366
350 - 400	0.000138402	4.823	4.255	2.503	0.301	0.135	0.015	0.313	0.048	0.001	0.001	0.010	0.126	0.319	0.602
400 - 470	8.32572e-05	6.705	6.185	3.666	0.325	0.150	0.051	0.322	0.081	0.001	0.000	0.040	0.316	0.506	0.441
470 - 550	4.45417e-05	11.07	10.08	5.992	0.282	0.099	0.141	0.416	0.146	0.003	0.000	0.027	1.145	0.028	0.450
550 - 650	2.06872e-05	18.31	17.87	8.722	0.451	0.156	0.405	0.649	0.336	0.006	0.001	0.024	0.774	0.783	0.690
650 - 900	1.20444e-05	27.39	26.52	9.893	0.528	0.726	1.752	0.852	0.610	0.008	0.002	0.066	1.270	0.929	0.905

Table A.9: Measured normalized ϕ^*_η cross-section on bare-level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 0.004	0.0349306	0.203	0.106	0.009	0.016	0.002	0.001	0.004	0.000	0.000	0.000	0.000	0.001	0.017	0.004
0.004 - 0.008	0.034552	0.192	0.104	0.009	0.013	0.004	0.001	0.004	0.000	0.000	0.000	0.000	0.037	0.017	0.004
0.008 - 0.012	0.0339599	0.169	0.096	0.010	0.014	0.001	0.000	0.004	0.000	0.000	0.000	0.000	0.005	0.017	0.004
0.012 - 0.016	0.0331271	0.189	0.107	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.010	0.018	0.004
0.016 - 0.02	0.0321708	0.190	0.105	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.013	0.018	0.003
0.02 - 0.024	0.0310276	0.191	0.110	0.010	0.014	0.001	0.000	0.004	0.000	0.000	0.000	0.001	0.018	0.019	0.003
0.024 - 0.029	0.0369623	0.164	0.097	0.008	0.011	0.003	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.015	0.003
0.029 - 0.034	0.0350223	0.165	0.099	0.009	0.009	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.005	0.015	0.003
0.034 - 0.039	0.0328932	0.179	0.110	0.009	0.010	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.009	0.015	0.003
0.039 - 0.045	0.0369478	0.150	0.097	0.008	0.009	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.006	0.012	0.003
0.045 - 0.051	0.0342078	0.161	0.099	0.009	0.009	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.011	0.003
0.051 - 0.057	0.0316385	0.177	0.112	0.009	0.008	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.018	0.009	0.004
0.057 - 0.064	0.0339537	0.165	0.104	0.008	0.005	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.015	0.006	0.004
0.064 - 0.072	0.0353948	0.159	0.099	0.007	0.003	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.009	0.003	0.004
0.072 - 0.081	0.0358216	0.164	0.094	0.007	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.004
0.081 - 0.091	0.035598	0.166	0.095	0.007	0.001	0.002	0.000	0.001	0.000	0.000	0.000	0.000	0.006	0.000	0.004
0.091 - 0.102	0.0346786	0.164	0.095	0.007	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.011	0.000	0.005
0.102 - 0.114	0.0332424	0.186	0.097	0.007	0.003	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.009	0.000	0.006
0.114 - 0.128	0.0339522	0.193	0.096	0.006	0.005	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.006
0.128 - 0.145	0.0351818	0.183	0.093	0.006	0.008	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.006
0.145 - 0.165	0.0347355	0.202	0.093	0.006	0.010	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.001	0.007
0.165 - 0.189	0.0344194	0.192	0.093	0.006	0.014	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.001	0.008
0.189 - 0.219	0.0345692	0.185	0.087	0.006	0.015	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.002	0.001	0.011
0.219 - 0.258	0.0345931	0.181	0.086	0.008	0.018	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.001	0.015
0.258 - 0.312	0.035026	0.172	0.084	0.009	0.018	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.003	0.005	0.017
0.312 - 0.391	0.0347651	0.160	0.082	0.010	0.017	0.001	0.001	0.003	0.000	0.000	0.000	0.000	0.008	0.010	0.021
0.391 - 0.524	0.0348434	0.150	0.077	0.012	0.014	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.006	0.015	0.024
0.524 - 0.695	0.0242315	0.204	0.102	0.020	0.017	0.001	0.004	0.005	0.000	0.000	0.000	0.000	0.002	0.028	0.037
0.695 - 0.918	0.016705	0.245	0.135	0.031	0.017	0.003	0.006	0.007	0.000	0.000	0.000	0.002	0.007	0.048	0.060
0.918 - 1.153	0.00955916	0.360	0.217	0.059	0.020	0.006	0.011	0.013	0.000	0.001	0.000	0.002	0.049	0.090	0.099
1.153 - 1.496	0.00746651	0.420	0.236	0.081	0.006	0.002	0.010	0.016	0.000	0.000	0.000	0.003	0.062	0.110	0.111
1.496 - 1.947	0.00491834	0.494	0.327	0.132	0.000	0.000	0.010	0.023	0.000	0.001	0.000	0.004	0.087	0.146	0.149
1.947 - 2.522	0.00311493	0.685	0.484	0.221	0.007	0.001	0.011	0.035	0.001	0.001	0.001	0.007	0.073	0.192	0.203
2.522 - 3.277	0.00208096	0.937	0.678	0.340	0.006	0.005	0.016	0.052	0.003	0.001	0.002	0.006	0.083	0.254	0.270
3.277 - 5	0.00201913	0.872	0.630	0.353	0.011	0.005	0.015	0.051	0.005	0.001	0.002	0.006	0.085	0.219	0.246
5 - 10	0.00168999	0.943	0.604	0.421	0.007	0.013	0.006	0.057	0.007	0.001	0.002	0.004	0.030	0.230	0.211

Table A.10: Measured normalized $\mathrm{p}_{\mathrm{T}}^{\mu\mu}$ cross-section on dressed-level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 2	0.0472729	0.273	0.204	0.022	0.169	0.090	0.001	0.008	0.000	0.000	0.000	0.001	0.058	0.047	0.008
2 - 4	0.10078	0.117	0.097	0.008	0.061	0.068	0.007	0.003	0.000	0.000	0.000	0.000	0.012	0.044	0.002
4 - 6	0.105826	0.101	0.090	0.008	0.060	0.067	0.000	0.003	0.000	0.000	0.000	0.000	0.017	0.000	0.003
6 - 8	0.0952221	0.120	0.101	0.008	0.059	0.037	0.001	0.003	0.000	0.000	0.000	0.000	0.002	0.010	0.003
8 - 10	0.0817378	0.129	0.113	0.009	0.046	0.023	0.001	0.003	0.000	0.000	0.000	0.000	0.001	0.014	0.003
10 - 12	0.0693586	0.154	0.131	0.010	0.029	0.081	0.007	0.002	0.000	0.000	0.000	0.000	0.003	0.016	0.004
12 - 14	0.0591619	0.177	0.146	0.010	0.010	0.100	0.005	0.002	0.000	0.000	0.000	0.000	0.006	0.024	0.004
14 - 16	0.0502303	0.196	0.167	0.010	0.004	0.097	0.002	0.002	0.000	0.000	0.000	0.000	0.010	0.007	0.004
16 - 18	0.043131	0.220	0.188	0.009	0.013	0.075	0.006	0.002	0.000	0.000	0.000	0.000	0.005	0.019	0.006
18 - 20	0.0369246	0.221	0.212	0.009	0.029	0.070	0.005	0.003	0.000	0.000	0.000	0.000	0.002	0.014	0.008
20 - 22.5	0.0392712	0.188	0.187	0.007	0.031	0.051	0.002	0.002	0.000	0.000	0.000	0.000	0.012	0.005	0.007
22.5 - 25	0.0329571	0.215	0.208	0.007	0.046	0.041	0.004	0.003	0.000	0.000	0.000	0.000	0.003	0.018	0.010
25 - 27.5	0.0280397	0.236	0.233	0.007	0.057	0.042	0.006	0.003	0.000	0.000	0.000	0.000	0.015	0.005	0.012
27.5 - 30	0.0238578	0.268	0.257	0.007	0.065	0.033	0.006	0.004	0.000	0.000	0.000	0.000	0.021	0.010	0.017
30 - 33	0.0242491	0.229	0.233	0.006	0.070	0.017	0.013	0.004	0.000	0.000	0.000	0.000	0.034	0.001	0.017
33 - 36	0.0203131	0.264	0.296	0.007	0.091	0.007	0.014	0.005	0.000	0.000	0.000	0.000	0.003	0.016	0.019
36 - 39	0.0171668	0.285	0.332	0.008	0.107	0.004	0.010	0.005	0.000	0.000	0.000	0.000	0.038	0.015	0.022
39 - 42	0.0145277	0.314	0.360	0.009	0.136	0.017	0.004	0.006	0.001	0.000	0.000	0.000	0.044	0.000	0.029
42 - 45	0.0123856	0.336	0.328	0.010	0.143	0.036	0.004	0.006	0.001	0.000	0.000	0.001	0.056	0.008	0.033
45 - 48	0.0104723	0.373	0.363	0.013	0.169	0.022	0.001	0.006	0.001	0.000	0.000	0.002	0.056	0.028	0.039
48 - 51	0.00888542	0.405	0.421	0.016	0.194	0.021	0.008	0.006	0.001	0.001	0.000	0.000	0.038	0.031	0.051
51 - 54	0.00755696	0.449	0.508	0.019	0.212	0.022	0.007	0.007	0.001	0.001	0.000	0.000	0.050	0.053	0.067
54 - 57	0.00644683	0.493	0.589	0.022	0.228	0.013	0.024	0.007	0.000	0.001	0.000	0.000	0.004	0.073	0.085
57 - 61	0.00748728	0.403	0.517	0.019	0.167	0.013	0.010	0.005	0.000	0.001	0.000	0.000	0.034	0.058	0.081
61 - 65	0.00631266	0.427	0.589	0.025	0.171	0.015	0.008	0.006	0.000	0.001	0.000	0.001	0.060	0.082	0.099
65 - 70	0.00672562	0.399	0.548	0.026	0.146	0.000	0.001	0.005	0.000	0.000	0.000	0.001	0.088	0.097	0.100
70 - 75	0.00562552	0.447	0.542	0.041	0.164	0.001	0.001	0.007	0.000	0.001	0.000	0.000	0.087	0.091	0.135
75 - 80	0.0047522	0.482	0.662	0.063	0.170	0.028	0.013	0.009	0.000	0.001	0.000	0.001	0.072	0.148	0.156
80 - 85	0.00403371	0.557	0.735	0.088	0.163	0.029	0.004	0.013	0.000	0.001	0.000	0.005	0.075	0.163	0.194
85 - 95	0.00649798	0.323	0.437	0.061	0.099	0.007	0.008	0.010	0.000	0.000	0.000	0.003	0.036	0.114	0.120
95 - 105	0.00483469	0.423	0.564	0.095	0.117	0.011	0.002	0.015	0.000	0.000	0.000	0.005	0.050	0.129	0.149
105 - 125	0.00653312	0.290	0.351	0.076	0.079	0.000	0.002	0.012	0.000	0.000	0.000	0.003	0.032	0.106	0.098
125 - 150	0.00459762	0.385	0.412	0.111	0.097	0.008	0.004	0.020	0.000	0.000	0.000	0.003	0.041	0.113	0.126
150 - 175	0.00260144	0.606	0.637	0.197	0.128	0.057	0.005	0.038	0.000	0.000	0.000	0.007	0.032	0.193	0.181
175 - 200	0.00150677	0.950	0.852	0.337	0.175	0.025	0.013	0.064	0.002	0.000	0.000	0.012	0.016	0.234	0.243
200 - 250	0.00151307	0.823	0.739	0.323	0.122	0.008	0.000	0.056	0.003	0.000	0.000	0.008	0.154	0.159	0.152
250 - 300	0.000598786	1.726	1.629	0.774	0.168	0.059	0.090	0.115	0.009	0.000	0.000	0.013	0.090	0.231	0.234
300 - 350	0.000294408	2.907	2.859	1.354	0.246	0.022	0.026	0.188	0.019	0.001	0.000	0.006	0.013	0.377	0.324
350 - 400	0.000143273	4.748	3.810	2.438	0.295	0.136	0.015	0.308	0.047	0.001	0.001	0.009	0.299	0.308	0.516
400 - 470	8.52153e-05	6.665	5.329	3.620	0.321	0.157	0.053	0.320	0.079	0.001	0.000	0.039	0.194	0.494	0.519
470 - 550	4.55063e-05	11.03	8.844	5.942	0.281	0.087	0.136	0.414	0.145	0.003	0.000	0.028	0.740	0.027	0.472
550 - 650	2.06171e-05	18.62	16.67	8.892	0.456	0.180	0.398	0.662	0.340	0.006	0.001	0.027	1.876	0.786	0.596
650 - 900	1.22673e-05	27.33	24.57	9.943	0.514	0.737	1.723	0.845	0.605	0.008	0.002	0.063	0.495	0.913	0.684

Table A.11: Measured normalized ϕ^*_η cross-section on dressed-level definition including a full breakdown of relative uncertainties. The first three uncertainties (Stat, MCStat as well as electron efficiencies) are bin-to-bin uncorrelated

Bin (GeV)	$d\sigma/\sigma$	Stat [%]	MCStat [%]	μ-Eff. [%]	μ-Scale [%]	μ-Res. [%]	Sagitta [%]	Eff ID [%]	Iso Iso [%]	Trig-er [%]	TTVA ID [%]	Z-Pos [%]	P.U. [%]	Model [%]	Bkg [%]
0 - 0.004	0.0349233	0.206	0.106	0.009	0.016	0.002	0.001	0.004	0.000	0.000	0.000	0.000	0.000	0.017	0.004
0.004 - 0.008	0.0345491	0.195	0.101	0.009	0.013	0.004	0.001	0.004	0.000	0.000	0.000	0.000	0.033	0.017	0.003
0.008 - 0.012	0.0339613	0.171	0.101	0.010	0.014	0.001	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.017	0.003
0.012 - 0.016	0.0331271	0.191	0.113	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.007	0.018	0.003
0.016 - 0.02	0.0321593	0.193	0.109	0.010	0.013	0.002	0.000	0.004	0.000	0.000	0.000	0.000	0.013	0.018	0.003
0.02 - 0.024	0.0310257	0.193	0.116	0.010	0.014	0.001	0.000	0.004	0.000	0.000	0.000	0.001	0.018	0.019	0.003
0.024 - 0.029	0.0369587	0.166	0.100	0.008	0.011	0.003	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.015	0.002
0.029 - 0.034	0.03503	0.166	0.104	0.009	0.009	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.006	0.015	0.003
0.034 - 0.039	0.0328939	0.181	0.112	0.009	0.010	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.009	0.015	0.003
0.039 - 0.045	0.0369368	0.152	0.100	0.008	0.009	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.008	0.012	0.003
0.045 - 0.051	0.0341967	0.162	0.105	0.009	0.009	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.011	0.004
0.051 - 0.057	0.0316565	0.178	0.121	0.009	0.008	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.019	0.009	0.004
0.057 - 0.064	0.0339555	0.166	0.113	0.008	0.005	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.016	0.006	0.004
0.064 - 0.072	0.0354126	0.160	0.106	0.007	0.003	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.010	0.003	0.004
0.072 - 0.081	0.0358407	0.165	0.099	0.007	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.002	0.004
0.081 - 0.091	0.0356026	0.167	0.100	0.007	0.001	0.002	0.000	0.001	0.000	0.000	0.000	0.000	0.005	0.000	0.004
0.091 - 0.102	0.0346878	0.165	0.103	0.007	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.013	0.000	0.005
0.102 - 0.114	0.0332578	0.187	0.107	0.007	0.003	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.010	0.000	0.005
0.114 - 0.128	0.0339584	0.194	0.106	0.006	0.005	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.005
0.128 - 0.145	0.0352118	0.183	0.104	0.006	0.008	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.004	0.000	0.006
0.145 - 0.165	0.0347468	0.202	0.105	0.005	0.010	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.002	0.001	0.007
0.165 - 0.189	0.0344395	0.192	0.104	0.006	0.014	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.002	0.001	0.009
0.189 - 0.219	0.034604	0.185	0.095	0.006	0.015	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.003	0.001	0.010
0.219 - 0.258	0.0346067	0.182	0.093	0.008	0.018	0.002	0.000	0.003	0.000	0.000	0.000	0.000	0.001	0.001	0.015
0.258 - 0.312	0.0350348	0.172	0.090	0.009	0.018	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.004	0.005	0.019
0.312 - 0.391	0.0347762	0.160	0.088	0.010	0.017	0.001	0.001	0.003	0.000	0.000	0.000	0.000	0.010	0.010	0.024
0.391 - 0.524	0.0348396	0.150	0.084	0.012	0.014	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.008	0.015	0.030
0.524 - 0.695	0.0242044	0.204	0.108	0.020	0.017	0.001	0.004	0.005	0.000	0.000	0.000	0.000	0.001	0.028	0.046
0.695 - 0.918	0.0166723	0.245	0.143	0.031	0.017	0.003	0.006	0.007	0.000	0.000	0.000	0.002	0.005	0.048	0.069
0.918 - 1.153	0.00952137	0.361	0.231	0.060	0.020	0.006	0.011	0.013	0.000	0.001	0.000	0.002	0.042	0.090	0.122
1.153 - 1.496	0.00743887	0.422	0.246	0.081	0.006	0.002	0.010	0.016	0.000	0.000	0.000	0.003	0.059	0.111	0.136
1.496 - 1.947	0.00489564	0.496	0.334	0.132	0.000	0.000	0.010	0.023	0.000	0.001	0.000	0.004	0.079	0.147	0.182
1.947 - 2.522	0.00310407	0.686	0.483	0.221	0.007	0.001	0.011	0.035	0.001	0.001	0.001	0.007	0.050	0.193	0.229
2.522 - 3.277	0.00207306	0.939	0.675	0.340	0.006	0.006	0.016	0.052	0.003	0.001	0.002	0.006	0.059	0.255	0.268
3.277 - 5	0.00201066	0.874	0.633	0.353	0.011	0.005	0.014	0.051	0.005	0.001	0.002	0.006	0.060	0.220	0.229
5 - 10	0.00168618	0.944	0.625	0.421	0.007	0.014	0.006	0.057	0.007	0.001	0.002	0.004	0.013	0.231	0.240

A.4.4 Combined fiducial cross section tables

Table A.12: Combination: measured normalized $p_{T}^{\ell\ell}$ cross-section on born level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 2	0.0484018	9.20699e-05	7.07504e-05
2 - 4	0.102286	5.85091e-05	8.13524e-05
4 - 6	0.106486	6.11488e-05	7.36234e-05
6 - 8	0.09468	6.29709e-05	7.48824e-05
8 - 10	0.08115	6.31765e-05	6.93918e-05
10 - 12	0.0686642	6.53704e-05	6.84137e-05
12 - 14	0.0582644	6.97449e-05	6.40802e-05
14 - 16	0.0496081	6.33438e-05	6.37153e-05
16 - 18	0.0424937	5.59555e-05	6.24328e-05
18 - 20	0.0366795	5.00285e-05	5.84013e-05
20 - 22.5	0.039002	4.54181e-05	5.35437e-05
22.5 - 25	0.0330001	4.22252e-05	5.04145e-05
25 - 27.5	0.0279983	4.11078e-05	4.77862e-05
27.5 - 30	0.0238948	3.9604e-05	4.51904e-05
30 - 33	0.0243127	3.84881e-05	4.05577e-05
33 - 36	0.020348	3.68572e-05	4.05719e-05
36 - 39	0.017258	3.42924e-05	3.80441e-05
39 - 42	0.0146294	3.15593e-05	3.42204e-05
42 - 45	0.0125113	2.89591e-05	2.86418e-05
45 - 48	0.0105687	2.6248e-05	2.69046e-05
48 - 51	0.00890156	2.48723e-05	2.57712e-05
51 - 54	0.00763575	2.26259e-05	2.4761e-05
54 - 57	0.00657764	2.19066e-05	2.40572e-05
57 - 61	0.00749142	1.98297e-05	2.35118e-05
61 - 65	0.00637974	1.89423e-05	2.19824e-05
65 - 70	0.00675707	1.81e-05	2.1917e-05
70 - 75	0.00566338	1.75167e-05	1.92242e-05
75 - 80	0.00479026	1.71679e-05	1.8926e-05
80 - 85	0.00408189	1.64693e-05	1.82576e-05
85 - 95	0.00654072	1.54701e-05	1.72583e-05
95 - 105	0.00483986	1.46129e-05	1.62734e-05
105 - 125	0.00658685	1.33498e-05	1.44866e-05
125 - 150	0.00465846	1.26382e-05	1.27902e-05
150 - 175	0.00262564	1.09681e-05	1.1255e-05
175 - 200	0.00153693	9.98559e-06	9.6362e-06
200 - 250	0.0015317	8.27e-06	8.44609e-06
250 - 300	0.000614663	6.77416e-06	7.13132e-06
300 - 350	0.000294738	5.22589e-06	5.69645e-06
350 - 400	0.000137497	4.03982e-06	4.1404e-06
400 - 470	8.81033e-05	2.96661e-06	3.61831e-06
470 - 550	4.33334e-05	2.26039e-06	3.18794e-06
550 - 650	1.9995e-05	1.62144e-06	2.45382e-06
650 - 900	1.29451e-05	1.26637e-06	2.06126e-06

Table A.13: Combination: measured normalized $p_T^{\ell\ell}$ cross-section on bare-level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 2	0.0475504	9.28213e-05	6.90362e-05
2 - 4	0.101353	6.08271e-05	8.10463e-05
4 - 6	0.106234	6.7458e-05	7.51727e-05
6 - 8	0.0952441	6.46773e-05	7.39169e-05
8 - 10	0.0819623	6.70158e-05	6.87693e-05
10 - 12	0.0695442	6.84462e-05	6.80029e-05
12 - 14	0.0591255	7.26186e-05	6.31039e-05
14 - 16	0.0503067	6.57963e-05	6.22881e-05
16 - 18	0.0429808	5.71014e-05	6.11359e-05
18 - 20	0.0368932	5.15593e-05	5.67534e-05
20 - 22.5	0.0391201	4.95935e-05	5.25987e-05
22.5 - 25	0.0329572	4.62382e-05	4.94949e-05
25 - 27.5	0.0278573	4.34168e-05	4.6355e-05
27.5 - 30	0.0237336	4.17985e-05	4.36414e-05
30 - 33	0.0240899	4.06327e-05	3.85737e-05
33 - 36	0.0201106	3.84799e-05	3.85503e-05
36 - 39	0.0169941	3.61286e-05	3.61579e-05
39 - 42	0.0143626	3.19283e-05	3.32069e-05
42 - 45	0.0122447	2.98261e-05	3.08178e-05
45 - 48	0.0103726	2.77177e-05	2.88819e-05
48 - 51	0.00872035	2.60888e-05	2.78669e-05
51 - 54	0.00752682	2.34673e-05	2.70728e-05
54 - 57	0.00649828	2.15414e-05	2.61866e-05
57 - 61	0.00740055	1.96582e-05	2.58495e-05
61 - 65	0.00632542	1.91145e-05	2.45588e-05
65 - 70	0.00672535	1.81426e-05	2.43102e-05
70 - 75	0.00563865	1.81245e-05	2.15272e-05
75 - 80	0.00476925	1.76793e-05	2.12656e-05
80 - 85	0.0040593	1.75944e-05	2.01854e-05
85 - 95	0.0065543	1.65207e-05	1.91029e-05
95 - 105	0.0048434	1.58768e-05	1.81138e-05
105 - 125	0.00664637	1.46545e-05	1.59316e-05
125 - 150	0.00469294	1.38932e-05	1.3845e-05
150 - 175	0.00262862	1.27508e-05	1.201e-05
175 - 200	0.00153471	1.14759e-05	9.83009e-06
200 - 250	0.00154258	9.41543e-06	8.78588e-06
250 - 300	0.000613504	7.84896e-06	7.40915e-06
300 - 350	0.000288272	5.95131e-06	6.04188e-06
350 - 400	0.000130209	4.62975e-06	4.30729e-06
400 - 470	8.42037e-05	3.29779e-06	3.80343e-06
470 - 550	4.23032e-05	2.63254e-06	3.35865e-06
550 - 650	1.78701e-05	1.89293e-06	2.59014e-06
650 - 900	1.09718e-05	1.27009e-06	2.05682e-06

Table A.14: Combination: measured normalized $p_T^{\ell\ell}$ cross-section on dressed-level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 2	0.047355	9.23093e-05	7.05793e-05
2 - 4	0.100749	6.43893e-05	8.28542e-05
4 - 6	0.105882	6.69077e-05	7.56658e-05
6 - 8	0.0949437	6.62176e-05	7.50979e-05
8 - 10	0.0817792	6.51265e-05	6.9209e-05
10 - 12	0.0694513	6.68967e-05	6.83804e-05
12 - 14	0.0590048	6.94298e-05	6.26724e-05
14 - 16	0.0502642	6.04069e-05	6.15742e-05
16 - 18	0.0429929	5.25937e-05	6.02655e-05
18 - 20	0.0370126	4.73065e-05	5.64064e-05
20 - 22.5	0.0392406	4.44757e-05	5.19931e-05
22.5 - 25	0.0331163	3.93991e-05	4.85914e-05
25 - 27.5	0.0279877	3.77813e-05	4.58413e-05
27.5 - 30	0.0238782	3.65074e-05	4.31238e-05
30 - 33	0.0242559	3.56876e-05	3.85115e-05
33 - 36	0.0202563	3.44084e-05	3.82916e-05
36 - 39	0.0171595	3.26224e-05	3.61864e-05
39 - 42	0.0145228	3.00257e-05	3.33881e-05
42 - 45	0.0124067	2.94057e-05	2.83512e-05
45 - 48	0.0104858	2.77067e-05	2.66867e-05
48 - 51	0.00883093	2.62202e-05	2.54403e-05
51 - 54	0.00757164	2.44488e-05	2.52077e-05
54 - 57	0.00653256	2.31017e-05	2.43891e-05
57 - 61	0.00744965	2.19206e-05	2.41531e-05
61 - 65	0.00633894	2.08386e-05	2.27876e-05
65 - 70	0.00672342	1.97514e-05	2.27463e-05
70 - 75	0.00563524	1.88784e-05	1.9893e-05
75 - 80	0.00476786	1.81977e-05	1.94481e-05
80 - 85	0.00405943	1.76324e-05	1.85114e-05
85 - 95	0.00651426	1.57865e-05	1.74488e-05
95 - 105	0.00481935	1.51005e-05	1.64947e-05
105 - 125	0.00655944	1.37809e-05	1.46021e-05
125 - 150	0.00463389	1.31152e-05	1.25412e-05
150 - 175	0.00261032	1.17764e-05	1.10288e-05
175 - 200	0.00152423	1.05991e-05	9.31331e-06
200 - 250	0.00152505	8.73811e-06	8.24053e-06
250 - 300	0.000612111	7.19088e-06	6.97137e-06
300 - 350	0.000293318	5.82887e-06	5.7904e-06
350 - 400	0.000136458	4.37717e-06	4.23041e-06
400 - 470	8.72444e-05	3.16199e-06	3.67545e-06
470 - 550	4.3269e-05	2.48122e-06	3.23724e-06
550 - 650	1.98859e-05	1.8606e-06	2.5468e-06
650 - 900	1.28455e-05	1.48521e-06	2.1452e-06

Table A.15: Combination: measured normalized ϕ^*_η cross-section on born level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 0.004	0.035224	3.3002e-05	4.22236e-05
0.004 - 0.008	0.0347745	3.43065e-05	4.08972e-05
0.008 - 0.012	0.0342777	3.35231e-05	3.77037e-05
0.012 - 0.016	0.0333618	3.27754e-05	3.9893e-05
0.016 - 0.02	0.032353	3.32583e-05	3.89072e-05
0.02 - 0.024	0.0311558	3.33819e-05	3.80512e-05
0.024 - 0.029	0.0370657	3.32104e-05	3.80942e-05
0.029 - 0.034	0.0351749	3.35412e-05	3.66745e-05
0.034 - 0.039	0.0330133	3.34642e-05	3.74524e-05
0.039 - 0.045	0.0369318	3.38181e-05	3.60054e-05
0.045 - 0.051	0.0342557	3.3463e-05	3.54671e-05
0.051 - 0.057	0.0316793	3.32858e-05	3.57894e-05
0.057 - 0.064	0.0339437	3.32204e-05	3.57051e-05
0.064 - 0.072	0.035285	3.31613e-05	3.54866e-05
0.072 - 0.081	0.0357135	3.34499e-05	3.64786e-05
0.081 - 0.091	0.0355081	3.32183e-05	3.64546e-05
0.091 - 0.102	0.0345578	3.32188e-05	3.54987e-05
0.102 - 0.114	0.0332071	3.25967e-05	3.72334e-05
0.114 - 0.128	0.0337539	3.21143e-05	3.80415e-05
0.128 - 0.145	0.0351125	3.21205e-05	3.77979e-05
0.145 - 0.165	0.034607	3.16505e-05	3.92337e-05
0.165 - 0.189	0.034277	3.22664e-05	3.78279e-05
0.189 - 0.219	0.0343892	3.18751e-05	3.65997e-05
0.219 - 0.258	0.0345321	3.23376e-05	3.52203e-05
0.258 - 0.312	0.0349403	3.14827e-05	3.36994e-05
0.312 - 0.391	0.0346584	3.00498e-05	3.07156e-05
0.391 - 0.524	0.0347049	2.7785e-05	2.91862e-05
0.524 - 0.695	0.0242012	2.55494e-05	2.71808e-05
0.695 - 0.918	0.016647	2.32721e-05	2.36343e-05
0.918 - 1.153	0.00951672	2.11372e-05	2.04394e-05
1.153 - 1.496	0.00743976	1.87117e-05	1.80328e-05
1.496 - 1.947	0.0048905	1.72572e-05	1.47492e-05
1.947 - 2.522	0.00310003	1.54292e-05	1.29212e-05
2.522 - 3.277	0.00206868	1.4279e-05	1.17199e-05
3.277 - 5	0.00202566	1.27274e-05	1.0355e-05
5 - 10	0.00168588	1.16021e-05	9.13053e-06

Table A.16: Combination: measured normalized ϕ^*_η cross-section on bare-level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 0.004	0.0349391	4.46912e-05	3.94318e-05
0.004 - 0.008	0.0345122	4.01181e-05	3.71815e-05
0.008 - 0.012	0.0340039	3.95384e-05	3.22906e-05
0.012 - 0.016	0.0331228	4.08665e-05	3.55249e-05
0.016 - 0.02	0.0321624	4.04856e-05	3.44312e-05
0.02 - 0.024	0.0309842	3.99389e-05	3.34194e-05
0.024 - 0.029	0.0369005	3.94617e-05	3.48077e-05
0.029 - 0.034	0.0349995	4.06997e-05	3.27768e-05
0.034 - 0.039	0.032911	3.90833e-05	3.40499e-05
0.039 - 0.045	0.03685	3.83277e-05	3.25704e-05
0.045 - 0.051	0.0342098	3.68061e-05	3.20686e-05
0.051 - 0.057	0.0316576	3.65624e-05	3.30752e-05
0.057 - 0.064	0.0339544	3.75391e-05	3.29982e-05
0.064 - 0.072	0.035338	3.76844e-05	3.27656e-05
0.072 - 0.081	0.0357531	3.79268e-05	3.37307e-05
0.081 - 0.091	0.0355645	3.85042e-05	3.3787e-05
0.091 - 0.102	0.0346552	3.89863e-05	3.2265e-05
0.102 - 0.114	0.0332685	3.96332e-05	3.45945e-05
0.114 - 0.128	0.0338758	3.84928e-05	3.63542e-05
0.128 - 0.145	0.0352091	3.87975e-05	3.60471e-05
0.145 - 0.165	0.0347445	3.92615e-05	3.83494e-05
0.165 - 0.189	0.034406	3.91013e-05	3.64897e-05
0.189 - 0.219	0.03451	3.89142e-05	3.49478e-05
0.219 - 0.258	0.0346401	3.78033e-05	3.38266e-05
0.258 - 0.312	0.0350531	3.64531e-05	3.25578e-05
0.312 - 0.391	0.0347728	3.54974e-05	2.99735e-05
0.391 - 0.524	0.0348692	3.27322e-05	2.84954e-05
0.524 - 0.695	0.0243506	3.02172e-05	2.67034e-05
0.695 - 0.918	0.0167975	2.65786e-05	2.28032e-05
0.918 - 1.153	0.00962036	2.40141e-05	1.96521e-05
1.153 - 1.496	0.00753356	2.12824e-05	1.76517e-05
1.496 - 1.947	0.00496001	1.90926e-05	1.41099e-05
1.947 - 2.522	0.00313391	1.7779e-05	1.24537e-05
2.522 - 3.277	0.00208345	1.67468e-05	1.1273e-05
3.277 - 5	0.00203526	1.55534e-05	9.836e-06
5 - 10	0.00168653	1.41101e-05	8.47548e-06

Table A.17: Combination: measured normalized ϕ^*_η cross-section on dressed-level definition including uncertainties. Results taken from Ref. [54].

Binning [GeV]	Value	Uncorrelated	Correlated
0 - 0.004	0.0349469	3.34703e-05	3.90875e-05
0.004 - 0.008	0.0345232	3.40107e-05	3.6728e-05
0.008 - 0.012	0.0340365	3.30096e-05	3.34266e-05
0.012 - 0.016	0.0331452	3.36301e-05	3.63187e-05
0.016 - 0.02	0.0321724	3.38353e-05	3.52204e-05
0.02 - 0.024	0.030996	3.41649e-05	3.44198e-05
0.024 - 0.029	0.0369003	3.39673e-05	3.54702e-05
0.029 - 0.034	0.0350242	3.47706e-05	3.41183e-05
0.034 - 0.039	0.0329308	3.47166e-05	3.46997e-05
0.039 - 0.045	0.0368541	3.51593e-05	3.34197e-05
0.045 - 0.051	0.0342214	3.45764e-05	3.30073e-05
0.051 - 0.057	0.0316649	3.37488e-05	3.41425e-05
0.057 - 0.064	0.0339602	3.46434e-05	3.41816e-05
0.064 - 0.072	0.0353447	3.41797e-05	3.38192e-05
0.072 - 0.081	0.0357716	3.40013e-05	3.45153e-05
0.081 - 0.091	0.0355837	3.31954e-05	3.45908e-05
0.091 - 0.102	0.0346689	3.31695e-05	3.36022e-05
0.102 - 0.114	0.0332996	3.24277e-05	3.55629e-05
0.114 - 0.128	0.0339018	3.17266e-05	3.66372e-05
0.128 - 0.145	0.0352444	3.15205e-05	3.6612e-05
0.145 - 0.165	0.0347754	3.13613e-05	3.82952e-05
0.165 - 0.189	0.0344335	3.14262e-05	3.67808e-05
0.189 - 0.219	0.0345371	3.13728e-05	3.5239e-05
0.219 - 0.258	0.0346829	3.08549e-05	3.37676e-05
0.258 - 0.312	0.03506	3.00196e-05	3.25203e-05
0.312 - 0.391	0.0347688	2.86299e-05	3.00543e-05
0.391 - 0.524	0.0348062	2.64584e-05	2.8694e-05
0.524 - 0.695	0.0242636	2.43756e-05	2.658e-05
0.695 - 0.918	0.016702	2.21945e-05	2.29776e-05
0.918 - 1.153	0.0095508	1.96696e-05	2.00184e-05
1.153 - 1.496	0.00746865	1.75353e-05	1.77467e-05
1.496 - 1.947	0.00491465	1.5756e-05	1.4499e-05
1.947 - 2.522	0.00311467	1.44229e-05	1.27839e-05
2.522 - 3.277	0.00207347	1.33426e-05	1.15793e-05
3.277 - 5	0.00202779	1.21578e-05	1.01343e-05
5 - 10	0.00168137	1.10963e-05	8.93557e-06

A.5 Sherpa generator plots

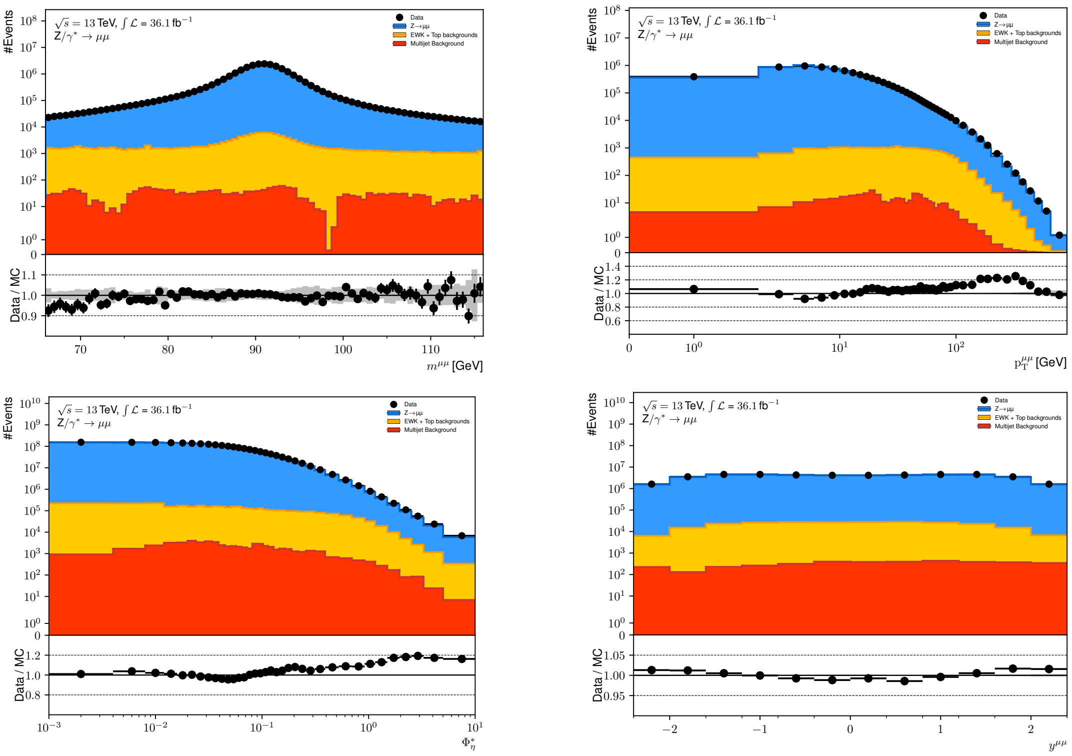


Figure A.1: Sherpa control plots for the Z mass $m_{\mu\mu}$ (upper left), Z transverse momentum $\mathrm{p}_\mathrm{T}^{\mu\mu}$ (upper right) as well as muon $\phi^*_{\mu\mu}$ (lower left) rapidty $y_T^{\mu\mu}$ (lower right). All plots show detector-level data. The MC samples are normalized as such that the MC luminosity fits the data luminosity.

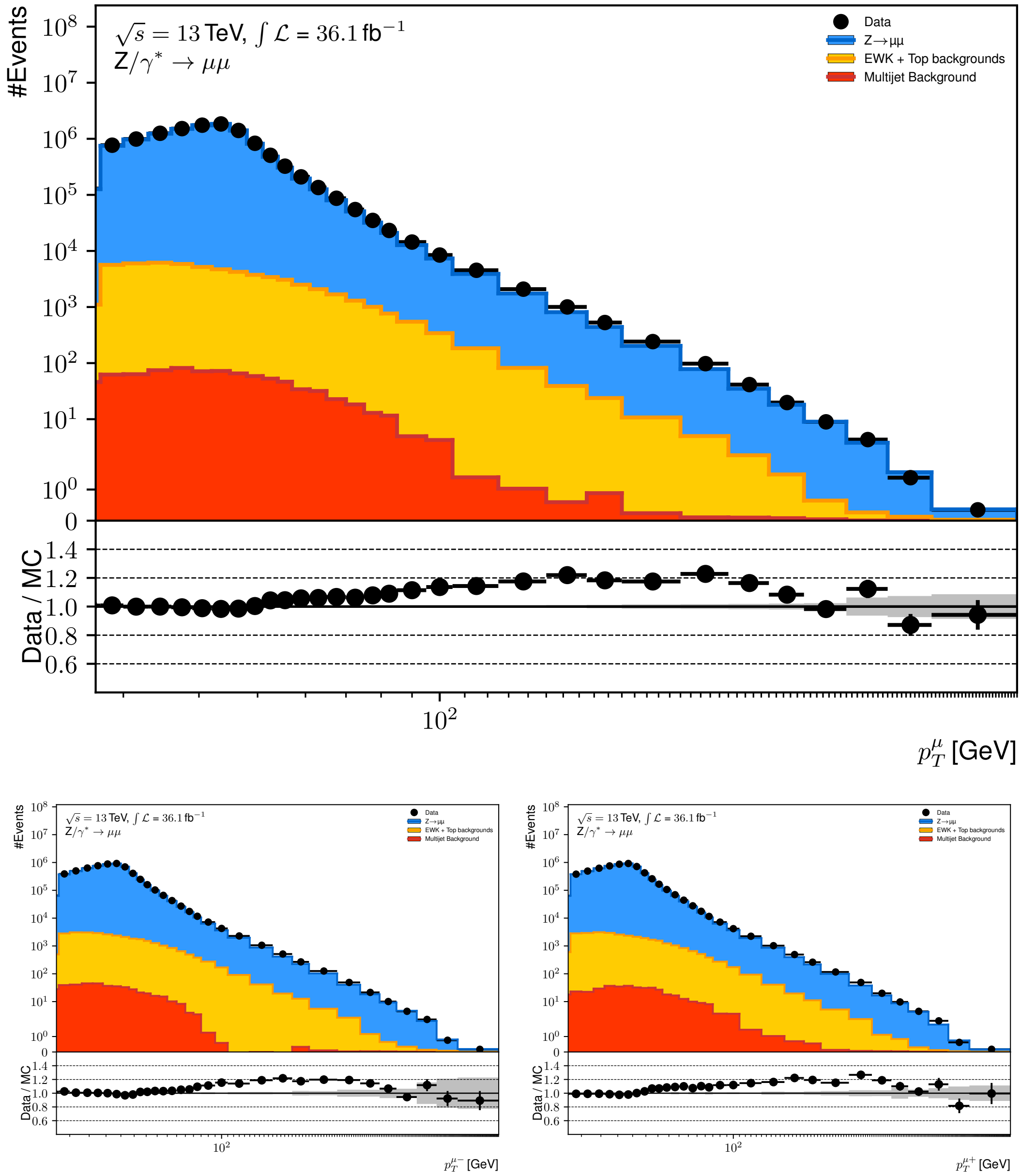


Figure A.2: Sherpa control plots of the muon transverse momentum p_T^μ (up) for all muons as well as $p_T^{\mu+}$ for μ^+ (lower left) and $p_T^{\mu+}$ for μ^- (lower right).

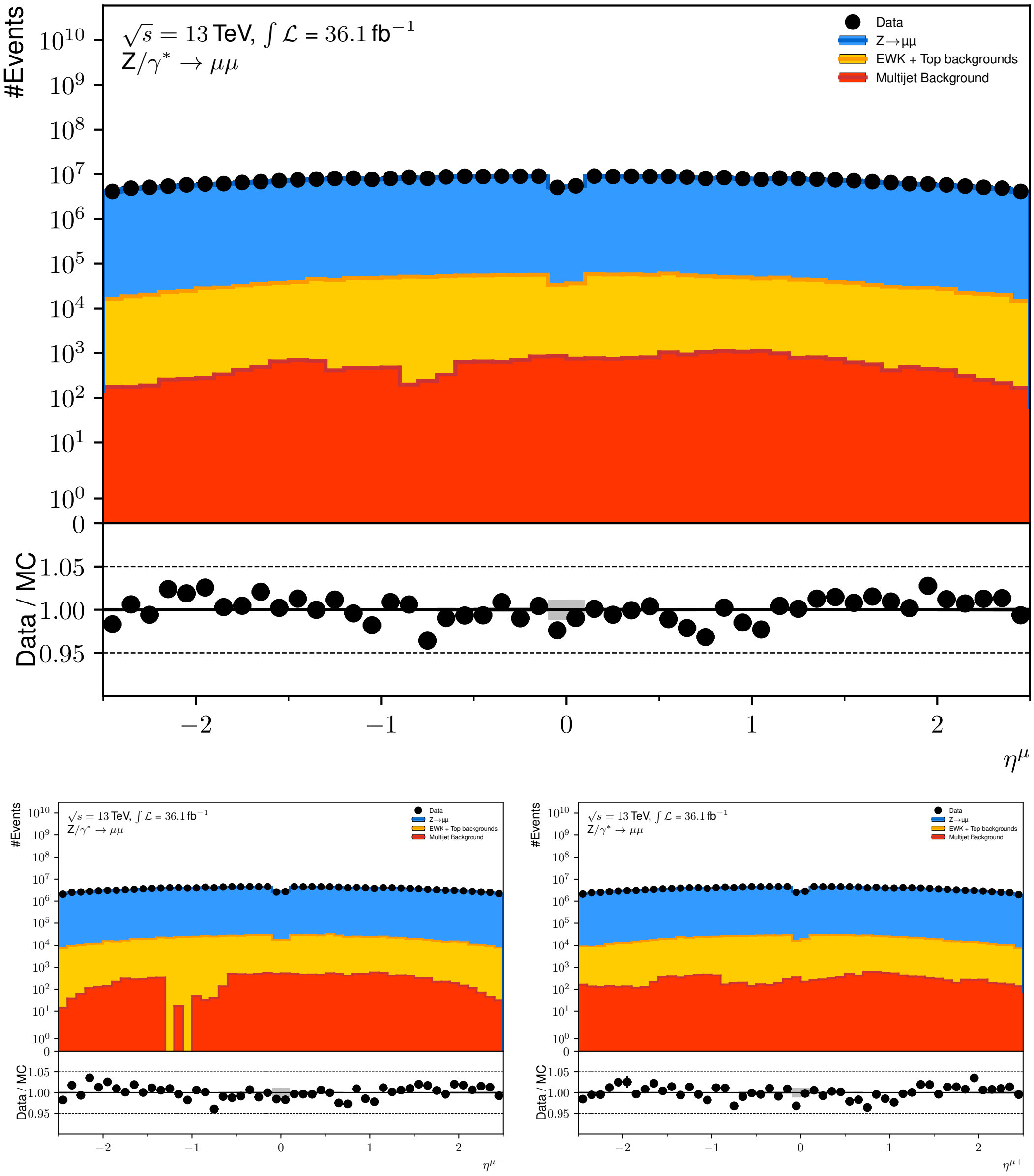


Figure A.3: Sherpa control plots of the muon pseudorapidity η_μ (up) for all muons as well as $\eta_{\mu+}$ for μ^+ (lower left) and $\eta_{\mu+}$ for μ^- (lower right).

A.6 Additional differential cross section plots

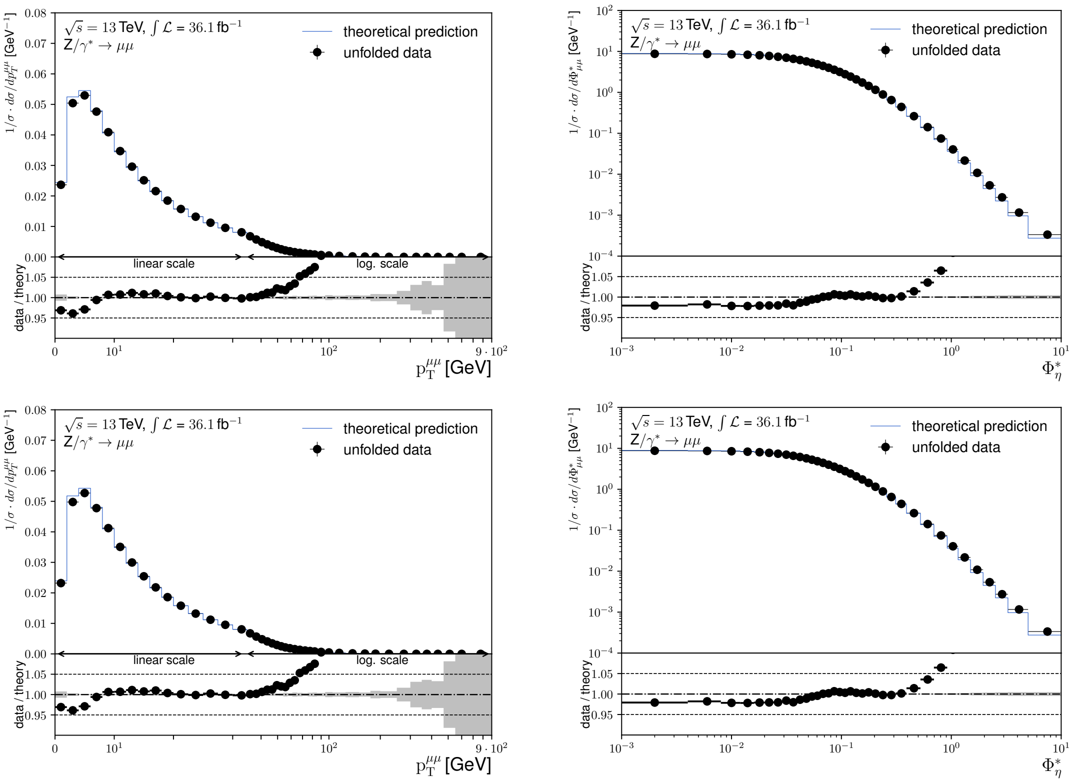


Figure A.4: Normalized differential cross-section $p_T(Z)$ (left) and Φ^* (right) for dressed-Level (upper row) and bare-level muons (lower row).

A.7 Corrections to this thesis

Title page:

- Removed blank page
- **Gutachter** changed to **Berichterstatter**
- Fixed alignment of last two rows

Content:

- Section 9.9.3: 2.% changed to 2 %
- Section 9.8: **9.9.4** changed to **Section 9.9.4**
- Section 9.8: resized image
- Section 8.2: **initial** changed to **ISR**
- Equation 8.18: $-\hat{\mathrm{p}}$ changed to $+\hat{\mathrm{p}}$
- Equation 8.5: removed f
- Equation 7.4 added vector arrows
- Section 9.12: removed line break due to shift in text
- Equation 9.20: $\mathbf{C_k}$ changed to $\mu_\mathbf{k}$
- Page 133, Item 1: $\mathbf{P}(\mathbf{x_i}, \mathbf{I})$ changed to $\mathbf{P}(\mu_\mathbf{i}, \mathbf{I})$
- Figure 3.3: corrected caption

Acronyms

Adam	**Ada**ptive **M**oment Estimation
ALICE	**A** **LHC** **I**on **C**ollider **E**xperiment
ANN	**A**rtificial **N**eural **N**etworks
ATLAS	**A** **T**oroidal **LHC** **A**pparatu**S**
BCM	**B**eam **C**onditions **M**onitor
CERN	European Organization for Nuclear Research
CMS	**C**ompact **M**uon **S**olenoid
CSC	**C**athode **S**trip **C**hamber
CTP	**C**entral **T**rigger **P**rocessor
EMEC	**E**lectro**m**agnetic **e**nd-cap **C**alorimeters
FNN	**F**eedforward **N**eural **N**etworks
FSR	**F**inal-**S**tate **R**adiation
GRL	**G**ood **R**un **L**ist
HERA	**H**adron-**E**lectron **R**ing **A**ccelerator
HLT	**H**igh-**L**evel **T**rigger
ID	**I**nner **D**etector
ISR	**I**nitial-**S**tate **R**adiation
LHC	**L**arge **H**adron **C**ollider
LHCb	**LHC** **B**eauty experiment
LINAC2	**Lin**ear **AC**celerator **2**
LO	**L**eading **O**rder
MC	**M**onte **C**arlo
MDT	**M**onitored **D**rift **T**ubes
MS	**M**uon **S**ystem

NLO	**N**ext-to-**L**eading **O**rder
NNLO	**N**ext-to-**N**ext-to-**L**eading **O**rder
PDF	**P**arton **D**istribution **F**unctions
PS	**P**roton **S**ynchrotron
PSB	**P**roton **S**ynchrotron **B**ooster
QCD	**Q**uantum **C**hromo**d**ynamics
QED	**Q**uantum **E**lectro**d**ynamics
RMS	**R**oot **M**ean **S**quare
RPC	**R**esistive **P**late **C**hamber
SCT	**S**ilicon **M**icrostrip **L**ayers
SK	**S**oft**K**iller
SM	Standard Model of Particle Physics
SPS	**S**uper **P**roton **S**ynchrotron
TGC	**T**hin **G**ap **C**hamber
TRT	**T**ransition **R**adiation **T**racker
TST	**T**rack **S**oft **T**erm
vdM	**v**an **d**er **M**eer

List of Figures

List of Tables

Bibliography

[1] T. Aaltonen, B. Álvarez González, S. Amerio, D. Amidei, A. Anastassov, A. Annovi, J. Antos, G. Apollinari, J. A. Appel, T. Arisawa, and et al., "Precise measurement of the w-boson mass with the cdf ii detector", Physical Review Letters **108**, `10.1103/physrevlett.108.151803` (2012).

[2] V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, G. D. Alexeev, G. Alkhazov, A. Alton, G. Alverson, M. Aoki, and et al., "Measurement of the w boson mass with the d0 detector", Physical Review Letters **108**, `10.1103/physrevlett.108.151804` (2012).

[3] Jasmin Abdallah, a Adragna, b Alexa, c Alves, d Amaral, f Ananiev, g Anderson, h Andresen, f Antonaki, i Batusov, j Bednar, k Behrens, e Bergeaas, l Biscarat, m Blanch, Georges Blanchot, Jacek Blocki, ab Bohm, l Boldea, and b Zenzh, "Mechanical construction and installation of the atlas tile calorimeter", Jinst **8**, `10.1088/1748-0221/8/11/T11001` (2013).

[4] R. Achenbach et al., "The ATLAS level-1 calorimeter trigger", JINST **3**, P03001 (2008).

[5] Agostinelli et al., "Geant4—a simulation toolkit", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250–303 (2003).

[6] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP **06**, 043 (2010).

[7] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, "NLO vector-boson production matched with shower in POWHEG", JHEP **07**, 060 (2008).

[8] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", Nuclear Physics B **126**, 298–318 (1977).

[9] Y. Arai, M. Asai, E. Barberio, T. Emura, J. Goldberg, K. Homma, M. Ikeno, M. Imori, K. Ishii, H. Ishiwaki, T. Kawamoto, T. Kobayashi, D. Lellouch, L. Levinson, N. Lupu, G. Mikenberg, M. Miyake, K. Nagai, T. Nagano, I. Nakamura, M. Nomachi, M. Nozaki, S. Odaka, T.K. Ohska, O. Sasaki, H. Shirasu, H. Takeda, T. Takeshita, S. Tanaka, and C. Yokoyama, "Timing optimization of thin gap chambers for the use in the atlas muon endcap trigger", Nuclear

Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **367**, Proceedings of the 7th International Wire Chamber Conference, 398–401 (1995).

[10] T. Argyropoulos, K. A. Assamagan, B. H. Benedict, V. Chernyatin, E. Cheu, J. Deng, A. Gordeev, I. Gough Eschrich, V. Grachev, K. Johns, S. Junarkar, A. Kandasamy, V. Kaushik, A. Khodinov, A. Lankford, X. Lei, R. Murillo Garcia, K. Nikolopoulos, P. O'Connor, C. L. Parnell-Lampen, W. Park, V. Polychronakos, R. Porter, M. Purohit, M. Schernau, B. K. Toggerson, and D. Tompkins, "Cathode strip chambers in atlas : installation, commissioning and in situ performance", in 2008 ieee nuclear science symposium conference record (Oct. 2008), pp. 2819–2824.

[11] S. Artz, B. Bauss, H. Boterenbrood, V. Buescher, Augusto Cerqueira, R. Degele, Sumandeep Dhaliwal, N. Ellis, Philippe Farthouat, G. Galster, Marco Ghibaudi, J. Glatzer, S. Haas, O. Igonkina, K. Jakobi, Peter Jansweijer, C. Kahra, A. Kaluza, Michiru Kaneda, and M. Zinser, "The atlas level-1 muon topological trigger information for run 2 of the lhc", Journal of Instrumentation **10**, `10.1088/1748-0221/10/02/C02027` (2015).

[12] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014.

[13] ATLAS Collaboration, "Measurement of angular correlations in Drell-Yan lepton pairs to probe Z/γ^* boson transverse momentum at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector", Phys. Lett. B **720**, 32 (2013).

[14] ATLAS Collaboration, "Measurement of the (W)-boson mass in (pp) collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector", Eur. Phys. J. C **78**, 110 (2018).

[15] ATLAS Collaboration, "Measurement of the Z/γ^* boson transverse momentum distribution in (pp) collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector", JHEP **09**, 145 (2014).

[16] ATLAS Collaboration, "Measurements of top-quark pair to Z-boson cross-section ratios at $\sqrt{s} = 13, 8, 7\,\text{TeV}$ with the ATLAS detector", JHEP **02**, 117 (2017).

[17] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$", Eur. Phys. J. C **76**, 292 (2016).

[18] ATLAS Collaboration, "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV", Eur. Phys. J. **C78**, 903 (2018).

[19] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012.

[20] *Atlas experiment - public results*, (July 2019) `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_201`.

[21] *ATLAS magnet system: Technical Design Report, 1*, Technical Design Report ATLAS (CERN, Geneva, 1997).

[22] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, "Precise prediction for the w-boson mass in the standard model", Physical Review D **69**, `10.1103/physrevd.69.053006` (2004).

[23] Richard D. Ball et al. (NNPDF), "Parton distributions for the LHC Run II", JHEP **04**, 040 (2015).

[24] Richard D. Ball et al. (NNPDF), "Parton distributions from high-precision collider data", Eur. Phys. J. C **77**, 663 (2017).

[25] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo (NNPDF), "Parton distributions with QED corrections", Nucl. Phys. B **877**, 290–320 (2013).

[26] V.D. Barger and R.J.N. Phillips, *Collider physics*, Frontiers in Physics (Avalon Publishing, 1996).

[27] P. Berta, L. Masetti, D. W. Miller, and M. Spousta, "Pileup and Underlying Event Mitigation with Iterative Constituent Subtraction", (2019).

[28] Peter Berta, Martin Spousta, David W. Miller, and Rupert Leitner, "Particle-level pileup subtraction for jets and jet shapes", Journal of High Energy Physics **2014**, `10.1007/jhep06(2014)092` (2014).

[29] Wojciech Bizon, Xuan Chen, Aude Gehrmann-De Ridder, Thomas Gehrmann, Nigel Glover, Alexander Huss, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Paolo Torrielli, "Fiducial distributions in Higgs and Drell-Yan production at N^{3}LL+NNLO", JHEP **12**, 132 (2018).

[30] Wojciech Bizon, Aude Gehrmann-De Ridder, Thomas Gehrmann, Nigel Glover, Alexander Huss, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Duncan M. Walker, "The transverse momentum spectrum of weak gauge bosons at N^{3}LL+NNLO", Eur. Phys. J. **C79**, 868 (2019).

[31] Wojciech Bizon, Pier Francesco Monni, Emanuele Re, Luca Rottoli, and Paolo Torrielli, "Momentum-space resummation for transverse observables and the Higgs $p_\perp$ at N^{3}LL+NNLO", JHEP **02**, 108 (2018).

[32] Enrico Bothmann et al., "Event Generation with Sherpa 2.2", SciPost Phys. **7**, 034 (2019).

[33] Enrico Bothmann, Marek Schönherr, and Steffen Schumann, "Reweighting QCD matrix-element and parton-shower calculations", Eur. Phys. J. C **76**, 590 (2016).

[34] Bernard Brickwedde, Benjamin Philip Nachman, and Matthias Schott, *Convolutional Neural Networks with Event Images for Pileup Mitigation with the ATLAS Detector*, tech. rep. ATL-COM-PHYS-2019-628 (CERN, Geneva, May 2019).

[35] Carsten Burgard, *Example: standard model of physics*, (Dec. 2016) `http://www.texample.net/tikz/examples/model-physics/` (visited on 12/31/2016).

[36] P Burian, P Broulim, B Bergmann, V Georgiev, S Pospisil, L Pusman, and J Zich, "Timepix3 detector network at ATLAS experiment", JINST **13**, C11024. 16 p (2018).

[37] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam, "Jet substructure as a new Higgs search channel at the LHC", Phys. Rev. Lett. **100**, 242001 (2008).

[38] William Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*, tech. rep. ATL-COM-SOFT-2015-119 (CERN, Geneva, May 2015).

[39] Matteo Cacciari and Gavin P. Salam, "Pileup subtraction using jet areas", Phys. Lett. **B659**, 119–126 (2008).

[40] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, "Softkiller, a particle-level pileup removal method", The European Physical Journal C **75**, `10.1140/epjc/s10052-015-3267-2` (2015).

[41] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, *The anti$-k_t$ jet clustering algorithm*, 2008.

[42] J M Campbell, J W Huston, and W J Stirling, "Hard interactions of quarks and gluons: a primer for lhc physics", Reports on Progress in Physics **70**, 89–193 (2006).

[43] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini, "Scattering Amplitudes with Open Loops", Phys. Rev. Lett. **108**, 111601 (2012).

[44] Giordano Cattani, "The resistive plate chambers of the atlas experiment: performance studies", Journal of Physics: Conference Series **280**, `10.1088/1742-6596/280/1/012001` (2011).

[45] ATLAS Collaboration, "ATLAS Calorimeter: Run 2 performance and Phase-II upgrades", PoS **EPS-HEP2017**, 485 (2017).

[46] ATLAS Collaboration, "ATLAS muon spectrometer: Technical design report", (1997).

[47] ATLAS Collaboration, "ATLAS pixel detector electronics and sensors", JINST **3**, P07007 (2008).

[48] ATLAS Collaboration, *Constituent-level pile-up mitigation techniques in ATLAS*, tech. rep. ATLAS-CONF-2017-065 (CERN, Geneva, Aug. 2017).

[49] ATLAS Collaboration, *Generalized Numerical Inversion: A Neural Network Approach to Jet Calibration*, tech. rep. ATL-PHYS-PUB-2018-013 (CERN, Geneva, July 2018).

[50] ATLAS Collaboration, "Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector", Eur. Phys. J. **C77**, [Erratum: Eur. Phys. J.C77,no.10,712(2017)], 580 (2017).

[51] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, tech. rep. ATLAS-CONF-2019-021 (CERN, Geneva, June 2019).

[52] ATLAS Collaboration, “Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data.”, Eur. Phys. J. C **74**, Comments: 21 pages plus author list + cover pages (34 pages total), 21 figures, 2 tables, submitted to EPJC, All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/ GROUPS/PHYSICS/-PAPERS/ PERF-2014-05/, 3130. 34 p (2014).

[53] ATLAS Collaboration, “Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, Eur. Phys. J. **C76**, 291 (2016).

[54] ATLAS Collaboration, *Measurement of the transverse momentum distribution of drell-yan lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ tev with the atlas detector*, 2019.

[55] ATLAS Collaboration, “Measurement of the w-boson mass in pp collisions at $\sqrt{s} = 7$ tev with the atlas detector”, The European Physical Journal C **78**, `10.1140/epjc/s10052-017-5475-4` (2018).

[56] ATLAS Collaboration, “Measurements of top-quark pair to z-boson cross-section ratios at $\sqrt{s}$ =13, 8, 7 tev with the atlas detector”, Journal of High Energy Physics **2017**, `10.1007/jhep02(2017)117` (2017).

[57] ATLAS Collaboration, *Monitored drift tubes in ATLAS*, tech. rep. ATL-M-PN-129 (CERN, Geneva, Sept. 1996).

[58] ATLAS Collaboration, “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV.”, Eur. Phys. J. C **76**, Comments: 27 pages inlcuding cover page plus author list (45 pages total), 12 figures, 3 tables, submitted to Eur. Phys. J. C. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/ PAPERS/PERF-2015-10/, 292. 45 p (2016).

[59] ATLAS Collaboration, “Performance of pile-up mitigation techniques for jets in *pp* collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, Eur. Phys. J. **C76**, 581 (2016).

[60] ATLAS Collaboration (ATLAS Collaboration), “Performance of the ATLAS Trigger System in 2015. Performance of the ATLAS Trigger System in 2015”, Eur. Phys. J. C **77**, 77 pages in total, author list starting page 61, 50 figures, 1 table. Published in Eur. Phys. J. C. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2016-01/, 317. 76 p (2016).

[61] ATLAS Collaboration, *Primary vertex reconstruction at the ATLAS experiment*, tech. rep. ATL-SOFT-PROC-2017-051. 4 (CERN, Geneva, Feb. 2017).

[62] ATLAS Collaboration, "The atlas experiment at the cern large hadron collider", JINST **S08003** (2008).

[63] ATLAS Collaboration, *The ATLAS Tile Calorimeter*, tech. rep. ATL-TILECAL-PROC-2015-002 (CERN, Geneva, Mar. 2015).

[64] ATLAS Collaboration, "The ATLAS Trigger in Run 2: Design, Menu, and Performance", PoS **EPS-HEP2017**, 525 (2017).

[65] ATLAS Collaboration, *The Design and Performance of the ATLAS Inner Detector Trigger for Run 2 LHC Collisions at* $\sqrt{s} = 13\ TeV$, tech. rep. ATL-DAQ-PROC-2016-004 (CERN, Geneva, Feb. 2016).

[66] ATLAS Collaboration, "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS", JINST **13**, P07017. 33 p (2018).

[67] ATLAS Collaboration, "The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim", (2010).

[68] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1", Eur. Phys. J. **C77**, 490 (2017).

[69] ATLAS Collaboration, *Tracking and Vertexing with the ATLAS Inner Detector in the LHC Run-2*, tech. rep. ATL-PHYS-PROC-2017-075 (CERN, Geneva, June 2017).

[70] The ALICE Collaboration, "The ALICE experiment at the CERN LHC", Journal of Instrumentation **3**, S08002–S08002 (2008).

[71] The CMS Collaboration, "The CMS experiment at the CERN LHC", Journal of Instrumentation **3**, S08004–S08004 (2008).

[72] The LHCb Collaboration, "The LHCb detector at the LHC", Journal of Instrumentation **3**, S08005–S08005 (2008).

[73] J C Collins, Davison Eugene Soper, and George F Sterman, "Transverse momentum distribution in Drell-Yan pair and W and Z boson production", Nucl. Phys. B **250**, 199–224. 37 p (1984).

[74] John C. Collins, Davison E. Soper, and George Sterman, *Factorization of hard processes in qcd*, 2004.

[75] Wikimedia Commons, *Summary of interactions between particles described by the standard model*, (2014) `https://commons.wikimedia.org/wiki/File:Elementary_particle_interactions_in_the_Standard_Model.png`.

[76] Aviv Cukierman and Benjamin Nachman, "Mathematical Properties of Numerical Inversion for Jet Calibrations", Nucl. Instrum. Meth. **A858**, 1–11 (2017).

[77] D. J. Lange, "The EvtGen particle decay simulation package", Nucl. Instrum. Meth. A **462**, 152 (2001).

[78] G. D'Agostini, "A Multidimensional unfolding method based on Bayes' theorem", Nucl. Instrum. Meth. **A362**, 487–498 (1995).

[79] G. D'Agostini, *Improved iterative bayesian unfolding*, 2010.

[80] N. Davidson, T. Przedzinski, and Z. Was, "PHOTOS Interface in C++: Technical and Physics Documentation", Comput. Phys. Commun. **199**, 86–101 (2016).

[81] Ansgar Denner, Stefan Dittmaier, and Lars Hofer, "Collier: a fortran-based complex one-loop library in extended regularizations", Comput. Phys. Commun. **212**, 220–238 (2017).

[82] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Mück, "Electroweak corrections to dilepton + jet production at hadron colliders", JHEP **06**, 069 (2011).

[83] Yuri 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**, [Zh. Eksp. Teor. Fiz.73,1216(1977)], 641–653 (1977).

[84] Sidney D. Drell and Tung-Mow Yan, "Massive lepton-pair production in hadron-hadron collisions at high energies", Phys. Rev. Lett. **25**, 316–320 (1970).

[85] *Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s}$ = 13 TeV*, tech. rep. ATL-PHYS-PUB-2015-051 (CERN, Geneva, Dec. 2015).

[86] K. Edmonds, S. Fleischmann, T. Lenz, C. Magass, J. Mechnich, and A. Salzburger, "The fast ATLAS track simulation (FATRAS)", (2008).

[87] "Electroweak measurements in electron–positron collisions at w-boson-pair energies at lep", Physics Reports **532**, Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP, 119–244 (2013).

[88] Stephen D Ellis, Christopher K Vermilion, and Jonathan R Walsh, "Recombination algorithms and jet substructure: pruning as a tool for heavy particle searches", Physical Review D **81**, 094023 (2010).

[89] Stephen D Ellis, Christopher K Vermilion, and Jonathan R Walsh, "Techniques for improved heavy particle searches with jet substructure", Physical Review D **80**, 051501 (2009).

[90] Stephen D. Ellis and Davision E. Soper, *Successive combination jet algorithm for hadron collisions*, 1993.

[91] Jens Erler and Matthias Schott, "Electroweak precision tests of the standard model after the discovery of the higgs boson", Progress in Particle and Nuclear Physics **106**, 68–119 (2019).

[92] Lyndon Evans and Philip Bryant, "LHC machine", Journal of Instrumentation **3**, S08001–S08001 (2008).

[93] F. Chollet, *Keras*, 2015.

[94] Fehling, *Standard model of fundamental particles and interactions*, (Aug. 2016) `https://web.archive.org/web/20160304133522/https://www.pha.jhu.edu/~dfehling/particle.gif`.

[95] Stefano Frixione, Paolo Nason, and Carlo Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, JHEP **11**, 070 (2007).

[96] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, JHEP **09**, 126 (2007).

[97] G. Arnison et al., “Experimental observation of lepton pairs of invariant mass around 95 gev/c^2 at the cern sps collider”, Physics Letters B **126**, 398–410 (1983).

[98] Jonathan R. Gaunt and W. James Stirling, “Double parton distributions incorporating perturbative qcd evolution and momentum and quark number sum rules”, Journal of High Energy Physics **2010**, `10.1007/jhep03(2010)005` (2010).

[99] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and T. A. Morgan, “NNLO QCD corrections for Drell-Yan p_T^Z and ϕ^* observables at the LHC”, JHEP **11**, 094 (2016); Erratum: JHEP **10**, 126 (2018).

[100] Alexander Glazov, *Machine learning as an instrument for data unfolding*, 2017.

[101] Tanju Gleisberg and Stefan Höche, “Comix, a new matrix element generator”, JHEP **12**, 039 (2008).

[102] Xavier Glorot and Yoshua Bengio, “Understanding the difficulty of training deep feedforward neural networks”, in Proceedings of the thirteenth international conference on artificial intelligence and statistics, Vol. 9, edited by Yee Whye Teh and Mike Titterington, Proceedings of Machine Learning Research (May 2010), pp. 249–256.

[103] Piotr Golonka and Zbigniew Was, “PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays”, Eur. Phys. J. C **45**, 97–107 (2006).

[104] V. N. Gribov and L. N. Lipatov, “Deep inelastic e p scattering in perturbation theory”, Sov. J. Nucl. Phys. **15**, [Yad. Fiz.15,781(1972)], 438–450 (1972).

[105] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, *Delving deep into rectifiers: surpassing human-level performance on imagenet classification*, 2015.

[106] Fabian Huegging, *The atlas pixel detector*, 2004.

[107] Sergey Ioffe and Christian Szegedy, *Batch normalization: accelerating deep network training by reducing internal covariate shift*, 2015.

[108] Diederik P. Kingma and Jimmy Ba, *Adam: a method for stochastic optimization*, 2014.

[109] Nicolas Maximilian Koehler (ATLAS), “ATLAS Muon Reconstruction Performance in LHC Run 2”, in Proceedings, 3rd Large Hadron Collider Physics Conference (LHCP 2015): St. Petersburg, Russia, August 31-September 5, 2015 (CERN, 2016), pp. 621–626.

[110] Nicolas Maximilian Kohler, “Searches for the Supersymmetric Partner of the Top Quark, Dark Matter and Dark Energy at the ATLAS Experiment”, Presented on 17 8 2018 (July 2019).

[111] Fabian Kohn, “Measurement of the charge asymmetry in top quark pair production in pp collision data at $\sqrt{s} = 7\,\text{TeV}$ using the ATLAS detector”, Presented 07 Mar 2012 (Apr. 2012).

[112] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz, “Pileup mitigation with machine learning (pumml)”, Journal of High Energy Physics **2017**, `10.1007/jhep12(2017)051` (2017).

[113] C Kourkoumelis and S Vourakis, “Hypatia—an online tool for atlas event visualization”, Physics Education **49**, 21 (2014).

[114] David Krohn, Matthew D. Schwartz, Matthew Low, and Lian-Tao Wang, “Jet Cleansing: Pileup Removal at High Luminosity”, Phys. Rev. **D90**, 065020 (2014).

[115] David Krohn, Jesse Thaler, and Lian-Tao Wang, “Jet Trimming”, JHEP **1002**, 084 (2010).

[116] Suman Kumar Kundu, Tanmay Sarkar, and Manas Maity, “Optimizing the parton shower model in PYTHIA with pp collision data at $\sqrt{s} = 13\,\text{TeV}$”, International Journal of Modern Physics A **34**, 1950219 (2019).

[117] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. -P. Yuan, “New parton distributions for collider physics”, Phys. Rev. D **82**, 074024 (2010).

[118] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler, “Soft Drop”, JHEP **05**, 146 (2014).

[119] J. M. Lindert et al., “Precise predictions for V+jets dark matter backgrounds”, Eur. Phys. J. C **77**, 829 (2017).

[120] Louis Lyons, Duncan Gibaut, and Peter Clifford, “How to Combine Correlated Estimates of a Single Physical Quantity”, Nucl. Instrum. Meth. A **270**, 110 (1988).

[121] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the lhc”, The European Physical Journal C **63**, 189–285 (2009).

[122] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC”, Eur. Phys. J. C **63**, 189–285 (2009).

[123] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and

Xiaoqiang Zheng, *TensorFlow: large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[124] B. Mehlig, *Artificial neural networks*, 2019.

[125] V. A. Mitsou, *The atlas transition radiation tracker*, 2003.

[126] Vasiliki A. Mitsou, *Atlas silicon microstrip tracker: operation and performance*, 2011.

[127] Esma Mobs, “The CERN accelerator complex - August 2018. Complexe des accelerateurs du CERN - Aout 2018”, General Photo (2018).

[128] Vinod Nair and Geoffrey E. Hinton, “Rectified linear units improve restricted boltzmann machines”, in Proceedings of the 27th international conference on international conference on machine learning, ICML’10 (2010), pp. 807–814.

[129] Paolo Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, JHEP **11**, 040 (2004).

[130] Paolo Nason and Giulia Zanderighi, “W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2”, Eur. Phys. J. C **74**, 2702 (2014).

[131] Y. E. NESTEROV, “A method for solving the convex programming problem with convergence rate $o(1/k^2)$”, Dokl. Akad. Nauk SSSR **269**, 543–547 (1983).

[132] Richard Nisius, “On the combination of correlated estimates of a physics observable”, Eur. Phys. J. C **74**, 3004 (2014).

[133] S. F. Novaes, *Standard model: an introduction*, 2000.

[134] K.A. Olive, “Review of particle physics”, Chinese Physics C **40**, 100001 (2016).

[135] P. Bagnaia et al., “Evidence for $Z^0 \to e^+e^-$ at the CERN pp collider”, Physics Letters B **129**, 130–140 (1983).

[136] Joao Pequenao, “Computer generated image of the whole ATLAS detector”, Mar. 2008.

[137] T Poikela, J Plosila, T Westerlund, M Campbell, M De Gaspari, X Llopart, V Gromov, R Kluit, M van Beuzekom, F Zappon, V Zivkovic, C Brezina, K Desch, Y Fu, and A Kruth, “Timepix3: a 65k channel hybrid pixel readout chip with simultaneous ToA/ToT and sparse readout”, Journal of Instrumentation **9**, C05013–C05013 (2014).

[138] Bogdan Povh, Klaus Rith, Christoph Scholz, and Frank Zetsche, *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte; 9th ed.* Springer-Lehrbuch (Springer, Heidelberg, 2014).

[139] J. Pumplin et al., “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”, JHEP **07**, 012 (2002).

[140] Ning Qian, “On the momentum term in gradient descent learning algorithms”, Neural Networks **12**, 145–151 (1999).

[141] *Quantum diaries*, (Nov. 2019) `https://www.quantumdiaries.org/wp-content/uploads/2011/05/exp-magnets.png`.

[142] Emanuele Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, Eur. Phys. J. C **71**, 1547 (2011).

[143] Imma Riu, “The atlas level-1 topological trigger performance in run 2”, Journal of Physics: Conference Series **898**, 032037 (2017).

[144] Frank Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain [j]”, Psychol. Review **65**, 386–408 (1958).

[145] Sebastian Ruder, *An overview of gradient descent optimization algorithms*, 2016.

[146] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Parallel distributed processing: explorations in the microstructure of cognition, vol. 1”, in , edited by David E. Rumelhart, James L. McClelland, and CORPORATE PDP Research Group (MIT Press, Cambridge, MA, USA, 1986) Chap. Learning Internal Representations by Error Propagation, pp. 318–362.

[147] Sampsonidou, Despoina and ATLAS, “Precise determination of the muon reconstruction efficiency in atlas at run–ii”, EPJ Web Conf. **164**, 08006 (2017).

[148] Jürgen Schmidhuber, “Deep learning in neural networks: an overview”, CoRR **abs/1404.7828** (2014).

[149] M Schönherr and F Krauss, “Soft photon radiation in particle decays in sherpa”, Journal of High Energy Physics **2008**, 018–018 (2008).

[150] Matthias Schott, G Duckeck, and Matthias Obermaier, *Study of the Sagitta Resolution of MDT-Chambers*, tech. rep. ATL-MUON-PUB-2006-010. ATL-COM-MUON-2006-004 (CERN, Geneva, Mar. 2006).

[151] Holger Schulz, “Measurement of the underlying event using track-based event shapes in $Z \to \ell^+\ell^-$ events with atlas”, PhD thesis (Humboldt-Universitaet zu Berlin, Mathematisch-Naturwissenschaftliche Fakultaet, 2015).

[152] Steffen Schumann and Frank Krauss, “A parton shower algorithm based on Catani–Seymour dipole factorisation”, JHEP **03**, 038 (2008).

[153] Osvaldo Simeone, *A very brief introduction to machine learning with applications to communication systems*, 2018.

[154] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, “An introduction to PYTHIA 8.2”, Comput. Phys. Commun. **191**, 159 (2015).

[155] Andre Sopczak, Babar Ali, Thanawat Asawatavonvanich, Jakub Begera, Benedikt Bergmann, Thomas Billoud, Petr Burian, Ivan Caicedo, Davide Caforio, Erik Heijne, Josef Janecek, Claude Leroy, Petr Manek, Kazuya Mochizuki, Yesid Mora, Josef Pacik, Costa Papadatos, Michal Platkevic, Stepan Polansky, Stanislav Pospisil, Michal Suk, and Zdenek Svoboda, "Precision Luminosity of LHC Proton-Proton Collisions at 13 TeV Using Hit-Counting with TPX Pixel Devices.", IEEE Trans. Nucl. Sci. **64**, Presented at the IEEE 2016 Nuclear Science Symposium, Strasbourg, France, 915. 10 p (2017).

[156] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", Journal of Machine Learning Research **15**, 1929–1958 (2014).

[157] Rende Steerenberg and Jorg Wenninger, "Operational challenges and performance of the LHC during Run II. Operation Challenges and Performance of the LHC During Run II", THA2WD01. 5 p (2018).

[158] W.J. Stirling, private communication, Feb. 3, 2020.

[159] M. et al. Tanabashi (Particle Data Group), "Review of particle physics", Phys. Rev. D **98**, 030001 (2018).

[160] Mark Thomson, *Modern particle physics* (Cambridge University Press, 2013).

[161] K. G. Tomiwa, "The performance of diphoton primary vertex reconstruction methods in $H \to \gamma\gamma + Met$ channel of ATLAS experiment", J. Phys. Conf. Ser. **889**, 012010 (2017).

[162] A. Valassi, "Combining correlated measurements of several different physical quantities", Nucl. Instrum. Meth. A **500**, 391–405 (2003).

[163] Samuel Webb, Bernard Brickwedde, Matthias Schott, Alexey Ezhilov, Oleg Fedin, Victor Solovyev, Denis Pudzha, and Jan Kretzschmar, *Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at* $\sqrt{s} = 13\ TeV$ *with the ATLAS detector*, tech. rep. ATL-COM-PHYS-2018-305 (CERN, Geneva, Mar. 2018).