Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering

DOCTORAL THESIS

Vector Boson Scattering and ZZ Production at ATLAS Detector

Prague 2021 Ondřej Penc

Bibliographic Entry	
Author	Ing. Ondřej Penc, Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Physics
Title of Dissertation	Vector Boson Scattering and ZZ Production at ATLAS Detector
Degree Programme	Experimental Nuclear and Particle Physics
Field of Study	Nuclear Engineering
Supervisor	doc. Alexander Kupčo, Ph.D., FZU - Institute of Physics of the Czech Academy of Sciences
Academic Year	2020/2021
Number of Pages	175
Keywords	HEP, VBS, Standard Model, ATLAS, Elektro-weak interaction

Bibliografický záznam	
Autor	Ing. Ondřej Penc, České vysoké učení technické v Praze, Fakulta jaderná a fyzikálně inženýrská, katedra fyziky
Název práce	Srážky vektorových bosonů v ZZ produkci na detektoru ATLAS
Studijní program	Experimentální jaderná a částicová fyzika
Studijní obor	Jaderné inženýrství
Školitel	doc. Alexander Kupčo, Ph.D., Fyzikální ústav AV ČR, v. v. i.
Akademický rok	2020/2021
Počet stran	175
Klíčová slova	HEP, VBS, Standardní model, ATLAS, Elektroslabá interakce

Abstract

This work is focused on the search of electro-weak production of two Z bosons accompanied by a pair of jets in proton-proton collisions with the ATLAS detector at the LHC in CERN. The effort aspires to be a basis for further research of vector boson scattering phenomena with two Z bosons in the final-state decaying into charged leptons. The search exploits ATLAS proton-proton collision data corresponding to $139\,\mathrm{fb}^{-1}$ of integrated luminosity, acquired at 13 TeV center-of-mass energy in proton-proton collisions during the Run 2 period of LHC. The background-only hypothesis is excluded with statistical significance of $5.8\,\sigma$, and the fiducial $ZZjj$ EW production cross-section is measured to be $\sigma_{\mathrm{EW}} = 0.36^{+0.09}_{-0.09}(\mathrm{stat.})^{+0.05}_{-0.05}(\mathrm{syst.})\,\mathrm{fb}$, with no significant deviation from the standard model prediction. This makes the measurement, together with simultaneous independent results from the ATLAS and CMS experiments [1, 2], the first observation of electro-weak production of two Z bosons in four-lepton channel.

Abstrakt

Tato práce je zaměřena na hledání elektroslabé produkce dvou Z bosonů doprovázených dvojicí sprškek při proton-proton srážkách v detektoru ATLAS na LHC v CERNu. Pokládá základ pro další výzkum rozptylu vektorových bosonů se dvěma Z bosony v konečném stavu, rozpadajících se na nabité leptony. Pátrání po těchto jevech využívá kolizní data experimentu ATLAS se $139\,\mathrm{fb}^{-1}$ integrované luminosity, získaná při 13 TeV srážkách protonů s protony během období Run 2 na LHC. Oproti statistickému pozadí jsou tyto jevy potvrzeny se statistickou významností $5.8\,\sigma$, a účinný průřez elektro-slabé produkce $ZZjj$ je změřen jako $\sigma_{\mathrm{EW}} = 0.36^{+0.09}_{-0.09}(\mathrm{stat.})^{+0.05}_{-0.05}(\mathrm{syst.})\,\mathrm{fb}$, bez významné odchylky od predikcí standardního modelu. Toto měření je spolu s nezávislými výsledky experimentů ATLAS a CMS [1, 2] prvním pozorováním elektro-slabé produkce dvou Z bosonů ve čtyřleptonovém kanálu.

Acknowledgement

I am happy to thank Saša Kupčo for supervising me, however, since he is not my supervisor from the beginning, I have to thank also to Vojtěch Petráček, Michal Marčišovský and Václav Vrba. The most thankful I am to my colleague Stefan Richter, who helped me in the beginning of my efforts and have shared his framework with me, on top of which I could build. Saša and Michal were also extremely helpful with the proof-reading, as well as my classmate Marek Matas without whose coaching in the final stage, the document would not be complete.

Big thanks belong also to my ex-wife Petra and my son Mikuláš, with who I spent the majority of my PhD studies period. I would like to thank also to my parents and my beloved family Jan, Helena, Zdeněk, Evička, Martin, Lucka, Veronika, Florián and Hugo. Special thanks belong to my brothers, Master of Time and Master of Wind, whose curiosity and enthusiasm helped me to find mine again.

Since I have a big number of people I want to thank to, I will make it shorter by groups: I want to thank to long time family friends, my CERN family, the center colleagues, Winter School mates, class mates, Aldebaran group, Nam gang, Maratea gang, ICNFP gang, CERN dancing collider people, LTS friends, couch surfers, and all friends around the world.

Contents

List of Abbreviations

AD	Antimatter Decelerator
ALICE	A Large Ion Collider Experiment
AMI	ATLAS Meta-data Interface
ATLAS	A Toroidal LHC Apparatus
BCM	Beam Conditions Monitor
BDTs	Boosted Decision Trees
BEH	Brout-Englert-Higgs
BSM	Beyond Standard Model
BT	Barrel Toroid
CDF	The Collider Detector at Fermilab
CERN	Conseil Européen pour la Recherche Nucléaire
CLIC	The Compact Linear Collider
CMS	Compact Muon Solenoid
CNGS	CERN Neutrinos to Gran Sasso
CP	Charge Parity
CSC	Cathode Strip Chamber
CS	Central Solenoid
CTP	Central Trigger Processor
DAQ	Data-Acquisition
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DQ	Data Quality
ECT	Endcap Toroid

EFT	Effective Field Theory
EF	Event Filter
ESD	Event Summary Data
EWSB	Electro-Weak Symmetry Breaking
EW	Electro-Weak
FCR	Fake Control Region
FE	Front-end electronics
FFR	Fake Factor Region
FTF	Fast Track Finder
FT	Fast Tracking
GSW	Glashow Salam Weinberg Theory
HEP	High Energy Physics
HL-LHC	High Luminosity LHC
HLT	High Level Trigger
IBL	Insertable B-Layer
ID	Inner Detector
ILC	The International Linear Collider
IP	Interaction Point
IT	Inner Triplet
JER	Jet Energy Resolution
JES	Jet Energy Scale
JVT	Jet Vertex Tagger
L1	Level-1
LB	Luminosity Blocks
LEP	Large Electron Positron collider
LHC	Large Hadron Collider
LINAC2	Linear Accelerator 2
LO	Leading Order

LUCID	Luminosity Measurement Using Cerenkov Integrating Detector
MC	Monte Carlo
MDT	Muon Drift Tube
ME	Matrix Element
MLE	Maximum Likelihood Estimate
MS	Muon Spectrometer
MVA	Multivariate Analysis
MoEDAL	Monopole and Exotics Detector at the LHC
NLO	next-to-leading Order
NNLO	next-to-next-to-leading Order
NP	Nuisance Parameters
OR	Overlap Removal
PD	Pixel Detector
PSB	Proton Synchrotron Booster
PS	Proton Synchrotron
PT	Precision Tracking
PoI	Parameter of Interest
QCD-CR	QCD Control Region
QCD	Quantum Chromodynamics
QGC	Quartic Gauge Coupling
QGP	Quark Gluon Plasma
RF	Radio Frequency
ROB	Readout Buffer
ROC	Receiver Operating Characteristic
ROD	Readout Driver
ROS	Readout System
RPC	Resistive Plate Chamber
RoIs	Regions of Interest

SCT	Semiconductor Tracker
SFOC	Same Flavor Opposite Charge
SLAC	Stanford Linear Accelerator Center
SM	Standard model
SPEAR	Stanford Positron Electron Asymmetric Rings
SPS	Super Proton Synchrotron
SR	Signal Region
TGC	Thin Gap Chamber
TMVA	Toolkit for Multivariate Analysis
TOTEM	TOTal Elastic and diffractive cross section Measurement
TP	Topological Processor
TRT	Transition Radiation Tracker
TTVA	Track-To-Vertex Association
TileCal	Tile-Calorimeter
UV	Ultra Violet
VBS	Vector Boson Scattering
WLCG	Worldwide LHC Computing Grid
xAOD	Extended Analysis Object Data

Chapter 1

Introduction

During the first data-taking period of the Large Hadron Collider (LHC), a long time missing puzzle-piece of the Standard Model (SM) was discovered. The observation of a new scalar particle during the search for the SM Higgs boson, performed by the ATLAS and CMS experiments in 2012 [3, 4], was a huge success of both theoretical and experimental particle physics. The theory demonstrated its predictive power by proposing the Brout-Englert-Higgs (BEH) mechanism, that suggested the existence of such a particle in 1964 [5, 6] and the experiments confirmed it by the discovery of this hard-to-find particle after almost a half of century.

All the ingredients participating in the Electro-Weak Symmetry Breaking (EWSB) of SM were collected, so, further investigations of the mechanism can advance in two main ways: precise measurements of properties of the acting particles or scrutiny of the processes associated to the EWSB. Since the fundamental properties of Electro-Weak (EW) bosons were already measured by the Large Electron Positron collider (LEP) and Tevatron, the former way would be focused primarily on the Higgs boson and precise measurements in the Higgs factories such as the ILC, CLIC, and FCC-ee, as well as in the high-luminosity LHC, during Run 3 of the LHC, or in another future collider [7–10]. The latter way is a newly opening field of the EW sector of SM, since the LHC provides large enough data sample to probe even the rare EW interactions.

Self-interaction between EWK intermediate vector bosons is directly connected with BEH mechanism and EWSB [11]. It is the existence of the Higgs boson that complements the rest of the SM in such a way, that it cancels out unphysical divergences in vector boson scattering (VBS) and restores unitarity that would otherwise be violated already at the energy scale of the LHC. Importance of the VBS processes lies also in the fact that some of the theories beyond the Standard Model (BSM) such as Supersymmetry [12], Composite Higgs [13], and Little Higgs [14] include mechanisms that make VBS differ from SM predictions at very high energies (over 100 TeV). Precise measurements of VBS cross-sections enable

not only a verification of the SM, but are also suitable for a BSM phenomena search.

The VBS processes in proton-proton collisions are accessible only through a measurement of EW production of the bosons accompanied by two jets. The EW production is a superior class of processes to VBS, where the VBS plays a significant role, sharing the importance and sensitivity to the phenomenology of both SM and BSM physics. In the final-state of processes like these, the jets play an important role for the VBS tagging. The two hadronic jets are remnants of quarks that were kicked-out of a proton after the emission of heavy vector bosons and are completing the typical signature of VBS processes. One of the most rare channels is the four-lepton $ZZjj$ VBS channel, too rare to be visible even in full Run 2 LHC data. Hence this work is focused on a measurement of $ZZjj$ EW production, to which the VBS is a part. This process has never been measured before, so, this thesis (and the related work published in reviewed journals) presents the first complete study of the $ZZjj$ EW production. The CMS collaboration carried out an independent study of the production with $139\,\mathrm{fb}^{-1}$ of data, which claims a four σ observation as well.

Similar studies have been carried out at the LHC with different vector bosons in a fully leptonic channel such as the $W^{\pm}W^{\pm}jj$ and $W^{\pm}Zjj$ processes, which found no deviations from the SM to this date [15, 16] and [17, 18]. The semileptonic channels $VVjj$ (V denotes Z or W boson) with one boson decaying to two jets still wait for evidence [19].

This work studies the four-lepton $ZZjj$ EW channel at unprecedented 13 TeV center-of-mass energy. Even though this channel suffers from a small cross-section, it provides a very clean final-state, with two Z bosons decaying into four leptons, excluding the semileptonic final-state with neutrinos. The main goal is the extraction of EW production significance over background accompanied by a measurement of inclusive production cross-section in VBS enhanced phase space. The measurement exploits data corresponding to $139\,\mathrm{fb}^{-1}$ of the integrated luminosity, acquired between years 2015-2018 in Run 2 of the LHC.

This thesis is closely related to publications [1, 20] where the author has significantly contributed. In the work [20], the author and others have analyzed the inclusive ZZ production in four-lepton channel at the ATLAS detector over the years 2015 and 2016 with $36.1\,\mathrm{fb}^{-1}$ of data. A differential cross-section for this process was measured as a function of 20 different variables which have theoretical and phenomenological implications since they are sensitive to the QCD nature of the process, parton distribution functions of the proton, as well as initial and final-state radiation effects. The author contributed to the area of jet analysis, acquired important experiences to be able to exploit them towards the goal of his work, the search for EW $ZZjj$ production. This work was published on behalf of

the ATLAS collaboration in the Physical Review D journal.

In the work [1], observation of the rare EW production of two Z bosons and a pair of jets in the four-lepton channel is presented. The author contributed primarily to the background estimate efforts. The results presented therein exclude the hypothesis of no EW production with a statistical significance of 5.5 σ and are consistent with SM predictions. This paper was submitted for publication in the Nature Physics journal and is currently in the process of being reviewed. The results were also presented at the conference ICHEP2020 in Prague. The analysis is repeated in this document, borrowing some of the intermediate results from the ATLAS collaboration analysis [1]. The result is replicated and even brought some new information. The presented work is a stand-alone effort using a comprehensive and independent framework developed by the author. Based on the framework used for the inclusive ZZ production measurement.

This document is divided into several parts. The measurement motivation is briefly discussed in Chapter 2, explaining the relation of VBS and EW production. The experimental setup is discussed in Chapter 3, focusing on the LHC as well as on the ATLAS detector. The analysis work itself starts with data and Monte Carlo samples, as the input to the measurement in Chapter 4 followed by Chapter 5 discussing the analysis workflow and its origin in the inclusive ZZ production. The executive part is started with an event selection in Chapter 6, followed by a background estimation in Chapter 7, and completed with a systematic uncertainties scrutiny in Chapter 8. The final have two parts, a cross-section measurement of inclusive $ZZjj$ production in Chapter 10 and an extraction of EW signal strength in Chapter 9.

The main analysis results extend the knowledge of SM, and the main analysis task addresses specifically the EW sector. However, the presented document is primarily a proof that the author mastered all important aspects of the analysis process and is eligible to lead a similar analysis belonging to the efforts in the EW sector of the SM and BSM searches.

Chapter 2

Motivation

An in-depth study of scattering of electro-weak (EW) massive gauge bosons, called Vector Boson Scattering (VBS), has several motivations. Brief discussion of some will be introduced here as an invitation to the Standard Model (SM) section, where corresponding theory fundamentals are discussed.

The mass of vector bosons implies a longitudinal polarization which cause a divergence of cross-section at high energies, if not compensated. The scattering realization via a Higgs boson exchange is significant in the VBS processes, playing the role of the compensator. And finally, all bosons are incorporated into the SM via an Electro-Weak Symmetry Breaking (EWSB) mechanism tailored to fit the non-Abelian feature of the SM.

The VBS can therefore be seen as a complex test of the SM, from the EWSB and non-Abelian perspective. Moreover, the VBS has a direct contribution of a four-vertex interaction, which becomes accessible.

2.1 Standard Model

To shed more light to the origin of the VBS and motivation of its measurement, necessary fundamentals of the SM will be described. This chapter introduces the relevant parts of the model as a theoretical description of the EW interaction and EWSB, while leaving the strong interaction details and phenomenological classification of the elementary particles to the reader, recommending to have a look for example in [21]. In the following text, the speed of light c and reduced Planck constant $\hbar$ are equal to one.

The SM Lagrangian is composed of several parts, describing the behavior of nature, so far, up to the energies achievable at the LHC:

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{QCD}}, \tag{2.1}$$

where the first two terms describe the EW interaction and EWSB mechanism introducing the Higgs boson, while the third and fourth describes the fermion mass generation through a Yukawa coupling and the strong interaction using quantum chromodynamics (QCD) theory. The subsequent sections briefly describe the EW unification, genesis of the EW gauge bosons, a way how the bosons acquire mass, and the derivation of the boson self-interaction and interaction with the Higgs field. Finally, the fermion mass generation by the Yukawa coupling is mentioned at the end. For a more complete description of the aforementioned, see [22].

2.1.1 Electroweak Theory

The EW theory was developed between years 1961 and 1969 [23–25] by Glashow, Salam, and Weinberg (GSW theory). As a result, the electromagnetic and weak interactions are unified exploiting the $SU(2)_{T_3}$ and $U(1)_Q$ symmetry groups into a new group $SU(2)_L \times U(1)_Y$ introducing a weak hypercharge

$$Y = Q - T_3, \tag{2.2}$$

where Q denotes a particle charge and T_3 a weak isospin. Fermion currents of the unified interaction couple only to the left-handed particles (denoted by L), which are arranged into $SU(2)$ doublets with the third component of the weak isospin $T_3 = \pm 1/2$. The right-handed fermions form $SU(2)$ singlets with $T_3 = 0$. The electroweak Lagrangian has a kinetic term, describing the gauge fields only; and an interaction term, describing their interaction with fermions, following a form

$$\mathcal{L}_{\mathrm{EW}} = \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{fermion}} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \sum_{f_L, f_R} \bar{\psi}_{f_{L,R}} i\gamma^\mu D_\mu^{f_{L,R}} \psi_{f_{L,R}} \tag{2.3}$$

with a covariant derivative

$$D_\mu^{f_{L,R}} = \partial_\mu - igA_\mu^a \frac{\tau^a}{2} - ig' Y_{L,R}^{(f)} B_\mu, \tag{2.4}$$

where the gauge part of the Lagrangian can be decomposed using

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon^{abc} A^b_\mu A^c_\nu \tag{2.5}$$

and

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu,$$

where γ and τ denote the Dirac and Pauli matrices. The triplet A^a_μ $(a = 1, 2, 3)$ denotes Yang–Mills fields corresponding to the weak isospin subgroup $SU(2)$, and B_μ represents a gauge field associated to the weak hypercharge subgroup $U(1)$. If the covariant derivative acts on a right-handed singlet (ψ_{f_R}), the non-Abelian part naturally disappears, since the $SU(2)$ generators τ^a are trivial. The left-handed doublets ψ_{f_L} are formed as

$$\psi_{\ell_L} = \begin{pmatrix} \nu^\ell_L \\ \ell_L \end{pmatrix}, \tag{2.6}$$

deliberately placing the neutrino to the top, which ensures the doublet constituents have the same hypercharge $Y_L = -0.5$, signaling they belong to the same lepton current. The ℓ denotes the lepton families, however the doublet can be formed of quarks as well, in the same manner placing the family member with higher charge on top. The coupling constants g and g', introduced in the Lagrangian through the covariant derivative (2.4), are related according to

$$g \sin\theta_W = g' \cos\theta_W = e \tag{2.7}$$

with the weak mixing angle θ_W and the electromagnetic coupling constant e. The A^a_μ and B_μ are rather abstract gauge fields without an association to real particles. However, the physical meaning can be revealed after application of an appropriate transformation based on linear combinations of the fields. The charged fields $W^\pm_\mu$ are defined by

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(A^1_\mu \mp A^2_\mu), \tag{2.8}$$

representing weak charged currents. A neutral weak current is represented by Z_μ and together with electromagnetic boson A_μ are obtained via an orthogonal transformation

$$\begin{aligned} B_\mu &= \cos\theta_W A_\mu - \sin\theta_W Z_\mu, \\ A^3_\mu &= \sin\theta_W A_\mu + \cos\theta_W Z_\mu. \end{aligned} \tag{2.9}$$

The last transformation recovers the massless photon of electromagnetic interaction at first glance hidden in the Lagrangian (2.3). The weak interaction is handled by three newly born real bosons $W^\pm$ and Z. The EW unification manifests itself already by a formulation of Equation (2.7). Two artificially added couplings g and g' relate to each other, and furthermore, to the coupling e. This in principle

represents the EW unification and birth of a new EW interaction with g coupling, denoted as α_{EW}.

Once the EW bosons are established, an investigation on their interactions is desirable. The gauge part of the electroweak Lagrangian (2.3) can be decomposed firstly to non-diagonal part choosing τ^1 and τ^2, which leads to a derivation of the charged current weak interaction Lagrangian, and the diagonal part with the third Pauli matrix and the Abelian part of covariant derivative. The diagonal part, after the transformation (2.9), leads to the neutral current weak interaction and electromagnetic interaction Lagrangians. The former is directly involved in the VBS via exchange of a Z boson.

The interactions of massive vector bosons are of a special interest in the VBS phenomenology since they define the scattering realizations. The corresponding Lagrangians are derived from the gauge part of the (2.3) after a decomposition using (2.5) and exclusion of the kinetic terms. The decomposition comes from a fact that the gauge field tensor can be expressed in terms of the commutator of covariant derivatives, namely

$$-igF_{\mu\nu} = [D_\mu, D_\nu].$$

Finally, the EW interaction couplings of the $W^\pm$, Z^0, and photon get a shape of

$$\begin{aligned}
\mathcal{L}_{WW\gamma} &= -ie(A_\mu W^-_\nu \overleftrightarrow{\partial^\mu} W^{+\mu} + W^-_\mu W^+_\nu \overleftrightarrow{\partial^\mu} A^\nu + W^+_\mu A_\nu \overleftrightarrow{\partial^\mu} W^{-\nu}),\\
\mathcal{L}_{WWZ} &= -ig\cos\theta_{\mathrm{W}}(Z_\mu W^-_\nu \overleftrightarrow{\partial^\mu} W^{+\mu} + W^-_\mu W^+_\nu \overleftrightarrow{\partial^\mu} Z^\nu + W^+_\mu Z_\nu \overleftrightarrow{\partial^\mu} W^{-\nu}),\\
\mathcal{L}_{WW\gamma\gamma} &= -e^2(W^-_\mu W^{+\mu} A_\nu A^\nu - W^-_\mu A^\mu W^+_\nu A^\nu),\\
\mathcal{L}_{WWZZ} &= -g^2\cos^2\theta_{\mathrm{W}}(W^-_\mu W^{+\mu} Z_\nu Z^\nu - W^-_\mu Z^\mu W^+_\nu Z^\nu),\\
\mathcal{L}_{WWWW} &= \frac{1}{2}g^2(W^-_\mu W^{-\mu} W^+_\nu W^{+\nu} - W^-_\mu W^{+\mu} W^-_\nu W^{+\nu}),\\
\mathcal{L}_{WWZ\gamma} &= g^2\sin\theta_{\mathrm{W}}\cos\theta_{\mathrm{W}}\times\\
&\quad (-2W^-_\mu W^{+\mu} A_\nu Z^\nu + W^-_\mu Z^\mu W^+_\nu A^\nu + W^-_\mu A^\mu W^+_\nu Z^\nu),
\end{aligned} \tag{2.10}$$

where the symbol $\overleftrightarrow{\partial}$ is defined by $f(\partial g) - (\partial f)g$. The terms in brackets define the vertex shape in sense of the Feynman diagrams, while the coefficient expresses the coupling strength.

The briefly sketched theory of the EW boson interactions, reveals a very interesting SM phenomena connected to the VBS, the electro-weak three- and four-vertices or triple and quartic gauge couplings (QGC). The triple gauge couplings were already measured at the Large Electron Positron collider [26–28] together with the QGC involving two photons ($WW\gamma\gamma$) [29, 30], as the first probe of the phenomenon. However, the couplings composed entirely of the EW bosons, typical for the VBS, start to be accessible only at the LHC, namely $WWWW$ and

$WWZZ$. The remaining $WWZ\gamma$ coupling also deserves attention, since an off-shell Z boson is experimentally indistinguishable from an off-shell γ particle. This is the reason why the $W\gamma$ and $Z\gamma$ are sometimes also considered as VBS channels. Since the VBS channels include a contribution of various types of couplings, it makes the scattering measurement a very sensitive test of the SM. Especially, when taking the couplings into account as a compensation of a divergent behavior of the intermediate vector boson (IVB) model in the high energy region.

The high energy misbehavior of Z and $W^{\pm}$ bosons is tightly connected with their non-zero mass implying a longitudinal polarization component. The mass of weak gauge bosons is directly generated by employing the second part part of the SM Lagrangian, $\mathcal{L}_{\text{Higgs}}$.

2.1.2 Electro-Weak Symmetry Breaking

The EW boson mass terms can not be added to the model simply by hand, since the gauge invariance as well as the renormalizability would be violated. The mass is acquired by adoption of a cognac bottle (or Goldstone) potential into the $SU(2) \times U(1)$ gauge theory. The mechanism was proposed by Brout, Englert, and Higgs in the 1960s [5, 6], called BEH or simply Higgs mechanism. The mechanism is tailored for the mass generation of three weak bosons and at least one massive scalar boson. Thus, it is clear that one has to start with at least four real scalar fields, from which three will become the so called Goldstone bosons which are to be "eaten" by the gauge bosons, as is spoken in a common physics "folklore". The fourth scalar turns into the Higgs boson. The four fields must obey the constraints of the $SU(2)$ symmetry, which can be done by coupling of scalars in complex fields and mounting them in a doublet representation. This rather a non-trivial composition leads to a weak isodoublet

$$\Phi = \begin{pmatrix} \varphi^{+} \\ \varphi^{0} \end{pmatrix} = \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix}, \tag{2.11}$$

where the upper pair of fields, denoted by a plus superscript, represents the fields carrying a positive charge and the lower neutral fields, following the same prescription as in Equation (2.6). Since the $SU(2)$ doublet corresponds to the weak isospin, together with the charge it determines the value of the weak hypercharge $Y_\Phi = 1/2$, which is associated to the $U(1)$ subgroup, according to the origin of Equation (2.2). The Higgs Lagrangian is based on the Klein-Gordon Lagrangian endowed with the $SU(2) \times U(1)$ covariant derivative (2.4) with $Y_L = -Y_\Phi$ in a form

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^{\dagger}(D^\mu \Phi) - V(\Phi), \tag{2.12}$$

with a potential V given by

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \tag{2.13}$$

where the scalar parameters μ and λ determine a shape of the potential as well as a vacuum expectation value v and the Higgs boson mass

$$v = \frac{\mu}{\sqrt{\lambda}} \text{ and } m_{\mathrm{H}}^2 = 2\lambda v^2. \tag{2.14}$$

The first step towards the generation of weak gauge boson masses is an application of the U-gauge fixation

$$\Phi(x) \to \Phi_U(x) = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H(x)) \end{pmatrix}, \tag{2.15}$$

which also defines $H(x)$, the Higgs field. In short, the fixation embodies three important steps:

1. reparametrization of the Φ doublet in a complex plane,
2. exploitation of a degree of freedom, coming from the Lagrangian invariance under local phase transformation, i.e. a choice of the phase to eliminate completely the angular fields corresponding to the Goldstone bosons,
3. shifting of the radial field to the local minimum (definition of v).

After some manipulations, the Lagrangian $\mathcal{L}_{Higgs}$ comprises a Higgs field kinetic, self-interaction, and mass terms. The remaining terms can be divided into two groups: the terms that do not consist of the Higgs field at all will give a rise of the massive weak gauge fields, the other terms govern the Higgs boson interactions with the electroweak gauge bosons.

After selection of terms without Higgs field H and application of the usual transformations as before (2.8) and (2.9), one is left with massive weak bosons

$$\begin{aligned} m_W &= \frac{1}{2} g v, \\ m_Z &= \frac{1}{2} (g^2 + g'^2)^{1/2} v. \end{aligned} \tag{2.16}$$

When the terms that consist of weak gauge and Higgs fields together are chosen, one can determine the interactions of W, Z, and Higgs bosons as

$$\mathcal{L}_{WWH} = g m_W W_\mu^- W^{+\mu} H,$$
$$\mathcal{L}_{ZZH} = \frac{g m_Z}{2\cos^2\theta_W} Z_\mu Z^\mu H. \tag{2.17}$$

The described mechanism is a minimal model of the EWSB as a part of the SM of particle physics. The symmetry breaking is in fact caused by a non-zero vacuum expectation value, caused by an unusual shape of the Goldstone potential. It is useful to emphasize that the weak gauge bosons acquire their mass thanks to the Goldstone bosons, massless unphysical bosonic fields originating from the complex reparametrization of the scalar isodoublet. The number of bosons is equal to the number of broken symmetry generators (corresponding to τ_1, τ_2, and τ_3), therefore since we have three massive gauge bosons, after invoking the BEH mechanism, three Goldstone bosons are spent for the mass generation, while leaving one scalar, the Higgs boson. The act of spending of the Goldstone bosons is artistically depicted by the author in Figure 2.1.

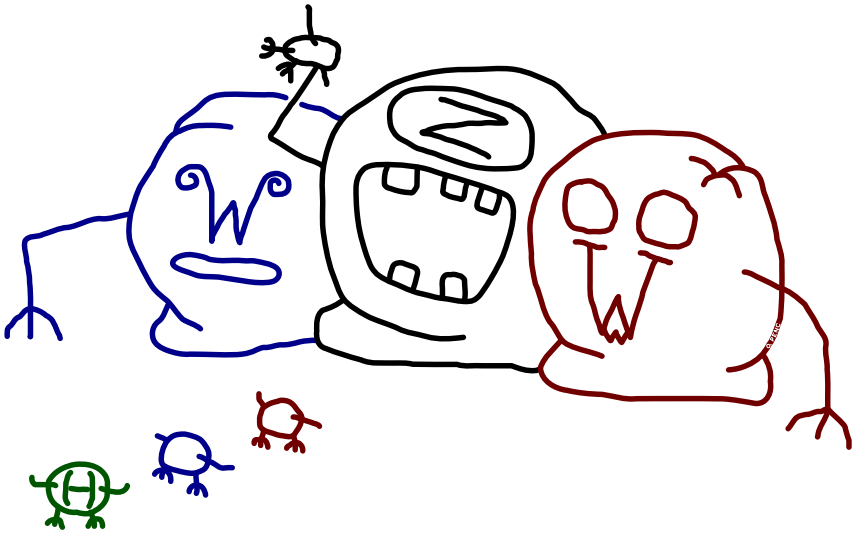


Figure 2.1: Artistic illustration of the act of spending of Goldstone bosons when three EW gauge bosons become heavy leaving one massive scalar field as a consequence of EWSB.

This section explains the origin of the EW gauge bosons masses their self-interaction and interaction with the Higgs boson. All this is necessary since this document study the EW VBS and touches the EWSB mechanism. All three massive electro-weak bosons W^+, W^-, and Z^0were incorporated into the SM by adoption of a non-Abelian gauge group $SU(2)$ along with an Abelian $U(1)$ into a Dirac

equation. A measurement of electro-weak boson interactions in VBS therefore probes the gauge structure and non-Abelian feature of the SM.

2.1.3 Mass Generation

Introduction of the $SU(2) \times U(1)$ symmetry to the SM has another consequence, the traditional mass term of Dirac equation can not be simply added due to the fact that the left-handed components form doublets while the right-handed are singlets which would violate the symmetry. The fermion masses have to be generated through interactions involving Higgs doublet employing a Yukawa-type coupling. Firstly invented by Yukawa for a description of interactions of nucleons in 1935 [31] (which later led to the prediction of a π meson). It can be said the mass of Higgs boson itself is generated the same way through the self-interaction terms. These facts amazingly close the circle and define the Higgs field interaction as a phenomenon generating the mass of all SM particles. For details on Yukawa couplings see Reference [32].

2.2 Vector Boson Scattering

The term Vector Boson Scattering or VBS is adopted from literature [33] and will be used as an abbreviation of massive EW gauge boson scattering throughout the document. The VBS channels are categorized according to the final-state bosons after the scattering.

The first examined VBS channel at the LHC was a same-sign W pair ($W^{\pm}W^{\pm}$) in a leptonic final-state, conducted as a search for an EW production of the bosons in proton-proton collisions. How the EW production and VBS are related is a subject of the following text. Nevertheless, the channel was celebrated for the first evidence of such production ever, in terms of a 3-σ statistical significance of the signal. The measurement was possible thanks to the same-sign requirement which suppresses a natural VBS background. The analysis was carried out by both ATLAS and CMS collaborations in the first years of LHC operation in 2012 [34, 35] at 8 TeV center-of-mass energy with 20 fb^{-1} of data.

2.2.1 ZZ Scattering

The EW vertices introduced in Equations 2.10 give a composition of all possible realization of VBS. The process at a tree level includes Feynman diagrams fundamentally divided into categories, involving

- triple gauge boson vertices in the s, t, and u channels,

- Higgs boson exchanges,
- and quartic gauge boson vertex,

using the Mandelstam notation. The SM does not allow the s channel realization of the VBS without the Higgs boson exchange in the ZZ VBS channel, due to an absent EW neutral coupling in Equations 2.10. The t channel realization can be mediated with a W boson, when initiated by a pair of Ws, or by the Higgs exchange. The quartic gauge coupling QGC contributes in a form of $WWZZ$. All diagrams in the Mandelstam channels are depicted in Figure 2.2 together with the diagrams of Higgs boson exchange, and QGC.

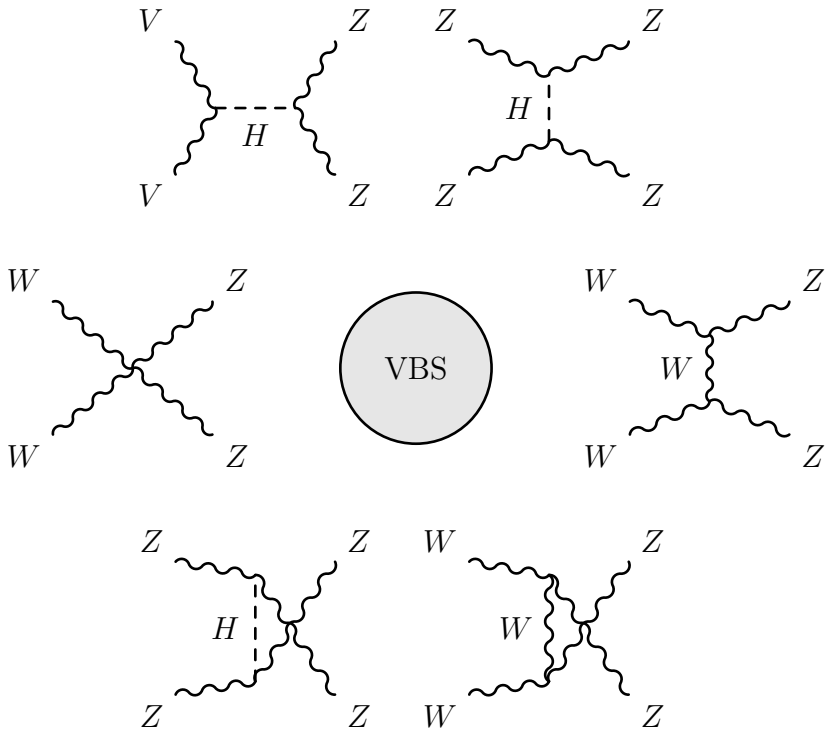


Figure 2.2: Leading order Feynman diagrams of the VBS contributing to the ZZ channel (V denotes W or Z). Any diagram can be inserted into the "VBS bubble" into Figure 2.3.

At the LHC, the scattering must be ignited via interaction of protons. The collision must be hard enough (with large transfer of momentum between proton constituents) to produce the vector bosons as mediators of the interaction. This requirement together with the small cross-section of inclusive ZZ production in proton-proton collisions and missing s channel in the ZZ VBS diagrams can illustrate why the phenomenon is so rare.

A leading order diagram representing a VBS of bosons V (denoting either W or Z) in the proton-proton collisions is depicted in Figure 2.3. Two quarks from colliding protons radiate two massive EW bosons to initiate the VBS. To get the ZZ channel diagrams one can simply replace the "VBS bubble" with any diagram from Figure 2.2.

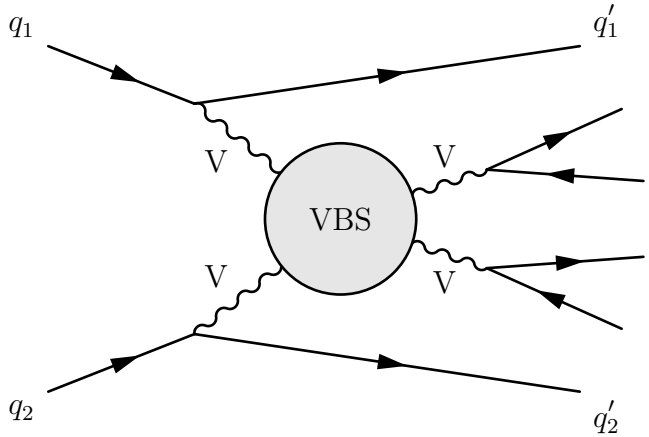


Figure 2.3: Feynman diagram of VBS in proton-proton collisions. The gray "VBS bubble" stands for any leading-order diagram with compatible external lines, i.e. diagrams shown in Figure 2.2.

The scattered quarks subsequently hadronize creating jets, an important signature crucial for VBS identification. The tagging jets are coming directly from the scattered protons, hence are expected to carry a significant amount of momentum directed to a forward kinematic region of the detector. Overall, the di-jet system is expected to be of large invariant mass with well separated constituents. In general, large mass and rapidity separation of the di-jet is a typical signature of any VBS process.

To complete the picture of VBS with other channels than the ZZ, one may consider putting diagrams from Figure 2.4 to the "VBS bubble" in Figure 2.3, as a representative sample of the WW and WZ channels.

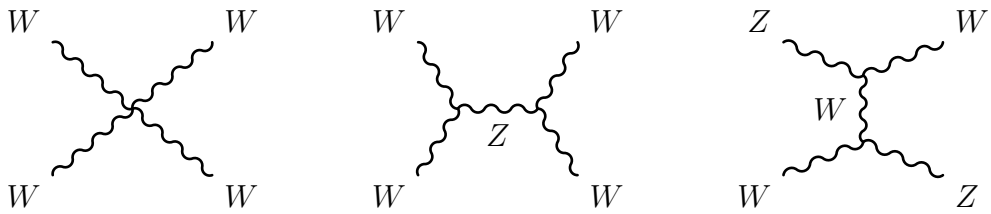


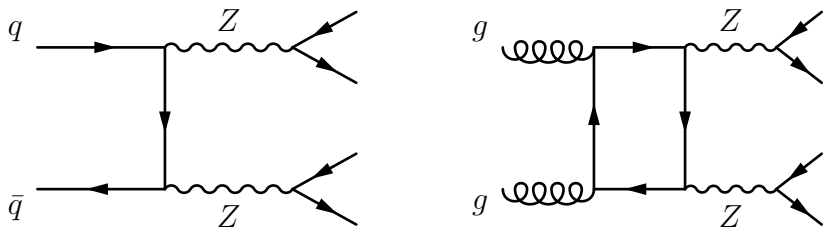
Figure 2.4: Representative set of the VBS Feynman diagrams contributing to the other channels besides the ZZ.

Since the VBS is a very rare process, it is logical to start with a more general set

of processes. An inclusive ZZ production is already well measured, representing a good starting point for understanding of the VBS as well as its natural background.

2.3 ZZ Production

The inclusive ZZ production comprises of many processes in which the VBS is hidden. The following text breaks down the composition and describes the usual categorization of the contributing processes. Several mechanisms deliver two Z bosons in the final-state, although only two dominant for proton-proton collision production will be mentioned for now. The first one is based on $q\bar{q}$ interactions, further called quark-induced production, while the second one is based on gg interaction with a quark loop as an intermediate step, called gluon-induced production. A diagram representing the quark-induced production is depicted in Figure 2.5a. It is a leading order (LO) process, therefore it dominates over the gluon-induced production shown in Figure 2.5b, which must employ the loop.



(a) **(b)**

Figure 2.5: Diagrams of ZZ production dominating in proton-proton collisions.

The processes depicted in Figure 2.5 do not contain interaction of the bosons, thus represents a natural VBS background. Since the quarks and gluons are precursors of the final-state boson production, the processes are conceptually called QCD production or QCD background, each sub-process is accordingly called quark- and gluon-induced background.

Once the VBS tagging jets are added to the final-state, according to the diagram in Figure 2.3, it can be said the ZZ production is primed towards the VBS.

2.4 $ZZjj$ Production

The $ZZjj$ production is still dominated by the QCD background, however the two-jets requirement with the VBS signature discussed in Section 2.2.1 makes

accessible a new set of processes called EW production. The set includes also the VBS, which is far from being distinguishable using the current available data sample. Therefore, the VBS efforts are focused on searches for EW production in different channels.

The EW production is often characterized by a notation of counting the power of EW coupling at the level of LO matrix element (ME). Hereby, sixth order of EW and zeroth order of QCD coupling in the ME is necessary for the EW production, while the QCD production demands only fourth order of EW and second order of QCD coupling. The processes shown in Figures 2.3 and 2.5 are in accordance with the orders of couplings, which justifies the notation and association of the VBS to the EW production. Example diagrams for non-VBS EW and QCD productions in $ZZjj$ final-state are shown in Figures 2.6a and 2.6b.

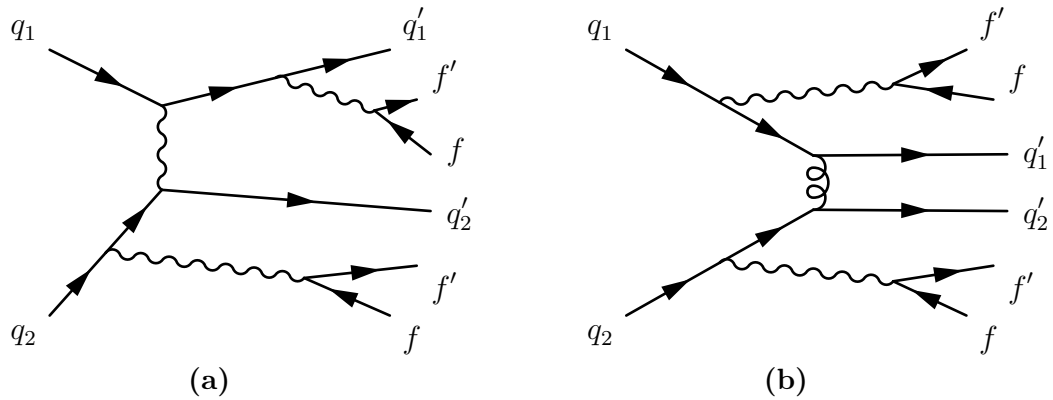


Figure 2.6: Typical representants of the non-VBS EW production (a), with an order of six of EW coupling, and the QCD production (b).

The VBS channels can be categorized by an EW-QCD ratio. For example the $W^{\pm}W^{\pm}jj$ channel has the ratio larger than one thanks to the same-sign requirement, which does not allow the bosons to be born from a single fermion current as can be seen in Figure 2.5a with Z bosons. This channel has the largest ratio among all, which explains why it is the first observed EW production channel. In comparison, the presented $ZZjj$ channel has the ratio below one, which makes it more challenging from the statistical population perspective. The trick with same-sign requirement can not be applied to channel with Z bosons for obvious reasons.

A pure VBS process is almost impossible to distinguish from other processes of EW production at the moment. A limited data collection does not allow to scrutinize the observable distributions which would help to further select the events. Such selection would have to employ finer selection criteria such as a jet-boson isolation or boson centrality.

A search for the $ZZjj$ EW production is the objective of this analysis, representing one of the first steps towards the future VBS studies. The EW production, as the superior class of processes to the VBS, share the aforementioned sensitivity to a possible divergence from the SM, therefore the $ZZjj$ EW production measurement is still a very good test of the SM.

2.5 VBS Research

The ZZ VBS channel was probed firstly by a search for $ZZjj$ EW production by CMS collaboration [36] using $36\,\text{fb}^{-1}$ of data, unfortunately without a significant success. The production was observed with a statistical significance of over 5-σ for the first time by the ATLAS [1] and over 4-σ by CMS [2] collaborations exploiting $139\,\text{fb}^{-1}$ of data at 13 TeV center-of-mass energy.

From historical point of view, the ZZ EW production was studied first in a $\gamma\gamma$ production in the LEP era in 1993 in [37] and further discussed in the light of the discovered Higgs boson in 2013 [38]. Monte Carlo (MC) simulated prediction of the $ZZjj$ EW production was discussed in [39].

Since the first evidence of the EW production of two massive vector bosons in the proton-proton collisions, the three-σ observation in the $W^{\pm}W^{\pm}jj$ leptonic channel in 2012, the VBS efforts flourish. Especially during the second operation period of the LHC, in which also this work was conducted. The $W^{\pm}W^{\pm}jj$ success was extended to a five-σ observation [15, 16] at 13 TeV energy with $36\,\text{fb}^{-1}$ of data. A $W^{\pm}Zjj$ EW leptonic channel celebrated a five-σ observation at the same time, exploiting the same data collected by the ATLAS [17]. The CMS reported the observation with an extended data collection afterwards [18]. The latter reference also includes a differential cross-section measurement in $W^{\pm}W^{\pm}jj$ channel. A semileptonic VV channel, considering the second V decaying hadronically, was conducted by ATLAS collaboration [19], unfortunately still waiting for at least three-sigma evidence. A complementary leptonic $W^{\pm}W^{\mp}jj$ channel has not been studied yet. The ATLAS results are in detail summarized by the author in [40]. Up to this date, no EW production searches have announced a deviation from the SM expectations.

2.6 Beyond the Standard Model

In general, all pioneer theories must be confronted with observed data. Also the GSW theoretical approach needed a confirmation from the experimental side. Especially because G. 't Hooft in 1971 proved a renormalizability of massless Yang-Mills fields as a general result and triggered a wave of construction of new models

based on non-Abelian gauge groups, from which the right one has to be chosen. The successful search for muon-neutrino scattering confirmed the existence of neutral currents by their first observation in 1973 and brought a firm base of the current SM to the fore. The discovery of the $W^{\pm}$ and Z^{0}bosons in 1983 then confirmed the SM as a realistic description of our universe. And the Higgs boson finally completed the whole picture in 2012. Nevertheless, the so far complete zoo of the particles do not imply there are no other undiscovered or hidden particles and do not prove the EWSB mechanism. Rephrased, there is no reason to believe the EWSB is realized in the minimal way, as is implemented in the SM. Moreover, very interesting phenomena like the dark matter, dark energy, or imbalance between matter and anti-matter in the universe [41] have so far no acceptable explanation. Therefore, it is necessary to study phenomena with a potential of pointing beyond the SM (BSM).

Many models of BSM physics can be considered as an alternative in the EWSB sector. Since the BSM physics is out of scope of this work, they will be only listed here as a reference for further studies: a composite Higgs boson model [42], pseudo-Nambu-Goldstone generation of the Higgs boson [43], an additional complex singlet to the SM [44], or a two Higgs doublet model [45], and the minimal Supersymmetric standard model itself [46]. The EWSB is directly connected to the VBS, as is shown in Section 2.1, therefore the listed models predict deviations from the SM in the yields of VBS and EW production.

A rather large number of alternative models offers many possibilities how to test the BSM physics. Evaluation of every BSM model by a comparison of its predictions to real data, a so called top-down approach, is therefore quite challenging. In the VBS efforts, it is common to use a model independent approach either based on anomalous couplings or more general Effective Field Theory, expanding the SM with effects of higher-dimensional operators suitable for SM fields [47, 48].

Nonetheless, the aim of this work is rather an observation of anticipated phenomena, then its interpretation. This probe will be followed by subsequent BSM efforts and if consistent with the SM prediction by limit settings of anomalous phenomena.

Chapter 3

Experiment

3.1 Introduction

A precise description of experimental settings is necessary to enable its repetition by other scientists, in order to confirm an observation. The description necessary for the presented analysis is possible only in a very condensed manner due to its enormous complexity and limited scope of this document. Therefore, it will be focused on the significant and interesting aspects with regard to the presented analysis, encouraging the reader to refer to the original documents, describing the Large Hadron Collider (LHC) [49] and ATLAS experiment [50], and additional documents describing the details, throughout the text. The project of proton-proton collision research at CERN is extremely unique outcome of the scientific cooperation which already cost a tremendous amount of person-power and financial support from the nations around the world. So, the possibility of repetition of the experiment is not only a question of its precise description.

As in all branches of human efforts, also in physics a certain level of competition exists. It is usually constructive and the physics field profits from it, for example when different experiments observe the same phenomenon, utilizing distinct physical detection principle at best, or simply replicating an experiment at another place. In the land of High Energy Physics (HEP), a famous example is the discovery of the J/ψ particle in Brookhaven National Laboratory in a proton-proton fixed target experiment and at the SLAC in the e^+e^-collider SPEAR simultaneously.

A priori doubling of experiments in the HEP is driven by a fact the accelerating machines are of huge proportions, so it is convenient to exploit the whole size of a machine by employing more experiments, not mentioning its enormous financial requirements. Examples from the collider physics efforts are: the experiments Underground Area 1 and 2 (UA1 and UA2), placed on the accelerator Super proton-antiproton Synchrotron (Sp$\bar{\text{p}}$S), which observed the W and Z bosons for

the first time simultaneously in 1983 [51, 52]; the experiments Delphi and Aleph, installed on the Large Electron Positron accelerator (LEP) [53], which for the first time observed the multiboson production (W pairs [54], Z pairs [55]) and did the first measurements of the triple and quartic gauge couplings [56]; and at last the experiments CDF and D0, stringed on the $p\bar{p}$ accelerator Tevatron [57], which discovered the top quark in 1995 [58, 59] and for the first time observed the QCD-initiated multiboson production [60, 61].

3.2 Large Hadron Collider

The LHC has several meanings, from the particle physics point of view it is just a particle collider, while from the accelerator physics perspective it is an accelerator as well as a storage ring. In the analysis part we will call it a collider, in the LHC description section rather a machine, a storage ring, or just a ring, because of the accelerating cavities make only a fraction of the ring body.

Primary function of the LHC is a research of proton-proton collisions, therefore the most important parameters of the machine are the center-of-mass energy of the collision $\sqrt{s}$, using Mandelstam variable, and luminosity $\mathcal{L}$. The latter expresses the machine capability to bring protons in a collision, representing an instantaneous amount of deliverable data in equation

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma_{\text{p-p}}\mathcal{L}(t). \tag{3.1}$$

where N is a number of observed events in the collisions and $\sigma_{\text{p-p}}$ is the proton-proton cross-section. While the energy parameter can represent the immediate discovery potential (J/ψ, W, Z), the luminosity is rather a key for the understanding of hidden and more complex physical mechanisms such as EWSB.

The LHC has started its operation on 18 September 2008 and already has a colorful operation history, divided into several periods: a commissioning period from 2008 to 2009, during which the LHC become the world's number one collider but also had to face childbirth troubles; a period from 2010 to 2012 called Run 1 (Run 1), when the paramount physics results were achieved [3, 4]; a period from 2013 to 2014 called Long Shutdown 1 (LS 1) dedicated to maintenance and moderate upgrade of the machine and its experiments; a period from 2015 to 2018 called Run 2 (Run 2), the second data taking period from which the presented analysis benefits; and since 2019 the Long Shutdown 2 (LS 2), a period for further maintenance and upgrades. The LHC will soon enter Run 3 period after which a major upgrade of the machine will take place resulting in a High Luminosity LHC (HL-LHC) [10]. The periods are summarized in Table 3.1 including delivered

data at various energies using integrated luminosity defined by the equation

$$\mathcal{L}_{\text{int.}} = \int \mathcal{L}(t)\mathrm{d}t. \tag{3.2}$$

Table 3.1: LHC operation periods in time

Years	Period	Description
2008-2009	Commissioning	Technical authorization of the LHC machine and detectors.
2010-2012	Run 1	Physics data-taking, $\sqrt{s}$ of 7 and 8 TeV, $\mathcal{L}_{\text{delivered}} = 28\,\text{fb}^{-1}$.
2013-2014	LS 1	LHC upgrade towards 13 TeV, general maintenance, minor upgrade of the LHC experiments.
2015-2018	Run 2	Physics data-taking, $\sqrt{s} = 13\,\text{TeV}$, $\mathcal{L}_{\text{delivered}} = 156\,\text{fb}^{-1}$.
2018-2021	LS 2	CERN accelerator complex consolidation, preparation for High Luminosity LHC, major upgrade of some LHC experiments.

3.2.1 Dimensions and Experiments

The CERN follows the experiment doubling strategy with the LHC by employing the ATLAS and CMS detectors on its ring. Enormous dimension of the ring with 27 km in circumference suggests to use even more experiments. The parameters of machine did not allow its construction in a hall or a building, the only possibility was to go underground. The solution minimizes the impact on the landscape and the soil provides a good shielding against cosmic radiation. The machine, together with its experiments, is placed roughly 100 m underground in a tunnel complex interconnected with the pre-accelerating stages. The ring has eight places called interaction points (IP), where the beam crossing is possible, however only four of these are active. Each of the IPs hosts an experiment or provides a service for the beam, such as a beam cleaning at IP3 and IP7, an acceleration at IP4, and a beam dump at IP6. In total the LHC engages four major experiments at the active IPs, namely:

- ATLAS - a general purpose detector placed at IP1,
- ALICE - a heavy-ion detector placed at IP2,
- CMS - a general purpose detector placed at IP5,

- LHCb - a b-physics specialized detector placed at IP8.

All the IPs are marked in an illustration of the LHC underground complex in Figure 3.1. The LHC colorful physics program is completed by three secondary experiments. TOTEM measures the collisions in the forward region primarily the $\sigma_{\text{p-p}}$, but also elastic scattering and diffraction dissociation of the protons, and is placed at IP5. LHCf simulates the cosmic rays in the atmosphere using the forward protons originating in IP1. The MoEDAL searches for magnetic monopoles and highly ionizing stable massive particles, occupying the complementary kinematic region of LHCb at IP8.

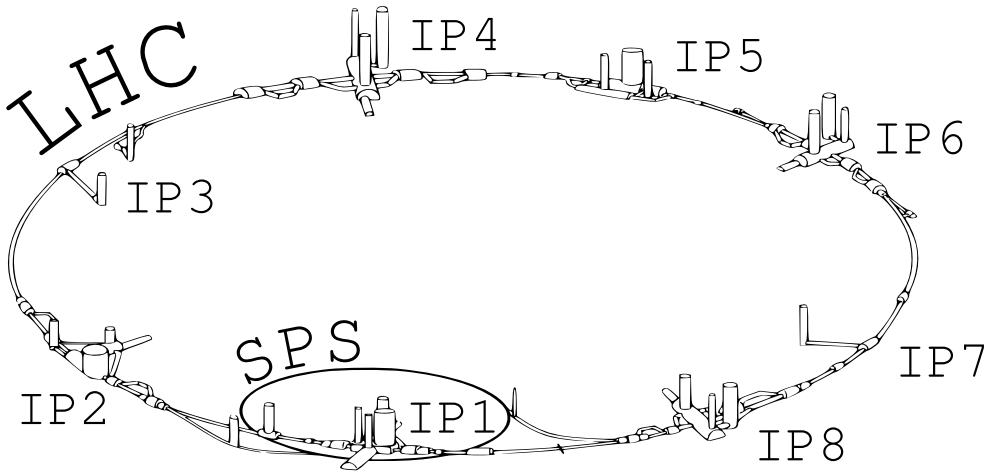


Figure 3.1: Sketch of the LHC underground complex with marked interaction points and an illustrated ring of Super Proton Synchrotron (SPS).

The ATLAS and CMS are high luminosity experiments employing a so called general purpose detector concept. These almost hermetical detectors are capable of measuring or setting limits on any SM particle in a wide momentum and angular range. Some of the design goals were a discovery of the SM Higgs boson, or a discovery of new particles, pointing to alternative theoretical models, such as the W' and Z' or the superpartners of the Super Symmetry (SUSY) model [62]. The LHCb experiment searches for new phenomena in decays of hadrons containing a b-quark [63], determining the weak mixing angles, exploring the CP violation (first observation of direct violation in Charm meson decay [64]), and studying "molecular" pentaquarks with meson-baryon substructure [65]. The ALICE experiment is optimized for the heavy-ion research in p-Pb and Pb-Pb collisions. The physics program of the experiment concerns mainly the hadron sub-structure, quark confinement, and quark-gluon plasma (QGP) research [66].

3.2.2 Proton Chain and Beam

The experimental setup for the presented analysis consists primarily of a tandem of the LHC storage ring and ATLAS detector. The ring delivers proton-proton collisions and the detector detects the collision products. This is a very simplified picture though, since a proton before the collision has to be accelerated in a chain of Linear Accelerator 2 (LINAC2), Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), and Super Proton Synchrotron (SPS) storage rings, depicted in Figure 3.2. The chain gradually accelerates a beam of protons, taking advantage of the long history of CERN accelerator physics, utilizing the machines from the previous physics programs as the PS or SPS. Besides of the LHC, the proton chain supplies additional experiments, for example Antimatter Decelerator (AD) [67], an experiment which studies properties of an anti-hydrogen molecule [68], and CERN Neutrinos to Gran Sasso (CNGS), a former experiment which for the first time observed a muon-tau neutrino oscillations [69].

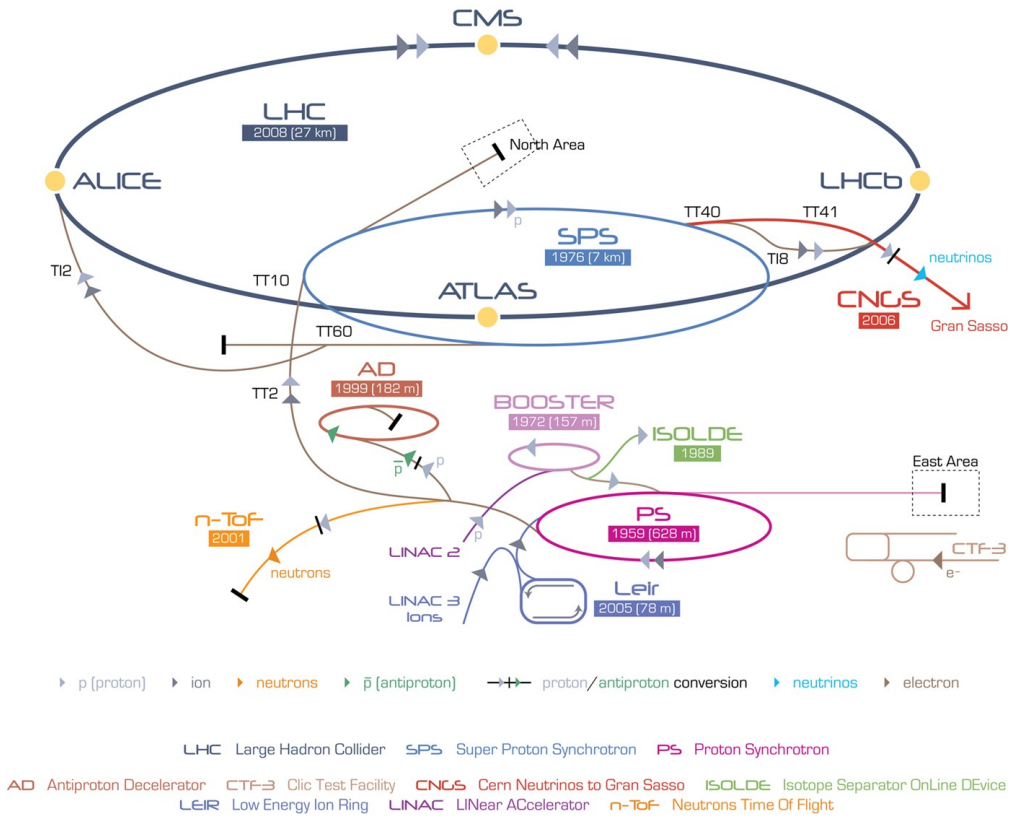


Figure 3.2: A schematic illustration of the chain of particle suppliers and accelerators at the CERN. The figure is completed with another proton beam users, i.e. the CNGS and AD experiments [70].

To increase the number of inelastic proton-proton collisions as much as possible, the LHC is nominally filled with 2808 bunches of approximately 115 billions protons per bunch (N_{p}), entirely called a beam. Two beams circulate in the LHC, and the bunches cross each other at frequency ($f_{\mathrm{b\text{-}c}}$) of 40 MHz causing up to 70 proton-proton inelastic interactions per a bunch-crossing, called pile-up interactions. The parameters result in the machine luminosity according to equation

$$\mathcal{L} = \frac{N_{\mathrm{p}}^2 f_{\mathrm{b\text{-}c}}}{4\pi\sigma_x\sigma_y}, \tag{3.3}$$

where the $\sigma_{\mathrm{x,y}}$ describe the transversal dimensions of the beam. The high frequency of bunch-crossing and the multiplicity of proton-proton collisions reflect both: the need for production of many experiment results, from which the rare and interesting realizations of physics can be found; and the probabilistic nature of the quantum mechanics, although having a minor effect since the beam size is roughly 17 µm at the IP. The designed luminosity of LHC is of order of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The 25 ns bunch timing is defined already by the PS. For a regular run, the PS is filled with 72 bunches forming a batch. The SPS can be filled with up to 4 batches, with a time gap between each, allowing the SPS injection system to react. The SPS batch tuple is called a train. The trains are injected to the LHC per triplets forming the whole filling scheme [71]. During Run 1 of the LHC, a regular beam was arranged of 39 batches of 72 bunches with a pattern 234 334 334 334 made of four train triplets with the first one made of two-, three-, and four-train [72], as can be seen in Figure 3.3.

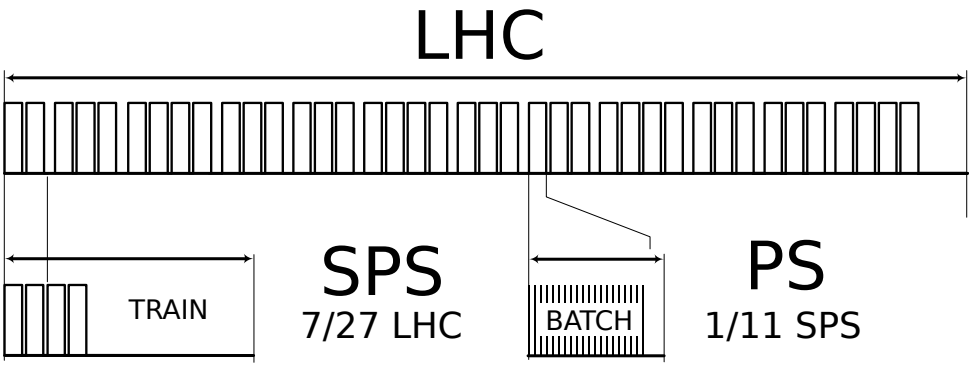


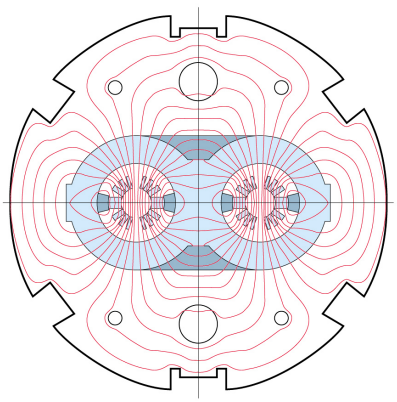
Figure 3.3: A regular filling scheme pattern for Run 1 of the LHC is depicted in the picture, together with the PS batch and SPS train sizes. Also lengths of the storage rings relative to the LHC are displayed.

3.2.3 Ring

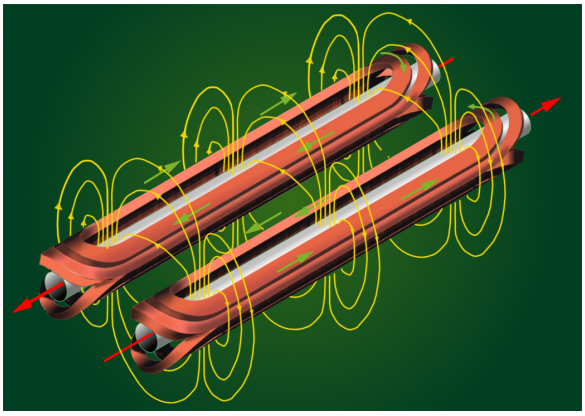
The LHC ring is geometrically not a precise circle, the shape constitutes rather a rounded octagon. The octagon vertices are called arcs and the sides are called edges. The edges have three parts: the central part, called an insertion, is sandwiched from both sides by matching sections and dispersion suppressors. From the machine perspective the experimental or service assemblage can be inserted into the ring as a module. Two beams circulate inside of the ring with the clockwise and counter-clockwise direction along a closed path called a reference trajectory.

3.2.4 Arcs and Dispersion

The arcs bend the beam keeping it on the reference trajectory and allowing a gradual acceleration by a single device, the main advantage of the circular colliders in comparison to the linear ones. A vertical magnetic field for the proton deflection is generated by dipole magnets with a radius of approximately 2.7 km, smaller than the 27 km ring radius (≈4.3 km). The deflection changes a particle velocity vector in the sense of direction and escalates a beam spread. The beam has a non-zero width and length, and has to be taken as a dynamic system with its own phase space. This means the protons enter a dipole magnet with a slightly different angles, momenta, and positions. As a result, the deflecting magnetic field (shown in a dipole magnet in Figure 3.4) increases the differences in the particle collection, causing a dispersion effect in the horizontal plane.



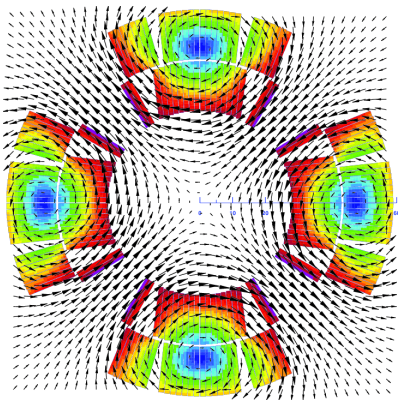
(a) An inter-section of the magnet in the transverse plane to the beam with a depicted magnetic flux map [73].



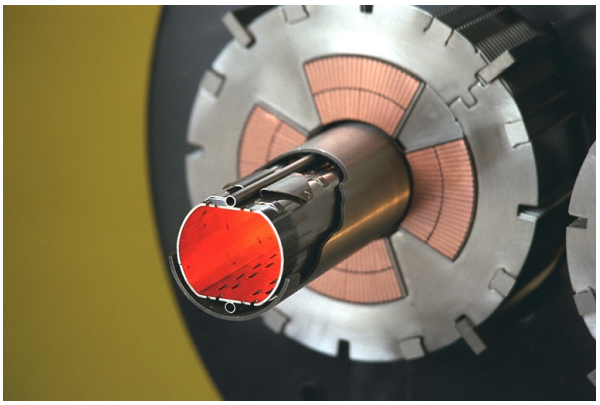
(b) A three dimensional illustration of a field generated by the magnet (yellow), with denoted electric current direction (green), and the beams order (red arrows) [74].

Figure 3.4: An illustration of the magnetic field of a twin aperture main dipole magnet used in the LHC ring.

A beam moving in a dipole magnet is rather stable in the horizontal plane, but destabilized in the vertical plane, taking into account the homogeneous charge of the particle ensemble. This is corrected by a number of quadrupole magnets (shown in Figure 3.5) in a focusing-defocusing F0D0 composition, where the "0" represents a zero focusing element or the deflecting dipole magnets. Simply, these are subsequential quadrupoles revolved by 90°, focusing and later defocusing the beam in horizontal plane and in the opposite order in the vertical plane.



(a) An inter-section of a single aperture quadrupole magnet in the transverse plane to the beam with a depicted magnetic flux map [75].



(b) A photograph detail of the left duct of a twin aperture quadrupole magnet. Four wiring defining the poles of the magnetic field are visible, together with an ovaloid beam screening and two cooling tubes in the beam pipe [76].

Figure 3.5: An illustration of the quadrupole magnets designed for the LHC ring.

The beam bending in the vicinity of an IP is undesired for the dispersion effect, therefore, the straight sections of the ring are needed to make space for a beam recovery. This justifies the octagon shape of the LEP tunnel, which was adjusted to an e^+e^-experiment. However, still more than suitable for the LHC purpose. It also explains the smaller radius of the main dipole magnets. The dispersion is an unwanted effect which needs to be suppressed, because at the interaction points the beam has to be focused and stable to maximize the probability of the proton-proton collisions at the designed center-of-mass energy. The other side of the coin, from the accelerator composition perspective, is, the linear colliders have a negligible dispersion effect.

The LHC is a continuous machine, which means its sections at a certain level blend into each other. A dispersion suppressor connected to an arc still consists of the main dipoles and starts the transition to the straight part of the ring. The dispersion effect needs to be suppressed in the arcs as well. While the ideal

infinitive for a suppressor would be "to eliminate", in case of an arc it would be "to sustain", the effect.

The suppression sequence of dipoles is interleaved with special single powered quadrupole magnets. These are supposed to cancel the horizontal dispersion introduced in the arcs as well as the one generated by the magnetic system at the IPs by adjusting a crossing angle for the beam alternation.

Another two essential purposes of the suppressors are to adapt the LHC reference trajectory to the geometry of the LEP tunnel, and to adjust the beam for a matching section and support the transition from the insertion optics to the periodical solution of arc and vice versa.

3.2.5 Theory and Beta Function

The accelerator physics base stone is a differential equation of motion with periodic focusing properties, so called Hill's equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}s^2} + K(s)x = 0, \tag{3.4}$$

where K is a periodic function of a distance s around a storage ring representing its focusing strength and the x represents a transversal displacement of a beam. A solution of Equation (3.4) leads to a periodic function which implies quasi-harmonic oscillations of the beam in a storage ring aperture as

$$x(s) = \sqrt{\epsilon\beta(s)}\cos\left[\phi(s) + \phi_0\right], \tag{3.5}$$

where the amplitude β is called an optical function, ϕ represents the phase advance of oscillations, and $\sqrt{\epsilon}$ and ϕ_0 are the integration constants. The solution is in principle a machine lattice and operation regime dependent, or in general, time and length dependent. Although it is typically considered from the maximal value perspective, so the envelope eliminates the time dependence. A solution including the dispersion effect would require to solve an inhomogeneous differential equation.

A beam is usually described using Twiss (or Courant-Snyder) parametrization using a phase space ellipse derived from the solution in Equation (3.5), parameters

$$\alpha(s) = -\frac{\beta'}{2}, \qquad \gamma(s) = \frac{1+\alpha^2}{\beta}, \tag{3.6}$$

and the first derivation of the solution

$$x(s)' = -\sqrt{\frac{\epsilon}{\beta(s)}}\left\{\alpha(s)\cos\left[\phi(s) + \phi_0\right] + \sin\left[\phi(s) + \phi_0\right]\right\}. \tag{3.7}$$

The ellipse parametrization is depicted in Figure 3.6 together with the equations justifying the choice of α and γ parameters. These highlight the proportionality of the beam size to the β-function and the beam dispersion to the γ, an inverse function of β-function. The ellipse area is defined by the emittance ϵ which acts as a constant due to Liouvile's theorem [77]. This implies a low β-function gives a smaller beam size although a larger beam dispersion and vice versa. In the transverse plane a beam experiences betatron oscillations, emerging primarily due to the beam bending in the arcs. The beam perturbations in the direction of motion are driven by synchrotron oscillations caused by the accelerating system. The beam reference trajectory is never fully achieved due to a misalignment and factory imperfections of magnets. Therefore, the trajectory must be corrected for the errors, giving a closed orbit. On the closed orbit the β-function is integrated, see Figure 3.7. The function represents one of the most important parameters of the machine.

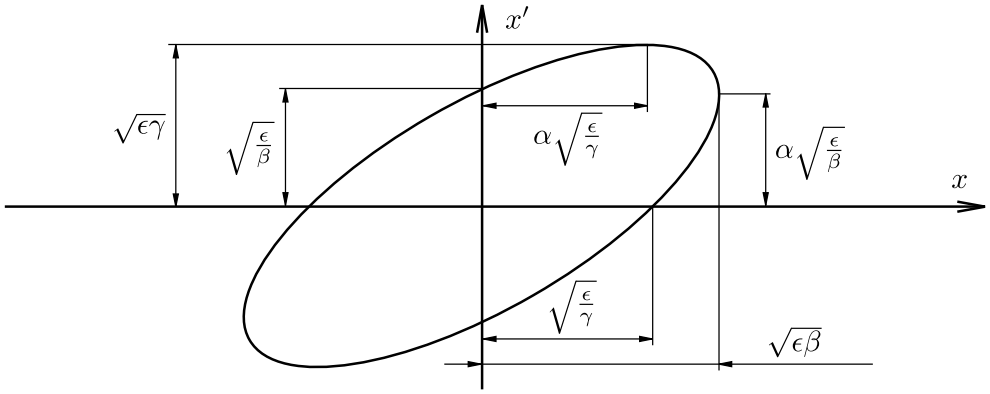


Figure 3.6: The phase space ellipse describing the transverse displacement of beam in the x direction and its derivative x', as a solution of Equation (3.4).

Since the β-function depends on the machine lattice, another essential storage ring parameter is a number of oscillations per turn. It is called a betatron tune and exists in both horizontal and vertical planes. The number stands additionally for a possibility of resonances, as betatron, synchrotron, skew, and combined synchro-betatron resonances. A storage ring as complex as the LHC has to deal with many sources of a non-linear force arising in multipole magnet correctors defining the real shape of the closed orbit. A threshold amplitude beyond which the betatron motion becomes unbounded is called a dynamic aperture [78]. As was mentioned, a bunch in the beam is a dynamic system. It embodies collective effects, arising from the electromagnetic interaction of the particles among themselves, with their environment (possibly including outside particles such as electrons), and with the other beam. Collective effects are also a function of the vacuum system geometry

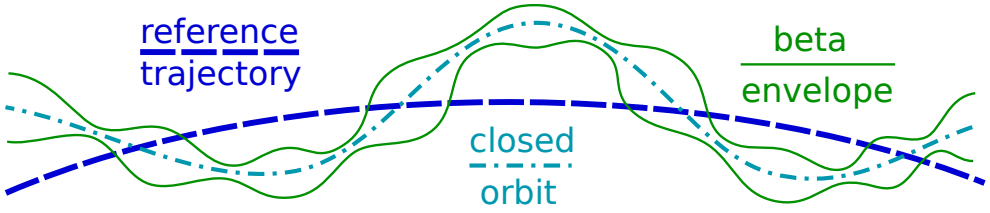


Figure 3.7: An illustration of a beam particles trajectory. The designed reference trajectory is corrected for magnet errors due to a misalignment and construction imperfections, becoming the closed orbit. The β-function is integrated to the closed orbit giving the real particle motion envelope.

and its surface properties. Coherent oscillations generate a charge in the vacuum system causing instabilities which must be suppressed by Landau damp correction magnets [79].

The β-function is periodical in the arcs because of the F0D0 cells. In the insertion regions the amplitude variance is rather small, and minimal in the low-β experimental insertions of the ATLAS and CMS. Increase of the focusing strength decreases the size of the beam envelope β and increases the tune and vice versa [80]. The transition between the periodic optic in an arc and the low-β optic in an insertion region is managed by a matching section. It is a natural follow-up of the dispersion suppressors, again a sequence of the quadrupole magnets, now without dipoles.

3.2.6 Insertion Regions

The straight sections of the ring have a length of 528 m, and have two categories, an experimental and utility insertion region. The utility insertions contain the beam control and maintenance appliances which demand a special adjustment of the beam. The experimental insertions have the active IP in the center which demands separation and recombination dipole magnets and a device for the final focusing of the beam before a collision. The final beam preparation for each experimental insertion region is made of three quadrupole magnetic systems used for a reduction of the optical β-function, called an Inner Triplet (IT). The triplet is connected to a matching section and represents the final stage of the IP sandwich towards the center.

Dipole magnets commence a collision by a gentle deflection of beams under a small angle towards the IP, resulting in a crossing angle of 285 µrad. In case of the low-β regions, the crossing-angle scheme for the ATLAS is in the vertical plane, while for the CMS in the horizontal plane. This approach minimizes an unwanted long-range beam-beam interactions since the two beams share a single vacuum

chamber at the experimental insertions. During a beam injection and ramp-up phase the collisions are avoided by a parallel deflection of the beams out of the crossing plane achieved by a magnetic bump.

The LHC has to count also with the magnetic systems of the LHCb and ALICE experiments. The magnetic fields of their spectrometers are perpendicular to the beam, asymmetrical, and directly affect the beam conditions. Therefore also compensating magnetic systems for these experiments have to be employed in the ring. This fact also constrains the choice of the beam crossing and separation scheme. Moreover, IP8 is shifted by 15 half RF wavelengths ($\approx 11.25\,\mathrm{m}$) towards IP7 in order to make enough space for the spectrometer magnet of the LHCb experiment.

3.2.7 Acceleration

The LHC acceleration system design profits from the loop shape of the machine. A single Radio Frequency (RF) system is reused many times at every turn of the beam.

Protons are injected to the LHC from the SPS at energy of 450 GeV and accelerated up to energy of 6.5 TeV afterwards. In Run 3, the maximal beam energy is planned to be the nominal 7 TeV. The storage ring needs the highest aperture at the injection energy. Once the particle ensemble is accelerated in the ring, due to the relativistic contraction the emittance is reduced and the beam size shrinks (i.e. the transverse beam size changes from 1.19 mm to 0.3 mm [79]). The machine optics has to adapt to different energies during the acceleration, taking in to account a parameter called chromaticity [81]. It is a parameter that describes how particles of a different momentum see a different focusing strength in the quadrupole magnets and, as a consequence, have a different betatron oscillation frequency [82]. Mathematically it is how a change of the relative momentum vary the tune.

The RF system utilizes superconducting acceleration cavities, generating an electromagnetic field of a nominal frequency of 400 MHz. Each bunch of a beam is carried in a RF bucket as on a wave. The dynamic electromagnetic field acts on the full length of a bunch, accelerating the "late" protons more and the "early" ones less. It keeps the particles together, although is the primary source of the synchrotron oscillations which needs to be maintained. In fact the LHC employs an exclusive RF system for each beam. Separated accelerating system is less demanding on the beam handling power than a wide aperture cavity system for both beams. The superconducting cavity in Figure 3.8 is larger than a regular copper one. The standard distance between the beams defined as 194 mm had to be increased to 420 mm, to enable two superconducting RF cavities accelerate the beams next to each other [79]. This is another point where the separation dipole magnets enter the scene, similarly as at the IPs. However, deflecting the beam in

the opposite direction. The RF system also controls the timing of the beams and is a point where the synchronization signals are distributed from.

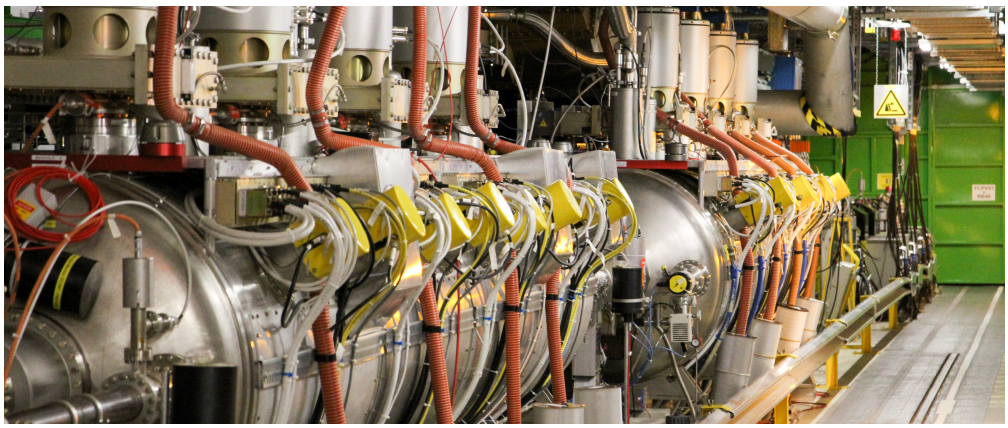


Figure 3.8: A photo of a part of the RF accelerating system, placed in a helium cryostat [83].

3.2.8 Beam Cleaning

The beam cleaning is achieved by various devices, but primarily by a combination of collimators and magnets. A collimator physically cuts off beam margins by its jaws actuated in and out to the beam pipe of the storage ring. Any possible hardware solution for the collimators can only resist a small fraction of the LHC beam energy [84]. Therefore a considerably precise and firm construction of the jaws had to be achieved, using carbon and tungsten based materials with a high melting point to withstand the extreme conditions of temperature and stress in case of accidental beam disturbances [85]. Particles outside of a tolerable momentum range must be absorbed in the momentum cleaning insertion before they can be lost in the arcs. The arcs protection rises a special interest, since the cold mass of the main dipole magnets is vulnerable to loss of its superconductivity, called a magnet quench. The jaws are outstretched between $6\,\sigma$ to $7\,\sigma$ from the beam center, supposing the σ to be a natural unit of the beam. Such as the standard deviation in a case of the Gaussian distribution of particles in the beam.

A multi-stage cleaning process is implemented in the LHC storage ring. A primary collimator intercepts a possible lost primary proton generating an on-momentum and off-momentum secondary proton halo. The secondary proton halo is intercepted by the secondary collimators which leak only a small tertiary halo. The last halo represents a negligible harm to the cold mass of the superconducting magnets. The tertiary collimators are used locally to provide an additional protection of sensitive devices [79]. The collimators absorb only a small fraction of the

energy from the lost protons. To avoid secondary particle showers carrying energy from the beam, additional fixed absorbers are installed in to the collimation system. The machine employs also two injection collimators at IP2 and IP8, installed to protect the ring against a mis-kicked beam during the beam injection. And collision collimators, installed to catch the debris from proton-proton collisions at the active IPs.

Another device installed in the cleaning insertion is called a scrapper. The scrapers are in the direct physical contact with the beam as well as the collimators, although are used for a beam halo diagnostics and therefore are placed before the collimation system. The scrappers do not use jaws rather a tiny probe designed to measure the beam. The LHC employs six scrappers for the vertical, horizontal, and momentum measurements of each beam.

3.2.9 Beam Dump

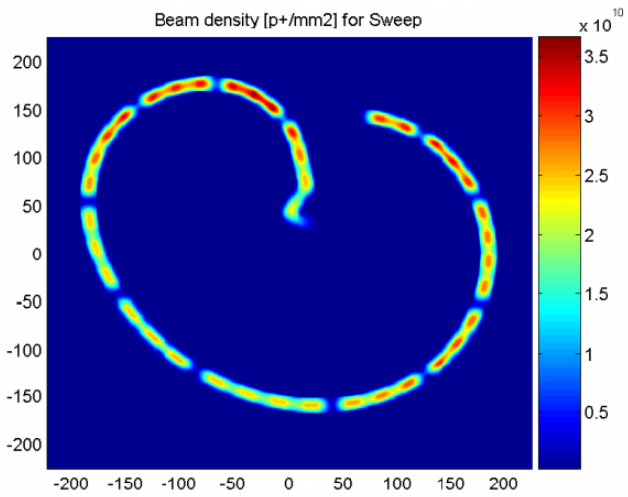
The beam energy stored in the LHC rings can be of up to 700 MJ (3×10^{14} of protons at 7 TeV). This is an unprecedented energy which can be highly dangerous if badly manipulated. Therefore, a reliable beam dump system has to be employed to protect the machine, in case of serious fault conditions arise. It is important to mention, the extremely challenging environment of the storage ring accumulates a large additional amount of energy in the magnets. So a possible mismanipulation of the beam can cause a subsequent release of this energy. A magnet quench can have dangerous consequences as the charge flowing through a magnet wiring has to suddenly face a nonzero resistivity. Such as a scenario that happened at the beginning of the LHC Run 1, when a superconducting magnet connection almost immediately evaporated, see Figure 3.9a.

To be dumped the beams are vertically extracted from the machine using a combination of horizontally deflecting fast-pulsed ("kicker") magnet and vertically-deflecting double steel septum magnets. The beam dump completion takes between two and three turns. Once the beam is sent to the beam dump, to withstand the whole power of its energy, it needs to be swept around as can be seen in Figure 3.9b. The figure deserves a comparison with Figure 3.3, where one can identify the regular Run 1 filling pattern.

3.2.10 Vacuum

The conditions under which the beam is stored and accelerated must be as free as possible of any disturbing effect. This includes particles from the environment which are eliminated by the ultrahigh vacuum with pressure below one nPa. Together with the insulation vacuum systems for superconducting magnets and helium distribution line, at pressure levels of an order of hundreds nPa, the LHC

(a) A photo of damaged magnet connection [86].



(b) A 2D image of the face of beam dump absorber (counting protons per area) after a 450 GeV beam with the regular Run 1 filling pattern was dumped [87].

Figure 3.9: Two possible scenarios of the release of energy from the LHC machine. Accidental release caused by the incident on 19 September 2008 (a), and a properly dumped beam (b).

vacuum complex is the largest in the world. It consists of roughly 104 km of pipes, 50 km for the insulation, and 54 km for the beam itself [79].

3.2.11 Magnets

The main role in the LHC no doubt is played by the magnets. The LHC has 1232 bending dipole superconducting magnets with nominal magnetic field of 8.33 T, cooled down to 1.85 K, making them colder than the universe with the average temperature of 2.73 K. Each arc employs 127 dipoles for the primary bending, the rest is placed in the dispersion suppressors, 8 dipoles each. Another 70 dipoles are operated for the separation and recombination of the beams, in a single and twin aperture implementations, and for the horizontal orbit correctors, as well as for the experimental spectrometers of the ALICE and LHCb and their compensators.

Once the beam is kept in the LHC ring, the quadrupole magnets are employed for its stabilization on the closed orbit. Each arc employs 45 magnets in the F0D0 lattice cells, 40 for a tune correction, and 8 for a skew correction of the beam. The remaining 32 main quadrupoles in the insertion regions finalize the periodic solution of the arcs. Each IT in fact contains 4 magnets, organized in three logical magnet blocks. The beam focusing is handled by 110 quadrupole magnets placed in

the dispersion suppressors and matching sections of the four experimental, and also the RF, and beam dump insertions. While another 168 quadrupoles are dedicated for the two cleaning insertions only.

In total, the LHC employs an incredible number of 9594 magnets, including the beam benders, horizontal and vertical orbit correctors, dispersion suppressors, beam separators and recombinators, skew and lattice correctors, beam kickers and bumpers, inner triplets, and experimental spectrometers and their compensators. The number of various magnets illustrates how complex the machine is. The largest fraction makes the 2464 sextupole, 1232 octupole, and 1232 decapole spool piece corrector magnets shadowing the main dipoles. A spool piece magnet corrects the field errors between the main dipole magnets, enables the maximal dynamic aperture, and defines the closed orbit. The machine employs even 8 dodecapole magnets for non-linear field harmonic corrections.

3.2.12 Collisions

Two identical beams are accelerated by the LHC in the opposite directions to enable a head-to-head collision and each consists of unpolarized protons. These facts decide the collision products are uniform along the polar angle of the beam as well as along the axial angle. Finally, this suggests a barrel shape characteristic of the ALTAS detector.

3.3 ATLAS Detector

A Toroidal Apparatus or the ATLAS detector encompasses IP1 of the LHC to detect, reconstruct, and measure almost any particle produced in a proton-proton collision. The only known particle capable of escaping the detector vast undetected is a neutrino. A solid-angle coverage is nearly 4π, limited only by the LHC beam pipe itself. The ATLAS is the largest of LHC detectors. It is approximately 44 m long barrel-like apparatus having 25 m in diameter, which weights 7000 t. This section again pinpoints only the important or interesting facts, for more details the reader should refer to the ATLAS technical design report [50].

3.3.1 Geometry

A cylindrical geometry of the detector where the axis is identified with the LHC beam gives a radial and forward-backward symmetry for the event reconstruction. The ATLAS uses a right-handed coordinate system with its origin at the nominal IP in the center of the detector and the z-axis along the beam direction. The x-axis

points from the IP to the center of the LHC ring, and the y-axis points upward. Cylindrical coordinates (ϕ, r) are used in the transverse (x, y) plane, ϕ being the azimuthal angle around the beam direction. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln[\tan(\theta/2)]$ and with advantage preferred as a kinematic variable in the detector since its difference is invariant under boosts along the z-axis. Angular distance is measured in units of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

3.3.2 Subdetector Composition

From inside out a particle originating in a collision can hit at first a pixel based, a microstrip based, and eventually a straw-tube based detector. All three entirely forming an Inner Detector (ID) serving as a device for reconstruction of tracks of charged particles with coverage up to $|\eta| = 2.5$. All subdetectors provide full coverage in the azimuthal angle. The ID is placed inside of a superconducting solenoid generating a magnetic field for the charge and momentum reconstruction of a particle [88].

The following set of sub-detectors handles the particle energy measurement. A finely segmented lead/liquid-argon (LAr) electromagnetic calorimeter is placed first, with a coverage up to $|\eta| = 3.2$. A steel/scintillator-tile hadronic calorimeter follows, securing the central region up to $|\eta| = 1.7$. Another LAr calorimeters in the endcaps enclose the forward regions and complete both the electromagnetic and hadronic calorimeter system coverage up to $|\eta| = 4.9$.

The whole apparatus is enclosed by a subdetector system dedicated to muons. A muon spectrometer (MS) consists of three layer system of precision drift tubes and cathode strip chambers providing the muon tracking in a range of $|\eta| = 2.7$ and resistive plate and thin gap chambers for fast triggering capability up to $|\eta| = 2.4$. The MS system is placed in a toroidal magnetic field. The MS is the only subdetector that does not provide a full coverage in the azimuthal angle, as a consequence of the ATLAS size and weight. The coverage is limited around $\phi = -\pi/2$ due to a firm steel supporting structure (feet), which does not allow a placing of the MS modules at the bottom of the detector body. All subdetectors are depicted together in the cylinder in Figure 3.10.

The order of individual subdetectors is utilized for particle identification. Simple rules correspond to the particle species:

- no other particle than a muon can reach the MS,
- only a hadronic particle or a muon, can reach the hadronic calorimeter,
- an electron and photon can reach at most the electromagnetic calorimeter,
- only a charged particle leaves a track in the ID,

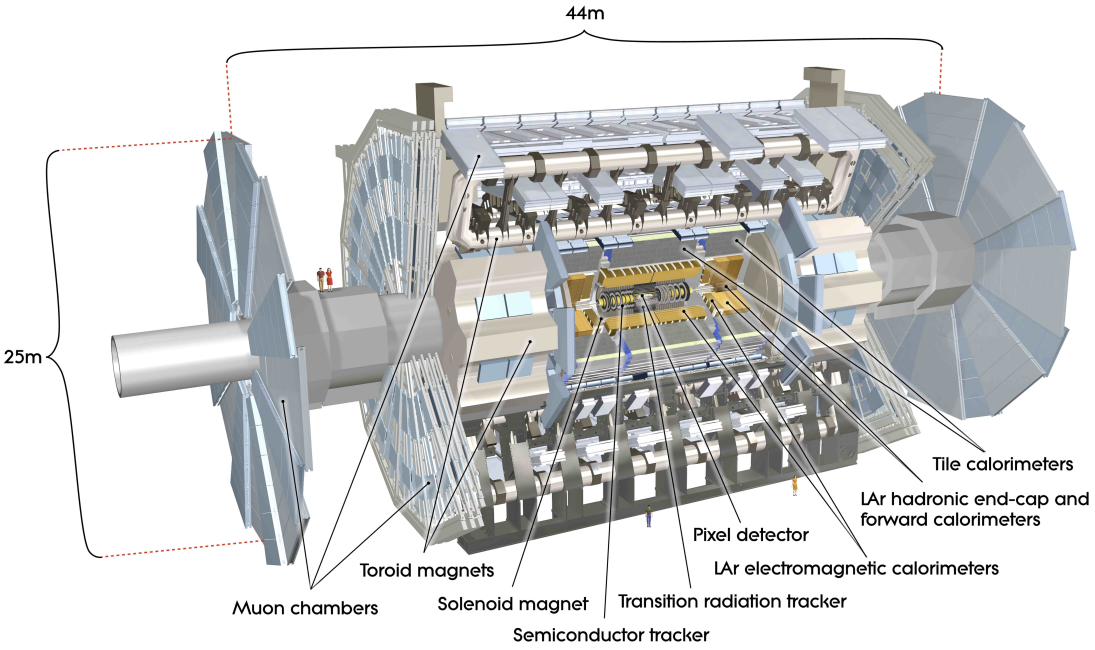


Figure 3.10: Overview of a subdetectors placement in the ATLAS cylinder. Basic dimensions and people silhouettes are placed in the picture, to illustrate the size of detector [89]

.

- a neutrino is the only particle which leaves the detector body undetected.

To summarize the bullet points, the detector is able to estimate neutrino properties, exploiting the information from all other reconstructed particles and the four-momentum conservation law. Taking into account the proton-proton collisions multiplicity, the ID has to face a considerably busy environment of charged particles [90] and the calorimeters need to be corrected for the pile-up jet contamination from other events.

3.3.3 Particle Energy and Magnetic System

As was mentioned, the ATLAS is a general purpose detector, therefore it can measure particle properties in a wide range of energies. From an order of GeV in the b-physics region of interest to an order of TeV, where one might expect the presence of exotic physics. It makes the detector a universal tool for HEP, which is projected to a rich physics program the ATLAS collaboration conducts, since it is partially able to cover also the programs of LHCb and ALICE.

The detector has to reconstruct the four-momentum of a particle, or at least

the position four-vector of energy deposit, for the subsequent event reconstruction. The calorimeter subsystem provides the energy measurement of a particle, in a destructive manner, which requires a good resolution in both detector angles ϕ and η.

The momentum measurement relies on the deflection of a charged particle in a magnetic field. Inner (ID) and outer (MS) tracking sub-detectors have their measurement resolution tightly connected with the magnetic field orientation. Intensive magnetic field flux density and a fine detector granularity in a bending plane define the good momentum resolution. These two properties also specify a detector saturation and the upper limit of the momentum range.

Both energy measurements has to cope with three-dimensional spatial information of signal. A calorimeter processes the signal to reconstruct the affected detector volume, where the particle energy has been absorbed. A tracking detector typically processes the signal from several detector layers and identifies hits of a particle to reconstruct its track. Both detectors process the information from layers, although the fact the layers of the momentum measurement are not adjacent to each other, makes its processing highly combinatorially demanding. This predestinates the calorimeters for usage in the fast filtration of event trigger, while the tracking detectors for the subsequent precision measurement.

The magnetic field for the ATLAS is provided by a unique magnetic system, shaped by a Central Solenoid (CS) and three surrounding toroids, a Barrel Toroid (BT) and two Endcap Toroids (ECT). The CS provides a peak magnetic field of 2.6 T for the ID particle tracking. The toroid magnets generate a magnetic field for the outer particle tracking of the MS in a range of peak values between 3.9 T to 4.1 T. The toroids provide a bending plane parallel to the beam direction. Together with the CS bending plane perpendicular to the beam, the ATLAS adopts a two-plane tracking scheme. This makes the detector unique in comparison to the other LHC experiments.

3.3.4 Inner Detector

The ID is composed of a set of three complementary detectors. The two innermost ones are based on the semiconductor technology, while the third one utilizes the transition radiation. A particle non-destructively interacts with the ID layers and tubes, generating a signal from which its track is reconstructed [91, 92].

The track is represented by the detector angles η and ϕ, impact parameters z_0 and d_0, and parameters connected to the track sagitta as the momentum and charge (sagitta direction). The angles determine the track direction and the impact parameters the track-to-beam perigee. The impact parameters are obtained by extrapolation of a reconstructed track to the center of detector resulting in a position at the beam axis z_0 and a distance from the axis d_0. The extrapolation

is also employed in the hard collision (primary) vertex identification, another vital purpose of the ID besides of the momentum measurement. A primary vertex is identified when more than one track share the same impact parameter values. A secondary vertex, belonging to a decay of c or b-quark mesons, is again identified by coincidence of more than one track although with high d_0 parameter.

The semiconductor based part of the detector is composed (from the center) of four hi-resolution pixel layers in the barrel and three layers in each endcap, followed by four semiconductor strip layers in the barrel and nine layers in each endcap. A longitudinal crosscut of the ID is depicted in Figure 3.11 with individual subdetectors illustrated as blocks, including exact location in R and z coordinates with highlighted η coverage. The silicon detectors cover the whole η-range of the ID ($|\eta| < 2.5$).

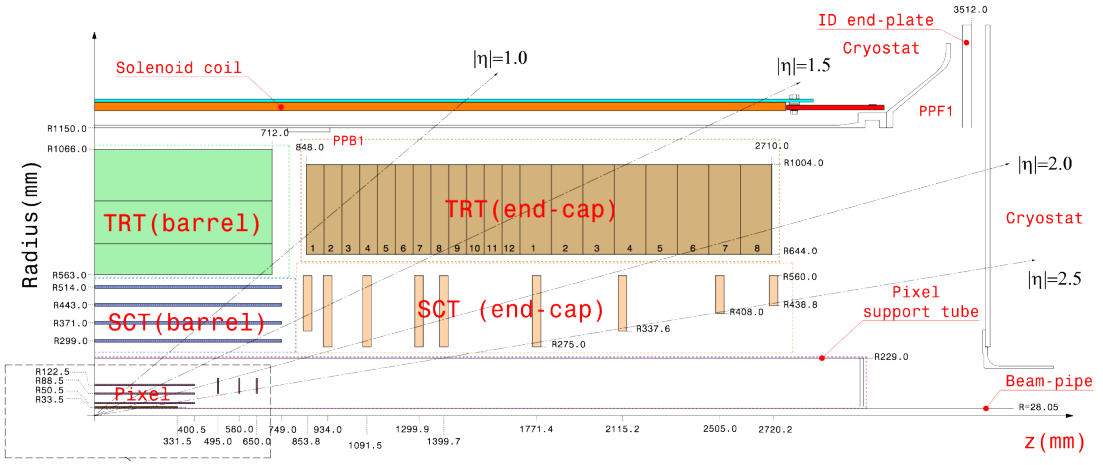


Figure 3.11: Technical drawing of the ID, the subdetectors are displayed as blocks in colors, dimensions are in mm. The figure also depicts construction elements, as a beam pipe, cryostat, or the CS coil. The η scale expresses the forward coverage of each subdetector [93].

The innermost layer of a Pixel Detector (PD) is an upgrade which was not in the original design. An Insertable Beam Layer (IBL) was inserted to the system after Run 1 and is placed as the closest layer to the LHC beam. The IBL improves the impact parameter resolution, enhances the vertex identification, and increases the tracking robustness [94]. The ϕ-R resolution of the multi-layer system of PD is 10 µm, corresponding to the track deflection plane. For the vertex identification the z resolution is more significant, being 115 µm. The very fine resolution of the PD is burdened with a challenge of readout of more than 80 million channels. All details regarding the PD can be find in Reference [95].

The next stage of the silicon detector consists of a Semiconductor Tracker (SCT). The SCT, based on stereo-angled doublets of silicon microstrip sensors with 80 µm

pitch [96] and an angle of 40 mrad, enhances the resolution perpendicular to the direction of the angle. It is a trade of between two orthogonal resolutions in the microstrip plane, preferring the one in the direction of track deflection (ϕ) against the other (z or R). The ϕ-R resolution in this case is 17 µm while the other one is only 580 µm [97, 98]. The secondary resolution corresponds to the z direction in the barrel, because of the sensors are placed on the cylindrical surface. The endcap sensors are placed fan-like on a disk, resulting in the secondary resolution to be in the R direction. The SCT needs a readout of roughly 6 million channels. The subdetector is in full detail described in the documentation.

A semiconductor detector is based on the electron-hole pair generation induced by a passage of a particle through the depletion region of a diode in the reverse regime. The mechanism principle, creation of the region, and various techniques related to the silicon detectors are well described for example in Reference [99].

A Transition Radiation Tracker (TRT) made of straw tubes, is engaged in the second part of the subdetector. The tubes are filled with xenon gas, ionized by an electrostatic field, to enable a detection of the radiation emitted by a particle transition over an active medium between the tubes. The medium capable of a transition radiation is composed of two adjacent materials of different dielectric constant [100]. The TRT barrel consists of 73 layers of the tubes [101] and the endcaps of 160 layers [102], providing 36 hits per a particle passage in average. The tube system enhances the tracking resolution up to $|\eta| = 2$ by 130 µm resolution in the ϕ-R plane only. The TRT possesses a number of 350 thousand channels [101].

Once a track is measured, the charge and momentum of a particle can be deduced. The information of a charged particle four-vector can be completed via a connection to the energy measurement acquired in the successive calorimetry stage.

3.3.5 Calorimeters

The calorimeter measurement is destructive for all visible particles, except for a muon due to its high mass, two hundred times higher than the electron mass.

Predominantly, a particle in the electromagnetic calorimeter undergoes an interaction with matter creating an electron-positron pair (photon), or just radiating a photon due to the *bremsstrahlung* effect (electron). A photon can be described as being in a superposition of an electron-positron loop, so, when it approaches a nucleus the loop can disintegrate into a leptonic pair. An electron just radiates on a nuclei in the metal matrix of absorber matter. The characteristic parameter of absorber in electromagnetic calorimeter is called a radiation length, the mean distance from the entering point of a particle at which the particle loses $1/e$ of its initial energy. The interaction with an absorber layers gradually decreases the particle energy, consequentially increases the number of emitted particles, primar-

ily in the longitudinal direction, and creates a particle shower. A shower gets wider and fades out as advancing through absorbing layers in the detector bulk, see an electron example in Figure 3.12. The longitudinal and lateral dimensions of the shower are examined and the particle energy measured according to the affected detector volume.

A similar principle works for the hadronic calorimeters as well, although the nuclear processes dominate the interaction. The hadronic shower is enriched by baryons and heavy fragments of the absorber in addition to the mostly presented mesons. Due to higher mass of the mesons, the characteristic parameter of hadronic absorber is defined as the mean distance from the entering point of a particle at which the particle undergoes inelastic interaction. The parameter is called a nuclear interaction length. The inelastic interaction disintegrates the incoming particle, produces a number of secondary particles, and reweighs the showering process back to the electromagnetic nature.

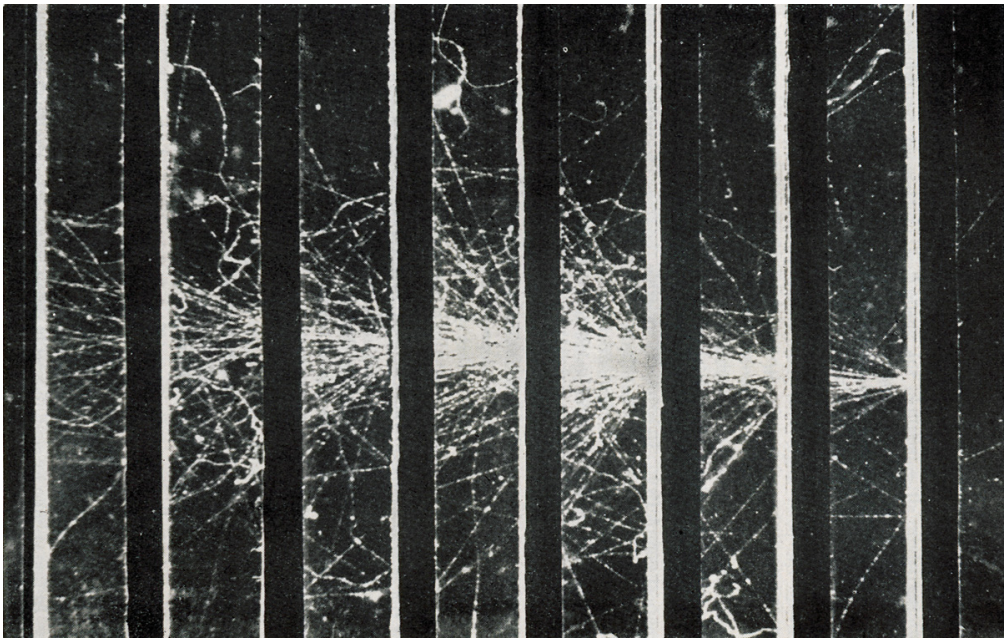
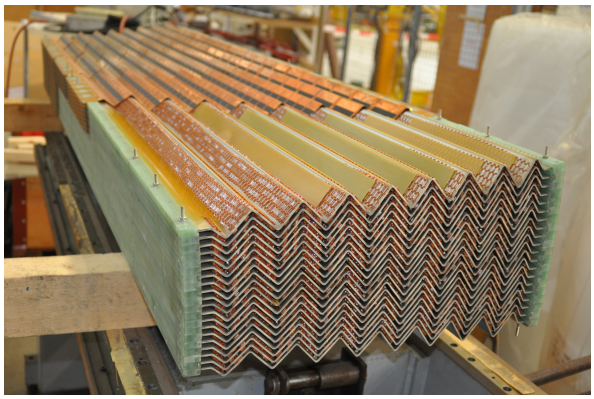


Figure 3.12: A particle shower initiated by a high-energy electron passing through calorimeter absorbers. (Courtesy: C. Y. Chao, American Association for the Advancement of Science).

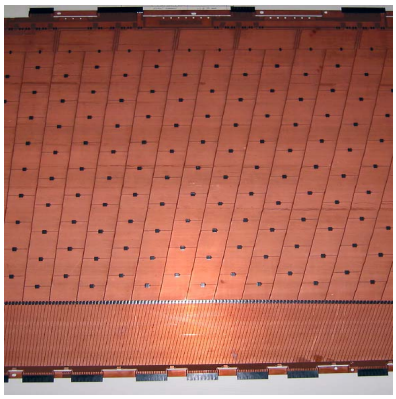
All calorimeters in the ATLAS are of a sampling type. It means the calorimeter mass is divided by high density material absorbers into segments in the azimuthal and rapidity angles. The segments are filled in with a substance capable of detection of the energy, released after an interaction of a particle with the absorber, a scintillator crystal (light signal) or a ionizable gas in electrostatic field (elec-

tric current signal). The ATLAS detector employs both signal generations, light in case of a Tile-Calorimeter (TileCal) and electric current in case of LAr based calorimeters.

The ATLAS electromagnetic calorimeter is composed of a barrel and two endcaps, sharing the same technology for the absorbers and electrodes stacking in an accordion shape [103]. The shape evolves in the direction out from the detector center, and is composed of lead absorbers interleaved with three copper electrodes separated from each other by insulating sheets of inter-gap spacers. The space between the absorbers and electrodes is filled by the argon element in a liquid state, for its stability and radiation hardness. The LAr also keeps the detector uniformity, thanks to its liquid form. On the other hand, the whole system has to be cooled down below the boiling point of the substance, so the whole calorimeter must be placed in a cryostat. The outer two electrodes are connected to a high voltage source, to ionize the LAr in the spacers, while the middle one is used for the signal readout. An accordion shaped calorimeter prototype in an assembly process can be seen in Figure 3.13a and the etched readout electrode before folding in Figure 3.13b.



(a) An expansion of the accordion shape from the left to right as the absorbers evolve from the detector center [104].



(b) An evolution of the pseudorapidity angle to the right is visible on the central electrode [105].

Figure 3.13: Photographs of the electromagnetic LAr calorimeter components. The accordion shaped section of a barrel calorimeter (a) and its readout electrode etched in a copper plate before a horizontal folding (b).

The hadronic calorimeter utilizes two technologies, the TileCal type in the center and the LAr type in the forward part. The central part is composed of a barrel and two extended endcaps each of three made of 64 wedges forming a ring around the electromagnetic calorimeter, with the wedge angle following to the polar

angle of detector. The calorimeter constitutes of steel absorbers, and scintillator tiles as the active medium. Tiles are placed perpendicular to the beam axis, sandwiched in the wedge, to increase the granularity in the η-direction. The signal light is sent by an optical fiber to a photo-multiplier and further processed by readout electronics. The calorimeter must be able to read out more than 460 thousands tiles [106].

The LAr part of the hadronic calorimeter is as well of the sampling type, utilizing a copper absorber, with a flat proposition (no accordion). The space between absorbers is divided by three double electrodes separated by insulating gap spacers, creating four independent drift zones for the LAr ionization. The electrostatic field is achieved by grounding of copper absorbers and connecting a high voltage to the nearer side of the next electrode. The other side of the electrode is again grounded, starting a new drifting zone. The middle electrode serves for the signal readout and carries an etched pad-structure defining the lateral segmentation of the detector. The electrode is sandwiched by a high voltage plates. The hadronic endcaps have to a readout 5632 channels only.

The LAr calorimeter has the characteristic time period defined as the length of a single ionization pulse composed of a peak, when the generated charge is collected on electrodes, and an inverse pedestal, when the system recombines. The period is called a drift time and the typical value for the calorimeter is 400 ns. The full period is too long to be read out, therefore only the peak using the first four 25 ns samples is considered in Run 2 [107]. The inverse pedestal makes a constant bias for measurements in the following bunch-crossings and has to be corrected, primarily in the forward region of the detector.

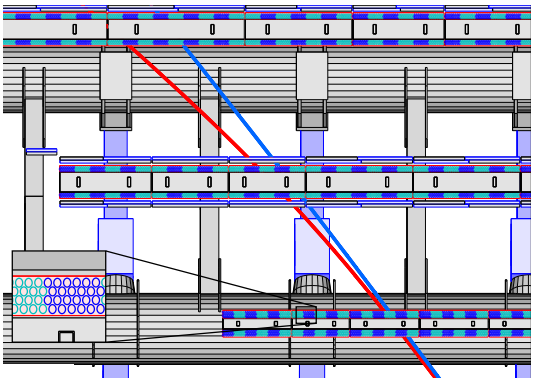
3.3.6 Muon Detectors

A muon can pass through the whole detector undestroyed, furthermore in each part observed i.e. ID and calorimeters. It primarily interacts with the electrons in the atomic matrix of detector matter. The electron-muon scattering angle is minimal due to the heavier muon. Accordingly, a muon ionizes much smaller volume of the detector than an electron, and no electromagnetic shower evolves in the matter. The ATLAS utilizes a gaseous MS of various types. The first two are fast MSs for an event trigger purpose, having three to four active layers. They adopt a technology of combination of multiple anode wires and cathode strips resulting in a Thin Gap Chamber (TGC) or plates for both electrodes as a Resistive Plate Chamber (RPC). The chambers provide an electrostatic field of a very high intensity supplied by 3 kV and 10 kV (TGC and RPC). The field causes a gas ionization and initiate the avalanche effect, a secondary outcome is an acceleration of the charge transport to the electrodes which makes the readout process faster. The RPCs are placed in the barrel while the TGCs are placed in the endcaps of

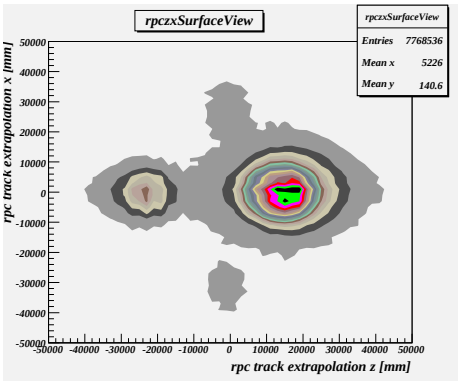
the ATLAS detector. The precision measurement of a muon track is delivered by another two MS types. A Muon Drift Tube (MDT) utilizes a technology of single wire tubes and a Cathode Strip Chamber (CSC) again combines multiple anode wires and cathode strips The MDT electric field is driven by voltage of approximately 3 kV.

To enable a very precise momentum measurement of muons the MDT is specialized for a measurement in the lateral plane to the beam. This corresponds to the fact of placing the MS in the toroid magnet. Since the MDT provides only a measurement of the coordinates in the deflection plane (η-R), the space-point information has to be completed by the azimuthal angle measurement (ϕ) from the RPC and TGC.

The MDT is a primary tracking detector with six multi-tube layers, covering the majority of the cylinder, see Figure 3.14a. The MDT have a limited rate of muon detection and since the forward detector region expects a higher number of muons passing the MS, the tubes would be saturated in a region of $|\eta| > 2$. Therefore, the nearest MDT layer to the beam is removed and the CSC with a higher detection rate provides the measurement capability instead [108].



(a) The MDT layers with a detail of multi-tube composition in the bottom left corner. The bottom two layers consist of four rows of tubes while the others of three. Two muon tracks with a different p_{T} are depicted by red and blue lines [50].



(b) The map clearly demonstrates a higher number of muons arrived to the detector from the secondary and primary shafts ($x = 0$) and some of them arrived even from the service elevators ($z = 0$) [109].

Figure 3.14: Figures related to the ATLAS muon detector. The MDT detector vertical crosscut in the magnetic deflection plane (a) and a contour map of a number of tracks reconstructed by the RPC and extrapolated to the ground level.

3.3.7 Trigger

The ATLAS trigger is the first system that processes raw data at the physics level. It is a real-time event processor making an online decision if to keep an event or not. Accordingly, it filters out the ordinary and well understood events for the yield of rare and notable ones (e.g., pure-jet vs. four-lepton event). The selected events are processed again later with all corrections and calibrations in an offline stage, hence the online-offline naming.

A hardware stage called a Level-1 (L1) trigger reduces an event rate from 40 MHz to 100 kHz based on the prompt information from the calorimeters and muon chambers. A High Level Trigger (HLT) software stage then reduces the rate to 1 kHz in average [110] using the full information from the L1 trigger and the track reconstruction of the ID and MS.

The ATLAS trigger system underwent a significant change of its architecture between Run 1 and Run 2. The former three-stage system was simplified to two stages only, consisting of the hardware L1 and software HLT step. While the L1 hardware stage was upgraded with hardware components as a Topological Processor (TP) and enhanced Central Trigger Processor (CTP) employing the FPGAs instead of custom electronics [111]. Two Run 1 software stages a Level-2 (L2) and Event Filter (EF) were merged into the current HLT. The merger simplified the network interconnection which avoids a repetition of some reconstruction steps. Further details on a trigger optimization, algorithms, and strategies for Run 2 can be found in References [112–114].

The L1 trigger processes the information from the fast subsystem of MS (RPC, TGC) and from the whole calorimeter system. The signals are pre-processed in a muon CTP interface and in calorimeter cluster and jet/energy processors. The CTP then makes the L1 trigger decision, which signal is distributed to the detector readouts drivers (ROD), to control if the data should be moved up for further processing or if should release the pipeline of ROD. The L1 trigger primarily counts the multiplicity and compare the energy threshold of physics objects and test the L1 requirements. Although it can not take into account the full information of pile-up multiplicity, the TP extracts information of jets and muons direction in space to achieve additional background rate reduction [111]. At the end, the L1 trigger generates a list of affected parts of the detector worth of a full readout, called Regions of Interest (RoIs). The RoIs are handed over to a RoI builder, an interface between the L1 trigger and HLT [115].

The HLT stage operates only within the RoIs delivered by the L1 trigger. In general, the goal is to extend the event information by the contribution from slower, although precise subdetectors such as the inner and outer tracking, the ID and MS. The trigger stage also exploits a measurement coincidence, for example a track and calorimeter cluster matching or a combination of a muon and ID

tracks. A key architectural strategy of the HLT is a step-wise event selection. The selection is based on a series of trigger decision steps called a trigger chain. The ATLAS trigger consider many physics signatures such as the electrons, muons, jets, photons, $E_{\mathrm{T}}^{\mathrm{miss}}$ and τ-leptons, and b-physics. Each signature requests tens of chains, taking into account high-p_{T} objects, multiplicity, isolation, exclusivity, specific combination, etc. An example template of a chain could be an RoI multiplicity, calorimeter cluster multiplicity and energy threshold, track multiplicity, track-cluster matching, object selection. The HLT processing is based on a successive requests for new data from the RoI, to decrease a network bandwidth in case of early rejected events Each step in the chain performs a feature extraction on newly transferred data refining the already acquired information of the event. It can reject an event in early stage to release the buffer resources. Because of the event reconstruction has to face a busy environment of multiple proton-proton collisions, a vertex finding and an object-vertex matching plays a decisive role in an event building.

One of the essential feature extractions in the HLT is the ID tracking and associated vertex finding. Its processing is divided into two sections, a Fast Tracking (FT) and Precision Tracking (PT). The main algorithm of the FT handles a pattern recognition and is known as a Fast Track Finder (FTF). The hits inside of a RoI are at first recognized as clusters in the detector layers and then combined with the geometrical information of detector to provide three-dimensional spacepoints. Afterwards, the algorithm performs a spacepoint-triplets finding within a subset of three different ID layers, to create basic track primitives. The final track is constructed by a merger of the triplets consistent with the same trajectory. The FT is finalized by a preliminary track fitting, outlying hits rejection, and an extension to the TRT subdetector [116]. The FTF is designed to generate medium quality tracks as quickly as possible and to provide a track seeding for the subsequent second section, employing the precision algorithms. The PT is an algorithm optimized for speed, producing results close to the quality of offline algorithms. It resolves a hit-to-track association ambiguity and removes duplicate tracks via a track scoring, to finally perform an accurate track fitting. Once the PT complete the processing as well the trigger system can make the final HLT decision. The vertex finding algorithm is based on a histogram approach. The tracking primitives are extrapolated to the center of detector and each fills a histogram associated to the z axis of the detector with its z_0 impact parameter. A histogram bin with the highest number of counts is marked as a vertex candidate.

3.3.8 Data Acquisition

From the ATLAS detector length of 44 m, speed of light, and 40 MHz bunch-crossing rate it can be calculated, the detector is usually occupied by particles

originated in up to three bunch-crossings. This makes the detector readout rather challenging.

The ATLAS Data-Acquisition (DAQ) chain is composed, with respect to the direction from the active medium of the detector to a long term data storage of Front-end electronics (FE), a Readout system (ROS), data collection network, ATLAS output data storage, and a Worldwide LHC Computing Grid (WLCG),

The FE handles the basic signal processing as amplification, filtration, shaping, multiplexing, compression, zero-suppression (a device for suppression of a continual sending of empty information, to relax a band width occupation), and derandomization (a hardware pipeline to overcome bursts of potentially interesting events). The FE directly provides data to the L1 trigger for a fast evaluation. The last element of the FE chain is the ROD which receives the L1 trigger decision back and passes the acquired data to the following stage. The decision is also responsible for the full RoI readout.

The ROS is a front-end of the data acquisition at software level. It provides the data security, correction, and further multiplexing, and buffers the input using a Readout Buffer (ROB).

Right after the ROS, the data collection network follows. It collects the data from all subdetectors, and provides a coherent piece of information of a requested event. The information is transferred to the HLT trigger to enable the final HLT decision making. If an event is accepted the associated data are sent to a local data storage and around the world through the WLCG [117]. The full DAQ system with interconnection to the trigger system is depicted in Figure 3.15.

3.3.9 Luminosity

One of the crucial points for all analyses performed on the data acquired by the ATLAS detector is the measurement of proton-proton collisions luminosity. It is a necessary fundamental for a cross-section measurement of any process, moreover its total uncertainty represents a limiting factor which has to be suppressed as much as possible. The luminosity could be calculated straight from the machine parameters, although resulting in a sizable uncertainty. Therefore, it must be measured directly at the IP. The luminosity at IP1 is measured by several methods and detectors [118].

An Absolute Luminosity for the ATLAS (ALFA) detector delivers the information of the proton-proton total cross-section itself, which is necessary for the luminosity calculation.

The detector utilizes the optical theorem for the extrapolation of measurement of elastic proton-proton scattering in the forward region to the total cross-section $\sigma_{\text{p-p}}$. Thus the luminosity can be deduced using Equation (3.1) [119].

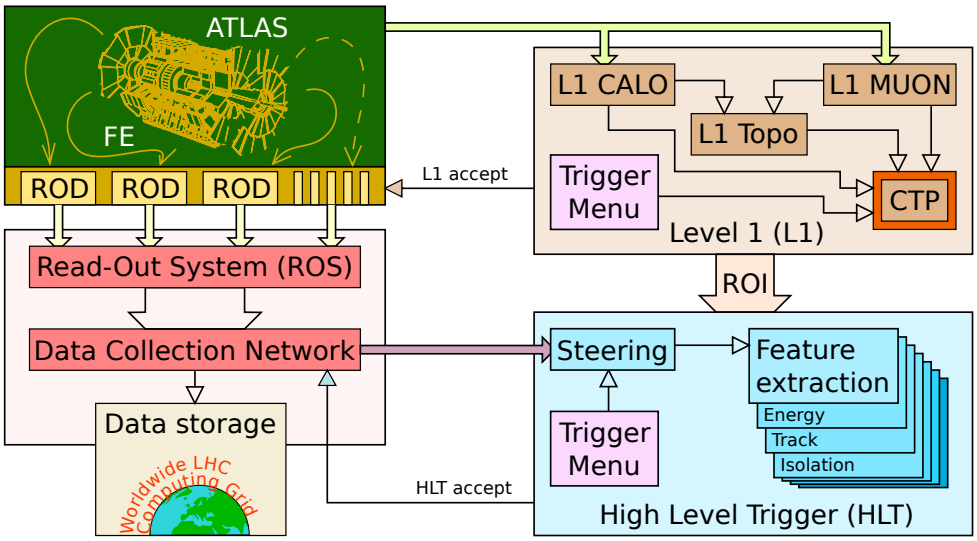


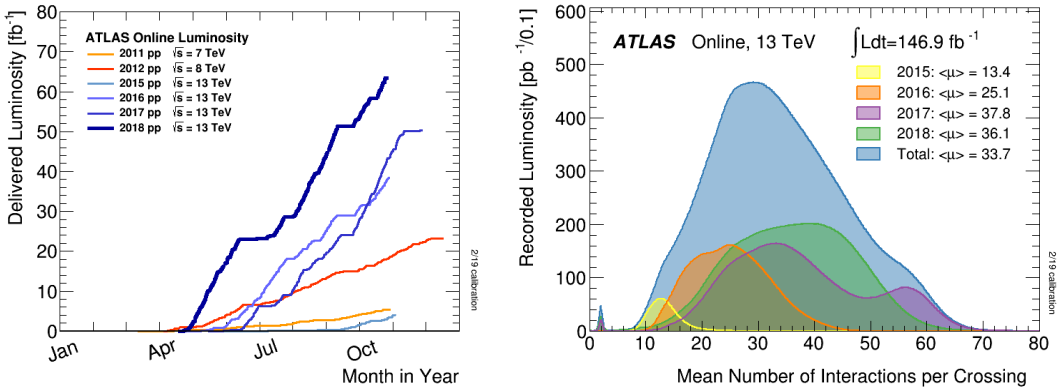
Figure 3.15: A schematic of the ATLAS DAQ and trigger interconnection. The DAQ chain is on the left side illustrating the data flow from the hardware FE up to the WLCG. The trigger architecture is depicted on the right side highlighting the two stage approach with the L1 trigger and HLT.

An instantaneous luminosity is measured by a Beam Condition Monitor (BCM) and a Luminosity Measurement Using Cerenkov Integrating Detector (LUCID). The BCM was originally installed as a protection of the PD against radiation dose caused by an accidental beam loss. The ATLAS primary luminosity monitor is the LUCID. According to the original design its purpose is a measurement of the instantaneous luminosity, together with a detection of the inelastic proton-proton scattering, and monitoring of the beam conditions as well.

The luminosity measurement is based on knowledge of the LHC revolution frequency, number of proton-proton collisions per bunch-crossing and a calibration constant in form of a cross-section. The last is calculated from the normalized beam density of beam in the horizontal and vertical planes using Equation (3.3). The density is measured in situ using a special regime of the LHC performing so called Van der Meer scan. It measures an interaction rate of the beams as a function of their displacement [120]. Granularity of the measurement is based on luminosity blocks (LB), a basic time unit for storing ATLAS luminosity information for physics. The blocks are defined by CTP and the duration is approximately one minute, depending on the experimental conditions [118]. The detector conditions are assumed to be constant during a whole LB. The measurement was upgraded to an enhanced LUCID 2 [121] version because of increased LHC center-of-mass energy of collisions to 13 TeV in Run 2. It brought an advanced luminosity integra-

tion and Run 2 luminosity determination with total uncertainty of only 1.7 % [122].

3.4 Data for analysis



(a) Cumulative plot of delivered luminosity by the LHC across years of data taking periods.

(b) Histogram of recorded luminosity by ATLAS detector as a function of mean number of inelastic interactions per bunch-crossing.

Figure 3.16: Delivered and recorded luminosity plots from multiple periods of the LHC operation [123].

The luminosity defines the amount of data that can be used for the analysis. The ATLAS collaboration measures both the delivered luminosity by the LHC and the recorded luminosity by the ATLAS detector, as can be seen in Figure 3.16. The dataset is enriched by different center-of-mass energies in Figure 3.16a shown for each year in both Run 1 and Run 2 periods. The recorded luminosity is shown in Figure 3.16b as a function of the mean number of pile-up interactions where one can see how the number increases across the years of Run 2. The recorded luminosity is smaller than the delivered one due to possible troubles in the detector operation and unavoidable circumstances. Although the ratio for Run 2 is 94 % that proofs a heavy-duty effort of the detector operation team and the whole collaboration. Once the data are collected the analysis itself can start, in the following chapter.

Chapter 4

Statistical Samples

The presented analysis must process extensive amount of data for a successful search for EW production of $ZZjj$, including VBS. The signal process is very rare and unfortunately it is overwhelmed by a sizable background.

The search is performed over a collection of recorded events, by the ATLAS detector, further referred to as real data or real sample. The real data are a primary input to the analysis and a place where the signal and background events compatible with the signal signature are contained.

A comparison of the real data with a theoretical model is necessary not only for the signal strength extraction, but also for a background estimate, detector systematic correction, and uncertainty determination. Therefore, simulated events must be processed in parallel with the primary input. Such events are produced by various event generators utilizing MC simulation of proton-proton collisions in accordance with the SM predictions. The simulated events are generated for individual processes, resulting in a number of simulated samples (referred to as MC samples).

Details of the real data sample are described in this chapter, the ATLAS data acquisition system is described in Chapter 3. Generation of the simulated samples is briefly described in this chapter, followed by a catalog of all samples utilized by the analysis.

4.1 Input Scheme

Raw ATLAS detector output is transformed to Event Summary Data (ESD), which contain complete information about the event reconstruction. Due to their huge size, ESD files are not suitable for physics analyses, therefore the data must be converted to a more flexible format. Extended Analysis Object Data (xAOD) format transforms the ESD into physical objects, directly readable by a data analysis

framework. The xAOD is a successor of simple AOD used during Run 1, improved with an ability to produce derivations (DAODs) of the original sample in order to make a number of smaller and more readily available samples tailored for particular groups of analyses [124]. The derivations are produced centrally by the ATLAS Derivation Production Group, so the analysis teams, sharing the same final-state, can benefit from it. Moreover, possible duplicate production by individual teams is avoided [125]. The presented analysis is considered as an EW research in a frame of the SM and is using "DAOD_STDM3" derivation tag.

The derivation used in this analysis includes events with at least two electron or muon leptons with p_{T} over 15 GeV and $|\eta|$ below 2.6 or events accepted by di-electron or di-muon triggers. The muons must pass at least basic track requirements [126] and the electrons must be matched to a cluster in the EM calorimeter (IsEMLoose) or be accepted by `LooseLH` identification working point [127]. The selection criteria correspond to, or are lower then, the signal selection criteria defined for the analysis.

The production steps of real samples are recorded in meta-data tags versioned by a four digit number. From the analysis perspective, two types of the tags are important:

- r1234 – reconstruction tag,
- p1234 – post processing tag,

where a number 1234 represents an example of tag version, the number can be even of five digits. The reconstruction tags record the conditions of detector during the data acquisition and the post-processing tag keeps information on all data included in the file, object containers, quality labeling, applied selection criteria, pile-up conditions etc. The "r" tag can be altered with "f" tag, stating the reconstruction was done by the WLCG root node (Tier-0) during the data acquisition. The tag is usually accompanied by "m" tag for Tier-0 merging, an intermediate step. A sample can be processed by more than one reconstruction or post processing steps. All information of any sample produced by the ATLAS collaboration can be found in ATLAS Meta-data Interface (AMI) database [128].

4.2 Recorded Data Samples

The presented analysis processes unprecedented amount of real events of proton-proton collisions at center-of-mass energy of 13 TeV. The data samples were collected during the whole Run 2 data-taking period, between years 2015 and 2018, excluding the runs before the LHC upgrade to 25 ns bunch-spacing, which were recorded in early 2015 with 50 ns bunch-spacing. The analysis feed utilize "good

run lists" for determination of the ATLAS data-taking runs with high data quality (DQ) usable for the physics research. The DQ signing is performed on per-LB basis, as well as the integrated luminosity measurement, see Section 3.3.9. All good samples together have total integrated luminosity of 139 fb^{-1}, broken down into years as shown in Table 4.1. The uncertainty in the integrated luminosity of the combined sample is 1.7 % [122], obtained using the LUCID-2 detector [121] for the primary luminosity measurements.

Period year	Integrated luminosity [fb^{-1}]	Physics runs
2015	3.2	66
2016	33.0	151
2017	44.3	186
2018	58.5	194

Table 4.1: Summary of data-taking periods of Run 2

The runs from the beginning and end of each data-taking period, as a representative sample, are listed in Table 4.2. In total, 597 runs were processed by the analysis framework.

Period	Run (period range)	Production steps
2015	279928	r9264_p3083_p4251
	284484	r9264_p3083_p4251
2016	297730	r9264_p3083_p4251
	311473	r9264_p3083_p4251
2017	325713	r10260_p3399_p4251
	340453	r10426_p3399_p4251
2018	348885	f937_m1972_p4251
	364292	f1002_m2037_p4251

Table 4.2: Summary of the ATLAS data-taking runs processed by the analysis framework. The listing shows the number of events taken during the run and production tags, including the AMI links.

4.3 Theory Prediction

A firm theory prediction is crucial for search analysis of a process with such small cross-section as the $ZZjj$ EW production. Moreover, the analysis represents a test of the SM, therefore the prediction accuracy is essential for evident confirmation (or exclusion) of the signal hypothesis.

4.3.1 Monte Carlo Simulation

The following description of MC simulations is very brief, oriented in the way of description of the method inputs rather than the embedded algorithmic techniques, for the sake of the experimental physicist point of view. A thorough description can be found in [129, 130].

The MC simulation method operates on stochastic variables in accordance with the particle physics nature. A fundamental element of the method is represented by the probability density function of object or process variables, from which the probability of obtaining a value from a certain interval can be deduced. The method utilizes the functions for the random sampling of variables, to make a decision on a particle decay, radiation, splitting, and new particle creation.

The simulation starts with generation of a hard process, which is described by squaring of the corresponding matrix element (ME) in case of leading order (LO) approximation. In higher orders, more advanced techniques are needed in order to get rid of ultra-violet divergences by employing regularization and renormalization techniques [131]. In case of hard proton-proton collisions, like at LHC, the particles entering the ME are not directly protons but their constituents, partons (quarks and gluons). In such collisions, proton is described by a set of Parton Distribution Functions (PDF) which give the probability of finding a parton of given flavor carrying a given fraction of proton's longitudinal momentum [132]. PDFs must be measured, since can not be derived from first principles from QCD Lagrangian as strong non-perturbative effects are involved. For many processes at the LHC the PDF uncertainties represent a significant part of uncertainty in the theoretical cross-section calculations.

Due to mass singularities presented in the QCD, the hard partons entering the ME radiate intensively additional partons. These parton showers can be still described perturbatively [133–135] and can be implemented in MC generators as stochastic process. Parton radiation makes PDFs dependent not only on fractional momentum x but on the virtuality of parton entering/leaving the hard process. The evolution in virtuality is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [136–139]. The parton showering depends on the hardness of interaction represented by the transferred momentum between protons. A hard interaction can produce high momentum quarks and subsequent

gluon radiation, resulting in parton showers driven by gluon self-coupling. The phenomena is described by the Parton Splitting Functions.

The hadronization is the late state of evolution of the simulated process, when the hadrons are formed out of the quarks and gluons, the perturbative QCD becomes invalid, and non perturbative approach must be employed in a form of phenomenological fragmentation models (Lund string [140, 141] used by PYTHIA, Cluster hadronization [142, 143] used by HERWIG, etc.). The functions encode the information on how partons produced in hard-scattering processes are turned into an observed colorless hadronic bound final-state [144].

The perturbative quantum field theory enters the simulation through the ME and corresponding higher order corrections.The analysis considers only the QCD corrections, however the EW corrections are starting to be studied as well [145]. The ME can be calculated in the first approximation at the leading order accuracy, using the tree-level feynman diagrams, usually sufficient for the modeling of distribution shapes. For a proper modeling of scale, higher order diagrams are necessary. Accuracy of next-to-leading order (NLO), or even next-to-next-to-leading order (NNLO) can be achieved by inclusion of all diagrams in the ME or usage of lower order diagrams with a subsequent re-weighting to the desired order. The latter case can lead, at first glance, to a whimsical result as a zero-jet process at NLO accuracy in QCD.

A MC generator tune represents the settings of input parameters optimized for the best agreement of simulated observables with the real data. A summary of validation and tuning of the MC generators for the LHC can be found in [146]. The settings are adjusted in several steps for initial and final state radiation of ME (start of the parton showering), considering underlying event, beam remnants, color re-connections using minimum bias events, and energy scaling [147]. An example of generator tune used in PYTHIA is called Perugia tune [148].

The MC simulation of proton-proton collision is a complex process. For the better understanding of the individual phases and mechanisms a representative illustration based on Feynman diagrams is shown in Figure 4.1. The illustration emphasizes a difference in decay complexity of an EW boson (~~~~), which is driven by the properties of the object, and a gluon (ooooo), driven by the showering and hadronization where the QCD itself takes a significant part (asymptotic freedom, gluon self-coupling, color charge, confinement etc.).

To sum up all inputs to the MC simulation, these parameters must be under control:

- renormalization and factorization scales, such as the QCD scales,
- parton distribution functions and splitting functions, as the perturbative input,

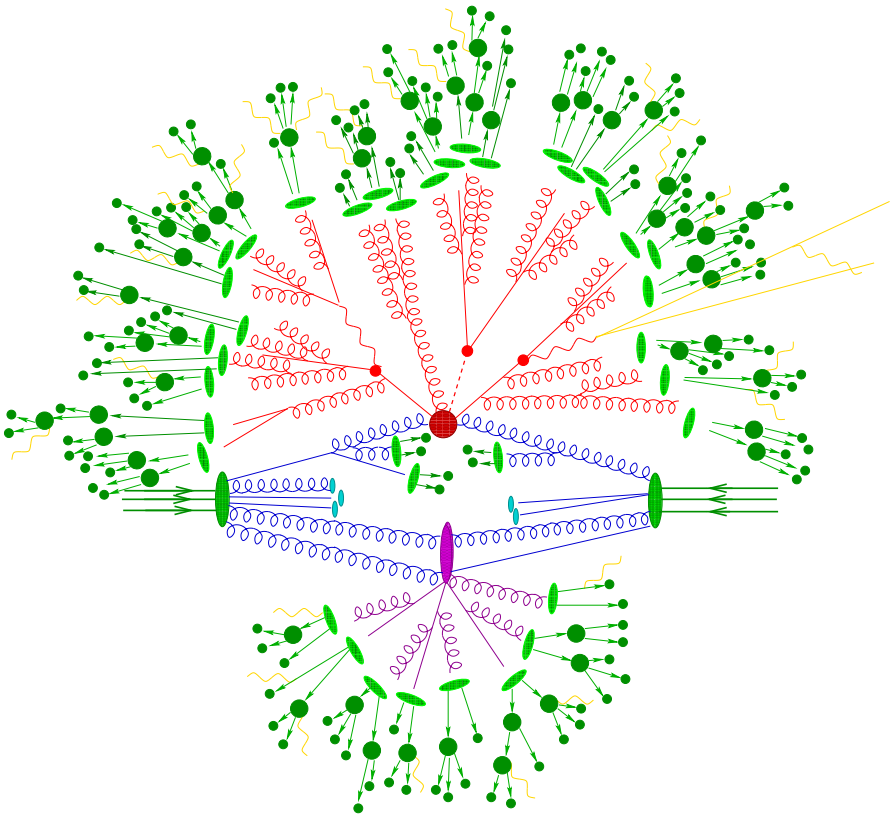


Figure 4.1: Diagrammatical illustration of a simulated proton-proton collision generated by the SHERPA event generator. Quarks and gluons (blue) of the beam protons (biggest green blobs) scatter creating a hard interaction (big red blob) and an underlying event or multi-parton interaction (purple blob). The emergence of color neutral remnants of the original protons is allowed through the hadronization (light green ellipses). The hard interaction produces new particles (red), radiating gluons driven by the parton showering. The particles decay into final-state leptons (yellow lines) or hadronize creating jets and photon radiation (yellow waves). In the end, the unstable hadrons can decay (dark green blobs). The illustration shows a $t\bar{t}H$ production as can be identified from the quarks and boson decays (small red blobs) [149].

- hadronization scale, minimum perturbative cutoff, collinearity cut-off, and fragmentation functions, as the non-perturbative input,
- and a broad array of parameters of the event generator framework, such as the generator tune.

The described event generation involves the proton-proton collision and immediate decays to the final state particles. It usually represents only the first step towards the theory prediction usable in a search analysis with the ATLAS detector,

because of another simulation is necessary to complete the picture, discussed in the next section. However, already the truth information can be used, for example to the phase space determination.

4.3.2 ATLAS Event Simulation

A simulation of the interaction of generated event final-state particles with the detector, projects the experimental setup into the MC simulation at reconstruction level. This step is necessary to make the artificial simulated events comparable with the measured data in the ESD format, enabling to model the systematic uncertainties. For this purpose, the ATLAS Monte Carlo Simulation Infrastructure was established [89].

The infrastructure includes the following simulation inputs and tools: a geometrical and material model of the ATLAS detector prepared in GEANT 4 (a simulation tool) [150], a simulation conditions database, detector calibration, event generator job options and interfaces, and an interconnection with the large-scale production infrastructure like the WLCG.

The model includes all detector aspects, especially the granularity of all sensitive parts (pixels, strips, tubes, tiles), the discontinuity of the measurement in a form of the layers in the ID and MS, and the detector volumes of calorimeters. Each volume has defined construction matter which enables a proper simulation of electromagnetic and hadronic shower creation in absorbers and detection process of scintillation or ionization. The transition radiation emission process is also included in the detection simulation. Detector geometry imperfections like barrel-endcap transition region or zero-eta cable shafts are modeled as well.

The simulation of event detection is even able to project realistic measurement calibration and conditions of the detector, as are the differences in temperature, misalignment, distortions, detector noise, and dead sensor elements. The LHC beam properties are also taken into account, including the event overlay for pile-up simulation, bunch crossing frequency, and in-time pileup.

The last step of the simulation of event passage through the detector is digitization. A process of transformation of the energy deposited in the sensitive regions of the detector into electric signal, in the same format as the real event would generate. It brings a great advantage in a sense that the same readout algorithms can be used for the simulated and real events, including the full reconstruction chain of the ATLAS DAQ and Trigger.

Besides sensors, the ATLAS model includes the parts of detector support systems, magnetic subsystem, and construction elements. It also represents dead material which can cause unwanted effects like *bremsstrahlung*, which must be simulated as well. The model is comprehensive including the detector skeleton,

sensor holding staves, magnet coils, cooling infrastructure, cable trays and shafts, but even single cables and circuit boards of readout electronics.

All simulation production steps can be summarized in three main phases with following sets of processes

- event generation - from the collisions, across the ME, parton showering, and hadronization to the final states,
- detector response - from the generated event passage through the detector, scintillation, and ionization to the signal digitization,
- reconstruction - from the detector readout to the event reconstruction chain of ATLAS DAQ and Trigger.

The simulated samples are described with the same meta-data tags as the real samples, described in the Section 4.1, although enriched with

- e1234 – event generation tag,
- s1234 – and simulation tag,

preceding the reconstruction and post-processing tags.

4.4 Simulated Samples

Due to the large time-scale of the real data collection and various pile-up multiplicities of the LHC collision events across the data collection years, the resulting MC sample is composed of several different pile-up profiles. The profiles are defined by the rates of pile-up multiplicities applied for runs in each data-taking period, as can be seen in Figure 3.16b.

While profiles for years 2015 and 2016 were combined in a single MC simulation campaign called "mc16a", another two profiles for years 2017 and 2018 are used separately for campaigns "mc16d" and "mc16e". Therefore, the simulation must be processed three times to realistically model the real combined sample. A summary of the pile-up profiles used for MC simulations is shown in Table 4.3, highlighting the production tags. It can be noticed that the tags of the real samples in Table 4.2 correspond to the pile-up profile "r" tags of identical year. Although, with exception of the year 2018, where the data were not re-processed, the first processing from Tier-0 is still used.

The existence of three pile-up profiles demands repetition of the simulation of individual MC samples as well as execution of the analysis code. Final combination is weighted by the integrated luminosity of each corresponding data-taking period, as shown in Table 4.1. The huge amount of processed data places an additional challenge on the analysis framework presented in this document.

Period year	Pile-up profile	Reconstruction tags
2015–2016	mc16a	r9364, r9315
2017	mc16d	r10201, r10210
2018	mc16e	r10724, r10726

Table 4.3: Summary information for pile-up profiles of Run 2

.

4.4.1 Signal Samples

The MC samples of the EW $ZZjj$ production are simulated using two different options, utilizing the MadGraph or Sherpa as the event generator, resulting in two independent samples. The MadGraph settings utilizes the Pythia for the parton showering and hadronization and EvtGen for the heavy flavor decays modeling. The second settings utilize the Sherpa generator for all simulation steps. The exact version of the individual tools applied for the simulation is listed in Table 4.4. Further details can be decoded from the production steps and a dataset number corresponding to the generator job options. The process description in the table considers leptons marked as ℓ to be electrons, muons, or taus (e, μ, or τ) or a corresponding anti-particle, unless stated otherwise. Tables mentioned later in the text may consider also ν to be (anti)neutrino from any family.

Dataset	Process	Generator(s)	Tune/PDF	Production steps
345071	$q\bar{q} \to \ell\ell\ell\ell jj$	MadGraph5_aMC@NLO(v2.3.3.p1) Pythia8(v8.212) EvtGen(v1.6.0)	A14 NNPDF3.0 LO	e5994_e5984_s3126_r9364_r9315_p3761 e6834_e5984_s3126_r10201_r10210_p4097 e6834_e5984_s3126_r10724_r10726_p4097
364283	$q\bar{q} \to \ell\ell\ell\ell jj$	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e6055_e5984_s3126_r9364_r9315_p4097 e6055_e5984_s3126_r10201_r10210_p4097 e6055_e5984_s3126_r10724_r10726_p4097
364250	$q\bar{q} \to \ell\ell\ell\ell$	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5894_s3126_r9364_r9315_p4097 e5894_e5984_s3126_r10201_r10210_p4097 e5894_e5984_s3126_r10724_r10726_p4097
345706	$gg \to \ell\ell\ell\ell$	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e6213_e5984_s3126_r9364_r9315_p4097 e6213_e5984_s3126_r10201_r10210_p4097 e6213_e5984_s3126_r10724_r10726_p4097
361603	$q\bar{q} \to \ell\ell\ell\ell$	Powheg(v2_r2819) Pythia8(v8.210) EvtGen(v1.2.0)	AZNLO CTEQ6L1	e4475_s3126_r9364_r9315_p4250 e4475_e5984_s3126_r10201_r10210_p4250 e4475_e5984_s3126_r10724_r10726_p4250

Table 4.4: Summary of the ZZ MC samples processed in this analysis framework. The listing includes the generated process description, utilized event generators, and their settings. Full information for the direct identification of datasets in the collaboration database is included, i.e. job option and AMI links.

The signal samples contain Feynman diagrams consisting of six EW couplings and zero QCD couplings, including the diagrams of Higgs boson exchange with off-shell Higgs boson, for details see Figures 2.3, 2.2 and 2.6a. The signal boson pairs decay to charged leptons, however, both samples include also a tri-boson case when the third boson decays hadronically into two jets. Events of the primary sample are generated with a ME at LO in perturbative QCD, using A14 tune and NNPDF3.0 LO PDF set [151]. A four-lepton invariant mass filter at the generator level is set at 130 GeV. The secondary sample was also generated at LO using NNPDF3.0 NNLO PDF set.

The QCD ZZ production MC sample is simulated using the SHERPA generator with a ME calculated at NLO in QCD with up to one jet and at LO with two or three jets. The sample includes only the quark-induced processes, and is used as a primary MC sample for modeling of the QCD $ZZjj$ production. The NNPDF3.0 NNLO PDF set is applied here as well. A di-lepton mass filter at the generator level is set at 4 GeV and the leading and sub-leading lepton p_{T} is required to be over 5 GeV. The last selection requirement is overlapped by the derivation selection criteria, so is mentioned here only to complete the information, and will not be mentioned again in the text.

The secondary sample for QCD ZZ production consists of gluon-induced processes and is produced by SHERPA generator. The sample is generated with a ME at LO in QCD with up to one jet using the NNPDF3.0 NNLO PDF set. The sample includes gluon-gluon fusion and interference diagrams. The four-lepton mass filter at the generator level is set at 130 GeV. The sample is reweighted by the analysis framework to NLO accuracy using a factor of 1.67 [152].

As a cross-check, another ZZ production MC sample was chosen. The sample is generated in POWHEG with cross-section precision at NLO in QCD (with no jets). The generator is set to AZNLO CTEQ6L1 tune [153], using the CT10 NLO PDF set. The di-lepton mass filter at the generator level is set at 4 GeV This sample contains only events with jets from the parton showering, therefore is suitable only for comparison studies without jet selection.

4.4.2 Background Samples

A number of MC simulated samples is used for the background estimation and validation of utilized methods. All samples are generated by the SHERPA generator using NNPDF3.0 NNLO PDF set, with exception of some $t\bar{t}$ cases.

The $WZjj$ production sample is generated with a ME calculated at LO in QCD and the WZ production sample at NLO with up to one jet and at LO with two or three jets. The samples are used for both background estimation and validation, listed in Table 4.5.

The WWZ, WZZ, and ZZZ MC samples, further called tri-boson samples,

Dataset	Process	Generator	PDF	Production steps
364284	$q\bar{q} \to \ell\ell\ell\nu jj$	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e6055_e5984_s3126_r9364_r9315_p4250 e6055_e5984_s3126_r10201_r10210_p4097 e6055_e5984_s3126_r10724_r10726_p4097
364253	$q\bar{q} \to \ell\ell\ell\nu$	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5916_s3126_r9364_r9315_p4097 e5916_e5984_s3126_r10201_r10210_p4097 e5916_e5984_s3126_r10724_r10726_p4097

Table 4.5: Summary of the WZ background MC samples. The listing includes the generated process description, utilized event generators, and their settings. Full information for the direct identification of datasets in the collaboration database is included, i.e. job option and AMI links.

are generated using a NLO ME in QCD with up to one jet and a LO ME with two jets. The samples are used exclusively for the background estimation and are listed in Table 4.6.

Dataset	Process	Generator	PDF	Production steps
364243	$q\bar{q} \to WWZ$ ($4\ell 2\nu$)	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5887_s3126_r9364_r9315_p4097 e5887_e5984_s3126_r10201_r10210_p4097 e5887_e5984_s3126_r10724_r10726_p4097
364245	$q\bar{q} \to WZZ$ ($5\ell 1\nu$)	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5887_s3126_r9364_r9315_p4097 e5887_e5984_s3126_r10201_r10210_p4097 e5887_e5984_s3126_r10724_r10726_p4097
364248	$q\bar{q} \to ZZZ$ ($4\ell 2\nu$)	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5887_s3126_r9364_r9315_p4097 e5887_e5984_s3126_r10201_r10210_p4097 e5887_e5984_s3126_r10724_r10726_p4097
364247	$q\bar{q} \to ZZZ$ ($6\ell 0\nu$)	Sherpa(v2.2.2)	NNPDF3.0 NNLO	e5887_s3126_r9364_r9315_p4097 e5887_e5984_s3126_r10201_r10210_p4097 e5887_e5984_s3126_r10724_r10726_p4097

Table 4.6: Summary of the triboson background MC samples. The listing includes the generated process description, utilized event generators, and their settings. Full information for the direct identification of datasets in the collaboration database is included, i.e. job option and AMI links.

The Z + jets MC samples are generated at cross-section accuracy of NLO in QCD with up to two jets and at LO with up to four jets, the reweighting to the higher order is done at the generator level. The process is sliced according to the jets flavor and highest value of a vectorial and scalar sum of p_{T} of all final state particles, including jets. The slicing results in a number of MC samples, therefore only a few representative ones are listed in Table 4.7, for each decay channel of simulated Z boson ($\ell^+\ell^-$). The samples are used for the validation of misidentified

background estimation.

Dataset range	Process	Generator	PDF	Production steps
364100–364113	$q\bar{q} \to Z/\gamma^* + \text{jets}$ (ee)	Sherpa(v2.2.1)	NNPDF3.0 NNLO	e5271_s3126_r9364_r9315_p4252 e5271_s3126_r10201_r10210_p4097 e5271_e5984_s3126_s3136_r10724_r10726_p4097
364114–364127	$q\bar{q} \to Z/\gamma^* + \text{jets}$ $(\mu\mu)$	Sherpa(v2.2.1)	NNPDF3.0 NNLO	e5299_s3126_r9364_r9315_p4252 e5299_s3126_r10201_r10210_p4252 e5299_e5984_s3126_s3136_r10724_r10726_p4097
364128–364141	$q\bar{q} \to Z/\gamma^* + \text{jets}$ $(\tau\tau)$	Sherpa(v2.2.1)	NNPDF3.0 NNLO	e5307_s3126_r9364_r9315_p4097 e5307_e5984_s3126_r10201_r10210_p4250 e5307_e5984_s3126_s3136_r10724_r10726_p4097

Table 4.7: Summary of a range of the Z + jets MC samples. The listing includes the generated process description, utilized event generators, and their settings. Full information for the direct identification of datasets in the collaboration database is included, i.e. job option and AMI links.

The MC samples of $t\bar{t}$ production are generated by various generators as can be seen in Table 4.8. The bare production is generated by POWHEG utilizing PYTHIA and EVTGEN with a NLO ME in QCD, using A14 tune and NNPDF2.3 LO PDF set [154]. The productions including one or two W bosons are generated by MADGRAPH with the same tandem of supporting tools and the tune and PDF set as in the previous case. The QCD accuracy of the diboson case is at NLO and the other at LO. The $t\bar{t}Z$ production is generated by SHERPA with a ME calculated at LO in perturbative QCD, using NNPDF3.0 NNLO PDF set. The bare production sample is used for the validation of the misidentified lepton background estimation. The other samples are included in the irreducible background estimation.

The signal samples in Section 4.4.1 are used also for estimation of background originating from $Z \to \tau\tau$ decay.

Dataset	Process	Generator(s)	Tune/PDF	Production steps
410472	$pp \to t\bar{t}$	Powheg Pythia8(v8.230) EvtGen(v1.6.0)	A14 NNPDF23LO	e6348_e5984_s3126_r9364_r9315_p4097 e6348_e5984_s3126_r10201_r10210_p4097 e6348_e5984_s3126_r10724_r10726_p4097
410142	$pp \to t\bar{t} + Z/\gamma^* \ (\ell\ell)$	Sherpa(v2.2.0)	NNPDF3.0 NNLO	e4686_e5984_s3126_r9364_r9315_p4097 e4686_e5984_s3126_r10201_r10210_p4097 e4686_e5984_s3126_r10724_r10726_p4097
410155	$pp \to t\bar{t} + W$	MadGraph5_aMC@NLO(v2.3.3.p0) Pythia8(v8.210) EvtGen(v1.2.0)	A14 NNPDF23LO	e5070_s3126_r9364_r9315_p4097 e5070_e5984_s3126_r10201_r10210_p4097 e5070_e5984_s3126_r10724_r10726_p4097
410081	$pp \to t\bar{t} + WW$	MadGraph5_aMC@NLO(v2.2.2.p6) Pythia8(v8.186) EvtGen(v1.2.0)	A14 NNPDF23LO	e4111_s3126_r9364_r9315_p4097 e4111_e5984_s3126_r10201_r10210_p4097 e4111_e5984_s3126_r10724_r10726_p4097

Table 4.8: Summary of the $t\bar{t}$ MC samples. The listing includes the generated process description, utilized event generators, and their settings. Full information for the direct identification of datasets in the collaboration database is included, i.e. job option and AMI links.

Chapter 5

Analysis Setup

The primary goal of this analysis is the observation of EW production of two Z bosons and a pair of jets in the four charged lepton channel. The signal must be extracted using an advanced statistical method, as it is overwhelmed by a large background.

As the first step towards the goal, a proper understanding of the signal and background must be achieved. Afterwards, an event selection enhancing the signal and suppressing the background contribution can be defined, using physical object kinematic requirements. The phase space defined by a set of requirements is called Signal Region (SR). Once further kinematic requirements become inefficient or divert too large portion of the signal from the SR, the input observable for the statistical method must be defined. Advanced selection methods are exploited for the definition using various Multivariate Analysis (MVA) techniques.

This chapter describes the analysis strategy, its origin, and signal-to-background comparison.

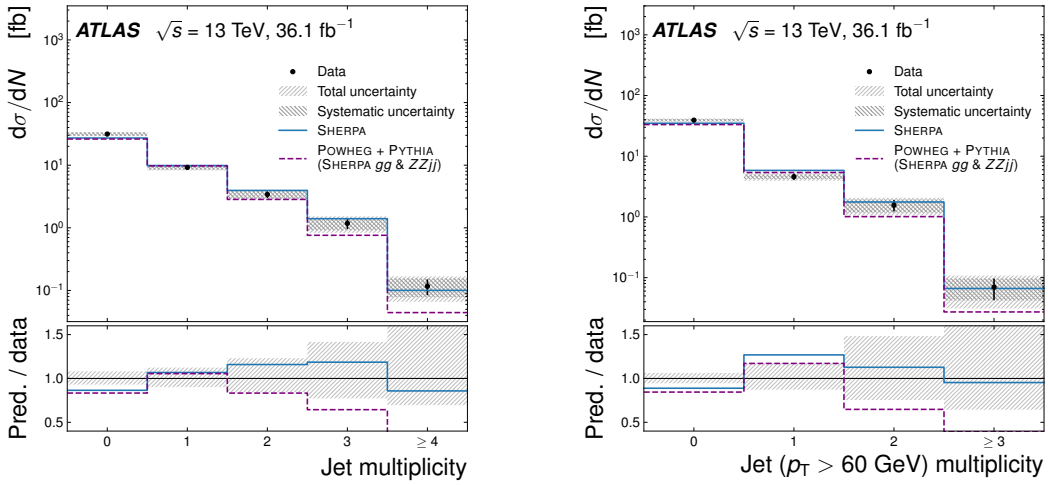
5.1 Inclusive ZZ Production

Since the final-state of interest consists of two Z boson and two jets, an inclusive ZZ production represents a natural starting point for the studies. The boson selection can be adopted with minor optimizations to maximize the signal yield, but the jet selection must be added for the VBS signature identification. The focus on EW production dramatically changes the original analysis since a dominant part of the inclusive ZZ production signal becomes a background overwhelming the signal.

The presented analysis is built up on top of a framework employed for the inclusive ZZ production cross-section measurement by the ATLAS collaboration with 36.1 fb^{-1} [20] of data (2015-2016), where the author significantly contributed.

Not only the similar final-state makes the ZZ production an ideal start, it also provides a good insight into utilized analysis techniques and serves as a good guide because of similar analysis work-flow.

The ZZ measurement also offers the first insight into the EW production study by examination of ZZ cross-section as a function of jet multiplicity, shown in Figure 5.1. The cross-section with two or more jets is of the interest, representing the upper limit of the expected result. As was anticipated, the process is very rare with the cross-section of approximately 4 and 1.5 fb, using all jets (Figure 5.1a) and jets with p_{T} over 60 GeV (Figure 5.1b) respectively.



(a) **(b)**

Figure 5.1: Cross-section of inclusive ZZ production as a function of jet multiplicity (a) and multiplicity considering only jets with p_{T} over 60 GeV (b). The histograms are taken from the starting point analysis [20].

5.2 EW ZZ Production

As is discussed in Section 2.2.1, the EW production of the bosons must be accompanied by two jets as a VBS signature. Typical observables used for VBS search analyses are a di-jet invariant mass m_{jj}, di-jet constituents separation in terms of rapidity defined in equation

$$\Delta y_{jj} = |y_{j_1} - y_{j_2}|,$$

and product of their rapidity values ($y_{j_1} \times y_{j_2}$). The latter requirement is in fact used for an identification of events with the jets at the opposite sides of the detector

by requesting a negative value. A survey of jet requirements used in other EW production channels (WW, WZ, and semi-leptonic VV) is shown in Table 5.1. For a complete comparison of the channels studied in the ATLAS collaboration, see [40]. The semileptonic channel considers one of the bosons to be decaying hadronically to two jets, the V is interchangeable with a W or Z boson.

	Jet selections		
VBS Channel	m_{jj} [GeV]	Δy_{jj} [–]	$y_{j_1} \times y_{j_2}$ [–]
$WWjj$	500	> 2	–
$WZjj$	500	–	< 0
$VVjj$ (semi)	400	–	< 0

Table 5.1: A survey of jet selection requirements used in other VBS channels than the ZZ, based on [40].

The survey helps with a choice of a set of observables most efficient for the selection of $ZZjj$ events. However, a direct comparison of the signal and QCD background, in the next section, suggests the final choice of the set and determination of the threshold values.

5.3 MC Truth Analysis

Before the search can be conducted, a thorough understanding of theoretical prediction of the target phenomena is necessary. The prediction is accessible through the MC simulations described in Chapter 4. In search for phenomena with a small cross-section, the simulated MC samples can be available faster than the real data with sufficient statistics can be collected. This is true especially in the case demanding full Run 2 data sample, just like in the presented analysis. The analysis of MC samples provides identification of the typical signature of signal and background processes, validation and optimization of the selection criteria. At a certain level, the simulated data-sample availability in advance naturally removes the observation bias.

5.3.1 Signal to Background Comparison

The comparison takes into account the EW production as the signal and the quark-induced QCD production as the most significant background. A preliminary phase space is composed of the inherited ZZ selection criteria and a requirement for at

least two accompanying jets for the follow-up study. The presence of two jets is a necessary prerequisite for a construction of the di-jet system for evaluation of the VBS signature described in the previous section. The di-jet is constructed from the two leading p_{T} jets of an event.

The comparison studies shapes of observables using a normalization of the distributions to unity. The set of observables listed in Table 5.1 is compared in Figure 5.2, together with a distribution of the leading p_{T} jet.

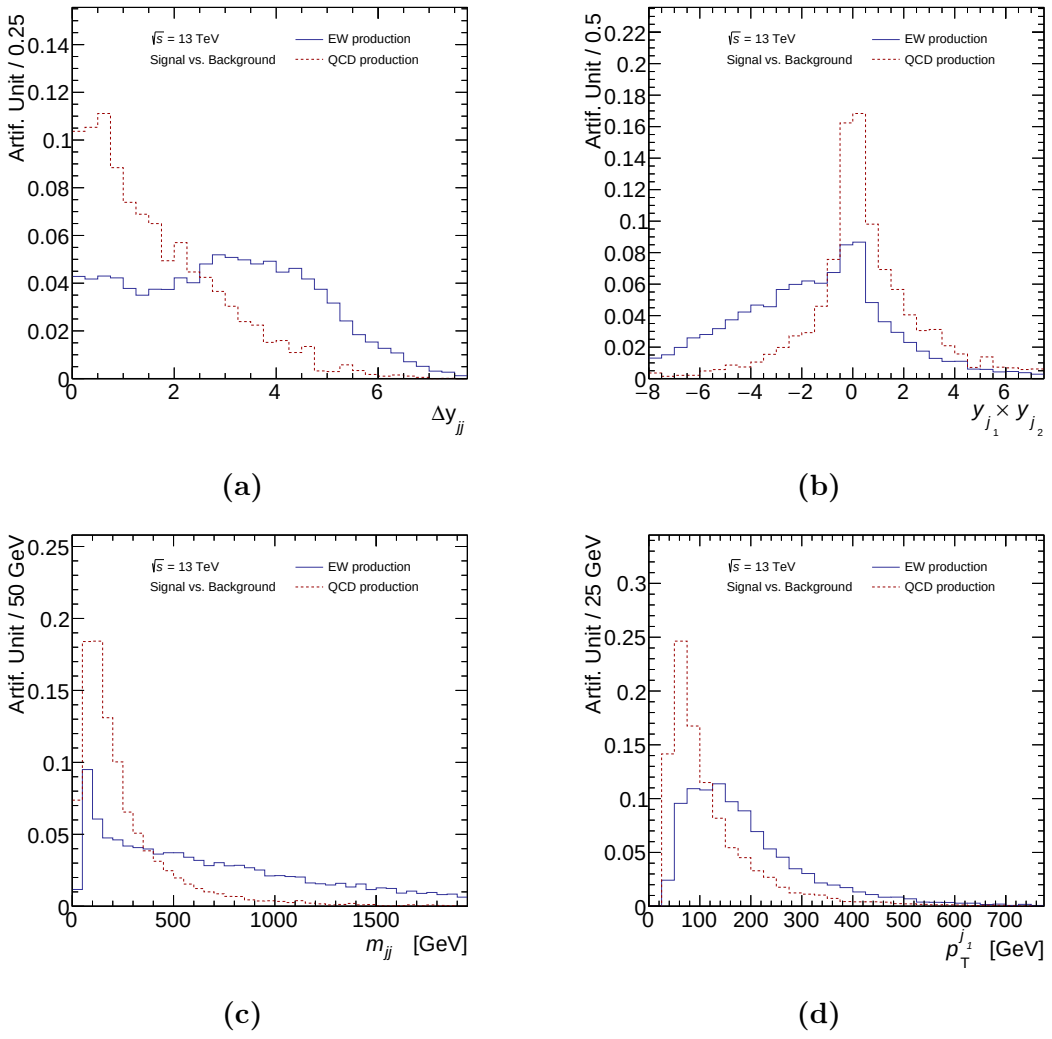


Figure 5.2: Shape comparison of the EW signal and quark-initiated QCD background distributions of MC simulated samples. Distributions as a function of Δy_{jj} (a), $y_{j_1} \times y_{j_2}$ (b), m_{jj} (c), and leading jet p_{T} (d) are shown. Distributions are normalized to unity.

The Figure 5.2a clearly shows the signal is enhanced when the di-jet constituents are detected at the opposite sides of the detector. It corresponds to the expectation of VBS events being driven by EW interaction. The shape comparison of rapidity separation in Figure 5.2b supports the previous statement as well. Both of the selection requirements in Table 5.1 corresponding to the findings are adopted "as is", for the ZZ channel. Simultaneity of both requirements can comfortably overcome the background peak around zero in $y_{j_1} \times y_{j_2}$ distribution. The di-jet mass comparison in Figure 5.2c proves the requirement as a reliable selection criteria. The last distribution of the comparison shows, that properties of single objects such as p_{T} are also worth of considering in the selection. This conclusion will be utilized later in the MVA based selection.

Finally, the requirements of VBS tagging di-jet are chosen to be $y_{j_1} \times y_{j_2} > 0$, $\Delta y_{jj} > 2$, and $m_{jj} > 300\,\mathrm{GeV}$. Due to the small $ZZjj$ cross-section, it was decided to take a smaller m_{jj} value as a threshold, in comparison to the other channels.

5.3.2 Signal Misconfiguration

The signal MC sample plays a crucial role in the research, since it models the $ZZjj$ EW production, thus it deserves a special attention. As was stated in Chapter 4, two centrally produced samples by the ATLAS collaboration (MadGraph, Sherpa) are available. During a WW MC modeling for VBS research in Reference [155], a problem with non-optimal settings of color flow in the MC prediction produced by Sherpa was discovered. Since the Sherpa MC sample stated in Table 4.4 uses the same version of the generator, an additional comparison is necessary to check the impact of the problem in the ZZ VBS channel. A comparison of both samples in terms of the aforementioned tagging jet observables is shown in Figure 5.3 at the top, together with plots from the modeling for WW VBS channel at the bottom.

The problem of Sherpa misconfiguration is manifested similarly in the ZZ VBS channel. The low mass region of m_{jj} distribution is overestimated, while the high mass region is underestimated. The Δy_{jj} distribution is overestimated near a value of two, while overestimated over three in absolute value. The trends of disagreement in both channels are very similar, thus the MadGraph MC sample is chosen as a theory prediction of signal for this analysis. The Sherpa MC sample will be utilized for systematic uncertainty determination in the later stage.

5.3.3 EW and QCD Production Interference

Both EW and QCD production have the same final-state, therefore the corresponding processes are not easy to distinguish. Moreover both processes interfere with each other.

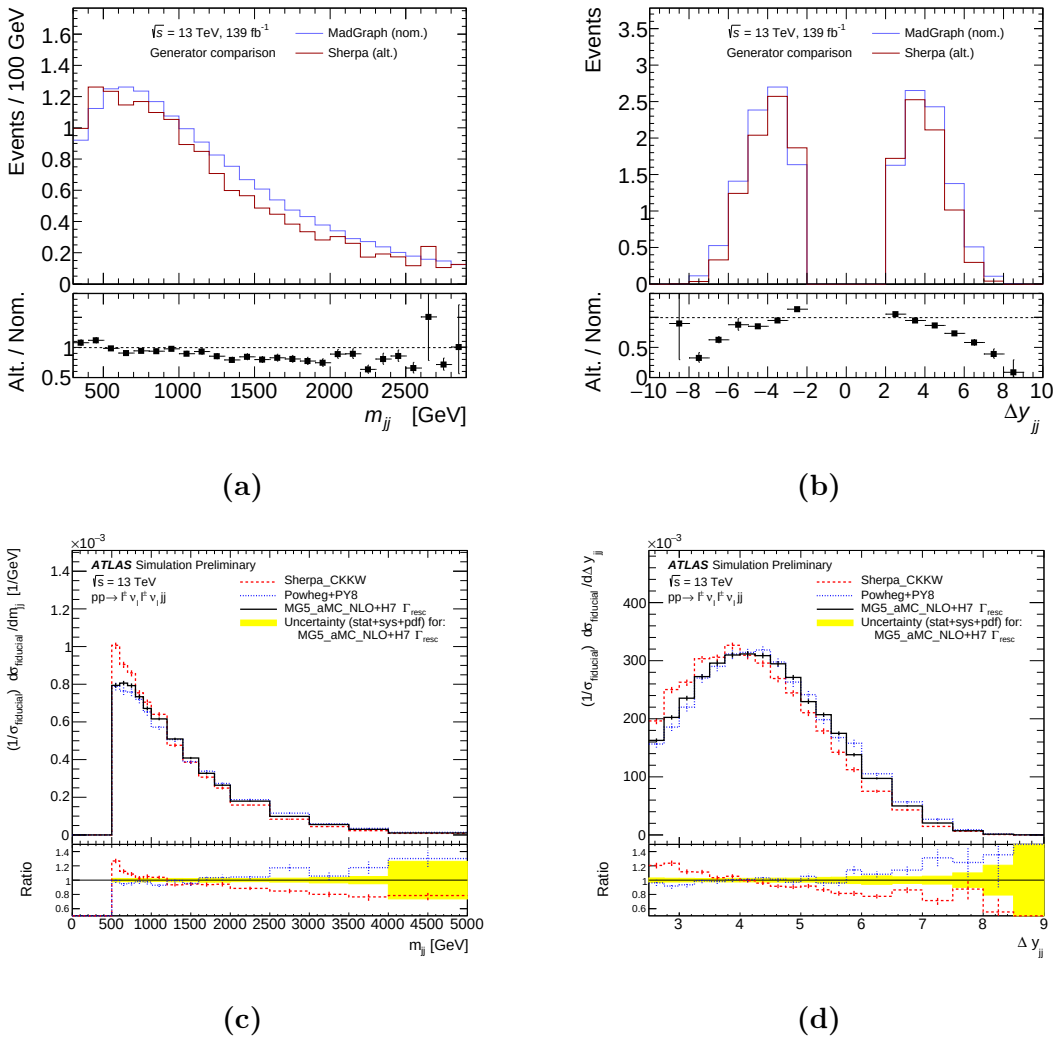


Figure 5.3: Comparison of MC simulated samples produced by the MadGraph and Sherpa generators. A comparison for $ZZjj$ channel is shown at the top, as a function of di-jet mass (a) and di-jet constituents rapidity separation (b). A comparison complemented also with the Powheg generator for $WWjj$ channel is shown at the bottom (c) and (d) [155].

Interference at the leading order arise from the squaring of matrix element including both processes, as a real part of the following formula:

$$|M_{\mathrm{EW}} + M_{\mathrm{QCD}}|^2 = |M_{\mathrm{EW}}|^2 + |M_{\mathrm{QCD}}|^2 + 2\mathrm{Re}(M_{\mathrm{EW}} + M_{\mathrm{QCD}}), \qquad (5.1)$$

where the M_{EW} and M_{QCD} represent matrix elements of EW and QCD production. A standard approach for determination of the interference amplitude is based on a

subtraction method. Unfortunately, the method requires a MC simulation of each case: EW, QCD, and inclusive production, which represents a bottleneck of the method, due to a CPU time expensive simulation.

An alternative approach exploits a relatively new feature of the MadGraph event generator, which is able to produce the interference MC simulated sample directly. The method is based on an event selection at the generator level, using properties of the ME. A theoretical interpretation of the interference modeling depends on counting of a QCD coupling order at the amplitude level. Using the coupling notation from Section 2.4, the M_{QCD} is of the second order of QCD coupling and the M_{EW} of the zeroth. The requested order at the amplitude level of the process will be of the second order as well. This is the key requirement for the interference sample production by the MadGraph generator.

The interference estimate is taken from the ATLAS $ZZjj$ EW production analysis as 6.8 % [1]. The subtraction method was successfully used for a cross-check of the result using limited statistics.

5.4 Data Processing

Statistical analysis of data requires to process a tremendous number of events, both real measured and MC simulated. This fact places another challenge in a sense of choice of the data processing approach. In the ATLAS collaboration, two approaches are considered typical: a mini-tree/n-tuple approach and histogram approach.

While the former fills a tuple of observables for each event keeping the event-observable association and fills the final histograms in a post-processing, the latter fills the histograms already during the selection stage. Mini-trees are more demanding in terms of memory usage in the selection stage of the analysis, however leaving a significant flexibility at the post-processing. This enables the analyzer to decide some final selection criteria, to create new combinations of observables, and to change the histogram binning at the final stage of the analysis. In the histogram approach, the analyzer has to define the selection criteria and binning in the beginning, which makes the keeping of full information of event obsolete, therefore saves the memory resources. It is wise to note that if the binning is fine enough, there is some flexibility also in this case, in a sense of the possibility of bin merger. Both approaches have certain pros and cons:

- Mini-tree approach:
  - pros: post-processing flexibility (choice of observables), MVA tuning, unconstrained binning;

  - cons: memory consumption, demanding systematic uncertainty evaluation.

- Histogram approach:
  - pros: less memory exhausting, systematic uncertainty evaluation accessible at the GRID nodes,
  - cons: requirement for an advanced user, not straight forward optimization process.

Since the analysis program at first served for the inclusive ZZ production analysis, which was rather a precision study with higher number of systematic uncertainties, and now it is serving for the presented search analysis, therefore both approaches are implemented and the previous conclusions are based on a direct comparison. The memory usage of both approaches is shown in Figure 5.4. The excessive memory consumption is obvious, which led to an implementation of a reduced mini-tree approach. Storage of the fundamental properties of physical objects, as the Lorentz vectors and labels, instead of intermediate objects as a di-jet or boson is the fundamental idea. The new approach saves a significant amount of memory, although only postpone a memory limit violation to a higher number of processed events. Moreover, such approach obviously demands an additional step in the post-processing which represents another source of potential error.

The uncertainty estimation of the analysis was successfully processed on the GRID, so, the recommendations can be summarized in a few points:

- the analysis should not exceed 50 considered sources of uncertainties (counting only sources demanding the execution of selection phase of the analysis algorithm, excluding luminosity etc.),
- the MC samples should be split into reasonable chunks of events, i.e. do not exceed a 10^5 of events per file,
- too large files transforms the problem character from memory to CPU time consumption,
- the GRID jobs are by default limited to approximately 90 hours, which can be negotiated if necessary.

The GRID processing has an advantage of using one submission system for real data samples and systematic uncertainty estimation with MC samples.

The memory limitation can be overcome by downloading of the processed sample to a private site and processing it "at home", but it disobey the WLCG *job-goes-where-the-data-are* concept. The author of the analysis prefers to use the

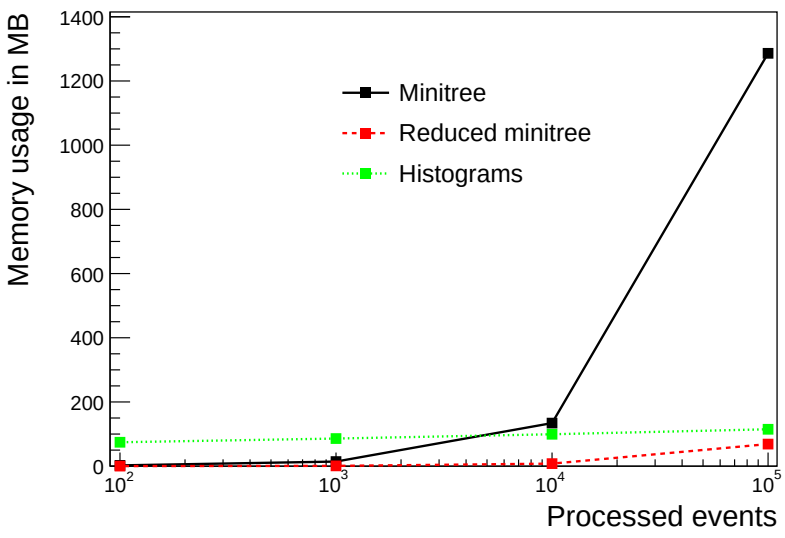


Figure 5.4: Memory usage comparison of the histogram and mini-tree analysis data storage approaches. A reduced version of the mini-tree approach is shown as well.

WLCG computing grid to prove that the state-of-the art physics can be done with the available resources.

Chapter 6

Event Selection

The presented analysis is focused on $ZZjj$ EW production in proton-proton collisions with four charged leptons in the final-state, called four-lepton channel. The fiducial phase space, defining the kinematic region of the measurement, follows the reconstruction criteria as much as possible. Results of the VBS analyses are usually not extrapolated to the whole phase space since the theoretical prediction has large uncertainty.

6.1 Fiducial Phase Space

The fiducial phase space definition has three parts, describing the leptons, vector bosons, and VBS tagging di-jet.

At the lepton level, the phase space requires at least four charged leptons, electrons or muons, originating from the primary vertex (prompt leptons) with $|\eta| < 2.7$. A four-momentum of each lepton is corrected with momenta of photons in a near vicinity in terms of the detector angular distance $\Delta R < 0.1$. The correction is called *dressing* and emulates the *bremsstrahlung* and other material-interaction effects.

The lepton pool must be able to form a pair of same flavor opposite charge (SFOC) di-leptons. The di-leptons represent Z boson candidates if the invariant mass m_{ll} is between 66 and 116 GeV. The constituents of resulting lepton quadruplet must pass a hierarchical transverse momentum (p_{T}) requirement defined from the most to least energetic as $p_{\mathrm{T}} > 20, 20, 10, 7$ GeV. The leptons must be isolated from each other by $\Delta R > 0.2$. If more than one quadruplet can be constructed, the one minimizing a distance from Z pole mass

$$\Delta m_{Z\text{-pole}} = |m_{ll} - m_Z| + |m_{l'l'} - m_Z| \tag{6.1}$$

is chosen for further processing. The Z pole mass value is obtained from the Review of Particle Physics from 2018 [156]. To finalize the boson level definition,

all possible di-leptons in the final quadruplet must have the invariant mass over 10 GeV as a quarkonia veto.

The definition of the VBS signature properties, described in Chapter 5, is utilized for the tagging jet kinematic criteria. Firstly, the phase space requires an event with at least two jets originated in the primary vertex. The jets are reconstructed using the anti-k_T algorithm with $R = 0.4$ [157], taking into account all particles with exception of the signal leptons described in the previous paragraph. Each jet must be isolated from the quadruplet leptons by $\Delta R > 0.2$; having p_{T} over 30 GeV in a central region, with $|\eta| < 2.4$; and p_{T} over 40 GeV in a forward region, with $2.4 < |\eta| < 4.5$. The candidate event must have the leading and sub-leading jet (j_1 and j_2) located at the opposite sides of the detector, satisfying equation $y_{j_1} \times y_{j_2} < 0$. The two jets form a di-jet system, which must must have well separated constituents with $\Delta y_{jj} > 2$ and invariant mass $m_{jj} > 300$ GeV. The definition of fiducial phase space is summarized in Table 6.1.

Category	Requirement	Parameter
Leptons	Momentum	$p_{\mathrm{T}} > 7$ GeV
	Pseudorapidity	$\|\eta\| < 2.7$
Jets	Clustering	`AntiKt4EMTopo`
	Central momentum	$p_{\mathrm{T}} > 30$ GeV ($\|\eta\| < 2.4$)
	Forward momentum	$p_{\mathrm{T}} > 40$ GeV ($2.4 < \|\eta\| < 4.5$)
	Pseudorapidity	$\|\eta\| < 4.5$
Overlap Removal	Lepton favoring	Remove jets if overlap with leptons $\Delta R = 0.2$
Quadruplet	Lepton pairing	Two SFOC pairs
	Hierarchy momentum	$p_{\mathrm{T}}^{\ell_1} > 20$ GeV, $p_{\mathrm{T}}^{\ell_2} > 20$ GeV, and $p_{\mathrm{T}}^{\ell_3} > 10$ GeV
	Separation	$\Delta R > 0.2$ between leptons in quadruplet
	Ranking	Minimal $\Delta m_{Z\text{-pole}}$, see Equation (6.1)
	Z window	$66 < m_{Z_1,Z_2} < 116$ GeV
	Quarkonia veto	$m_{ll} > 10$ GeV (any di-lepton)
Di-jet Tagging	Ranking	Highest scalar p_T sum
	Detector sides	$y_{j_1} \times y_{j_2} < 0$
	Rapidity separation	$\Delta y_{jj} > 2$
	Mass	$m_{jj} > 300$ GeV

Table 6.1: Fiducial phase space definition of the measurement. The object and event selection criteria applied in the four lepton channel of the analysis. The m_{ll} notation represents an invariant mass of any di-lepton in the quadruplet, while the m_{jj} represents the mass of a di-jet of the two most energetic jets.

6.2 Event Reconstruction and Selection

The event reconstruction and selection is accomplished simultaneously by the analysis algorithm, since an early rejection of useless event can save the data processing time. The design is divided into three steps, similarly to the fiducial phase space:

- an object selection considers electrons, muons, and jets;
- quadruplet reconstruction builds up the Z boson candidates;
- tagging di-jet reconstruction identifies the VBS signature.

6.3 Lepton Selection

The lepton selection is divided into two phases: starting with baseline requirements for individual leptons, from which the quadruplet is reconstructed; while leaving fine selection requirements to the final stage. The reason is a modeling of lepton misidentification which needs a space for loosening the final criteria. All objects are reconstructed using ATLAS combined performance tools, securing the reconstruction quality, enabling uncertainty extraction, and providing a scale factors, accounting identification and reconstruction efficiencies.

6.3.1 Electrons

The electron objects are required to be reconstructed inside of the ID. The ID tracks are associated to EM calorimeter clusters resulting in a better combined reconstruction. The baseline requirements are $|\eta| < 2.47$ and $p_{\mathrm{T}} > 7\,\mathrm{GeV}$. The combined performance electron quality tagging is set to `VeryLooseLH` working point. The tagging utilize likelihood based methods as stated in [158]. In addition, electron objects with EM clusters of poor quality are rejected. A longitudinal impact parameter $|z_0 \sin\theta|$ is calculated from the detector polar angle θ and a distance of the track from the primary vertex z_0. The parameter must be smaller than 0.5 mm ensuring an association with the primary vertex.

Selection criteria of isolation and transverse impact parameter are not applied for the baseline objects, because of their utilization in a stage of the background modeling.

6.3.2 Muons

The muon objects are of four types, according to the sub-detectors they are reconstructed in. The common types are:

`Combined`	reconstructed independently in the ID and MS and combined in the final stage,
`SegTagged`	reconstructed in the ID and matched to at least one segment in the MS
`CaloTagged`	reconstructed in the ID and matched to a cluster in calorimeter with a footprint of minimum ionizing particle,
`StandAlone`	reconstructed in the MS only, this type is also called extrapolated.

At the baseline level, the muons can be of any type. The muons are reconstructed up to $|\eta| < 2.7$ thanks to the `StandAlone` type of muon reconstruction, exploiting the full η coverage of the MS. A general p_{T} requirement is set to 7 GeV with an extension for `CaloTagged` type to 15 GeV, since the MS information is completely missing. This muon type has the smallest purity of all. The combined performance muon quality tagging is set to `Loose` working point [126]. The longitudinal impact parameter threshold is set alike the electron case. Since the cosmic radiation easily penetrates the ATLAS shielding and Earth crust, a cosmic rejection criteria are invoked employing a transverse impact parameter d_0. Muon tracks with higher value than 1 mm most likely do not originate in the primary vertex and are rejected as a cosmic radiation.

6.3.3 Triggers

The low level requirements defining a decision for the online filtering of the proton-proton collision events, described in Section 3.3.7, are called triggers. Their naming follows a notation composed of: a number of objects (if not displayed a single object is expected), an object type (`e` or `mu`), and a transverse momentum threshold in GeV, after an underscore (`_`), a quality requirement can follow (i.e. `lhmedium` or `iloose`) [159, 160]. The trigger can be further specified by a L1 trigger requirement or a request to be reconstructed only in MS. The names can be chained, standing for a multi-object trigger.

The analysis must employ a number of triggers because of exploiting data of four years collection, during which the trigger menu underwent a certain evolution. The utilized triggers are set up for electrons, muons, and their combinations, listed in Table 6.2 for year 2015. The triggers used in of subsequent years are listed in Appendix in Tables 13.2, 13.1, and 13.3. A certain trigger is applicable only to a range of LHC runs, represented by data-taking periods denoted by a letter-number couple in the last column.

Object	Trigger name	Data-taking periods
Electron		
	e24_lhmedium_L1EM20VH	All
	e60_lhmedium	All
	e120_lhloose	All
Multi-electron		
	2e12_lhloose_L12EM10VH	All
	e17_lhloose_2e9_lhloose	All
Muon		
	mu20_iloose_L1MU15	All
	mu40	All
	mu60_0eta105_msonly	All
Multi-muon		
	2mu10	All
	3mu6	All
	3mu6_msonly	All
	mu18_2mu4noL1	All
	mu18_mu8noL1	All
Mix		
	e17_lhloose_mu14	D3-D6
	e7_lhmedium_mu24	D3-D6
	2e12_lhloose_mu10	D3-D6
	e12_lhloose_2mu10	D3-D6

Table 6.2: Triggers utilized in the analysis for the data collected during year 2015. The triggers are applicable only to some periods of the data-taking.

6.3.4 Jets

The jets are reconstructed using three-dimensional topological clusters reconstructed in the calorimeter, which are then matched to tracks in the ID. The clusters are required to be of a good quality. Details of this procedure are discussed in [161]. The ID range naturally splits the jets selection to central with $|\eta|$ up to 2.4 and forward region within a range of $2.4 < |\eta| < 4.5$. The ID significantly improves the pile-up jet identification thanks to advanced vertex association using a Jet Vertex Tagger (JVT). Thus central region has lowered p_{T} threshold of 30 GeV in comparison to the forward region with $p_{\mathrm{T}} > 40$ GeV. The JVT is a multivariate discriminant with an output range between minus one and plus one, compatible with a pile-up (-1) and hard-scatter (1) jet prediction [162]. The discriminating threshold is set to

0.59, considering jets up to 60 GeV in p_{T}. In general, a probability of finding a pile-up jet with higher p_{T} is negligible.

6.3.5 Overlap Removal

Once the basic objects are selected, an overlap removal (OR) must be invoked to avoid double-counting. The OR prescription is following:

- if two electrons share a track or have overlapping calorimeter clusters:
  - remove the electron with lower p_{T};
- if an electron and muon share a track:
  - remove the muon if `CaloTagged`,
  - otherwise remove the electron;
- if a lepton is inside of a jet cone defined as $\Delta R = 0.4$:
  - remove the jet.

It can be noticed that a jet can not remove a lepton in any case, so, the utilized OR working point is called lepton favoring.

6.4 Boson Reconstruction

The reconstruction of Z bosons starts with a pairing of SFOC leptons and coupling them into a quadruplet. The quadruplet leptons of already baseline quality must pass a hierarchical p_{T} requirement with p_{T} over 20, 20, and 10 GeV for three most energetic leptons. Not more that two electrons can fail identification `LooseLH` working point and not more then one muon can be of `StandAlone` or `CaloTagged` type. At this point the best quadruplet is chosen according to Equation (6.1).

6.4.1 Quadruplet Quality

At a level of best quadruplet, finally, the isolation, electron identification, and transverse impact parameter are evaluated. These parameters are discussed in the Section 7.4, where the modeling of misidentified lepton background requires they would not affect the choice of the best quadruplet.

All signal leptons must pass an isolation working point set to `FixedCutLoose`, having transverse impact parameter significance $d_0/\sigma(d_0) < 5$ and 3 for electrons and muons respectively. An angular separation between quadruplet leptons must

be $\Delta R > 0.2$. Moreover, all electrons must pass `LooseLH` identification working point.

At the end, all possible di-leptons must pass the quarkonia veto, $m_{ll} > 10\,\text{GeV}$, and the Z candidate di-leptons must have the invariant mass in a range of $66 < m_{Z_1,Z_2} < 116\,\text{GeV}$.

6.5 VBS Tagging

The di-jet tagging requires at least two jets associated to the event. If more than two jets are available, the leading and sub-leading ones are chosen to reconstruct the tagging di-jet. The constituents must be from the opposite side of the detector satisfying $y_{j_1} \times y_{j_2} < 0$, and must be well separated with $\Delta y_{jj} > 2$, and having invariant mass of $m_{jj} > 300\,\text{GeV}$.

6.6 Preliminary VBS Selection

The described selection criteria define the SR representing a pre-VBS selection stage where the signal is enhanced, although the QCD background is still dominant. All the requirements are summarized in Table 6.3. A comparison of MC prediction and observed data is shown in Figure 6.1

6.7 Multivariate Analysis

Further application of linear cuts on top of the pre-VBS selection does not bring effective signal separation any more. Relative amount of the signal in the final sample can not be easily increased without a significant loss of the overall statistical population, therefore, advanced methods for the signal discrimination must be employed.

MVA techniques are adopted for signal-background discrimination in the analysis. Despite of the ineffectivity of the simple kinematics cuts on observables, their distributions are still usable for the signal differentiation. The signal and background processes have different fundamentals, which affects the shape of distributions and enables to identify the signal.

A practice in common VBS efforts shown, a system of subsequent cuts organized in a tree topology, called decision tree, is suitable for the signal-background classification. In HEP community, Boosted Decision Trees (BDTs) receive a significant popularity among the search analyses. The ATLAS collaboration VBS efforts almost exclusively employ TMVA [163] toolkit, the presented analysis however prefers Scikit-learn [164] toolkit to challenge the monopoly.

Category	Requirement	Parameter
Event Pre-selection	Trigger	Details in Section 6.3.3
Electrons	Momentum	$p_{\mathrm{T}} > 7\,\mathrm{GeV}$
	Pseudorapidity	$\lvert\eta\rvert < 2.47$
	Identification	`VeryLooseLH` working point
	Quality	No bad cluster
	Impact parameter	$\lvert z_0 \sin\theta\rvert < 0.5\,\mathrm{mm}$
Muons	Momentum	$p_{\mathrm{T}} > 7\,\mathrm{GeV}$ (15 GeV if `CaloTagged` muon)
	Pseudorapidity	$\lvert\eta\rvert < 2.7$
	Identification	`Loose` working point
	Impact parameter	$\lvert z_0 \sin\theta\rvert < 0.5\,\mathrm{mm}$ (if not `StandAlone`)
	Cosmic rejection	$d_0 < 1\,\mathrm{mm}$ (if not `StandAlone`)
Jets	Clustering	`AntiKt4EMTopo`
	Central momentum	$p_{\mathrm{T}} > 30\,\mathrm{GeV}$ ($\lvert\eta\rvert < 2.4$)
	Forward momentum	$p_{\mathrm{T}} > 40\,\mathrm{GeV}$ ($2.4 < \lvert\eta\rvert < 4.5$)
	Pseudorapidity	$\lvert\eta\rvert < 4.5$
	Pile-up jet	JVT < 0.59
Overlap Removal	Lepton favoring	Remove jets if overlap with leptons $\Delta R = 0.2$
Quadruplet Selection	Lepton pairing	Two SFOC pairs
	Hierarchy momentum	$p_{\mathrm{T}}^{\ell_1} > 20\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\ell_2} > 20\,\mathrm{GeV}$, and $p_{\mathrm{T}}^{\ell_3} > 10\,\mathrm{GeV}$
	Electron identification	At least two electrons of `LooseLH` identification
	Muon quality	Maximally one `StandAlone` or `CaloTagged` muon
	Ranking	Smallest $\Delta m_{Z\text{-pole}}$, see Equation (6.1)
Quadruplet Quality	Separation	$\Delta R > 0.2$ between leptons in quadruplet
	Isolation	`FixedCutLoose` working point
	Electron identification	`LooseLH` working point
	Impact parameter	$d_0/\sigma(d_0) < 5$ (3) for e (μ)
	Z window	$66 < m_{Z_1,Z_2} < 116\,\mathrm{GeV}$
	Quarkonia veto	$m_{ll} > 10\,\mathrm{GeV}$ (any dilepton)
Di-jet Tagging	Ranking	Highest scalar p_T sum
	Detector sides	$y_{j_1} \times y_{j_2} < 0$
	Rapidity separation	$\Delta y_{jj} > 2$
	Mass	$m_{jj} > 300\,\mathrm{GeV}$

Table 6.3: Signal region definition of the measurement. The object and event selection criteria applied in the reconstruction stage of the analysis are listed. The m_{ll} notation represents an invariant mass of any di-lepton in the quadruplet, while the m_{jj} represents the mass of a di-jet of the two most energetic jets.

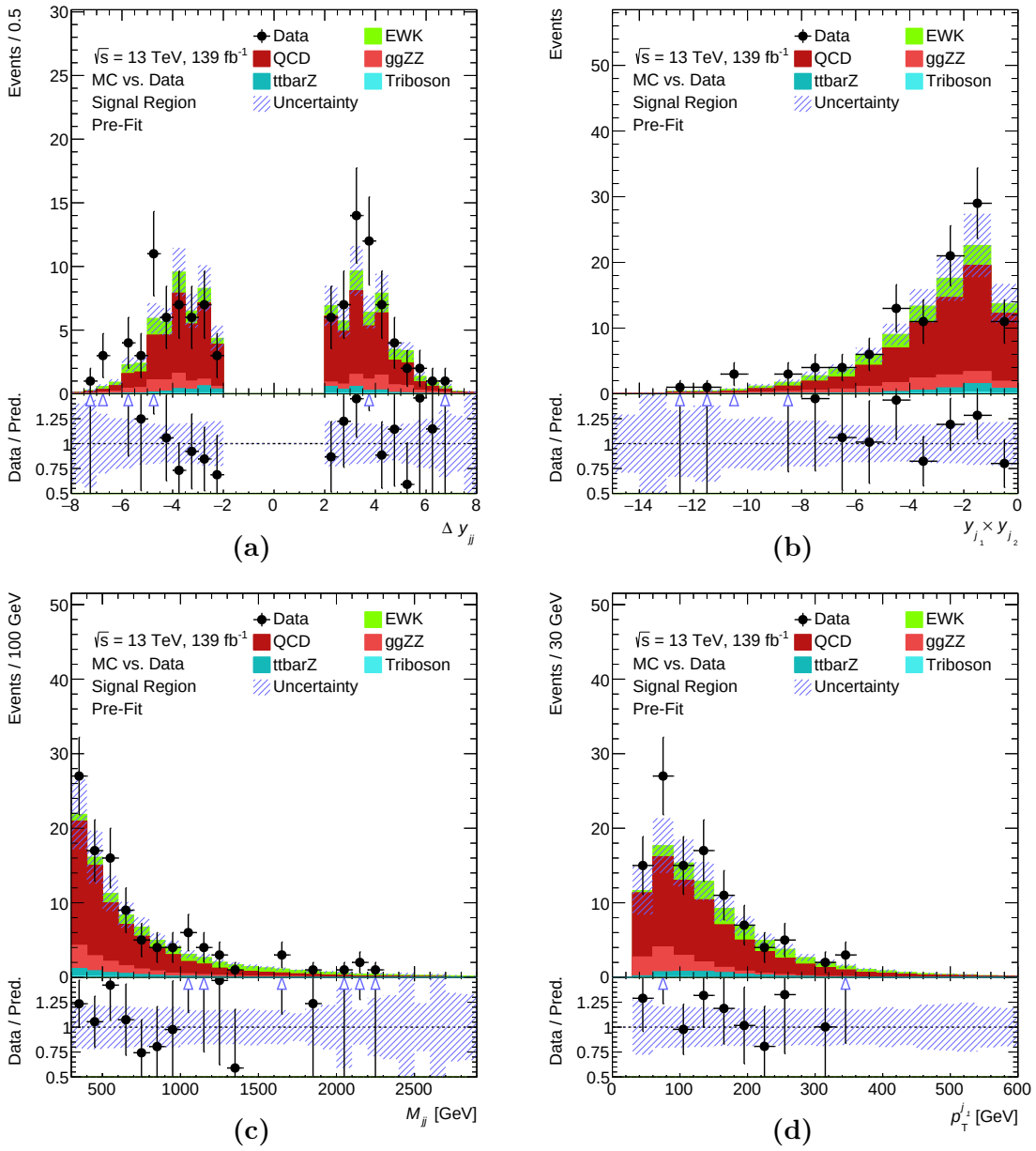


Figure 6.1: Comparison of the MC prediction and observed data in the SR. Distributions as a function of Δy_{jj} (a), $y_{j_1} \times y_{j_2}$ (b), m_{jj} (c), and leading jet p_{T} (d) are shown. Both statistic and systematic uncertainties are included.

6.7.1 Method

The classification task is transformed into a regression of continuous variable in a range between minus one and plus one. Closer to plus one the value is, more likely the evaluated event belongs to the signal group. Towards the negative values, a background event is expected. In general, the value quantifies the measure of signal and background probability. The resulting continuous distribution is called a BDT score. The score is similar to the JVT parameter introduced in Section 6.3.4.

The BDTs methods belong to a supervised machine learning class. The input samples are labeled as signal or background, so the algorithm is provided with a truth feedback, which is utilized for the training. The trees are employed as weak learners, creating an ensemble with a high number of classifiers, sometimes called a forest. The final decision is based on a collective decision of the trees, representing the boost. The trees are trained in an iterative manner, reflecting the current success of the classification in each iteration. Two approaches are utilized for the boosting, an adaptive approach uses a weighting of the classifiers and samples, while a gradient approach exploits an optimal classifier.

Adaptive Boost assigns a weight to individual trees according to their error in classification. More successful the tree is the higher vote-weight it obtains. Each iteration creates a set of new trees, but only the best tree is kept for the final decision. The decision is based on votes of all best trees.

Moreover, the events have another weight, a sample-weight, which is increased in a case of their wrong classification by the best tree in the current iteration. So, the next iteration will be focused on them. Such events will be less likely wrongly identified in the next iterations.

Once the maximum number of trees is constructed, or a vote-weight of a certain number of the latest best trees do not satisfy a condition the method is considered trained.

Gradient Boost uses a different approach, which is at first glance simpler. It starts with a mean value of the events collection, representing an initial guess of the whole ensemble. In every iteration, the algorithm constructs an optimal classifier. It is the best possible, although still a weak classifier. The classifier is based on calculation of a loss function using pseudo-residual values. The values represent a measure of wrong classification. The optimal classifier is inhibited by a control parameter called learning rate, which slows down the boosting for a better overall result. The final decision is performed as a sum of contributions of all relevant trees.

Once the maximum number of trees is constructed, or the pseudo-residuals do not decrease for a certain number of iterations, the training is finished. By con-

struction the method is not bounded to absolute value of one, because of the trees addition, therefore the hyperbolic tangent function is used to fulfill this requirement.

The algorithms are explained very briefly, for a thorough understanding, it is recommended to study the original papers for Adaptive [165] and Gradient Boosting [166]. The latter also describes the learning rate effect on the boosting. It is also good to emphasize, that the weighting and construction of an optimal classifier is possible only because of the training is supervised. The two methods are further denoted as BDT and GBDT. From the analysis perspective, it is important to understand the behavior of the input parameters which are of two types, associated to the weak learner and to the boosting.

6.7.2 Parameters

The weak learner parameters are associated to the tree properties. Tree depth is a maximal number of subsequent decisions and is chosen to three for BDT and five for GBDT. Leaf size is a minimal number of events affected by a decision in a leaf node. The parameter is proportional to the whole sample and is optimized to 2 %.

Boosting parameters affect all trees, the learning process, and final decision: forest size, a maximum number of trees in the ensemble, is set to 800 for BDT and 1000 for GBDT; learning rate, a weight of new tree in the Gradient Boosting is set to 0.1; and beta function, a measure of vote- and sample-weighting, in its meaning comparable to the learning rate in Adaptive Boosting is set to 0.5.

The introduced MVA parameters were optimized in case of the Scikit-learn tool, the TMVA classifier was borrowed directly from the ATLAS measurement [1].

6.7.3 MVA Input

The training is supervised using the MC samples of $ZZjj$ EW and $ZZjj$ QCD production (including quark- and gluon-induced) as the signal and background labeled events. The samples of all MC campaigns (mc16a, mc16d, and mc16e described in Section 4.4) are used as the input to enlarge the statistical population. The events across the campaigns differ only in the pile-up profile, which is corrected by a proportional re-weighting by a corresponding integrated luminosity (see Table 4.1). The training and testing phase of the MVA methods is executed on two independent samples. The combined MC sample input is split to avoid the input bias. Two thirds are used for the training and the rest for testing of the resulting classifier.

The input variables are categorized into five classes:

tagging di-jet the familiar observables used for the VBS signature definition m_{jj}, Δy_{jj}, and $y_{j_1} \times y_{j_2}$;

quadruplet the energy observables of the reconstructed Z bosons p_{T}^{ZZ} and m_{ZZ};

vector-boson a p_{T} of boson closer to the $m_{Z\text{-pole}}$, $p_{\mathrm{T}}^{Z_1}$, and centrality of both bosons $y^*_{Z_1}$ and $y^*_{Z_2}$;

single object a p_{T} of the leading and sub-leading jets and the third most energetic lepton, $p_{\mathrm{T}}^{j_1}$, $p_{\mathrm{T}}^{j_2}$, and $p_{\mathrm{T}}^{\ell_3}$;

global a fraction of p_{T} and a scalar sum of transverse momentum H_{T} of $ZZjj$ system, $p_{\mathrm{T}}^{ZZjj}/H_{\mathrm{T}}^{ZZjj}$;

where the y^*_Z denotes a boson centrality between jets, defined as

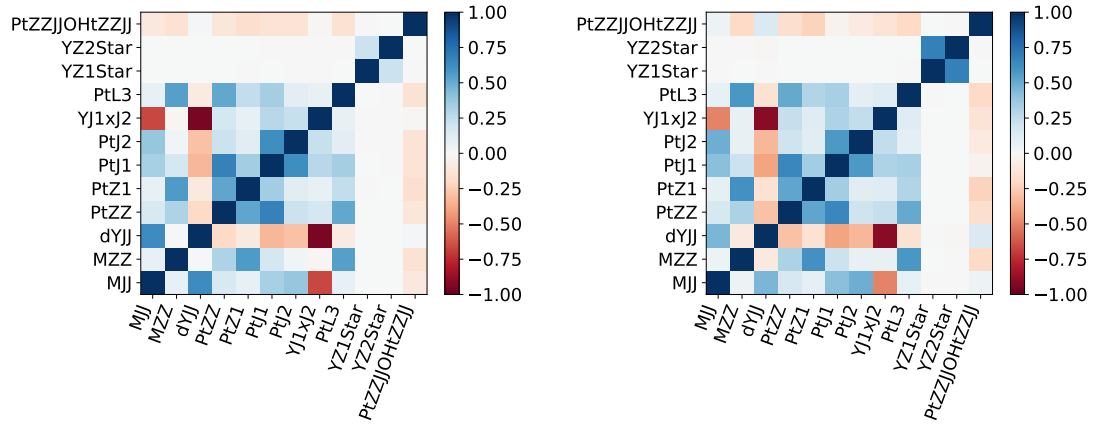
$$y^*_Z = y_Z - \frac{y_{j_1} + y_{j_2}}{2}.$$

The physical observables used as variables in the MVA methods are listed in Table 6.4. The variables are considered as the input to the MVA methods rather than a physical observables, thus the physical units are dropped. A physical meaning of the variables does not play a role in the prediction nor the training, thus the units are dropped.

MVA Variable	Physical Observable	MVA Variable	Physical Observable
`MJJ`	m_{jj}	`PtZ1`	$p_{\mathrm{T}}^{Z_1}$
`dYJJ`	Δy_{jj}	`YZ1Star`	$y^*_{Z_1}$
`YJ1xJ2`	$y_{j_1} \times y_{j_2}$	`YZ2Star`	$y^*_{Z_2}$
`MZZ`	m_{ZZ}	`PtJ1`	$p_{\mathrm{T}}^{j_1}$
`PtZZ`	p_{T}^{ZZ}	`PtJ2`	$p_{\mathrm{T}}^{j_2}$
`PtZZJJOHtZZJJ`	$p_{\mathrm{T}}^{ZZjj}/H_{\mathrm{T}}^{ZZjj}$	`PtL3`	$p_{\mathrm{T}}^{\ell_3}$

Table 6.4: Naming of the MVA variables associated to the physical observables.

The MVA variables are chosen to be predominantly uncorrelated with exception of the tagging di-jet variables which are tolerated with some level of correlation. The correlations were evaluated for both signal and background inputs, shown in Figure 6.2, utilizing the Pearson correlation coefficient. It can be noticed that, for example, the variables of centrality (Z^*) are quite correlated in the background input, although much less correlated in the signal.



(a) **(b)**

Figure 6.2: Visualization of the correlations of MVA input variables in the signal (a) and background (b) categories. The variables to physical observables association is defined in Table 6.4.

The tagging di-jet variables are put in relation utilizing scatter diagrams in Figure 6.3. The anti-correlation is clearly visible in both diagrams. Although the density of points in Figure 6.3a is much higher at the region with small values, while the signal has almost uniform distribution in the full range. The anti-correlation in Figure 6.3b is caused by the fact the both variables affect each other creating a clearly visible contour, however the signal distribution is again more uniform around the contour than the background. All three variables, even when correlated, still have a discriminating power. Therefore, they are kept in the set of input variables.

The scrutiny of the MVA input variables is completed with a signal to background comparison of distribution shapes of all variables shown in Figure 6.4. The comparison shows the distributions have different shapes, although not significant enough to apply an additional simple linear cut. The `PtZZJJOHtZZJJ` indicates the signal events have a low ratio. The differences are rather mild in case of the p_{T} related variables, while the tagging di-jet still have decent distinctive potential, even after already applied di-jet cuts in the pre-VBS selection. Finally, the centrality variables `YZ1Star` and `YZ2Star` reveal the bosons are rather centered between the jets in case of the signal.

6.7.4 Performance and Validation

Before the classifier performance can be evaluated, the machine learning process must be validated employing an over-training control. A classifier is trained and

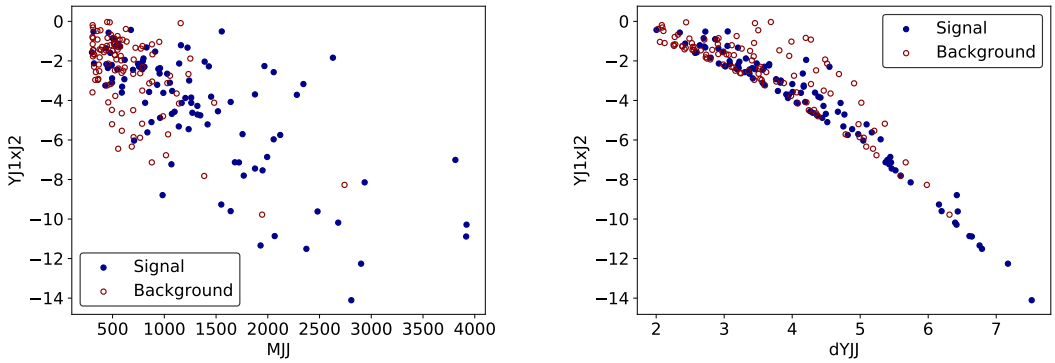


Figure 6.3: Correlation of the di-jet variables visualized using scatter plots.The `MJJ` is related to `YJ1xJ2` in (a) and `dYJJ` is related to `YJ1xJ2` in (b).

tested on different input samples, however the BDT score should be the same for both samples as well as for both categories, the signal and background. The BDT scores are compared in Figure 6.5, where one can see the distributions are very matching. The classifier trained on the training sample behaves the same way on the testing sample.

The usual performance test of a machine learning classifier is a receiver operating characteristic (ROC) curve. The curve puts in relation false positive and true positive rates. In principle, all well predicted samples are accounted below the curve while all wrong guesses are over the line, i.e. true and false negatives. The area under ROC curve is calculated for the three tested cases BDT (Scikit Learn) 0.862, GBDT (Scikit Learn) 0.865, and GBDT (TMVA) 0.854. The results are very similar for all the methods and tools, as can be seen in a direct comparison of the ROCs, in Figure 6.6, which implies they are comparable.

6.8 Summary

The Scikit-learn GBDT was chosen as the MVA method for further analysis steps. All MVA approaches are comparable, so, the TMVA monopoly in the ATLAS collaboration is rather historical and due to an advanced integration of the ROOT tools in the analysis frameworks. In general, it can be expected the Scikit-learn tools will become more popular with an integration of python into the analysis frameworks and ROOT, which is already in process.

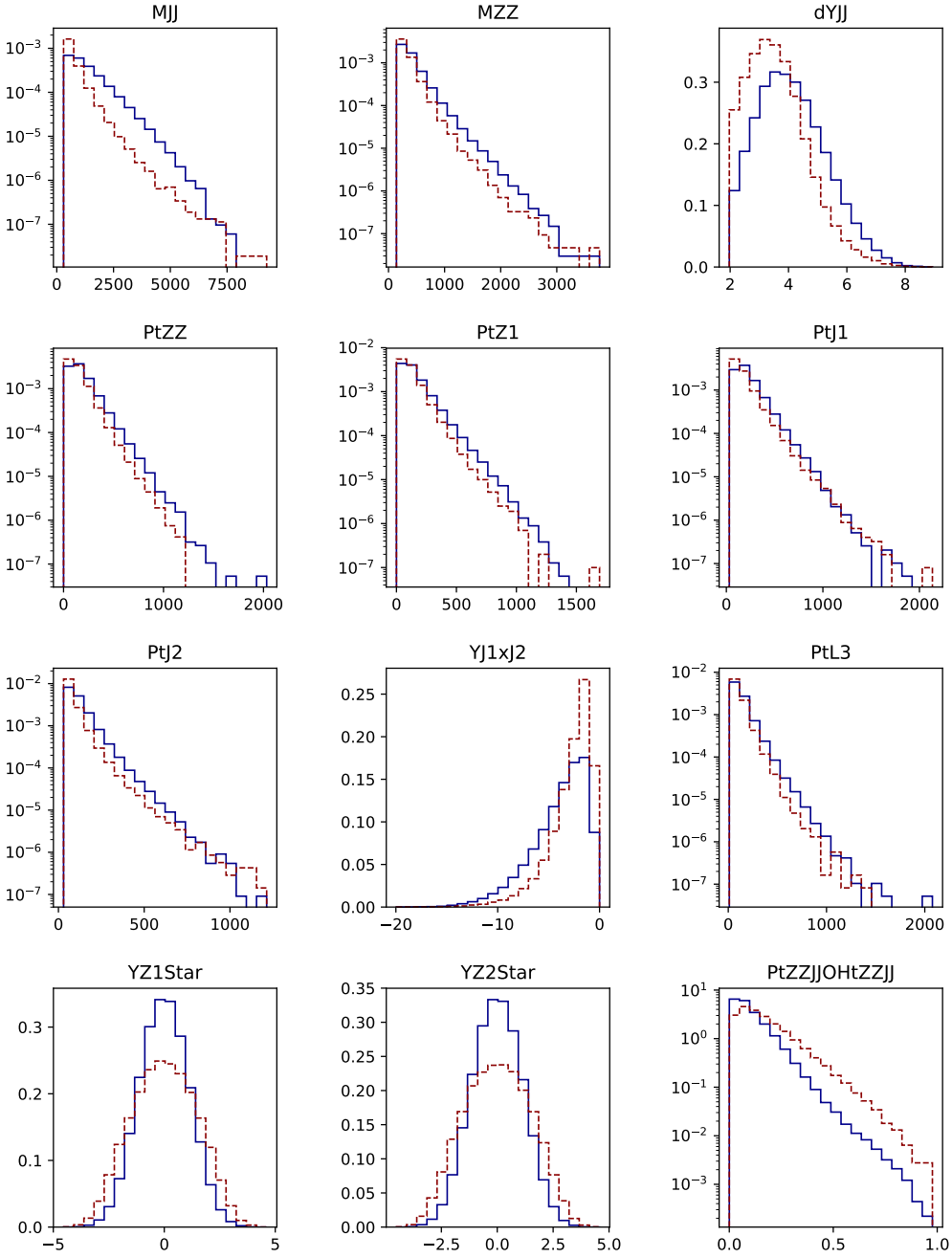


Figure 6.4: Comparison of signal (solid blue line) and background (dashed red line) MVA inputs. The distributions are normalized to unity. The mass and momentum related variables are shown in a logarithmic scale.

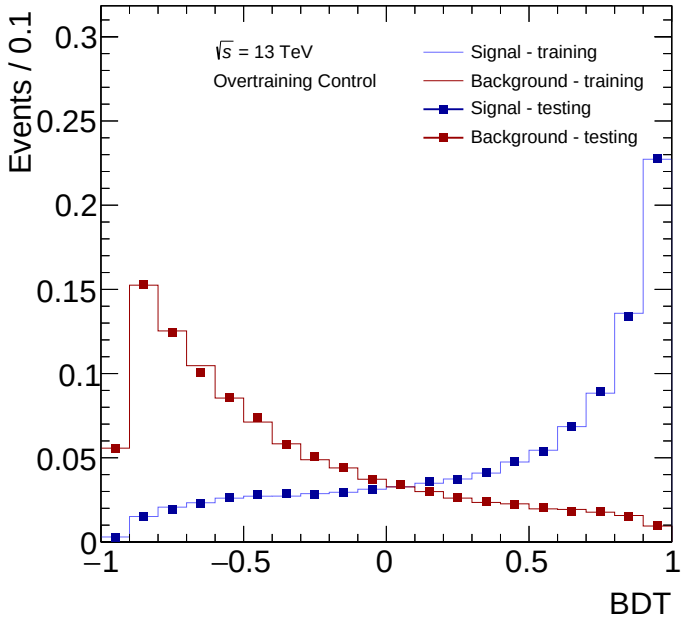


Figure 6.5: Comparison of BDT scores evaluated on the training and testing input samples for the signal and background categories.

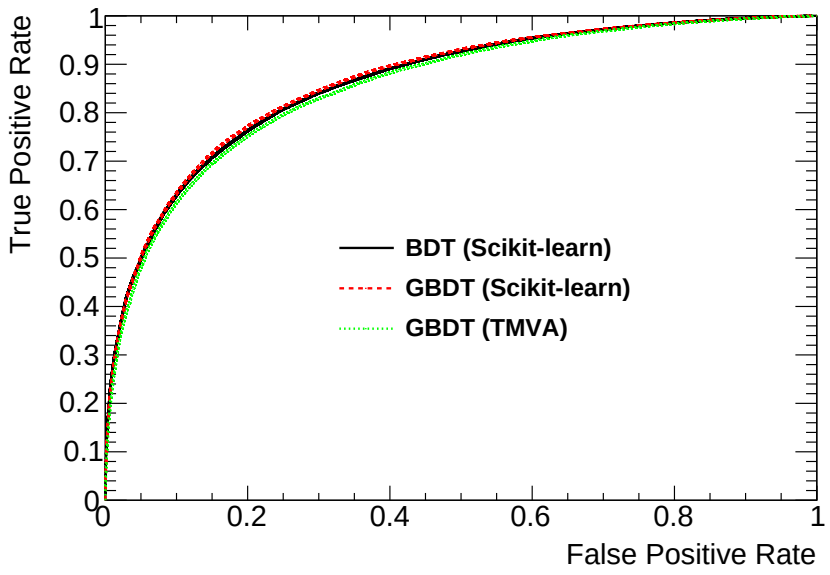


Figure 6.6: ROC curve of BDT and GBDT classifiers implemented in the Scikit-learn and GBDT implemented in the TMVA.

Chapter 7

Background

7.1 Introduction

The presented analysis is build up on the base-stone of inclusive ZZ to four lepton production analysis [20], which is very clean channel, therefore the EW production search background arising from non-ZZ processes will be small. On the other hand, the channel is overwhelmed by the QCD production.

Non-ZZ processes with at least four genuine leptons can pass the signal selection criteria (described in Chapter 6) and can not be easily filtered out without an inevitable loss of portion of the small VBS signal. Such processes are categorized as irreducible and are called prompt background throughout the document. The prompt background is fully modeled by MC simulated samples and its contribution is subtracted from the SR yield.

The EW-QCD ratio of the presented channel is very small, which makes the QCD-induced processes dominant not only over the signal, but also over remaining backgrounds. The processes also produce prompt leptons, although due to its large amount can not be simply subtracted using the MC simulated samples. Advanced selection techniques must be applied to separate the background, therefore it does not belong to the pure irreducible background category. However, it can be categorized as the natural-VBS background and will be called QCD background in the following text.

The third type of background comes from the processes that are able to produce only a fraction of prompt leptons, the rest are misidentified objects or leptons originated in a jet. Such leptons are not genuine in the sense of four-lepton processes and are called fake leptons (or just fakes). Hadronizing quarks or gluons from the busy environment of proton-proton collisions are the primary source of fake leptons.

7.2 Prompt Background

The prompt background is composed primarily of the tri-boson and $t\bar{t}Z$ production, but in principle also rear processes like $t\bar{t}WW$ production can contribute. The ZZ production itself contributes as well, when one of the bosons decays to tau-leptons which later decays leptonically. The multiboson background is composed of WWZ, WZZ, and ZZZ production with at least four charged leptons in the final state. Semi-leptonic ZZV production with four charged leptons and hadronically decaying V boson (W or Z) is a part of signal, having six EW couplings (described in Section 4.4.1), therefore it is excluded from the background. All background contributions are estimated using the MC simulation, the corresponding MC samples description can be found in Section 4.4.2. The individual background contributions to the SR are listed in the left column of Table 7.1, together with the statistical and systematic uncertainty components added in quadrature. The systematic uncertainty includes experimental and theoretical components, in detail described in Chapter 8.

MC sample	Event yield
$t\bar{t}$Z	$3.86^{+0.71}_{-0.64}$
Tri-boson	$0.61^{+0.74}_{-0.38}$
$ZZ \to (\ell\ell\tau\tau)$	$0.15^{+0.24}_{-0.08}$
$t\bar{t}WW$	$0.02^{+0.01}_{-0.02}$
Total	$4.64^{+1.05}_{-0.76}$

Table 7.1: Prompt background event yield estimate in the SR scaled to 139 fb^{-1}, split into individual processes, including statistical and systematic uncertainties.

7.3 QCD Background

The QCD background is considered only in the case of EW production search, in the inclusive cross-section measurement the processes are included in the signal.

The quark-induced processes of the QCD background are dominant over the gluon-induced ones approximately by an order of magnitude, due to the quark-loop included in the diagram of the latter (see Figure 2.5b). The background contribution is partially modeled by the MC simulation, using the distribution shape. The normalization is determined by a data-driven approach maximizing a likelihood function in a background control region; ideally, that under the assumed statistical model the observed data are the most probable. The approach utilizes

a profile likelihood fit, described in Section 9.2 dedicated to the signal significance extraction. The utilized MC samples are fully described in Section 4.4.1.

The control region for the data-driven approach was chosen in a way enhancing the background against signal and is called QCD Control Region (QCD-CR). The QCD-CR is defined by inversion of two jet selection criteria of the SR, the di-jet mass (m_{jj}) and the rapidity separation of the jets (Δy_{jj}). The inversion ensures the orthogonality between the regions (SR and QCD-CR) to avoid double-counting, other criteria are the same. The yields of QCD background processes are listed in Table 7.2 for both SR and QCD-CR. The prompt background processes contribute to the QCD-CR by $9.48^{+1.29}_{-0.92}$ events. A comparison of the MC prediction and observed data in a newly defined region is shown in Figure 7.1.

MC sample	Event yield	
	SR	QCD-CR
Quark-induced	$60.58^{+20.37}_{-14.13}$	$114.81^{+34.31}_{-24.69}$
Gluon-induced	$11.13^{+5.44}_{-4.22}$	$14.75^{+7.14}_{-5.50}$
Total	$71.71^{+21.08}_{-14.75}$	$129.56^{+35.05}_{-25.3}$

Table 7.2: QCD background event yield estimate in the SR and QCD-CR scaled to $139\,\text{fb}^{-1}$, split into individual processes, including statistical and systematic uncertainties.

7.4 Fakes Background

In case of the $ZZjj$ EW production, the fakes background is overwhelmed by the QCD background, however, it becomes significant in the measurement of the inclusive production. Due to difficulties of the MC modeling of object misidentification, a data-driven approach was chosen to estimate the amount of the background.

Processes contributing to the fakes background can be grouped into two categories according to their final-states, with one fake as N_{RRRF} and with two as N_{RRFF}. Typical processes contributing to the first category are WZ + jets and $t\bar{t}W$. The processes of the second category are Z + jets, $t\bar{t}$, and WW + jets. A contribution of processes with three fake leptons (i.e. W + jets) are assumed negligible and are unlikely to pass the selection criteria in the SR. The background estimation is focused on the two most contributing processes, Z + jets and $t\bar{t}$. MC samples supporting the data-driven method are described in Section 4.4.2.

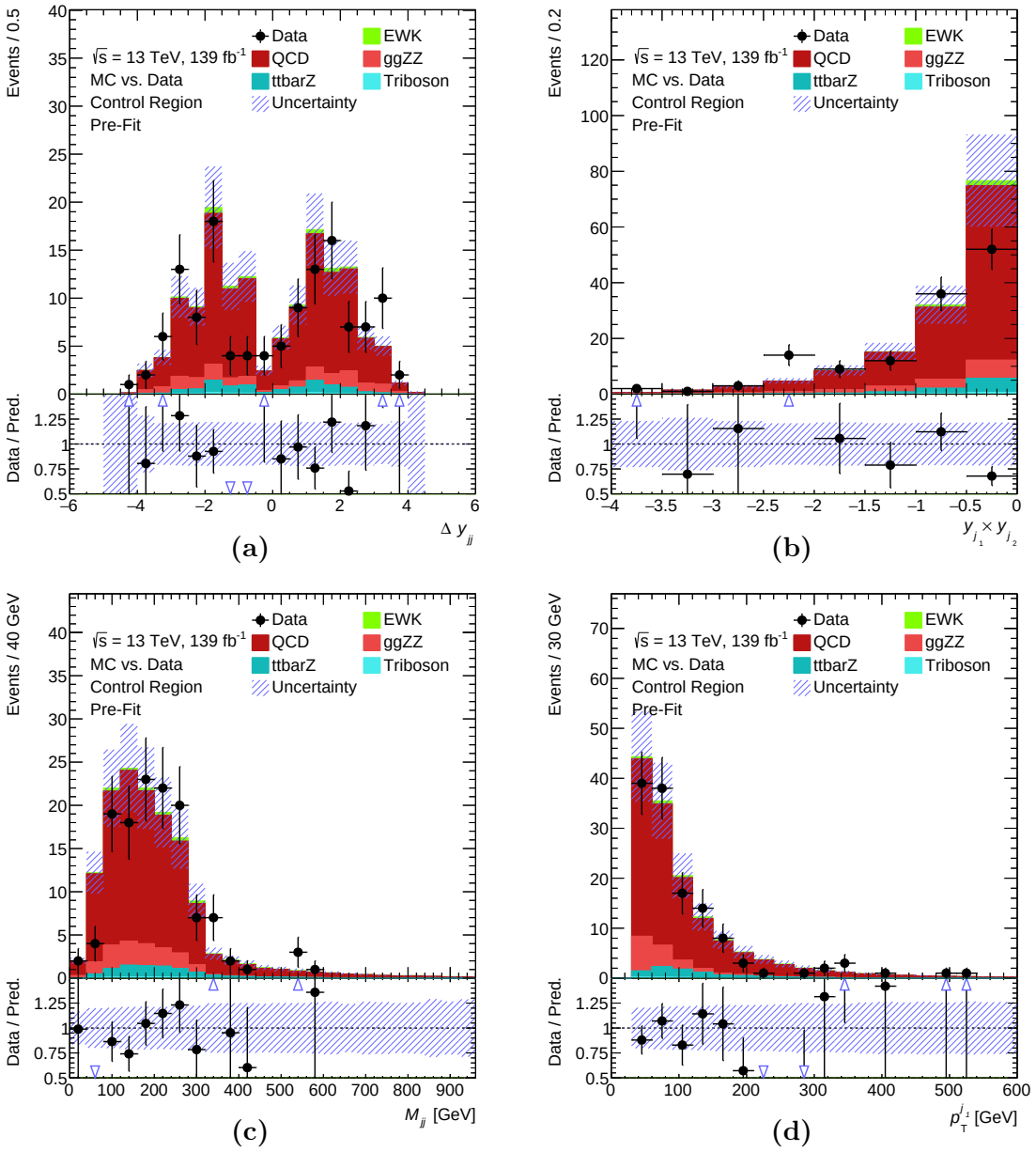


Figure 7.1: Comparison of the MC prediction and observed data in the QCD-CR. Distributions as a function of Δy_{jj} (a), $y_{j_1} \times y_{j_2}$ (b), m_{jj} (c), and leading jet p_{T} (d) are shown. Both statistic and systematic uncertainties are included.

7.4.1 Background Estimation Technique

The background estimation method is based on extrapolation of the fake leptons contribution from a fake-enriched data sample to the SR. Using this sample, a relation between the lepton selection and fake lepton rate is derived. The relation is called fake factor and is used for the projection of leptons misidentification to the SR. The background estimation approach is called fake factor method.

Collection of the data sample is achieved in a kinematic region, further called Fake Factor Region (FFR), carefully chosen to contain the Fake leptons with high purity. Two distinct FFRs are used, with fake leptons originated in Z + jets and $t\bar{t}$ processes.

The fake factor (F_{Fake}) is defined as an efficiency (f) that a fake lepton passes the SR selection requirements over an efficiency ($\bar{f}$) that a Fake lepton passes a loosened selection as:

$$F_{\text{Fake}} = \frac{f}{\bar{f}} = \frac{N_G}{N_P}\,, \tag{7.1}$$

where N_G is a number of leptons that passed the signal lepton selection (denoted as *good* leptons by G) and N_P is a number of leptons that passed the loosened selection (denoted as *poor* leptons by P). Both types of leptons are acquired in the FFR, therefore are assumed to be fakes. The loosened selection (or definition of a *poor* lepton) must not overlap with the signal selection (or definition of a *good* lepton), which is done by inversion of chosen selection criteria. The inversion is made separately for electrons and muons in the following lepton properties:

- the isolation or likelihood-based identification for electrons, an electron failing the identification criterion base working point `LooseLH` must pass the next lower, `VeryLooseLH`;
- the isolation or impact parameter significance for muons, a cut $|d_0/\sigma(d_0)| < 10$ on muons failing the second criterion is applied in addition.

The loosened kinematic region is established in both FFRs and the SR using the same selection criteria and is inclusive, accepting the cases when both criteria are not satisfied. The region adjacent to the SR is of a special interest, the factors from both FFRs will be applied there. Further in text, the region is called Fake Control Region (FCR). The inversion is applied in the Event Selection part of the SR, after the best quadruplet is chosen although before the di-jet selection. Orthogonality, secured by the inversion is necessary to avoid double-counting of the *good* leptons in the denominator of the F_{Fake}. For the context of the FCR and SR see Table 7.3.

While the SR counts only the events with *good* leptons as N_{GGGG} (the ideal number the method tries to correct), the FCR includes events with one and two

Category	Cut Name	Requirement
... (Object and Quadruplet Selection of the SR)		
Event Selection	*Good* electrons G_e	Pass isolation and pass identification
	Poor electrons P_e	Fail isolation or fail identification
	Good muons G_μ	Pass isolation and pass $\lvert d_0/\sigma(d_0)\rvert < 3$
	Poor muons P_μ	Fail isolation or pass $3 < \lvert d_0/\sigma(d_0)\rvert < 10$
	Lepton counting ($\ell = e, \mu$)	$P_\ell = 2$ and $G_\ell = 2$, or $P_\ell = 1$ and $G_\ell = 3$
Overlap Removal		Remove jets with $\Delta R_{\ell\text{-}j} < 0.4$ to *poor* leptons in the selected quadruplet
... (Dijet Selection of the SR)		

Table 7.3: FCR definition is similar to the SR, therefore only the part where the inversion is applied is shown. For definition of isolation and identification working points, refer to the SR definition in Table 6.3. The preceding and succeeding selection steps are represented by dots.

poor leptons in the counting (N_{GGGP} and N_{GGPP}). If the picture is completed with the efficiency e of a real lepton (R) passing the signal lepton selection and a complementary efficiency for passing the loosened selection $\bar{e} = (1 - e)$. Together, with already mentioned efficiencies f and $\bar{f} = (1 - f)$, considering a fake lepton (F), the whole spectrum of possibilities can be constructed, starting with

$$N_{GGGG} = e_1 e_2 e_3 e_4 N_{RRRR},$$

a relation between number of events with all *good* leptons and number of events with all real leptons, going across

$$N_{GGGG} = f_1 e_2 e_3 e_4 N_{FRRR},$$

a relation showing the first Fake lepton passed the signal selection, nearing to the end with

$$N_{GGPP} = f_1 e_2 \bar{f}_3 \bar{e}_4 N_{FRFR},$$

a relation having all possible efficiencies with two real and two fake leptons passing SR and FCR selections. Using all possibilities, one can construct a matrix relating the truth classification to the reconstructed cases. Assuming the efficiency e is equal to one, the matrix can be outlined as:

$$\begin{pmatrix} N_{GGGG} \\ N_{PGGG} \\ N_{GPGG} \\ \vdots \\ N_{GGPP} \end{pmatrix} = \begin{pmatrix} 1 & f_1 & f_2 & f_3 & f_4 & f_1 f_2 & f_1 f_3 & f_1 f_4 & f_2 f_3 & f_2 f_4 & f_3 f_4 \\ 0 & \bar{f}_1 & 0 & 0 & 0 & \bar{f}_1 f_2 & \bar{f}_1 f_3 & \bar{f}_1 f_4 & 0 & 0 & 0 \\ 0 & 0 & \bar{f}_2 & 0 & 0 & f_1 \bar{f}_2 & 0 & 0 & \bar{f}_2 f_3 & \bar{f}_2 f_4 & 0 \\ & & & & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \bar{f}_3 \bar{f}_4 \end{pmatrix} \begin{pmatrix} N_{RRRR} \\ N_{FRRR} \\ N_{RFRR} \\ \vdots \\ N_{RRFF} \end{pmatrix}.$$

Since we are looking for a reduction term for the reconstructed objects by the truth objects, in shape of $N_{GGGG} - N_{RRRR}$, the first line of the matrix will be elaborated. Using the Gauss elimination, the line gets a shape of

$$\begin{aligned} N_{GGGG} = N_{RRRR} + N_{PGGG}\frac{f_1}{\bar{f}_1} + N_{GPGG}\frac{f_2}{\bar{f}_2} + \cdots + N_{GGGP}\frac{f_4}{\bar{f}_4} \\ -N_{PPGG}\frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} - N_{PGPG}\frac{f_1}{\bar{f}_1}\frac{f_3}{\bar{f}_3} - \cdots - N_{GGPP}\frac{f_3}{\bar{f}_3}\frac{f_4}{\bar{f}_4}. \end{aligned}$$

The order of leptons does not play a role in the statistical analysis, so the result can be summed up to a form of

$$N_{\text{Fake}} = \left(N_{gggp} - N^{ZZ}_{gggp}\right)\frac{f}{\bar{f}} - \left(N_{ggpp} - N^{ZZ}_{ggpp}\right) \times \left(\frac{f}{\bar{f}}\right)^2 , \tag{7.2}$$

where the terms N_{gggp} and N_{ggpp} are the numbers of events with one and two *poor* leptons regardless of their position in the lepton quadruplet, and the reducible term on the left hand side. A detailed derivation of the equation can be found in [167]. The N^{ZZ} terms are added artificially as a compensation of the real leptons contribution in the FCR and as a support for the initial assumption of $\bar{e} = 0$ resp. $e = 1$. The compensation is derived using the MC samples.

7.4.2 Fake Factor Calculation

The F_{Fake} is calculated for two major contributing processes to the FCR, therefore two FFRs need to be defined for $Z+\text{jets}$ and $t\bar{t}$ processes and two factors will be produced by the method. Both factors are calculated using Equation (7.1) with a correction for the real leptons coming from the ZZ production, and corrections for the other background processes as $WZ+\text{jets}$ and the major process reciprocally, resulting in

$$F^{Z+\text{jets}}_{\text{Fake}} = \frac{N_G - N_G^{WZ,\,ZZ,\,t\bar{t}}}{N_P - N_P^{WZ,\,ZZ,\,t\bar{t}}} \quad \text{and} \quad F^{t\bar{t}}_{\text{Fake}} = \frac{N_G - N_G^{WZ,\,ZZ,\,Z+\text{jets}}}{N_P - N_P^{WZ,\,ZZ,\,Z+\text{jets}}} , \tag{7.3}$$

for the $Z+\text{jets}$ and $t\bar{t}$ FFRs.

***Z*+jets fake factor region** The region definition is built up on the SR baseline object selection criteria, defined in Table 6.3, to keep the modeling of the fake contribution as accurate as possible. The Z boson candidate is found using a SFOC lepton pair made of *good* leptons with high p_{T}. The di-lepton mass must not differ more than 20 GeV from the Z pole mass. In case of more di-leptons are constructed, the one with the smallest difference is chosen. Only events with one

additional lepton to the di-lepton are taken for the F_{Fake} calculation. Evaluation of the *additional* leptons and possible jets encloses the FFR definition, summarized in Table 7.4. The *additional* lepton is evaluated as *good* or *poor* according to the same prescription used in the FCR. The event selection requires the missing transverse energy below 25 GeV to suppress the processes including a W boson.

Category	Cut Name	Requirement
	... (Baseline Object Selection of SR)	
Di-lepton Selection	Pairing	SFOC pair of leptons
	Lepton isolation	`FixedCutLoose` working point
	Impact parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$ (3) for electron (muon)
	Electron identification	`LooseLH` working point
	Lepton momentum	$p_{\text{T}} > 25\,\text{GeV}$
	Z window	$71 < m_Z < 101\,\text{GeV}$
	Ranking	Smallest $\lvert m_{\ell\ell} - m_Z\rvert$
Event Selection	Missing energy	$E_{\text{T}}^{\text{miss}} < 25\,\text{GeV}$
	Lepton counting	$N_\ell = 3$
Additional Lepton	*Good*/*poor* lepton	See Table 7.3
Jets Selection	Overlap Removal	Remove jets with $\Delta R_{\ell\text{-}j} < 0.4$ to *good*/*poor* leptons

Table 7.4: FFR definition for Z + jets processes. The definition is composed of the SR baseline object selection, Z boson tagging, and *additional* leptons evaluation.

The fake factor for Z+jets processes is studied as a function of pseudo-rapidity η and transverse momentum p_{T}, shown in a two-dimensional plot in Figure 7.2a, separately as a function of each in Figures 7.2b and 13.5c, and as a function of a jet multiplicity in Figure 13.5d for electrons. The same set of plots is shown for muons in Figure 7.3. The one-dimensional projection is calculated from the two-dimensional plot as a mean value of the other axis bins. The electron and muon F_{Fake} differ in the shape and scale, which can be seen in the central region of η distribution in Figures 7.2b and 7.3b. The contrast is primarily driven by their characteristic reconstruction in the detector and different physical source. Fake lepton contribution to muons is likely due to imperfection of the isolation criterion. Imperfection of identification during the object reconstruction is behind the contribution to electrons (photons, jets). For detail of the factor calculation, the individual terms used in Equation (7.3) are shown in Figures 7.2e and 7.2f for electrons and in Figures 7.3e and 7.3e for muons. The one-dimensional plots also include a factor based exclusively on the MC simulation. The factor calculation is the same, however, MC samples (Z + jets, $t\bar{t}$, WZ, and ZZ) are used instead of the real data.

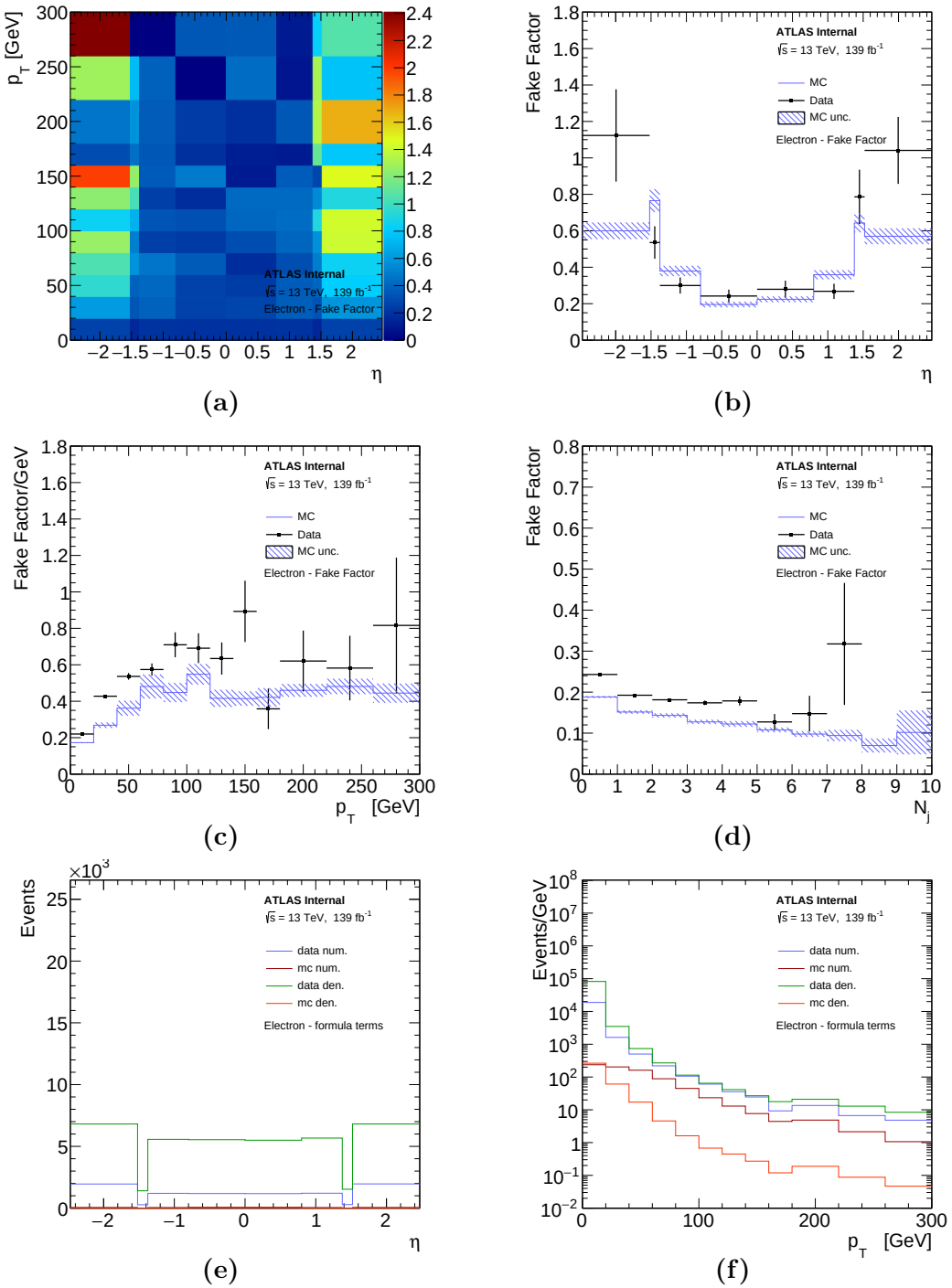


Figure 7.2: Summary of the Z + jets fake factor for electrons. The factor is shown as a function of η and p_{T} together in (a), separately in (b) and (c), and as a function of the number of jets in (d). *additional* electrons η and p_{T} distributions from FFR, as they contribute to Equation (7.3), are shown in (e) and (f) [168].

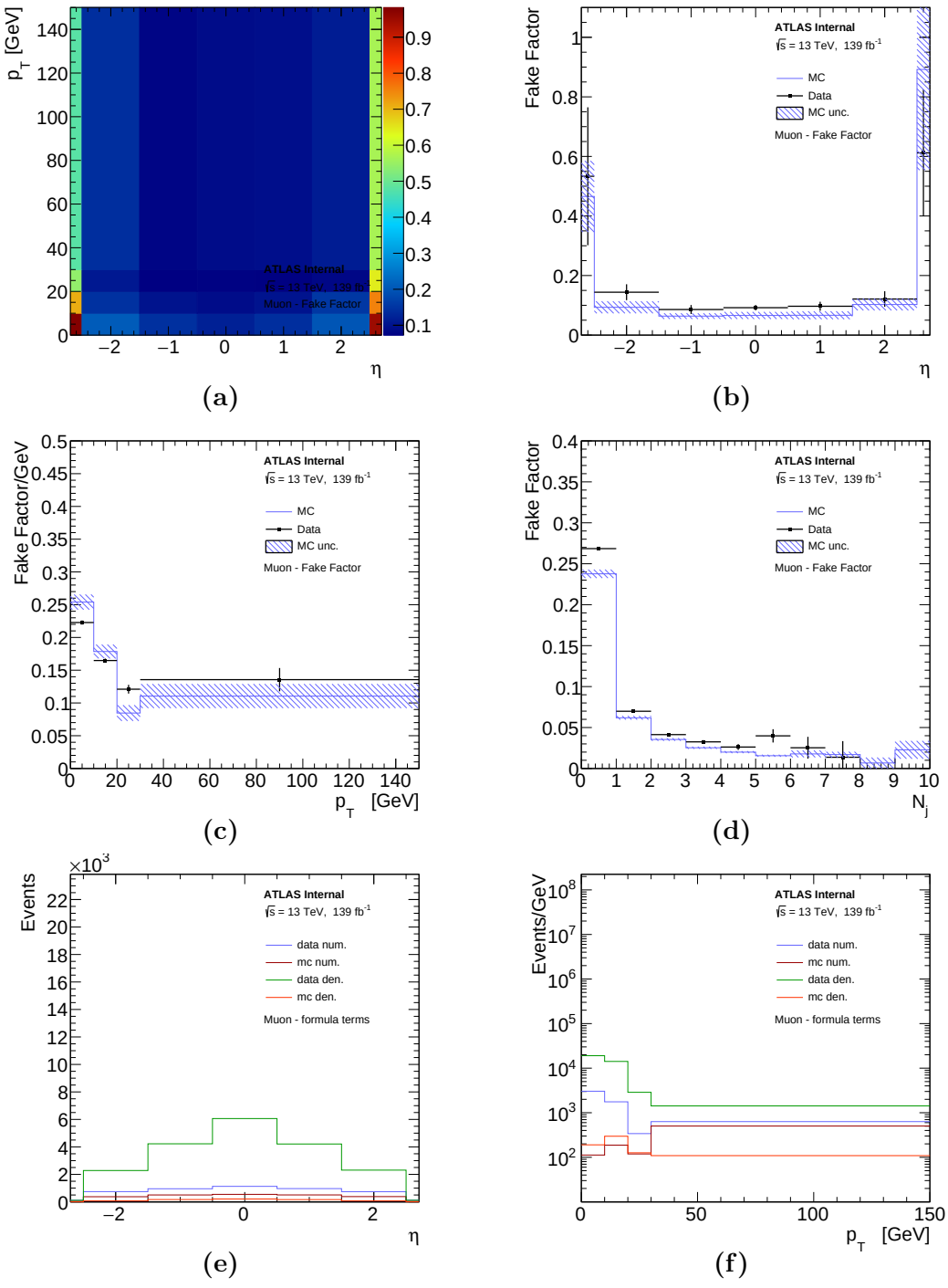


Figure 7.3: Summary of the Z + jets fake factor for muons. The factor is shown as a function of η and p_{T} together in (a), separately in (b) and (c), and as a function of the number of jets in (d). *additional* muon η and p_{T} distributions from FFR, as they contribute to Equation (7.3), are shown in (e) and (f) [168].

The electron F_{Fake} increases towards the high p_{T} region, where the real leptons are expected to dominate. The low $|\eta|$ region F_{Fake} is rather constant, while in the high region increases up to one. In the higher $|\eta|$ region the calculation faces increasing statistical uncertainty, which is expected. The electron binning is adjusted according to the barrel-endcap transition region of electromagnetic calorimeter ($1.37 < |\eta| < 1.52$), to monitor the effect on the factor. Eventually, the factor is consistent with the barrel and endcap cases.

The F_{Fake} slowly increases in the high p_{T} region of muons as well, despite the fact the MC correction in numerator starts to play a significant role, as can be seen in Figure 7.3f. The muon F_{Fake} is stable as a function of η, except of the region between 2.5 and 2.7 where the muon reconstruction quality is decreased due to the limited range of the ID.

In general, the fake factor can also be dependent on the jets multiplicity which is displayed in Figure 13.5d for electrons and in Figure 7.3d for muons. The dependence is rather constant for two or more jets, therefore the jet multiplicity is not taken into account for the calculation. The final overlap removal in the selection (Table 7.4) ensures that the fake leptons are not double counted as jets. The distributions of the *additional* leptons are shown in Appendix in Figures 13.1 and 13.2.

$t\bar{t}$ fake factor region The FFR for $t\bar{t}$ process is also based on the SR baseline object criteria. The selection chain starts with reconstruction of the W bosons using an electron-muon pair composed of *good* leptons with high p_{T}. The non-uniform pairing suppresses WW backgrounds. If more than one pair is available, the one with the smallest impact parameter significance is chosen. The rest of decay products demands additional requirements: the neutrinos $E_{\text{T}}^{\text{miss}} > 50\,\text{GeV}$ and the b-quarks at least one b-tagged jet. The events must have 3 leptons in total, resulting in one *additional* lepton. The last selection step requires W transverse mass (m_{T}^{W}) below 60 GeV to suppress $t\bar{t}W$ process contribution. The mass is defined by equation

$$m_{\text{T}}^{W} = \sqrt{2p_{\text{T}}^{\ell_3} E_{\text{T}}^{\text{miss}} \left[1 - \cos\left(\Delta\phi\left(p_{\text{T}}^{\ell_3}, E_{\text{T}}^{\text{miss}}\right)\right)\right]} \tag{7.4}$$

with the *additional* lepton ℓ_3 and missing transverse momentum $E_{\text{T}}^{\text{miss}}$ as input. All selection steps are summarized in Table 7.5.

The $t\bar{t}$ fake factor is constructed for electrons and muons independently as well. The resulting factors are displayed in Figure 7.4, on the left hand side for electrons and the right for muons. The set of plots is composed of a two-dimensional function of η and p_{T} and separately as two one-dimensional functions of each. The plots share the binning of the Z + jets factors for a straight forward combination in the

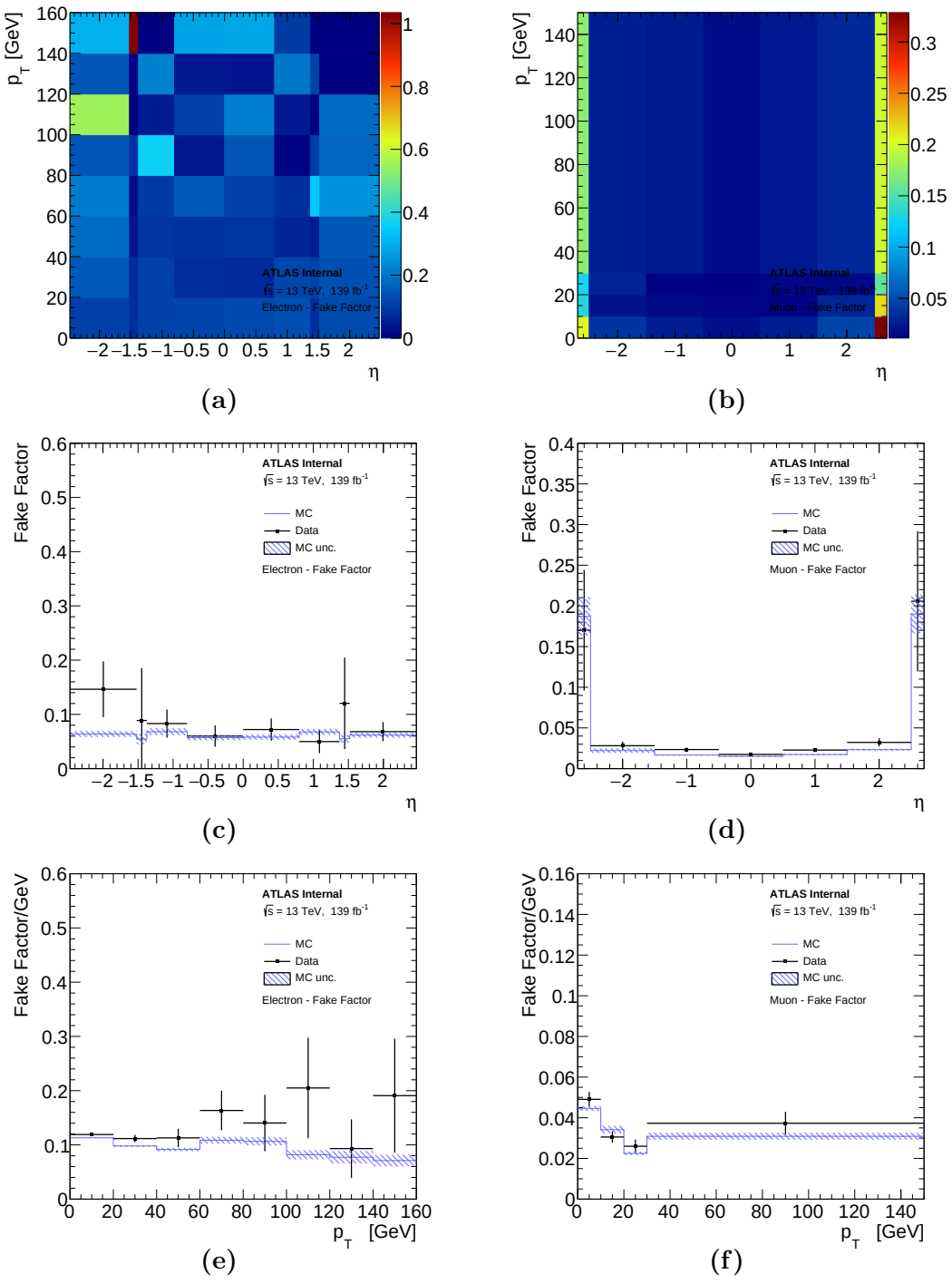


Figure 7.4: Fake factor of $t\bar{t}$ process for electrons as a two-dimensional function of η and p_T (a) and separately as a function of η (c) and p_{T} (e). The same set of plots is shown for muons on the right hand side (b), (d) and (f) [168].

Category	Cut Name	Requirement
	... (Baseline Object Selection of SR)	
Di-lepton Selection	Pairing	Electron-muon di-leptons
	Lepton isolation	`FixedCutLoose` working point
	Impact parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$ (3) for electron (muon)
	Electron identification	`LooseLH` working point
	Lepton momentum	$p_{\mathrm{T}} > 25\,\mathrm{GeV}$
	Ranking	Made of leptons with the smallest $\lvert d_0/\sigma(d_0)\rvert$
Event Selection	Missing Energy	$E_{\mathrm{T}}^{\mathrm{miss}} > 50\,\mathrm{GeV}$
	Jet flavor	At least one b-tagged jet
	Lepton momentum	$p_{\mathrm{T}} > 28\,\mathrm{GeV}$ (at least one)
	Lepton counting	$N_\ell = 3$
	W boson mass	$m_T^W < 60\,\mathrm{GeV}$ for three *good* leptons, see Equation (7.4)
Additional Lepton	*Good*/*poor* lepton	See Table 7.3
Jets Selection	Overlap Removal	Remove jets with $\Delta R_{\ell\text{-}j} < 0.4$ to *good*/*poor* leptons

Table 7.5: FFR definition for $t\bar{t}$ process. Definition is composed of SR baseline object selection, $t\bar{t}$ tagging, and *additional* leptons evaluation.

final stage of background estimation. The distributions of the *additional* leptons are shown in Appendix in Figures 13.3 and 13.4.

The electron factor is almost constant in both η and p_{T} distributions in Figure 7.4c and 7.4e. The uncertainty is higher towards higher values of $|\eta|$ and p_{T}. The muon factor has a very similar shape as the Z + jets case, although having around four times smaller scale, see Figure 7.4d.

Fake Factor Combination The factors calculated from the Z + jets and $t\bar{t}$ processes using Equations 7.3 are combined into the final fake factor according to their contributions to the FCR in the following lepton channels; electron ($eeee$), mixed ($ee\mu\mu$), and muon ($\mu\mu\mu\mu$). Each bin of the final factor is composed as a weighted arithmetic mean of the two input factors. Ratios of Z + jets over $t\bar{t}$ contributions, used for the combination, are 2.59, 0.95, and 0.74 for the electron, mixed, and muon channel. A proportion of the processes can be read off, for example, from the di-jet mass distribution in Figure 7.5. The plot also highlights the dominance of both processes over the others (WZ, ZZ). A comparison of the factors before combination is shown in Appendix in Figure 13.5.

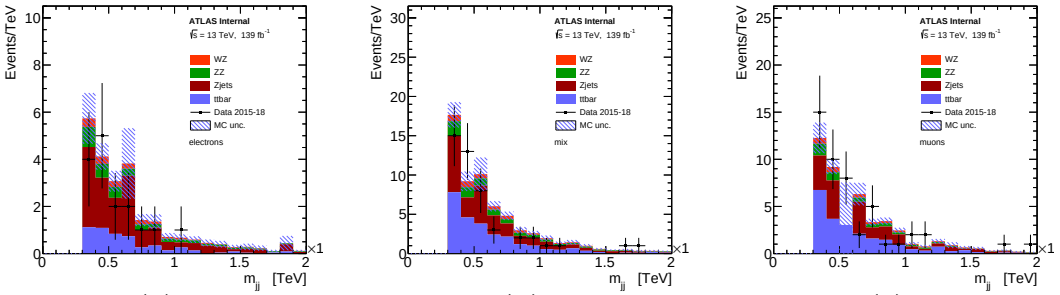


Figure 7.5: Dijet mass distribution in the FCR splited into the electron (a), mixed (b), and muon (c) channel. Only statistical uncertainty is included [168].

7.4.3 Background Yields Calculation

Once the fake factor is calculated, event yields in the FCR (N_{ggxx}) need to be extracted as the next ingredient to Equation (7.2). In Table 7.6, raw counts of the real events with one and two *poor* leptons are displayed together with counts for the MC correction (N^{ZZ}). The event yields are sorted into channels, and split by the

Counter	Channel				
	Electron	Mixed			Muon
		e-Fake(s)	μ-Fake(s)	e- and μ-Fake	
N_{gggp}	8.0	7.0	10.0	–	14.0
N^{ZZ}_{gggp}	3.3	3.7	3.7	–	4.7
N_{ggpp}	8.0	7.0	17.0	9.0	33.0
N^{ZZ}_{ggpp}	0.3	0.1	0.1	0.3	0.3

Table 7.6: FCR observed event yields split into the considered four-lepton channels. The first column is identified with the input terms of Equation (7.2).

flavor of the *poor* lepton(s) in the mixed channel. The nominal fakes background estimate is derived using the combined two-dimensional fake factor applied to the events of FCR. The estimate is shown together with the intermediate results in Table 7.7 accompanied by the statistical uncertainty. The yields in the mixed channel with one *poor* electron and one *poor* muon in the same event are not considered, since for such a case an appropriate fake factor does not exist.

Term	Channel			Inclusive
	Electron	Mixed	Muon	
$N_{gggp} \times F$ (+)	1.75 ± 0.64	2.45 ± 0.73	0.87 ± 0.24	5.07 ± 1.00
$N^{ZZ}_{gggp} \times F$ (−)	0.70 ± 0.03	0.96 ± 0.03	0.20 ± 0.01	1.86 ± 0.05
$N_{ggpp} \times F^2$ (−)	0.40 ± 0.14	0.47 ± 0.11	0.11 ± 0.02	0.98 ± 0.18
$N^{ZZ}_{ggpp} \times F^2$ (+)	0.02 ± 0.01	0.01 ± 0.00	0.00 ± 0.00	0.03 ± 0.01
N_{Fake} (Sum)	0.68 ± 0.65	1.02 ± 0.74	0.57 ± 0.24	2.27 ± 1.02

Table 7.7: Fakes background event yield estimate scaled to $139\,\text{fb}^{-1}$, broken down into terms of Equation (7.2) and split into four-lepton channels. The terms are complemented with a sign (in brackets), displaying their orientation within the final sum. The statistical uncertainty is shown along the values.

7.4.4 Systematic Uncertainties

The systematic uncertainty of the fakes background estimate originates in three main sources.

The first source is the fake factor itself. The uncertainty contribution is modeled by a propagation of its statistical uncertainty (higher and lower), reduction of the binning to a single bin approach, and alternation to the MC simulation based single-bin factor. The latter is a very conservative approach, since the MC samples were intended only for the validation of the original factor. The extended uncertainty set is driven by a small statistical sample of events in the FCR. Surprisingly, the MC based factor uncertainty resides within the range defined by the variation of statistical uncertainty.

The second source comes from the definition of inverted lepton selection. Benefiting from the fact both electron and muon selection inversions include the isolation criterion, the value of isolation was varied up and down by 50 %. The variation affects FCR and both FFRs.

The third source is the MC correction. The MC simulation weights of corresponding samples were uniformly varied up and down by 30% to cover the real lepton correction uncertainty. A similar effect would have a variation of the process cross sections.

The fakes background is not dominant for the EW $ZZjj$ production, thus the extended variation set, listed in Table 7.8, can be included in the final systematic uncertainty without significant affection of the measurement accuracy. The individual variations do not shift the estimate in the same direction for each channel. Hence the variations are symmetrized in the final stage.

The resulting fakes background estimate expects $2.27 \pm 1.02(\text{stat.}) \pm 0.86(\text{syst.})$ events in the SR. The largest uncertainty components are the variation of isolation

(−25 %) and MC correction (14 %). The systematic component of total uncertainty accounts for 37 % while the statistic component for 45 % of the nominal estimate.

Nominal estimate	0.68	1.02	0.57	2.27
Varied estimate				
Stat. unc. of F_{Fake} varied down	+0.02	−0.15	−0.06	−0.19
Stat. unc. of F_{Fake} varied up	−0.02	+0.15	+0.06	+0.18
Single bin F_{Fake}	−0.03	−0.43	+0.08	−0.38
MC based single bin F_{Fake}	−0.14	+0.39	−0.13	+0.12
Isolation varied down	+0.26	−0.47	−0.35	−0.56
Isolation varied up	+0.04	+0.08	−0.01	+0.12
MC corr. varied down	+0.02	+0.03	+0.27	+0.31
MC corr. varied up	−0.02	−0.04	−0.25	−0.31
Total systematic variation	±0.30	±0.78	±0.54	±0.86

Table 7.8: Fakes background systematic uncertainties summary. The values are shown as an absolute deviation from the nominal estimate. The total systematic variation is a result of quadratic addition of the individual variations and symmetrized.

7.4.5 Closure Test

Unfortunately, due to a small contribution of $Z+\text{jets}$ and $t\bar{t}$ into the SR a classic closure test of the method can not be done without loosing the selection criteria (omitting the jet selection). However, all contributions of the processes with less than four prompt leptons ($Z+\text{jets}$, $t\bar{t}$, WZ, $t\bar{t}W$) can serve as a theory prediction of the background with $0.68^{+0.80}_{-0.68}$ events in the SR. Although compatible with zero, the prediction can be compared to the MC simulation based fakes background estimate, which is not data-driven and can serve for the closure with estimate of 2.39 ± 1.12 events in the SR (includes only statistical uncertainty). Both values are in agreement within the uncertainty range.

Another closure test of the method can be taken from the ZZ to four lepton cross-section measurement [20] as the loosened version without jet selection. The method was also tested in the WZ to three charged-leptons cross-section measurement in [169].

7.5 Background Summary

Total background estimation includes the prompt, QCD, and fakes background contributions, scaled to $139\,\text{fb}^{-1}$ of the recorded data sample. As was plotted at the beginning of this chapter in Section 7.3, the QCD background is dominant

over the EW production as well as the remaining background. The prompt and fakes background event yield estimates are comparable. The fakes background is dominated by the statistic, while the other two by the theoretical uncertainties. Estimates of all backgrounds are listed in Table 7.9, including statistical and systematic uncertainties.

Background	Estimate
Prompt	$4.64^{+1.05}_{-0.76}$
QCD	$71.71^{+21.08}_{-14.75}$
Fakes	$2.27^{+1.33}_{-1.33}$
Total	$78.62^{+21.15}_{-14.83}$

Table 7.9: Summary of the background event yield estimates in the SR scaled to $139\,\text{fb}^{-1}$, including statistical and systematic uncertainties.

A total background for the inclusive cross-section measurement accounts for 7 % of the whole yield. The EW production search background amounts of 82 % with the EW-QCD ratio of 0.24. All numbers are related to the SR.

Chapter 8

Systematic Uncertainties

A number of systematic uncertainties affect the accuracy of the analysis results. The sources are divided to two categories: *experimental*, rising from the measurement of physical quantities, reconstruction methods of physical objects, and instrumentals of the experiment; and *theoretical*, originating from the physics modeling, proton structure, fundamental QCD parameter (α_{QCD}), parton showering, and hadronization.

The uncertainties mentioned in this chapter, are evaluated primarily for the theory predictions given by the $ZZjj$ EW production defined as the signal and the quark- and gluon-induced $ZZjj$ QCD production defined as the QCD background in Chapter 7. These processes are dominant in the SR and their MC samples were used for the training of the MVA methods, described in Chapter 6.

The uncertainty calculation exploits the combined MC sample for full Run 2 scaled to $139\,\mathrm{fb}^{-1}$. Total numbers are derived from the BDT score distribution in SR after application of the uncertainty estimation techniques, such as symmetrization etc. The BDT score is preferred since it serves as an input for the signal strength extraction in the final stage of the analysis.

8.1 Experimental Uncertainties

Accounted experimental systematic uncertainties are listed according to the source objects for electrons, muons, and jets. Uncertainties characteristic for the experiment are listed separately, namely for the integrated luminosity and pile-up.

The momentum measurement of physical objects plays the most important role in the measurement. The uncertainty effect is applied in a form of smearing and scale variation of associated object properties before the event selection. Therefore, the observables subjected to selection criteria can change and cause an event would be shifted beyond the acceptance of SR.

Another type of systematic uncertainties is applied at the end of the selection process in a form of a scale factor. Such uncertainties are associated with several efficiencies (reconstruction, isolation etc.) changing the final selected event weight.

8.1.1 Electrons

The momentum related uncertainties consider the energy resolution and energy scale of the measured electron. The uncertainties are generated by a fully correlated model providing a single parameter for each, denoted as `EG_RESOLUTION_ALL` and `EG_SCALE_ALL`.

The efficiency related uncertainties consider electron identification, isolation, and reconstruction. The electron trigger efficiency uncertainty is not considered, since the efficiency for the $ZZjj$ final state passing the ATLAS trigger menu is above 99 %. The three uncertainties are denoted as `EL_EFF_ID`, `EL_EFF_Iso`, and `EL_EFF_Reco`.

Details of Run 2 electron uncertainties can be found in [158]

8.1.2 Muons

The muon momentum scale variation is covered by a single uncertainty `MUON_SCALE`, while the smearing is independent for tracks in the Inner Detector (ID) denoted as `MUON_ID` and Muon Spectrometer (MS) denoted as `MUON_MS`. The charge dependent smearing of momentum is based on track sagitta corrections covered by `MUON_SAGITTARHO` uncertainty, complemented by a residual charge bias after the correction addressed by `MUON_SAGITTA_RESBIAS` uncertainty.

The efficiency uncertainties are considered for isolation, reconstruction, and Track-to-vertex Association (TTVA). All three sources are evaluated individually for the statistical uncertainty as: `MUON_EFF_ISO_STAT`, `MUON_EFF_RECO_STAT`, and `MUON_EFF_TTVA_STAT`; and for the systematic components of uncertainty as: `MUON_EFF_ISO_SYS`, `MUON_EFF_RECO_SYS`, and `MUON_EFF_TTVA_SYS`. In addition, the reconstruction based uncertainty is evaluated for low transverse momenta (p_{T}) for both components as `MUON_EFF_RECO_STAT_LOWPT` and `MUON_EFF_RECO_SYS_LOWPT`. The reconstruction methods and their associated uncertainties are described in [126]. The muon trigger efficiency for the studied final-state is again nearing 100 %, thus not included.

8.1.3 Jets

A Jet Energy Scale (JES) uncertainty set is combined from results of a number of auxiliary measurements. Several *in situ* techniques are utilized, for example, the evaluation of transverse momentum balance between a jet and reference object

such as a Z boson or photon [170, 171], a study of multiple proton-proton interactions, high-transverse momentum jets, quark- and gluon-initiated jets [172], near-by and isolated jets [173], and others, resulting in approximately 100 uncertainty parameters. The JES working point chosen for the analysis consists of 8 uncertainties recommended for full ATLAS Run 2 analyses.

The reduced set of uncertainties comes from a pseudorapidity (η) intercalibration, flavor response, multijet-balance and Z/γ + jet calibration, when the last two are grouped to reduce overall demands on CPU time expensive systematic uncertainty processing [174]. The η-intercalibration uncertainty rise from a forward jets calibration based on di-jet events with one jet in the central region of detector. The uncertainty is divided into negative and positive pseudorapidity, `JET_EtaInt_NC_negEta` and `JET_EtaInt_NC_posEta`, and also for high energy and a case determined by data acquired in 2018 only, `JET_EtaInt_NC_highE` and `JET_EtaInt_NC_2018data`. The flavor response considers gluon-initiated jets and is reflected in `JET_Flavor_Response` uncertainty. The grouping of uncertainties results in three parameters denoted as `JET_GroupedNP_1-3`.

The second dominant source of uncertainty originates in a Jet Energy Resolution (JER). The methods are fundamentally similar to the techniques mentioned for JES, exploiting energy resolution techniques of a di-jet system performed by previous experiments at Sp$\bar{\text{p}}$S and Tevatron [175, 176] like the di-jet balance and bisector method. JER working point uses a reduced set of systematic uncertainties with eight parameters.

The JER uncertainty depend on the calorimeter resolution, noise, and bias and are grouped to six primary parameters `JET_JER_EffectiveNP_1-6` and a complementary `JET_JER_EffectiveNP_7restTerm` parameter [161]. The set of uncertainties is finalized with a parameter `JET_JER_DataVsMC_MC16` derived from a comparison of real and simulated data using `MC16` prescription.

Uncertainty associated with an identification of pile-up jets has a form of scale factor. The identification is based on the association of a jet with the primary vertex of evaluated event, using the JVT [162] with the uncertainty denoted as `JVT_EFF`.

8.1.4 Collision Multiplicity

A global pile-up scale factor uncertainty, derived from the profiles of particular MC modeling campaign (described in Section 4.4), is denoted as `PRW_DATASF`. The pileup reweighting variation [177] of MC samples is applied on the ratio between the predicted and measured inelastic proton-proton cross-section in fiducial phase space defined in [178].

8.1.5 Luminosity

Uncertainty on the integrated luminosity affects all MC samples and was determined to be 1.7 % for ATLAS collaboration analyses exploiting the complete data sample from Run 2 data-taking [122]. Luminosity measurement is briefly explained in Section 3.3.9.

8.1.6 Summary

All experimental uncertainties considered in the analysis are listed in Table 8.1 relative to the nominal yield in the SR. The EW production signal is listed twice for the nominal theory prediction generated by MadGraph and the alternative by Sherpa. The QCD background is listed separately for the quark- and gluon-induced QCD productions, both generated by Sherpa. Details of the four MC samples of the predictions are listed in Table 4.4.

Both signal predictions are dominated by the electron identification uncertainty (`EL_EFF_ID`) with approximately 2 %, however the alternative experiences too high uncertainty of 3.3 % in case of the pile-up reweighting (`PRW_DATASF`), which is accounted for the Sherpa misconfiguration described in Section 5.3.2.

The background predictions are dominated by the jet energy scale uncertainties (`JET_Grouped1-2`) with approximately 8 % in quark- and 9.5 % in gluon-induced case. The signal predictions have smaller jet related uncertainties, because of quite demanding requirements on the di-jet system, which is constructed from the two leading jets of event. This makes the jet energy smearing variation less likely to change the SR yield.

8.2 Theoretical Uncertainties

The theoretical uncertainties are bound to the MC sample generation as well as the theory they are modeling. Input parameters consider QCD scales, parton distribution functions, parton showering parameters and even a different event generator itself.

8.2.1 Strong Coupling

Uncertainty on the accuracy of strong coupling (α_S), called `ALPHA_S`, is calculated by a variation of the value itself. A value chosen by ATLAS collaboration for Run 2 of LHC corresponds to 0.118 at energy of Z boson mass, in accordance with Reference [179], listed in Review of Particle Physics from 2016 [180]. Considered up and down variations are 0.119 and 0.117. Unfortunately, this option is available only for MC samples produced by Sherpa generator.

Uncertainty	EW signal		QCD background	
	Nominal	Alternative	Quark-induced	Gluon-induced
EG_RESOLUTION_ALL	0.0	$^{+0.0}_{-0.1}$	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.0}$
EG_SCALE_ALL	$^{+0.0}_{-0.1}$	$^{+0.0}_{-0.1}$	0.0	$^{+0.1}_{-0.0}$
EL_EFF_ID	$^{+2.2}_{-2.1}$	$^{+2.1}_{-2.0}$	$^{+2.0}_{-1.9}$	2.1
EL_EFF_Iso	0.2	0.2	0.2	0.2
EL_EFF_Reco	0.5	0.5	0.5	0.5
MUON_EFF_ISO_STAT	0.1	0.1	0.0	0.0
MUON_EFF_ISO_SYS	0.5	0.5	0.5	0.5
MUON_EFF_RECO_STAT	0.2	0.2	0.2	0.2
MUON_EFF_RECO_STAT_LOWPT	0.0	0.0	0.0	0.0
MUON_EFF_RECO_SYS	0.6	0.6	0.6	0.6
MUON_EFF_RECO_SYS_LOWPT	0.1	0.1	0.1	0.1
MUON_EFF_TTVA_STAT	0.1	0.1	0.1	0.1
MUON_EFF_TTVA_SYS	0.1	0.1	0.1	0.1
MUON_ID	0.0	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	0.0
MUON_MS	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	$^{+0.0}_{-0.1}$
MUON_SAGITTA_RESBIAS	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$
MUON_SAGITTA_RHO	$^{+0.1}_{-0.0}$	0.0	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$
MUON_SCALE	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.0}$	0.1	$^{+0.0}_{-0.1}$
JET_EtaInt_NC_2018data	$^{+0.1}_{-0.0}$	$^{+0.2}_{-0.0}$	0.2	$^{+0.3}_{-0.2}$
JET_EtaInt_NC_highE	0.0	0.0	0.0	0.0
JET_EtaInt_NC_negEta	0.0	0.0	0.0	$^{+0.0}_{-0.1}$
JET_EtaInt_NC_posEta	0.0	0.0	0.0	0.0
JET_Flavor_Response	0.5	$^{+0.5}_{-0.4}$	$^{+3.2}_{-2.9}$	$^{+3.8}_{-3.6}$
JET_GroupedNP_1	$^{+1.2}_{-1.1}$	1.0	$^{+8.1}_{-6.4}$	$^{+9.1}_{-7.9}$
JET_GroupedNP_2	1.2	1.2	$^{+8.1}_{-6.5}$	$^{+9.4}_{-8.4}$
JET_GroupedNP_3	0.1	0.1	$^{+0.4}_{-0.3}$	$^{+0.2}_{-0.3}$
JET_JER_DataVsMC_MC16	0.0	$^{+0.3}_{-0.0}$	$^{+0.4}_{-0.0}$	$^{+0.4}_{-0.0}$
JET_JER_EffectiveNP_1	$^{+0.2}_{-0.0}$	$^{+0.3}_{-0.0}$	$^{+5.3}_{-0.0}$	$^{+4.0}_{-0.0}$
JET_JER_EffectiveNP_2	$^{+0.1}_{-0.0}$	$^{+0.3}_{-0.0}$	$^{+3.7}_{-0.0}$	$^{+2.6}_{-0.0}$
JET_JER_EffectiveNP_3	$^{+0.1}_{-0.0}$	$^{+0.4}_{-0.0}$	$^{+2.8}_{-0.0}$	$^{+2.2}_{-0.0}$
JET_JER_EffectiveNP_4	0.0	$^{+0.3}_{-0.0}$	$^{+1.7}_{-0.0}$	$^{+0.6}_{-0.0}$
JET_JER_EffectiveNP_5	0.0	$^{+0.2}_{-0.0}$	$^{+1.0}_{-0.0}$	$^{+0.4}_{-0.0}$
JET_JER_EffectiveNP_6	0.0	$^{+0.1}_{-0.0}$	$^{+0.7}_{-0.0}$	$^{+0.4}_{-0.0}$
JET_JER_EffectiveNP_7restTerm	$^{+0.1}_{-0.0}$	$^{+0.3}_{-0.0}$	$^{+1.6}_{-0.0}$	$^{+0.6}_{-0.0}$
JVT_EFF	0.1	0.1	0.1	0.2
PRW_DATASF	$^{+0.6}_{-0.7}$	$^{+3.5}_{-3.7}$	$^{+2.6}_{-1.9}$	$^{+2.4}_{-1.9}$
LUMI	1.7	1.7	1.7	1.7

Table 8.1: Summary of experimental uncertainties of the nominal and alternative theory prediction of the signal, as well as of the QCD background predictions. Uncertainties are shown relative to the nominal yield in the SR in %.

8.2.2 Parton Distribution Functions

Treatment of uncertainties on PDFs considers only variation sets recommended for Run 2 of LHC elaborated in Reference [179]. Both, a nominal PDF set uncertainty `PDF` and PDF set variation uncertainty `ALT_PDF` are included. Further mentioned PDF sets are accessible in LHAPDF database documented in [181]. The PDF uncertainties are calculated only for ZZ production MC samples listed in Table 4.4 as well as Tri-boson in Table 4.6. The MC simulation option with MadGraph generator employs the NNPDF3.0 LO PDF set [151] for the nominal, while the CT10 LO [182] and MMHT2014 LO [183] for the uncertainty estimation. The option with Sherpa generator employs NNPDF3.0 NNLO as the nominal set, whereas the CT14 NNLO [184] and MMHT2014 NNLO as the uncertainty input.

The nominal PDF sets consist of approximately a hundred variations, computed by individual MC replicas (with varied PDF parameters), which need to be combined in to the final uncertainty. A combination formula follows the prescription from Reference [181] for N replicas with central value F_0 as

$$\sigma_F^+ = \sigma_F^- = \left[\frac{1}{N-1}\sum_{k=1}^{N}(F_k - F_0)^2\right]^{1/2}.$$

The alternative PDF uncertainty is calculated from the maximal deviation of central values of alternative PDF sets and symmetrized around the nominal.

8.2.3 QCD Scales

The uncertainty on QCD scales employs a standard set of simultaneous variations of factorization and renormalization scales by a factor of two up and down, excluding the cases of contra-variation, as can be seen in Table 8.2. An uncertainty `QCD_SCALE` is constructed using a maximal deviation of the variations from the nominal value, resulting in a asymmetrical envelope.

	$0.5\mu_\mathrm{r}$	$1\mu_\mathrm{r}$	$2\mu_\mathrm{r}$
$0.5\mu_\mathrm{f}$	✓	✓	✗
$1\mu_\mathrm{f}$	✓	✓	✓
$2\mu_\mathrm{f}$	✗	✓	✓

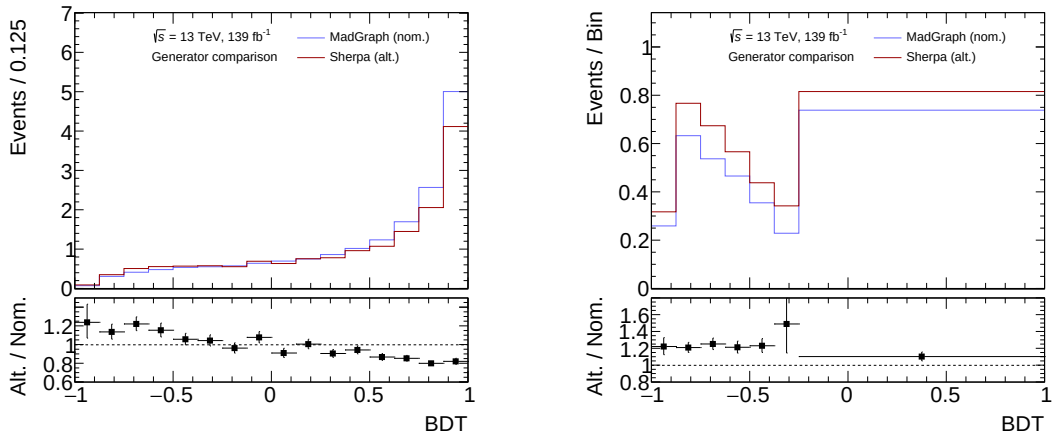
Table 8.2: Standard set of the QCD factorization (μ_f) and renormalization (μ_r) scales variations with the nominal value at the center.

8.2.4 Cross Section

Due to a scaling of the gluon-induced QCD background sample to NLO by a factor 1.67, an additional uncertainty on the process cross-section of 25 % is applied, as is recommended in [152]. A cross-section uncertainty is also applied to the Triboson MC samples, following WZ EW production research [17]. The uncertainty is denoted by `XSECT`.

8.2.5 Alternative Generator

Although, the theory prediction produced by the SHERPA generator contains the misconfiguration (described in Section 5.3.2), it is the only alternative available. The sample is not reliable for the theory prediction, but can be used for estimation of the event generator uncertainty, denoted as `ALT_GEN`. The uncertainty is calculated as a difference between the nominal (MADGRAPH) and alternative (SHERPA) theory predictions and symmetrized around the nominal. A comparison of both predictions as a function of the BDT score, can be seen in Figure 8.1. The alternative generator underestimates the higher region of the BDT score in the SR and globally overestimates the yields in the QCD-CR, which is in agreement with the findings from Section 5.3.2.



(a) **(b)**

Figure 8.1: Comparison of MC simulated signal prediction produced by nominal (MADGRAPH) and alternative (SHERPA) generators.

8.2.6 Parton Showering

The parton showering uncertainty is taken from the ATLAS research of $ZZjj$ EW production [1]. The uncertainty is determined using different generators, PYTHIA 8 and HERWIG 7, and is denoted by `PS`. The uncertainty was calculated for the signal using a private MC samples and the resulting shape was reused for estimation of the quark-induced QCD background as well. This approach was chosen due to a large number of events necessary for a direct calculation. Overall uncertainty was estimated to be 3.3 %.

8.2.7 Summary

All theoretical uncertainties are summarized in Table 8.3 relative to the nominal yield in the SR.

The nominal signal prediction is dominated by the alternative generator uncertainty (`ALT_GEN`) with 12 %. The alternative prediction is dominated by the QCD scales uncertainty (`QCD_SCALE`) with 7 %. The QCD coupling uncertainty (`ALPHA_S`) is minimal for the theoretical predictions of EW production, which is expected.

The QCD background theory predictions are dominated by the QCD scales uncertainty as well, with 30 % for quark-induced and 40 % for the gluon-induced QCD background to which also the cross-section uncertainty significantly contributes.

Uncertainty	EW signal		QCD background	
	Nominal	Alternative	Quark-induced	Gluon-induced
`ALPHA_S`	0.0	$^{+0.1}_{-0.2}$	$^{+2.5}_{-2.4}$	$^{+1.7}_{-1.9}$
`PDF`	5.9	2.0	1.2	1.6
`QCD_SCALE`	$^{+6.5}_{-5.9}$	$^{+7.2}_{-6.3}$	$^{+30.1}_{-20.7}$	$^{+39.2}_{-25.5}$
`ALT_PDF`	0.9	2.3	2.1	0.5
`ALT_GEN`	12.4	–	–	–
`XSECT`	–	–	–	25.0
`PS`	3.3	–	–	–

Table 8.3: Summary of theoretical uncertainties of the nominal and alternative theory prediction of the signal, as well as of the QCD background predictions. Uncertainties are shown relative to the nominal yield in the SR in %.

8.3 Uncertainties Comparison

All uncertainties described in the chapter are grouped and displayed in Table 8.4. The instrumental uncertainties are naturally grouped, according to the sections in the description. Luminosity & pile-up group includes also the `JVT_EFF`, QCD parameters group consists of the QCD coupling and QCD scales, and Generator & parton showering were grouped since the uncertainty on parton showering is also based on the simulation tools comparison.

Uncertainty	EW signal		QCD background	
	Nominal	Alternative	Quark-induced	Gluon-induced
MC sample statistics	0.3	1.3	0.7	0.9
Electron efficiency	2.2	$^{+2.2}_{-2.1}$	$^{+2.1}_{-2.0}$	$^{+2.2}_{-2.1}$
Electron scale & resolution	$^{+0.0}_{-0.1}$	$^{+0.0}_{-0.1}$	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.0}$
Muon efficiency	0.8	0.8	0.8	0.8
Muon scale & resolution	$^{+0.2}_{-0.0}$	$^{+0.2}_{-0.0}$	$^{+0.2}_{-0.1}$	0.2
Jet energy scale	$^{+1.8}_{-1.7}$	$^{+1.7}_{-1.6}$	$^{+11.8}_{-9.6}$	$^{+13.6}_{-12.1}$
Jet energy resolution	$^{+0.2}_{-0.0}$	$^{+0.8}_{-0.0}$	$^{+7.5}_{-0.0}$	$^{+5.3}_{-0.0}$
Luminosity & pileup	1.8	$^{+3.9}_{-4.0}$	$^{+3.1}_{-2.5}$	$^{+2.9}_{-2.5}$
Total - experimental	$^{+3.5}_{-3.4}$	5.1	$^{+14.5}_{-10.2}$	$^{+15.1}_{-12.6}$
QCD parameters	$^{+6.5}_{-5.9}$	$^{+7.2}_{-6.3}$	$^{+30.2}_{-20.8}$	$^{+39.2}_{-25.6}$
Cross section & PDF	6.0	3.0	2.4	25.1
Generator & parton showering	12.8	–	–	–
Total - theoretical	$^{+15.7}_{-15.4}$	$^{+7.9}_{-7.2}$	$^{+30.3}_{-21.0}$	$^{+46.5}_{-35.8}$
Total	$^{+15.9}_{-15.7}$	$^{+9.3}_{-8.7}$	$^{+33.6}_{-23.3}$	$^{+48.9}_{-37.9}$

Table 8.4: Summary of all uncertainties considered in the analysis for the nominal and alternative theory prediction of the signal, as well as of the QCD background predictions. Uncertainties are grouped to categories and shown relative to the nominal yield in the SR in %.

On the signal side, the dominant source of uncertainty comes from the Generator & parton showering, which is tightly connected with the Sherpa misconfiguration. The second dominant source comes from the QCD scale uncertainty.

The background is dominated by the jet and theory uncertainties, which is expected since the necessary NNLO accuracy is achieved by the scaling from lower orders. The jet uncertainty difference between signal and background is expected since the EW production consists primarily of jets from the hard process.

Finally, a comparison of total uncertainties of the experimental and theoretical sides significantly emphasizes the theory uncertainty.

Chapter 9

EW Production Observation

9.1 Introduction

The main goal of the presented analysis is the first observation of the $ZZjj$ EW production in the four-lepton channel. The production represents a signal, whose strength is being extracted together with a statistical significance of exclusion of the background-only hypothesis. The significance is extracted in two ways: as expected significance, not using the real events in the SR, an approach further denoted as blinded; and as observed significance, using a complete set of events. A utilized statistical method is based on a maximum likelihood estimation exploiting a profile likelihood ratio. A process of optimization of the method configuration was executed as blinded to avoid the observer bias. The method principles, maximum likelihood estimator finding (fitting), and optimization are described in the chapter together with the uncertainties treatment and observed significances.

9.2 Method

The method builds up the statistical model using a binned likelihood function considering a Poisson distribution in every bin of input observable. The expected value for each bin n_i of a real distribution is parameterized by a single parameter μ encoding the signal strength in formula

$$E(n_i) = \mu s_i + b_i, \tag{9.1}$$

where s_i represents the theoretical prediction of signal in the i-th bin, b_i is reserved for the background yields. The variable μ represents so called parameter of interest (PoI). The yield predictions s_i and b_i are functions of secondary parameters, the so-called nuisance parameters (NP), such as a total background yield and shape,

primarily constrained by real data. Another group of NPs are rising from systematic uncertainties, constrained by auxiliary measurements. The presented analysis include number of NP coming from different sources, further denoted by a vector $\boldsymbol{\theta}$. The likelihood function for a SR input histogram with N bins is defined by

$$L_{\text{PoI}}(\mu, \boldsymbol{\theta} \,|\, n_1 \dots n_N) = \prod_{i=1}^{N} \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))}.$$

The systematic uncertainty NP has to be added using an additional likelihood function in a form of Gaussian distribution. The NP function adds a new dimension and defines a profile of the original model, therefore the resulting function is called profile likelihood. The profile broadens the likelihood function which reflects the effect of systematic uncertainty as a loss of information about the parameter of interest. The presented analysis includes a number of NP, which formulates a multi-dimensional problem. The profiling results in a likelihood function of the SR in a formula

$$L_{\text{SR}}(\mu, \boldsymbol{\theta} \,|\, \boldsymbol{n}) = L_{\text{PoI}}(\mu, \boldsymbol{\theta} \,|\, \boldsymbol{n}) L_{\text{NP}}(\boldsymbol{\theta}).$$

The input for the NP likelihood function is delivered in a form of plus and minus one σ (standard deviation) variation of uncertainty, having the mean value identical to the nominal value and probabilistic character. For details on the Gaussian distribution treatment see [185].

The background modeling NPs are constrained by the real and simulated data in a dedicated control region called QCD Control Region (QCD-CR) already mentioned in Section 7.3. The region is constructed in a way to enhance the QCD background with EW-QCD ratio only, 0.025 an order of magnitude less than in the SR. Therefore, the likelihood function of QCD-CR is not considered μ dependent and is defined similarly as in the case of SR resulting in a formula for the overall likelihood function

$$L(\mu, \boldsymbol{\theta} \,|\, \boldsymbol{n}) = L_{\text{SR}}(\mu, \boldsymbol{\theta} \,|\, \boldsymbol{n}) L_{\text{QCD-CR}}(\boldsymbol{\theta} \,|\, \boldsymbol{n}). \tag{9.2}$$

The expected value in Equation (9.1) is defined in accordance to a null hypothesis formulation with $\mu = 0$ (H_0), where no signal events contribute to the SR, and an alternative hypothesis with $\mu > 0$ (H_1), where some events in SR have an origin in the EW $ZZjj$ production. A hypothesized value is tested with a profile likelihood ratio defined as

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})},$$

where the $\hat{\hat{\boldsymbol{\theta}}}$ denotes a value of $\boldsymbol{\theta}$ that maximizes L for a specified μ, thus is a function of the parameter of interest. The numerator is called conditional maximum likelihood function. The $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are maximum likelihood estimators simultaneously maximizing the function in denominator. Process of finding the estimators is called fitting in the following text.

The λ function expresses a compatibility between the observed data and μ in a range between zero and one. To obtain a measure of incompatibility of the μ with the null hypothesis another sample statistic is used, in a form of equation

$$t_\mu = -2 \ln \lambda(\mu).$$

A minimum value of zero ($t_\mu = 0$) is compatible with the $\hat{\boldsymbol{\theta}}$, while a value of one gives one σ range around the minimum. According to the Wilks and Wald theorems the statistic follows a non-central chi-square distribution [186], which allows a calculation of equivalent significance as a significance of H_0 rejection.

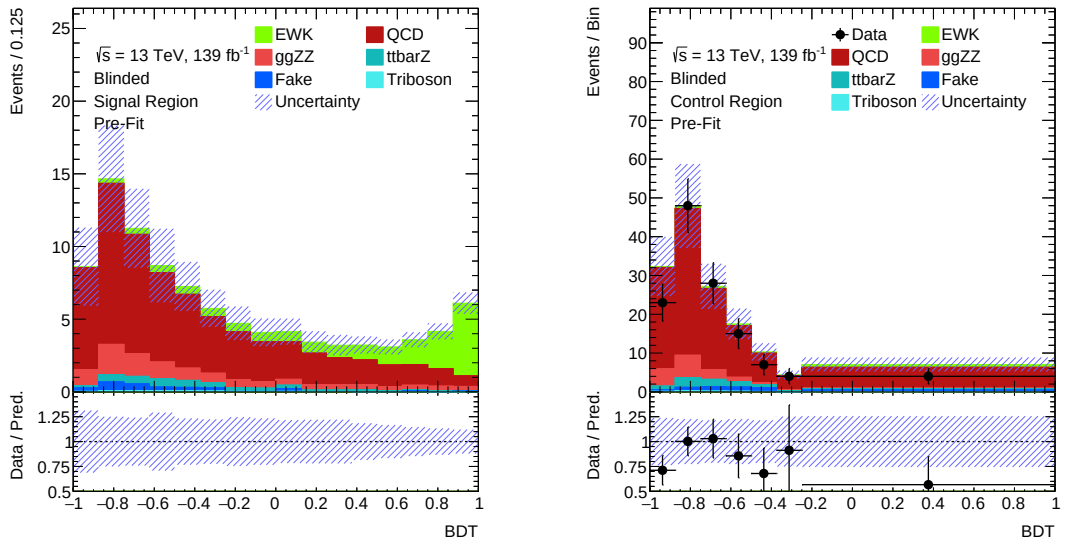
9.3 Fit Input

For the fitting procedure, a framework developed in the ATLAS Collaboration called TRExFitter is used. The framework is based on the ROOT Data Analysis Toolkit [187].

The BDT score distribution is taken as an input for the fitting because of the advanced capability of signal-background separation. The BDT score generation is described in Section 6.7. Pre-fit distributions of theoretical predictions in the SR and both real data and predictions in the QCD-CR are shown in Figure 9.1. The binning is adopted from [188]. Due to time-expensive processing of MC samples for systematic uncertainty evaluation, only samples contributing by at least one percent of the major contributing sample are included in the fitting. The major process is the quark-induced $ZZjj$ QCD production, so the limit is 0.6 event in the SR. Yields of individual processes are summarized in table Table 9.1, the observed yield in the SR is blinded. The blinding disables the real data input in the SR, therefore, it is replaced with the Asimov data generated by a likelihood defined by the total maximum likelihood estimator.

9.4 Uncertainty Treatment

All uncertainties listed in Chapter 8 are included in the fitting procedure and treated as correlated between the SR and QCD-CR ((9.2)), in the beginning of optimization. The process of optimization is treated with blinded data in the SR.



(a) **(b)**

Figure 9.1: BDT score distributions in the SR (a) and QCDCR (b).

MC sample	Event yield	
	SR	QCD-CR
EW $ZZjj$	$17.52^{+2.74}_{-2.69}$	3.22 ± 0.68
QCD $ZZjj$ (Quark-induced)	$60.58^{+20.37}_{-14.13}$	$114.81^{+34.31}_{-24.69}$
QCD $ZZjj$ (Gluon-induced)	$11.13^{+5.44}_{-4.22}$	$14.75^{+7.14}_{-5.50}$
$t\bar{t}$Z	$3.86^{+0.35}_{-0.26}$	$8.21^{+0.33}_{-0.38}$
Fakes background	$2.27^{+1.33}_{-1.33}$	$4.75^{+2.55}_{-2.55}$
Tri-boson	$0.61^{+0.20}_{-0.17}$	$0.97^{+0.31}_{-0.26}$
MC Total	$95.97^{+21.31}_{-15.05}$	$146.71^{+35.15}_{-25.44}$
Real data	–	129.00

Table 9.1: The SR and QCD-CR observed and predicted event yields composition split by individual processes. The values are scaled to $139\,\text{fb}^{-1}$ including statistical and systematic uncertainties in the simulated data and statistical uncertainty in the real data.

In the following text, a different scenarios of uncertainty treatment are probed. A nominal scenario use both shape and normalization for each systematic uncertainty. Alternatives exclude the normalization of theoretical uncertainties of the EW signal and quark-induced QCD background predictions towards a pure shape variation used for example in [15]. The normalization is dropped individually for

- uncertainties of EW signal prediction,
- quark-induced QCD background prediction,
- and for both uncertainties simultaneously.

Effect of the options on the expected significance is minimal (at the third digit of the significance). Impact on the test statistic t_μ is shown in Figure 9.2 for the normalization exclusion in the case of EW signal prediction. The statistic function has changed to be more symmetrized having an exponential character towards the high values. The changes exclusive to the background prediction uncertainties has no effect on the significance neither on the statistic. The test statistic is

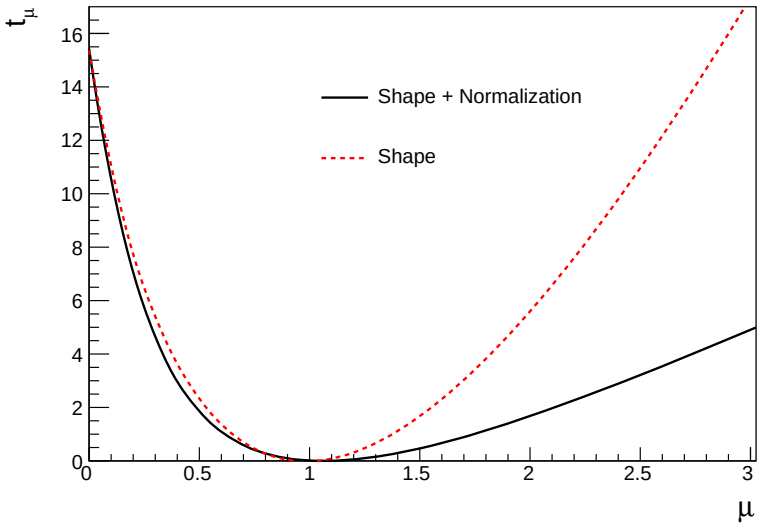


Figure 9.2: Graph of test statistic t_μ affected by exclusion of normalization of theoretical uncertainties of EWK signal sample.

expected to have a quadratic-like behavior around minimum, which is restored by the exclusion. Therefore, the change of theory uncertainties (normalization drop) exclusive for the EW prediction is kept in the fitting setup.

The post-fit distributions experience a decrease in the overall uncertainty in the QCD-CR and a decrease in lower bins in the SR, visible in Figure 9.1 and 9.3. The decrease is caused by a constraint from the data in QCD-CR. On the other hand, an increase in the high BDT region is clearly visible in the SR, which is caused by the fitting itself. However, the decrease may point to an over-constraint despite the fact, it is not visible in pre- and post-fit pull plots.

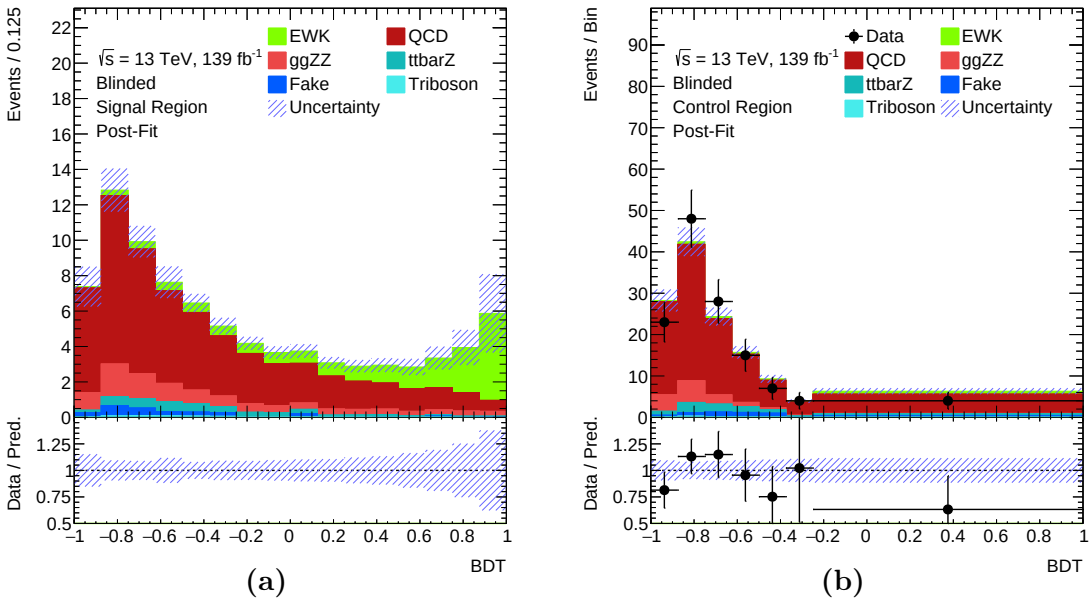


Figure 9.3: BDT score distributions after fitting in the SR (a) and QCDCR (b).

The over-constraint behavior, visible in the BDT plot uncertainties, is caused by an uncertainty which can be identified by a test of un-correlation between the SR an QCD-CR of three major theoretical uncertainties:

- renormalization and factorization scales uncertainty of the quark-induced QCD background prediction,
- the same uncertainty of the EW signal prediction,
- and the alternative generator uncertainty of the EW signal prediction.

The effect of the changes on the expected significance, t_μ statistic, and pull pots are negligible with exception of the QCD scales uncertainty un-correlation in the QCD background prediction. The expected significance was decreased from 3.9 to 3.7σ and uncertainty pull plots become slightly constrained, see Figure 9.5. The un-correlation also slightly decreased the test statistic curve, shown in Figure 9.4.

An obvious explanation is that the QCD-CR data sample has small statistic uncertainty (sufficient statistic population), however the prediction has a large, or over-conservative, QCD scale uncertainty, see Figure 9.1b. So the fit is constrained by data from the QCD-CR. This is also in accordance with a ranking plot, showing pre-fit and post-fit impact of input nuisance parameters to the parameter of interest, in Figure 9.8. The first two parameters are connected with the discussed QCD scale uncertainty and μ_{QCD} fitting the QCD-CR (by default set to one according to (9.1)). The next three parameters are connected with the jet energy scale and electron isolation uncertainties.

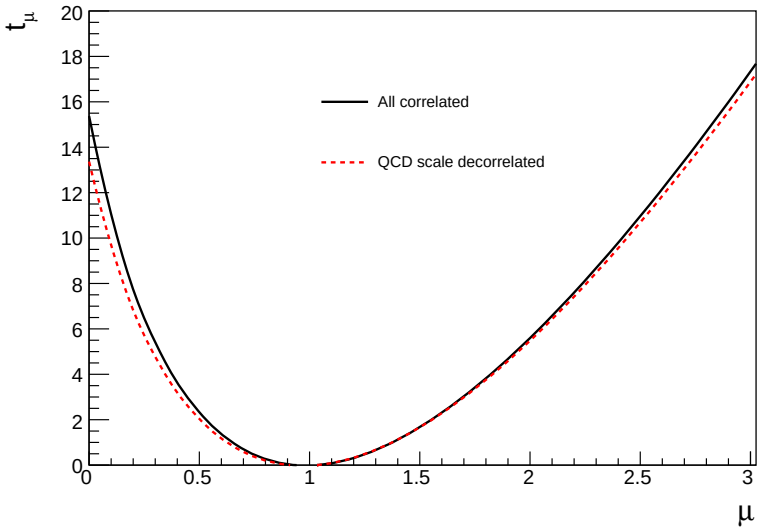
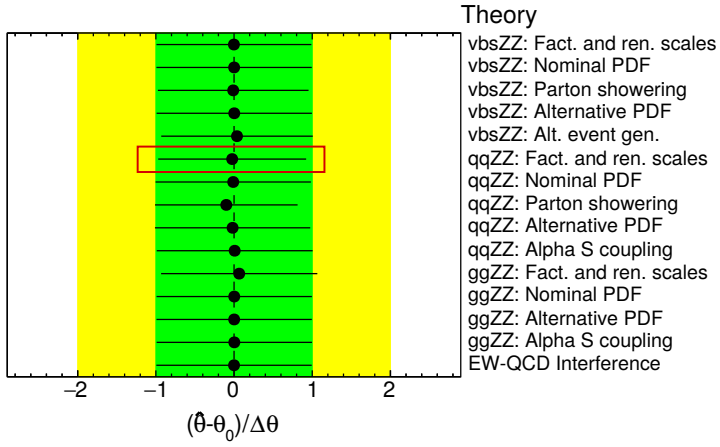


Figure 9.4: Graph of test statistic t_μ affected by un-correlation of QCD factorization and renormalization scale uncertainty between the SR and QCD-CR.

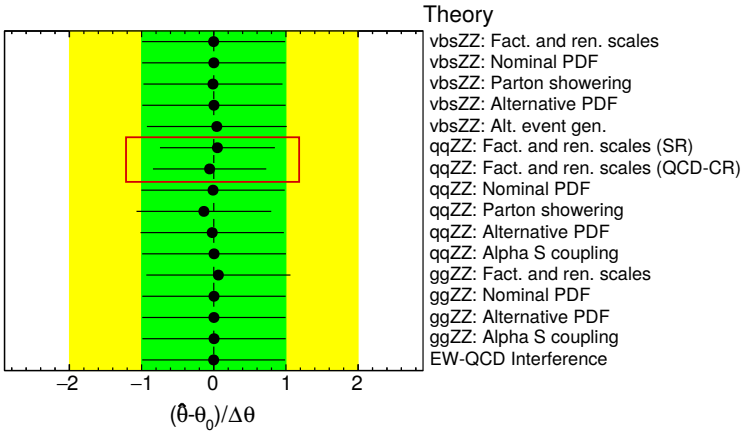
The expected significance is not very much affected neither by the normalization nor the un-correlation change. Therefore, the changes are kept and the signal strength extraction is prepared for the unblinding with promising 3.7σ of expected significance.

9.5 Fitting Results

For the final fitting, the configuration with the dropped normalization of theory uncertainties of the EW signal prediction and un-correlated QCD scales uncertainty between the SR and QCD-CR of the QCD background prediction was used.



(a)



(b)

Figure 9.5: Theory uncertainty pull plots after fitting. Plots with correlated (a) and un-corrlated (b) QCD factorization and renormalization scale uncertainty (highlighted by a rectangle) between the SR and QCD-CR.

The yields of both regions after fitting are listed in Table 9.3. The prediction is in a better agreement with the QCD-CR event yield prediction, when compared to the event yield before fitting in Table 9.1. The effect of systematic uncertainty on the t_μ statistic is displayed in Figure 9.6, where a drop in the expected significance is clearly visible. The BDT distributions suffer from an over-constrain in the case of the QCD background, which is caused by the real data in the QCD-CR. Hence only the lower BDT bins of the SR are affected. The higher BDT bins with presence of majority of the EW signal are not affected, shown in Figure 9.7. The observed significance of exclusion of the background-only hypothesis was enumerated to

$5.8\,\sigma$ against expected significance of $3.7\,\sigma$.

A signal strength μ_{EW} of the observed data is calculated together with a control-signal strength μ_{QCD} considering contributions from both quark- and gluon-induced QCD background instead. The strengths are calculated also in the blind regime nearing value of one because of using the Asimov data, although not exactly because the QCD-CR still contains the observed data. An important control is the expected strength in the SR, which should be almost exactly at the value of one. All calculated strengths are listed in Table 9.2, for the case including all uncertainties as well as using the statistical uncertainly only.

	Signal Strength			
	All uncertainties		Statistical only	
	μ_{EW}	μ_{QCD}	μ_{EW}	μ_{QCD}
Expected	0.98 ± 0.36	0.90 ± 0.24	0.99 ± 0.35	0.87 ± 0.08
Observed	1.78 ± 0.46	0.88 ± 0.20	1.84 ± 0.44	0.88 ± 0.08

Table 9.2: Expected and observed signal strengths of the EW signal and QCD background (quark- and gluon-induced together). Statistical and systematic uncertainties are included.

MC sample	Event yield	
	SR	QCD-CR
EW $ZZjj$	32.98 ± 10.39	5.87 ± 1.94
QCD $ZZjj$ (Quark-induced)	58.10 ± 9.75	98.30 ± 11.62
QCD $ZZjj$ (Gluon-induced)	9.96 ± 3.57	13.20 ± 4.81
$t\bar{t}$Z	3.86 ± 0.12	8.21 ± 0.20
Fakes Background	1.97 ± 1.50	4.31 ± 2.49
Tri-boson	0.61 ± 0.02	0.97 ± 0.03
Total MC	107.48 ± 11.96	130.87 ± 11.32
Real Data	107	129.00

Table 9.3: The SR and QCD-CR predicted and observed event yields after fitting. The composition is split by individual processes. The values are scaled to $139\,\text{fb}^{-1}$ including statistical and systematic uncertainties in simulated data and statistical uncertainty in the real data.

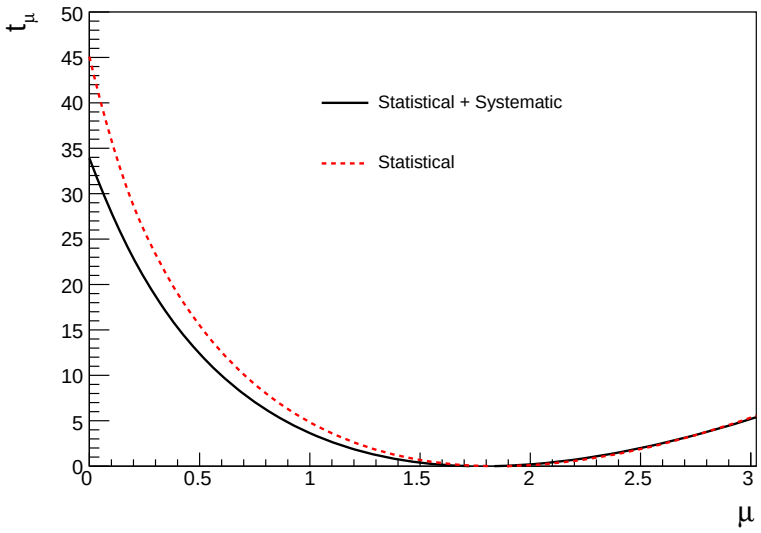


Figure 9.6: The test statistic t_μ affected only by the statistical uncertainty and by both statistical and systematic uncertainty.

9.6 EW Cross Section Extraction

The EW cross-section can be extracted using the signal strength which can be expressed as a ratio of the observed event yield over the MC predicted event yield. The ratio can be further identified with the measured fiducial cross-section over the predicted cross-section as in equation

$$\mu_{ZZjj\ \mathrm{EW}} = \frac{N_{\mathrm{data}}^{\mathrm{signal}}}{N_{\mathrm{MC}}^{\mathrm{signal}}} = \frac{\sigma_{ZZjj\ \mathrm{EW}}^{\mathrm{fid}}}{\sigma_{ZZjj\ \mathrm{EW}}^{\mathrm{pred}}}.$$

The predicted cross-section ($\sigma_{ZZjj\ \mathrm{EW}}^{\mathrm{pred}}$) is extracted directly from the fiducial phase space to be

$$\sigma_{ZZjj\ \mathrm{EW}}^{\mathrm{pred}} = 0.20 \pm 0.02\,\mathrm{fb}.$$

When multiplied by the observed signal strength including all uncertainties, shown in Table 9.2, the fiducial EW $ZZjj$ production cross-section is obtained to be

$$\sigma_{ZZjj\ \mathrm{EW}}^{\mathrm{fid}} = 0.36^{+0.09}_{-0.09}(\mathrm{stat.})^{+0.05}_{-0.05}(\mathrm{syst.})\,\mathrm{fb}.$$

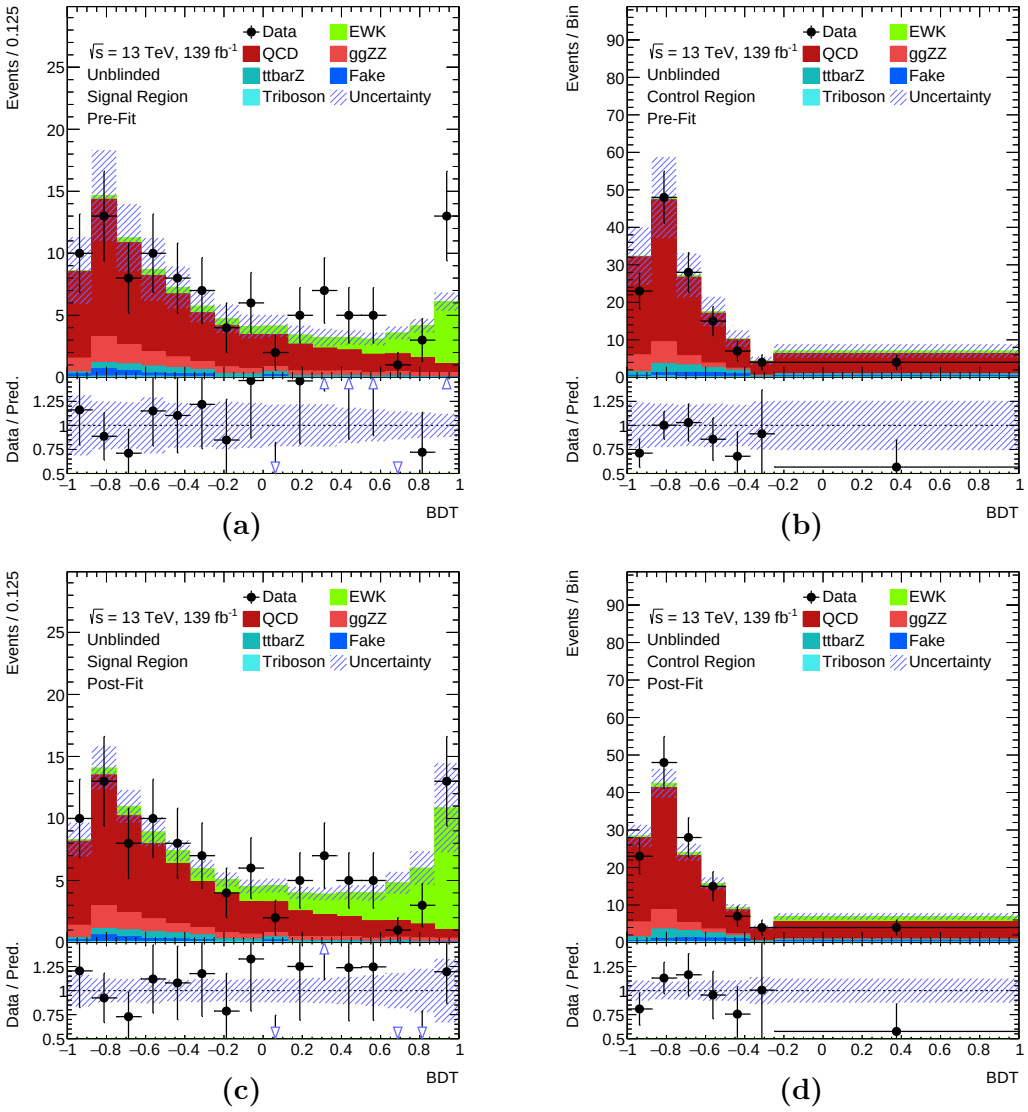


Figure 9.7: Unblinded BDT distributions of the SR (a) and QCD-CR (b) before fitting and after fitting (c) and (d).

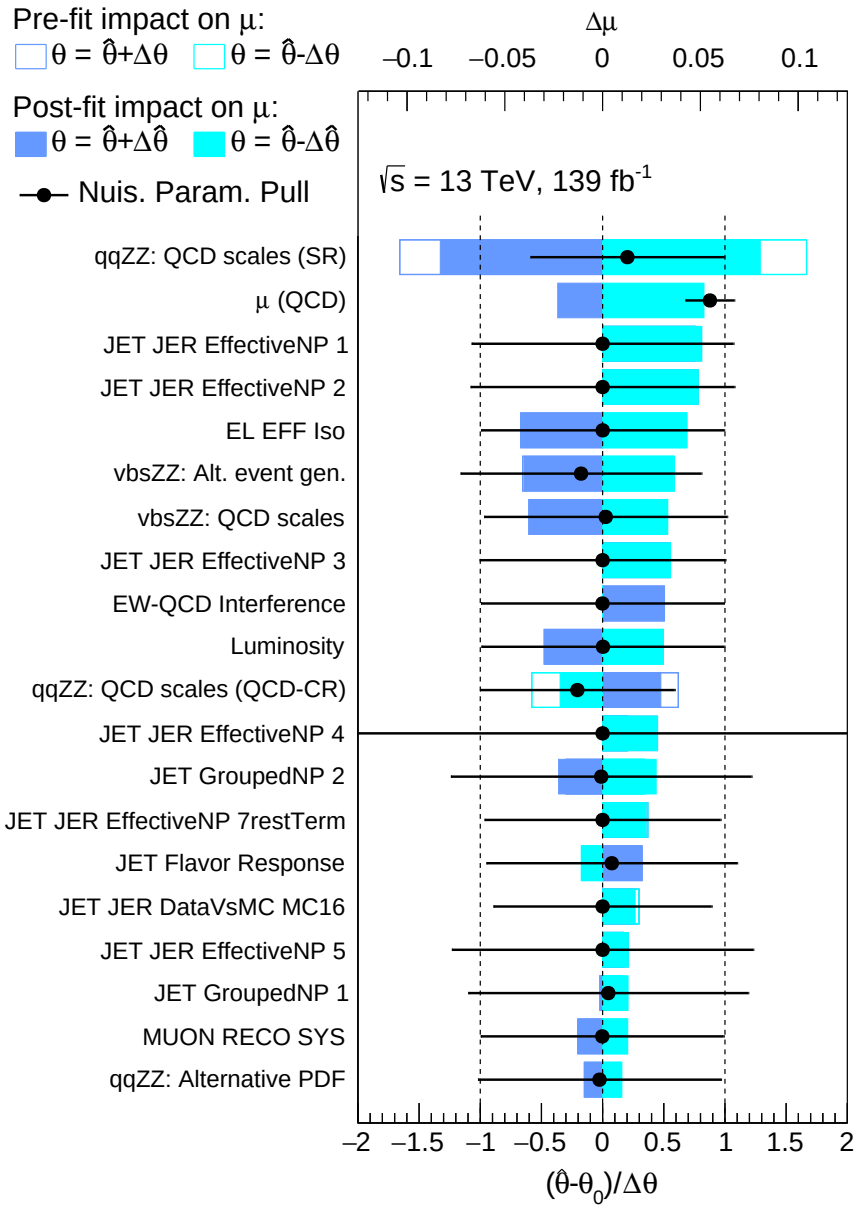


Figure 9.8: Ranking of impact of the nuisance parameters on the parameter of interest μ_{EW}. The empty boxes represent the impact before fitting, while the full boxes after fitting.

Chapter 10

Inclusive Production Measurement

The primary result of this analysis is accompanied with a cross-section measurement of inclusive $ZZjj$ production. The cross-section is calculated in the fiducial phase space defined in Table 6.1, using corresponding definition of the SR in Table 6.3.

10.1 Correction Factor

The measurement is corrected for the detector imperfections estimated by a fraction of the predicted event yield in the SR (N_{rec}) and truth event yield in the fiducial phase space (N_{fid}) in equation

$$C_{\text{ZZjj}} = \frac{N_{\text{rec}}}{N_{\text{fid}}}. \tag{10.1}$$

The factor was estimated for each inclusive production sample separately, as well as in combination, see Table 10.1.

Uncertainty	EW production	QCD production		Inclusive production
		Quark-induced	Gluon-induced	
C_{ZZjj}	$0.64^{+0.07}_{-0.08}$	$0.70^{+0.10}_{-0.07}$	$0.76^{+0.11}_{-0.09}$	$0.69^{+0.07}_{-0.05}$
Proportion	0.21	0.67	0.11	1.00

Table 10.1: Correction factor broken down to the individual processes of inclusive $ZZjj$ production. The result is combined in the last column using the proportions of individual processes.

The input yields of Equation (10.1) with an uncertainty variation applied are used to obtain a varied factor, which deviation from the nominal C_{ZZjj} is employed

as the uncertainty. The experimental uncertainties dominates over the theoretical, since the latter are largely correlated for the fiducial phase space and reconstructed SR, while the former affects only the SR, as can be seen in Table 10.2.

Uncertainty	EW production	QCD production		Inclusive production
		Quark-induced	Gluon-induced	
Statistical	0.0	0.1	0.1	± 0.1
Experimental	3.0	$^{+14.4}_{-10.0}$	$^{+15.0}_{-12.4}$	$^{+9.9}_{-6.9}$
Theoretical	0.1	$^{+2.7}_{-2.9}$	0.4	$^{+1.8}_{-2.0}$
Total	3.0	$^{+14.7}_{-10.4}$	$^{+15.0}_{-12.5}$	$^{+10.1}_{-7.2}$

Table 10.2: Summary of the correction factor uncertainties. The numbers are shown relative to the resulting factor in %.

The calculation of factor is tested for sensitivity on event migration effect, by a counting of the number of events that passed fiducial but failed reconstruction criteria and vice versa. Amount of such events was five orders of magnitude smaller than the case when an event passed both.

10.2 Cross Section

In comparison to the EW cross-section, the inclusive case has much smaller background since the QCD productions are considered as signal. So the task is diametrically opposite, the signal is overwhelming the signal. The cross-section is calculated straightforward from the equation

$$\sigma_{\text{fid}} = \frac{N_{\text{obs}} - N_{\text{bkg}}}{C_{\text{ZZjj}}\,\mathcal{L}},$$

where N_{obs} and N_{bkg} denote the numbers of observed events and predicted background events in the SR. The background event yield prediction is 6.91, obtained from Table 7.9 by exclusion of the QCD background. Only the prompt and fakes backgrounds are considered, estimated by MC modeling and data-driven methods, for details see Chapter 7. Details of modeling and accuracy of the background samples can be found in Section 4.4.2. A number of observed events is counted to be 107, taken from Table 9.3. The equation denominator consists of integrated luminosity equal to $138.97\,\text{fb}^{-1}$ and a correction factor C_{ZZjj}.

The measurement comprises all the uncertainties listed in Chapter 8, which are propagated to the result through variables N_{bkg} and C_{ZZjj}. The systematic

uncertainty is calculated according to equation

$$\Delta\sigma_{\text{fid}} = \frac{1}{C_{\text{ZZjj}}\,\mathcal{L}}\sqrt{{\Delta N^{i}_{\text{bkg}}}^{2} + \left(\frac{N_{\text{obs}} - N_{\text{bkg}}}{C_{\text{ZZjj}}}\right)^{2}{\Delta C^{i}_{\text{ZZjj}}}^{2} + \left(\frac{N_{\text{obs}} - N_{\text{bkg}}}{\mathcal{L}}\right)^{2}\Delta\mathcal{L}^{2}},$$

where the $\Delta N^{i}_{\text{bkg}}$ and $\Delta C^{i}_{\text{ZZjj}}$ represent the vectors of uncertainties affecting the N_{bkg} and C_{ZZjj} variables, composed of all uncertainties with exception of EW production generator and parton showering uncertainty. The EW production processes are minor in the inclusive production.

10.3 Result

The fiducial cross section was calculated using $139\,\text{fb}^{-1}$ of data at 13 TeV of center-of-mass energy, resulting in

$$\sigma^{\text{fid.}}_{ZZjj} = 1.05^{+0.12}_{-0.12}(\text{stat.})^{+0.11}_{-0.08}(\text{exp.})^{+0.02}_{-0.02}(\text{theo.})\,\text{fb}.$$

The systematic uncertainty is divided to experimental and theoretical component to show the experimental side dominates. The theoretical uncertainty is much smaller, since the MC modeled background is small and the corresponding uncertainties are correlated for N_{bkg} and C_{ZZjj}. The predicted cross-section is determined to be $0.93^{+0.22}_{-0.15}\,\text{fb}$.

As a cross check, the correction factor is calculated also for a ZZ production phase space, the SR without VBS di-jet selection stage, as $0.61^{+0.01}_{-0.01}$. The original factor calculated for the inclusive ZZ production cross-section measurement was determined as $0.60^{+0.02}_{-0.02}$ [20].

Chapter 11

Discussion

The results of presented analysis are compared to the published ATLAS and CMS collaboration articles [1, 2]. Both fiducial EW and inclusive (EW+QCD) $ZZjj$ production cross-sections as well as the significance of rejection of the QCD-only hypothesis in the $ZZjj$ production are subject of the comparison. Both observed and expected values are shown to assess a possible excess with respect to the SM. This analysis and the ATLAS collaboration use the same theoretical predictions generated at LO for EW production in MadGraph and at NLO for QCD productions in Sherpa. For details on the MC samples consult Section 5.3.2. The CMS collaboration has all predictions generated in MadGraph with the same precision corresponding to each production.

The inclusive fiducial cross-sections are shown in Table 11.1. The CMS utilizes more tight VBS selection from the m_{jj} and $\Delta\eta_{jj}$ perspective, while leaving the requirement on jets being from the different sides of detector. Moreover, the CMS uses an additional cut on a four-lepton mass over 180 GeV. All results are comparable in terms of the uncertainty, where the statistical component of uncertainty is dominant. Both the presented and ATLAS analyses experience slightly underestimated theoretical prediction, while the CMS analysis overestimated. The cross-section measured by ATLAS collaboration is approximatelly 20 % higher then the result of this work, since using a fiducial phase space adapted for a combined measurement with the two charged leptons and two neutrinos channel, shortly two-lepton channel. The phase space has higher acceptance of the Z bosons using a wider window for their invariant masses.

The significances of rejection of the QCD-only hypothesis in the $ZZjj$ production, as a result of statistical fit (described in Chapter 9), are shown in Table 11.2. The presented and ATLAS analyses exceed 5-σ observation of EW production, while the CMS reached 4-σ [1, 2]. The CMS result was acquired in a loosened VBS selection with $m_{jj} < 100$ GeV, requiring four-lepton mass over 180 GeV. All analyses experience underestimated theoretical prediction.

	Measured	Predicted	Note
This Work	$1.05^{+0.11}_{-0.11}(\text{stat.})^{+0.11}_{-0.08}(\text{syst.})$	$0.93^{+0.22}_{-0.15}$	$m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
ATLAS	$1.27^{+0.12}_{-0.12}(\text{stat.})^{+0.08}_{-0.08}(\text{syst.})$	$1.14^{+0.20}_{-0.20}$	$m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
CMS	$1.00^{+0.12}_{-0.11}(\text{stat.})^{+0.07}_{-0.07}(\text{syst.})$	$1.21^{+0.09}_{-0.09}$	$m_{jj} > 400\,\text{GeV}$, $\Delta\eta_{jj} > 2.4$, and $m_{4l} > 180\,\text{GeV}$

Table 11.1: Fiducial inclusive $ZZjj$ production cross-sections measured in this analysis and by the ATLAS and CMS collaborations [1, 2]. In the last column a significant differences in the fiducial phase space definitions of individual measurements are noted.

	Measured	Predicted	Note
This Work	$5.8\,\sigma$	$3.7\,\sigma$	$m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
ATLAS	$5.5\,\sigma$	$3.9\,\sigma$	$m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
CMS	$4.0\,\sigma$	$3.5\,\sigma$	$m_{jj} > 100\,\text{GeV}$, and $m_{4l} > 180\,\text{GeV}$

Table 11.2: Significances of rejection of the QCD-only hypothesis in $ZZjj$ production extracted in this analysis and by the ATLAS and CMS collaborations [1, 2]. In the last column a significant differences in the fiducial phase space definitions of individual measurements are noted.

Using the statistical fit, EW signal strength was derived and utilized for fiducial EW production cross-section extraction. The resulting cross-sections are listed in Table 11.3, unfortunately not directly comparable, since the ATLAS carried out the measurement as a combination of four-lepton and two-lepton channels ($\ell\ell\ell\ell$ and $\nu\nu\ell\ell$). Nonetheless, using the prediction from the thesis, one can estimate the ATLAS single channel ($\ell\ell\ell\ell$) result as $0.30 \pm 0.09\,\text{fb}$, using the signal strength 1.5 ± 0.4 [1].

	Measured [fb]	Predicted [fb]	Note
This Work	$0.36^{+0.09}_{-0.09}(\text{stat.})^{+0.05}_{-0.05}(\text{syst.})$	0.2 ± 0.02	$\ell\ell\ell\ell$, $m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
ATLAS	$0.82^{+0.18}_{-0.18}(\text{stat.})^{+0.10}_{-0.10}(\text{syst.})$	0.61 ± 0.03	$\ell\ell\ell\ell + \nu\nu\ell\ell$, $m_{jj} > 300\,\text{GeV}$, $\Delta\eta_{jj} > 2.0$, and $y_{j_1} \times y_{j_2} < 0$
CMS	$0.33^{+0.11}_{-0.10}(\text{stat.})^{+0.04}_{-0.03}(\text{syst.})$	0.28 ± 0.02	$\ell\ell\ell\ell$, $m_{jj} > 100\,\text{GeV}$, and $m_{4l} > 180\,\text{GeV}$
CMS	$0.18^{+0.07}_{-0.06}(\text{stat.})^{+0.02}_{-0.01}(\text{syst.})$	0.20 ± 0.01	$\ell\ell\ell\ell$, $m_{jj} > 400\,\text{GeV}$, $\Delta\eta_{jj} > 2.4$, and $m_{4l} > 180\,\text{GeV}$

Table 11.3: Fiducial EW $ZZjj$ production cross-section measured in this analysis and by the ATLAS and CMS collaborations [1, 2]. In the last column a corresponding final-state channel and significant differences in the fiducial phase space definitions of individual measurements are noted.

To summarize, no analysis observed a deviation from the SM, all are dominated by statistical component of uncertainty, and finally, all three for the first time

observed the existence of $ZZjj$ EW production in proton-proton collisions. The EW cross-section results unambiguously identified the four-lepton channel as the most rare in the VBS channels.

Chapter 12

Conclusions

The presented work, together with the associated article [1], is the first complete study of the EW production of two Z bosons associated with two jets in a VBS topology in the four-lepton channel. The precise knowledge of this process is crucial for the completion and understanding of the SM, as well as for checking the validity of some of the BSM theories.

The importance of these processes for the SM stems from the fact that the BEH mechanism, proposed in 1964 [5, 6] and experimentally verified in 2012 [3, 4], predicts precise cancellation of amplitudes induced by the Higgs boson and restores unitarity of the VBS processes which would otherwise be broken at the energies of LHC.

The importance of the knowledge of VBS cross-sections for the BSM searches comes from the fact that many of these theories predict deviations from the SM at high energies [12–14]. Precise measurements of processes like four-lepton $ZZjj$ EW production either help to verify or put tension onto such models and guide our further steps in the searches for new physics.

The work that is presented in this thesis probed the rarest processes of VBS, the one where two Z-bosons decay into two charged lepton pairs. Measurement of this process has proven to be troublesome due to its small cross section, however, the four-lepton production in the final-state offers a very clean signal that can be extracted from the background.

Furthermore, VBS processes in proton-proton collisions are usually accompanied by a production of two jets resulting from the initial partons that radiate the heavy vector bosons. These partons are kinematically shifted by the emission, get ejected from the incoming proton and appear in the final-state as hadronic jets. These jets are proven valuable for tagging of the VBS processes.

On top of the ZZ selection based on lepton pairing, the VBS tagging di-jet requirements were employed. The di-jet must have invariant mass over 300 GeV, constituents reconstructed in opposite sides of the detector, and being well sep-

arated with pseudorapidity difference of over two. For further VBS tagging, the boson centrality with respect to the tagging jets was identified as promising. Based on the selection criteria a pre-VBS signal region was established, in which a fiducial inclusive production cross-section was measured to be $1.05^{+0.11}_{-0.11}(\text{stat.})^{+0.11}_{-0.08}(\text{syst.})$ fb. The result is compatible with the SM and in a good agreement with both ATLAS and CMS collaboration measurements.

Unlike the other VBS searches and analyses at ATLAS, the author did not use the TMVA in the framework of ROOT for the machine-learning applications. Instead, the Scikit-learn was utilized which is a part of the Python machine learning libraries, and was proven to be a good alternative to the TMVA.

A possibility of no $ZZjj$ EW production in the four-lepton channel was studied since it could be a possible sign of BSM physics. The analysis found that this hypothesis is not valid with 5.8 σ and that the result is consistent with the SM prediction.Together, a fiducial EW production cross-section was measured to be $0.36^{+0.09}_{-0.09}(\text{stat.})^{+0.05}_{-0.05}(\text{syst.})$ fb. The result represents the first measurement of a single four-lepton channel with a five-σ significance, since the ATLAS collaboration published only an inclusive measurement with the two-lepton channel and CMS reached only four-σ.

The work connected with this thesis resulted in two articles. The measurement of inclusive ZZ production in 4-lepton channel [20] was published in Physical Review D and served as a starting point for this work. The second article [1] on EWK $ZZjj$ production was submitted to Nature Physics and includes also the two-lepton channel in addition to the work presented in this thesis.

Studies like the one presented in this thesis, help to solidify our understanding of the SM and rigorously check the extent to which this theory is valid. There are still many missing puzzle pieces that pave the way to a complete and unified theory of physics. However, the work that was presented in this thesis, elevated a part of the unknown features of the ZZ VBS processes and added them to the list of successes of the SM.

The measurements of EW production of all $VVjj$ channels represents an opening gate towards the VBS scattering. The $ZZjj$ channel is dominated by statistical uncertainty having up to twenty events in unprecedented amount of $139\,\text{fb}^{-1}$ of data. However, the expected $300\,\text{fb}^{-1}$ of additional data from the coming Run 3 will make the results more accurate and enable a serious differential cross-section measurement. Even polarization studies will become feasible.

Chapter 13

Appendix

Object	Trigger name	Data-taking periods
Electron		
	e26_lhtight_nod0_ivarloose	B-End
	e60_lhmedium_nod0	B-End
	e140_lhloose_nod0	B-End
	e300_etcut	B-End
Multi-electron		
	2e17_lhvloose_nod0_L12EM15VHI	B-End
	2e24_lhvloose_nod0	B-End
	e24_lhvloose_nod0_2e12_lhvloose_nod0_L1EM20VH_3EM341649	B-End
Muon		
	mu26_ivarmedium	B-End
	mu50	B-End
	mu60_0eta105_msonly	B-End
Multi-muon		
	2mu14	B-End
	mu22_mu8noL1	B-End
	mu22_mu8noL1_calotag_0eta010	B-End
	mu20_2mu4noL1	B-End
	3mu4	B-End
	3mu6	B-End
	3mu6_msonly	B-End
	4mu4	B-End
Mix		
	e17_lhloose_nod0_mu14	B-End
	e26_lhmedium_nod0_mu8noL1	B-End
	e7_lhmedium_nod0_mu24	B-End
	e12_lhloose_nod0_2mu10	B-End
	2e12_lhloose_nod0_mu10	B-End

Table 13.1: Triggers utilized in the analysis for the data collected during year 2017. The triggers are applicable only to some periods of data-taking.

Object	Trigger name	Data-taking periods
Electron		
	e24_lhtight_nod0_ivarloose	A-D3
	e24_lhmedium_nod0_L1EM20VH	A-D3
	HLT_e24_lhmedium_nod0_L1EM20VH	A-D3
	e26_lhtight_nod0_ivarloose	D3-End
	e60_lhmedium_nod0	D3-End
	e60_medium	D3-End
	e140_lhloose_nod0	D3-End
	e300_etcut	D3-End
Multi-electron		
	2e15_lhvloose_nod0_L12EM13VH	A-D3
	e17_lhloose_nod0_2e9_lhloose_nod0	All
	2e17_lhvloose_nod0	D4-End
	e17_lhloose_nod0_2e10_lhloose_nod0_L1EM15VH_3EM8V1481	G-End
	e20_lhmedium_nod0_2g10_loose	G-End
	e20_lhmedium_nod0_2g10_loose_L1EM15VH_3EM8VH	G-End
Muon		
	mu24_iloose	A
	mu24_ivarloose	A
	mu40	A
	mu50	All
	mu24_ivarmedium	B-E
	mu24_imedium	B-E
	mu26_ivarmedium	D4-End
	mu26_imedium	D4-G1
Multi-muon		
	HLT_2mu10	A
	HLT_2mu10_nomucomb	A
	HLT_mu20_mu8noL1	A-E
	HLT_mu20_2mu4noL1	All
	HLT_3mu4	A
	HLT_3mu4	I4-End
	HLT_mu6_2mu4	A-D3
	HLT_mu20_nomucomb_mu6noL1_nscan03	A-D3
	HLT_mu11_nomucomb_2mu4noL1_nscan03_L1MU11_2MU6	A-D3
	HLT_mu20_msonly_mu10noL1_msonly_nscan05_noComb	A-D3
	HLT_mu11_nomucomb_2mu4noL1_nscan03_L1MU11_2MU6_bTau	A-D3
	HLT_mu11_nomucomb_mu6noL1_nscan03_L1MU11_2MU6_bTau	A-D3
	HLT_mu6_nomucomb_2mu4_nomucomb_bTau_L1MU6_3MU4	A-D3
	HLT_2mu6_nomucomb_mu4_nomucomb_bTau_L12MU6_3MU4	A-D3
	HLT_2mu14	B-End
	HLT_2mu14_nomucomb	B-End
	HLT_3mu6	B-End
	HLT_mu22_mu8noL1	D4-End
	HLT_3mu6_msonly	D4-End
Mix		
	HLT_e17_lhloose_nod0_mu14	All
	HLT_e24_lhmedium_nod0_L1EM20VHI_mu8noL1	A-D3
	HLT_e7_lhmedium_nod0_mu24	All
	HLT_e12_lhloose_nod0_2mu10	All
	HLT_2e12_lhloose_nod0_mu10	All
	HLT_e26_lhmedium_nod0_L1EM22VHI_mu8noL1	D4-End

Table 13.2: Triggers utilized in the analysis for the data collected during year 2016. The triggers are applicable only to some periods of data-taking.

Object	Trigger name	Data-taking periods
Electron		
	e26_lhtight_nod0_ivarloose	B-End
	e26_lhtight_nod0	B-End
	e60_lhmedium_nod0	B-End
	e140_lhloose_nod0	B-End
	e300_etcut	B-End
Multi-electron		
	2e17_lhvloose_nod0_L12EM15VHI	B-End
	2e24_lhvloose_nod0	B-End
	e24_lhvloose_nod0_2e12_lhvloose_nod0_L1EM20VH_3EM364485	B-End
Muon		
	mu26_ivarmedium	B-End
	mu50	B-End
	mu60_0eta105_msonly	B-End
Multi-muon		
	2mu14	B-End
	mu22_mu8noL1	B-End
	mu20_2mu4noL1	B-End
	3mu6	B-End
Mix		
	e17_lhloose_nod0_mu14	B-End
	e26_lhmedium_nod0_mu8noL1	B-End
	e7_lhmedium_nod0_mu24	B-End
	e12_lhloose_nod0_2mu10	B-End
	2e12_lhloose_nod0_mu10	B-End

Table 13.3: Triggers utilized in the analysis for the data collected during year 2018. The triggers are applicable only to some periods of data-taking.

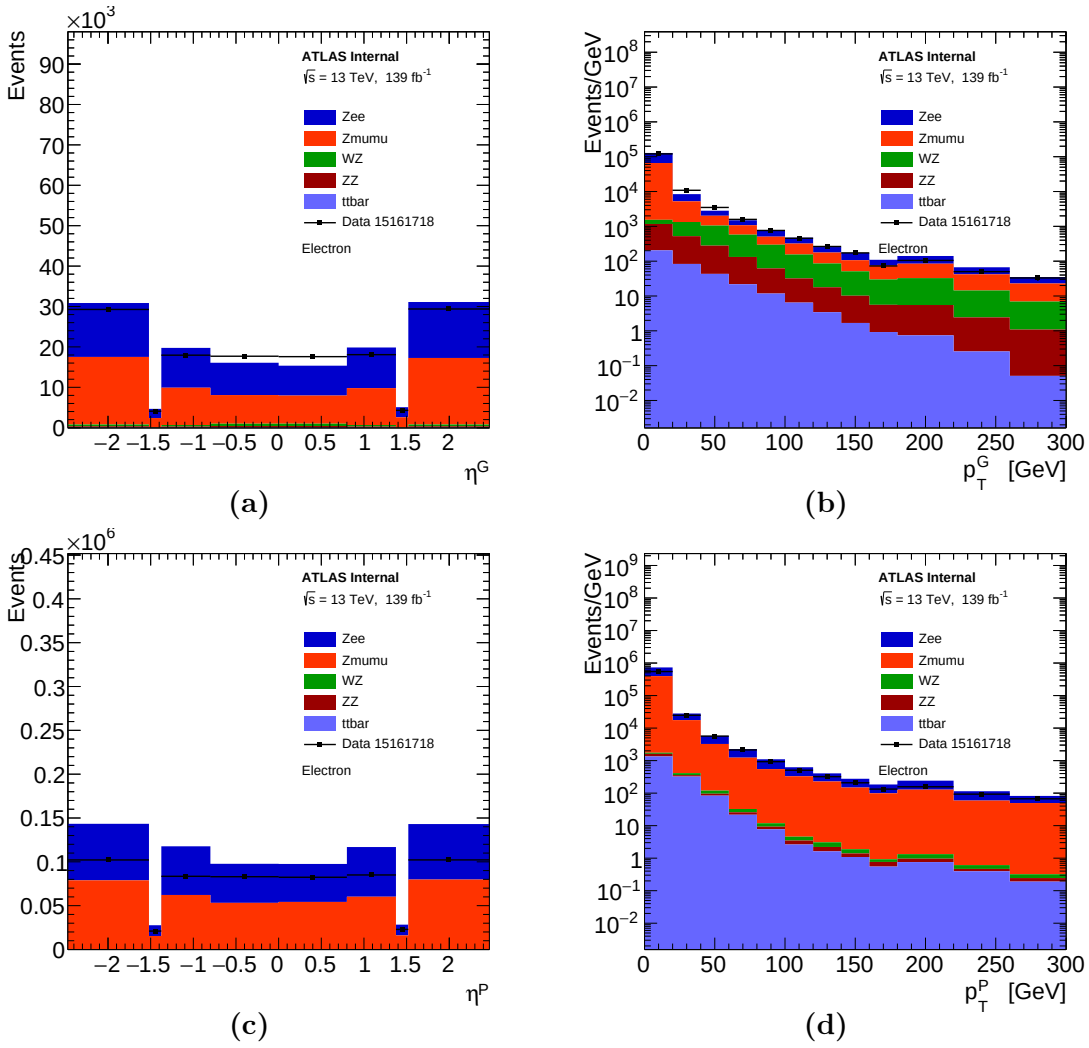


Figure 13.1: Fake factor of the Z + jets region: *additional* electron η and p_{T} distributions. Histograms (a) and (b) show the *good* electron while the (c) and (d) refer to the case of *poor* electron distributions [168].

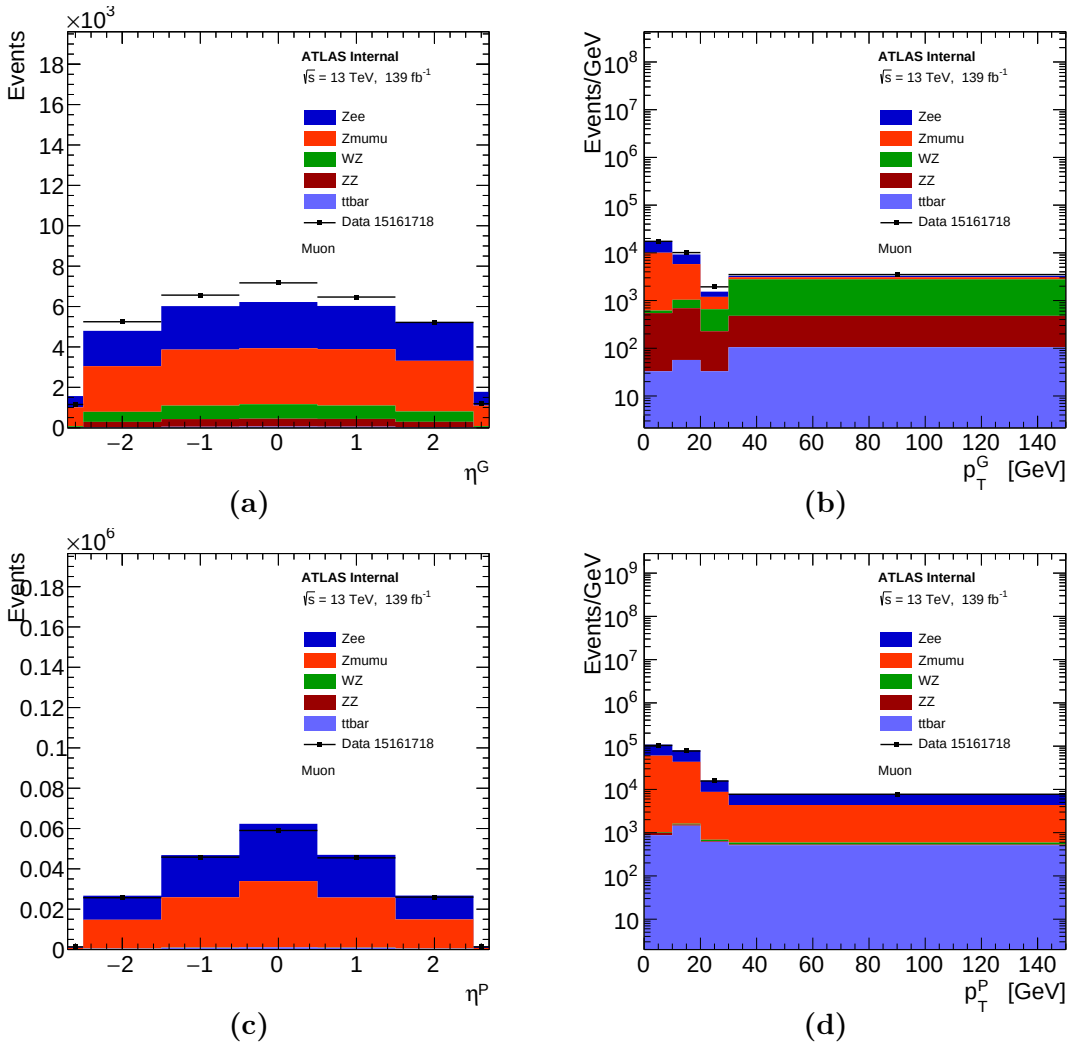


Figure 13.2: Fake factor of the $Z+$jets region: *additional* muon η and p_{T} distributions. Histograms (a) and (b) show the *good* muon while the (c) and (d) refer to the case of *poor* muon distributions [168].

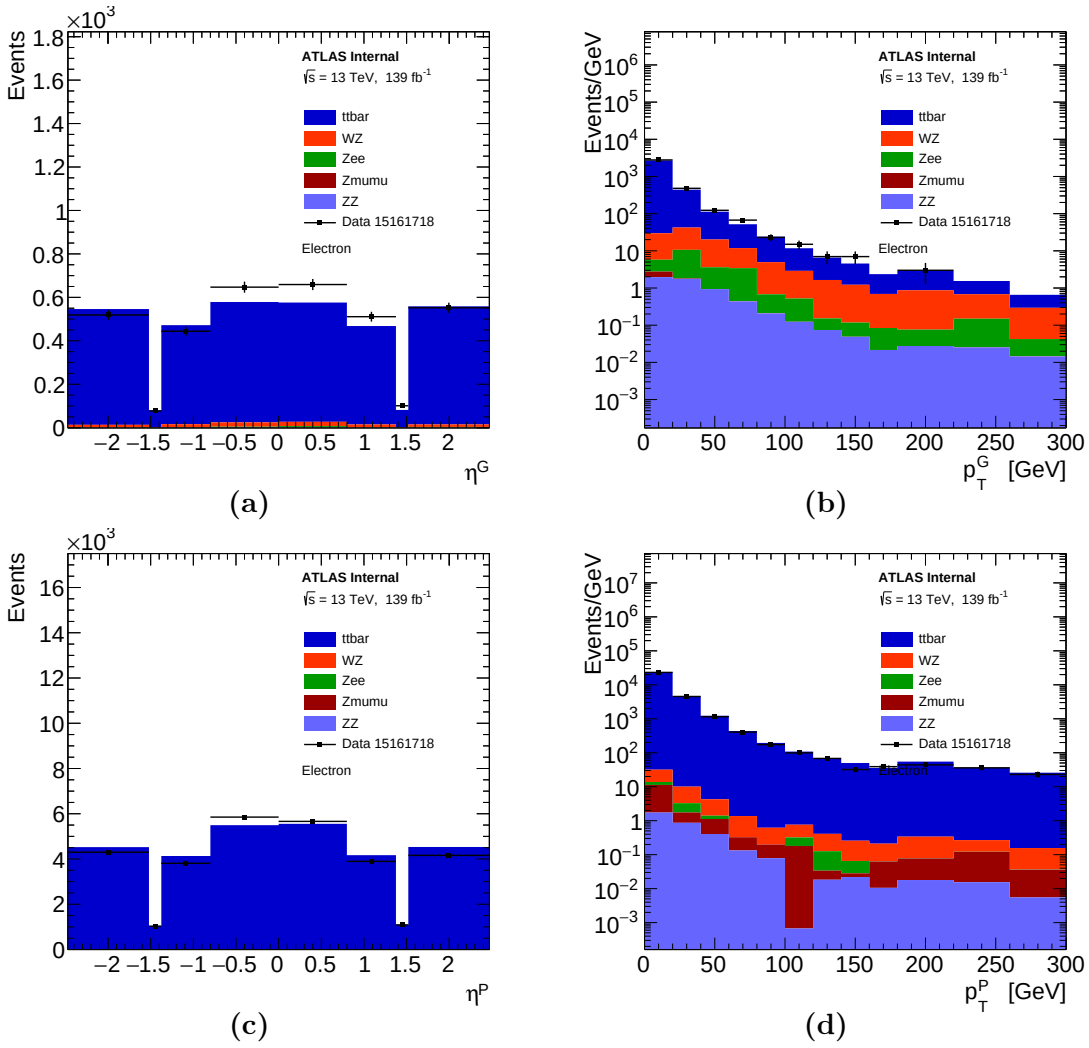


Figure 13.3: Fake factor of the $t\bar{t}$ region: *additional* electron η and p_{T} distribution. Histograms (a) and (b) show the *good* electron while the (c) and (d) refer to the case of *poor* electron distributions [168].

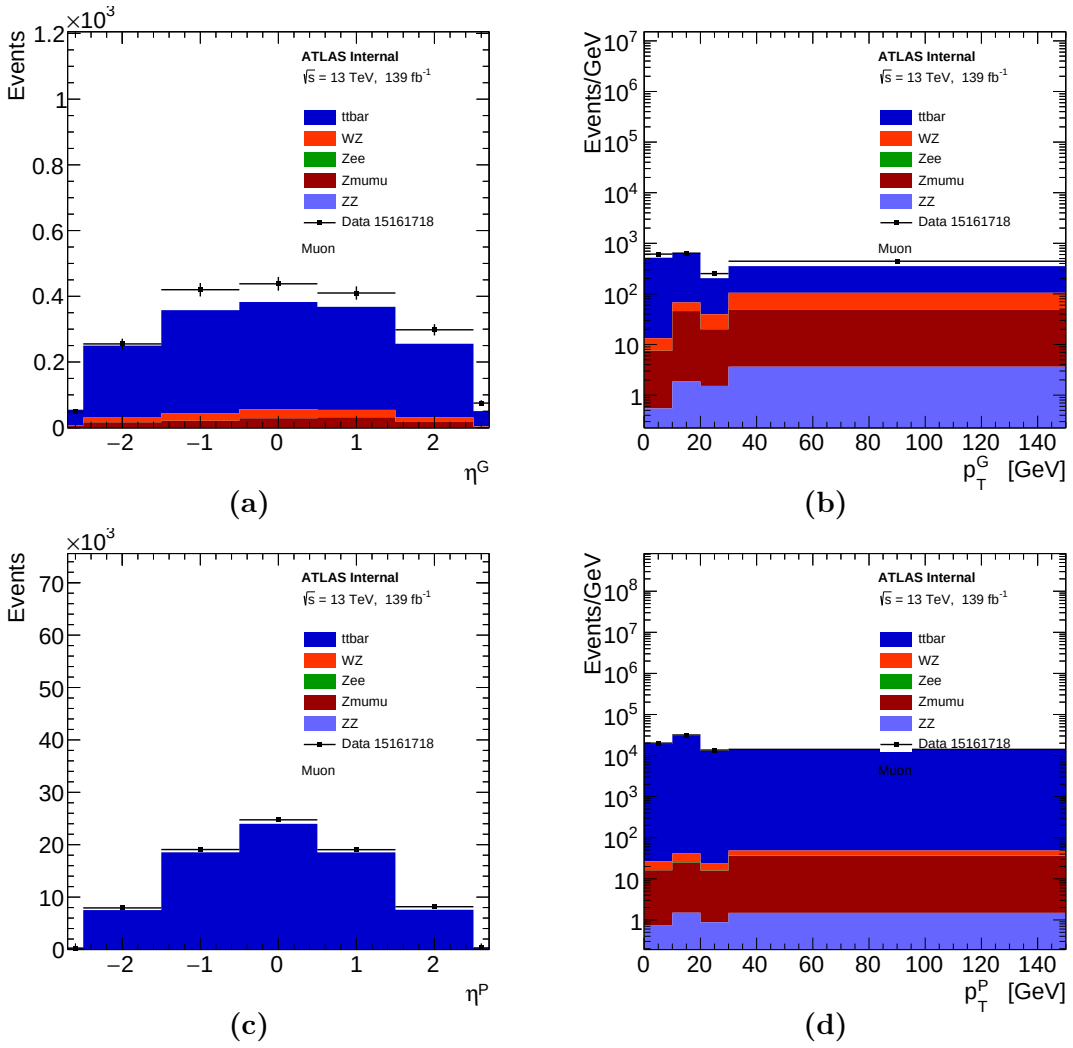


Figure 13.4: Fake factor of the $t\bar{t}$ region: *additional* muon η and p_{T} distributions. Histograms (a) and (b) show the *good* muon while the (c) and (d) refer to the case of *poor* muon distributions [168].

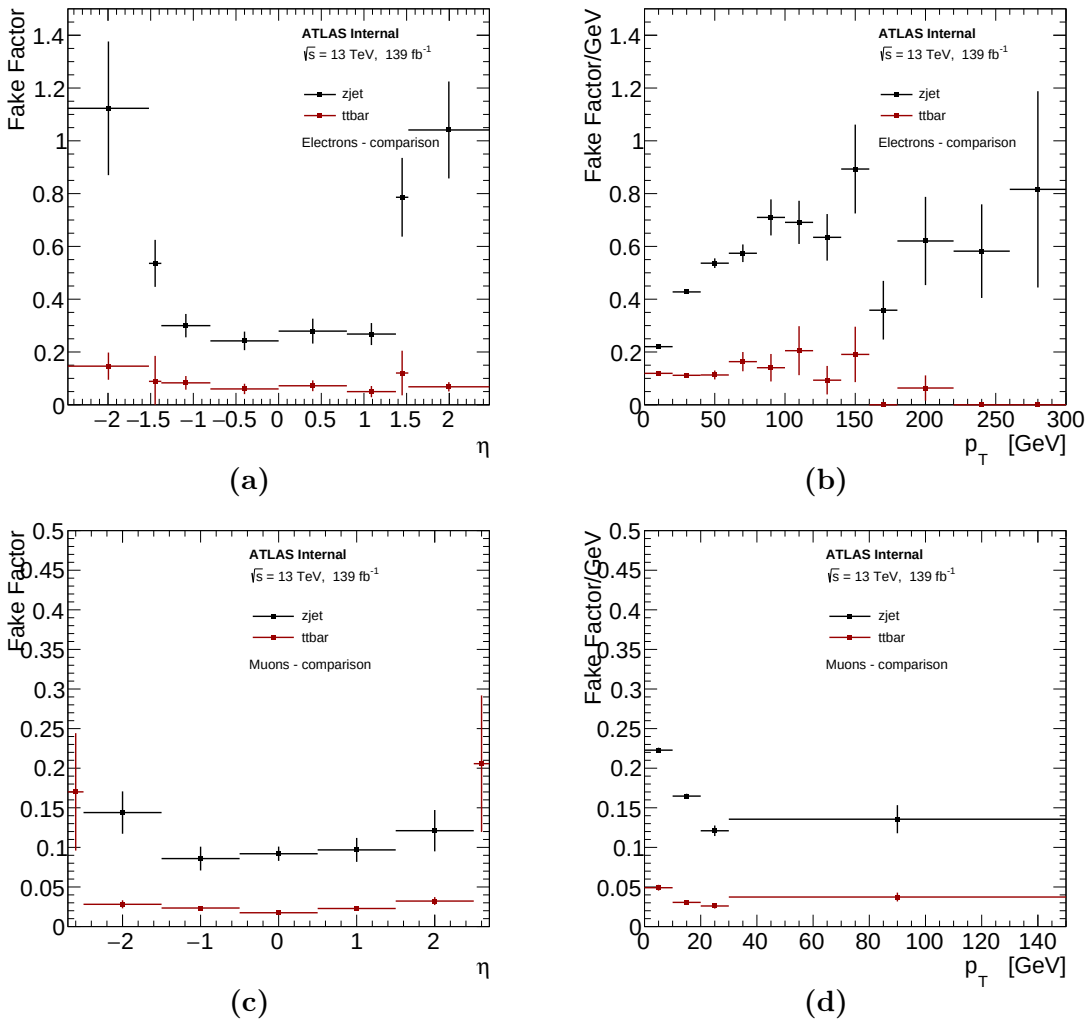


Figure 13.5: Comparison of the Z + jets and $t\bar{t}$ fake factors in η and p_{T} distributions [168].

Bibliography

[1] G. Aad et al., "Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC", *Submitted to Nature Physics* (2020), URL: https://arxiv.org/abs/2004.10612.

[2] A. M. Sirunyan et al., "Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* 812 (2021), p. 135992.

[3] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* 716 (2012), pp. 1–29.

[4] S. Chatrchyan et al., "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* 716 (2012), pp. 30–61.

[5] P. Higgs, "Broken symmetries, massless particles and gauge fields", *Physics Letters* 12 (1964), pp. 132 –133, URL: http://www.sciencedirect.com/science/article/pii/0031916364911369.

[6] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* 13 (1964), pp. 321–323, URL: http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[7] "The International Linear Collider Technical Design Report - Volume 2: Physics" (2013), ed. by H. Baer et al.

[8] "Physics and Detectors at CLIC: CLIC Conceptual Design Report" (2012), ed. by L. Linssen et al.

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

[10] "High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary Design Report" (2015), ed. by G Apollinari et al.

[11] B. W. Lee, C. Quigg, and H. B. Thacker, "The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass", *Phys. Rev. Lett.* 38 (1977), pp. 883–885.

[12] A. Djouadi, "Implications of the Higgs discovery for the MSSM", *Eur. Phys. J. C* 74 (2014), p. 2704.

[13] R. Contino et al., "On the effect of resonances in composite Higgs phenomenology", *JHEP* 10 (2011), p. 081.

[14] N. Arkani-Hamed et al., "The Littlest Higgs", *JHEP* 07 (2002), p. 034.

[15] M. Aaboud et al., "Observation of electroweak production of a same-sign W boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. Lett.* 123 (2019), p. 161801.

[16] A. M. Sirunyan et al., "Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* 120 (2018), p. 081801.

[17] M. Aaboud et al., "Observation of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* 793 (2019), pp. 469–492.

[18] A. M. Sirunyan et al., "Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* 809 (2020), p. 135710.

[19] G. Aad et al., "Search for the electroweak diboson production in association with a high-mass dijet system in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* 100 (2019), p. 032007.

[20] M. Aaboud et al., "$ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ cross-section measurements and search for anomalous triple gauge couplings in 13 TeV pp collisions with the ATLAS detector", *Phys. Rev. D* 97 (2018), p. 032005.

[21] W. N. Cottingham and D. A. Greenwood, *An introduction to the standard model of particle physics*, Cambridge University Press, 2007.

[22] J. Hořejší, *Fundamentals of electroweak theory*, Prague: The Karolinum Press, 2002, URL: https://cds.cern.ch/record/1477734.

[23] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* 22 (1961), pp. 579 –588, URL: http://www.sciencedirect.com/science/article/pii/0029558261904692.

[24] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266, URL: http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[25] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* C680519 (1968), pp. 367–377.

[26] LEP TGC Working Group (ALEPH, DELPHI, L3, OPAL), "A Combination of Preliminary Results on Gauge Boson Couplings Measured by the LEP Experiments", *LEPEWWG/TGC/2003-01(2003)* (2004), URL: http://lepewwg.web.cern.ch/LEPEWWG/lepww/tgc.

[27] P. Achard et al., "Study of the process at {LEP} and limits on triple neutral-gauge-boson couplings", *Physics Letters B* 597 (2004), pp. 119 – 130, URL: http://www.sciencedirect.com/science/article/pii/S0370269304010093.

[28] P. Achard et al., "Measurement of triple-gauge-boson couplings of the W boson at {LEP}", *Physics Letters B* 586 (2004), pp. 151 –166, URL: http://www.sciencedirect.com/science/article/pii/S0370269304003430.

[29] P. Achard et al., "Study of the $W^+W^-\gamma$ process and limits on anomalous quartic gauge boson couplings at {LEP}", *Physics Letters B* 527 (2002), pp. 29 –38, URL: http://www.sciencedirect.com/science/article/pii/S037026930201167X.

[30] G. Abbiendi et al., "Constraints on anomalous quartic gauge boson couplings from $\nu\bar{\nu}\gamma\gamma$ and $q\bar{q}\gamma\gamma$ events at CERN LEP2", *Phys. Rev. D* 70 (2004), p. 032005, URL: http://link.aps.org/doi/10.1103/PhysRevD.70.032005.

[31] H. Yukawa, "On the Interaction of Elementary Particles I", *Proc. Phys. Math. Soc. Jap.* 17 (1935), pp. 48–57.

[32] J. Horejsi, *Fundamentals of electroweak theory*, 2002.

[33] J. Reuter, W. Kilian, and M. Sekulla, "Simplified Models for New Physics in Vector Boson Scattering - Input for Snowmass 2013" (2013).

[34] G. Aad et al., "Evidence for Electroweak Production of $W^\pm W^\pm jj$ in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector", *Phys. Rev. Lett.* 113 (2014), p. 141803.

[35] V. Khachatryan et al., "Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets", *Phys. Rev. Lett.* 114 (2015), p. 051801.

[36] A. M. Sirunyan et al., "Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* 774 (2017), pp. 682–705.

[37] R. Rosenfeld, "Can gamma gamma —> Z(L) Z(L) serve as a probe of the electroweak symmetry breaking sector?", *Mod. Phys. Lett.* A9 (1994), pp. 735–748.

[38] R. L. Delgado, A. Dobado, and F. J. Llanes-Estrada, "One-loop W_LW_L and Z_LZ_L scattering from the electroweak Chiral Lagrangian with a light Higgs-like scalar", *JHEP* 02 (2014), p. 121.

[39] B. Jäger, A. Karlberg, and G. Zanderighi, "Electroweak ZZjj production in the Standard Model and beyond in the POWHEG-BOX V2", *Journal of High Energy Physics* 2014 (2014), pp. 1–19, URL: http://dx.doi.org/10.1007/JHEP03(2014)141.

[40] O. Penc, "Observation and measurements of vector-boson scattering with the ATLAS detector", *Phys. Scripta* 95 (2020), p. 084014.

[41] N. Aghanim et al., "Planck 2018 results. VI. Cosmological parameters", *Astron. Astrophys.* 641 (2020), A6.

[42] B. Holdom, "Accidental Higgs boson", *Phys. Rev. D* 90 (2014), p. 015004, URL: http://link.aps.org/doi/10.1103/PhysRevD.90.015004.

[43] M. Carena, L. Da Rold, and E. Pontón, "Minimal composite Higgs models at the LHC", *Journal of High Energy Physics* 2014 (2014), pp. 1–58, URL: http://dx.doi.org/10.1007/JHEP06(2014)159.

[44] V. Barger et al., "Complex Singlet Extension of the Standard Model", *Phys. Rev.* D79 (2009), p. 015018.

[45] A. Broggio et al., "Limiting two-Higgs-doublet models", *Journal of High Energy Physics* 2014 (2014), pp. 1–18, URL: http://dx.doi.org/10.1007/JHEP11(2014)058.

[46] A. Djouadi, "The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model", *Phys. Rept.* 459 (2008), pp. 1–241.

[47] C. Degrande, O. Eboli, B. Feigl, et al., "Monte Carlo tools for studies of non-standard electroweak gauge boson interactions in multi-boson processes: A Snowmass White Paper" (2013), URL: https://inspirehep.net/record/1256129/files/arXiv:1309.7890.pdf.

[48] W. Kilian, "Electroweak symmetry breaking: The bottom-up approach", *Springer Tracts Mod. Phys.* 198 (2003), pp. 1–113.

[49] A. Breskin and R. Voss, *The CERN Large Hadron Collider: Accelerator and Experiments*, Geneva: CERN, 2009, URL: https://cds.cern.ch/record/1244506.

[50] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* 3 (2008), S08003.

[51] G. Arnison et al., "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV", *Phys. Lett.* 122B (1983), [,611(1983)], pp. 103–116.

[52] G. Arnison et al., "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider", *Phys. Lett.* 126B (1983), pp. 398–410.

[53] J. R. J. Bennett et al., "Design Concept for a 100-GeV e+ e- Storage Ring (LEP)" (1977).

[54] "First W-pairs at LEP2. Premières paires de W au LEP2", *CERN Bulletin* (1996), pp. 1–3, URL: https://cds.cern.ch/record/1726532.

[55] M. Acciarri et al., "Study of neutral current four fermion and Z Z production in e+ e- collisions at S**(1/2) = 183-GeV", *Phys. Lett.* B450 (1999), pp. 281–293.

[56] "A Combination of Preliminary Results on Gauge Boson Couplings Measured by the LEP experiments" (2003).

[57] T. Group, "Design Report Tevatron 1 project" (1984).

[58] F. Abe et al., "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631.

[59] S. Abachi et al., "Observation of the top quark", *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637.

[60] V. M. Abazov et al., "Measurement of the WW production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. Lett.* 94 (2005), [Erratum: Phys. Rev. Lett.100,139901(2008)], p. 151801.

[61] T. Aaltonen et al., "First Measurement of ZZ Production in panti-p Collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev. Lett.* 100 (2008), p. 201801.

[62] S. P. Martin, "A Supersymmetry primer" (1997), [Adv. Ser. Direct. High Energy Phys.18,1(1998)], pp. 1–98.

[63] M. Neubert, "Introduction to B physics", *Proceedings, Summer School in Particle Physics: Trieste, Italy, June 21-July 9, 1999*, 2000, pp. 244–295.

[64] R. Aaij et al., "Observation of CP Violation in Charm Decays", *Phys. Rev. Lett.* 122 (2019), p. 211803.

[65] R. Aaij et al., "Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$", *Phys. Rev. Lett.* 122 (2019), p. 222001.

[66] F. Noferini, "ALICE Highlights", *MDPI Proc.* 13 (2019), p. 6.

[67] S. Maury, "The Antiproton Decelerator: AD", *Hyperfine Interact.* 109 (1997), pp. 43–52.

[68] M. Ahmadi et al., "Investigation of the fine structure of antihydrogen", *Nature* 578 (2020), pp. 375–380.

[69] N. Agafonova et al., "Observation of a first ν_τ candidate in the OPERA experiment in the CNGS beam", *Phys. Lett. B* 691 (2010), pp. 138–145.

[70] C. Bulletin, "The particle suppliers. Les fournisseurs de particules" (2010), p. 03, URL: https://cds.cern.ch/record/1255151.

[71] G. Arduini et al., *Alternative bunch filling schemes for the LHC*, tech. rep., 2007, URL: http://cds.cern.ch/record/1075485.

[72] R Bailey and P. Collier, *Standard Filling Schemes for Various LHC Operation Modes*, tech. rep., Geneva: CERN, 2003, URL: https://cds.cern.ch/record/691782.

[73] J.-L. Caron, "Computed magnetic flux in the LHC dipole.. Champ magnetique (calcule) dans le dipole LHC.", AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998, URL: https://cds.cern.ch/record/841503.

[74] J.-L. Caron, "Magnetic field induced by the LHC dipole's superconducting coils.. Champ magnetique cree par les bobines superconductrices des aimants dipolaires du LHC.", AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998, URL: https://cds.cern.ch/record/841511.

[75] F. Sonnemann, "Resistive transition and protection of LHC superconducting cables and magnets", PhD thesis, RWTH Aachen U., 2001.

[76] J.-L. Caron, "Beam pipe.. Chambre a vide.", AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1999, URL: https://cds.cern.ch/record/842472.

[77] B. Holzer, "Introduction to Transverse Beam Dynamics", *CERN Accelerator School: Course on Superconductivity for Accelerators*, 2013, pp. 27–45.

[78] S. Turner, ed., *CAS-CERN Accelerator School and Uppsala University: Course on advanced accelerator physics, Uppsala, Sweden, Sep 18-29 1989: Proceedings*, 1990.

[79] *LHC Design Report Vol.1: The LHC Main Ring*, tech. rep., 2004.

[80] F. S. J. Gomez-Cadenas, *NUFACT'08 School*, URL: http://benasque.org/2008nufact/, (accessed: 02.04.2020).

[81] A Verdier, "Chromaticity (particle accelerators)" (1995), URL: https://cds.cern.ch/record/302491.

[82] S. Guiducci, "Chromaticity", *CAS-CERN Accelerator School: 5th general accelerator physics course, Jyvaskyla, Finland, 7-18 Sep 1992: Proceedings. 2 vol.* 1992, pp. 191–206.

[83] B. P. Bielawski, "LHC point 4 cavern containing the RF equipment" (2017), General Photo, URL: https://cds.cern.ch/record/2293083.

[84] R. W. Assmann et al., *Summary of the CERN Meeting on Absorbers and Collimators for the LHC*, tech. rep., Geneva: CERN, 2002, URL: https://cds.cern.ch/record/691937.

[85] G. Gobbi et al., "Novel LHC collimator materials: High-energy Hadron beam impact tests and nondestructive postirradiation examination", *Mech. Adv. Mat. Struct.* (2019).

[86] M. Brice, "Damage of the LHC magnets in sector 3-4 of the LHC, provoked by the incident which happened on 19 September 2008. Dommages sur les aimants dans le secteur 3-4 du LHC, provoqués par l'incident survenu le 19 septembre 2008", 2008, URL: https://cds.cern.ch/record/1185822.

[87] B. Goddard et al., "A large diameter entrance window for the LHC beam dump line", *Conf. Proc. C* 0505161 (2005), p. 1698.

[88] A Yamamoto et al., "The ATLAS central solenoid", *Nucl. Instrum. Methods Phys. Res., A* 584 (2007), pp. 53–74, URL: https://cds.cern.ch/record/1069672.

[89] G. Aad et al., "The ATLAS Simulation Infrastructure", *Eur. Phys. J. C* 70 (2010), pp. 823–874.

[90] M. Aaboud et al., "Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2", *Eur. Phys. J. C* 77 (2017), p. 673.

[91] *ATLAS inner detector: Technical design report. Vol. 1*, tech. rep., 1997.

[92] "ATLAS inner detector: Technical design report. Vol. 2" (1997).

[93] M. Aaboud et al., "Study of the material of the ATLAS inner detector for Run 2 of the LHC", *JINST* 12 (2017), P12009.

[94] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, tech. rep., 2010.

[95] M. Alam et al., *ATLAS pixel detector: Technical design report*, tech. rep., 1998.

[96] A. Ahmad et al., "The Silicon microstrip sensors of the ATLAS semiconductor tracker", *Nucl. Instrum. Meth. A* 578 (2007), pp. 98–118.

[97] A. Abdesselam et al., "The barrel modules of the ATLAS semiconductor tracker", *Nucl. Instrum. Meth. A* 568 (2006), pp. 642–671.

[98] A. Abdesselam et al., "The ATLAS semiconductor tracker end-cap module", *Nucl. Instrum. Meth. A* 575 (2007), pp. 353–389.

[99] G. Lutz, *Semiconductor radiation detectors: device physics*, Berlin: Springer, 1999, URL: https://cds.cern.ch/record/411172.

[100] A. Bingul, "The ATLAS TRT and its performance at LHC", *J. Phys. Conf. Ser.* 347 (2012), p. 012025.

[101] E. Abat et al., "The ATLAS TRT barrel detector", *JINST* 3 (2008), P02014.

[102] E. Abat et al., "The ATLAS TRT end-cap detectors", *JINST* 3 (2008), P10003.

[103] "ATLAS liquid argon calorimeter: Technical design report" (1996).

[104] "ATLAS Liquid Argon Calorimeter 2m prototype. Module Zero du calorimètre à Argon liquide d'ATLAS", 1990, URL: https://cds.cern.ch/record/2254084.

[105] B. Aubert et al., "Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter", *Nucl. Instrum. Meth.* A558 (2006), pp. 388–418.

[106] "ATLAS tile calorimeter: Technical design report" (1996).

[107] S. Morgenstern, "ATLAS LAr calorimeter performance in LHC Run-2", *Nucl. Instrum. Meth. A* 936 (2019), ed. by G. Batignani et al., pp. 86–89.

[108] "ATLAS muon spectrometer: Technical design report" (1997).

[109] M. Bianco, "ATLAS RPC commissioning status and cosmic ray test results", *eConf* C080625 (2008), ed. by S. Ciprini, G. Mantovani, and G. Anzivino, p. 0035.

[110] A. Airapetian et al., "ATLAS: Detector and physics performance technical design report. Volume 1" (1999).

[111] E. Simioni et al., "Topological and Central Trigger Processor for 2014 LHC luminosities", 2012, URL: http://cdsweb.cern.ch/record/1459209/files/ATL-DAQ-PROC-2012-041.pdf.

[112] *Studies for the development of the Inner Detector trigger algorithms at ATLAS*, tech. rep., Geneva: CERN, 2013, URL: https://cds.cern.ch/record/1602918.

[113] O. Penc, "The performance and development for the Inner Detector Trigger algorithms at ATLAS", *J. Phys. Conf. Ser.* 608 (2015), p. 012008.

[114] O. Penc, "The design and performance of the ATLAS Inner Detector trigger for Run 2", *Proceedings, 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC 2015): San Diego, California, United States*, 2016, p. 7581784.

[115] R. Blair et al., "The ATLAS High Level Trigger Region of Interest Builder", *JINST* 3 (2008), P04001.

[116] *ATLAS high-level trigger, data acquisition and controls: Technical design report*, tech. rep., 2003.

[117] I. Bird et al., "Update of the Computing Models of the WLCG and the LHC Experiments" (2014).

[118] M. Aaboud et al., "Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC", *Eur. Phys. J.* C76 (2016), p. 653.

[119] P. Jenni et al., "ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity" (2008).

[120] P. Grafström and W. Kozanecki, "Luminosity determination at proton colliders", *Prog. Part. Nucl. Phys.* 81 (2015), pp. 97–148.

[121] G. Avoni et al., "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS", *JINST* 13 (2018), P07017.

[122] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, tech. rep., Geneva: CERN, 2019, URL: http://cds.cern.ch/record/2677054.

[123] *Luminosity public results - Run 2*, Accessed: 2020-07-15, Geneva, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[124] A. Buckley et al., "Implementation of the ATLAS Run 2 event data model", *J. Phys. Conf. Ser.* 664 (2015), p. 072045.

[125] J. Catmore et al., "A new petabyte-scale data derivation framework for ATLAS", *J. Phys. Conf. Ser.* 664 (2015), p. 072007.

[126] G. Aad et al., "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV", *Eur. Phys. J. C* 76 (2016), p. 292.

[127] M. Aaboud et al., "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* 79 (2019), p. 639.

[128] S. Albrand, J. Fulachier, and F. Lambert, "The ATLAS metadata interface", *J. Phys. Conf. Ser.* 219 (2010), ed. by J. Gruntorad and M. Lokajicek, p. 042030.

[129] B. Webber, "Monte Carlo Simulation of Hard Hadronic Processes", *Ann. Rev. Nucl. Part. Sci.* 36 (1986), pp. 253–286.

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

[131] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*, Cambridge University Press, 2005.

[132] D. E. Soper, "Parton distribution functions", *Nucl. Phys. B Proc. Suppl.* 53 (1997), ed. by C. Bernard et al., pp. 69–80.

[133] G. T. Bodwin, "Factorization of the Drell-Yan Cross-Section in Perturbation Theory", *Phys. Rev. D* 31 (1985), [Erratum: Phys.Rev.D 34, 3932 (1986)], p. 2616.

[134] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization for Short Distance Hadron - Hadron Scattering", *Nucl. Phys. B* 261 (1985), pp. 104–142.

[135] J. C. Collins, D. E. Soper, and G. F. Sterman, "Soft Gluons and Factorization", *Nucl. Phys. B* 308 (1988), pp. 833–856.

[136] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[137] V. Gribov and L. Lipatov, "Deep inelastic e p scattering in perturbation theory", *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[138] L. Lipatov, "The parton model and perturbation theory", *Sov. J. Nucl. Phys.* 20 (1975), pp. 94–102.

[139] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* 126 (1977), pp. 298–318.

[140] B. Andersson et al., "Parton Fragmentation and String Dynamics", *Phys. Rept.* 97 (1983), pp. 31–145.

[141] T. Sjostrand, "Jet Fragmentation of Nearby Partons", *Nucl. Phys. B* 248 (1984), pp. 469–502.

[142] J.-C. Winter, F. Krauss, and G. Soff, "A Modified cluster hadronization model", *Eur. Phys. J. C* 36 (2004), pp. 381–395.

[143] A. Kupco, "Cluster hadronization in HERWIG 5.9", *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, 1998.

[144] R. Field and R. Feynman, "Quark Elastic Scattering as a Source of High Transverse Momentum Mesons", *Phys. Rev. D* 15 (1977), pp. 2590–2616.

[145] A. Denner et al., "NLO QCD and EW corrections to vector-boson scattering into ZZ at the LHC", *JHEP* 11 (2020), p. 110.

[146] A. Buckley et al., "Monte Carlo event generator validation and tuning for the LHC", *PoS* ACAT08 (2008), ed. by T. Speer, F. Carminati, and M. Werlen, p. 112.

[147] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys. Rev. D* 82 (2010), p. 074018.

[148] P. Z. Skands, "The Perugia Tunes", *1st International Workshop on Multiple Partonic Interactions at the LHC*, 2009, pp. 284–297.

[149] F. Siegert, "Monte-Carlo event generation for the LHC", PhD thesis, Durham U., 2010.

[150] S. Agostinelli et al., "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303.

[151] R. D. Ball et al., "Parton distributions for the LHC Run II", *JHEP* 04 (2015), p. 040.

[152] F. Caola et al., "QCD corrections to ZZ production in gluon fusion at the LHC", *Phys. Rev. D* 92 (2015), p. 094028.

[153] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* 07 (2002), p. 012.

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

[155] "Modelling of the vector boson scattering process $pp \to W^{\pm}W^{\pm}jj$ in Monte Carlo generators in ATLAS" (2019).

[156] M. Tanabashi et al., "Review of Particle Physics", *Phys. Rev. D* 98 (2018), p. 030001.

[157] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* 04 (2008), p. 063.

[158] M. Aaboud et al., "Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data", *JINST* 14 (2019), P03017.

[159] T. Hryn'ova and K. Nagano, *Trigger Menu Strategy for Run 2*, tech. rep., Geneva: CERN, 2014, URL: https://cds.cern.ch/record/1703730.

[160] T. Berger-Hryn'ova et al., *Physics Uses and Hardware Constraints of the L1 Topological Trigger*, tech. rep., Geneva: CERN, 2014, URL: https://cds.cern.ch/record/1645921.

[161] G. Aad et al., "Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector" (2020).

[162] "Tagging and suppression of pileup jets" (2014).

[163] A. Hocker et al., "TMVA - Toolkit for Multivariate Data Analysis" (2007).

[164] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[165] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *J. Comput. Syst. Sci.* 55 (1997), pp. 119–139.

[166] J. H. Friedman, "Stochastic gradient boosting", *Comput. Stat. Data Anal.* 38 (2002), pp. 367–378.

[167] M. Becker, "Measurement of ZZ production and search for anomalous neutral triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment", PhD thesis, Mainz U., 2018.

[168] A. Collaboration, *Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC*, tech. rep., Paper / Conf. Note Draft for STDM-2017-19, Geneva: CERN, 2019, URL: https://cds.cern.ch/record/2679741.

[169] M. Aaboud et al., "Measurement of the $W^{\pm}Z$ boson pair-production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector", *Phys. Lett. B* 762 (2016), pp. 1–22.

[170] "Jet energy scale and its systematic uncertainty in proton-proton collisions at sqrt(s)=7 TeV with ATLAS 2011 data" (2013).

[171] "Probing the measurement of jet energies with the ATLAS detector using Z+jet events from proton-proton collisions at sqrt(s) = 7 TeV" (2012).

[172] "Light-quark and Gluon Jets in ATLAS" (2011).

[173] "Close-by Jet Effects on Jet Energy Scale Calibration in pp Collisions at sqrt(s)=7 TeV with the ATLAS Detector" (2011).

[174] M. Aaboud et al., "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* 96 (2017), p. 072002.

[175] P. Bagnaia et al., "Measurement of Production and Properties of Jets at the CERN anti-p p Collider", *Z. Phys. C* 20 (1983), pp. 117–134.

[176] B. Abbott et al., "High-p_T jets in $\bar{p}p$ collisions at $\sqrt{s} = 630$ GeV and 1800 GeV", *Phys. Rev. D* 64 (2001), p. 032003.

[177] W. Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*, tech. rep., Geneva: CERN, 2015, URL: https://cds.cern.ch/record/2014726.

[178] M. Aaboud et al., "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.* 117 (2016), p. 182002.

[179] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys. G* 43 (2016), p. 023001.

[180] C. Patrignani et al., "Review of Particle Physics", *Chin. Phys. C* 40 (2016), p. 100001.

[181] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* 75 (2015), p. 132.

[182] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev. D* 82 (2010), p. 074024.

[183] L. A. Harland-Lang et al., "Parton distributions in the LHC era: MMHT 2014 PDFs", *Eur. Phys. J. C* 75 (2015), p. 204.

[184] S. Dulat et al., "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev. D* 93 (2016), p. 033006.

[185] R. D. Cousins, J. T. Linnemann, and J. Tucker, "Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process", *Nucl. Instrum. Meth. A* 595 (2008), pp. 480–501.

[186] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* 71 (2011), [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554.

[187] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth. A* 389 (1997), ed. by M. Werlen and D. Perret-Gallix, pp. 81–86.

[188] "Observation of electroweak production of two jets in association with a Z-boson pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector" (2019).